



Thèse

2011

Open Access

This version of the publication is provided by the author(s) and made available in accordance with the copyright holder(s).

In silico pharmacology in drug discovery: methods for binding affinity prediction, identifying protein-protein interaction breakers, and finding targets for small molecules

Westermaier, Yvonne

How to cite

WESTERMAIER, Yvonne. In silico pharmacology in drug discovery: methods for binding affinity prediction, identifying protein-protein interaction breakers, and finding targets for small molecules. Doctoral Thesis, 2011. doi: 10.13097/archive-ouverte/unige:136333

This publication URL: <https://archive-ouverte.unige.ch/unige:136333>

Publication DOI: [10.13097/archive-ouverte/unige:136333](https://doi.org/10.13097/archive-ouverte/unige:136333)

UNIVERSITÉ DE GENÈVE

Section des Sciences Pharmaceutiques

Laboratoire de Biochimie Pharmaceutique

FACULTÉ DES SCIENCES

Professeur Leonardo Scapozza

***In Silico* Pharmacology in Drug Discovery:
Methods for Binding Affinity Prediction, Identifying
Protein-Protein Interaction Breakers, and Finding
Targets for Small Molecules**

THÈSE

présentée à la Faculté des sciences de l'Université de Genève
pour obtenir le grade de Docteur ès sciences, mention interdisciplinaire

par

Yvonne Westermaier

d'Urnäsch (AR)

et

de Munich (Allemagne)

Thèse N° 4295

GENÈVE

2011



**UNIVERSITÉ
DE GENÈVE**

FACULTÉ DES SCIENCES

**Doctorat ès sciences
Mention interdisciplinaire**

Thèse de *Madame Yvonne WESTERMAIER*

intitulée :

**"In silico Pharmacology in Drug Discovery:
Methods for Binding Affinity Prediction,
Identifying Protein-Protein Interaction Breakers, and
Finding Targets for Small Molecules"**

La Faculté des sciences, sur le préavis de Messieurs L. SCAPOZZA, professeur ordinaire et directeur de thèse (Section des sciences pharmaceutiques), S. MORO, professeur associé (Molecular Modeling Section, Università degli Studi di Padova, Italia), D. ROGNAN, docteur (Structural Chemogenomics Group, Laboratory of Therapeutic Innovation, Centre National de la Recherche Scientifique, Université Louis Pasteur, Strasbourg, France) et G. SCHNEIDER, professeur (Institute of Pharmaceutical Sciences, Eidgenössische Technische Hochschule Zürich, Suisse), autorise l'impression de la présente thèse, sans exprimer d'opinion sur les propositions qui y sont énoncées.

Genève, le 10 mars 2011

Thèse 4295 -

Le Doyen, Jean-Marc TRISCONE

N.B. - La thèse doit porter la déclaration précédente et remplir les conditions énumérées dans les "Informations relatives aux thèses de doctorat à l'Université de Genève".

Dedicated to

My parents and my sister Flavia

"If we were to name the most powerful assumption of all, which leads one on and on in an attempt to understand life, it is that all things are made of atoms, and that everything that living things do can be understood in terms of the jiggings and wiggings of atoms."

Dr. Richard Feynman, American physicist, 1963

CURRICULUM VITAE

Yvonne Barbara Annina WESTERMAIER

Born on the 16th of December 1980 in Lucerne (CH)

Nationalities: Swiss and German

Permanent address: Via Vitgé 10, 7017 Flims (CH)

Phone: +41 081 911 44 84

e-mail: yvonne.westermaier@gmail.com

EDUCATION

- 2011: Dr ès sciences, mention interdisciplinaire, Universités de Genève et Lausanne (CH)
- 2005: Diplôme d'Etudes Avancées (DEA) en biologie moléculaire et cellulaire, mention biotechnologie, option bioinformatique, Université Louis Pasteur I, Strasbourg (FR)
- 2005: Trinational Bioengineer's degree in Biotechnology (FR/DE/CH)
- 2001-2002: 1. und 2. Vordiplom in Biologie, Universität Basel (CH)
- 2000: Swiss Matura with Ancient Greek, Latin and History of Art, Klosterschule Disentis, Disentis (CH)

PUBLICATIONS

Westermaier Y, Barril X, and Scapozza L (2015): *Virtual screening: An in silico tool for interlacing the chemical universe with the proteome*. Methods. 71, 44-57.

Westermaier Y, Veurink M, Riis-Johannessen T, Guinchard S, Gurny R, and Scapozza L (2013): *Identification of aggregation breakers for bevacizumab (Avastin®) self-association through similarity searching and interaction studies*. Eur J Pharm Biopharm. 85, 773-80.

Veurink M, **Westermaier Y**, Gurny R, and Scapozza L (2013): *Breaking the aggregation of the monoclonal antibody bevacizumab (Avastin®) by dexamethasone phosphate: insights from molecular modelling and asymmetrical flow field-flow fractionation*. Pharm Res. 30, 1176-87.

Martic M, Pernot L, **Westermaier Y**, Perozzo R, Kraljevic TG, Kristafor S, Raic-Malic S, Scapozza L, and Ametamey S (2011): *Synthesis, crystal structure, and in vitro biological evaluation of C-6 pyrimidine derivatives: new lead structures for monitoring gene expression in vivo*. Nucleosides Nucleotides Nucleic Acids. 30, 293-315.

Kristafor S, Novakovic I, Gazivoda Kraljevic T, Kraljevic Pavelic S, Lucin P, **Westermaier Y**, Pernot L, Scapozza L, Ametamey SM, and Raic-Malic S (2011): *A new N-methyl thymine derivative comprising a dihydroxyisobutenyl unit as ligand for thymidine kinase of herpes simplex virus type 1 (HSV-1 TK)*. Bioorg Med Chem Lett. 21, 6161-5.

Jaquet V, Marcoux J, Forest E, Leidal KG, McCormick S, **Westermaier Y**, Perozzo R, Plastre O, Fioraso-Cartier L, Diebold B, Scapozza L, Nauseef WM, Fieschi F, Krause KH, and Bedard K (2011): *NADPH oxidase (NOX) isoforms are inhibited by celastrol with a dual mode of action*. Br J Pharmacol. 164, 507-20.

Westermaier Y. Meeting report for Thomson Reuters Scientific: *New technologies: Nanomedicine, proteomics and drug targeting*. Frontiers in Medicinal Chemistry Meeting 2009, Joint-German Swiss Meeting on Medicinal Chemistry, Heidelberg (DE).

PATENTS

Gurny R, Scapozza L, **Westermaier Y**, and Veurink M. *Stabilized intact antibody formulations, related methods and uses thereof*. WO 2011/121559 A2.

Gurny R, Scapozza L, **Westermaier Y**, and Veurink M. *Stabilized intact antibody formulations, related methods and uses thereof*. WO 2011/121560 A2.

PRESENTATIONS

Y. Westermaier and L. Scapozza: *A combined in silico-in vitro fishing for intracellular targets of phenazopyridine*. Swiss Chemical Society (SCS) Fall Meeting 2010, Swiss Federal Institute of Technology Zurich (ETHZ), 16th of September 2010, Zurich (CH).

Y. Westermaier and L. Scapozza: *The role of cheminformatics in target finding*. 24^{ème} séminaire en sciences pharmaceutiques, 3^{ème} cycle en sciences pharmaceutiques, "Frontiers in drug discovery & development", 8th of September 2009, Zermatt (CH).

Y. Westermaier and L. Scapozza: *Target identification and PET-tracer design by means of computational approaches*. PhD day of the School of Pharmaceutical Sciences Geneva-Lausanne, 23rd of June 2009, Hermance (CH).

Y. Westermaier and L. Scapozza: *In silico target discovery in neural stem cell differentiation*. symposium of the Società Ticinese delle Scienze Biomediche e Chimiche (STSBC), Meeting of the Union of the Swiss Societies for Experimental Biology, 6th of February 2008, Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne (CH).

Y. Westermaier and L. Scapozza: *Homology modelling of Nox2*. Institut de Biologie Structurale, Université Joseph-Fourier, 1 October 2007, Grenoble (FR).

POSTERS

Y. Westermaier, S. Ametamey, and L. Scapozza: *In silico prediction of PET tracer affinity using molecular dynamics*. MIPTec 2009, Basel (CH).

Y. Westermaier, S. Ametamey, and L. Scapozza: *Insights into the dynamics of type I and type II thymidine kinases*. Swiss Chemical Society (SCS) Fall Meeting 2009, Swiss Federal Institute of Technology Lausanne, Lausanne (CH).

Y. Westermaier, S. Ametamey, and L. Scapozza: *Towards a better understanding of PET tracer affinity using molecular dynamics*. 2nd Swiss Pharma Science Day, 2009, Berne (CH) and 24^{ème} séminaire en sciences pharmaceutiques. "Frontiers in drug discovery and development", 2009, Zermatt (CH).

Y. Westermaier and L. Scapozza: *Dynamics of HSV-1 TK applied to PET tracer design for HSV-1 TK-based gene therapy*. Frontiers in Medicinal Chemistry 2009, Joint German-Swiss Meeting on Medicinal Chemistry, Heidelberg (DE).

Y. Westermaier, L. Pernot, M. Martic, S. Ametamey, and L. Scapozza: *Designing PET tracers for HSV-1 TK-based gene therapy using molecular dynamics*. 1st Swiss Pharma Science Day, Zurich (CH).

Y. Westermaier, L. Pernot, M. Martic, S. Ametamey, and L. Scapozza: *Designing PET tracers for HSV-1 TK-based gene therapy using molecular dynamics*. Fall Meeting of the Swiss Chemical Society (SCS), University of Zurich, Zurich (CH).

Y. Westermaier, L. Pernot, M. Martic, S. Ametamey, and L. Scapozza: *Molecular dynamics as a means to designing PET tracers for HSV-1 TK-based gene therapy*. XXth International Symposium on Medicinal Chemistry of the European Federation of Medicinal Chemistry (EFMC-ISMIC) 2008, Vienna (AT).

Y. Westermaier, D. Vitel, D. Suter, K.-H. Krause, and L. Scapozza: *In silico target discovery in neural stem cell differentiation*. Meeting of the Union of the Swiss Societies for Experimental Biology (USGEB) 2008, Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne (CH).

L. Pernot, **Y. Westermaier**, R. Perozzo, M. Martic, P.A. Schubiger, S. Ametamey, and L. Scapozza: *Designing PET tracers for HSV-1 TK-based gene therapy using structural biology and molecular dynamics*. Meeting of the Swiss Crystallographic Society, Zurich (CH).

V. Jaquet, K. Bedard, **Y. Westermaier**, L. Fioraso-Cartier, B. Mottironi, L. Scapozza, and K.-H. Krause: *Celastrol, a novel inhibitor of NOX enzymes*. Gordon Research Conference: NOX Family NADPH Oxidases, New London (NH, USA).

Y. Westermaier, D. Vitel, D. Suter, K.-H. Krause, and L. Scapozza: *In silico target finding in neural stem cell differentiation: Protocol validation*. 22e séminaire en sciences pharmaceutiques, 3e cycle en sciences pharmaceutiques, Zermatt (CH).

Y. Westermaier, D. Vitel, D. M. Suter, K.-H. Krause, and L. Scapozza: *Validation of inverse virtual screening for target finding in neural stem cell differentiation*. Annual Meeting of the Swiss Society of Pharmacology and Toxicology (SSPT), Bern (CH).

Y. Westermaier, V. Jaquet, K.-H. Krause, L. Scapozza: *Virtual Screening on the NADPH oxidase subunit p47^{phox}: Identification of potential protein-protein interaction breakers*. 21e séminaire en sciences pharmaceutiques, 3e cycle en sciences pharmaceutiques, Zermatt (CH).

AWARDS

- | | |
|-------|--|
| 2010: | 2 nd best oral presentation in the Division of Medicinal Chemistry at the Swiss Chemical Society Fall Meeting 2010, Zurich (CH) |
| 2008: | Best Poster Award, Swiss Society of Oncology, Meeting of the Union of the Swiss Societies for Experimental Biology (USGEB) 2008, Lausanne (CH) |
| 2007: | 3 rd Novartis Poster Price, Swiss Society of Pharmacology and Toxicology (SSPT) Annual Meeting, Berne (CH) |

RÉSUMÉ

Plusieurs aspects de la conception de médicaments assistée par ordinateur sont considérés dans la présente thèse. Les approches et méthodes utilisées peuvent être rangées sous l'appellation générique "pharmacologie *in silico*", puisqu'elles participent de l'étude de l'action d'agents médicamenteux, au niveau moléculaire, par ordinateur. L'introduction donne un aperçu de la prédiction *in silico* d'affinité de liaison, de l'étude des interactions protéine-protéine, et de la chimio-informatique.

Dans le premier chapitre, la technique dénommée « steered molecular dynamics (SMD) » est appliquée au calcul d'affinité de substrats micromolaires de la thymidine kinase 1 du virus de l'herpès simplex (HSV-1 TK). Le but de cette technique est d'aider à la conception de traceurs de tomographie par émission de positron (TEP) basés sur la HSV-1 TK. De tels composés pourraient être utilisés dans le suivi de l'activité de l'enzyme en thérapie cellulaire ou dans des techniques d'imagerie appliquées aux tumeurs. En analogie avec les expériences de l'extraction d'un atome par microscopie à force atomique, la SMD peut être utilisée pour extraire une molécule d'un site actif d'une enzyme. Ainsi, le travail appliqué est proportionnel à l'affinité de liaison. Suivant cette méthode, il est démontré pour la première fois que des substrats porteurs de motifs pyrimidiques ayant des affinités connues peuvent être distingués les uns des autres. En effet, des différences d'affinité aussi faibles que 4.3 μM peuvent être reproduites. De plus, l'affinité de substrats analogues (fluorés) peut être prédite. Les mouvements majeurs de structures cristallines de la HSV-1 TK liée à différents substrats sont observés pendant une dynamique moléculaire classique d'une durée de 5 ns et analysés par analyse en composantes principales (ACP). Il est alors révélé que le cœur de l'enzyme comprenant les résidus catalytiques reste stable. Le domaine LID, qui se rabat sur le substrat après son entrée dans le site de liaison, est au contraire flexible et montre un mouvement caractéristique dans 2 dimensions, ce qui est en faveur d'une ouverture en charnière, codée dans la structure fermée.

Dans le champ de l'étude de la stabilisation des anticorps monoclonaux thérapeutiques, le deuxième chapitre est consacré à la découverte *in silico* et à la validation *in vitro* de molécules interférant avec des phénomènes d'agrégation. Elles ont fait l'objet deux brevets (WO 2011/121559 A2 and WO 2011/121560 A2). Ce travail se base sur une demande provisoire de brevet antérieure (US 9155745 B2), qui indique que les corticostéroïdes dexaméthasone phosphate et bétaméthasone phosphate stabilisent l'anticorps bévacizumab en réduisant sa propension à former des dimères ou encore en inversant les agrégats déjà formés pour les ramener essentiellement à l'état monomérique. Selon une modélisation par homologie, deux modèles de bévacizumab ont été construits: un modèle monomérique et un modèle d'agrégation. Ce dernier correspond à une forme dimérisée. Bévacizumab est un anticorps monoclonal humanisé composé de deux régions Fab responsables de son activité et d'une région Fc dérivée d'IgG1. Le modèle d'agrégation a révélé une région de contact unique entre le Fc d'un monomère et une région Fab du monomère adjacent, laquelle est présente dans des anticorps sur le marché comme trastuzumab (Herceptin®) et rituximab (MabThera®). Différents stéroïdes ont été « dockés » sur la région Fc du monomère. Leur capacité à interférer avec les interactions entre monomères a été vérifiée sur le modèle d'agrégation et validée expérimentalement par AFFF-MALS (Asymmetric Field-Flow Fractionation coupled to Multi-Angle Light Scattering). Les stéroïdes prédits et validés incluent la bétaméthasone phosphate, en tant qu'interférant avec l'agrégation, et la triamcinolone acétonide phosphate, comme molécule sans effet sur l'agrégation. Conçue d'après « scaffold hopping » à partir de molécules connues pour interférer avec l'agrégation, l'AMP a été obtenue comme composé prometteur pour présenter également de telles propriétés d'interférence, ce qui a été confirmé par docking et par validation expérimentale. Puisque la région d'interaction primaire est conservée parmi les anticorps du type IgG1, cette molécule pourrait être appliquée de manière plus générale afin de prévenir l'agrégation d'anticorps.

Le troisième chapitre se focalise sur l'identification d'inhibiteurs - de faible poids moléculaire ou de nature peptidique - d'isoformes de la NADPH oxydase (NOX). Dans la partie centrale du chapitre, une interface spécifique est étudiée: celle d'interaction protéine-protéine cristallisée, importante dans l'assemblage et l'activation phagocytaire de NOX2. Cette interface n'est pour le moment ciblée par aucun inhibiteur connu. 43 inhibiteurs potentiels de faible poids moléculaire ont été identifiés en combinant une investigation guidée par pharmacophores avec un criblage virtuel basé sur une structure cristalline. Des modes de liaisons potentiels ont été suggérés pour un lead expérimental de faible poids moléculaire (célastrol). Des mesures de fluorescence intrinsèque ainsi que la spectrométrie de masse à conditions natives ont confirmé que célastrol inhibe l'interaction protéine-

protéine ciblée. Dans le but d'améliorer les affinités de liaison de peptidomimétiques ciblant la même interface protéine-protéine, une analyse d'interaction basée sur GRID a été conduite afin d'identifier des substituants favorables, combinant des acides aminés naturels avec des acides aminés non naturels. Les peptides prédits *in silico* ont des valeurs de CI50 comparables au peptide natif. Par ailleurs, un mode de liaison a été prédit pour le rétropeptide du peptide natif.

Des données expérimentales suggèrent que la phénazopyridine induit la différenciation homogène de cellules souches embryonnaires du type CGR8 en cellules neuronales. A partir de ce constat, le chapitre 4 traite de l'identification de cibles de cette molécule, à partir de relations structures-activités ou encore par criblage *in silico* inverse. Ces investigations viennent compléter des tests d'inhibition fonctionnelle et des tests de liaisons compétitives. Après une validation non biaisée et une validation basée sur les connaissances disponibles, un seuil a été défini à 10% du classement par score d'interaction comme critère à l'élimination de cibles non spécifiques. Ensuite, la phénazopyridine a été « dockée » dans plus de 5 000 structures cristallines de protéines, le docking du ligand co-cristallisé dans chaque protéine servant de contrôle. Les critères de sélection comprenaient: (i) le classement par énergie d'interaction de la phénazopyridine, (ii) l'écart quadratique moyen du ligand de contrôle, (iii) l'existence d'interactions cibles clés communes à la phénazopyridine et au ligand de contrôle ainsi que (iv) des résultats de publications détaillant le contexte biologique. Des tests fonctionnels d'inhibition et des tests de liaison compétitive ont servi de validation expérimentale. En appliquant trois des quatre critères de sélection au criblage virtuel, 80% des cibles identifiées expérimentalement ont été trouvés, générant de cette façon un premier profil de sélectivité pour la phénazopyridine.

En conclusion, l'itération de la modélisation moléculaire et de la validation expérimentale a prouvé son utilité dans l'obtention des résultats décrits dans chaque chapitre de la présente thèse. Elle constitue de ce fait une méthode digne d'intérêt sur un plan général.

SUMMARY

The present thesis addresses different areas of computational drug design. The approaches used can be summarized as “*in silico* pharmacology”, because they are all related to studying drug action computationally. The introduction provides an overview on computational binding affinity prediction, protein-protein interactions and cheminformatics.

In chapter 1, steered molecular dynamics (SMD) is applied to the affinity calculation of μM substrates of the *Herpes Simplex Virus 1* Thymidine Kinase (HSV-1 TK) to foster the design of HSV-1 TK-based Positron Emission Tomography (PET) tracers to be used in monitoring TK activity in stem cell therapy or cancer imaging. In analogy to the single atom pulling experiments in atomic force microscopy, SMD can be used to pull out a molecule from an enzyme's active site, with the applied work being proportional to the binding affinity. It is shown for the first time that pyrimidine-based substrates with known substrate affinities can be distinguished, that affinity differences as small as 4.3 μM can be reproduced and that the affinity of (fluorinated) substrate analogues can be predicted using this method. Main motions of holo crystal structures of HSV-1 TK during 5 ns classical molecular dynamics are analyzed by a principal component analysis, revealing that the enzyme core harbouring the residues involved in catalysis stays stable. In contrast, the LID domain (closing over the substrate upon its entry) is flexible and shows a distinct movement in two directions, which suggests that the hinge-like opening is “encoded” in the closed holo structure.

In the context of stabilizing therapeutic monoclonal antibodies, chapter 2 deals with the *in silico* discovery and *in vitro* validation of small molecule aggregation breakers which have been recently patented (WO 2011/121559 A2 and WO 2011/121560 A2) and is based on a previous patent application (US 9155745 B2) stating that the corticosteroids dexamethasone phosphate and betamethasone phosphate stabilize the antibody bevacizumab by reducing its propensity to form dimers or reverting already formed aggregates into an essentially monomeric state. Using homology modelling, a monomer and an aggregation model of bevacizumab were built, the latter representing a dimer. Bevacizumab is a humanized monoclonal antibody formed by two Fab regions responsible for its activity and one Fc region derived from IgG1. The aggregation model revealed a unique contact region between the Fc of one monomer and one Fab of the adjacent monomer, which is conserved in marketed IgG1 antibodies like trastuzumab (Herceptin®) and rituximab (MabThera®). Different

steroids were docked onto the monomer Fc region, checked for their propensity of hindering the contact between monomers in the aggregation model and experimentally validated by Asymmetric Field-Flow Fractionation coupled to Multi-Angle Light Scattering (AFFF-MALS). The predicted and validated steroids include betamethasone phosphate as an aggregation breaker and triamcinolone acetonide phosphate as a compound with no effect on aggregation. Based on a scaffold hopping approach with aggregation breaking molecules, AMP was obtained as a hit and confirmed by both docking and experimental validation to be an aggregation breaker. Because of the conserved primary aggregation region in IgG1 antibodies, this molecule could be broadly applicable to current antibody aggregation issues.

Chapter 3 focuses on identifying small molecular weight and peptide inhibitors of NADPH oxidase (NOX) isoforms. In the main part of this chapter, a crystallized protein-protein interaction interface of importance in the assembly and activation of the phagocyte NOX2 is assessed. This interface is so far not targeted by known inhibitors. 43 potential small molecular weight inhibitors were identified by combining a pharmacophore search with structure-based virtual screening. A potential binding orientation was suggested for an experimental small molecular hit (celastrol). Subsequent intrinsic fluorescence measurements and native mass spectrometry confirmed that celastrol is inhibiting the targeted protein-protein interaction. To improve the binding affinities for peptidomimetics targeting the same protein-protein interface, a GRID-based interaction analysis was carried out to spot favourable substitutions of natural with non-natural amino acids. Remarkably, the *in vitro* predicted peptides had IC₅₀ values comparable to the native peptide. Besides, a putative binding mode of the retropeptide of the wild type peptide was predicted.

Based on experimental data suggesting that phenazopyridine induces the homogeneous differentiation of embryonic CGR8 stem cells into neuronal cells and the need for target finding, structure-activity relationships (SARs) and inverse *in silico* screening were applied as complements to functional inhibition tests and competitive binding assays in chapter 4. After an unbiased and a knowledge-based docking validation, a threshold at 10% of the score-based ranking could be defined as criterion to eliminate unspecific targets. Subsequently, phenazopyridine was docked into over 5000 protein crystal structures. Docking the corresponding co-crystallized ligand into each protein served as a control. Selection criteria included (i) the interaction energy-based ranking of phenazopyridine, (ii) the control ligand RMSD, (iii) shared key target interactions between phenazopyridine and the control as well as (iv) publications reporting on the biological context. Experimental validation included functional inhibition tests and competitive binding assays. By applying three of the four selection criteria to *in silico* screening, 80% of the experimentally identified targets were retrieved, providing a rough selectivity profile of phenazopyridine.

Iterating molecular modelling and experimental validation has proven to be beneficial for the outcome of each chapter and should generally be applied as often as possible.

TABLE OF CONTENTS

Curriculums vitae	4
Résumé.....	7
Summary.....	9
Introduction - Strengths and pitfalls of computational methods addressing binding affinity prediction, protein-protein interactions and cheminformatics	11
0.1 Computational binding affinity prediction	11
0.2 Protein-protein interactions.....	30
0.3 Cheminformatics	38
0.4 References.....	45
Chapter 1 – <i>In silico</i> determination of micromolar substrate affinities using steered molecular dynamics	61
Abstract	61
1.1 Introduction.....	61
1.2 Methods.....	68
1.3 Results and discussion.....	71
1.4 Conclusion.....	90
1.5 Acknowledgements	91
1.6 References.....	92
Chapter 2 – Stabilization of therapeutic monoclonal antibodies using small molecule aggregation breakers	100
Abstract	100
2.1 Background	100
2.2 Methods.....	104
2.3 Results and discussion.....	109
2.4 Conclusions.....	127
2.5 Acknowledgements	127
2.6 References.....	128
Chapter 3 - NADPH oxidases	130
Abstract	130
3.1 Introduction to the biological context	130
3.2 Targeting SH3 domains.....	137
3.3 The importance of proline.....	140
3.4 Small molecule NOX inhibitors	140
3.5 Peptidomimetics	142
3.6 Targeting the p47 ^{phox} -p22 ^{phox} interaction	144
3.7 Methods.....	146
3.8 Results and discussion.....	150
3.9 Conclusions.....	167
3.10 Acknowledgements	167
3.11 References.....	168
Chapter 4 - <i>In silico</i> screening for targets of phenazopyridine: state of the art, validation and optimization of scoring evaluation/target selection	179
Abstract	179
Introduction.....	179
4.1 Methods.....	183
4.2 Results and discussion.....	190
4.3 Extension of the analysis.....	206
4.4 Experimental validation of kinases	207
4.5 Conclusions.....	207
4.5 Acknowledgements	209
4.6 References.....	209
Chapter 5 - Conclusions	217
Acknowledgements	218
Abbreviations used in this thesis	219
Annex.....	219

INTRODUCTION - STRENGTHS AND PITFALLS OF COMPUTATIONAL METHODS ADDRESSING BINDING AFFINITY PREDICTION, PROTEIN-PROTEIN INTERACTIONS AND CHEMINFORMATICS

The present thesis covers four main projects, each of them covering different approaches of computational drug design involving small organic molecules, peptides, and proteins (including enzymes and antibodies).

The introduction aims at discussing the advantages and limitations of these approaches. In its first part, the current status of *in silico* binding affinity predictions is reviewed as a complement to chapter 1, where the substrate-binding strength of a human and a viral thymidine kinase are assessed in view of developing substrate analogues as Positron Emission Tomography (PET) tracers.

The second part provides an overview of how protein-protein interactions can be broken and is thought as an introduction to chapters 2 and 3. In chapter 2, antibody aggregation is blocked with small molecules, while in chapter 3, the assembly and activation of the NADPH oxidase 2 (NOX2) is inhibited by peptides or a small molecule.

Finally, the third part of the introduction concentrates on ligand- and target-based cheminformatics approaches, interconnecting chapters 2 to 4, the latter being dedicated to *in silico* target discovery for a small molecule with a known *in vitro* effect on stem cell differentiation.

0.1 COMPUTATIONAL BINDING AFFINITY PREDICTION

In silico affinity prediction constitutes one of the challenges of computational drug design. The introduction to this thesis aims at giving a brief overview of the status of the field, highlighting strengths and pitfalls of cutting-edge affinity prediction methods. By extending an exhaustive review provided by Gohlke and Klebe ¹, the subjects of molecular recognition, intra- and intermolecular interactions, and computational binding affinity prediction methods are treated. Furthermore, a short excursus on complementary experimental affinity determination methods is provided.

0.1.1 MOLECULAR RECOGNITION AS A STARTING POINT FOR DRUG DISCOVERY

The concept of molecular recognition as the basis of signal transduction goes back to several pioneers:

- a) E. Fischer (1852-1919), known for his "lock-and-key" principle ² on shape or chemical complementarities between an enzyme and a ligand in enzymatic reactions (Figure 1).

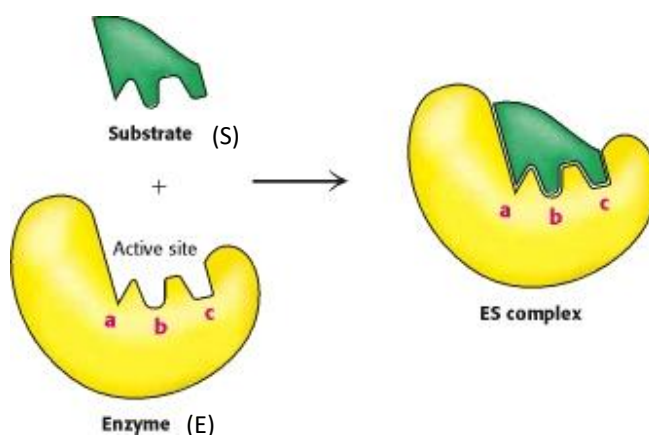


Figure 1: The lock-and-key principle. An enzyme and a substrate share shape or chemical complementarities and fit like a lock and a key.

- b) P. Ehrlich (1854-1915), who stated that "*corpora non agunt nisi fixata*" ³, which could be interpreted as binding and electrostatic surface complementarities being a prerequisite for drug action.
- c) D. E. Koshland (1920-2007), the inventor of the "induced fit" model (1958) ⁴, which implicates a structural adaptation of both a substrate and an enzyme as a precondition for binding.

- d) L. Pauling (1901-1994), who found that enzymes are complementary to the transition state of a substrate, which leads to a decrease in activation energy and thus an increased reaction rate ⁵. This finding refines the “lock-and-key” principle in the sense that the transition state of the substrate, and not its ground state, has the highest affinity to an enzyme.

0.1.1.1 3D STRUCTURES

With the development of methods for the 3D structure determination of macromolecules, M. F. Perutz ⁶ and J. C. Kendrew ⁷ contributed widely to the understanding of molecular recognition, making comparative analyses of protein-ligand complexes with affinity data ^{8, 9} possible. Nowadays, mainly structure-determination techniques such as X-ray crystallography or Nuclear Magnetic Resonance (NMR) spectroscopy are used. The public domain structure repository Protein Data Bank (PDB) ^{10, 11} undergoes currently an exponential growth (Figure 2) and contains mainly structures of proteins and/or nucleic acids.

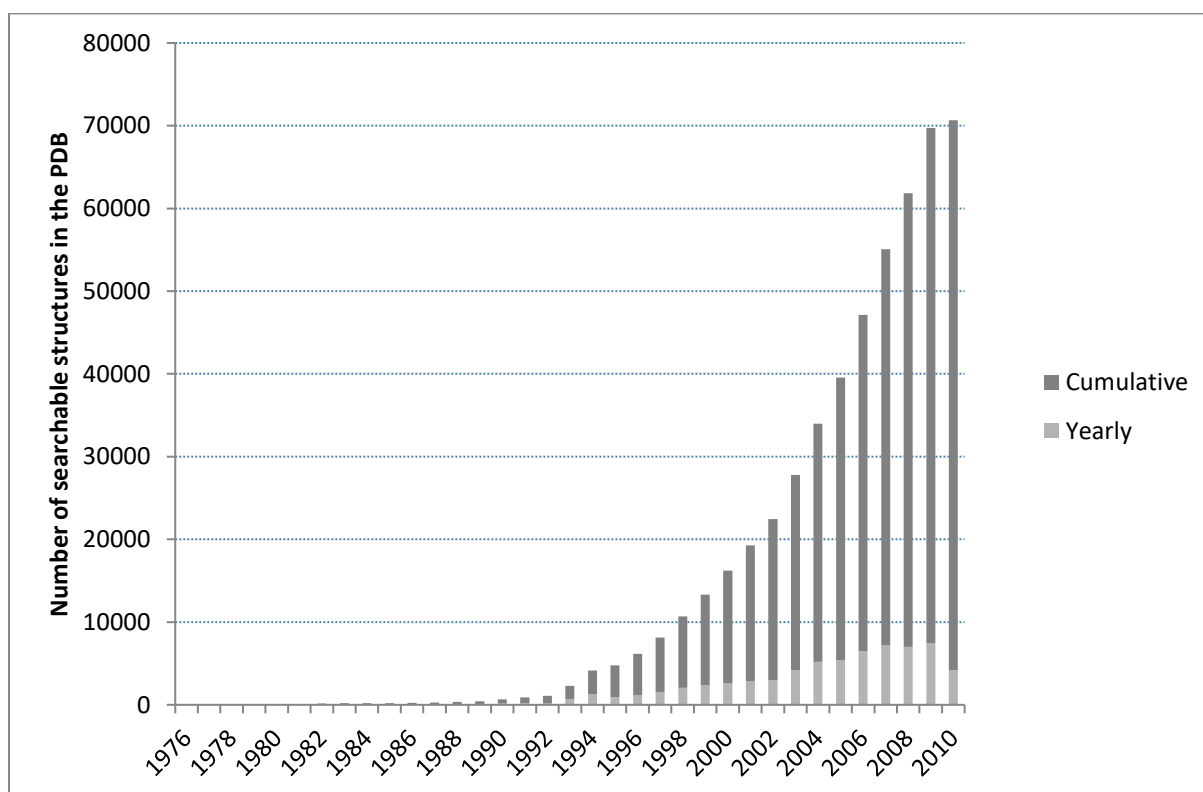


Figure 2: The exponential growth of the Protein Data Bank (PDB) since the deposition of the first structures in 1976. The searchable structures over time vary according to structure depositions and obsolete structure removals. In this stacked 2D graph, the cumulative number of structures is reported in dark grey and the yearly “newcomers” are overlaid in light grey on the former. The numbers for 2010 include structures released as of the 20th of July 2010. Data were obtained from <http://www.pdb.org/pdb/statistics/contentGrowthChart.do?content=total&seqid=100>.

Using X-ray protein-ligand complexes as a reference for drug design efforts, one needs to keep in mind that the qualities of the structures are related to their accuracies. The accuracy depends upon the reliability factor (R-factor), which indicates the standard deviation of the atom positions between the refined crystallographic model and the experimental X-ray diffraction data, and upon the resolution, which is the highest resolvable peak in the diffraction pattern. Resolution values between 1.8 and 3 Å are usual, allowing the determination of all atoms except the H-atoms. An atomic resolution, including the determination of H-positions, would correspond to < 1.2 Å. Knowing that oxygen and nitrogen diffract similarly, the correct side chains termini of Asn and Gln are assigned based on the surrounding residues and the geometry of H-atoms (Hs) of surrounding water molecules. Mobile Hs as in His or hydroxyl groups need to be guessed from the environment. Moreover, crystal structures are the result of averages over atomic fluctuations representing the thermal motion around equilibrium. This is of

importance when performing docking. Crystal contacts between neighbouring unit cells of the crystal may influence the contact zone conformation. The term contact zone describes the first point of nucleation as the very first event in crystallization. This zone is crucial for explaining the mechanism of action of small molecule aggregation breakers in chapter 2.

0.1.1.2 BINDING AFFINITY IN DRUG DISCOVERY

In a disease-relevant (protein) target-centred drug discovery process, binding affinity plays a crucial role, as it is the main criterion for moving from a hit compound to a lead. Generally, a lead is defined as a compound with considerable biological activity, neither being extremely polar nor lipophilic, neither bearing toxic nor reactive groups leading to irreversibly binding, with analogues indicating that structural modification may alter activity¹². In the further optimization steps towards a drug, affinity still needs to be improved, along with target selectivity and pharmacokinetic (adsorption, distribution, metabolism, excretion and toxicity (ADMET)) properties.

0.1.1.3 IN VITRO VERSUS IN SILICO AFFINITY SCREENING

In the early hit discovery steps, *in vitro* high throughput screening (HTS) of large compound libraries against one or a few therapeutic targets is traditionally performed, although rather disappointing hit rates (i.e. percentages of true active molecules) of usually less than 0.1% are achieved^{13, 14}. Setting up criteria for compound selection and prioritization is thus very important¹⁵. Alternatively to HTS or as a lower cost and time expense complement^{16, 17}, targeted approaches like virtual screening (VS) or rational drug design (RDD) are viable strategies in the presence of reliable 3D target data, revealing the spatial and energetic criteria responsible for protein-ligand binding¹⁸. Thus, HTS is technology-driven and hampered by phenomena like limited solubility or aggregation, while VS is knowledge-driven; however, progress in the VS field is still impeded by the quality of the scoring function used to discriminate true hits from non-binders¹⁹.

0.1.1.4 VIRTUAL SCREENING

VS techniques can be divided into target- and ligand-based approaches.

In *target- or structure-based VS*, compound databases are searched for matching the physicochemical properties of a given protein binding site, with pre- or post-screening filters helping to funnel down the number of compounds to the experimental test sets (Figure 3). Common requirements are either lead- or drug-likeness. Compared to drug-like compounds, leads display generally more room for improvements of physicochemical properties and resemble rather fragments than drugs. Drug-likeness is empirically defined by Lipinski's "rule of 5"²⁰ (no more than 5 H-bond donors, no more than 10 H-bond acceptors, a molecular weight (MW) of maximally 500 Da and a logP < 5). A typical target-based VS protocol is summarized in Figure 4.

Ligand-based VS uses the information provided by one or several known binders to a target to extract similar compounds from databases, either by similarity and substructure searching²¹, pharmacophores^{22, 23} or shape matching²⁴. Pharmacophores are 3D representations of features essential for affinity, used to retrieve compounds conforming to these rules from a database²⁵ and to generate pseudo-receptors²⁶, which are simplified active site models based on a series of active compounds. Using all these techniques, detecting remote similarity (structurally dissimilar, but similar activity) is challenging²⁷, also because it is always difficult to establish a threshold that indicates valid similarity relationships.

The advantages and disadvantages of target- and ligand-based virtual screening are contrasted in Table 1.

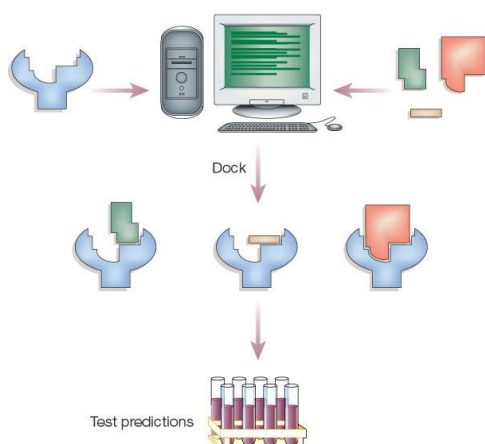


Figure 3: Structure-based virtual screening: In silico funnelling down the set of molecules to be tested experimentally on a particular protein, using docking programs like e.g. Dock. The Figure was taken from <http://www.nature.com/nature/journal/v432/n7019/images/nature03197-f2.2.jpg>.

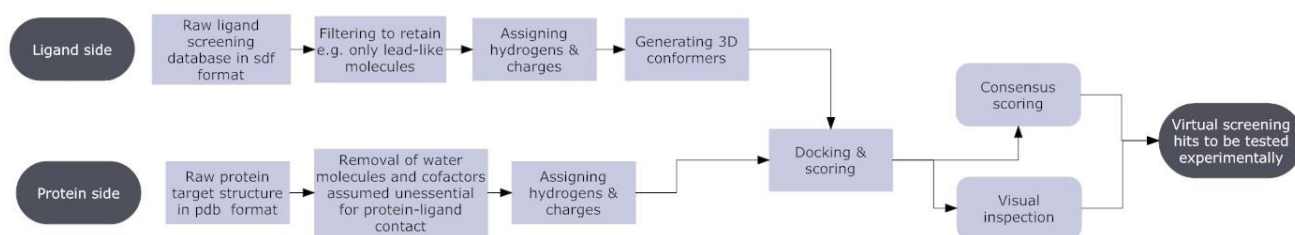


Figure 4: Flow-chart of a virtual screening procedure and analysis.

Table 1: Advantages and disadvantages of target- and ligand-based virtual screening.

	Structure-based virtual screening	Ligand-based virtual screening
Advantages	- Not restricted to known ligands - Target information considered - Low false negative hit rate	- Computationally not demanding - Target information not required
Disadvantages	- At least a homology model of target required - Computationally demanding - Inherits the limitations of scoring functions - Induced fit phenomena not sufficiently accounted for - High false positive hit rate	- No additional information from target - Only chemotypes of known ligands considered 3D structure of ligands dependent on method of conformer generation

0.1.1.5 RATIONAL DRUG DESIGN (RDD)

RDD relies on some 3D knowledge on the binding mode or mode of action, providing ideas on ligand series to be synthesised and tested. RDD results hint at further computer-assisted improvements. Perhaps the most known example of a drug developed by such approaches is the sialidase inhibitor oseltamavir²⁸, better known as Tamiflu®.

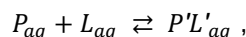
Ligand-based approaches to RDD without prior knowledge of the protein target 3D structure include quantitative structure activity relationships (QSAR), pharmacophore- or pseudo-receptor models. In QSAR, a correlation between the structure and the affinity of a series of biologically active compounds is attempted; yielding affinity predictions for compounds whose affinity is unknown.

In contrast, target-based approaches to RDD rely on crystal structures of proteins or models thereof. Molecules can be grown from fragments or compounds can be docked into the binding site, but both approaches crucially depend on a reliable, i.e. accurate and fast prediction of binding affinity. One needs first to generate reasonable ligand binding modes (configuration generation problem) and then identify the binding modes that correspond best to experimental data based on an *in silico* estimation (structure and affinity prediction problem) ²⁹. The first issue is solved, at least for less flexible binding pockets without water-mediated hydrogen bonds and protonation state changes of the protein upon binding ³⁰, while identifying and predicting the affinity of near-native ligand conformations is still not always trivial.

0.1.2 INTRA- AND INTERMOLECULAR CONTRIBUTIONS TO BINDING AFFINITY

Binding affinity characterizes the interaction strength of ligands with their respective binding site. High affinity ligands are stronger bound to a protein than low affinity ligands and stay longer at the binding site. Moreover, a lower concentration is sufficient for occupying a binding site and triggering a physiological response.

Binding affinity is a synonym for the Gibbs free energy of binding. In simple terms, it expresses the tendency of a molecule to interact with another one or the stability of the formed complex with respect to the free species ³¹. For reversible (i.e. non-covalent) binding, the association of a protein (P) with a ligand (L) in an aqueous (aq), ion-containing environment can be written as:



with the ' sign denoting possible conformational changes upon binding. The binding free energy difference between the bound and the unbound state is a sum of enthalpic components (breaking and formation of H-bonds, formation of specific hydrophobic contacts), and entropic components (release of water from hydrophobic surfaces to solvent, loss of conformational mobility of both the protein and the ligand) ³². Assuming a thermodynamic equilibrium, the Gibbs free energy of binding ΔG reflects the energy change upon binding and is related to the equilibrium constant K_{eq} via:

$$\Delta G = \Delta G_{complex} - (\Delta G_{protein} + \Delta G_{ligand}) = -RT \ln(K_{eq}) = -RT \ln\left(\frac{k_{on}}{k_{off}}\right) = \Delta H - T\Delta S ,$$

where $K_{eq} = \frac{k_{on}}{k_{off}} = K_a = \frac{1}{K_d}$, with k_{on} being the on-rate, k_{off} the off-rate, K_a the association constant and K_d the dissociation constant; R is the gas constant ($R = 1.99 \text{ cal/mol} \cdot K$), T the absolute temperature in K, ΔH the enthalpy change and ΔS the entropy change upon binding. The values of K_a or K_d , ΔH and ΔS can be obtained computationally. While ΔH is calculated using molecular mechanics, ΔS results from the Maxwell-Boltzmann equation. Archakov et al. nicely explain the role of enthalpy and entropy: ΔH depends on H-bond formation, electrostatic and Van der Waals (VdW) interactions, whereas ΔS depends on changes in the degrees of freedom related to backbone or solvent conformation, solvent and association entropy ³³. Because ΔH and ΔS often have an opposite behaviour called enthalpy-entropy compensation ³⁴, small changes in ΔG are only noticed.

Experimental K_d values fall in the range between 10^{-2} and 10^{-12} kJ/mol (at 298 K), corresponding to a ΔG between -10 and -70 kJ/mol in aqueous solution ^{8, 9, 32}. Thus, a difference in ΔG of 5.7 kJ/mol alters the K_d by one order of magnitude ¹. It was shown that ΔG increases by ca. 6.3 kJ/mol per atom for compounds with up to 15 heavy atoms and that ΔG is only weakly dependent on the molecular weight above this atom number ³⁵. Kuntz et al. ³⁵ stress that the largest binding interactions per atom are associated with metals, small anions, and ligands forming covalent bonds. Ionisable groups in biological systems are generally distributed on several atoms and electrostatic interactions with the binding site play a crucial role, especially in cases where only few other interaction points are seen (important for a phosphate-Lys interaction in chapter 2).

The physical basis for computational affinity prediction is laid by so called force fields. A force field is a collection of the main interatomic parameters that describe the potential energy of a system of atoms. These parameters originate from crystal structures, infrared spectroscopy, density measurements, vaporization enthalpies (i.e. the energies required to transform given substance quantities into gas at given pressures) or solvation energies. The simplest molecular mechanics (MM)-based force fields

consider VdW interactions and electrostatics only, leaving entropy and solvent effects aside. Obeying Newton's laws, protein and ligand atoms are represented by hard spheres connected by springs (covalent bonds). In a constant force field, the potential energy U is a function of the coordinates. For a simple bond stretching, the force F acting on the mass is $F = -kl$, where k stands for the spring constant, l_0 for the starting position, and l for the final position:

$$U(l) = - \int_{l_0}^l F dl = - \int_{l_0}^l (-kl) dl = \frac{1}{2} k (l - l_0)^2$$

A full force field equation contains a combination of bonded (angles, bonds and torsions) and non-bonded interaction terms (Lennard-Jones potential and Coulomb equation). In a similar way to the bond interaction term mentioned above, the other terms can be derived.

The AMBER (Assisted Model Building and Energy Refinement)³⁶ potential energy function used in chapter 1 is given as an example of a force field equation:

$$U(r^N) = \underbrace{\sum_{bond} \frac{1}{2} k_b (l_i - l_{i,0})^2 + \sum_{angle} \frac{1}{2} k_a (\theta_i - \theta_{i,0})^2 + \sum_{torsion} \frac{1}{2} U_n [1 + \cos(n\omega - \gamma)]}_{\text{BONDED INTERACTIONS}} +$$

$$\underbrace{\sum_{i=1}^N \sum_{j=i+1}^N \left\{ \underbrace{\epsilon_{ij} \left[\underbrace{\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12}}_{\text{repulsion}} - 2 \underbrace{\left(\frac{\sigma_{ij}}{r_{ij}} \right)^6}_{\text{attraction}} }_{\text{Van der Waals}} + \underbrace{\frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}}_{\text{Coulomb}} \right\}}_{\text{NONBONDED INTERACTIONS}}$$

U is a function of r (the coordinates) and N (the number of atoms). The first term of the equation represents the energy between covalently bonded atoms, the second the energy due to the angles between the electron orbitals, and the third the torsional bond energy due to the bond order, i.e. the influence of neighbouring atoms or lone electron pairs.

Among the nonbonded interactions, VdW forces are determined by long-range attractive dispersion forces (represented by $\frac{1}{r^6}$) and short-range Pauli repulsion due to overlapping electron orbitals ($\frac{1}{r^{12}}$). σ_{ij} is the empirically determined equilibrium distance, which is multiplied by the factor 2, r_{ij} is the inter-atomic distance and ϵ is the well depth. The electrostatic interactions are accounted for by the Coulomb term, varying as the inverse of r_{ij} with the partial charges q_i and q_j . ϵ_0 represents the electric constant. The AMBER force field was calibrated against quantum mechanics data and neutron diffraction as well as microwave spectroscopy data. The parameters of the force field are stored in parameter files, of which FF99 (parameters for peptides, proteins, and nucleic acids) and GAFF (Generalized AMBER Force Field: parameters for small organic molecules) are used in chapter 1.

The force field equation stated provides a first insight into important protein-ligand interactions. Besides the mentioned bonded and nonbonded interactions (VdW (0.1.2.3) and electrostatic interactions (0.1.2.1)), H-bonds (0.1.2.2), hydrophobic interactions (0.2.2.1), solvation and desolvation (0.1.2.4), conformational changes (0.1.2.5) as well as additivity, cooperativity, and enthalpy-entropy compensation (0.1.2.6) influence affinity and are discussed in the indicated sections of the introduction.

From a medicinal chemist point of view, looking at binding affinity in terms of average functional group contributions to drug-receptor interactions can be informative. Already in 1984, a ranking of such contributions was provided (Table 2), confirming that charged groups bind more strongly than polar groups, which in turn bind more tightly than non-polar groups³⁷. The implications of a strong, complementary binding contribution of a phosphate and a Lys to the disruption of antibody aggregation are discussed in chapter 2.

Table 2: Calculated average functional group contributions in kcal/mol according to ³⁷.

Contribution in kcal/mol per					
Charged groups	CO ₂ ⁻ : 8.2	PO ₄ ²⁻ : 10.0	N ⁺ : 11.5		
Polar groups	N: 1.2	OH: 2.5	CO: 3.4	O, S or ethers: 1.1	Halogens: 1.3
Nonpolar groups	C(sp ²): 0.7	C(sp ³): 0.8			

0.1.2.1 ELECTROSTATICS

As the most important contributors to non-covalent binding, electrostatic interactions include salt bridges, hydrogen bonds, dipole-dipole interactions, or metal ions. Salt bridges are formed between ligand moieties and protein side chains bearing an opposite charge of the same magnitude. Considering a physiological pH of 7.4, Arg and Lys side chains are protonated, while the carboxyl groups of Asp and Glu are deprotonated. Except for Lys, these charges are often delocalized. The protonated terminal amino group of Lys with a pK_a of 10.8 ³⁸ are of particular importance in chapter 2, where the concept of “buried complementary charges” ³⁹ intervenes. His side chain protonation needs to be looked at on a case-by-case basis given the local atomic environment. Cation- π and the recently discovered anion- π ⁴⁰ interactions can be regarded as special cases of a salt bridge. Because H-bonds are of utmost importance, an own subchapter (0.1.2.2) is dedicated to them. Interactions of the dipole-dipole or metal ion type are not further treated.

0.1.2.2 HYDROGEN BONDS

Together with hydrophobic contacts, hydrogen bonds (H-bonds) keep secondary protein structures in their 3D fold and assure the formation of ligand-protein complexes (except for excessively hydrophobic compounds like retinol binding to the retinol binding protein ⁴¹). H-bonds are generally accepted to positively influence protein stabilisation ⁴². Per H-bond contributions can be estimated via model compounds or site-directed mutagenesis. With the latter method however, also other interactions are altered and eventually the solvation of the binding site changes. For isosteric replacements, one needs to consider side chain entropy and the hydrophobic effect (i.e. the tendency of H₂O to exclude nonpolar molecules discussed in 0.2.2.1), roughly estimated to be 1.5 ± 1.0 kcal/mol for a buried, uncharged H-bond ⁴².

Fine-tuning an H-bond with respect to its prevalent covalent, electrostatic or VdW character is possible, given that an H-bond falls in an energy range from 40 to approximately 0.25 kcal/mol ⁴³. The characteristics of electrostatic (N-H $\cdots$ O and O-H $\cdots$ O) and weaker, dispersive H-bonds (C-H $\cdots$ O ⁴⁴, C-H $\cdots$ N and CH $\cdots\pi$) have been studied in high-resolution crystal structures ⁴⁵, with the electrostatic ones being preferred in protein-ligand complexes and dominating packing, and the dispersive ones being most abundant in H-bonds to water or hydrophobic environments. The considerable entropy loss upon trapping water in a binding site can be compensated by a gain in enthalpy whenever water-mediated H-bonds are formed. A water molecule has thus to form as many H-bonds as possible at an optimal distance and angle. Typically, the N $\cdots$ O distance is between 2.8 and 3.2 Å and the angle N-H $\cdots$ O larger than 150°, while the angle C=O $\cdots$ H has a broader range between 100 and 180°. The importance of water in binding is further corroborated by its presence in binding sites of protein-ligand complexes, but absence in the holo form ⁴⁶. Specificity of recognition is generally augmented by using more interactions and reversibility of binding is guaranteed by the existence of rather flexible weak interactions that can easily be broken. However, where water-mediated interactions between a ligand and a protein significantly enhance affinity, replacing such water with an additional H-bond might result in affinity loss, because the new H-bond cannot compensate for the lower desolvation energy ⁴⁷.

The H-bonding strength depends inversely upon the surrounding dielectric constant, which means that buried H-bonds contribute stronger to affinity than more solvent-exposed ones ⁴⁸. As an example, only 1-2% of the buried NH and C=O of a protein amide bond do not contribute to an H-bond ⁴⁹, meaning that such “single” polar groups are highly detrimental to binding ⁴⁸. The importance of H-bonds in ligand orientation can be explained by the well-known X-ray structures of dihydrofolate (DHF) or methotrexate (MTX) bound to dihydrofolate reductase. Although at a first glance, DHF and MTX look very similar, a superposition based on the maximum common substructure (Figure 5a) does not yield

the highest match of H-bonding functions. While in Figure 5a, they share only 3 donor / acceptor orientations, 5 functions are shared in Figure 5b ⁴¹.

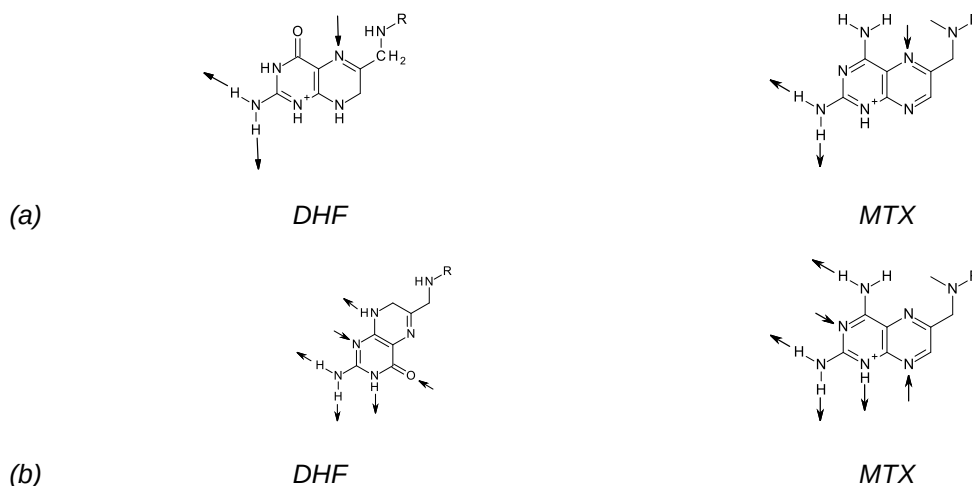


Figure 5: Dihydrofolate (DHF) and methotrexate (MTX) look very similar, but share more H-bond donor and acceptor functions (arrows) if they are not just simply overlaid based on the maximal common substructure.

A further example of the directional nature of H-bonds being crucial for ligand recognition includes the four nucleic acids of which DNA is composed, where a difference of only one H-bond is decisive of the base pair matching. With respect to the variable geometrical preferences of different H-bond types, Isostar (<http://isostar.ccdc.cam.ac.uk/html/isostar.html>), a library of non-bonded interactions ⁵⁰, provides useful insights. Notably aromatic nitrogens such as in purines or pyrimidines exhibit a strong directionality (chapters 1 and 4), while moderate geometrical preferences are found for phosphate groups ⁵¹ (chapter 2).

Regarding the maximal number of H-bonds, expressed as the number of polar atoms per total number of atoms, it is quite high for small, tight complexes like biotin-streptavidin ²⁰. The affinity of streptavidin for biotin has to date one of the strongest known non-covalent binding energies ($K_d \sim 10^{-15}$ M ⁵²). This “superaffinity” arises from 5 hydrophobically encircled H-bonds, which were shown by a 10 ns molecular dynamics (MD) to be responsible for atypical entropic and enthalpic hydration penalties of the apo structure. The complex was stabilized by a 5-membered water ring assuring the maintenance of a maximal number of hydrogen bonds ⁵³. This result provided an explanation as to why implicit solvent treatment via e.g. the solvent-accessible surface area (SASA), the Poisson-Boltzmann (PB) equation or the Generalized Born (GB) equation underestimates the binding affinity for this particular case, while explicit methods lie within experimental accuracy ⁵⁴⁻⁵⁶. The PB equation is a differential equation that describes electrostatic interactions between molecules in solution, while the GB approximation is an approximation to the exact solution of the PB equation, where the protein is treated as sphere and the dielectric constant at the protein surface differs from the one in the protein interior.

As a general rule however, no clear correlation exists between the number of hydrogen bonds and affinity ⁸, also because the strength of H-bonds varies considerably from 2-6 kJmol⁻¹ for neutral ones (2-15 times better affinity) to 20 kJ/mol (3000 times better affinity) for ionic ones ⁵⁷. The insufficient understanding of H-bonds is nevertheless not the only obstacle to reliable affinity predictions, given that desolvation (0.1.2.4) and the hydrophobic effect (0.2.2.1) need further studying as well.

0.1.2.3 VAN DER WAALS (VDW) INTERACTIONS

With respect to attractive/repulsive VdW interactions in protein crystals, it has been noted that “empty space is wasted space” ⁵⁸. This concept is also valid for protein-ligand interactions ⁵⁹. In contrast to disappointing affinity estimations based on the number of H-bonds (0.1.2.2), maximal VdW interactions can be estimated easily based on a simple, solvent-free model ^{60, 61}. With higher atom numbers, “self-shielding” reduces the average atom contribution as a function of r^6 , r being the atom

radius, and interaction ranges can be defined as a function of the ligand geometry. Maximizing surface/volume ratios is thus achieved with branched, extended molecules, typical examples being protease inhibitors ⁶² (chapter 3). However, conformational changes upon binding and solvation effects are usually not considered in such models.

0.1.2.4 SOLVATION AND DESOLVATION

Molecular association in an aqueous environment is a balance between solute and solvent interactions; upon binding, solvent-solute interactions are replaced by solute-solute interactions ⁶³. A majority of water molecules implicated in stabilizing protein-ligand interactions behaves similarly ⁶⁴, forming compact tetrahedral-shaped networks. These so-called interstitial water molecules generally prefer cavities on the protein to cavities on the ligand, probably reflecting the lower conformational freedom of protein cavities which are more conserved and energetically favourable.

0.1.2.5 CONFORMATIONAL CHANGES

Conformational changes of a protein upon ligand binding are thought to be mainly driven by hydrophobic interactions ⁵⁷. During such an association, entropy is changed due to a reduction of the conformational mobility, i.e. the internal rotations. Looking at the ligand and the protein only, each of them loses 3 degrees of translation and rotation and 6 vibrational degrees are gained. Such an entropy loss was estimated for amino acids to be 0 for Ala, Gly and Pro and 8.7 kJ/mol for Gln ^{65, 66}.

0.1.2.6 ADDITIVITY, COOPERATIVITY, AND ENTHALPY-ENTROPY COMPENSATION

A frequent approach to the understanding of the total ΔG is the concept of the additivity of contributions of different chemical groups ^{37, 67}. However, in statistical thermodynamics, ΔG is a system's property, which cannot be uncoupled from the system's configuration. Thus, additive ΔG approaches are questionable with respect to the underlying physical meaning. Approaches geared towards splitting ΔG into physicochemical contributions are an alternative.

These contributions are not always additive, as the burial of hydrophobic side chains can favour cooperatively near-by electrostatic interactions ^{68, 69}. When looking at the highest-affinity protein-ligand interactions, additivity of interactions is not true, while it holds true for less affine ligands. Brooijmans et al. stated that the lack of increase in affinity with higher interface size is likely due to non-thermodynamic factors, such as functional and evolutionary constraints ⁷⁰.

Enthalpy-entropy compensation, where the enthalpy and entropy changes cancel each other out, is responsible for the fact that in protein-ligand complexes, binding is enhanced by a loss of degrees of freedom.

0.1.3 DOCKING

Searching for a ligand that can fit both geometrically and energetically into the binding site of a protein is called molecular docking, a problem with thousands of degrees of freedom. Docking programs use either regression- (0.1.4.3.4) or knowledge-based (0.1.4.3.5) scoring functions to rank ligand poses. The protein-ligand complex to be formed should display a high complementarity with respect to steric and surface properties and the ligand should adopt an energetically favourable conformation. In other words, any docking program can be divided into 2 main parts: (i) docking, i.e. searching the conformational space for low energy conformations of a ligand in the context of a given protein active site and (ii) scoring, the interaction energy function evaluating the protein-ligand interactions. As noted by Brooijmans and Kuntz, the former must search the potential energy landscape in enough detail to find the global energy minimum, while the latter should represent the ligand-protein interactions as accurate as possible, assessing both steric and chemical complementarity between a ligand and a protein ⁷¹. However, there is always a dilemma between the accuracy of the scoring function and the speed of the docking program and therefore compromises must be made.

0.1.3.1 DOCKING ALGORITHMS

A good docking algorithm requires an energy function that is selective, in that it clearly differentiates correctly docked structures from misdocked ones, and that is efficient, meaning that a correctly docked structure can be identified quickly ⁷². Leach et al. provide a comprehensive overview on the successes and gaps in docking and scoring ⁷³. When assessing a program's ability to reproduce the ligand's X-ray pose, an ideal docking/scoring combination should result in a single solution close to the ligand's X-ray position (in terms of the heavy atom root-mean square deviation (RMSD)) ⁷⁴. However, when comparing protein-ligand docking programs, it is difficult to draw conclusions of general applicability ⁷⁵ as the docking performance of most programs is highly dependent on the specific characteristics of both the binding site and the ligand ⁷⁶⁻⁸⁵. Different docking/scoring combinations have been extensively assessed in the past ^{80, 83, 86, 87}, with the probably most extensive being the assessment conducted by Warren et al. on 10 docking programs and 37 scoring functions ⁸⁷. This study shows that ligand conformations similar to the crystallographic one could be obtained and that the active compounds were retrieved among the pharmaceutically relevant decoys, but that success was still target-dependent and none of the docking programs/scoring functions made a useful ligand binding affinity prediction.

A better weighting of hydrophobic interactions, an accurate penalisation of unfavourable interactions, a better representation of ligand solvation, and a proper parameterisation of π -stacking, cation- π as well as transition metal interactions are needed. Because tautomeric forms differ in shape, functional groups, surface, and H-bonding patterns ⁸⁸, active sites and ligands should be carefully protonated based on the presence of structural water molecules in a crystal structure, i.e. water molecules lying at H-bonding distance to both the protein and the ligand.

Upon binding, a small molecule loses translational and rotational entropy. To reduce the conformational space to be searched upon docking for speed and efficiency reasons, usually all internal degrees of freedom except three translational and rotational degrees of freedom are kept fixed. However, induced fit upon ligand binding is not at all considered or accounted for in a limited way ⁴, as the binding site is predefined and only some amino acid side chains are eventually allowed to rotate. Docking algorithms usually take ligand flexibility into account and are either systematic, stochastic and deterministic search algorithms ⁷¹ (0.1.4.2.7). Moderate molecular flexibility can contribute to a favourable change in binding energy in two ways: by optimizing the non-covalent protein-ligand interaction (contributing to a favourable change in enthalpy), or by increasing the entropy (or minimizing the decrease in entropy) upon binding by releasing interfacial water and increasing flexibility in parts of the protein or the ligand ⁸⁹, which can eventually lead to enthalpy-entropy compensation. Some protein flexibility can be accounted for by rotamer libraries of protein side chains or by considering several protein conformers from crystal structures ⁹⁰, Monte Carlo (MC) or Molecular dynamics (MD) simulations ⁹¹ or by using the ICM-flexible receptor docking algorithm (IFREDA), where discrete protein conformations are used as input ⁹². Minimal rotations of side chains generally assure good protein-ligand shape complementarity, mimicking the motions found in nature ⁸⁹. One has however to keep in mind that the docking accuracy decreases substantially for ligands with 8 or more rotatable bonds or when using an average or apo protein structure ⁹³.

0.1.4 THEORETICAL BINDING AFFINITY PREDICTION

Protein-ligand interaction strength is commonly characterized in terms of binding affinity, high affinity being associated to higher stabilizing forces between the two binding partners and longer ligand resident times as compared to low affinity. In an enzymatic reaction, the dissociation constant K_d or the inhibition constant K_i are mostly taken as a measure for affinity. IC_{50} values, which are also used, are known to vary depending on experimental conditions, contrarily to the absolute K_i values. IC_{50} values represent the concentration at which the activity of an inhibitor is half the maximum. They can only be related to K_i values in case of a competitive inhibition via the Cheng-Prusoff equation ⁹⁴:

$$K_i = \frac{IC_{50}}{1 + [L]/K_d}, \text{ where } [L] \text{ is the unbound ligand concentration and } K_d \text{ the dissociation constant.}$$

The already mentioned enthalpy-entropy compensation (0.1.2.6) is a main hurdle to the prediction of binding free energy. However, by using structural data to predict the enthalpy, calculated entropy, and heat capacity (i.e. the amount of heat inducing a temperature / entropy change in a protein-ligand

system) values for protein-peptide binding were in good agreement with experimental data (reviewed by Murphy ⁹⁵). Most commonly, the heat capacity and the enthalpy changes are seen as a function of polar and apolar accessible surface area changes with empirical coefficients. When calculating the total entropy change, solvent effects, backbone, and side chain rearrangements as well as changes in the translational, rotational and vibrational degrees of freedom upon binding need to be accounted for.

When looking at methods for theoretical affinity prediction, *ab initio* quantum mechanics are more precise than semi-empirical methods, because they explicitly consider the electronic configuration, with the drawbacks of being time-consuming and limited to small systems. Next on the affinity prediction precision scale come MM-based methods, which are discussed briefly in 0.1.4.2. The basics of free energy perturbation, thermodynamic integration, linear energy calculations, steered molecular dynamics (applied for the first time to affinity prediction in chapter 1), force field-, regression- and knowledge-based methods are also explained. The regression- and knowledge-based methods yield scoring functions used in molecular docking (0.1.3).

0.1.4.1 AFFINITY PREDICTION WHEN TARGET STRUCTURE IS UNKNOWN

In the absence of a protein structure, a similar biological activity is assumed to correspond to a chemical similarity in the ligands. The precision of ligand-based affinity prediction methods is compared to other methods in Figure 6. Ligand-based methods assessing binding affinity qualitatively include fingerprints (0.3.1.1, protein-ligand interaction types coded in bit strings) and pharmacophores (0.1.1.4), while QSAR yields qualitative predictions, provided the training set with a wide range of presumably “bio-active” ⁹⁶ conformations and the initial compound alignments are adequate. Comparative Molecular Field Analysis (CoMFA) ⁹⁶ is a commonly used 3D-QSAR technique which evaluates molecules with respect to steric (VdW) and electrostatic (Coulomb) molecular interaction fields (MIF), reflecting enthalpic contributions, but assuming that entropic terms are constant across the compound series studied. Alternatively, Comparative Molecular Similarity Analysis (CoMSiA) ⁹⁷ can be carried out. It is based on Gaussian functions for steric, electrostatic, and hydrophobic interactions and avoids steep potentials close to molecules found in CoMFA. Efforts to increase the dimensionality of 3D-QSAR exist, such as 4D- (i.e. 3D- and conformational sampling) and higher dimension-QSAR ⁹⁸. Another MIF method, GRID (chapter 4), considers also hydrophobic probes and probes forming one ⁹⁹ or several H-bonds ¹⁰⁰. In a fundamentally different method, called YAK ¹⁰¹, amino acids are arranged around a ligand in such a way as to optimally reflect its affinity, yielding a pseudoreceptor representation of the active site.

0.1.4.2 AFFINITY PREDICTION WHEN TARGET STRUCTURE IS KNOWN

Among the MM-based methods, MM-PBSA (Molecular Mechanics Poisson Boltzmann Surface Area) and MM-GBSA (Molecular Mechanics Generalized Born Surface Area) were successfully applied to affinity prediction ¹⁰². These methods regard ΔG as the difference of the free energy of the solvated complex minus the sum of the free solvated protein and the free solvated ligand, respectively. ΔG is obtained from MD snapshots as the sum of force-field related potential energies (bonds, angles, torsions, VdW and electrostatic interactions), as well as the polar and non-polar solvation free energy ¹⁰³. With respect to polar solvation free energy on the one hand, in MM-PBSA, the Poisson Boltzmann equation is used, while in MM-GBSA the so-called Generalized Born approximation is applied to treat the solvent implicitly. On the other hand, non-polar contributions are calculated as being linearly dependent on the SASA.

Considering a system of a ligand and a protein allowed to form a complex, the free energy of all the components describes the reaction and association propensity of the whole system. This property is commonly calculated using MD or MC techniques, with “roadblocks” being errors in the systems’ energy representation and insufficient sampling of the relevant (low free energy) conformations of the system ⁶³. Ways to quantify the protein-ligand interactions obtained by such means can be divided into four accuracy levels: (i) thermodynamic methods, (ii) force field-based methods, (iii) ligand-based methods (0.1.4.1), and (iv) scoring functions. With respect to (i), the accuracy still depends upon a reliable parameterisation of energetic terms, such as dielectric constants and VdW weighting factors ¹⁰⁴. Due to the extensive calculation time, they are so far not applicable to standard protein-ligand docking (0.1.3), but to the refinement of docking-derived affinity estimates.

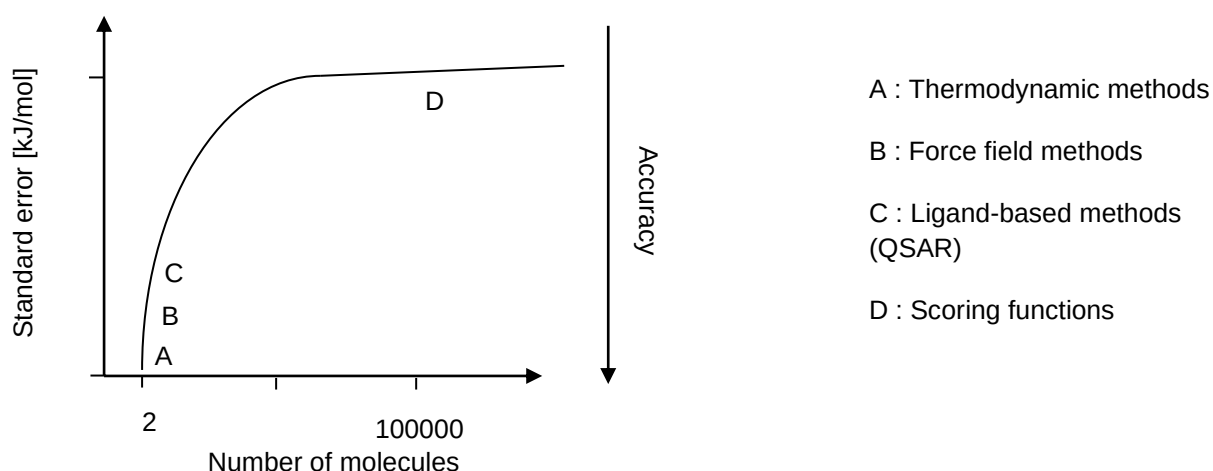


Figure 6: Qualitative comparison of different *in silico* binding affinity prediction methods. Thermodynamic methods are most accurate, followed by force field-based methods, ligand-based methods (QSAR), and scoring functions used in molecular docking.

0.1.4.2.1 FREE ENERGY PERTURBATION (FEP), POTENTIAL OF MEAN FORCE (PMF) AND THERMODYNAMIC INTEGRATION (TI)

The statistical mechanics behind the techniques cited in this paragraph are extensively reviewed in view of predicting free energies in ¹⁰⁵. Before dynamic studies were feasible, the infinite order free energy perturbation (FEP) formula was derived by Zwanzig ¹⁰⁶ as $\Delta G_{AB} = -RT \ln \langle \exp(-\Delta H_{AB}/RT) \rangle_A$, where ΔG_{AB} is the free energy difference between systems A and B, R the universal gas constant, T the temperature, $\Delta H_{AB} = \Delta H_B - \Delta H_A$ the small Hamiltonian difference and $\langle \dots \rangle_A$ an ensemble average of A. This equation is helpful in calculating relative binding free energies between the free and the bound species (protein and ligand) from ensemble averages of the system extracted from MC or MD runs. The calculations of absolute ones would be inaccurate, provided that both MC and MD techniques are hampered by poor sampling. FEP ¹⁰⁷ with explicit solvent is currently the most powerful and promising approach to protein-ligand binding energy prediction ⁶³, yielding nearly quantitative free energy changes in solution while at the same time allowing for dynamics of the system ¹⁰⁵. Its precision is however limited by the requirement of ΔH_{AB} being small, which depends on the force field used and the sampling. Improved sampling can be achieved by $\Delta G_{A+B \rightarrow AB}$ (forward) and $\Delta G_{AB \rightarrow A+B}$ (reverse) sampling, as the most accurate forward value is calculated by averaging the total free energies of $\Delta G_{A+B \rightarrow AB}$ and $\Delta G_{AB \rightarrow A+B}$ ¹⁰⁸. The difference between the ΔG s obtained from forward and reverse calculations is called hysteresis and the total error in ΔG is either half the hysteresis ¹⁰⁹ or the standard deviation from repeated runs ¹¹⁰. Besides, free energy convergence may be checked with much longer simulations ^{111, 112}.

Free energy changes can also be calculated with respect to a reaction coordinate or a bond distance. The second case is known as a PMF. Contrarily to the above equation, the geometries are changing (ξ and $\xi + \Delta\xi$): $\Delta G_{\xi \rightarrow \xi + \Delta\xi} = -RT \ln \langle \exp(-(\Delta H_\xi - H_{\xi + \Delta\xi})/RT) \rangle_\xi$. Usual methods to calculate the PMF consider first the free ligand in bulk solvent, restrain it to achieve the bound conformation, and translate it then into the binding site. From the PMF, the association free energy is calculated by integrating the above equation, provided $\Delta\xi$ is given. PMF calculation is applicable to cases where FEP fails, e.g. for a strong enthalpy-entropy compensation, strong electrostatic interactions, and a high ligand flexibility in the unbound state. Because the accuracy of FEP results is “just” limited by the computational effort for conformational sampling, with increasing computer power, protein conformational flexibility should become addressable.

An alternative to FEP yielding the individual energy contributions is thermodynamic integration (TI) ¹⁰⁷, which is adapted to systems with large conformational changes. In this method, ligand flexibility is enhanced by lowering the rotational barriers, causing a large sampling.

0.1.4.2.2 LINEAR INTERACTION ENERGY CALCULATIONS

Linear interaction energy (LIE) calculations are a compromise between the accuracy and the speed of ΔG predictions^{113, 114}, yielding estimates with a precision of about 1 kcal/mol¹¹⁵⁻¹¹⁷. The approach can also be applied to docking/scoring issues^{118, 119}. In this method, no information on intermediate states is needed. A methodological overview can be found in e.g.¹¹⁹. Additive contributions to ΔG can be obtained from MM calculations based on the differences in VdW energy, electrostatic energy, and a solute cavity penalty of the protein-bound ligand and the unbound ligand in solution. For both states, starting from either crystal structures or docking poses, conformational sampling is performed independently via MD or MC techniques to obtain ensemble averages. Each component of ΔG is weighted according to an appropriate training set of compounds with known affinity and the model then applied to a huge test set, one example being a study on bromobenzimidazole inhibitors of the casein kinase 2 (CK2)¹²⁰.

0.1.4.2.3 STEERED MOLECULAR DYNAMICS (SMD)

Because binding and unbinding are rare processes in a protein-ligand system, steered molecular dynamics (SMD) can be used to force the ligand in a functionally sound direction. An in-depth introduction into SMD is provided by Izrailev et al.¹²¹. SMD can be applied to affinity prediction given the following conditions: a) a reliable set of experimental affinity data for close ligand analogues, b) the corresponding high resolution protein-ligand crystal structures, and c) a crystal structure or homology model of the open, still ligand-bound conformation are at hand (so as not to strongly distort the protein structure upon unbinding). Given these prerequisites, open conformations of all protein-ligand complexes can be modelled and submitted to SMD. The maximal unbinding force (or work) at ligand exit from the protein, obtained for each ligand, can then be related to experimentally determined affinity data. The obtained regression line serves as a basis for the binding affinity determination of ligand analogues which have been previously docked into the open protein conformation. This method has been used in chapter 1 to predict the affinity of potential thymidine kinase substrates.

0.1.4.2.4 REGRESSION-BASED METHODS

Regression-based or empirical scoring functions rely on a statistical frequency analysis of protein-ligand interactions, with the underlying assumption that favourable interactions occur more frequently than less favourable ones. In other words, the total enthalpy is assumed to be additive (0.1.2.6). These approaches deal with finding mathematical functions that can be fitted to experimental binding data. They include e.g. partial least square fitting, multiple linear regression or artificial neuronal network approaches and rely usually on X-ray protein-ligand complexes, for which the experimental ligand affinity is known. They are empirical because their accuracy and transferability depends largely on the compilation of the training set¹²² and because they represent the most frequently observed interaction types. The pioneer scoring function of Böhm included H-bonds, ionic interactions, buried non-polar surface regions, and the loss of (intra)molecular mobility. It was based on 45 complexes with a cross-validated standard deviation of 9.3 kJ/mol¹²³, a value which was lowered by 0.5 kJ/mol by adding 37 complexes, aromatic interactions and unfavourable electrostatics¹²⁴. Lessons learned in subsequent scoring functions such as ChemScore¹²⁵ and Score¹²⁶ included: an increase of the number of complexes considered in both, a refinement of intramolecular flexibility in Chemscore and an inclusion of water molecules crucial for binding in Score. The training set dependency is an advantage for target-tailored scoring functions¹²⁷⁻¹²⁹ but is a general disadvantage when the scoring function is applied to other protein-ligand complexes. Although an universal and highly accurate scoring function may still be out of reach, current approaches are often capable of yielding reasonable correlations for sets of similar ligands¹³⁰.

Despite the fact that many features involved in binding are not fully understood, the additive approach is a good approximation for empirical scoring functions, although they fail to describe cooperativity in detail¹³¹. Thus, empirical scoring functions have been successfully used to correctly rank virtual screening results, and many cases where the correct ligand pose has been predicted can be found in the literature¹³²⁻¹³⁵.

0.1.4.2.5 KNOWLEDGE-BASED APPROACHES

In contrast to empirical scoring functions, which heavily rely on the types of interactions used in the training set and on the way binding affinities have been determined, knowledge-based scoring functions rely on “knowledge” derived from a large dataset of protein-ligand complexes, i.e. the frequency of occurrence of individual contacts is presumed to be a measure of their energetic contribution³². Comparing the performance of knowledge-based and empirical scoring functions, the former are better in extracting binding poses similar to experimentally confirmed ones, while the latter are better in the actual affinity prediction, provided that a fairly accurate binding geometry is given¹⁸. The knowledge-based approach was particularly successful in predicting affinities for inhibitors of the well-studied Human Immunodeficiency Virus HIV-1 protease^{136, 137}. Perhaps the most renowned knowledge-based scoring functions are PMF (potential of mean force)¹⁰⁴ and DrugScore¹³⁸. The first considers implicit water contributions and protein-ligand distances up to 12 Å and yielded a deviation of 1.8 log K_i from experimental values in its initial implementation. The second relies on a large protein-ligand interaction database (ReliBase) and achieved the same affinity prediction accuracy¹³⁹ as PMF. Above, DrugScore includes solvation and entropy contributions in an implicit manner and does not require any assumptions on protonation states, since only contacts to non-hydrogen atoms are regarded¹³⁹.

Knowledge-based scoring functions are not necessarily more exact than empirical scoring functions but rely less on the training set and the regression model used to weight each interaction. Both types of scoring functions depend however upon the resolution of the crystal structures considered for defining the scoring function and the precision of the method used for experimentally determining binding affinity.

0.1.4.2.6 FROM GENERAL TOWARDS TARGETED SCORING FUNCTIONS

Among the potentials used for quantifying protein-ligand contacts, there is a trend from general towards targeted scoring functions, like the Astex Statistical Potential targeted towards cyclin dependent kinase 2 (CDK2)¹⁴⁰. This is in line with the increasing interest in cheminformatics approaches (0.3), which aim at matching the ligand with the target space and vice versa. Statistical (i.e. empirical) potentials seem particularly suited for a targeted approach. They rely on the comparison of the observed number of contacts between atom types with the number of contacts one would expect if there were no inter-atomic interactions. Such a scoring function can be adapted whenever new structures become available, rendering it even more targeted. Additionally, it is much fuzzier than a pharmacophore and can be applied to compounds with a less conserved binding motif¹⁴⁰.

0.1.4.2.6 IMPROVING SCORING FUNCTIONS

Knowledge- and regression-based scoring functions are the fastest scoring functions but rely on the quality and number of experimental affinities. Thus, the more data are available, the better the scoring function. With respect to improving scoring function predictability, especially in virtual screening (0.1.1.4) (chapters 3 and 4), a popular concept is consensus scoring^{78, 79, 141}. In this technique, several scoring functions are applied, which weight features of the ligand in a different way. If the ligand is scored by all scoring functions in e.g. the top 10 %, it will get a high consensus score. In this way, the bias of each single scoring function is downweighted. A less pragmatic, more quantitative method correlates the outputs of several scoring functions with experimentally determined affinity data by partial least square analysis¹⁴². Otherwise, unfavourable interaction results can be filtered out⁴⁸. However, assessments of scoring functions generally agree that most of them perform well in recognizing near-native from misdocked solutions, but the prediction of binding affinities remains problematic¹⁴³. Reliable binding affinity prediction with docking is not yet achieved, as shown by Warren⁸⁷. A few exceptions¹⁴⁴ exist however. The problem with scoring can be simply put as a subtraction problem of large, rather inaccurate numbers (the interaction energy of the complex minus the energy required for complex formation).

For developing more reliable scoring functions, curated test sets of binding data (preferentially K_d values) complemented with microcalorimetric data are increasingly used. Databases comprising such test sets are compiled in Table 3.

Table 3: Suitable databases for developing scoring functions.

Database name	URL
BRENDA (BRAunschweig ENzyme DAtabase) ¹⁴⁵⁻¹⁴⁷	http://www.brenda.unikoeln.de
BindingDB ¹⁴⁸	www.bindingdb.org
PDBCal ¹⁴⁹	http://www.pdbcal.org/
AffinDB ¹⁵⁰	http://pc1664.pharmazie.uni-marburg.de/affinity/
BindingMOAD (Mother Of All Databases) ^{151, 152}	http://www.bindingmoad.org
PDBbind ^{153, 154}	http://www.pdbbind.org
Ligand-Protein Database ¹⁵⁵	http://lpdb.scripps.edu
ProLINT (Protein Ligand INTERactions; restricted to phosphatases and kinases) ¹⁵⁶	http://dna01.bse.kyutech.ac.jp/jouhou/prolint/prolint.html
Protein Ligand Database ¹⁵⁷ (PLD; containing experimental and calculated binding energies, obtained by a knowledge-based method ^{158, 159})	http://www-mitchell.ch.cam.ac.uk/pld/

The PDBCal, the AffinDB, the BindingMOAD, the PDBbind, and the Protein-Ligand Database are the most suitable for scoring function development.

0.1.4.2.7 DOCKING PROGRAMS USED IN THIS THESIS

Among the search algorithms explicitly accounting for ligand flexibility, systematic, stochastic, and deterministic algorithms can be mentioned.

Systematic searches are either of the anchor-and-growth, incremental construction or matching algorithm type.

In the *anchor-and-grow* method, an anchor region is defined in the ligand with the underlying assumption that the anchor position observed in the intact ligand is among the lowest-energy anchor positions when evaluated independently of the rest of the ligand.

Subsequently, the ligand is constructed around the anchor fragment in the binding site of the protein. DOCK was the first docking program being developed in the 1980ies ⁷⁶. In the DOCK 4.0 algorithm ¹⁶⁰ and further developments ^{161, 162}, the anchor is placed based on steric complementarity and flexible parts are grown incrementally with the help of a matching algorithm (detailed later), allowing for rigid or flexible ligand docking.

In *incremental construction*, a ligand is broken into fragments at the rotatable bonds and is then assembled piecewise in the protein binding site, starting from usually the largest core fragment. After having added a fragment in all possible conformations to the core, only the best placements are taken further ¹⁶³. This method has been invented as the anchor-and-growth algorithm does not handle conformational flexibility correctly. FlexX ¹⁶³ docks the first fragment based on chemical interaction models, instead of just steric complementarity, using a “pose-clustering” algorithm merging similar ligand conformations into the binding site. The base fragment must be large enough to make specific interactions with the protein, so that a preference in binding orientation can be determined. The increase in size of the base fragment is said to be directly correlated to the increasing probability of predicting its binding mode correctly ¹⁶³. Fragments forming H-bonds or salt bridges are added first, as these are directional interactions which are *per se* precisely defined. Improved FlexX versions dock hydrophobic ligands better ¹⁶⁴ and include discrete water molecules that could guide the positioning of the ligand ¹⁶⁵. Similarly to FlexX, Surflex also includes an incremental construction algorithm, combining the Hammerhead scoring function and a search engine that relies on a surface-based molecular similarity method as a means to rapidly generate suitable putative poses for molecular fragments ¹⁶⁶.

A *matching algorithm* is used in DOCK 4.0 to assist the anchor-and-grow algorithm in aligning structural/chemical features of both the ligand and the protein by performing complementarity pairing (donor/acceptor or positively/negatively charged).

Stochastic searches are either based on simulated annealing or genetic algorithms.

Simulated annealing introduces random thermal motion within a force field and then decreases the temperature and thus the thermal motion over time. The random motion of the ligand allows for the exploration of the global search space, while the decreasing temperature drives the system to a stable, (local) energy minimum. The most widely used simulated annealing procedure is the Metropolis-Monte Carlo simulated annealing algorithm representing the whole ligand by several parameters determining its location, orientation and conformation, which are randomly perturbed. The obtained score is kept if it is better than one of the previous round; otherwise it is kept with a probability depending on the difference in score and effective temperature (known as the Metropolis criterion), allowing to overcome small energy barriers.

Genetic algorithms such as the one of GOLD (Genetic Optimization for Ligand Docking)^{91, 167, 168} can be termed *in silico* evolutionary approaches. In a genetic algorithm, initial ligand conformations are encoded in chromosomes containing genes representing ligand translation, rotation, and torsion. The chromosomes can migrate among subpopulations to improve the conformational search and are assigned a fitness score estimating the free energy of binding. The higher the score, the likelier it is that the chromosomes are combined ("crossover") and/or altered ("mutation") to form child chromosomes, which in turn are evaluated for interaction energies. Iteration stops usually when a population "converges" on a docking pose. Genetic algorithms are generally effective at generating many low-energy conformers, with no guarantee that a global minimum is found. Already in the initial GOLD publication⁹¹, where a ligand was docked into its cognate binding site, reproducing the experimental binding mode was achieved in 71% of the cases. For fast docking speeds (1-2 min/compound), the Goldscore^{167, 169} predicts binding energies with a standard deviation of ca. 10.5 kJ/mol¹⁶⁸ as a sum of H-bonds derived from donor/acceptor pairs and geometries, internal VdW energies, and steric interactions between the ligand and the protein^{91, 167}. With respect to the active site "preference" of GOLD, it is well known that it performs best in docking ligands into hydrophilic active sites⁸⁴.

Deterministic approaches are expected to be reproducible and aim at reducing the energy of the protein-ligand system either by energy minimizations or MD. Energy minimization can lead to conformations which are trapped in a local minimum and never reach the global energy minimum of the energy surface. With MD, the global minimum can be reached, but it is difficult to estimate *a priori* how much calculation efforts must be made. In every docking program used in this thesis, ligand minimizations are involved.

A summary of the docking programs used, and their search algorithms is provided in Table 4, followed by a listing of their strengths, weaknesses, and purposes (Table 5) in each chapter of this thesis.

Table 4: Docking programs used in this thesis and their search algorithms.

Program	Type of search algorithm	Systematic			Stochastic	Deterministic
		Anchor-and-growth	Incremental construction	Matching algorithm	Genetic algorithm	Energy minimization
Dock	Systematic	x		x		x
FlexX			x			x
GOLD	Stochastic				x	x
Surflex	Mixed		x		x	x

Table 5: Strengths and weaknesses of the docking programs used in this thesis, adapted from ⁸³.

Program	Strengths	Weaknesses	Use in this thesis
<i>Dock</i>	Small binding sites; open cavities; small hydrophobic ligands	Docking and scoring of flexible or highly polar ligands	Chapter 3: Virtual screening on the p47 ^{phox} subunit of a NADPH oxidase
<i>FlexX</i>	Small binding sites; small hydrophobic ligands	Docking and scoring of very flexible ligands	Chapter 2: Systematic docking of rather rigid corticosteroids and similar molecules on an antibody surface
<i>GOLD</i>	Small binding sites; small hydrophobic ligands	Ranking very polar ligands; ranking ligands in large cavities	Chapter 1: Docking of thymidine kinase substrate analogues; chapter 4: Inverse virtual screening with a small ligand
<i>Surflex</i>	Large / open cavities; small binding sites; very flexible ligands	Low speed for large ligands	Chapter 3: Docking of celastrol

0.1.5 EXPERIMENTAL BINDING AFFINITY MEASUREMENTS

Because experimental affinity measurements are an important prerequisite and/or complement to *in silico* predictions, some of these approaches are summarized in Table 6 on the next three pages. Indirect methods with respect to ΔH include equilibrium dialysis, ultrafiltration, high-performance affinity chromatography, surface plasmon resonance, dual polarization interferometry, radiometric competitive binding assays, fluorescence spectroscopy, NMR, mass spectrometry (MS), and atomic force microscopy. Direct measurements of ΔH in contrast can be achieved with isothermal titration calorimetry (ITC) or differential scanning calorimetry.

0.1.6 CONCLUSION ON AFFINITY PREDICTION

The above discussion showed that striving for high affinity is a must in drug design, but that predicting affinity is still a challenge. Available *in silico* methods continue to be a compromise between accuracy and speed. On the one hand, the most exact methods based on thermodynamics are very calculation intensive. On the other hand, in virtual screening of thousands of molecules against a protein target, standard errors in the order of 10 kJ/mol can commonly be expected. The steered molecular dynamics-based affinity prediction method developed in chapter 1 for potential thymidine kinase substrates is an intermediate method in terms of accuracy and speed.

Table 6: The features of indirect and direct affinity measurements. a) Indirect methods with respect to ΔH .

Technique	Advantages	Disadvantages	Constants determined
Equilibrium dialysis	Simple, cheap, more specific than ultrafiltration	Low through-put, non-specific absorption, intermediate to high protein consumption	K_d
Ultrafiltration	Physiological protein and ligand concentrations, shorter measurement times than dialysis	Less precise K_d values than with equilibrium dialysis	K_d
High-performance affinity chromatography	Small amounts of ligand and proteins Quantitative and qualitative study of protein-ligand-interactions in solution Kinetic binding constants have been shown to be very close to those determined in solution ^{170, 171} .	Medium-throughput determination of ligand binding only	k_{on} , k_{off} and thus K_d via $\frac{k_{on}}{k_{off}} = K_d$
Surface Plasmon resonance (SPR)	Real-time monitoring of binding events in non-radiolabelled conditions	Immobilization needed	k_{on} , k_{off} and calculation of K_d
Dual polarization interferometry (DPI)	Real-time, high-resolution measurements of molecular size, density and mass. Quantification of any conformational change concomitantly with measuring reaction rates, affinities and thermodynamics	Analytical tools to detect relationship between conformational change and function are currently missing.	K_d
Radiometric competitive binding assays (fluorescence/UV)	Specificity, sensitivity	Radiolabelling is necessary	K_i
Fluorescence spectroscopy	The fluorescence change is proportional to the concentration of the moiety to which the fluorophore is attached (the ligand or the protein) and the maximal fluorescence change is proportional to the total concentration of the fluorescing species, rendering the calculation of binding constants possible.	Potential artefacts due to the auto-fluorescence of the ligand	K_d
Alpha Screen® Technology	High-throughput detection of binding, beads-based amplification of signal	The laser equipment needed for creating the beam for the excitation of the beads is costly.	K_d
Nuclear magnetic resonance (NMR)	Binding detected through chemical shifts or magnetization changes	High amounts of expressed protein needed	K_d
Mass spectrometry (MS)	Measuring e.g. the unbinding as a function of acceleration voltage ¹⁷²	Rather low throughput	K_d
Atomic force microscopy (AFM)	Protein-ligand unbinding forces can be converted to K_d values ¹⁷³	Very low throughput	K_d

Table 6: The features of indirect and direct affinity measurements. b) Direct methods with respect to ΔH .

Technique	Advantages	Disadvantages	Constants determined
Isothermal titration calorimetry (ITC)	Direct measurements of heat quantities upon complex formation, giving a clear picture of the basic physical forces implied and yielding the K_a (and thus the ΔG) and the stoichiometry. The heat capacity change is an approximate measure of the surface area buried in an association reaction and can be used to predict conformational rearrangements in associating protein molecules ¹⁷⁴ . Used to generate reliable data for the development of empirical scoring functions ¹⁷⁵ . A database containing ITC data including enthalpy, entropy, heat capacity (enthalpy change per temperature change), dissociation constant and 3D structure information was published: PDBCal ¹⁴⁹ .	Rather low throughput	ΔH , K_d
Differential scanning calorimetry (DSC) ¹⁷⁶	Incremental temperature changes are indicative of intramolecular packing and conformational changes. Used to generate reliable data for the development of empirical scoring functions ¹⁷⁵ .	Rather low throughput	ΔH , K_d

0.2 PROTEIN-PROTEIN INTERACTIONS

From a drug design perspective, targeting protein-protein interactions is much more challenging than addressing enzymes (receptors), for which most (substrate or transition state) inhibitors have been developed in the past ^{177, 178}. This has several reasons. Protein-protein interfaces are large, and it is often not intuitive to identify the key residues which could interact with a small molecular weight or peptide inhibitor. Moreover, the surface patches involved in protein-protein interactions are frequently not adjacent to each other, which is a challenge for designing peptide inhibitors. Because of the flatness of these interfaces, drug-like small molecules are however not always a good starting point for inhibitor development either. Lead compound identification is thus hampered in cases where no naturally occurring protein-binding small molecule exists ¹⁷⁹. The development of suitable biological assays to detect protein-protein interactions like Surface Plasmon Resonance (SPR), functional reporter gene or biophysical assays (ultracentrifugation or light scattering for assessing the quaternary structure) ¹⁸⁰ might also be tedious. Antibodies are facing the same challenges as protein-proteins interactions, but do have an advantage in that they combine a high degree of specificity with diversity, provided by the six complementarity determining regions (CDRs) and the shape and physicochemical properties of the epitopes, usually 4-7 residues in length ¹⁸¹.

Approaches used to tackle these challenges via modulating protein-protein interactions involve direct inhibition (e.g. via an inhibitor mimicking the protein binding partner or a competitive inhibitor interfering with the formation of the interface) or indirect inhibition (i.e. allosteric binding).

Protein-binding peptides can modulate (i.e. activate or inhibit) protein networks. When modifying a binding peptide, several strategies exist: (i) Optimizing the core or (ii) the flanking interactions, thereby mostly changing side chains or less frequently the backbone of a peptide. Considering flanking interactions, Sood et al. ¹⁸² demonstrated *in silico* that a simultaneous optimization of 3 to 9 additional residues recovers both the conformation and the sequences of naturally occurring binding peptides. These predictions were experimentally confirmed for p53 peptides binding to the N-terminal domain of the human oncoprotein Mdm2. The authors suggested that better peptide affinities could be achieved with a protein target displaying a well-defined hydrophobic pocket and a less flat surface, as well as with enhanced conformational sampling of the correct backbone. This might be achieved by incorporating flexibility via an ensemble of conformers collected from explicit solvent MD simulations ¹⁸³.

Peptide-based inhibitors of dimerisation are used to modify regulatory processes or to inhibit enzyme activation, especially in the context of antibiotic, antiviral and antiparasitic drugs ³³. Examples for peptidic antivirals are inhibitors of the HIV-1 protease, reverse transcriptase and integrase ¹⁸⁴⁻¹⁸⁷ or the DNA polymerase of herpes simplex virus ¹⁸⁸. They generally resemble dimer interfaces and have mostly low mM, but sometimes also low nM affinities ^{189, 190}. Besides, stapled peptides mimicking α -helices ^{191, 192} can be used as dimerisation inhibitors (0.2.4). Such peptides are characterized by efficient cell penetration, good pharmacokinetics, high affinity binding to large target proteins and excellent stability.

In contrast, dimerizers assist the formation of a protein-protein complex by binding with low nM affinity ¹⁹³ to each of the proteins with a specific anchor group being connected to a linker of 5-16 atoms ^{193, 194}, its length depending on the specificity requirements. Such dimerisers are thought to have applications as inducers of proliferation ¹⁹⁵ or enhancers of targeted apoptosis (e.g. in gene therapy) ^{193, 196}.

The preceding examples clearly show that protein-protein interface assembly can be blocked with either a small molecule or a peptide to modulate protein activity. Their development is supported by methods dissecting interaction hot spots ¹⁹⁷, recognizing characteristic motifs ¹⁹⁸ and mapping the interface ³³.

Targeting protein-protein interactions has several advantages and disadvantages, which are discussed in the following. Marketed drugs and research successes ¹⁹⁹ have convincingly shown that it is feasible to inhibit protein-protein interactions, but success preceded know-how ²⁰⁰ on the mechanism of action. An example are the vinca alkaloids used since the 1950s in the treatment of cancer, which were only shown in 1996 to bind to β -tubulin controlling thereby the polymerization of α - β -tubulin and mitotic spindle formation ²⁰¹, and therefore exerting an antimitotic action.

The conservation of protein modules involved in signalling networks suggests that protein interfaces have been optimized by evolution for function²⁰². The properties of such modules determine the specificity of binding^{203, 204}, with affinities ranging from as low as millimolar to as high as femtomolar²⁰⁵. This substantial spread of affinity might be an advantage for modulating cell signalling but can also be a disadvantage for drug design. However, new modes of actions^{179, 199} are discovered, examples being macrolactones directed against proline-rich regions of regulatory domains of several kinases²⁰⁶ or peptide antagonists of erythropoietin^{207, 208}. Therapeutic areas with an involvement of protein-protein interactions of the proline-rich type include e.g. cancer, allergy, asthma, osteoporosis and viral illnesses^{209, 210}.

0.2.1 INTERFACE TYPES

When looking at a protein-protein interface, complementary chemical properties are found, which facilitate molecular recognition. At the interface, functionally important residues are usually conserved among a series of proteins, while residues determining specificity are distinct. Most conserved polar residues make the interface rigid, minimizing binding entropy, due to the decrease in conformational flexibility²¹¹. Native and near-native interfaces can be computationally predicted.

While assessing an interface, one needs to consider a series of factors. First, its druggability depends on its 3D structure and physicochemical characteristics. Hajduk et al. presented a method for generating druggability indices derived from NMR-based screening data. Advantageously, the entire protein can be monitored for perturbations upon adding a low molecular weight drug to the protein surface²¹², providing information on binding site location and druggability. The contact surface of protein-protein interfaces might involve from 6 up to 30% of a monomer surface, varying from 550 to 4900 Å²³³, with an average of 800 Å²^{213, 214}. Concerning the composition of such surface patches, several amino acids occur more frequently: Arg, His, Asn, Trp, Tyr, Ser^{214, 215} and hydrophobic ones^{213, 216, 217}, having a secondary structure composed of 47% random coils, 36% α -helices and 17% β -sheets²¹³. Second, it has to be known whether the participants in the interface are homo- or hetero-oligomers²¹⁸, oligomers being the native state of proteins²¹⁹. This information is important because homodimer interfaces are generally larger than heterodimer ones. Further, interfaces larger than 1500 Å² tend to undergo conformational changes upon complexation^{220, 221}. The burial on both sides of the interface has been estimated to lie between 750 and 1500 Å²²²². However, a sample of 20 marketed drugs targeting interfaces features solvent-accessible surface areas in the range of 150 to 500 Å²²⁰⁰ only, which means probably that the above estimation of burial is too high. Third, it must be assessed whether the interaction is permanent or transient. The contact surfaces of permanent stable complexes display more hydrophobic interfaces²¹⁸ or resemble domain contacts^{33, 223}. Temporary complexes in contrast exhibit either (i) a well-defined hydrophobic core surrounded by a ring of polar groups including water or (ii) short hydrophobic patches, polar amino acids and intersubunit H-bonds²²⁴. Membrane proteins involved in transient complexes such as cytochromes or NADPH-specific flavoproteins constitute somehow an exception, as these interactions are formed by hydrophobic interactions with the membrane^{216, 217}. Such hydrophobic cavities can be exploited for gaining affinity for e.g. peptide inhibitors (chapter 3).

The flatness of an interface needs also to be considered. The questions arising are: Is there a chance for one protein to be partly buried in the other to establish a stronger interaction? Are there some cavities to be explored? Cavities represent only about 10% of the total interface³³. Analyzing the flexibility and complementarity of the interface might help, owing to the fact that homodimers and permanent heterodimers are known to be more complementary than transient heterodimers²¹³. Therefore, the first two should be more druggable than the last. In other words, "recognition spaces"¹⁶⁹ defined by e.g. the residue interface propensity, hydrophobicity, planarity, and solvent-accessible surface area²²⁵ are particularly striking in true homodimers. With respect to structural rearrangements upon binding, pre-existing conformation equilibria in monomers are likely to guide the selection of protein-protein interactions²²⁶. Information on water molecules maintaining close packing could hint at an appropriately designed inhibitor replacing these water molecules.

0.2.2 HOT SPOTS

Different protein regions on the surface contribute to a different extent to binding. The concept of unbound, pre-organized binding hot spots being able to act in an additive manner²²⁷ supports the

uneven distribution of residues important for binding. It was introduced by Clackson and Wells²⁰⁷. An extension with alanine scanning mutagenesis data (www.asedb.org²²⁸) by Bogan and Thornton²²⁹ showed that (i) the binding free energy of side chains is poorly correlated with the change in buried surface area; (ii) the sequestration of residues from bulk solvent is a necessary condition for large energy contributions; and that (iii) Trp, Tyr, and Arg are significantly enriched in hotspots. Alanine scanning mutagenesis allows for the discovery of a hot spot residue where a mutation of an amino acid to Ala gives rise to a typically tenfold or higher drop in affinity¹⁹⁷. Such a hot spot mutation can be seen as a perturbation, i.e. a loss of optimal VdW contacts, electrostatic pairings, buried non-polar surface area, a small conformational change, a local unfolding, an increased entropy of the unbound states, aggregation or global unfolding¹⁹⁷. Using alanine scanning, it was shown that hot spot residues usually cluster at the centre of interfaces²²⁹, possibly indicating that peripheral side chains are more easily replaced by e.g. water and that occlusion of bulk solvent leads to high affinity (as in the case of biotin-streptavidin, 0.1.2.2).

Because hot spots contribute the most to the binding affinity, they are crucial for recognition and subsequent binding¹⁹⁷. Of course, this is only seen to some extent when comparing crystal structures of the same target bound to different ligands, where the same amino acids interact in different 3D situations, implying conformational changes in loops and hinges. One example is the Fc region of IgG1 with different peptides bound (PDB ID: 1DN2). The peptides bind to the same favoured hot spot. This hot spot displays an involuted, hydrophobic core and forms H-bonds as well as salt bridges towards the periphery, adapting to the ligand which is bound²³⁰. As a general rule, it has been noted for hot spots in protein-protein interfaces that the core residues are more buried than the surrounding rim²³¹. Despite the hydrophobic character of protein surfaces, polar interactions play an essential role in hot spots²⁰². Having studied 97 protein-protein interfaces, Hu et al.²⁰² demonstrated that (i) over the total number of interface residues, hydrophobic residues have higher propensities than polar ones and (ii) that in terms of conserved residues, polar residues dominate. Conserved polar (His, Asn, Gln, Thr, Ser) or partially polar residues (Phe, Met) are more enriched in this study than in alanine scanning experiments. The most conserved polar residues confer rigidity to minimize the entropic cost on binding, whereas the surrounding residues provide a flexible cushion²¹¹. There is some disagreement in the literature with respect to which residues are mostly found in the core, but Trp, Arg, Tyr and His could be the consensus. Larger interfaces tend to display more dispersed hot spots with flexible regions in between than smaller ones. Looking at protein-protein interfaces in different protein families, affinity and specificity were found to not necessarily be coupled²¹¹.

A potent inhibitor is expected to cover most of the area of the hot spot, and to have a better or equal shape complementarity to the protein than its protein partner. In other words, it should contain chemical groups which mimic the key interactions (the most important ones for ΔG) made by the competing subunit and chemical groups¹⁷⁸. Small molecules are found to bind most readily to the deepest cleft on the surface²³³. These new contacts are favourable in terms of enthalpy, leading to a better affinity.

Examples of hot spots identified in protein-protein interfaces can be found in¹⁹⁷. Nevertheless, not all interfaces possess hot spots (e.g. no hot spots are found in an interferon binding to its receptor, and protein L binding to an immunoglobulin light chain^{234, 235}). In these cases, main-chain interactions are predominant.

0.2.2 UNRAVELING THE FORCES BEHIND PROTEIN-PROTEIN INTERACTIONS

For protein-protein interactions, the dissociation constant K_d is commonly in the interval between 10^{-4} and 10^{-14} M, i.e. 4 orders of magnitude broader than for small molecules and corresponding to a difference in ΔG of 6-19 kcal/mol²³⁶. The forces governing protein-protein interactions include the hydrophobic effect, electrostatics, steric effects, and VdW attraction/repulsion. Hydrophobicity in protein-protein interactions is treated separately in 0.2.2.1. Electrostatics play also a considerable stabilizing role²³⁷⁻²³⁹, considering that there are up to 12 charged residues per interface²⁴⁰. In the case of charged residues, desolvation can be compensated because of the so-called electrostatic complementarity²⁴¹, which is due to charge-charge and hydrophilic stabilizing interactions with the surroundings²⁴⁰. It was postulated that this type of complementarity determines the complex lifetime²⁴². The involved surfaces are largely complementary in shape and characterized by the juxtaposition of amino acids with opposed character. Complementarity is necessary, but dynamic side chain arrangements upon binding can compensate for it to some extent. 56% non-polar, 29% neutral, and

15% charged groups are thought to make up protein-protein interfaces ²²⁰, which are generally more rigid than the remaining surface ²⁴³.

Regarding H-bonds, sidechains are mostly involved. Their number is proportional to the contact area of each participating subunit, being in the order of 10 per interface and displaying normal or weak strength ²⁴⁴. Water molecules forming networks at the interface further stabilize it by “shielding” charges and increasing shape and charge complementarity ²⁴¹. Further insight into the role of interstitial water can be gained from ²⁴⁵.

0.2.2.1 HYDROPHOBICITY

Upon binding, changes in ligand and protein solvation contribute essentially to the binding energy, because in water, large hydrophobic surfaces are energetically unfavourable (displaying energies of 12-30 cal/molÅ ²⁴⁶) and are therefore shielded as much as possible from bulk solvent, a phenomenon that is known as “the hydrophobic effect”. There is a cooperative effect of H-bonds when hydrophobic residues are buried. The magnitude of the hydrophobic effect can e.g. be assessed by solvent transfer experiments, where the solubility is measured as a function of temperature, free energy calculations or the measurement of binding constants for congeneric ligands differing in e.g. a methyl group ¹³¹. A frequently used method relies on the calculation of the SASA. Hydrophobicity is often overestimated using this approach. In fact, only hydrophobic surface regions that are not covered by a first hydration shell of water molecules can interact with bulk water ²⁴⁶. By taking the first hydration shell into account, the polar parts and their hydrophobic surroundings are shielded, reducing the SASA from 60% to 40% ²⁴⁷ and disrupting thereby the continuous hydrophobic surface area. Consequently, only the hydrophobic patches still accessible can interact with another protein or peptide.

The role of aromatic π - π interactions between aromatic ring systems in hydrophobic interactions is discussed next. As early as in 1985, these interactions were analyzed in terms of geometry, non bonded interaction energy, secondary structure location and their environment, with the main findings still valid today: (i) about 80% of aromatic side chains interact with 3 or more aromatic side chains, (ii) phenyl centroids are separated by 4.5 to 7 Å, (iii) dihedral angles approaching 90° are most common, and (iv) a typical π - π interaction energy is in the range between -1 and -2 kcal/mol (energetically favourable), depending on the environment of the aromatic pair ²⁴⁸. The high incidence of π - π interactions stated above speaks in favour of their protein stabilizing role. However, solvent-accessible aromatic side chains can only pair by displacing one or more water molecules of the first hydration shell ²⁴⁹.

Unravelling the nature of π - π interactions ²⁵⁰ is important due to their influence on binding affinities ²⁵¹, DNA base stacking, and tertiary structure stabilization upon folding in proteins ²⁵². A competition between face-to-edge (aromatic rings perpendicular), face-to-face (edge-displaced ²⁵³, i.e. with the rings parallel displaced), or sandwich (i.e. with the rings superposed) arrangements is usually found. π - π interactions can for example be investigated with a combined MD/*ab initio* approach. Gervasio et al. ²⁵⁴ searched the potential energy surface of pairs consisting of Trp, Phe or Tyr for local minima and performed then *ab initio* binding energy calculations. They stated that electrostatic and VdW interactions rationalize interactions among aromatics, while charge transfer and polarization phenomena are small. The competition of stacked and T-shaped interactions was studied in detail by MD followed by PMF calculations and compared to experimentally resolved crystal structures, revealing that Tyr-Tyr interactions are predominantly found on the surface, while Phe-Tyr and Tyr-Tyr interactions are more frequent in the protein hydrophobic core ²⁵⁵.

0.2.3 CHOOSING A PROTEIN-PROTEIN INTERFACE TO TARGET

Protein-protein interaction sites can e.g. be predicted by detecting sequence similarity and projecting the target sequence onto a homologous 3D structure, phylogenetic analysis of functionally linked proteins or mapping of metabolic pathways. Aspects facilitating the choice of a suitable target interface include according to Gadek et al. ²⁰⁰ the surface area of the interface, its amino acid composition, particularly the existence of hot spots, the effect of long-range electrostatics on recognition and kinetics, and the characteristics of the interface that may allow the binding region to be located without direct experimental knowledge of the complex geometry. The important role of long-range electrostatics in protein-protein recognition is reviewed by Janin et al. ²⁵⁶: Electrostatic (i.e. charge-

altering) mutations outside of the binding site increase k_{on} without affecting k_{off} , provoking diffusion-controlled protein-protein associations²⁵⁷. Nowadays, identifying homodimers interaction sites without information on the interface can be achieved to an accuracy of over 70%, taking the propensity of amino acids, the accessible surface area, a protrusion index, the planarity of the interfacial region, and hydrophobicity into account²¹³.

A possible strategy to select protein-protein interfaces as drug targets considers 1) the structure, 2) the cavity, 3) the hydrophobicity, 4) the size, and 5) the complementarity¹⁷⁸. At protein-protein interfaces, pockets can open and close over time. With respect to point 1), the comparison of apo and holo proteins is important to identify amino acid burial. In respect of cavity identification (point 2), pocket openings in protein-protein interfaces can e.g. be assessed by applying a pocket detection algorithm to transient pockets identified from MD snapshots. In this way, Eyrisch et al. identified transient pockets involved in protein-protein interactions²⁵⁸ and investigated the influence of solvent polarity and backbone rearrangements on the formation of pockets on protein surfaces²⁵⁹. They found that large pockets not detectable in the crystal structures of the apo protein opened frequently on the protein surfaces during standard 10 ns MD simulations at room temperature. The pockets detected from MD snapshots with a binding site prediction algorithm were then used to successfully dock inhibitor molecules (RMSD < 2 Å compared to crystal pose). With respect to 3), stability of the dimer and interaction affinity is increased with the number of hydrophobic substituents. This can be explained by the hydrophobic effect (0.2.2.1), but still some solubility should be kept to avoid unspecific aggregates. It is well-known that backbone H-bonds are essentially shielded from the surrounding solvent by hydrophobic residues. According to a finding of Fernández et al.²⁶⁰, the backbone H-bonds which are still solvent-accessible display a so-called packing defect in that they are not wrapped, but undergo a considerable stabilization with water. The size (4) and complementarity (5) of the interface have already been discussed above (0.2.1).

Summarizing, for the interface to be druggable, it should be able to accommodate an inhibitor having the right size to interact with key residues and being small enough for a good bioavailability.

0.2.3.1 BINDING SITE PREDICTIONS

Binding site predictions (reviewed in²⁶¹ and²⁶²) can be based on geometric/energetic criteria or docking and scoring of fragments or small molecules.

Geometric methods do not provide a direct physical meaning and rely usually on shape descriptors²⁶³⁻²⁷². Protein cavities can also be detected by a grid-based algorithm, where a grid point lies in a cavity if it has at least two neighbouring grid points also located in this cavity²⁷³⁻²⁷⁶. Conveniently, the DrugSite algorithm allows for detecting potential binding sites in both apo- and holo 3D structures²⁷⁷.

An example of an energy-based method is Q-SiteFinder²⁷⁸, where the interaction energy between the protein and a VdW probe is clustered according to spatial proximity and ranked according to the sum of interaction energies. The energy-based program GRID²⁷⁹ in turns assesses the propensity of a set of probes (functional groups) to interact with the protein surface. It is known to work well for enzymatic binding pockets but has not often been applied to protein surfaces. One promising application of GRID to SH3 protein domains yielded a 100-fold higher affinity and 1000-fold better selectivity for novel proline-rich peptides targeting Abl-SH3 and Fyn-SH3²⁸⁰. Another application of GRID to a peptide binding to a tandem SH3 domain of a NADPH oxidase is presented in chapter 3.

Both conserved and distinct, less frequent binding site features among different binding pockets can be found with e.g. "inexact graph-matching"²⁸¹, where the functional relationships among proteins are detected solely based on structure. As representations of the geometrical arrangement and physicochemical properties of a binding site, pseudocentres are used, i.e. a compressed representation of areas on the cavity surface where protein-ligand interactions occur. Such descriptors, which are nodes in a graph, are affected by mutations or conformational flexibility in the cavity. The discriminative power of this method applicable to selectivity issues was convincingly shown for the closely related serine proteases trypsin and thrombin, where two features only present in thrombin were found²⁸¹.

Among geometric- or energy-based binding site prediction methods assessing the size and shape, only few check for druggability, i.e. the propensity to accommodate drug-like molecules (reviewed in

²⁸²), e.g. DrugSite ²⁸³ or PocketPicker ²⁸⁴. As an example, the PocketPicker algorithm translates the shape and buriedness of potential binding sites into autocorrelation vectors ²⁸⁵ allowing for efficient alignment-free pocket comparisons. Suitable descriptors are derived from crystal binding pockets of known drugs with available affinity data or apo structures using self-organizing map clustering and are also applicable to orphan targets.

Force field-based fragment mapping or docking have a better physical basis than geometry-based approaches but use more computational resources and sometimes show difficulties to discriminate correct binding sites. Once a (druggable) cavity is identified, it can be aligned and compared to other druggable binding sites with e.g. SiteAlign ²⁸⁶. This is a cavity fingerprint based on 8 topological and physicochemical properties, which is robust enough to detect local similarities among active sites of different sizes, discriminate between protein subfamilies, and recover the known targets of promiscuous ligands for virtual screening ²⁸⁶. Alternatives are sequence and fold-independent 3D cavity comparison tools ²⁸⁷⁻²⁹³. Because none of the many existing methods is perfect, a consensus has to be found among several binding site predictions.

When combining the conservation of amino acids, solvation, electrostatics and hydrophobicity, about 70% of correct binding site predictions were achieved in several studies ²⁹⁴⁻²⁹⁷. Such predictions are based on the assumption that structural and functional similarity correspond ²⁹⁸ as a result of evolutionary pressure. On the fold level, similarity does not always correlate with functional similarity ²⁹⁹. To overcome this problem, molecular surfaces (such as the Connolly surface ³⁰⁰) can be used to search for similarity in well-studied, orphan or hypothetical, i.e. non-functionally annotated proteins ^{287, 291, 301, 302}. An extension of such an approach, also taking advantage of electrostatic surface potentials ³⁰², is accessible at <http://eF-site.hgc.jp/eF-seek/>.

Even if the number of 3D protein-protein crystal structures increased with the Structural Genomics project ³⁰³, it is unfortunately still very low ³⁰⁴ compared to the number of protein networks identified by MS ³⁰⁵. To overcome this gap, docking methods aiming at predicting the structure of protein-protein complexes based on individual 3D structures are valuable.

0.2.3.2 PROTEIN-PROTEIN DOCKING

Protein-protein docking is usually applied to the following questions: Assuming that two proteins bind *in vivo*, in which way and how strong do they interact with each other? If on the contrary they do not bind, what is the underlying reason and how can non-binders be converted into binders? As protein-protein docking is relying on physical principles, even proteins of unknown function can be docked, if their structure has been determined experimentally or theoretically. To limit the conformational space to be searched, rigid body docking is mostly performed and the stability of local minima tested ³⁰⁶. In spite of the considerable backbone and side-chain mobility of proteins, incorporating protein flexibility can be achieved with MD snapshots, rotameric searches and principal component analysis (PCA) ³⁰⁷.

Protein-protein docking follows a 2-step protocol: (i) An initial global search, followed by subsequent steps that improve the initial predictions of contacts ³⁰⁸ and (ii) confining the conformational search by usually keeping the larger protein fixed, similarly to the protein in protein-ligand docking. Then, the obtained docking solutions are clustered, re-ranked, and refined. In this respect, a combination of rescoring and refinement was shown to significantly improve protein docking performance ³⁰⁹. Benchmarks ³¹⁰ for the testing of several docking programs are available and international tests (CAPRI - Critical Assessment of PRediction of Interactions, <http://www.ebi.ac.uk/msd-srv/capri/>) are done on a regular basis.

Examples of protein-protein docking program are Haddock, an acronym of High Ambiguity Driven protein-protein DOCKing" ^{311, 312} or the online protein-protein docking tool RosettaDock, requiring two input proteins and a starting location for the search, yielding low-energy conformations (<http://rosettadock.graylab.jhu.edu>) ³¹³.

Firedock (<http://bioinfo3d.cs.tau.ac.il/FireDock>) ³¹⁴ is the first web server for flexible refinement and scoring of protein-protein docking solutions. It was shown that the docking protocol was successful in refining and rescoring solutions from benchmark cases. Its main steps include side-chain optimization, rigid-body minimization, scoring, and re-ranking to identify the near-native refined solutions.

Because the interface surface is large, it is not always trivial to identify the potential binding site. In such a case, MD is the method of choice for predicting or refining a binding site and for assessing the flexibility and exposedness of amino acid side chains of potential hot spots i) alone, ii) together with a protein binding partner or iii) a potential inhibitor. As noted by DeLano¹⁹⁷, given that no general algorithm has been developed yet to predict binding hot spots on the basis of their shape or composition only^{213, 222}, MD may be essential for predicting binding interface energetics considering structural plasticity³¹⁵.

0.2.3.3 EXPERIMENTAL APPROACHES TO ASSESS THE BINDING REGION AND TO MEASURE BINDING AFFINITY

Experimental approaches to study especially small molecule inhibitors of protein-protein interactions and inhibitors acting on upstream enzymes include e.g. cell-based translocation assays. Alternatively, so-called complementation assays are more tuned towards a signal upon protein-protein association^{316, 317}. After a successful assessment, experimental validation to understand the binding region can start. Co-immunoprecipitation, cDNA expression libraries, a yeast-2-hybrid³¹⁸ or yeast-3-hybrid screen³¹⁹, site-directed mutagenesis or peptide mapping confirms the hot spot residues. Using any type of chemical library, the search for competitive inhibitors might be based on Enzyme-Linked Immunosorbent Assays (ELISA), fluorescence or radiolabelling. Otherwise, binding assays might be performed on the target protein with analytical ultracentrifugation, SPR (e.g. Biacore) or ¹H-¹⁵N HSQC NMR (Heteronuclear Single Quantum Correlation NMR, indicating the backbone N-H bonds), yielding information on binding. With NMR, detection of strong binders is less of a problem, as the signal-to-noise ratio is high. One can also envisage to determine the stoichiometry of binding between a protein and a small molecule with $K_d < 1 \mu\text{M}$ by analytical ultracentrifugation³²⁰.

0.2.4 SUCCESSFULLY TARGETED INTERFACES

A survey on small molecule protein-protein inhibitors revealed that the chemical space of active compounds is accessible from chemical databases for well-studied targets only, while for others the diversity space needs to be expanded in a focused way³²¹, e.g. by diversity-oriented synthesis³²². In the following, examples of interfaces which have been successfully targeted with low molecular weight compounds or peptides are given.

Small molecular weight compounds addressing interfaces have been extensively reviewed^{179, 180, 199, 200, 323}. At first sight, such an inhibition of a protein-protein complex seems unlikely, provided the difference in size between a small molecule and a protein, but the small molecule-protein association needs to be seen as an equilibrium process²⁰⁰. Cochran et al. explained that the disruption of the protein-protein interaction by a small molecule has either a favourable enthalpic component (by stabilizing the remaining protein) or an entropic component favouring the exchange of the protein interaction partner by a small molecule³²⁴.

Mostly high affinity, but also low affinity protein complexes exist, e.g. cytokines^{325, 326} and components of electron transfer complexes that are signalling effectors with high charge complementarity. The antifungal and anti-angiogenic compound UCS15A was presented in 2002 as the first non-peptidic, small molecular agent capable of disrupting SH3-mediated protein-protein interactions, e.g. in Src³²⁷, holding promise as a lead for targeting proline-rich ligand-mediated protein-protein interactions²⁰⁶. UCS15A inhibited proline-rich SH3 interactions, but did not disrupt pre-existing complexes. It is assumed that *in vivo*, free proline-rich SH3 proteins may be far fewer than *in vitro*³²⁸. Atareh et al.³²⁹ performed virtual screening for proline-rich inhibitors of Src-SH3 using GOLD⁹¹ and the ZINC database³³⁰. Experimental validation of top scorers by fluorescence polarization revealed a benzoquinoline with μM affinity, which was used as a starting point for SAR. It was found that the benzoquinoline makes interactions with the surface of the SH3, and that an aliphatic linker and H-bond acceptor alkoxy groups on phenyl are essential for activity. Allosteric inhibitors of iNOS dimerisation, the mammalian inducible nitric oxide synthase, catalyzing the oxidation of L-arginine to NO and L-citrulline in the presence of NADPH^{331, 332}, are based on a pyrimidine-imidazole scaffold. Binding of an inhibitor to heme disrupts the functionally important dimerisation^{333, 334} (PDB ID: 1DD7). Another example is the nerve growth factor (NGF), a homodimer causing neurite growth or cell death depending on the receptor associated with it. Selective binding of the antagonist Ro 028-2750 causes a conformational change of one monomer at low, and of both monomers at high concentrations³³⁵.

Competitive antagonists of the interleukin 2 receptor, the principal mediator of the T-helper immune response, are other examples. The amphipathic binding site of this receptor displays a largely hydrophilic/acidic and a hydrophobic/basic side, to which a cytokine or an antagonist can adapt. Intriguingly, affinity and specificity can be handled independently within such an interface ³³⁶.

Peptide inhibitors of dimerisation are reviewed by Berg ¹⁷⁹, and shortly mentioned in the following. Because the catalytic substrate-binding pocket of HIV-1 protease is only formed in the homodimer, mimicking the interdigitating termini of the enzyme by peptide sequences is a way to counteract dimerisation. Another well-known target is the tumor suppressor p53 – oncoprotein hdm2 interface. In non-cancerous cells, p53 is the guardian of genomic integrity, while in many cancer cells; p53 is inactivated through interaction with the overexpressed hdm2. Peptides disrupting their interaction are rich in α -helices and mimic the p53 binding to a hydrophobic groove of hdm2.

The frequently encountered protein-protein interfaces that utilize an α -helix binding groove might be particularly amenable to small-molecule drug discovery ³²³ because they mimic the physicochemical properties of an α -helix. Organic molecules with helical shape include e.g. biphenyls, allenes, alkenene cycloalkanes, spiranes ³³⁷, terphenyls ³³⁸ or stapled peptides ^{191, 192} (Figure 7).

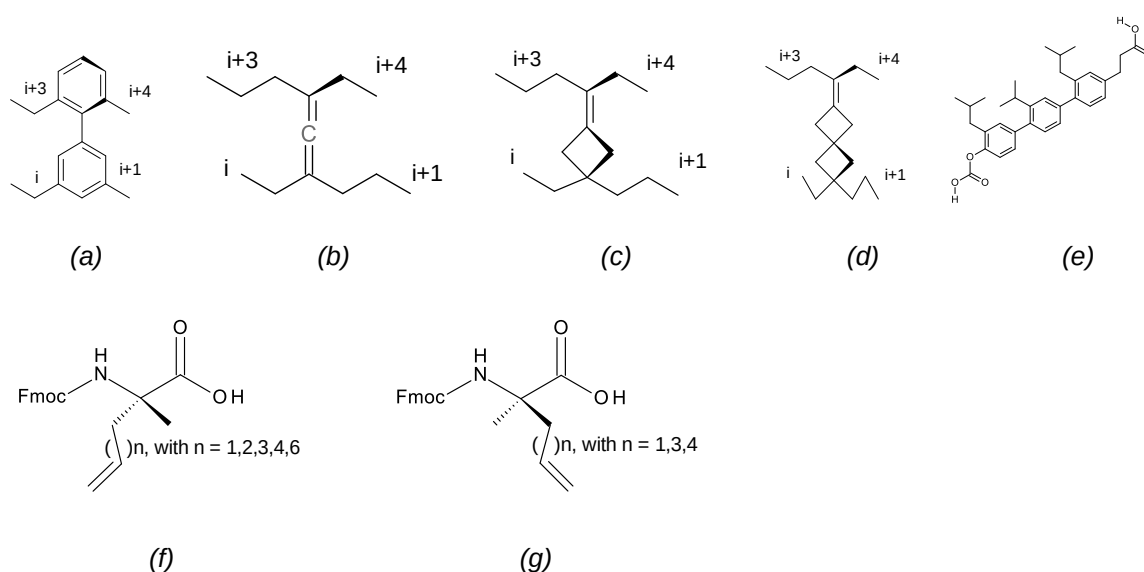


Figure 7: Organic molecules with helical shape include e.g. biphenyls (a), allenes (b), alkenene cycloalkanes (c), spiranes (d), terphenyls (e) or stapled peptides (f-g). One turn of an α -helix goes from i to $i+3$. (a)-(d) were adapted from ³³⁷, (e) from ³³⁸ and (f-g) from ³³⁹.

As an example, terphenyl derivatives disrupt the hexameric assembly of gp41 that would lead to HIV-1 fusion with host cells ³³⁸ or inhibit the binding of the helical region of BAK to the B-cell lymphoma anti-apoptotic protein Bcl-X_L ^{340, 341}. An NMR structure provided evidence for the α -helical BAK peptide binding to Bcl in a hydrophobic groove ³⁴⁰. Subsequently, evidence for small molecules inhibiting this association was provided with virtual screening ^{342, 343}, high throughput screening ³⁴⁴, ligand-based design ³⁴¹, fluorescence changes and ITC ^{345, 346}, while HSQC-NMR and SARs ^{344, 347} were performed for related compounds. An overview of the diversity of such small molecular weight inhibitors is provided by Zeitlin et al. ³⁴⁸.

A successful *in silico-in vitro* approach of inhibiting a protein-membrane interaction, involving the discoidin C2 domain of coagulation factor Va and the plasma membrane ³⁴⁹, was disclosed by Sperandio et al. ³⁵⁰ and included docking, functional assays and SPR measurements.

0.2.5 SH3: A CONSERVED PROTEIN DOMAIN

Eukaryotic signalling is characterized by the separation of catalysis from target recognition ³⁵¹. Conserved modular protein domains which interact cooperatively are key regulators of signal transduction. Upon binding of a molecule (protein, peptide or small non-peptidic compound) to a protein target, the latter is activated and promotes signalling leading to a cellular effect affecting cell

cycle progression, gene expression, cytoskeletal architecture, protein trafficking, adhesion, migration, and metabolism²⁰³. The essential players in this game are 50-100 amino acids long, functional domains.

In the following, only SH3 domains are discussed. Peptide sequences mimicking the natural binding partner of a tandem SH3 domain in the phagocyte NADPH oxidase are designed for higher affinity in chapter 3. The modular nature of SH3 domains in this enzyme³⁵² is known to play an important role in recruiting several cytosolic proteins required for enzyme activation. SH3 domains show a preference for conserved proline-rich motifs, where the surrounding amino acids determine the remaining binding specificity. The amino acid Pro is unique with its cyclised side chain conformation, providing highly discriminatory recognition without requiring high affinities³⁵³. SH3 domains consist of three loops separated by β -sheets: the RT loop containing critical Arg (R) and Thr (T) residues, the n-Src loop (named after a long insertion in the neuronal form of Src), and the distal loop. SH3 domains participate in tyrosine kinase signalling, cytoskeleton arrangement, and NADPH oxidase assembly. Peptides binding to SH3 are composed of about 10 amino acids with a preference for Pro^{354, 355} displaying K_d values in the range of 5-100 μ M³⁵⁶. Crystallographic evidence^{355, 357} shows that they form left-handed type II helices with three amino acids ("spines") per turn, two contacting the SH3 and one for stabilization²⁰³. SH3 binding peptides can bind in two different orientations owing to their pseudo-symmetry: N- to C-terminal (class I) or C- to N-terminal (class II). In this manner, the number of potential binding partners is increased, and the complex formation is organized in a distinct way. However, each SH3 has a preference for a specific binding orientation³⁵⁵. Specificity is ensured by interactions between non-prolines in the peptide and two SH3 loops surrounding the hydrophobic core.

0.2.6 CONCLUSION ON PROTEIN-PROTEIN INTERACTIONS

Although protein-protein interfaces are less well defined and more challenging than enzyme binding sites, this part of the introduction reviewing the literature highlights that targeting such interfaces with small molecules or peptides is feasible. Chapters 2 and 3 of this thesis add some further evidence.

0.3 CHEMINFORMATICS

The term cheminformatics is now over 10 years old but is arguably just a new name for a known field at the interface of chemistry and informatics. In 1998, Brown stated that the use of information technology and management has become a critical part of the drug discovery process³⁵⁸. Cheminformatics is the mixing of those information resources to transform data into information and information into knowledge for the intended purpose of making better decisions faster in the area of drug lead identification and optimization³⁵⁸. Later on, Manly et al. highlighted the impact of informatics and computational chemistry on synthesis and screening³⁵⁹. The way cheminformatics proceeds can be illustrated by a "knowledge pyramid"³⁶⁰ (Figure 8): From data (the basement), information is obtained and knowledge gained (the tip of the pyramid) along the axis numbers-facts-rules.

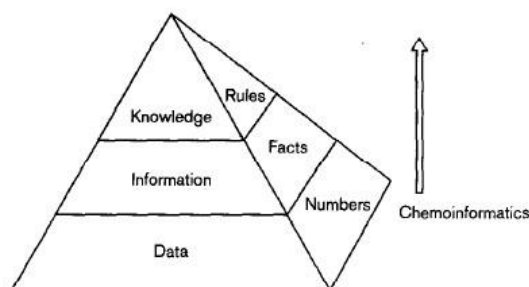


Figure 8: The knowledge pyramid. The Figure was taken from³⁶⁰.

Being based on information technology, the discipline of cheminformatics relies on compound libraries, representative target libraries and complementary experimental validation. Its basic principle is exemplified by completing a 2-dimensional matrix (Figure 9) containing binding constants (K_i or IC_{50}

values) or functional effects (e.g. EC₅₀ values). Each row represents different compounds being able to bind to one protein target and each column shows different targets of one compound. In this way, target or ligand selectivity can be assessed in a comparative manner.

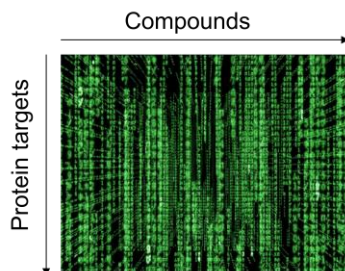


Figure 9: The 2-D matrix to be filled with information in cheminformatics. The matrix picture came from http://www.filetransit.com/images/screen/53ed891374a1cbfb63e1e7d11471004d_Matrix_Mania.jpg and is part of the movie “The Matrix” (USA, 1992).

Ultimately, cheminformatics attempts to match the chemical with the target space. In general, a distinction between ligand- and target-based cheminformatics is made. Both approaches are reviewed, while applications thereof can be found in chapters 2 to 4 of this thesis.

0.3.1 LIGAND-BASED CHEMINFORMATICS

In the search for new molecular entities (NMEs), a complete sampling of the chemical space is impossible, given that the number of putative compounds with a molecular weight < 500 Da is estimated to 10²⁰⁰ (as a comparison, the number of atoms in the universe is estimated to 10⁷⁸ - 10⁸²), out of which 10⁶⁰ are assumed to be drug-like³⁶¹. In ligand-based cheminformatics, one starts from annotated ligand libraries or privileged structures and performs ligand-based *in silico* screening, using 2- or 3-D descriptors of ligand characteristics to identify molecules sharing the same target profile. The screening procedure may e.g. be pharmacophore screening (chapters 2 and 3). A way of data mining is (quantitative) structure-activity relationship analysis ((Q)SAR). QSAR deals with the correlation of molecular features that determine biological activity and the prediction of compound potency, using a test set of related compounds with known activity, but does not allow for an extrapolation to novel chemotypes.

0.3.1.1 MEANS OF COMPARING LIGANDS

Ligand comparison or classification provides useful insights into the ligand side of a protein-ligand interactions. Classification of small molecules is often based on fingerprint-based similarity searches.

Interaction fingerprints are bit strings that convert 3D protein-ligand interactions into 1D bit representations. They are usually assessed by the Tanimoto coefficient (a similarity coefficient, used in chapter 2)³⁶², the Euclidian distance (a dissimilarity coefficient) or the Tversky coefficient (a composite coefficient)³⁶³. The mostly used Tanimoto coefficient is defined as $TC = \frac{C}{A+B-C}$, where A is the number of features in molecule X but not in Y, B their number in Y but not in X, and C their number in both A and B. The Tversky coefficient contains two user-defined parameters. If they are set to 1, this coefficient equals the Tanimoto coefficient. The Euclidian distance is given by $ED = \sqrt{A+B-2C}$. Besides 1D fingerprints, 2D and 3D fingerprints exist. The additional dimension in 2D fingerprints encodes the presence or absence of fragments, while 3D fingerprints consider multiple conformers.

Fingerprint approaches are either based on residues, structure or knowledge. Examples from the literature of each category are provided in the following.

Residue-based interaction fingerprints built from kinase interaction profiles were used to decipher inhibitor selectivity³⁶⁴. Structural interaction fingerprints served to design protein kinase family targeted chemical libraries³⁶⁵. Among the kinases studied, 20% of the interactions were conserved, and could serve as a basic kinase-like filter in virtual screening. Clustering the fingerprints according to the Tanimoto coefficient into a dendrogram showed four distinct clusters: ATP analogues, p38

inhibitors, cyclin-dependent kinase 2 (CDK2) inhibitors and “DFG-out” (inactive conformation) binders such as Gleevec®³⁶⁶.

Knowledge-based interaction fingerprints in turn can be used to improve the effectiveness of fast scoring functions³⁶⁷. They are also suitable to post-process fragment or scaffold docking poses. It was demonstrated that they performed better in (i) quantifying docking success than the often used root mean square deviation (RMSD from the ligand's X-ray pose) and (ii) enriching true actives than four frequently used scoring functions (FlexX, Glide, GOLD, and Surflex)³⁶⁸.

Other similarity search methods are 3D pharmacophores (0.1.1.4) or shape-based searches such as the one used and detailed in chapter 2. An extension of similarity searches based on homologous protein targets is called “Similog keys”, where atom triplets are not only noted to be present or absent as in fingerprint representations, but also counted. This method allows for searching for initially unknown ligands of homologous protein targets³⁶⁹.

Advantageously, similarity methods can also be applied to fast ligand superposition. Nowadays, they are at a performance level comparable to the well-established protein-ligand docking for binding mode assessment and molecular database screening²¹. An improvement of ligand alignments could be achieved by including knowledge on the binding site and docking may be performed more effectively if similarity with known ligand binding modes is enforced³⁷⁰.

Other classification methods not based on fingerprints rely on statistics, like Bayesian analysis, where the model “learns” how to distinguish known biologically active from inactive “background” molecules. Using such a Bayesian model, Xia et al.³⁷¹ classified kinase inhibitors. Their method is rapid and modifiable, shows meaningful enrichment (8-fold enrichment over random screening for several specific kinases belonging to multiple kinase classes), and is useful in detecting structurally unrelated compounds to known actives or novel targets.

Moving from a similarity to a functionality search, scaffold-hopping deals with finding isofunctional but structurally dissimilar molecular entities³⁷²⁻³⁷⁵.

Yet another means to explore first the chemical and then the target space is related to fragment screening, aiming at probing a large chemical space with minimal synthesis efforts. In analogy to the “rule of 5”²⁰ for drug-like molecules (0.1.1.4), the “rule of 3”³⁷⁶ is applied to fragments. This rule states that fragments that are developed towards a lead and a drug should display the following characteristics: A molecular weight < 300 Da, less than 3 H-bond donors, less than 3 H-bond acceptors and a clogP below 3. Additionally, a number of rotatable bonds ≤ 3 and polar surface area $\leq 60 \text{ \AA}^2$ might also be useful. An appropriate starting fragment should be easily accessible from a synthesis point of view and amendable to chemical modification. In fragment-based drug discovery³⁷⁷, biologically active low molecular weight hits normally having an activity in the millimolar range are identified by NMR³⁷⁸, X-ray crystallography^{379, 380} or tethering (Cys mutations, stable complex formation with disulfide bond forming fragments, and fragment detection via MS)³⁸¹. Fragments of less than 200 Da that are binding undergo the following steps: (i) evolution, (ii) linking, (iii) self-assembly and (iv) optimization³⁸² to yield compounds enriched for target binding that are more lead-like, less hydrophobic, and display more room to improve drug-likeness³⁸³. By adding additional functional groups in approach (i), binding affinity is improved. By linking (ii) fragments with moderate affinity for a target, a significant gain in affinity is achieved. In step (iii), fragments can “grow” inside the target, and in step (iv) ADMET properties are improved. A successful example is the development of nanomolar thrombin inhibitor leads³⁸⁴. The RECAP approach (REtrosynthetic Combinatorial Analysis Procedure) and related techniques used for identifying privileged fragments for combinatorial chemistry are somewhat the reverse of fragment-based drug discovery, as molecules are fragmented *in silico* based on chemical knowledge³⁸⁵.

0.3.2 TARGET-BASED CHEMINFORMATICS

Exploring the target space is an alternative for discovering new NMEs. Hit finding is generally enhanced by targets organized into superfamilies³⁸⁶ and target-directed molecular masterkeys³⁸⁷, showing similar pharmacophores binding to similar targets. These pharmacophores are usually built upon modifications of so-called privileged structures, which are scaffolds being able to bind multiple distinct protein classes³⁸⁸.

In a drug discovery context, therapeutically relevant targets have to be both disease modifying and druggable²⁸⁵. They should be assessed for these properties early on and anti-targets producing off-target effects should be identified as well. But how many druggable proteins do exist? The estimates differ somewhat in the literature. A survey of the database InterPro (<http://www.ebi.ac.uk/interpro/>) revealed > 3000 druggable (low molecular weight, disease relevant) protein domains³⁸⁹, of which about 600 could be targets of molecular drugs¹⁷⁷. Taking alternative splicing into account, the number of potential binding sites is even higher, probably being dependent on the size of the metabolome¹⁷⁷. Hopkins & Groom¹⁷⁷ and Russ & Lampel³⁹⁰ speak of more than 2000 drug targets with at least one drug candidate in clinical trials.

Target-based cheminformatics is either sequence- or structure-based. Target-based methods starting from sequences require a multiple alignment of the targets investigated and are tailored to target families with some lack of crystal structures. Typical examples are cavity trees for the family of G-protein coupled receptors (GPCRs)³⁹¹, which are membrane proteins that are extremely difficult to crystallize. Structure-based approaches depend upon high-resolution X-ray crystal structures, which can also be used as a basis for homology modelling of closely related proteins devoid of crystal structures (chapter 2). They include molecular interaction fields (MIFs), surface comparisons, the identification of packing defects²⁶⁰, and the characterization of binding pockets in human proteins²⁸³. MIFs are applied in chapter 3 to the identification of affinity-improving non-natural amino acid substitutions of a peptide. Packing defects or dehydrons^{392, 393} are H-bonds with a propensity to become dehydrated³⁹⁴. Fernández²⁶⁰ nicely showed that imatinib could gain in binding specificity and affinity towards the tyrosine kinase Bcr-Abl when non-conserved packing defects were erased by methylations. In this context, specificity refers to the appropriate shape complementarity to the target protein compared with other targets. Unspecific binding may sometimes be linked to unwanted off-target effects, but can also be an opportunity for compound repurposing by selectively optimizing side activities³⁹⁵. Selectivity refers to the preferential binding in terms of binding affinity to a target whose active site is distinct from other related proteins.

Systematically screening 3D structures for selectivity-related differences has been proposed³⁹⁶ and applied to kinases³⁹⁷. Focusing on the recognition properties of the ATP binding site, protein structures were aligned, interaction energies calculated using different GRID probes²⁷⁹, and stored in vectors yielding a variance-covariance matrix, which was submitted to different PCA analyses³⁹⁸. The projection plots or “target family landscapes” represent a classification of the binding sites according to the binding pattern. As many more structures will be solved, the significance of such approaches will increase.

When only apo protein structures are available, atom-based fingerprints of known ligand-binding pockets can be used to “train” an algorithm to yield a pharmacophore image of important hot spots. This has been shown to be a viable alternative to VS by molecular docking³⁹⁹. Kauvar et al.⁴⁰⁰ demonstrated how an affinity fingerprint database can be trained for screening purposes. An affinity fingerprint can be defined as a series of IC₅₀ values of the same compound against various targets. Based on a training database covering more than 4 log units of affinity, statistical weighting coefficients were obtained by regression. 70% of the affinities in the range between 1 and 10 µM were correctly predicted using this method. Such an approach is useful in assessing cross-reactivities in view of potential toxicity.

Target “fishing” is often referred to as chemoproteomics and can be achieved *in vitro* by cross-linking the substrate to an affinity chromatography column or *in silico* by inverse virtual screening (chapter 4). Figure 10 illustrates inverse screening as a cartoon. The modeller (penguin) handles the docking algorithm (rope) with a small molecule of interest (the yellow fish). The bait “attracts” potential protein targets (red fishes). Other baits stocked in a ligand database (next to the ice hole) are ready to be used, potentially attracting other “predator” proteins.

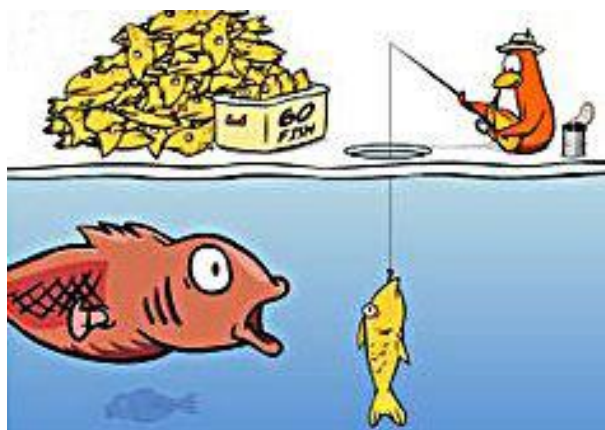


Figure 10: A cartoon view of inverse screening. The modeller (penguin) handles the docking algorithm (rope) with a small molecule (yellow fish). Target proteins (reddish fishes) are attracted.

Ligand- and target-based cheminformatics are interconnected: If the same ligand binds to structurally distant protein binding pockets, usually similar residues at spatially equivalent positions are responsible for the main protein-ligand interactions. Conversely, if different ligands can accommodate in closely related pockets, side chain movements or water-mediated interactions are generally involved. Several studies on adenine binding sites have however revealed that these statements are not always applicable. Complementary shape and electrostatic properties between the adenylate group and the protein can be achieved via a number of alternative amino acid residue arrangements, without these residues being conserved⁴⁰¹⁻⁴⁰³. Thus, the structural diversity of adenine binding pockets is larger than assumed.

0.3.2.1 ORGANIZING THE TARGET SPACE

A way to reveal the substrates or inhibitors of related protein targets are dendrograms based exclusively on small-molecule selectivity (Figure 11b). Vieth et al. published the first such dendrogram for 43 kinases (Figure 11b)⁴⁰⁴. It differs from traditional, sequence-based clustering (Figure 11a) mostly in the higher-level groupings on the smaller clusters and remains comparable for related targets.

Thus, highly related kinases are, on average, inhibited comparably by small molecules. This is an important message for target selection. Common binding patterns for kinases can also be deduced from MIFs obtained from crystallized ATP binding sites, as shown for experimentally tested cyclin-dependent kinase inhibitors³⁹⁷. However, *in vitro* binding assays led to an interaction map between clinical inhibitors and their corresponding kinases, with the specificity varying widely and not being strongly correlated with chemical structure or the identity of the intended target⁴⁰⁵. Alternatively, *in silico* profiling can be thought of. In the case of nuclear receptors, appropriate descriptors storing information taken from a literature-annotated chemical library were used to define a linkage map for active molecules and to retrieve potential targets of nuclear receptors, spotting off-target residual affinities and potential side effects⁴⁰⁶.

In Table 7, protein-ligand interaction databases (available online as of the 3rd of July 2010) are listed, which are useful to obtain protein-ligand interaction diagrams, ligand comparisons with known drugs, functional and catalytic binding sites, binding site alignments as well as repositories of PDB heteroatoms and information on homologous proteins.

Quite many of these databases are up to date, but an automatic update coupled to releases of the PDB would be the ideal case.

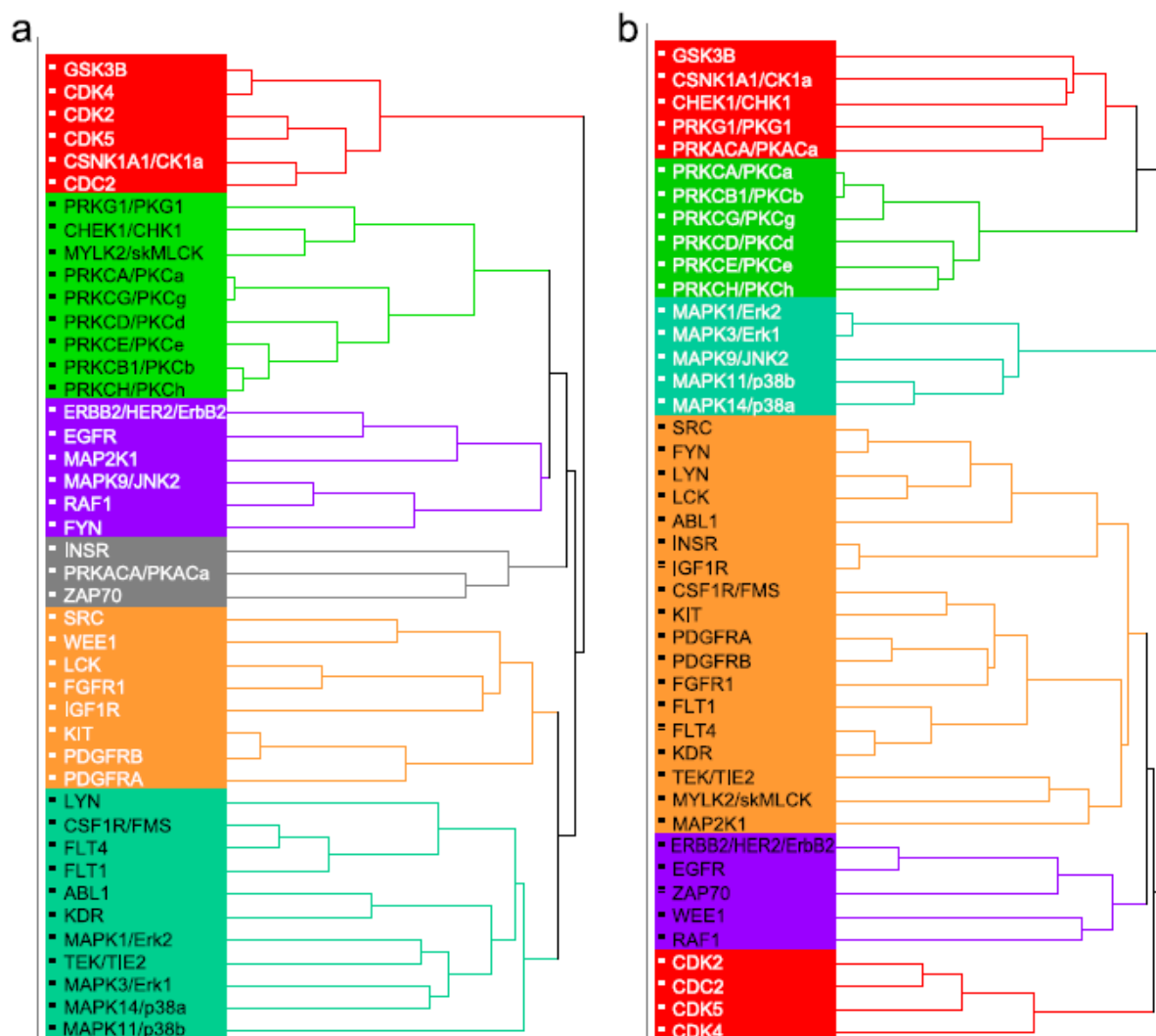


Figure 11: Traditional, sequence-based kinase clustering (a) versus small-molecule based kinase clustering (b). The Figure was taken from ⁴⁰⁴. The full kinase names corresponding to the abbreviations can be found in ⁴⁰⁴.

0.3.3 THE SPECTRUM OF CHEMINFORMATICS

The field of cheminformatics covers a broad spectrum from databases over efficient search to prediction methods. Ligand databases are aimed at molecular similarity and diversity analysis, as well as compound classification and selection, while target databases facilitate target identification. Data screening encompasses data mining, machine learning or VS, using 1-D to 3-D chemical file formats. Predictions might focus on drug-likeness, ADME or toxicity/off-target effects using chemical structures and properties or *in vivo* compound characteristics. In case of representative datasets, quantitative structure-activity or -property predictions can be performed.

Nowadays, it is current practice to screen compound collections, targeted or focused libraries against classes of related proteins (target families), while screening a large protein library with a few small molecules is still uncommon and treated in depth in chapter 4. To define chemical spaces, molecular descriptors are used, which reflect the molecular structure or physical/chemical properties. The principle of similar property is widely used in this context, stating that molecules having similar structures and properties should also exhibit similar activity. It is also commonly assumed that molecules being located closely together in chemical space tend to be functionally related.

Table 7: Useful protein-ligand interaction databases.

Database	URL	Last update	Description
PDBsum ⁴⁰⁷	http://www.ebi.ac.uk/thornton-srv/databases/pdbsum	Linked to PDB	At-a-glance overview of each crystal structure deposited in the PDB with detailed secondary structure PROMOTIF ⁴⁰⁸ analyses, summary Procheck results on the stereochemical “quality”, highly conserved PROSITE ⁴⁰⁹⁻⁴¹¹ patterns, and LIGPLOT ⁴¹² diagrams of protein-ligand interactions
CREDO ⁴¹³	http://www-cryst.bioc.cam.ac.uk/credo	Not known	Protein-ligand interactions are coded as structural interaction fingerprints. Approved drugs (from the DrugBank ^{414, 415} and the KEGG LIGAND database ⁴¹⁶) are automatically recognized. Structural alignments, sequence to PDB structure mappings, substructure searches as well as simultaneous molecular shape and electrostatics comparisons are available.
MMDB ⁴¹⁷	http://www.ncbi.nlm.nih.gov/sites/entrez?db=structure	Linked to PDB	Molecular Modeling Database: Search for homologous proteins via a vector alignment search tool; structure-function associations; ligands are browsable in PubChem
PDBSite ⁴¹⁸	http://srs6.bionet.nsc.ru/srs6-bin/cgi-bin/wgetz?-page+LibInfo+-newId+-lib+PDBSITE	2007	Functional protein sites (catalytic, ligand binding, post-transcriptional modification, protein-protein interaction site, etc.); PDBSiteScan for structural site comparison
SitesBase ⁴¹⁹	http://www.modelling.leeds.ac.uk/sb/	2006	Devised to study structure-function relationships; 3D alignments of binding sites together with a structural similarity score (relative to other structures in the database)
Catalytic Site Atlas ⁴²⁰	http://www.ebi.ac.uk/thornton-srv/databases/CSA/	2010	Catalytic residues are obtained based on a PDB code, a Swiss-Prot code, an Enzyme Classification (EC) number or a literature search
LigBase ⁴²¹	http://modbase.compbio.ucsf.edu/ligbase/	Not known	Structural database of aligned ligand binding sites
ProBiSite	http://protein.bio.puc.cl/carde-x/databases/pdb/index.html	2002	Binding site 3-5 Å around PDB ligand; mostly not drug-like; not up-to-date (2002 PDB release)
MSDsite ⁴²²	http://www.ebi.ac.uk/msd-srv/msdchem/cgi-bin/cgi.pl	2010	Geometrical and chemical environment of ligands and protein-protein interactions. PROSITE motifs ⁴⁰⁹⁻⁴¹¹ .
eF-site ⁴²³	http://ef-site.hgc.jp/eF-site/index.jsp	Not known	Molecular surfaces of protein functional sites, displaying the electrostatic potentials and hydrophobic properties together on the Connolly surfaces ³⁰⁰ of the binding sites
SuperLigands ⁴²⁴	http://bioinf.charite.de/superligands/	2004	Bond information; topological similarity assessed by Tanimoto coefficient ³⁷⁵ ; superposition of 3D ligands; similar drugs (SuperDrug: http://bioinf.charite.de/superdrug).
HIC-Up	http://xray.bmc.uu.se/hicup	2008	Hetero-compound Information Centre in Uppsala: Coordinates and visualization of components of pdb entries like amino acids, ligands or ion; for crystal structure refinement
Het_PDB Navi ⁴²⁵	http://hetpdbnavi.nagahama-i-bio.ac.jp/index.php	2010	Navigator for ligands (heteroatom entries) found in the PDB

0.3.4 CONCLUSION ON CHEMINFORMATICS

Concluding, fully matching the ligand with the target space is still unrealistic, but cheminformatic tools can already handle a tremendous amount of data and are extremely helpful in generating predictions fostering RDD. In respect of target discovery, cheminformatics can address many targets simultaneously, identifying unexpected off-target effects, addressing components of signalling pathways, and assigning biological functions (deorphanizing). In this way, it facilitates biological target validation, reducing the set of proteins to be tested experimentally.

0.4 REFERENCES

- Gohlke, H. & Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew Chem Int Ed Engl* **41**, 2644-76 (2002).
- Fischer, E. Einfluss der Configuration auf die Wirkung der Enzyme. *Ber. Dt. Chem. Ges.* **7**, 2985-2993 (1894).
- Ehrlich, P. *Lancet* **2**, 445-451 (1913).
- Koshland, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc Natl Acad Sci U S A* **44**, 98-104 (1958).
- Pauling, L. Nature of forces between large molecules of biological interest. *Nature* **161**, 707-9 (1948).
- Muirhead, H. & Perutz, M. F. Structure of Haemoglobin. a Three-Dimensional Fourier Synthesis of Reduced Human Haemoglobin at 5-5 Å Resolution. *Nature* **199**, 633-8 (1963).
- Kendrew, J. C. et al. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181**, 662-6 (1958).
- Boehm, H. J. & Klebe, G. What Can We Learn from Molecular Recognition in Protein- Ligand Complexes for the Design of New Drugs? *Angew Chem Int Ed Engl* **35**, 2588-2614 (1996).
- Babine, R. E. & Bender, S. L. Molecular Recognition of Proteinminus signLigand Complexes: Applications to Drug Design. *Chem Rev* **97**, 1359-1472 (1997).
- Berman, H. M. et al. The Protein Data Bank. *Nucleic Acids Res* **28**, 235-42 (2000).
- Berman, H. M. et al. The Protein Data Bank. *Acta Crystallogr D Biol Crystallogr* **58**, 899-907 (2002).
- Kubinyi, H. In Search for New Leads. *EFMC Yearbook* 14-28 (2003).
- Sills, M. A. Future considerations in HTS: the acute effect of chronic dilemmas *Drug Discov Today* **3**, 304-312 (1998).
- Lahana, R. How many leads from HTS? *Drug Discov Today* **4**, 447-448 (1999).
- Oprea, T. I. Property distribution of drug-related chemical databases. *J Comput Aided Mol Des* **14**, 251-64 (2000).
- Doman, T. N. et al. Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J Med Chem* **45**, 2213-21 (2002).
- Paiva, A. M. et al. Inhibitors of dihydrodipicolinate reductase, a key enzyme of the diaminopimelate pathway of *Mycobacterium tuberculosis*. *Biochim Biophys Acta* **1545**, 67-77 (2001).
- Klebe, G. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discov Today* **11**, 580-94 (2006).
- Lyne, P. D. Structure-based virtual screening: an overview. *Drug Discov Today* **7**, 1047-55 (2002).
- Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* **46**, 3-26 (2001).
- Mestres, J. K., R. M. A. Similarity versus docking in 3D virtual screening *Perspectives in Drug Discovery and Design* **20**, 191-207 (2000).
- Mason, J. S., Good, A. C. & Martin, E. J. 3-D pharmacophores in drug discovery. *Curr Pharm Des* **7**, 567-97 (2001).
- Sun, H. Pharmacophore-based virtual screening. *Curr Med Chem* **15**, 1018-24 (2008).
- Nicholls, A. et al. Molecular shape and medicinal chemistry: a perspective. *J Med Chem* **53**, 3862-86.
- Marriott, D. P., Dougall, I. G., Meghani, P., Liu, Y. J. & Flower, D. R. Lead generation using pharmacophore mapping and three-dimensional database searching: application to muscarinic M(3) receptor antagonists. *J Med Chem* **42**, 3210-6 (1999).

26. Tanrikulu, Y. & Schneider, G. Pseudoreceptor models in drug design: bridging ligand- and receptor-based virtual screening. *Nat Rev Drug Discov* **7**, 667-77 (2008).
27. Bajorath, J. Virtual screening in drug discovery: methods, expectations and reality. *Curr Drug Discov Technol*, 24-28 (2002).
28. Wutzler, P. & Vogel, G. Neuraminidase inhibitors in the treatment of influenza A and B-- overview and case reports. *Infection* **28**, 261-6 (2000).
29. Verkhivker, G. M. et al. Deciphering common failures in molecular docking of ligand-protein complexes. *J Comput Aided Mol Des* **14**, 731-51 (2000).
30. Carlson, H. A. & McCammon, J. A. Accommodating protein flexibility in computational drug design. *Mol Pharmacol* **57**, 213-8 (2000).
31. Gilson, M. K., Given, J. A., Bush, B. L. & McCammon, J. A. The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys J* **72**, 1047-69 (1997).
32. Boehm, H. J. & Stahl, M. The Use of Scoring Functions in Drug Discovery Applications. *Rev Comput Chem* **13**, 14-87 (2002).
33. Archakov, A. I. et al. Protein-protein interactions as a target for drugs in proteomics. *Proteomics* **3**, 380-91 (2003).
34. Brady, G. P. & Sharp, K. A. Entropy in protein folding and in protein-protein interactions. *Curr Opin Struct Biol* **7**, 215-21 (1997).
35. Kuntz, I. D., Chen, K., Sharp, K. A. & Kollman, P. A. The maximal affinity of ligands. *Proc Natl Acad Sci U S A* **96**, 9997-10002 (1999).
36. Case, D. A. et al. The Amber biomolecular simulation programs. *J Comput Chem* **26**, 1668-88 (2005).
37. Andrews, P. R., Craik, D. J. & Martin, J. L. Functional group contributions to drug-receptor interactions. *J Med Chem* **27**, 1648-57 (1984).
38. Dawson, R. M. C., Elliott, D. C., Elliott, W. H. & Jones, K. M. Data for Biochemical Research, 3rd ed. *Oxford University Press, Oxford* (1989).
39. Chong, L. T., Dempster, S. E., Hendsch, Z. S., Lee, L. P. & Tidor, B. Computation of electrostatic complements to proteins: a case of charge stabilized binding. *Protein Sci* **7**, 206-10 (1998).
40. Dawson, R. E. et al. Experimental evidence for the functional relevance of anion- π interactions. *Nat Chem* **2**, 533-8 (2010).
41. Kubinyi, H. Hydrogen Bonding: The Last Mystery in Drug Design? in: Pharmacokinetic Optimization in Drug Research (ed. B. Testa, H. van de Waterbeemd, G. Folkers, R. Guy). *Helvetica Chimica Acta*, 513-525 (2001).
42. Pace, C. N., Shirley, B. A., McNutt, M. & Gajiwala, K. Forces contributing to the conformational stability of proteins. *Faseb J* **10**, 75-83 (1996).
43. Abraham, M. H. et al. Application of hydrogen bonding calculations in property based drug design. *Drug Discov Today* **7**, 1056-63 (2002).
44. Pierce, A. C., Sandretto, K. L. & Bemis, G. W. Kinase inhibitors and the case for CH...O hydrogen bonds in protein-ligand binding. *Proteins* **49**, 567-76 (2002).
45. Sarkhel, S. & Desiraju, G. R. N-H...O, O-H...O, and C-H...O hydrogen bonds in protein-ligand complexes: strong and weak interactions in molecular recognition. *Proteins* **54**, 247-59 (2004).
46. Ni, H., Sotriffer, C. A. & McCammon, J. A. Ordered water and ligand mobility in the HIV-1 integrase-5CITEP complex: a molecular dynamics study. *J Med Chem* **44**, 3043-7 (2001).
47. Mikol, V., Papageorgiou, C. & Borer, X. The role of water molecules in the structure-based design of (5-hydroxynorvaline)-2-cyclosporin: synthesis, biological activity, and crystallographic analysis with cyclophilin A. *J Med Chem* **38**, 3361-7 (1995).
48. Stahl, M. & Bohm, H. J. Development of filter functions for protein-ligand docking. *J Mol Graph Model* **16**, 121-32 (1998).
49. McDonald, I. K. & Thornton, J. M. Satisfying hydrogen bonding potential in proteins. *J Mol Biol* **238**, 777-93 (1994).
50. Bruno, I. J. et al. IsoStar: a library of information about nonbonded interactions. *J Comput Aided Mol Des* **11**, 525-37 (1997).
51. Lommerse, J. P. & Taylor, R. Characterising non-covalent interactions with the Cambridge Structural Database. *J Enzyme Inhib* **11**, 223-43 (1997).
52. Weber, P. C., Ohlendorf, D. H., Wendoloski, J. J. & Salemme, F. R. Structural origins of high-affinity biotin binding to streptavidin. *Science* **243**, 85-8 (1989).
53. Young, T., Abel, R., Kim, B., Berne, B. J. & Friesner, R. A. Motifs for molecular recognition exploiting hydrophobic enclosure in protein-ligand binding. *Proc Natl Acad Sci U S A* **104**, 808-13 (2007).

54. Lazaridis, T., Masunov, A. & Gandolfo, F. Contributions to the binding free energy of ligands to avidin and streptavidin. *Proteins* **47**, 194-208 (2002).
55. Luo, H. & Sharp, K. On the calculation of absolute macromolecular binding free energies. *Proc Natl Acad Sci U S A* **99**, 10399-404 (2002).
56. Dixit, S. B. & Chipot, C. Can absolute free energies of association be estimated from molecular mechanical simulations? The biotin-streptavidin system revisited. *J Phys Chem A* **105**, 9795-9799 (2001).
57. Davis, A. M. & Teague, S. J. Hydrogen bonding, hydrophobic interactions, and failure of the rigid receptor hypothesis. *Angew Chem Int Ed Engl* **38**, 736-749 (1999).
58. Dunitz, J. D. & Gavezzotti, A. Attractions and repulsions in molecular crystals: What can be learned from the crystal structures of condensed ring aromatic hydrocarbons? *Acc Chem Res* **32**, 677-684 (1999).
59. Dragovich, P. S. et al. Structure-based design of novel, urea-containing FKBP12 inhibitors. *J Med Chem* **39**, 1872-84 (1996).
60. Cornell, W. D. et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *JACS* **117**, 5179-5197 (1995).
61. Kaminski, G. & Jorgensen, W. L. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Phys Chem* **100**, 18010-18013 (1996).
62. Kim, E. E. et al. Crystal structure of HIV-1 protease in complex with VX-478, a potent and orally bioavailable inhibitor of the enzyme. *JACS* **117**, 1181-1182 (1995).
63. Kollman, P. A. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem Rev* **93**, 2395-2417 (1993).
64. Poornima, C. S. & Dean, P. M. Hydration in drug design. 1. Multiple hydrogen-bonding features of water molecules in mediating protein-ligand interactions. *J Comput Aided Mol Des* **9**, 500-12 (1995).
65. Pickett, S. D. & Sternberg, M. J. Empirical scale of side-chain conformational entropy in protein folding. *J Mol Biol* **231**, 825-39 (1993).
66. Doig, A. J. & Sternberg, M. J. Side-chain conformational entropy in protein folding. *Protein Sci* **4**, 2247-51 (1995).
67. Lau, W. F. & Pettitt, B. M. Selective elimination of interactions: a method for assessing thermodynamic contributions to ligand binding with application to rhinovirus antivirals. *J Med Chem* **32**, 2542-7 (1989).
68. Williams, D. H., Maguire, A. J., Tsuzuki, W. & Westwell, M. S. An analysis of the origins of a cooperative binding energy of dimerisation. *Science* **280**, 711-4 (1998).
69. Sharman, G. J., Searle, M. S., Benhamu, B., Groves, P. & Williams, D. H. Burial of hydrocarbon causes cooperative enhancement of electrostatic binding. *Angew Chem Int Ed Engl* **34**, 1483-1485 (1995).
70. Brooijmans, N., Sharp, K. A. & Kuntz, I. D. Stability of macromolecular complexes. *Proteins* **48**, 645-53 (2002).
71. Brooijmans, N. & Kuntz, I. D. Molecular recognition and docking algorithms. *Annu Rev Biophys Biomol Struct* **32**, 335-73 (2003).
72. Vieth, M., Hirst, J. D., Dominy, B. N., Daigler, H. & Brooks, C. L., 3rd. Assessing Search Strategies for Flexible Docking. *J Comput Chem* **19**, 1623-1631 (1998).
73. Leach, A. R., Shoichet, B. K. & Peishoff, C. E. Prediction of protein-ligand interactions. Docking and scoring: successes and gaps. *J Med Chem* **49**, 5851-5 (2006).
74. McConkey, B. J., Sobolev, V. & Edelman, M. The performance of current methods in ligand-protein docking. *Current Science* **83**, 845-856 (2002).
75. Cole, J. C., Murray, C. W., Nissink, J. W., Taylor, R. D. & Taylor, R. Comparing protein-ligand docking programs is difficult. *Proteins* **60**, 325-32 (2005).
76. Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R. & Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J Mol Biol* **161**, 269-88 (1982).
77. Schulz-Gasch, T. & Stahl, M. Binding site characteristics in structure-based virtual screening: evaluation of current docking tools. *J Mol Model* **9**, 47-57 (2003).
78. Charifson, P. S., Corkery, J. J., Murcko, M. A. & Walters, W. P. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J Med Chem* **42**, 5100-9 (1999).
79. Clark, R. D., Strizhev, A., Leonard, J. M., Blake, J. F. & Matthew, J. B. Consensus scoring for ligand/protein interactions. *J Mol Graph Model* **20**, 281-95 (2002).
80. Stahl, M. & Rarey, M. Detailed analysis of scoring functions for virtual screening. *J Med Chem* **44**, 1035-42 (2001).

81. Halperin, I., Ma, B., Wolfson, H. & Nussinov, R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins* **47**, 409-43 (2002).
82. Bissantz, C., Folkers, G. & Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J Med Chem* **43**, 4759-67 (2000).
83. Kellenberger, E., Rodrigo, J., Muller, P. & Rognan, D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins* **57**, 225-42 (2004).
84. Kontoyianni, M., McClellan, L. M. & Sokol, G. S. Evaluation of docking performance: comparative data on docking algorithms. *J Med Chem* **47**, 558-65 (2004).
85. Sousa, S. F., Fernandes, P. A. & Ramos, M. J. Protein-ligand docking: current status and future challenges. *Proteins* **65**, 15-26 (2006).
86. Bursulaya, B. D., Totrov, M., Abagyan, R. & Brooks, C. L., 3rd. Comparative study of several algorithms for flexible ligand docking. *J Comput Aided Mol Des* **17**, 755-63 (2003).
87. Warren, G. L. et al. A critical assessment of docking programs and scoring functions. *J Med Chem* **49**, 5912-31 (2006).
88. Pospisil, P., Ballmer, P., Scapozza, L. & Folkers, G. Tautomerism in computer-aided drug design. *J Recept Signal Transduct Res* **23**, 361-71 (2003).
89. Zavodszky, M. I. & Kuhn, L. A. Side-chain flexibility in protein-ligand binding: the minimal rotation hypothesis. *Protein Sci* **14**, 1104-14 (2005).
90. Merlitz, H., Burghardt, B. & Wenzel, W. Impact of receptor conformation on in-silico screening performance. *Chem Phys Lett* **390**, 500-505 (2004).
91. Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* **267**, 727-48 (1997).
92. Cavasotto, C. N. & Abagyan, R. A. Protein flexibility in ligand docking and virtual screening to protein kinases. *J Mol Biol* **337**, 209-25 (2004).
93. Erickson, J. A., Jalaie, M., Robertson, D. H., Lewis, R. A. & Vieth, M. Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy. *J Med Chem* **47**, 45-55 (2004).
94. Cheng, Y. & Prusoff, W. H. Relationship between the inhibition constant (K_1) and the concentration of inhibitor which causes 50 per cent inhibition (I_{50}) of an enzymatic reaction. *Biochem Pharmacol* **22**, 3099-108 (1973).
95. Murphy, K. P. Predicting binding energetics from structure: looking beyond ΔG degrees. *Med Res Rev* **19**, 333-9 (1999).
96. Cramer, R. D. I., Patterson, D. E., Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *JACS*, 5959-5967 (1988).
97. Klebe, G., Abraham, U. & Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J Med Chem* **37**, 4130-46 (1994).
98. Vedani, A. & Dobler, M. 5D-QSAR: the key for simulating induced fit? *J Med Chem* **45**, 2139-49 (2002).
99. Wade, R. C., Clark, K. J. & Goodford, P. J. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 1. Ligand probe groups with the ability to form two hydrogen bonds. *J Med Chem* **36**, 140-7 (1993).
100. Wade, R. C. & Goodford, P. J. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 2. Ligand probe groups with the ability to form more than two hydrogen bonds. *J Med Chem* **36**, 148-56 (1993).
101. Vedani, A., Zbinden, P., Snyder, J. P. & Greenidge, P. A. Pseudoreceptor modeling: The construction of three-dimensional receptor surrogates. *JACS* **117**, 4987-4994 (1995).
102. Kollman, P. A. et al. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res* **33**, 889-97 (2000).
103. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912-34 (2002).
104. Muegge, I. & Martin, Y. C. A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J Med Chem* **42**, 791-804 (1999).
105. Reynolds, C. A., King, P. M. & Richards, W. G. Free energy calculations in molecular biophysics. *Molecular Physics*, 251-276 (1992).
106. Zwanzig, R. W. High-temperature equation of state by a perturbation method .1. nonpolar gases. *J Chem Phys* **22**, 1420-1426 (1954).

107. Kirkwood, J. G. Statistical mechanics of fluid mixtures. *J Chem Phys* **3**, 300-313 (1935).
108. Saito, M. & Nakamura, H. Hydration free energy calculations by the acceptance ratio method. *J Comput Chem* **11**, 76-81 (2004).
109. Pearlman, D. A. & Kollman, P. A. A new method for carrying out free energy perturbation calculations: Dynamically modified windows. *J Chem Phys* **90**, 2460-2471 (1989).
110. Singh, U. C., Brown, F. K., Bash, P. A. & Kollman, P. A. An approach to the application of free energy perturbation methods using molecular dynamics: applications to the transformations of CH₃OH - CH₃CH₃, H₃O⁺ - NH₄⁺, glycine - alanine, and alanine - phenylalanine in aqueous solution and to H₃O⁺(H₂O)₃ - NH₄⁺(H₂O)₃ in the gas phase. *JACS* **109**, 1607-1614 (1987).
111. Dang, L. X., Pearlman, D. A. & Kollman, P. A. Why do A.T base pairs inhibit Z-DNA formation? *Proc Natl Acad Sci U S A* **87**, 4630-4 (1990).
112. Mitchell, M. J. & McCammon, J. A. Free energy difference calculations by thermodynamic integration: Difficulties in obtaining a precise value. *J Comput Chem* **12**, 271-275 (2004).
113. Aqvist, J., Medina, C. & Samuelsson, J. E. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng* **7**, 385-91 (1994).
114. Hansson, T. & Aqvist, J. Estimation of binding free energies for HIV proteinase inhibitors by molecular dynamics simulations. *Protein Eng* **8**, 1137-44 (1995).
115. Brandsdal, B. O., Aqvist, J. & Smalas, A. O. Computational analysis of binding of P1 variants to trypsin. *Protein Sci* **10**, 1584-95 (2001).
116. Graffner-Nordberg, M., Kolmodin, K., Aqvist, J., Queener, S. F. & Hallberg, A. Design, synthesis, computational prediction, and biological evaluation of ester soft drugs as inhibitors of dihydrofolate reductase from *Pneumocystis carinii*. *J Med Chem* **44**, 2391-402 (2001).
117. Huang, D. & Caflisch, A. Efficient evaluation of binding free energy using continuum electrostatics solvation. *J Med Chem* **47**, 5791-7 (2004).
118. Ljungberg, K. B. et al. Computational modelling of inhibitor binding to human thrombin. *Eur J Pharm Sci* **12**, 441-6 (2001).
119. Aqvist, J., Luzhkov, V. B. & Brandsdal, B. O. Ligand binding affinities from MD simulations. *Acc Chem Res* **35**, 358-65 (2002).
120. Bortolato, A. & Moro, S. In silico binding free energy predictability by using the linear interaction energy (LIE) method: bromobenzimidazole CK2 inhibitors as a case study. *J Chem Inf Model* **47**, 572-82 (2007).
121. Izrailev, S. et al. Steered molecular dynamics. *Lecture Notes in Computational Science and Engineering; Computational Molecular Dynamics: Challenges, Methods, Ideas; Springer Verlag, New York* **4**, 39-65 (1998).
122. Ajay & Murcko, M. A. Computational methods to predict binding free energy in ligand-receptor complexes. *J Med Chem* **38**, 4953-67 (1995).
123. Bohm, H. J. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J Comput Aided Mol Des* **8**, 243-56 (1994).
124. Bohm, H. J. Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs. *J Comput Aided Mol Des* **12**, 309-23 (1998).
125. Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V. & Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des* **11**, 425-45 (1997).
126. Wang, R., Liu, L., Lai, L. H. & Tang, Y. Q. SCORE: A new empirical method for estimating the binding affinity of a protein-ligand complex. *J Mol Model* **4**, 379-394 (1998).
127. Bohacek, R. S. & McMartin, C. Multiple highly diverse structures complementary to enzyme binding sites: results of extensive application of a de novo design method incorporating combinatorial growth. *JACS* **116**, 5560-5571 (1994).
128. Rognan, D., Lauemoller, S. L., Holm, A., Buus, S. & Tschinke, V. Predicting binding affinities of protein ligands from three-dimensional models: application to peptide binding to class I major histocompatibility proteins. *J Med Chem* **42**, 4650-8 (1999).
129. Kasper, P., Christen, P. & Gehring, H. Empirical calculation of the relative free energies of peptide binding to the molecular chaperone DnaK. *Proteins* **40**, 185-92 (2000).
130. Moitessier, N., Englebienne, P., Lee, D., Lawandi, J. & Corbeil, C. R. Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *Br J Pharmacol* **153 Suppl 1**, S7-26 (2008).
131. Williams, D. H. & Bardsley, B. Estimating binding constants – The hydrophobic effect and cooperativity *Perspectives in Drug Discovery and Design* **17**, 43-59 (1996).

132. Timmers, L. F. et al. Structural studies of human purine nucleoside phosphorylase: towards a new specific empirical scoring function. *Arch Biochem Biophys* **479**, 28-38 (2008).
133. Wang, R., Lu, Y. & Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J Med Chem* **46**, 2287-303 (2003).
134. Wang, R., Lu, Y., Fang, X. & Wang, S. An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein-ligand complexes. *J Chem Inf Comput Sci* **44**, 2114-25 (2004).
135. Caceres, R. A. et al. Molecular modeling and dynamics simulations of PNP from *Streptococcus agalactiae*. *Bioorg Med Chem* **16**, 4984-93 (2008).
136. Sharp, K. A., Nicholls, A., Friedman, R. & Honig, B. Extracting hydrophobic free energies from experimental data: relationship to protein folding and theoretical models. *Biochemistry* **30**, 9686-97 (1991).
137. Wallqvist, A., Jernigan, R. L. & Covell, D. G. A preference-based free-energy parameterization of enzyme-inhibitor binding. Applications to HIV-1-protease inhibitor design. *Protein Sci* **4**, 1881-903 (1995).
138. Gohlke, H., Hendlich, M. & Klebe, G. Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol* **295**, 337-56 (2000).
139. Gohlke, H., Hendlich, M. & Klebe, G. Predicting binding modes, binding affinities and "hot spots" for protein-ligand complexes using a knowledge-based scoring function. *Perspectives in Drug Discovery and Design* **20**, 115-144 (2000).
140. Mooij, W. T. & Verdonk, M. L. General and targeted statistical potentials for protein-ligand interactions. *Proteins* **61**, 272-87 (2005).
141. Paul, N. & Rognan, D. ConsDock: A new program for the consensus analysis of protein-ligand interactions. *Proteins* **47**, 521-33 (2002).
142. Terp, G. E., Johansen, B. N., Christensen, I. T. & Jorgensen, F. S. A new concept for multidimensional selection of ligand conformations (MultiSelect) and multidimensional scoring (MultiScore) of protein-ligand binding affinities. *J Med Chem* **44**, 2333-43 (2001).
143. Ferrara, P., Gohlke, H., Price, D. J., Klebe, G. & Brooks, C. L., 3rd. Assessing scoring functions for protein-ligand interactions. *J Med Chem* **47**, 3032-47 (2004).
144. Moitessier, N., Therrien, E. & Hanessian, S. A method for induced-fit docking, scoring, and ranking of flexible ligands. Application to peptidic and pseudopeptidic beta-secretase (BACE 1) inhibitors. *J Med Chem* **49**, 5885-94 (2006).
145. Schomburg, I. et al. BRENDA: a resource for enzyme data and metabolic information. *Trends Biochem Sci* **27**, 54-6 (2002).
146. Schomburg, I., Chang, A. & Schomburg, D. BRENDA, enzyme data and metabolic information. *Nucleic Acids Res* **30**, 47-9 (2002).
147. Pharkya, P., Nikolaev, E. V. & Maranas, C. D. Review of the BRENDA Database. *Metab Eng* **5**, 71-3 (2003).
148. Liu, T., Lin, Y., Wen, X., Jorissen, R. N. & Gilson, M. K. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res* **35**, D198-201 (2007).
149. Li, L., Dantzer, J. J., Nowacki, J., O'Callaghan, B. J. & Meroueh, S. O. PDBcal: a comprehensive dataset for receptor-ligand interactions with three-dimensional structures and binding thermodynamics from isothermal titration calorimetry. *Chem Biol Drug Des* **71**, 529-32 (2008).
150. Block, P., Sotriffer, C. A., Dramburg, I. & Klebe, G. AffinDB: a freely accessible database of affinities for protein-ligand complexes from the PDB. *Nucleic Acids Res* **34**, D522-6 (2006).
151. Benson, M. L. et al. Binding MOAD, a high-quality protein-ligand database. *Nucleic Acids Res* **36**, D674-8 (2008).
152. Hu, L., Benson, M. L., Smith, R. D., Lerner, M. G. & Carlson, H. A. Binding MOAD (Mother Of All Databases). *Proteins* **60**, 333-40 (2005).
153. Wang, R., Fang, X., Lu, Y. & Wang, S. The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures. *J Med Chem* **47**, 2977-80 (2004).
154. Wang, R., Fang, X., Lu, Y., Yang, C. Y. & Wang, S. The PDBbind database: methodologies and updates. *J Med Chem* **48**, 4111-9 (2005).
155. Roche, O., Kiyama, R. & Brooks, C. L., 3rd. Ligand-protein database: linking protein-ligand complex structures to binding data. *J Med Chem* **44**, 3592-8 (2001).
156. Ahmad, S. et al. Protein-Ligand Interactions: ProLINT Database and QSAR Analysis *Genome Informatics* **14**, 537-538 (2003).

157. Puvanendrapillai, D. & Mitchell, J. B. L/D Protein Ligand Database (PLD): additional understanding of the nature and specificity of protein-ligand complexes. *Bioinformatics* **19**, 1856-7 (2003).
158. Mitchell, J. B. O., Laskowski, R. A., Alex, A. & Thornton, J. M. BLEEP - Potential of Mean Force Describing Protein-Ligand Interactions: I. Generating Potential. *J Comput Chem* **20**, 1165-1176 (1999).
159. Mitchell, J. B. O., Laskowski, R. A., Alex, A., Foster, M. J. & Thornton, J. M. BLEEP - Potential of Mean Force Describing Protein-Ligand Interactions: II. Calculation of Binding Energies and Comparison with Experimental Data. *J Comput Chem* **20**, 1177-1185 (1999).
160. Ewing, T. J., Makino, S., Skillman, A. G. & Kuntz, I. D. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J Comput Aided Mol Des* **15**, 411-28 (2001).
161. Moustakas, D. T. et al. Development and validation of a modular, extensible docking program: DOCK 5. *J Comput Aided Mol Des* **20**, 601-19 (2006).
162. Lang, P. T. et al. DOCK 6: combining techniques to model RNA-small molecule complexes. *Rna* **15**, 1219-30 (2009).
163. Rarey, M., Kramer, B., Lengauer, T. & Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J Mol Biol* **261**, 470-89 (1996).
164. Rarey, M., Kramer, B. & Lengauer, T. Docking of hydrophobic ligands with interaction-based matching algorithms. *Bioinformatics* **15**, 243-50 (1999).
165. Rarey, M., Kramer, B. & Lengauer, T. The particle concept: placing discrete water molecules during protein-ligand docking predictions. *Proteins* **34**, 17-28 (1999).
166. Jain, A. N. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *J Med Chem* **46**, 499-511 (2003).
167. Jones, G., Willett, P. & Glen, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J Mol Biol* **245**, 43-53 (1995).
168. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W. & Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins* **52**, 609-23 (2003).
169. Jones, S. & Thornton, J. M. Protein-protein interactions: a review of protein dimer structures. *Prog Biophys Mol Biol* **63**, 31-65 (1995).
170. Yang, J. & Hage, D. S. Effect of mobile phase composition on the binding kinetics of chiral solutes on a protein-based high-performance liquid chromatography column: interactions of D- and L-tryptophan with immobilized human serum albumin. *J Chromatogr A* **766**, 15-25 (1997).
171. Talbert, A. M., Tranter, G. E., Holmes, E. & Francis, P. L. Determination of drug-plasma protein binding kinetics and equilibria by chromatographic profiling: exemplification of the method using L-tryptophan and albumin. *Anal Chem* **74**, 446-52 (2002).
172. Veenstra, T. D. Electrospray ionization mass spectrometry in the study of biomolecular non-covalent interactions. *Biophys Chem* **79**, 63-79 (1999).
173. Ratcliff, G. C. & Erie, D. A. A novel single-molecule study to determine protein--protein association constants. *J Am Chem Soc* **123**, 5632-5 (2001).
174. Perozzo, R., Jelesarov, I., Bosshard, H. R., Folkers, G. & Scapozza, L. Compulsory order of substrate binding to herpes simplex virus type 1 thymidine kinase. A calorimetric study. *J Biol Chem* **275**, 16139-45 (2000).
175. de Azevedo, W. F., Jr. & Dias, R. Experimental approaches to evaluate the thermodynamics of protein-drug interactions. *Curr Drug Targets* **9**, 1071-6 (2008).
176. Jelesarov, I. & Bosshard, H. R. Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. *J Mol Recognit* **12**, 3-18 (1999).
177. Hopkins, A. L. & Groom, C. R. The druggable genome. *Nat Rev Drug Discov* **1**, 727-30 (2002).
178. Chene, P. Drugs targeting protein-protein interactions. *ChemMedChem* **1**, 400-11 (2006).
179. Berg, T. Modulation of protein-protein interactions with small organic molecules. *Angew Chem Int Ed Engl* **42**, 2462-81 (2003).
180. Yin, H. & Hamilton, A. D. Strategies for targeting protein-protein interactions with synthetic agents. *Angew Chem Int Ed Engl* **44**, 4130-63 (2005).
181. Geysen, H. M., Barteling, S. J. & Meloen, R. H. Small peptides induce antibodies with a sequence and structural requirement for binding antigen comparable to antibodies raised against the native protein. *Proc Natl Acad Sci U S A* **82**, 178-82 (1985).
182. Sood, V. D. & Baker, D. Recapitulation and design of protein binding peptide structures and sequences. *J Mol Biol* **357**, 917-27 (2006).

183. Li, L., Liang, S., Pilcher, M. M. & Meroueh, S. O. Incorporating receptor flexibility in the molecular design of protein interfaces. *Protein Eng Des Sel* **22**, 575-86 (2009).
184. Zutshi, R., Brickner, M. & Chmielewski, J. Inhibiting the assembly of protein-protein interfaces. *Curr Opin Chem Biol* **2**, 62-6 (1998).
185. Bouras, A. et al. Design, synthesis, and evaluation of conformationally constrained tongs, new inhibitors of HIV-1 protease dimerisation. *J Med Chem* **42**, 957-62 (1999).
186. Morris, M. C. et al. A new potent HIV-1 reverse transcriptase inhibitor. A synthetic peptide derived from the interface subunit domains. *J Biol Chem* **274**, 24941-6 (1999).
187. Sourgen, F. et al. A synthetic peptide from the human immunodeficiency virus type-1 integrase exhibits coiled-coil properties and interferes with the in vitro integration activity of the enzyme. Correlated biochemical and spectroscopic results. *Eur J Biochem* **240**, 765-73 (1996).
188. Digard, P. et al. Specific inhibition of herpes simplex virus DNA polymerase by helical peptides corresponding to the subunit interface. *Proc Natl Acad Sci U S A* **92**, 1456-60 (1995).
189. Schramm, H. J. et al. Lipopeptides as dimerisation inhibitors of HIV-1 protease. *Biol Chem* **380**, 593-6 (1999).
190. Shultz, M. D. & Chmielewski, J. Probing the role of interfacial residues in a dimerisation inhibitor of HIV-1 protease. *Bioorg Med Chem Lett* **9**, 2431-6 (1999).
191. Verdine, G. L. The combinatorial chemistry of nature. *Nature* **384**, 11-3 (1996).
192. Schafmeister, C. E., Po, J. & Verdine, G. L. An all-hydrocarbon crosslinking system for enhancing the helicity and metabolic stability of peptides. *JACS* **122**, 5891-5892 (2000).
193. Amara, J. F. et al. A versatile synthetic dimeriser for the regulation of protein-protein interactions. *Proc Natl Acad Sci U S A* **94**, 10618-23 (1997).
194. Keenan, T. et al. Synthesis and activity of bivalent FKBP12 ligands for the regulated dimerisation of proteins. *Bioorg Med Chem* **6**, 1309-35 (1998).
195. Blau, C. A., Peterson, K. R., Drachman, J. G. & Spencer, D. M. A proliferation switch for genetically modified cells. *Proc Natl Acad Sci U S A* **94**, 3076-81 (1997).
196. MacCorkle, R. A., Freeman, K. W. & Spencer, D. M. Synthetic activation of caspases: artificial death switches. *Proc Natl Acad Sci U S A* **95**, 3655-60 (1998).
197. DeLano, W. L. Unraveling hot spots in binding interfaces: progress and challenges. *Curr Opin Struct Biol* **12**, 14-20 (2002).
198. Fernandez, A. & Scott, R. Dehydron: a structurally encoded signal for protein interaction. *Biophys J* **85**, 1914-28 (2003).
199. Toogood, P. L. Inhibition of protein-protein association by small molecules: approaches and progress. *J Med Chem* **45**, 1543-58 (2002).
200. Gadek, T. R. & Nicholas, J. B. Small molecule antagonists of proteins. *Biochem Pharmacol* **65**, 1-8 (2003).
201. Rai, S. S. & Wolff, J. Localization of the vinblastine-binding site on beta-tubulin. *J Biol Chem* **271**, 14707-11 (1996).
202. Hu, Z., Ma, B., Wolfson, H. & Nussinov, R. Conservation of polar residues as hot spots at protein interfaces. *Proteins* **39**, 331-42 (2000).
203. Pawson, T. Protein modules and signalling networks. *Nature* **373**, 573-80 (1995).
204. Hunter, T. Signaling--2000 and beyond. *Cell* **100**, 113-27 (2000).
205. Reichmann, D., Rahat, O., Cohen, M., Neuvirth, H. & Schreiber, G. The molecular architecture of protein-protein binding sites. *Curr Opin Struct Biol* **17**, 67-76 (2007).
206. Oneyama, C. et al. Synthetic inhibitors of proline-rich ligand-mediated protein-protein interaction: potent analogs of UCS15A. *Chem Biol* **10**, 443-51 (2003).
207. Clackson, T. & Wells, J. A. A hot spot of binding energy in a hormone-receptor interface. *Science* **267**, 383-6 (1995).
208. Cwirla, S. E. et al. Peptide agonist of the thrombopoietin receptor as potent as the natural cytokine. *Science* **276**, 1696-9 (1997).
209. Dalgarno, D. C., Botfield, M. C. & Rickles, R. J. SH3 domains and drug design: ligands, structure, and biological function. *Biopolymers* **43**, 383-400 (1997).
210. Kay, B. K., Williamson, M. P. & Sudol, M. The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *Faseb J* **14**, 231-41 (2000).
211. Ma, B., Elkayam, T., Wolfson, H. & Nussinov, R. Protein-protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proc Natl Acad Sci U S A* **100**, 5772-7 (2003).

212. Hajduk, P. J., Huth, J. R. & Fesik, S. W. Druggability indices for protein targets derived from NMR-based screening data. *J Med Chem* **48**, 2518-25 (2005).
213. Jones, S. & Thornton, J. M. Principles of protein-protein interactions. *Proc Natl Acad Sci U S A* **93**, 13-20 (1996).
214. Stites, W. E. Protein-protein interactions: Interface Structure, Binding Thermodynamics, and Mutational Analysis. *Chem Rev* **97**, 1233-1250 (1997).
215. Davies, D. R. & Cohen, G. H. Interactions of protein antigens with antibodies. *Proc Natl Acad Sci U S A* **93**, 7-12 (1996).
216. Ivanov, Y. D. et al. The optical biosensor studies on the role of hydrophobic tails of NADPH-cytochrome P450 reductase and cytochromes P450 2B4 and b5 upon productive complex formation within a monomeric reconstituted system. *Arch Biochem Biophys* **362**, 87-93 (1999).
217. Ivanov, Y. D., Kanaeva, I. P. & Archakov, A. I. Optical biosensor study of ternary complex formation in a cytochrome P450 2B4 system. *Biochem Biophys Res Commun* **273**, 750-2 (2000).
218. Nooren, I. M. & Thornton, J. M. Diversity of protein-protein interactions. *Embo J* **22**, 3486-92 (2003).
219. Jones, S. & Thornton, J. M. Protein-protein recognition. *Kleathous C. (Ed.), Oxford University Press, New York*, 33-59 (2000).
220. Wodak, S. J. & Janin, J. Structural basis of macromolecular recognition. *Adv Protein Chem* **61**, 9-73 (2002).
221. Nooren, I. M. & Thornton, J. M. Structural characterisation and functional significance of transient protein-protein interactions. *J Mol Biol* **325**, 991-1018 (2003).
222. Lo Conte, L., Chothia, C. & Janin, J. The atomic structure of protein-protein recognition sites. *J Mol Biol* **285**, 2177-98 (1999).
223. Tsai, C. J., Xu, D. & Nussinov, R. Structural motifs at protein-protein interfaces: protein cores versus two-state and three-state model complexes. *Protein Sci* **6**, 1793-805 (1997).
224. Larsen, T. A., Olson, A. J. & Goodsell, D. S. Morphology of protein-protein interfaces. *Structure* **6**, 421-7 (1998).
225. Dimasi, N. Crystal structure of the C-terminal SH3 domain of the adaptor protein GADS in complex with SLP-76 motif peptide reveals a unique SH3-SH3 interaction. *Int J Biochem Cell Biol* **39**, 109-23 (2007).
226. Tobí, D. & Bahar, I. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc Natl Acad Sci U S A* **102**, 18908-13 (2005).
227. Keskin, O., Ma, B., Rogale, K., Gunasekaran, K. & Nussinov, R. Protein-protein interactions: organization, cooperativity and mapping in a bottom-up Systems Biology approach. *Phys Biol* **2**, S24-35 (2005).
228. Thorn, K. S. & Bogan, A. A. ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. *Bioinformatics* **17**, 284-5 (2001).
229. Bogan, A. A. & Thorn, K. S. Anatomy of hot spots in protein interfaces. *J Mol Biol* **280**, 1-9 (1998).
230. DeLano, W. L., Ultsch, M. H., de Vos, A. M. & Wells, J. A. Convergent solutions to binding at a protein-protein interface. *Science* **287**, 1279-83 (2000).
231. Chakrabarti, P. & Janin, J. Dissecting protein-protein recognition sites. *Proteins* **47**, 334-43 (2002).
232. Verkhivker, G. M. et al. Monte Carlo simulations of the peptide recognition at the consensus binding site of the constant fragment of human immunoglobulin G: the energy landscape analysis of a hot spot at the intermolecular interface. *Proteins* **48**, 539-57 (2002).
233. Vajda, S. & Guarnieri, F. Characterization of protein-ligand interaction sites using experimental and computational methods. *Curr Opin Drug Discov Devel* **9**, 354-62 (2006).
234. Jaitin, D. A. et al. Inquiring into the differential action of interferons (IFNs): an IFN- α 2 mutant with enhanced affinity to IFNAR1 is functionally similar to IFN- β . *Mol Cell Biol* **26**, 1888-97 (2006).
235. Svensson, H. G. et al. Contributions of amino acid side chains to the kinetics and thermodynamics of the bivalent binding of protein L to Ig kappa light chain. *Biochemistry* **43**, 2445-57 (2004).
236. Janin, J. Protein-protein recognition. *Kleathous C. (Ed.), Oxford University Press, New York*, 1-32 (2000).
237. Stevens, J. M., Armstrong, R. N. & Dirr, H. W. Electrostatic interactions affecting the active site of class sigma glutathione S-transferase. *Biochem J* **347 Pt 1**, 193-7 (2000).

-
238. Sheinerman, F. B., Norel, R. & Honig, B. Electrostatic aspects of protein-protein interactions. *Curr Opin Struct Biol* **10**, 153-9 (2000).
239. Ivanov, Y. D. et al. Molecular recognition in the p450cam monooxygenase system: direct monitoring of protein-protein interactions by using optical biosensor. *Arch Biochem Biophys* **391**, 255-64 (2001).
240. Xu, D., Lin, S. L. & Nussinov, R. Protein binding versus protein folding: the role of hydrophilic bridges in protein associations. *J Mol Biol* **265**, 68-84 (1997).
241. McCoy, A. J., Chandana Epa, V. & Colman, P. M. Electrostatic complementarity at protein/protein interfaces. *J Mol Biol* **268**, 570-84 (1997).
242. Archakov, A. I. & Ivanov, Y. D. The optical biosensor study of protein-protein interactions within cytochromes P450 containing monooxygenase systems. C. Nicolini (Ed.), *Biophysics of electron transfer and molecular bioelectronics*, Plenum Publication, USA, 173-194 (1999).
243. Cole, C. & Warwicker, J. Side-chain conformational entropy at protein-protein interfaces. *Protein Sci* **11**, 2860-70 (2002).
244. Xu, D., Tsai, C. J. & Nussinov, R. Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein Eng* **10**, 999-1012 (1997).
245. Raschke, T. M. Water structure and interactions with protein surfaces. *Curr Opin Struct Biol* **16**, 152-9 (2006).
246. Eisenhaber, F. Hydrophobic regions on protein surfaces. *Perspectives in Drug Discovery and Design* **17**, 27-42 (1999).
247. Eisenhaber, F. & Argos, P. Hydrophobic regions on protein surfaces: definition based on hydration shell structure and a quick method for their computation. *Protein Eng* **9**, 1121-33 (1996).
248. Burley, S. K. & Petsko, G. A. Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science* **229**, 23-8 (1985).
249. Vinogradov, S. N. & Linnell, R. H. Hydrogen Bonding. Van Nostrand Reinhold (Ed.), New York, 319 pages (1971).
250. Hunter, C. A. & Sanders, J. K. M. The nature of pi-pi interactions. *JACS* **112**, 5525-5534 (1990).
251. Muehldorf, A., Engen, D. V., Warner, J. C. & Hamilton, A. D. Aromatic-aromatic interactions in molecular recognition: A family of artificial receptors for thymine that shows both face-to-face and edge- to-face orientations. *JACS* **110**, 6561-6562 (1988).
252. Mitchell, J. B., Nandi, C. L., McDonald, I. K., Thornton, J. M. & Price, S. L. Amino/aromatic interactions in proteins: is the evidence stacked against hydrogen bonding? *J Mol Biol* **239**, 315-31 (1994).
253. Hobza, P. Potential energy surface for the benzene dimer. Results of ab initio CCSD(T) calculations show two nearly isoenergetic structures: T-shaped and parallel-displaced. *J Phys Chem* **100**, 18790-18794 (1996).
254. Gervasio, F. L., Chelli, R., Procacci, P. & Schettino, V. The nature of intermolecular interactions between aromatic amino acid residues. *Proteins* **48**, 117-25 (2002).
255. Chelli, R., Gervasio, F. L., Procacci, P. & Schettino, V. Stacking and T-shape competition in aromatic-aromatic amino acid interactions. *J Am Chem Soc* **124**, 6133-43 (2002).
256. Janin, J. The kinetics of protein-protein recognition. *Proteins* **28**, 153-61 (1997).
257. Selzer, T., Albeck, S. & Schreiber, G. Rational design of faster associating and tighter binding protein complexes. *Nat Struct Biol* **7**, 537-41 (2000).
258. Eyrisch, S. & Helms, V. Transient pockets on protein surfaces involved in protein-protein interaction. *J Med Chem* **50**, 3457-64 (2007).
259. Eyrisch, S. & Helms, V. What induces pocket openings on protein surface patches involved in protein-protein interactions? *J Comput Aided Mol Des* **23**, 73-86 (2009).
260. Fernandez, A. Incomplete protein packing as a selectivity filter in drug design. *Structure* **13**, 1829-36 (2005).
261. Campbell, S. J., Gold, N. D., Jackson, R. M. & Westhead, D. R. Ligand binding: functional site location, similarity and docking. *Curr Opin Struct Biol* **13**, 389-95 (2003).
262. Sotriffer, C. & Klebe, G. Identification and mapping of small-molecule binding sites in proteins: computational tools for structure-based drug design. *Farmaco* **57**, 243-51 (2002).
263. Liang, J., Edelsbrunner, H. & Woodward, C. Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci* **7**, 1884-97 (1998).
264. Binkowski, T. A., Naghibzadeh, S. & Liang, J. CASTp: Computed Atlas of Surface Topography of proteins. *Nucleic Acids Res* **31**, 3352-5 (2003).

265. Peters, K. P., Fauck, J. & Frommel, C. The automatic search for ligand binding sites in proteins of known three-dimensional structure using only geometric criteria. *J Mol Biol* **256**, 201-13 (1996).
266. Laskowski, R. A. SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J Mol Graph* **13**, 323-30, 307-8 (1995).
267. Laskowski, R. A., Luscombe, N. M., Swindells, M. B. & Thornton, J. M. Protein clefts in molecular recognition and function. *Protein Sci* **5**, 2438-52 (1996).
268. Glaser, F., Morris, R. J., Najmanovich, R. J., Laskowski, R. A. & Thornton, J. M. A method for localizing ligand binding pockets in protein structures. *Proteins* **62**, 479-88 (2006).
269. Brady, G. P., Jr. & Stouten, P. F. Fast prediction and visualization of protein binding pockets with PASS. *J Comput Aided Mol Des* **14**, 383-401 (2000).
270. Coleman, R. G. & Sharp, K. A. Travel depth, a new shape descriptor for macromolecules: application to ligand binding. *J Mol Biol* **362**, 441-58 (2006).
271. Kleywegt, G. J. & Jones, T. A. Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr D Biol Crystallogr* **50**, 178-85 (1994).
272. Ho, C. M. & Marshall, G. R. Cavity search: an algorithm for the isolation and display of cavity-like binding regions. *J Comput Aided Mol Des* **4**, 337-54 (1990).
273. Exner, T. E., Keil, M., Moeckel, G. & Brickmann, J. Identification of substrate channels and protein cavities. *J Mol Model* **4**, 340-343 (1998).
274. Levitt, D. G. & Banaszak, L. J. POCKET: a computer graphics method for identifying and displaying protein cavities and their surrounding amino acids. *J Mol Graph* **10**, 229-34 (1992).
275. Hendlich, M., Rippmann, F. & Barnickel, G. LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. *J Mol Graph Model* **15**, 359-63, 389 (1997).
276. Huang, B. & Schroeder, M. LIGSITEcsc: predicting ligand binding sites using the Connolly surface and degree of conservation. *BMC Struct Biol* **6**, 19 (2006).
277. An, J., Totrov, M. & Abagyan, R. Comprehensive identification of "druggable" protein ligand binding sites. *Genome Inform* **15**, 31-41 (2004).
278. Laurie, A. T. & Jackson, R. M. Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics* **21**, 1908-16 (2005).
279. Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem* **28**, 849-57 (1985).
280. Pisabarro, M. T. & Serrano, L. Rational design of specific high-affinity peptide ligands for the Abl-SH3 domain. *Biochemistry* **35**, 10634-40 (1996).
281. Weskamp, N., Hullermeier, E., Kuhn, D. & Klebe, G. Multiple graph alignment for the structural analysis of protein active sites. *IEEE/ACM Trans Comput Biol Bioinform* **4**, 310-20 (2007).
282. Hajduk, P. J., Huth, J. R. & Tse, C. Predicting protein druggability. *Drug Discov Today* **10**, 1675-82 (2005).
283. An, J., Totrov, M. & Abagyan, R. Pocketome via comprehensive identification and classification of ligand binding envelopes. *Mol Cell Proteomics* **4**, 752-61 (2005).
284. Weisel, M., Proschak, E. & Schneider, G. PocketPicker: analysis of ligand binding-sites with shape descriptors. *Chem Cent J* **1**, 7 (2007).
285. Weisel, M., Proschak, E., Kriegl, J. M. & Schneider, G. Form follows function: shape analysis of protein cavities for receptor-based drug design. *Proteomics* **9**, 451-9 (2009).
286. Schalon, C., Surgand, J. S., Kellenberger, E. & Rognan, D. A simple and fuzzy method to align and compare druggable ligand-binding sites. *Proteins* **71**, 1755-78 (2008).
287. Kinoshita, K., Furui, J. & Nakamura, H. Identification of protein functions from a molecular surface database, eF-site. *J Struct Funct Genomics* **2**, 9-22 (2002).
288. Schmitt, S., Kuhn, D. & Klebe, G. A new method to detect related function among proteins independent of sequence and fold homology. *J Mol Biol* **323**, 387-406 (2002).
289. Jambon, M., Imbert, A., Deleage, G. & Geourjon, C. A new bioinformatic approach to detect common 3D sites in protein structures. *Proteins* **52**, 137-45 (2003).
290. Brakoulias, A. & Jackson, R. M. Towards a structural classification of phosphate binding sites in protein-nucleotide complexes: an automated all-against-all structural comparison using geometric matching. *Proteins* **56**, 250-60 (2004).
291. Shulman-Peleg, A., Nussinov, R. & Wolfson, H. J. Recognition of functional sites in protein structures. *J Mol Biol* **339**, 607-33 (2004).
292. Ferre, F., Ausiello, G., Zanzoni, A. & Helmer-Citterich, M. Functional annotation by identification of local surface similarities: a novel tool for structural genomics. *BMC Bioinformatics* **6**, 194 (2005).

293. Powers, R. et al. Comparison of protein active site structures for functional annotation of proteins and drug design. *Proteins* **65**, 124-35 (2006).
294. Zhou, H. X. & Shan, Y. Prediction of protein interaction sites from sequence profile and residue neighbor list. *Proteins* **44**, 336-43 (2001).
295. Caffrey, D. R., Somaroo, S., Hughes, J. D., Mintseris, J. & Huang, E. S. Are protein-protein interfaces more conserved in sequence than the rest of the protein surface? *Protein Sci* **13**, 190-202 (2004).
296. Neuvirth, H., Raz, R. & Schreiber, G. ProMate: a structure based prediction program to identify the location of protein-protein binding sites. *J Mol Biol* **338**, 181-99 (2004).
297. Bordner, A. J. & Abagyan, R. Statistical analysis and prediction of protein-protein interfaces. *Proteins* **60**, 353-66 (2005).
298. Watson, J. D., Laskowski, R. A. & Thornton, J. M. Predicting protein function from sequence and structural data. *Curr Opin Struct Biol* **15**, 275-84 (2005).
299. Todd, A. E., Orengo, C. A. & Thornton, J. M. Evolution of function in protein superfamilies, from a structural perspective. *J Mol Biol* **307**, 1113-43 (2001).
300. Connolly, M. L. Solvent-accessible surfaces of proteins and nucleic acids. *Science* **221**, 709-13 (1983).
301. Kinoshita, K. & Nakamura, H. Identification of the ligand binding sites on the molecular surface of proteins. *Protein Sci* **14**, 711-8 (2005).
302. Kinoshita, K., Murakami, Y. & Nakamura, H. eF-seek: prediction of the functional sites of proteins by searching for similar electrostatic potential and molecular surface shape. *Nucleic Acids Res* **35**, W398-402 (2007).
303. Grabowski, M., Joachimiak, A., Otwinowski, Z. & Minor, W. Structural genomics: keeping up with expanding knowledge of the protein universe. *Curr Opin Struct Biol* **17**, 347-53 (2007).
304. Aloy, P. & Russell, R. B. Structural systems biology: modelling protein interactions. *Nat Rev Mol Cell Biol* **7**, 188-97 (2006).
305. Brehme, M. et al. Charting the molecular network of the drug target Bcr-Abl. *Proc Natl Acad Sci U S A* **106**, 7414-9 (2009).
306. Kozakov, D., Schueler-Furman, O. & Vajda, S. Discrimination of near-native structures in protein-protein docking by testing the stability of local minima. *Proteins* **72**, 993-1004 (2008).
307. Ritchie, D. W. Recent progress and future directions in protein-protein docking. *Curr Protein Pept Sci* **9**, 1-15 (2008).
308. Smith, G. R. & Sternberg, M. J. Prediction of protein-protein interactions by docking methods. *Curr Opin Struct Biol* **12**, 28-35 (2002).
309. Pierce, B. & Weng, Z. A combination of rescoring and refinement significantly improves protein docking performance. *Proteins* **72**, 270-9 (2008).
310. Hwang, H., Pierce, B., Mintseris, J., Janin, J. & Weng, Z. Protein-protein docking benchmark version 3.0. *Proteins* **73**, 705-9 (2008).
311. Dominguez, C., Boelens, R. & Bonvin, A. M. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* **125**, 1731-7 (2003).
312. de Vries, S. J. et al. HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins* **69**, 726-33 (2007).
313. Lyskov, S. & Gray, J. J. The RosettaDock server for local protein-protein docking. *Nucleic Acids Res* **36**, W233-8 (2008).
314. Mashiach, E., Schneidman-Duhovny, D., Andrusier, N., Nussinov, R. & Wolfson, H. J. FireDock: a web server for fast interaction refinement in molecular docking. *Nucleic Acids Res* **36**, W229-32 (2008).
315. Najmanovich, R., Kuttner, J., Sobolev, V. & Edelman, M. Side-chain flexibility in proteins upon ligand binding. *Proteins* **39**, 261-8 (2000).
316. Kaihara, A., Kawai, Y., Sato, M., Ozawa, T. & Umezawa, Y. Locating a protein-protein interaction in living cells via split Renilla luciferase complementation. *Anal Chem* **75**, 4176-81 (2003).
317. Yu, H. et al. Measuring drug action in the cellular context using protein-fragment complementation assays. *Assay Drug Dev Technol* **1**, 811-22 (2003).
318. Auerbach, D., Thaminy, S., Hottiger, M. O. & Stagljar, I. The post-genomic era of interactive proteomics: facts and perspectives. *Proteomics* **2**, 611-23 (2002).
319. Zhang, B., Kraemer, B., SenGupta, D., Fields, S. & Wickens, M. Yeast three-hybrid system to detect and analyze interactions between RNA and protein. *Methods Enzymol* **306**, 93-113 (1999).

320. Arkin, M. & Lear, J. D. A new data analysis method to determine binding constants of small molecules to proteins using equilibrium analytical ultracentrifugation with absorption optics. *Anal Biochem* **299**, 98-107 (2001).
321. Pagliaro, L. et al. Emerging classes of protein-protein interaction inhibitors and new tools for their development. *Curr Opin Chem Biol* **8**, 442-9 (2004).
322. Schreiber, S. L. Chemical genetics. *Chem Eng News* **81**, 51-61 (2003).
323. Arkin, M. R. & Wells, J. A. Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat Rev Drug Discov* **3**, 301-17 (2004).
324. Cochran, A. G. Antagonists of protein-protein interactions. *Chem Biol* **7**, R85-94 (2000).
325. Prudencio, M. & Ubbink, M. Transient complexes of redox proteins: structural and dynamic details from NMR studies. *J Mol Recognit* **17**, 524-39 (2004).
326. Roisman, L. C., Jaitin, D. A., Baker, D. P. & Schreiber, G. Mutational analysis of the IFNAR1 binding site on IFNalpha2 reveals the architecture of a weak ligand-receptor binding-site. *J Mol Biol* **353**, 271-81 (2005).
327. Sharma, S. V. et al. UCS15A, a non-kinase inhibitor of Src signal transduction. *Oncogene* **20**, 2068-79 (2001).
328. Oneyama, C., Nakano, H. & Sharma, S. V. UCS15A, a novel small molecule, SH3 domain-mediated protein-protein interaction blocking drug. *Oncogene* **21**, 2037-50 (2002).
329. Atatreh, N. et al. In silico screening and biological evaluation of inhibitors of Src-SH3 domain interaction with a proline-rich ligand. *Bioorg Med Chem Lett* **18**, 1217-22 (2008).
330. Irwin, J. J. & Shoichet, B. K. ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model* **45**, 177-82 (2005).
331. Griffith, O. W. & Stuehr, D. J. Nitric oxide synthases: properties and catalytic mechanism. *Annu Rev Physiol* **57**, 707-36 (1995).
332. Masters, B. S. et al. Neuronal nitric oxide synthase, a modular enzyme formed by convergent evolution: structure studies of a cysteine thiolate-ligated heme protein that hydroxylates L-arginine to produce NO. as a cellular signal. *Faseb J* **10**, 552-8 (1996).
333. Blasko, E. et al. Mechanistic studies with potent and selective inducible nitric-oxide synthase dimerisation inhibitors. *J Biol Chem* **277**, 295-302 (2002).
334. McMillan, K. et al. Allosteric inhibitors of inducible nitric oxide synthase dimerisation discovered via combinatorial chemistry. *Proc Natl Acad Sci U S A* **97**, 1506-11 (2000).
335. Niederhauser, O. et al. NGF ligand alters NGF signaling via p75(NTR) and trkA. *J Neurosci Res* **61**, 263-72 (2000).
336. Kraich, M. et al. A modular interface of IL-4 allows for scalable affinity without affecting specificity for the IL-4 receptor. *BMC Biol* **4**, 13 (2006).
337. Jacoby, E. Biphenyls as potential mimetics of protein alpha-helix. *Bioorg Med Chem Lett* **12**, 891-3 (2002).
338. Ernst, J. T. et al. Design of a protein surface antagonist based on alpha-helix mimicry: inhibition of gp41 assembly and viral fusion. *Angew Chem Int Ed Engl* **41**, 278-81 (2002).
339. Gaudio, A. C., Richards, W. G. & Takahata, Y. QSAR and molecular graphics analysis of N2-phenylguanines as inhibitors of herpes simplex virus thymidine kinases. *J Mol Graph Model* **18**, 33-41 (2000).
340. Sattler, M. et al. Structure of Bcl-xL-Bak peptide complex: recognition between regulators of apoptosis. *Science* **275**, 983-6 (1997).
341. Kutzki, O. et al. Development of a potent Bcl-x(L) antagonist based on alpha-helix mimicry. *J Am Chem Soc* **124**, 11838-9 (2002).
342. Wang, J. L. et al. Structure-based discovery of an organic compound that binds Bcl-2 protein and induces apoptosis of tumor cells. *Proc Natl Acad Sci U S A* **97**, 7124-9 (2000).
343. Enyedy, I. J. et al. Discovery of small-molecule inhibitors of Bcl-2 through structure-based computer screening. *J Med Chem* **44**, 4313-24 (2001).
344. Degterev, A. et al. Identification of small-molecule inhibitors of interaction between the BH3 domain and Bcl-xL. *Nat Cell Biol* **3**, 173-82 (2001).
345. Kim, K. M. et al. Biophysical characterization of recombinant human Bcl-2 and its interactions with an inhibitory ligand, antimycin A. *Biochemistry* **40**, 4911-22 (2001).
346. Tzung, S. P. et al. Antimycin A mimics a cell-death-inducing Bcl-2 homology domain 3. *Nat Cell Biol* **3**, 183-91 (2001).
347. Lugovskoy, A. A. et al. A novel approach for characterizing protein ligand complexes: molecular basis for specificity of small-molecule Bcl-2 inhibitors. *J Am Chem Soc* **124**, 1234-40 (2002).

348. Zeitlin, B. D., Zeitlin, I. J. & Nor, J. E. Expanding circle of inhibition: small-molecule inhibitors of Bcl-2 as anticancer cell and antiangiogenic agents. *J Clin Oncol* **26**, 4180-8 (2008).
349. Segers, K. et al. Design of protein membrane interaction inhibitors by virtual ligand screening, proof of concept with the C2 domain of factor V. *Proc Natl Acad Sci U S A* **104**, 12697-702 (2007).
350. Sperandio, O., Miteva, M. A., Segers, K., Nicolaes, G. A. & Villoutreix, B. O. Screening Outside the Catalytic Site: Inhibition of Macromolecular Inter-actions Through Structure-Based Virtual Ligand Screening Experiments. *Open Biochem J* **2**, 29-37 (2008).
351. Kuriyan, J. & Cowburn, D. Modular peptide recognition domains in eukaryotic signaling. *Annu Rev Biophys Biomol Struct* **26**, 259-88 (1997).
352. McPhail, L. C. SH3-dependent assembly of the phagocyte NADPH oxidase. *J Exp Med* **180**, 2011-5 (1994).
353. Ball, L. J., Kuhne, R., Schneider-Mergener, J. & Oschkinat, H. Recognition of proline-rich motifs by protein-protein-interaction domains. *Angew Chem Int Ed Engl* **44**, 2852-69 (2005).
354. Ren, R., Mayer, B. J., Cicchetti, P. & Baltimore, D. Identification of a ten-amino acid proline-rich SH3 binding site. *Science* **259**, 1157-61 (1993).
355. Yu, H. et al. Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell* **76**, 933-45 (1994).
356. Chen, J. K., Lane, W. S., Brauer, A. W., Tanaka, A. & Schreiber, S. L. Biased Combinatorial Libraries: Novel Ligands for the SH3 Domain of Phosphatidylinositol 3-Kinase. *JACS* **115**, 12591-12592 (1993).
357. Musacchio, A., Saraste, M. & Wilmanns, M. High-resolution crystal structures of tyrosine kinase SH3 domains complexed with proline-rich peptides. *Nat Struct Biol* **1**, 546-51 (1994).
358. Brown, F. K. *Annual Reports in Medicinal Chemistry* **33**, 375-384 (1998).
359. Manly, C. J., Louise-May, S. & Hammer, J. D. The impact of informatics and computational chemistry on synthesis and screening. *Drug Discov Today* **6**, 1101-1110 (2001).
360. Hann, M. & Green, R. Chemoinformatics--a new name for an old problem? *Curr Opin Chem Biol* **3**, 379-83 (1999).
361. Dean, P. M., Zanders, E. D. & Bailey, D. S. Industrial-scale, genomics-based drug design and discovery. *Trends Biotechnol* **19**, 288-92 (2001).
362. Holliday, J. D., Hu, C. Y. & Willett, P. Grouping of coefficients for the calculation of inter-molecular similarity and dissimilarity using 2D fragment bit-strings. *Comb Chem High Throughput Screen* **5**, 155-66 (2002).
363. Senger, S. Using Tversky similarity searches for core hopping: finding the needles in the haystack. *J Chem Inf Model* **49**, 1514-24 (2009).
364. Deng, Z., Chuaqui, C. & Singh, J. Structural interaction fingerprint (SIFt): a novel method for analyzing three-dimensional protein-ligand binding interactions. *J Med Chem* **47**, 337-44 (2004).
365. Singh, J., Deng, Z., Narale, G. & Chuaqui, C. Structural interaction fingerprints: a new approach to organizing, mining, analyzing, and designing protein-small molecule complexes. *Chem Biol Drug Des* **67**, 5-12 (2006).
366. Schindler, T. et al. Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science* **289**, 1938-42 (2000).
367. Mpamhanga, C. P., Chen, B., McLay, I. M. & Willett, P. Knowledge-based interaction fingerprint scoring: a simple method for improving the effectiveness of fast scoring functions. *J Chem Inf Model* **46**, 686-98 (2006).
368. Marcou, G. & Rognan, D. Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J Chem Inf Model* **47**, 195-207 (2007).
369. Schuffenhauer, A., Floersheim, P., Acklin, P. & Jacoby, E. Similarity metrics for ligands reflecting the similarity of the target proteins. *J Chem Inf Comput Sci* **43**, 391-405 (2003).
370. Fradera, X. et al. Ligand-induced changes in the binding sites of proteins. *Bioinformatics* **18**, 939-48 (2002).
371. Xia, X., Maliski, E. G., Gallant, P. & Rogers, D. Classification of kinase inhibitors using a Bayesian model. *J Med Chem* **47**, 4463-70 (2004).
372. Schneider, G., Neidhart, W., Giller, T. & Schmid, G. "Scaffold-Hopping" by Topological Pharmacophore Search: A Contribution to Virtual Screening. *Angew Chem Int Ed Engl* **38**, 2894-2896 (1999).
373. Schneider, G. et al. Virtual Screening for Bioactive Molecules by Evolutionary De Novo Design Special thanks to Neil R. Taylor for his help in preparation of the manuscript. *Angew Chem Int Ed Engl* **39**, 4130-4133 (2000).

374. Jenkins, J. L., Glick, M. & Davies, J. W. A 3D similarity method for scaffold hopping from known drugs or natural ligands to new chemotypes. *J Med Chem* **47**, 6144-59 (2004).
375. Renner, S. & Schneider, G. Scaffold-hopping potential of ligand-based similarity concepts. *ChemMedChem* **1**, 181-5 (2006).
376. Congreve, M., Carr, R., Murray, C. & Jhoti, H. A 'rule of three' for fragment-based lead discovery? *Drug Discov Today* **8**, 876-7 (2003).
377. Verdonk, M. L. & Hartshorn, M. J. Structure-guided fragment screening for lead discovery. *Curr Opin Drug Discov Devel* **7**, 404-10 (2004).
378. Shuker, S. B., Hajduk, P. J., Meadows, R. P. & Fesik, S. W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **274**, 1531-4 (1996).
379. Nienaber, V. L. et al. Discovering novel ligands for macromolecules using X-ray crystallographic screening. *Nat Biotechnol* **18**, 1105-8 (2000).
380. Carr, R. & Jhoti, H. Structure-based screening of low-affinity compounds. *Drug Discov Today* **7**, 522-7 (2002).
381. Erlanson, D. A. et al. In situ assembly of enzyme inhibitors using extended tethering. *Nat Biotechnol* **21**, 308-14 (2003).
382. Rees, D. C., Congreve, M., Murray, C. W. & Carr, R. Fragment-based lead discovery. *Nat Rev Drug Discov* **3**, 660-72 (2004).
383. Oprea, T. I., Davis, A. M., Teague, S. J. & Leeson, P. D. Is there a difference between leads and drugs? A historical perspective. *J Chem Inf Comput Sci* **41**, 1308-15 (2001).
384. Howard, N. et al. Application of fragment screening and fragment linking to the discovery of novel thrombin inhibitors. *J Med Chem* **49**, 1346-55 (2006).
385. Lewell, X. Q., Judd, D. B., Watson, S. P. & Hann, M. M. RECAP--retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. *J Chem Inf Comput Sci* **38**, 511-22 (1998).
386. Frye, S. V. Structure-activity relationship homology (SARAH): a conceptual framework for drug discovery in the genomic era. *Chem Biol* **6**, R3-7 (1999).
387. Muller, G. Medicinal chemistry of target family-directed masterkeys. *Drug Discov Today* **8**, 681-91 (2003).
388. Evans, B. E. et al. Methods for drug discovery: development of potent, selective, orally effective cholecystokinin antagonists. *J Med Chem* **31**, 2235-46 (1988).
389. Mulder, N. J. et al. InterPro, progress and status in 2005. *Nucleic Acids Res* **33**, D201-5 (2005).
390. Russ, A. P. & Lampel, S. The druggable genome: an update. *Drug Discov Today* **10**, 1607-10 (2005).
391. Surgand, J. S., Rodrigo, J., Kellenberger, E. & Rognan, D. A chemogenomic analysis of the transmembrane binding cavity of human G-protein-coupled receptors. *Proteins* **62**, 509-38 (2006).
392. Fernandez, A. & Berry, R. S. Molecular dimension explored in evolution to promote proteomic complexity. *Proc Natl Acad Sci U S A* **101**, 13460-5 (2004).
393. Deremble, C. & Lavery, R. Macromolecular recognition. *Curr Opin Struct Biol* **15**, 171-5 (2005).
394. Fernandez, A. & Scott, L. R. Adherence of packing defects in soluble proteins. *Phys Rev Lett* **91**, 018102 (2003).
395. Wermuth, C. G. Selective optimization of side activities: another way for drug discovery. *J Med Chem* **47**, 1303-14 (2004).
396. Kastenholz, M. A., Pastor, M., Cruciani, G., Haaksma, E. E. & Fox, T. GRID/CPCA: a new computational tool to design selective ligands. *J Med Chem* **43**, 3033-44 (2000).
397. Naumann, T. & Matter, H. Structural classification of protein kinases using 3D molecular interaction field analysis of their ligand binding sites: target family landscapes. *J Med Chem* **45**, 2366-78 (2002).
398. Westerhuis, J. A., Kourti, T. & MacGregor, J. F. Analysis of multiblock and hierarchical PCA and PLS models. *J Chemometrics* **12**, 301-321 (1998).
399. Barillari, C., Marcou, G. & Rognan, D. Hot-spots-guided receptor-based pharmacophores (HS-Pharm): a knowledge-based approach to identify ligand-anchoring atoms in protein cavities and prioritize structure-based pharmacophores. *J Chem Inf Model* **48**, 1396-410 (2008).
400. Kauvar, L. M. et al. Predicting ligand binding to proteins by affinity fingerprinting. *Chem Biol* **2**, 107-18 (1995).

401. Denessiouk, K. A., Lehtonen, J. V., Korpela, T. & Johnson, M. S. Two "unrelated" families of ATP-dependent enzymes share extensive structural similarities about their cofactor binding sites. *Protein Sci* **7**, 1136-46 (1998).
402. Denessiouk, K. A. & Johnson, M. S. When fold is not important: a common structural framework for adenine and AMP binding in 12 unrelated protein families. *Proteins* **38**, 310-26 (2000).
403. Denessiouk, K. A., Rantanen, V. V. & Johnson, M. S. Adenine recognition: a motif present in ATP-, CoA-, NAD-, NADP-, and FAD-dependent proteins. *Proteins* **44**, 282-91 (2001).
404. Vieth, M. et al. Kinomics-structural biology and chemogenomics of kinase inhibitors and targets. *Biochim Biophys Acta* **1697**, 243-57 (2004).
405. Fabian, M. A. et al. A small molecule-kinase interaction map for clinical kinase inhibitors. *Nat Biotechnol* **23**, 329-36 (2005).
406. Mestres, J., Martin-Couce, L., Gregori-Puigjane, E., Cases, M. & Boyer, S. Ligand-based approach to in silico pharmacology: nuclear receptor profiling. *J Chem Inf Model* **46**, 2725-36 (2006).
407. Laskowski, R. A. PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res* **29**, 221-2 (2001).
408. Hutchinson, E. G. & Thornton, J. M. PROMOTIF--a program to identify and analyze structural motifs in proteins. *Protein Sci* **5**, 212-20 (1996).
409. Hofmann, K., Bucher, P., Falquet, L. & Bairoch, A. The PROSITE database, its status in 1999. *Nucleic Acids Res* **27**, 215-9 (1999).
410. Falquet, L. et al. The PROSITE database, its status in 2002. *Nucleic Acids Res* **30**, 235-8 (2002).
411. Kasuya, A. & Thornton, J. M. Three-dimensional structure analysis of PROSITE patterns. *J Mol Biol* **286**, 1673-91 (1999).
412. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng* **8**, 127-34 (1995).
413. Schreyer, A. & Blundell, T. CREDO: a protein-ligand interaction database for drug discovery. *Chem Biol Drug Des* **73**, 157-67 (2009).
414. Wishart, D. S. et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res* **34**, D668-72 (2006).
415. Wishart, D. S. et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res* **36**, D901-6 (2008).
416. Goto, S., Okuno, Y., Hattori, M., Nishioka, T. & Kanehisa, M. LIGAND: database of chemical compounds and reactions in biological pathways. *Nucleic Acids Res* **30**, 402-4 (2002).
417. Chen, J. et al. MMDB: Entrez's 3D-structure database. *Nucleic Acids Res* **31**, 474-7 (2003).
418. Ivanisenko, V. A., Pintus, S. S., Grigorovich, D. A. & Kolchanov, N. A. PDBSITE: a database of the 3D structure of protein functional sites. *Nucleic Acids Res* **33**, D183-7 (2005).
419. Gold, N. D. & Jackson, R. M. SitesBase: a database for structure-based protein-ligand binding site comparisons. *Nucleic Acids Res* **34**, D231-4 (2006).
420. Porter, C. T., Bartlett, G. J. & Thornton, J. M. The Catalytic Site Atlas: a resource of catalytic sites and residues identified in enzymes using structural data. *Nucleic Acids Res* **32**, D129-33 (2004).
421. Stuart, A. C., Ilyin, V. A. & Sali, A. LigBase: a database of families of aligned ligand binding sites in known protein sequences and structures. *Bioinformatics* **18**, 200-1 (2002).
422. Golovin, A., Dimitropoulos, D., Oldfield, T., Rachedi, A. & Henrick, K. MSDsite: a database search and retrieval system for the analysis and viewing of bound ligands and active sites. *Proteins* **58**, 190-9 (2005).
423. Kinoshita, K. & Nakamura, H. eF-site and PDBjViewer: database and viewer for protein functional sites. *Bioinformatics* **20**, 1329-30 (2004).
424. Michalsky, E., Dunkel, M., Goede, A. & Preissner, R. SuperLigands - a database of ligand structures derived from the Protein Data Bank. *BMC Bioinformatics* **6**, 122 (2005).
425. Yamaguchi, A. et al. Het-PDB Navi.: a database for protein-small molecule interactions. *J Biochem* **135**, 79-84 (2004).

CHAPTER 1 – *IN SILICO* DETERMINATION OF MICROMOLAR SUBSTRATE AFFINITIES USING STEERED MOLECULAR DYNAMICS

ABSTRACT

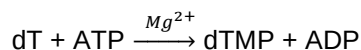
Rational drug design is based on deciphering of protein-ligand interactions, particularly the calculation of the binding free energy. In analogy to single atom pulling experiments in atomic force microscopy, steered molecular dynamics (SMD) can be used to pull out a molecule from an enzyme's active site, with the applied force (work) being proportional to the binding affinity. Colizzi et al. have recently shown for the first time that such an approach can be used to discern active from inactive inhibitors of the *Plasmodium falciparum* β -hydroxyacyl-acyl carrier protein dehydratase⁴²⁶, which is promising for automated *in silico* inhibitor selection⁴²⁷.

In the present work, the above method is applied to affinity calculations of μM substrates of the *Herpes Simplex Virus 1* Thymidine Kinase (HSV-1 TK) to foster the design of HSV-1 TK-based Positron Emission Tomography (PET) tracers to be used in monitoring TK activity in stem cell therapy or cancer imaging. At the origin of this work was the finding that substrates of both HSV-1 TK and the human TK 1 (hTK1) docked into each enzyme did not yield an activity score-based ranking correlating with known K_i values, revealing that docking alone was insufficient to explain substrate affinity. Classical Molecular Dynamics (CMD) was used to assess the stability of both enzymes. Subsequently, main motions were extracted from the CMDs with Principal Component Analysis (PCA). Interestingly, this analysis shows that the LID domain known to open upon substrate entry undergoes a clear hinge movement. Accessible substrate-bound conformations were then obtained from a closed holo crystal structure by SMD, possible substrate egress routes identified by Random Acceleration MD (RAMD) and one of these directions then used to pull the substrate out of the enzyme's embrace by SMD. The *in silico* pulling work needed for substrate extraction from HSV-1 TK correlates perfectly with experimentally known μM substrate affinities, allowing for distinguishing binding affinity differences of 4.3 μM , and is used to predict the affinities of (fluorinated) substrate analogues. In comparison, pulling substrates out of hTK1, which has a narrower substrate acceptance than HSV-1 TK, clearly shows that the method can be applied if the compounds carry modifications at both the base and the sugar. Thus, the here presented approach is useful in guiding PET tracer design prior to synthesis.

1.1 INTRODUCTION

1.1.1 INTRODUCING THE PLAYERS: THYMIDINE KINASES (TKS)

Thymidine kinases (TKs) catalyze the rate-limiting step of the DNA salvage pathway, which is important in DNA repair. It constitutes an alternative to the *de novo* DNA synthesis for pyrimidines⁴²⁸. TK phosphorylates deoxythymidine (dT), provided by food intake and imported by carrier proteins into cells, in the presence of adenosine 5'-triphosphate (ATP) and Mg^{2+} to thymidine monophosphate (dTMP) and adenosine monophosphate (AMP):



During this reaction, the γ -phosphate of ATP is transferred to the 5'-OH of the ribose of dT. dTMP is further phosphorylated in the presence of ATP to deoxythymidine diphosphate (dTDP) by the thymidylate kinase and to deoxythymidine triphosphate (dTTP) by the nucleoside diphosphate kinase. dTTP remains trapped inside the cell due to its negative charge and is subsequently incorporated into DNA by a DNA polymerase. The salvage pathway is counteracted by the enzyme 5'-nucleotidase (reviewed in⁴²⁹), hydrolysing the ester bond between the 5'-carbon and the phosphate in nucleoside monophosphates. TKs are known to be stabilized by the substrate dT⁴³⁰, inhibited by the product dTMP and feedback inhibited by dTTP⁴³¹⁻⁴³³.

According to their length, two different TK types can be distinguished, the long and the short TKs. Their properties are summarized in Table 8 together with the representatives further studied in this work, the *Herpes Simplex Virus 1* TK (HSV-1 TK) and the human thymidine kinase 1 (hTK1).

Table 8: Comparison of type I and type II TKs compiled from ⁴²⁹ and the references found therein.

Features	Long TKs (Type I)	Short TKs (Type II)
Molecular weight	40 kDa	25 kDa
Active as	Homodimers	Homotetramers
Location	Nucleus or cytosol	Cytosol
Expression	Constantly expressed	In replicating cells
Substrates	Pyrimidines, purines, 5-substituted nucleotide analogues	Deoxythymidine (dT), deoxyuridine (dU)
Example	Herpes Simplex Virus 1 TK (HSV-1 TK)	Human TK 1 (hTK1)

Long TKs such as HSV-1 TK are characterized by a broad substrate specificity, while short TKs such as hTK1 display a much narrower range of substrates. This is because the former category bind substrates mainly via side chains, allowing them to adapt, whereas the latter bind them via the rather rigid backbone. While HSV-1 TK phosphorylates pyrimidines, purines and substrate analogues with bulky 5-substitutions, hTK1 indeed only phosphorylates its natural substrates dT and deoxyuridine (dU), and substrate analogues with minor changes in the sugar moiety ⁴³⁴, i.e. small 5- (halogens or ethyl) or 3'-substitutions (H, halogens or azido) ^{429, 435-438}, or larger (not too bulky ⁴³⁹) substitutions at the N3 of the base ⁴⁴⁰⁻⁴⁴⁸. The therapeutically most relevant substitution consists in the introduction of an azido group at C3' leading to 3'-azido-3'-deoxythymidine (AZT), an anti-HIV drug ⁴⁴⁹ that is phosphorylated by hTK1 ⁴³⁵.

1.1.2 THERAPEUTIC APPLICATIONS INVOLVING TKs

Antiviral therapy was the first therapeutic application exploiting the difference in substrate acceptance between hTK1 and HSV-1 TK: Antiviral purine-based prodrugs are exclusively phosphorylated by HSV-1 TK and act as viral DNA chain terminators in their triphosphorylated form.

Within the gene therapy approach, the most studied ⁴⁵⁰⁻⁴⁵³ and most effective ⁴⁵⁴ enzyme/prodrug approach is HSV-1 TK/ganciclovir (GCV), with GCV having a suitable pharmacodynamic/pharmacokinetic (PD/PK) profile ^{455, 456}, but leading to several complications such as e.g. the T-cell mediated recognition of the host as "foreign" ⁴⁵⁷, and haematological toxicity ⁴⁵⁸.

Both hTK1 and HSV-1 TK can be used in non-invasive imaging via PET. Because hTK1 is only highly expressed in replicating cells during the S-phase of the cell cycle ⁴⁵⁹, cancer cell proliferation can be followed with e.g. [¹⁸F]-fluoro-3'-deoxy-3'-fluorothymidine ([¹⁸F]FLT) ⁴⁶⁰⁻⁴⁶². HSV-1 TK serves as a reporter gene in gene therapy, where the *in vivo* transduction efficiency, i.e. its location, extent and duration of expression can be monitored ⁴⁶³. Briefly, a radio-labelled substrate gets mono-phosphorylated by HSV-1 TK and remains trapped due to its highly negative charge after two subsequent phosphorylations in cells expressing HSV-1 TK. The sequestered product is assumed to have the same transport and metabolism properties as the natural substrate. ⁴⁶⁴ HSV-1 TK has been experimentally validated ⁴⁶⁵⁻⁴⁶⁸ as a tool for PET imaging, is known to be constitutively expressed ⁴⁶⁹, ⁴⁷⁰ and radio-labelled probes with different efficacies exist ⁴⁶⁵, e.g. fluorinated purines ^{471, 472} and pyrimidines ⁴⁷³. Pyrimidine-based HSV-1 TK substrates have no or a lower bystander effect ⁴⁷⁴ (i.e. killing of non-transfected cells when co-cultured with transfected cells) compared to purine-based ones ⁴⁷⁵ and are therefore preferable for PET tracing. Suitable characteristics for a HSV-1 TK radiotracer are summarized by Gambhir et al. ⁴⁷⁶. To assess experimentally whether a non-radioactive precursor is a substrate, UV-spectrophotometric ⁴⁷⁷ or HPLC assays ^{478, 479} are usually performed, while K_i values are obtained from kinetics measurements monitoring the conversion of labelled dT into dTMP in presence of various concentrations of a non-labelled (potential) substrate under Michaelis Menten conditions ⁴⁸⁰.

1.1.3 STRUCTURAL FEATURES OF TKs AND THEIR LIGANDS

HSV-1 TK (enzyme classification (EC) number: 2.7.1.21) counts 379 amino acids organized in 15 α -helices and 7 β -strands, phosphorylates both pyrimidine- and purine-based substrates, has a dual thymidine/thymidylate kinase function and was the first deoxyribonucleoside kinase to have its dimeric crystal structure solved ^{481, 482}. It features a five-stranded β -sheet at the substrate-binding core of each monomer, which is surrounded by helices and the ATP binding site is located in a loop ^{438, 482, 483}. This

loop is called the P-loop or glycine-rich loop (residues 56-63) and forms together with Arg216, Arg220, and Arg222 an anion hole that accommodates the β - and γ -phosphate of ATP^{481, 482, 484-486}. While ATP is solvent-exposed, the natural substrate dT is entirely inside the protein, but does not take advantage of the full binding site⁴⁸⁵ (Figure 12a). dT binding needs to happen before ATP binding, a finding that was corroborated by ITC measurements¹⁷⁴. HSV-1 TK shares the 5 central β -sheets and several surrounding α -helices, a nucleotide monophosphate binding domain and the LID domain with the isofunctional adenylate kinase (AdK). AdK is probably the most used protein for assessing large scale (apo/holo) conformational changes. Importantly, the LID of HSV-1 TK is Lys- and Arg-rich and its conformational changes influence the phosphorylation rate similarly to AdK⁴⁸⁷. Unique HSV-1 TK features in comparison to AdK include the N-terminal 45 amino acids, a partially mobile segment of 72 amino acids (250-322) and 29 C-terminal residues. HSV-1 TK residues involved in induced fit movements and dT binding include Ile97, Met128, Arg222, Pro223, Gly224, and Glu225⁴⁸⁰. They are conserved in many TKs^{488, 489}.

Binding of dT to HSV-1 TK is characterized by an extensive hydrogen bonding network (Figure 12b), among which the Watson-Crick-like interaction with Gln125 is notable, as its side chain orientation depends on the substrate type bound. Indeed, a flip by 180° is observed when comparing the pyrimidine with the purine binding mode⁴⁸¹. Thus, this residue acts as a “chemical chameleon”, adapting to two different substrate classes.

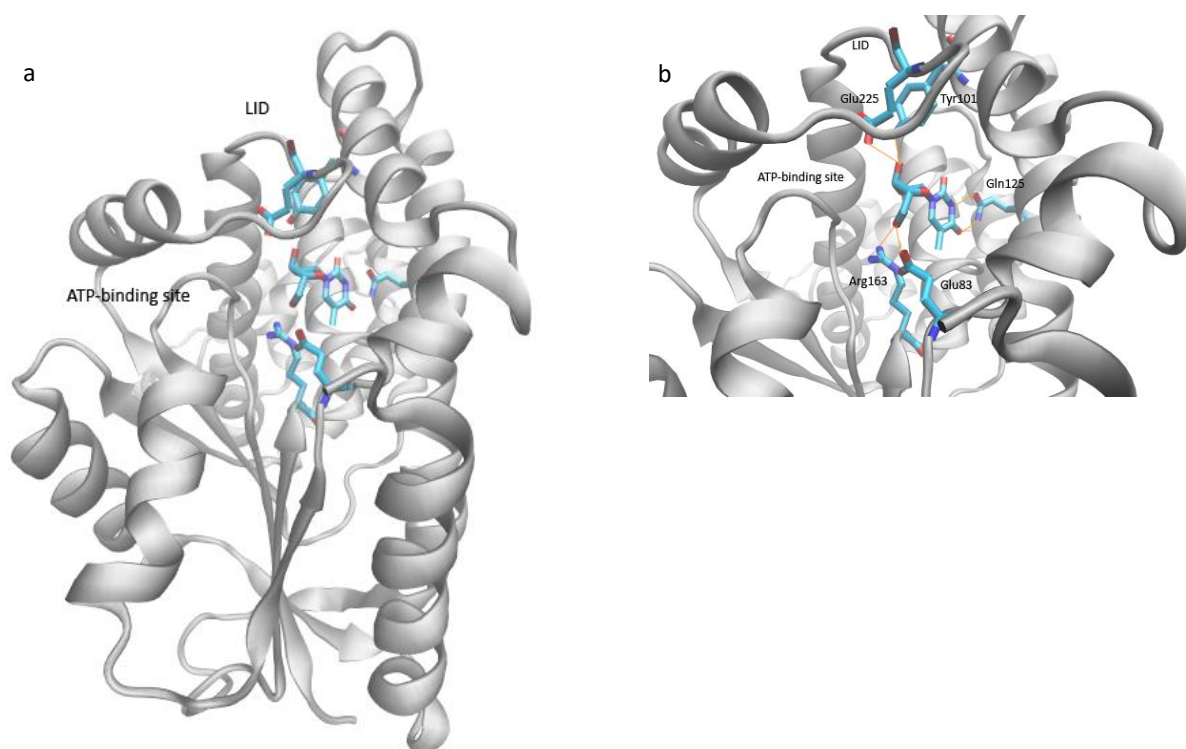


Figure 12: a) Overall structure of the HSV-1 TK monomer in complex with dT (PDB ID: 1KIM). The protein backbone appears in grey. Both dT and its H-bonding residues are depicted as sticks with cyan carbon atoms. b) The H-bonding network of HSV-1 TK involved in dT binding. The orange lines show the H-bonds between dT and five binding site residues of HSV-1 TK. H-bonds are usually 2.8 to 3.2 Å long and display an angle of 140°.

However, given the broad substrate acceptance of HSV-1 TK, H-bonds are not the only contributors to molecular recognition and binding. HSV-1 TK is tolerant with respect to modifications in the sugar moiety, which provide further anchors in contrast to the directionality dictating base⁴⁸⁶. The complete list of H-bonds and Van der Waals (VdW) contacts between HSV-1 TK and dT is provided in Table 9.

Table 9: Stabilizing VdW and H-bond interactions between dT and HSV-1 TK in the crystal structure 1KIM according to ⁴⁸⁶.

Type of interaction	dT		HSV-1 TK
VdW contacts	Base	Whole base	Met128, Ile100, Tyr172
H-bonds		2-carbonyl	2 water molecules interacting with the guanidinium of Arg176, the hydroxyl of Tyr101, and the side chain carbonyl of Gln125
		3-NH	Amide of Gln125
		4-carbonyl	Amide of Gln125
VdW contacts	Ribose	5-methyl	Side chains of Trp88, Tyr132, Arg163, relatively close also to Tyr172, and Ala167
H-bonds		3'OH	Tyr101, Glu225
		5'OH	Glu83; water-mediated H-bond to Arg163 H-bonding via another water to Asp162

Sandwich-like VdW interactions occur with Met128 and Tyr172 ⁴⁷⁵; the first is steric- electrostatic and the second essentially electrostatic according to density functional studies ⁴⁹⁰. Additional site-directed mutagenesis and ITC measurements showed a role of the residue triad His58/Met128/Tyr172 in substrate acceptance and Glu225 (part of the LID domain) in catalytic activity ⁴⁷⁸. By carrying out *ab initio* calculations on HSV-1 TK ligands and complementary biochemical studies, Sulpizi et al. ⁴⁹¹ noted that the electric dipole of ribose mimicking moieties pointed towards Glu225 and had a stabilizing role, while both Arg220 and Arg222 formed an anion hole and Glu83 acted as a base. Mutational studies further clarified the role of several amino acids: The residue Thr63 is located in front of the P-loop and involved in phosphoryl transfer and magnesium fixation ⁴⁸⁰, decreasing the size of Phe101 leads to a better acceptance of S-MCT (Table 12) and 5'-analogues as substrates, mutating Gln125 can lead to a pyrimidine-purine shift or a loss of phosphorylation, Ala167Phe affects pyrimidine phosphorylation ⁴⁹², Ala168Thr confers resistance ⁴⁸⁰ or hinders substrates with large C5-substituents to enter ⁴⁹³, Arg176Gln disturbs the electrostatic balance, and Glu225Leu ⁴⁷⁸ or Cys336Tyr ⁴⁸⁰ disturb the location of the LID domain.

Among different fluoroalkylated pyrimidine derivatives, especially 5-fluoroalkylated ^{494, 495} and C6-alkylated derivatives with the fluorination in the ribose mimicking side chain ⁴⁷³ hold promise as HSV-1 TK based PET tracers. The focus here is on the C6-alkylated pyrimidines DHBT and NMeDHBT, which were predicted by docking to share the same key interactions with HSV-1 TK as dT and confirmed by UV/VIS HPLC as not being substrates of hTK1 ⁴⁷³. Indeed, docking into the active site of hTK1 revealed that these compounds do not fit properly into it, as binding of the base via the backbone allows for less adaptation than in HSV-1 TK. In HSV-1 TK, the OH groups of both NMeDHBT and NMeFHBT mimic the 5'OH of dT. They point towards the catalytic residues Glu83 and Arg163, while the adjacent OHs are oriented towards the LID. This arrangement, which is beneficial to phosphorylation, has in the meantime been confirmed by a crystal structure of HSV-1 TK with NMeDHBT (PDB ID: 3F0T). Like in other holo structures of HSV-1 TK, Arg163 is the only residue situated in a disallowed region of Ramachandran plot. Compared to HSV-1 TK with DHBT (PDB ID: 1E2P), an additional water-mediated interaction between the dihydroxy-isobutyl and the side chains of Glu225 and Tyr101 is found. Together with the steric effect introduced by the Me group, the affinity of NMeDHBT towards HSV-1 TK is slightly increased in comparison to DHBT (Table 12). Both DHBT and NMeDHBT are not cytotoxic and have better *in vitro* binding affinities for HSV-1 TK than existing PET probes ^{473, 496}. Of the two OH groups, one can be replaced by ¹⁸F and one can be phosphorylated by HSV-1 TK. While fluorination of DHBT was rendered impossible by intramolecular cyclisation, the methylation in NMeDHBT effectively prevented it ⁴⁹⁷. Other compounds with the fluorine in the side chain (SMA1, 2, and 3 in Table 13) are hypothesized to be better taken up into cells *in vitro* than NMeFHBT, to have a higher *in vivo* stability and were confirmed by docking to assume a phosphorylation enabling conformation. A side-product of NMeFHBT, 6-(1,3-dihydroxyisobutenyl)-N-1-methylthymine (SK 7-8) was deduced from 1- and 2-dimensional ¹H- and ¹³C-NMR spectra. Its stereostructure was confirmed by X-ray crystallography and it was shown to selectively phosphorylate HSV-1 TK, but not hTK1 (L. Pernot, personal communication), which could also be seen in a later docking.

hTK1 is a homotetramer in the active state. TK1 crystal structures of the following species have been solved so far: *Homo sapiens* (hTK1) ^{434, 498}, *Vaccinia virus* ^{499, 500}, *Clostridium acetobutylicum* (Ca) ⁵⁰¹, *Bacillus cereus* (Bc) ⁵⁰², *Bacillus anthracis* (Ba) ⁵⁰², *Thermotoga maritima* (Tm) ⁵⁰⁰ and *Ureaplasma urealyticum* ⁵⁰². The highest (66%) sequence identity and most similar fold exist between the human and the *Vaccinia virus* TKs. Each subunit of hTK1 consists of two domains, one α/β -domain and a small zinc binding ⁵⁰³ domain, composed of Cys153, Cys156, Cys185, and Cys188 which are part of a rather flexible loop (Leu166-Lys180). This loop is stabilized by the zinc coordination and H-bond interactions, and resembles the zinc-binding motif of adenylate kinases (AdK) ⁵⁰⁴. The protein core is formed by a 6-stranded β -sheet surrounded by five α - and four 3_{10} -helices ⁴³⁴. The active site is located between the zinc binding domain and the lasso domain (Cys153-Asn234), which is in a closed conformation in the two human structures because of a stabilizing H-bond between dTTP and Val172. Bacterial TKs of Ca (1XX6: ADP) ⁵⁰¹, Bc ⁵⁰² (2I9R: apo) and Tm (2ORW) ⁵⁰⁰ are characterized by more open lasso domains. Especially the different Tm TK structures ⁵⁰⁵ and complementary fluorescence spectroscopy and mutagenesis studies showed that the structure remains largely unchanged upon substrate binding, except for the lasso loop and zinc binding domain. Here, the Ca TK structure was taken to model a semi-open conformation of hTK1, similarly to ^{447, 506}. The active site residues of hTK1 and their functions are listed in Table 10.

Table 10: Active site residues of hTK1 and their function according to ⁴³⁴.

Residue(s)	Function
Met28	Keeping dTTP in place, hydrophobic interactions
Asp58	H-bond to 3'OH of dT, contributes to the reaction velocity by positioning the catalytic residue Arg60
Arg60 and Glu98	Catalytic couple, present in all mammalian deoxyribonucleotide kinases; Arg60 fixes phosphate of TMP (equivalent to the α -phosphate of dTTP)
Ser33 and Asp98	Magnesium fixation
Phe101	Stacking against base of dT; reason why N-MCT can still bind, but S-MCT not
Val106	Mutations at this position can induce a permanent catalytic activity ⁵⁰⁷
Leu124	Keeping dTTP in place; hydrophobic interactions
Phe128 (amide)	H-bond with O4 of dT
Phe133	Stacking against base of dT
Thr163	Points towards C5; large substitutions as in BVDU not tolerated ⁴³⁵
Val172 (carbonyl)	H-bond with N3 of dT
Val174 (amide)	H-bond with O2 of dT
Tyr181	Stacking against base of dT

Summarizing, protein-ligand interactions between HSV-1 TK and a substrate involve exclusively amino acid side chains, while in hTK1 the substrate's base is solely interacting with the backbone. Important substrate and phosphate binding residues in both HSV-1 TK and hTK1 are given in Table 11.

Table 11: Substrate and phosphate binding residues in both HSV-1 TK and hTK1, with * denoting backbone interactions (adapted from M.V. Fogli's thesis ⁵⁰⁸).

Interactions	HSV-1 TK	hTK1
H-bonds	Gln125	Val172*, Phe128*, Val174*
VdW interactions	Glu83, Ile100, Tyr101, Met128, Tyr132, Ala168, Arg176	Met28, Thr127, Thr163
Stacking	Trp88, Ala124, Met128, Tyr172	Phe101, Phe133, Tyr181
O3' H-bonds	Tyr101, Glu225	Asp58, Gly176*
O5' surroundings	His58, Glu83, Ile97, Arg163, Arg222	Met28, Arg60, Glu98, Ile175
O5' phosphate	Met60, Lys62, Thr63, Thr64, Arg220, Arg222	Phe29*, Lys32, Ser33, Thr35

1.1.4 MOLECULAR DYNAMICS

Molecular dynamics (MD) deal with simulating the time evolution of a macromolecular system, e.g. its motions in space over time according to Newton's equation of motion:

$$\frac{dr_i(t)}{dt} = \frac{p_i(t)}{m_i} = f_i(t) = -\nabla U(r),$$

where (r) are the Cartesian coordinates of the atom $i = 1, \dots, N$, p the momentum, t the time, m the mass, f the force acting on i and $-\nabla U(r)$ the gradient of the potential energy U , specified by the force field.

While biophysical methods provide insights into static protein-ligand complexes, MD provides a complete dynamic picture on macromolecular processes, yielding information on atomic fluctuations, eventual conformational changes, and functionally relevant states. Indeed, dynamics are essential to function or stated with Feynman's provocative words: *If we were to name the most powerful assumption of all, which leads one on and on in an attempt to understand life, it is that all things are made of atoms, and that everything that living things do can be understood in terms of the jiggings and wiggings of atoms* ⁵⁰⁹. More specifically, MD is essential for understanding the lowering of the activation free energy in terms of the enzyme's structure and flexibility ⁵¹⁰.

Nowadays, the upper still feasible time scale limit for atomistic simulations is the μ s time scale ^{511, 512}. In this work, MD trajectories are analyzed using the concept of essential dynamics ^{513, 514}. According to this concept, configurations can be separated into (i) an "essential" subspace of only a few degrees of freedom comprising most of the positional fluctuations and (ii) the remaining space. While the degrees of freedom belonging to (i) describe motions relevant for protein function, the ones of (ii) represent merely irrelevant local fluctuations ⁵¹³. Simulations of a few hundreds of ps are generally long enough to provide a stable and statistically reliable ⁵¹⁵⁻⁵¹⁷ distinction between the essential and near constraints subspaces ⁵¹⁷.

Here, potentially functionally relevant protein motions are extracted from MD trajectories by two different means: Principal Component Analysis (PCA) and Normal Mode Analysis (NMA). While PCA chooses the degrees of freedom that contribute most to the atomic displacements observed during MD, NMA aims at obtaining uncoupled, local degrees of freedom. PCA has a major advantage over NMA in that it also considers motions resulting from visits to multiple minima ⁵¹⁸. NMA generates a set of conformational ensembles, representing vibrations around local minima, in a much faster way than conventional MD, but using random displacements. It is an established technique for the study of proteins ⁵¹⁹⁻⁵²¹, particularly ligand-induced conformational changes ^{520, 522}.

In analogy to the closely related AdK, TKs are thought to undergo large-scale rearrangements of their LID domain, closing over the substrate dT and the cofactor ATP. Indeed, the open-closed transition in nucleoside kinases by which an Arg is brought into a catalytically competent position has been extensively discussed ⁵²³⁻⁵²⁵. The benchmark system for computationally assessing such transitions is the *E. coli* AdK, because both its apo (PDB ID: 4AKE) and its holo crystal structures (PDB ID: 1AKE) are known. In the presence of Mg^{2+} , AdK transfers the γ -phosphate of ATP to AMP, yielding two molecules of ADP. At least two distinct AdK conformations exist in solution ⁵²⁶, one binding ATP and AMP and one binding the two ADP molecules. AdK closure and reopening was determined to be the rate-limiting step ⁵²⁷. For both AdK and TKs, the LID closure over the core creates a solvent-occluding environment beneficial to phosphoryl transfer. In the case of AdK, these LID motions are in the order of ns ⁵²⁸, and core movements in the order of ps ^{529, 530}. Similar timescales are assumed for TKs. The major findings with respect to AdK dynamics are summarized in the following and could provide a basis for studies on TKs. The full transition pathway of AdK from closed to open has been identified ⁵³¹, with the overall structural integrity being maintained and AdK behaving like a rigid body with flexible domains ^{226, 532-534}. In spite of a substantial energy difference between the open and closed state ^{535, 536}, the apo enzyme is capable of switching between these states in the absence of bound nucleotides ^{534, 537, 538}. It appears that AMP and the LID control this switch ⁵³⁹. This is consistent with a study combining mutagenesis on mesophilic and thermophilic AdKs and differential scanning calorimetry showing that the controls of stability and activity are well separated, with the core domain governing overall stability, while the binding of AMP and the LID domain control the temperature dependence of the catalytic activity ⁵⁴⁰. It seems probable that also other enzymes than AdK ^{533, 541}

possess an innate ability for an open-closed interconversion^{534, 542}. Using NMR relaxation and a 10 ns MD, Henzler-Wildman et al.⁵⁴³ found that ps to ns atomic fluctuations in the hinge of AdK facilitate the large-scale slower LID motions that produce a catalytically competent state.⁵⁴³ Furthermore, a PCA analysis showed that the apo AdK encodes its holo form, meaning that it provides an ensemble of apo conformations to the substrates that select the ones relevant for binding⁵⁴⁴.

A still not fully closed debate is whether AdK undergoes (i) an induced fit, (ii) a lock-and-key movement, (iii) displays a “pre-existing equilibrium” (iv) or “encodes” the fluctuations necessary for ligand selection and binding.⁵⁴⁴ The first case has for a long time been favoured⁴⁸³, because the enzyme undergoes large conformational changes, which are consistent with the notion of protein plasticity. The second variant can be discarded, because it would only imply minor movements in the binding site and the ligand upon binding. The third possibility encompasses that a native protein adopts conformations ranging from apo to holo-like, with the ligand selecting the conformation that would shift the equilibrium to the bound state (conformational selection)⁵⁴⁵. In the “encoding hypothesis”, the ligand does not select the conformation, but the apo conformations contain the necessary information for ligand binding^{544, 546-559}. (iii) and eventually (iv) seem to be possible for proteins such as AdK or TKs that undergo large rearrangements between the apo and holo forms (e.g. occlusion of the substrate(s) from solvent).

1.1.5 STUDYING UNBINDING BY STEERED MOLECULAR DYNAMICS (SMD) AND RANDOM ACCELERATION MOLECULAR DYNAMICS (RAMD)

Rare events or slow processes are difficult to study by MD, because MD runs are usually limited to nanoseconds. Metadynamics, introduced by Laio and Parrinello^{560, 561}, is one solution to this problem, as the technique is very useful in identifying transition states on the energy surface, together with transition-associated mechanisms. Metadynamics is a modification of standard MD in that harmonic restraints are applied over time on collective variables of interest, e.g. the distance from the ligand to the binding site, and in that the system is prohibited to visit the same conformation twice, leading to a forced exploration of the free energy surface⁵⁶². The focus of this work however is on another *in silico* method to sample rare events. The unbinding with a large energy barrier can be assessed by steered molecular dynamics (SMD) and the related random acceleration molecular dynamics (RAMD). RAMD is used to define a biologically sound ligand exit pathway, which is then used as fixed ligand unbinding direction in SMD.

Steered molecular dynamics (SMD)

Contrarily to classical molecular dynamics (CMD), SMD forces protein-ligand systems out of equilibrium to evolve along a predefined ligand exit path. In this way, the unbinding process can be studied at timescales accessible to MD techniques. Typically, the ligand is attached to a virtual harmonic spring of a given stiffness and pulled out at a constant speed. Predefining the pulling-out pathway is advantageous given that force or work profiles can yield an estimation of the pathways' likelihood and a free energy profile can directly be compared to the one of other pathways and eventually experimental k_{on} or k_{off} values^{563, 564}. Izrailev et al.¹²¹ provide an in-depth introduction to SMD, which can be seen as the *in silico* counterpart to atomic force microscopy (AFM⁵⁶⁵). By comparing *in vitro* with *in silico* pulling experiments, Sotomayor et al. realized that *in vitro* AFM pulling experiments enable the characterization of the mechanical response of biological matter at the nm scale, whereas they do not reveal the molecular mechanics underlying mechanical function, which can only be unravelled by *in silico* pulling⁵⁶⁶. Put in other words, ligand binding/unbinding is “steered” by a series of protein-ligand interactions and the shape of the free energy landscape dictates binding/unbinding kinetics.

Since the first application of SMD to the unbinding of biotin from avidin and streptavidin^{12 567}, it has gradually become a tool for protein-ligand recognition and interaction⁵⁶⁶. Other examples where this technique is applied to ligand unbinding^{121, 568-574}, peptide elasticity¹⁴, unfolding^{575, 576} or refolding⁵⁷⁷, and conformational preferences of peptides⁵⁷⁸ have been published. Direct applications of SMD to drug design purposes include e.g. the discrimination of active from inactive inhibitors⁴²⁶ and the design of a more potent imatinib analogues binding to the inactive conformation of the tyrosine kinases c-KIT and Abl⁵⁷⁹. In the latter publication, it is shown that imatinib dissociation is more favourable via the ATP channel lined by more flexible residues than the wider allosteric binding pocket

and that a potent imatinib analogue should not display a bulky hydrophobic head, but eventually a large polar tail.

Random acceleration molecular dynamics (RAMD)

The pulling-out direction can be identified by random expulsion molecular dynamics, an extension of SMD sometimes also termed RAMD, invented by Lüdemann et al.⁵⁸¹, featuring a low computational cost and an objective pathway search. Contrarily to SMD, the force applied to the ligand is randomly oriented, kept for a given time step, unless the ligand bumps into the protein, and then changed. Examples of successful disclosures of ligand pathways can e.g. be found in⁵⁸¹⁻⁵⁸⁵. However, the high-velocity ligand expulsions tend to favour straight expulsion trajectories and the observed relative frequencies of different pathways can therefore be biased⁵⁸⁶.

Combining docking with MD

While docking is applied to fast ligand conformational sampling (mostly treating the protein as rigid), MD is used for exhaustive protein-ligand conformational sampling, which is time-consuming and carries the risk of getting trapped in local minima. In a review by Alonso et al.⁵⁶², the advantages of MD in a docking context are especially seen in pre-docking protein preparation, structure optimization, refinement of docked protein-ligand complexes and binding free energy calculations. Combining docking and MD using metadynamics for flexible docking in water, Gervasio et al.⁵⁸⁷ could show that the whole energy surface including all minima and barriers could be used to study molecular recognition in detail. Less calculation intensive coarse grained metadynamics involving only the degrees of freedom necessary for docking/undocking are however already able to distinguish crystallographic from non crystallographic protein-ligand conformations and to find the transition state in ligand-protein binding⁵⁸⁸.

1.1.6 MAIN QUESTIONS ADDRESSED IN THIS CHAPTER FOR HSV-1 TK AND HTK1

The main questions tackled in this chapter include:

- Are the minimized holo crystal structures stable over 5 ns CMDs?
- Can main motions be extracted from these CMDs using PCA? If so, which motions are most important for catalysis?
- Can the substrate extraction work obtained by SMD be correlated with experimentally determined affinity data and can potential binding affinities of (fluorinated) substrate analogues be extrapolated?

1.2. METHODS

1.2.1 GENERAL PROTEIN-LIGAND SETUP FOR DYNAMICS

Crystallized holo HSV-1 TK structures (Table 12) were downloaded from the PDB^{10, 11}. Monomer A was retained with its associated substrate and the water molecules involved in water-mediated interactions between the substrate and HSV-1 TK (three in the case of HSV-1 TK/dT (PDB ID: 1KIM), two for HSV-1 TK/DHBT (PDB ID: 1E2P) and five for HSV-1 TK/NMeDHBT (PDB ID: 3F0T)) were kept using Sybyl 7.3 (Tripos, Inc.). Non-resolved loops were modelled using the same program. For HSV-1 TK/dT, the loops involved Val70 to Asp77, Gly148 to Pro154 and Gly264 to Pro280. The homologies between the effective HSV-1 TK loop sequences and their best match in the loop database were comprised between 60 and 69%, with associated backbone root mean square deviations (RMSDs) in the range of 0.2-0.3 Å. The loops modelled for HSV-1 TK/DHBT concerned residues Leu69 to Arg75, Ser149 to Pro154 and Gly264 to Pro280, with homologies between 56 and 66% and RMSDs ranging from 0.05 to 0.5 Å. For HSV-1 TK/NMeDHBT, the loops encompassing Ala147 to Pro154 and Gly264 to Pro280 had homologies of 74.51 and 63%, respectively, with RMSD values of 0.3 Å in both cases. The first modelled loop in HSV-1 TK/dT and HSV-1 TK/DHBT is present in the crystal structure of HSV-1 TK/NMeDHBT but was not taken over for these two structures because it was resolved later. Nevertheless, this loop is reasonably similar in the three cases. Hydrogens were then added to the protein-ligand complex using the biopolymer menu of Sybyl 7.3. The complexes were saved in Amber PDB format and each ligand was extracted, checked for correct protonation as well as bond types, and then saved in mol2 format. The further ligand preparation was done in AMBER9.0³⁶, using the

Antechamber and the Parmchk modules. With Antechamber, standard ligand topology files were generated, while with Parmchk, non-standard parameters were treated and adjusted in case of an error message. Subsequently, tleap was started using the AMBER force field FF99 and the general AMBER force field (GAFF) parameters, the Antechamber, Parmchk and protein-ligand complex files loaded, and the complex solved in a 12 Å thick layer of TIP3P water, forming a rectangular box. The excess charge of the system was then neutralized by adding counterions. The saved solvated protein-ligand complex was then reloaded, and the resulting AMBER coordinate and topology files were stored.

The hTK1 ligands dTMP, dTDP and dTTP were parameterised separately. Published phosphotyrosine parameters⁵⁸⁹ were applied to the monophosphate part of dTMP. For the phosphate parts of dTDP and dTTP, ADP and ATP parameters available at <http://www.pharmacy.manchester.ac.uk/bryce/amber#cof> were used. The Zn²⁺ ion held in place by four Cys and additional water molecules in hTK1 had to be parameterised carefully. The Zn²⁺ library file was generated from the Zn²⁺ related entries in ions94.lib and the residue name was changed to ZNA. In the complex file, ZN was also changed to ZNA. Knowing that Zn²⁺ is coordinated by four Cys in hTK1, a frcmod (Parmchk) file was created by taking the bond, angle, dihedral angle, and non-bonded parameters from the thesis of Ross Walker⁵⁹⁰ and the Amber mailing list (<http://archive.ambermd.org/>). tleap was launched as before, and the zinc library as well as parameter files red in together with the ligand antechamber and parmchk files. The remaining setup procedure followed the already described steps.

Each protein-ligand complex underwent a 50 ps NPT (stable number (N) of particles, pressure (P) and temperature (T)) minimization-equilibration with a pressure at 1 bar, a temperature of 300 K, a timestep of 2 fs, a Langevin thermostat at 5 ps⁻¹, periodic boundary conditions, and Particle-Mesh Ewald electrostatics (PMEGridSpacing parameter set to always less than 1 Å) in NAMD2.6⁵⁹¹. The cutoff for VdW and short-range electrostatic interactions was set to 12 Å. The coordinates were saved every 0.5 ps and the pressure monitored every 0.2 ps. The total energy of the system was stable at the end of each minimization-equilibration.

1.2.2 CLASSICAL MOLECULAR DYNAMICS, PRINCIPAL COMPONENT ANALYSIS, AND NORMAL MODE ANALYSIS

To assess the general stability of three holo HSV-1 TK and one holo hTK1 minimized structures in a water bath, 5 ns long CMDs were performed. Subsequent PCA and NMA analyses served to extract main motions from these CMDs.

In total, ten 5 ns long NPT CMDs were done in NAMD2.6 (constant N, P (1 bar) and T (300K)): HSV-1 TK (dT), (DHBT) and (NMeDHBT) were repeated three times each, starting from the same minimization file for each repetition, and hTK1 (dTTP) was run once. Then, these 10 CMDs were prepared for PCA. Initially, the frames from each production trajectory were aligned to the corresponding crystal structure to remove non-essential motions. Next, the average structure of the trajectory served as a template for a second alignment round. Both alignments were carried out in Wordom⁵⁹², yielding a trajectory with minimized internal protein dynamics. To reduce the data processing and computation time, each 100th frame (corresponding to every 50th ps) was extracted from this trajectory as a pdb file. Considering only the atomic positions of the C α atoms, both the average pdb and the reduced number of frames were then used to construct a covariance matrix. The diagonal of this matrix represents the variance of the movement of the different C α atoms as compared to the average structure. The rest of the matrix shows the covariance of the movements of pairs of C α atoms. The PCA was calculated separately on each 100th frame (0 ns to 5 ns, 50 frames in total) for the x, y and z coordinates of all C α atoms with customized scripts using the open source statistical program R 2.0.9 (<http://www.r-project.org/>). It was chosen to split the obtained covariance matrix according to the x-, y- and z-coordinates, to analyse the movement behaviour along these axes separately with respect to the average structure. The dimensions of each matrix were checked to match the corresponding number of C α atoms. As an example, in the case of HSV-1 TK (dT), each matrix had 330 columns and 330 rows, i.e. 330 dimensions. From the 10 CMDs, totally 30 matrices were generated, and each matrix projected on a plane which best represented the covariance in two dimensions. NMA was performed on the C α atoms of each monomer A of the three HSV-1 TK crystal structures, using the online NMA tool of the Norwegian Bioinformatics Platform at <http://services.cbu.uib.no/tools/normalmodes>.

1.2.3 OBTAINING AN OPEN, STILL SUBSTRATE BOUND CONFORMATION BY SMD

Because all holo crystal structures used are in a closed conformation and initial standard SMD trials on extracting substrates from HSV-1 TK led to a strong distortion of α -helices and the LID domain, an open, still substrate bound conformation had to be obtained by SMD. SMD has already been used to bias trajectories in such a way as to obtain intermediate conformations between apo and holo crystal structures (reviewed in ⁵⁹³). In the absence of a partially open holo or open apo crystal structure of HSV-1 TK, SMD was applied to obtain open, still substrate bound conformations of each protein-ligand complex. For the SMDs in NAMD2.6, the external force along a predefined pulling direction was defined by using $C(t) = C(t_0) + vt$, where C corresponded to the 3D coordinates, t to the time, t_0 to the starting time and v to the velocity (kept constant here). The force constant was set to $k = 7$ kcal/mol $\cdot\text{\AA}^2$. During a 200 ps SMD, the LID domain (encompassing residues Lys219 to Arg226) was gradually opened while keeping the C α of Gln125 (involved in substrate H-bonding) fixed and defining the C α of Gln221 as movable atom. All the protein atoms except the ones belonging to the LID and the preceding and following residues up to different flanking glycines were restrained. Three combinations of non-restrained residues were tried: Gly205 to Gly226, Gly206 to Gly254, and Gly216 to Gly227. To see whether the substrate dT had an influence on LID opening, the three restraining protocols were applied to both i) HSV-1 TK with dT and ii) HSV-1 TK from which dT had been deleted, reflecting a “pseudo” apo structure. The stability of the open conformation was then assessed with a standard 1 ns NPT MD. For hTK1, the lasso loop covering dTTP was modelled in Sybyl7.3, so as to match the conformation of the more open loop seen in the crystal structure of the *Thermatoga maritima* TK (PDB ID: 2ORW), which provides more space for the adenine of ATP ⁵⁰⁰.

1.2.4 RAMD

50 ps of RAMD were also performed under NPT conditions in NAMD2.6. The external force used to move the substrate was defined by $F_{\text{external}} = f_0 \vec{r}$, where f_0 is the constant magnitude of the randomly chosen force and $\vec{r}$ a randomly oriented unit vector. During the simulation, a minimum velocity v_{min} is maintained during a number of steps m over a minimum distance r_{min} with a timestep Δt : $v_{\text{min}} = r_{\text{min}}/m\Delta t$. Different combinations of f_0 , m , and r_{min} were taken over from Vashisth et al. ⁵⁸⁵ and the optimal conditions under which the substrate exit rate was maximal were searched.

1.2.5 DOCKING

To obtain acceptable starting binding poses for the substrate analogues which are not crystallized in complex with HSV-1 TK or hTK1, docking was performed using GOLD 4.0 ⁹¹. This program has three main advantages: It explores the full range of ligand conformational flexibility, accounts for some protein flexibility via side chain rotameric libraries, and allows for including crystallographic water molecules into the active site, which could potentially be involved in ligand binding. Additionally, its docking accuracy and its discrimination of known TK ligands are superior to other commonly used docking programs ⁸³.

GOLD was run including water molecules involved in water-mediated interactions ⁵⁹⁴, taking as input crystal structure either 1KIM (HSV-1 TK crystallized with dT), 1E2P (HSV-1 TK crystallized with DHBT) or 3F0T (HSV-1 TK crystallized with NMeDHBT), depending on the similarity of the compound to be docked with the cognate ligand of each crystal structure. Each ligand was prepared in Sybyl7.3 by taking either dT, DHBT or NMeDHBT as template. The correct protonation, atom and bond types as well as Gasteiger-Huckel charges were assigned, a standard 100 step Powell minimization done, and each ligand saved as mol2 file. In the GOLD configuration file, the active site was defined with a radius of 6.5 $\text{\AA}$ around the centre of mass (previously calculated in Sybyl7.3) of the experimentally determined ligand. Flipping of planar nitrogens and rotation of carboxylic OH was enabled and the retained water molecules could spin so as to adapt their H-bond mediating capacities to the docking poses. 10 poses were generated per docking run and early termination was allowed if the heavy atom RMSD of the three first poses was within 1.5 $\text{\AA}$ RMSD. For anything else, default parameters were taken. As a control, dT, DHBT and NMeDHBT were redocked into their cognate crystal structure, yielding RMSDs of 0.5, 0.7, and 0.8 $\text{\AA}$, respectively. Each best-scored docking pose was retained, the GOLD lone pairs removed, protonation, atom and bond types checked, saved in mol2 format in Sybyl7.3, and subsequently used for generating the protein-ligand complex file.

1.2.6 UNBINDING SMDs

Depending on the crystal structure used for docking, the corresponding CMD stabilized holo structure was loaded into Sybyl, all water molecules except the H-bond mediating ones removed and the cognate ligand replaced by the above prepared mol2 ligand. Both the ligand and the complex preparation in AMBER were then done as described for the general procedure.

The constant direction for the unbinding SMDs for all the substrate analogues was deduced from the biologically most sound exit direction found by RAMD. The protein-ligand setup for all complexes was carried out as described in 1.2.1. The minimization-equilibration protocol was the same as for the CMDs, except that protein's centre of mass motion was not allowed anymore. For each protein-ligand system, a SMD configuration file together with a tcl script were adapted from H. Vashisth, who kindly provided them. k was set to 7 kcal/molÅ² and v to 1.5 Å/ps. By means of the tcl script, the force values obtained for each timestep were converted into accumulated work values, which could potentially directly be used to calculate a PMF via Jarzynski's equality⁵⁶³ or related formula⁵⁹⁵. For getting an idea on the reproducibility of the time-dependent work values for each protein-ligand system, 20 repetitions were carried out for each and the median values calculated. Pyrimidine-based HSV-1 TK substrates with known K_i values obtained earlier from competition experiments (Table 12) and their work evolution over time served to establish a correlation. This correlation was then used to extrapolate binding affinities from SMD-based work-time profiles of substrate analogues with unknown K_i values (Table 13).

1.2.7 COMPARED DATASETS

The known HSV-1 TK substrates depicted in Table 12 have all been crystallized previously, while the compounds with unknown K_i values (Table 13) were docked into HSV-1 TK according to the procedure detailed in 1.2.5.

1.3 RESULTS AND DISCUSSION

1.3.1 CLASSICAL MOLECULAR DYNAMICS (CMD)

Stability assessment

As expected for NPT CMDs, the pressure (P) and the temperature (T) do not significantly deviate from the input values 1 bar and 300 K. The total energy of the system decreases at the beginning of each CMD and stabilizes after 1 ns. The stability of the proteins during the CMDs was analyzed in terms of RMSDs from each corresponding crystal structure and compared to the B-factors. Considering the C α atom RMSD over time, all dynamics for both HSV-1 TK and hTK1 reached a plateau (corresponding to the mean RMSD of 2.2 Å for HSV-1 TK and 1.4 Å for hTK1) after approximately 1 ns and fluctuated in the crystal structure resolution range until the end of the simulations. Figure 13a depicts the RMSD time course for both a HSV-1 TK (dT) and an hTK1 (dTTP) simulation and is shown as a representative example of the other CMDs.

The C α atom RMSDs were then computed per amino acid and compared to the B-factors, which indicate the flexibility of the residues within the crystal. As an example, for an atom with an average B-factor of 20 Å², the mean displacement from the equilibrium position amounts to 0.5 Å¹.

Figure 13b displays a comparison of the B-factors and RMSDs per residue for HSV-1 TK. Overall, the RMSD can be taken as a measure for assessing the intrinsic movement of single amino acids, because the RMSD show the same tendency as the B-factors (including the ones of some modelled loops which could not be resolved experimentally because of their high flexibility). Additionally, the MD-derived structures represent the dynamic behavior around the energy minimum: Structural deviations from the starting structure are highest for the N-terminal residues, the LID domain (residues 219-226) and a flexible segment around residue 270, but very low (less than 2 Å) for the core, which is consistent with the crystallographic B-factors, even when compared to structures with different ligands bound. This stability is a prerequisite for the rest of this study.

Table 12: HSV-1 TK substrates with known experimental K_i values, obtained from substrate competition experiments.

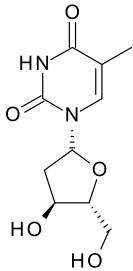
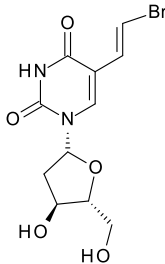
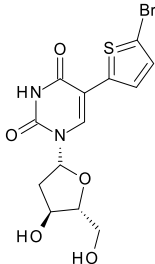
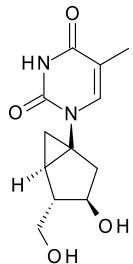
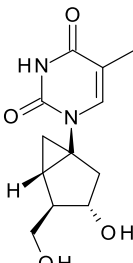
Substrate name	Substrate abbreviation	Structure	K_i or K_m	Crystal structure	Ref.
(2-deoxy)thymidine	dT		$0.2 \pm 0.05 \mu\text{M}$	1KIM	486, 596
5-bromovinyldeoxyuridine	BVD		$0.3 \mu\text{M}$ (IC_{50})	1KI8	486, 597
5-bromothieryl-deoxyuridine	BTD		$3.5 \mu\text{M}$ (IC_{50})	1KI4	
(N)-MCT, North-methanocarbathymidine, 3'-endo-methanocarbathymidine	NCT		$11.4 \pm 1.6 \mu\text{M}$	1E2K	596
(S)-MCT, South-methanocarbathymidine, 3'-exo-methanocarbathymidine	SCT		$4.1 \pm 2.4 \mu\text{M}$	1OF1	477, 481, 482

Table 12 (continued): HSV-1 TK substrates with known experimental K_i values, obtained from substrate competition experiments.

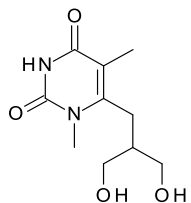
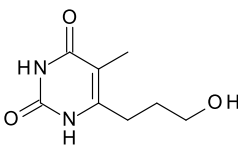
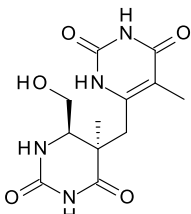
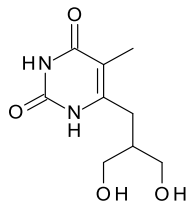
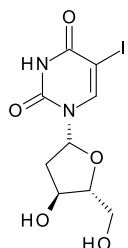
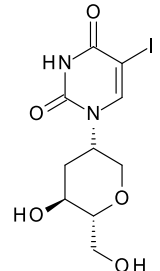
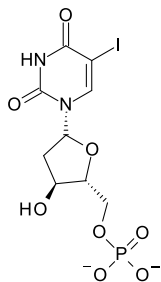
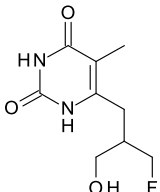
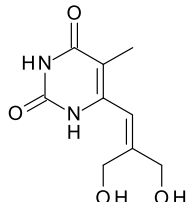
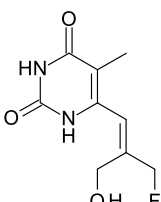
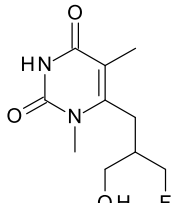
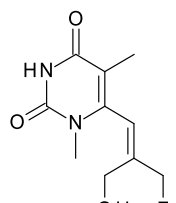
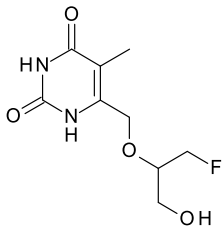
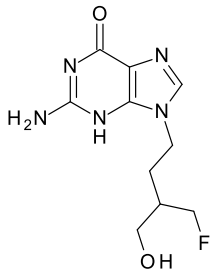
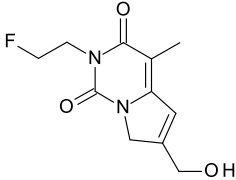
Substrate name	Substrate abbreviation	Structure	K_i or K_m	Crystal structure	Ref.
N-methyl-6-(1,3-dihydroxy-isobutyl)thymidine	NMeDHT		$10 \pm 0.3 \mu\text{M}$	3F0T	497
6-hydroxypropylthymidine	HPT		$26.6 \pm 0.4 \mu\text{M}$	1E2M	598
6-[[4-(hydroxymethyl)-5-methyl-2,6-dioxohexahydropyrimidin-5-yl]methyl]-5-methylpyrimidine-2,4(1H,3H)-dione	RCA		$30.9 \pm 1.8 \mu\text{M}$	1E2N	598
6-(1,3-dihydroxy-isobutyl)thymidine	DHBT		$35.3 \pm 1.3 \mu\text{M}$	1E2P	473, 598

Table 13: Potential and known HSV-1 TK substrates with K_i values to be predicted (continues on next two pages).

Substrate name	Substrate abbreviation	Structure	Crystal structure	Ref.
5-iododeoxyuridine	IDU		1KI7	486
1',5'-anhydro-2',3'-dideoxy-2'-(5-iodouracil-1-YLD-arabino-hexitol	AHU		1KI6	486

5-iodo-deoxyuridine-monophosphate	5IU		3VTK	485
6-[2-(fluoromethyl)-3-hydroxy-propyl]-5-methyl-1H-pyrimidine-2,4-dione	FHBT		None	None
6-[3-hydroxy-2-(hydroxymethyl)prop-1-enyl]-5-methyl-1H-pyrimidine-2,4-dione	SK-7,8		None	None
6-[(E)-2-(fluoromethyl)-3-hydroxy-prop-1-enyl]-5-methyl-1H-pyrimidine-2,4-dione	SMA2		None	None
6-[2-(fluoromethyl)-3-hydroxy-propyl]-1,5-dimethyl-pyrimidine-2,4-dione	NMeFHBT		None	None
6-[(E)-2-(fluoromethyl)-3-hydroxy-prop-1-enyl]-1,5-dimethyl-pyrimidine-2,4-dione	SMA1		None	None

6-((1-fluoro-3-hydroxypropan-2-yloxy)methyl)-5-methylpyrimidine-2,4(1H,3H)-dione	FHOMP		None	None
2-amino-9-[3-(fluoranylmethyl)-4-hydroxybutyl]-3H-purin-6-one	FHBG		None	None
2-(2-fluoroethyl)-6-(hydroxymethyl)-4-methyl-7H-pyrrolo[1,2-c]pyrimidine-1,3-dione	SMA3		None	None

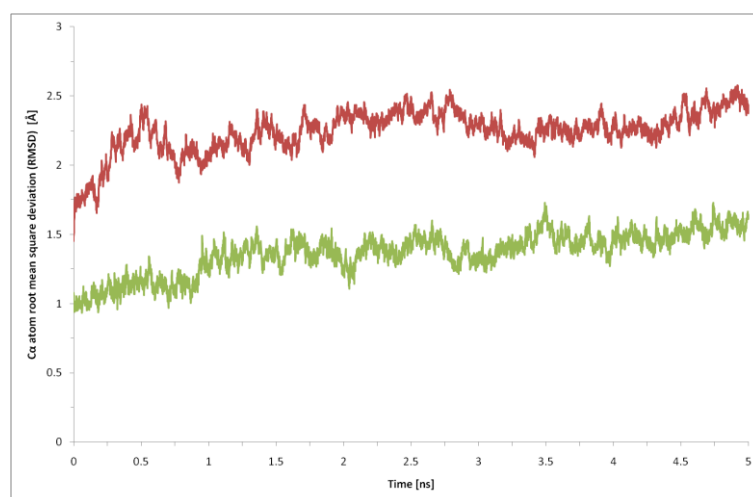
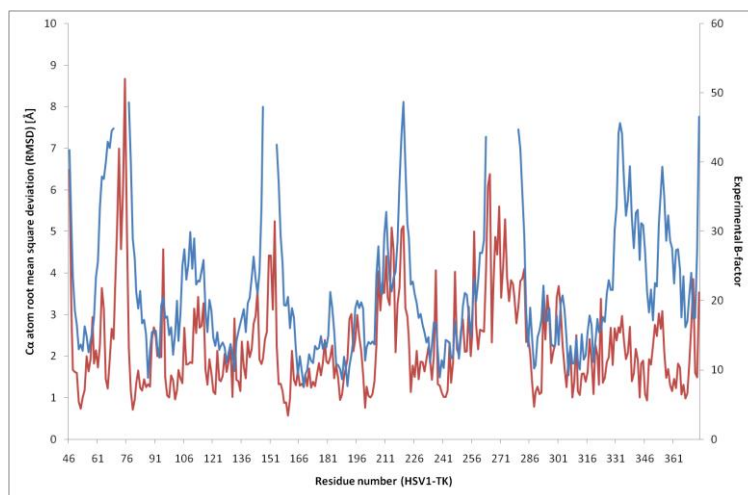
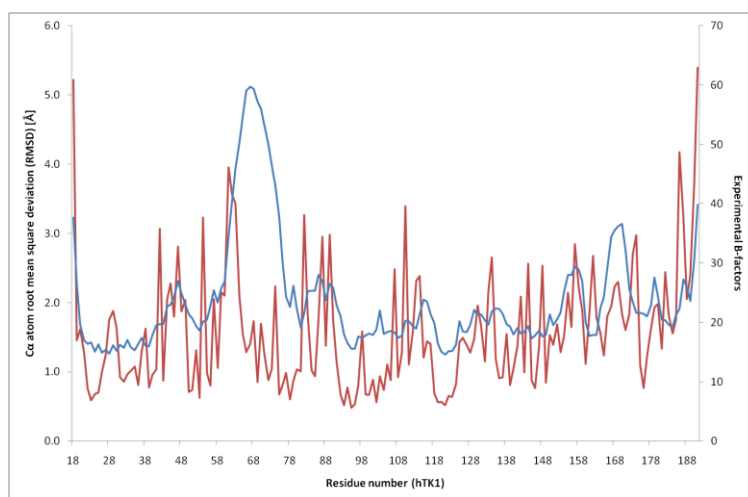


Figure 13: a) Stability assessment for HSV-1 TK (dT) in red and hTK1 (dTTP) in green. The Ca atom root mean square deviation (RMSD) [Å] from each crystal structure is plotted versus the time [ns]. The mean RMSD is 2.2 Å for HSV-1 TK and 1.4 Å for hTK1.



b) Movement assessment for HSV-1 TK.



c) Movement assessment for hTK1.

Figure 13 (continued): Movement assessment for HSV-1 TK (b) and hTK1 (c) in terms of RMSD (red) and B-factors (blue) per residue.

Figure 13c illustrates the same type of analysis for hTK1. Here, the RMSDs and corresponding B-factors do not always share the same tendency. Nevertheless, the main minima and maxima behave similarly.

Extraction of main movements from CMDs: Principal Component Analysis

To spot the main movements of the C α atoms during the 5 ns CMDs and to remove unessential translations (“background noise”), different PCAs were performed on the x-, y and z-coordinates, respectively. After having superposed the trajectory of the CMD onto the average structure, a covariance matrix was computed. In a next step, this matrix was projected along the two dimensions reflecting most of the movement. Table 14 a and b shows the percentages of the total movement accounted for by the sum of principal component 1 (PC1) and 2 (PC2). Remarkably, all the percentages are over 60% and very often approaching 90%. This means that a large proportion of the movement is represented by the projections.

Table 14 a: The percentage of the total movement accounted for by the most important principal components (PC1 and PC2) for each CMD type. For HSV-1 TK with each substrate (dT, DHBT, and NMeDHBT, respectively), the CMD was run three times starting from the same minimization-equilibration output conformation for each protein-ligand system.

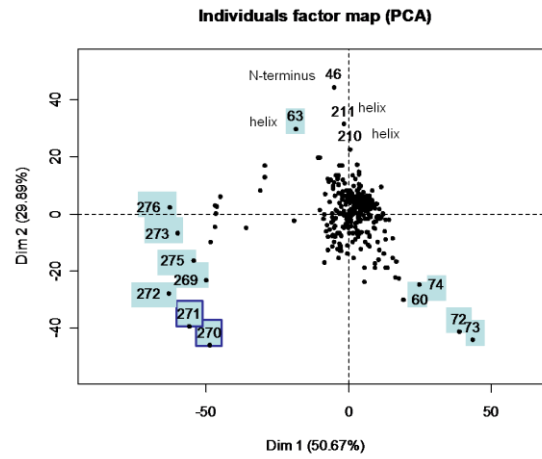
CMD type	HSV-1 TK with dT			HSV-1 TK with DHBT			HSV-1 TK with NMeDHBT		
x-axis (run)	x (1)	x (2)	x (3)	x (1)	x (2)	x (3)	x (1)	x (2)	x (3)
PC1 + PC2	80.56%	80.91%	75.05%	87.53%	90.44%	86.34%	79.55%	66.78%	62.21%
y-axis (run)	y (1)	y (2)	y (3)	y (1)	y (2)	y (3)	y (1)	y (2)	y (3)
PC1 + PC2	69.63%	85.56%	73.31%	84.89%	89.79%	63.74%	78.53%	86.92%	77.70%
z-axis (run)	z (1)	z (2)	z (3)	z (1)	z (2)	z (3)	z (1)	z (2)	z (3)
PC1 + PC2	78.38%	84.94%	70.17%	88.04%	74.83%	83.31%	81.35%	75.46%	79.95%

Table 14 b: The percentage of the total movement accounted for by the most important principal components (PC1 and PC2) for the CMD of hTK1 with dTTP.

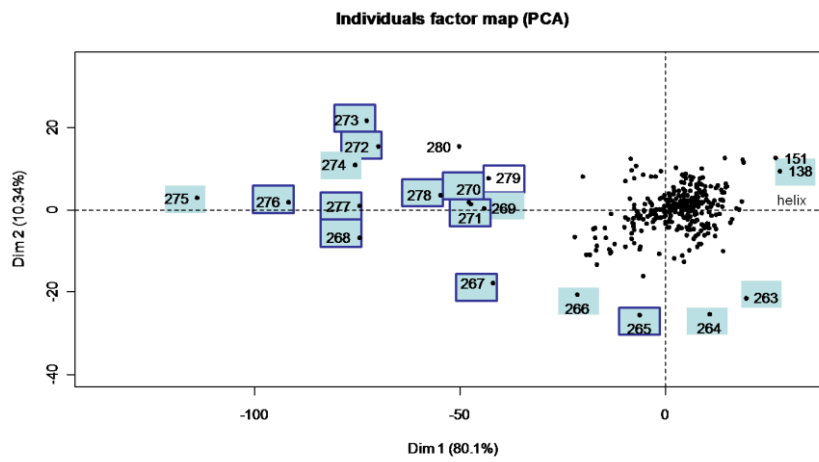
CMD type	hTK1 with dTTP
x-axis	X
PC1 + PC2	72.07%
y-axis	Y
PC1 + PC2	77.75%
z-axis	Z
PC1 + PC2	89.59%

For explaining the projections presented in the following, the points nestling to the origin are not of interest, as they do not represent motions deviating significantly from the average. For outliers either along PC1, PC2 or both, the corresponding $\cos^2$ values have been computed in the program R 2.0.9 as a measure of the quality of the projection. A $\cos^2$ value is calculated by the following formula: $\cos^2 = \frac{\|g-a\|^2}{\|g-a'\|^2}$, where g is the barycentre of the points in 2D, a is any point in 3D and a' is the projection of point a in 2D. The pipes denote a Euclidian norm. A perfect projection would have a $\cos^2$ value of 1, with the motion being entirely explained by PC1 and 2. A threshold of (at least) 0.9 was set to retain only very good projections for the analysis (highlighted in blue in the PCA graphs).

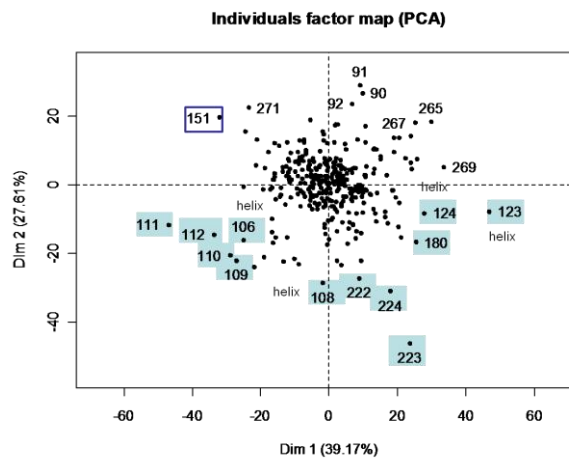
Only one representative for each protein-ligand system (HSV-1 TK/dT and hTK1/dTTP) and direction (x, y, and z) is shown. The remaining graphs can be found in the annex. In the PCA graphs, PC1 corresponds to Dim1 (projection dimension 1) and PC2 to Dim2 (projection dimension 2). The label of each point corresponds to the residue number of the considered TK. Points without a label are either situated in a very crowded area of the scatter plot or do neither belong to an α -helix, nor a β -sheet, nor an amino acid segment with a particular function. Of interest in the analysis of the PCA projections are the outliers, which can be clearly distinguished from the bulk of the points and are located opposite to other outliers with respect to one or both projection axes.



a) HSV-1 TK/dT, x-direction (run 1)



b) HSV-1 TK/DHBT, x-direction (run 2)



c) HSV-1 TK/NMeDHBT, x-direction (run 2)

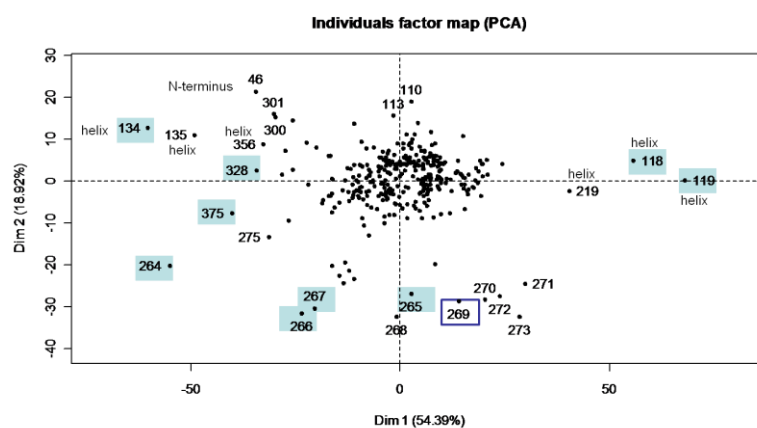
(Figure legend on the next page)

Figure 14a, b and c: Projections according to the first two principal components (PC1 = Dim1 and PC2 = Dim2) for representative runs of the 5 ns CMD with respect to the x-coordinates of the C α atoms of HSV-1 TK/dT (a), HSV-1 TK/DHBT (b) and HSV-1 TK/NMeDHBT (c), respectively. Each point with a distinct projection is labelled according to the corresponding C α residue number in HSV-1 TK and annotated if it belongs to a terminus, a secondary structure element or if it is known to have a specific role in the protein. Points contained in cyan boxes reflect an excellent projection quality ($\cos^2 \leq 0.9$) and the blue frame marks points which display a distinct projection in the repeated MDs. This and the following Figures were generated with the statistical program R.

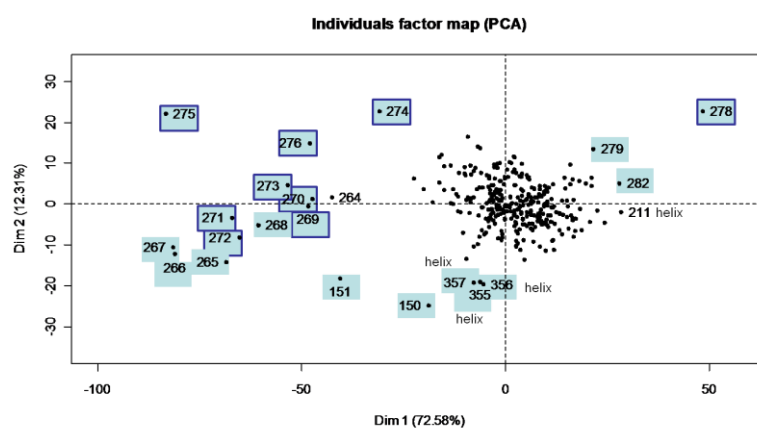
Considering the movement according to the x coordinates of the C α atoms for the three substrates, many of the outlier points correspond either i) to the LID domain (residues 219-226), ii) to helical structures, iii) the N-terminus ($\cos^2$ value of 0.89) or iii) to parts not resolved in the crystal structures and modelled as described in the methods. The modelled parts are termed segment 1 (residues 70-77 in HSV-1 TK/dT and 69-75 in HSV-1 TK/DHBT), 2 (residues 148-154 in HSV-1 TK/dT, 149-154 in HSV-1 TK/DHBT, and 147-154 in HSV-1 TK/NMeDHBT) and 3 (residues 264-280 in all three crystal structures) in the following. Particularly segment 3 occurs in all the PCAs. It belongs to a partially flexible part (residues 250-322) situated after the LID domain in sequence. Its frequent distinct behaviour seen in the PCA graphs confirms the indirect observation of extremely high or absent B-factors in the crystal structures, but also the direct observation of elevated C α atom RMSDs in this region. Concerning the three LID domain residues (residues 222-224) appearing in the HSV-1 TK/NMeDHBT PCA, they are characterized by an opposed x-axis movement with respect to segment 2, but a relatively similar movement compared to residues 123 and 124. It seems therefore likely that the three amino acids preceding the H-bonding residue Glu225 move in a concerted way with amino acids 123 and 124, which are located just before the H-bonding residue Gln125. Because this behaviour with respect to the x-axis is only seen in two of the three HSV-1 TK/NMeDHBT PCAs, but neither in HSV-1 TK/DHBT nor HSV-1 TK/dT, it can be hypothesized that the water-mediated interaction between the dihydroxy-isobutyl moiety and the side chains of Glu225 and Tyr101 (as seen in the crystal structure) creates a network of water-mediated stabilizing H-bond interactions extending towards Gln125.

Concerning the movements along the y-axis for the three HSV-1 TK CMDs, a LID residue (labelled as 219) appears in HSV-1 TK/dT only, where its movement is opposed to the N-terminus on PC1 and to segment 3 on both PC1 and PC2. In the other two graphs, residue 151 (which belongs to segment 2) is found close to some segment 3 residues, meaning that their y-movement is associated. Generally, outliers with an α -helical character seem to “prefer” only certain spots. However, a direct quantitative comparison is not possible due to the different projection axes. Pairwise (i) or higher order (ii) comparisons quantifying the strength of a relationship or opposition would nevertheless have been feasible, either via Kendall’s rank correlation coefficient (for i) or via a Kruskal-Wallis test (for ii), but go beyond the scope of this work.

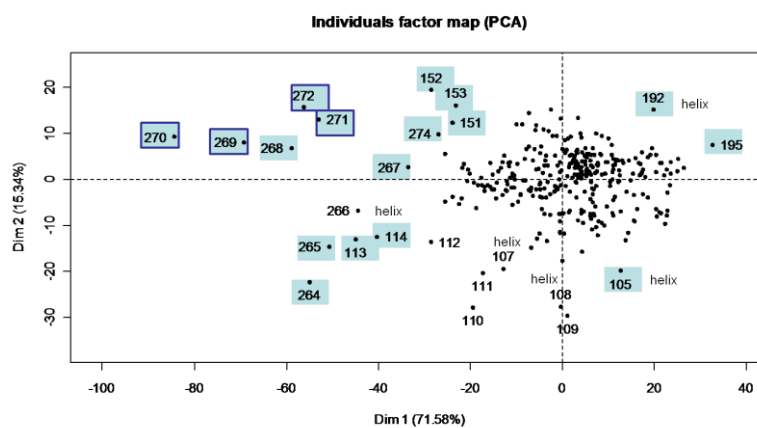
Looking at the z-movement behaviour of HSV-1 TK (dT), the segment 3 outliers with the amino acid numbers 270 to 273, which are conserved among all three HSV-1 TK/dT CMDs in z-direction, move as one unit. The C α atoms preceding and following them are spread around them and have mostly $\cos^2$ values higher than 0.9, indicating an excellent projection of the points. The behaviour of residue 220 of the LID domain seems associated to the segment 3 movement with respect to PC1. A similar association is observed for residue 219 with residues 270 to 272 (part of segment 3), according to both PC1 and PC2. In the graphs corresponding to HSV-1 TK/DHBT, segment 3 is more widespread, but can be clearly distinguished from segment 2 with respect to PC2. Again, C α “outliers” involved in helical structures tend to cluster.



a) HSV-1 TK/dT, y-direction (run 3)

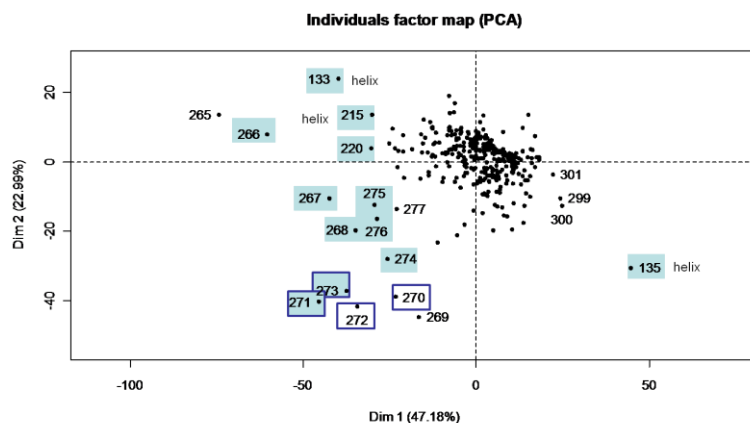


b) HSV-1 TK/DHBT, y-direction (run 2)

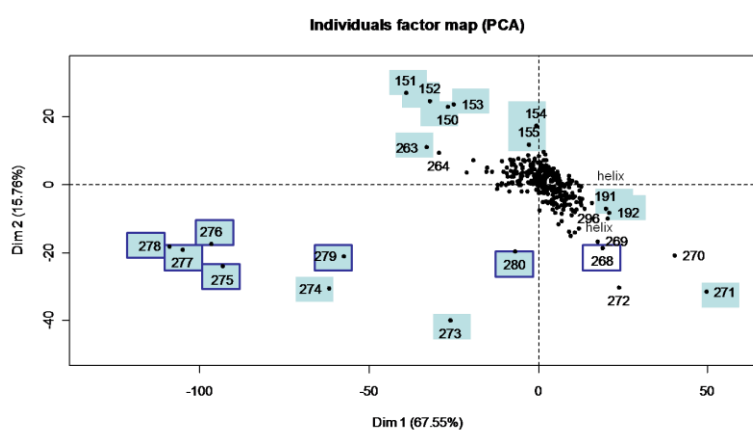


c) HSV-1 TK/NMeDHBT, y-direction (run 2)

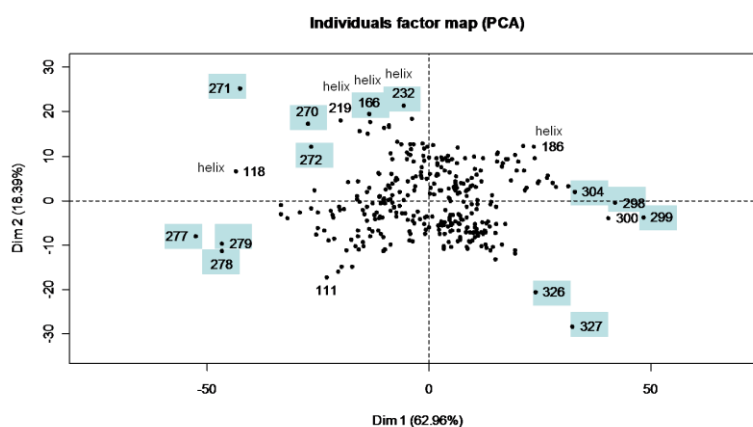
Figure 15a, b and c: Projections according to the first two principal components (PC1 = Dim1 and PC2 = Dim2) for representative runs of the 5 ns CMD with respect to the y-coordinates of the C α atoms of HSV-1 TK/dT (a), HSV-1 TK/DHBT (b), and HSV-1 TK/NMeDHBT (c), respectively. The same colouring and labelling criteria as for Figure 14a, b, and c apply.



a) HSV-1 TK/dT, z-direction (run 3)



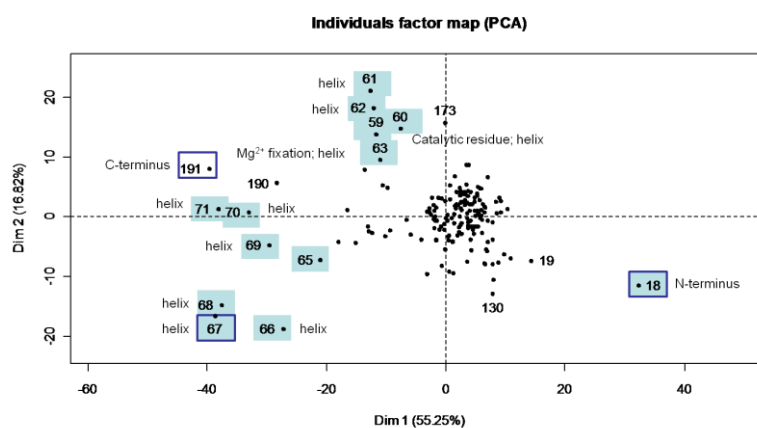
b) HSV-1 TK/DHBT, z-direction (run 3)



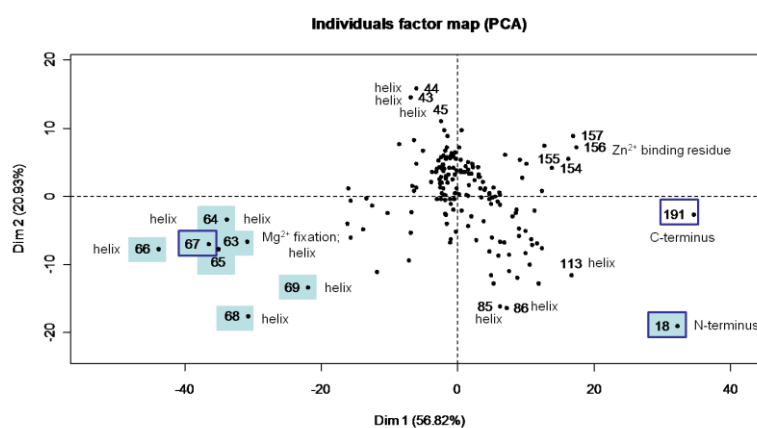
c) HSV-1 TK/NMeDHBT, z-direction (run 1)

Figure 16 a, b and c: Projections according to the first two principal components (PC1 = Dim1 and PC2 = Dim2) for representative runs of the 5 ns CMD with respect to the z-coordinates of the Ca atoms of HSV-1 TK/dT (a), HSV-1 TK/DHBT (b), and HSV-1 TK/NMeDHBT (c), respectively. The same colouring and labelling criteria as for Figure 14a, b, and c apply.

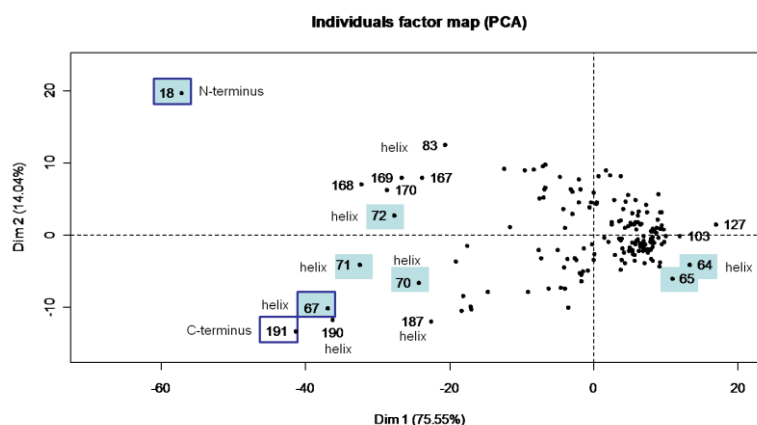
The PCAs obtained from the CMD on hTK1 in complex with dTTP are considered next.



a) hTK1/dTTP, x-direction



b) hTK1/dTTP, y-direction



c) hTK1/dTTP, z-direction

Figure 12a, b and c: Projections according to the first 2 principal components (PC1 = Dim1 and PC2 = Dim2) of the 5 ns CMD regarding the x-, y- and, z-coordinates of the $C\alpha$ atoms of hTK1/dTTP, respectively. The same colouring and labelling criteria as for Figure 14a, b, and c apply.

During CMD, the Zn^{2+} coordination in hTK1 is kept intact, which is important for the structural integrity of the enzyme. With regard to the movements of hTK1 (dTTP) in the x-, y- and z-direction, it can be stated that the N- and C-termini can be distinguished most clearly for the x- and z-movements and that outliers being part of a helical secondary structure are grouped.

Of importance for hTK1 function is that the movements in the x-direction of the catalytic Arg60 and the Mg^{2+} -fixing residue Ser63 are distinct from the bulk of the core residues in the centre of the graphs, but located relatively close to each other, implying that their motion is coupled.

The lasso domain of hTK1 covers residues 153 to 191. With respect to movements relative to the x-axis, residue 173 moves more in a more similar way to the C-terminus than to the N-terminus, which is farther away in the crystal structure. This is also true for residues 154 to 157 considering the y-movements, especially the Zn^{2+} -binding residue Cys156. Moreover, the movements of these residues contrast with the ones of the Mg^{2+} -coordinating residue on both PC1 and PC2. The movements of the lasso residues 167 to 170 along the z-axis are rather distinct from the ones of the N- and C-termini, in contrast to what has been observed for lasso residues in the x- and y-directions.

Summarizing the PCA results for both HSV-1 TK and hTK1, the $\text{C}\alpha$ backbone atoms of the core residues involved in H-bonds with the substrate do not markedly diverge from the average structure of each CMD, regardless of the substrate bound. "Outliers" in the projection are furthermore never belonging to the central β -sheet, but rather to flexible parts (mainly the LID domain and segment 3 in the case of HSV-1 TK) or the surrounding α -helices. This contrasts with the general knowledge that the flexibility of residues in α -helices is mostly smaller than in β -sheets or loops. Thus, the particular stability of the β -sheet core seems to be an intrinsic TK feature, whose core movements are in analogy to AdK^{529, 530} assumed to happen in the order of ps. To be able to state whether the core stability is partially a result of a bound ligand, apo HSV-1 TK and hTK1 CMDs and PCAs should be done.

Most importantly, the LID domain in HSV-1 TK moves in opposite directions to segment 3 with respect to the x- and y-axes but has a similar behaviour in the z-direction, meaning that they move as a unit along this axis. Therefore, it can be assumed that the LID movement has a hinge-like character upon substrate entry/exit. To further corroborate and refine this crucial statement, a PCA on an apo HSV-1 TK CMD should be done and could eventually be completed with a PCA on a holo HSV-1 TK CMD containing both the natural substrate dT and the cofactor ATP or the TK reaction products dTMP and ADP.

Clearly, obtaining tendencies like the mentioned ones by PCAs from 30 independent CMD runs as short as 5 ns is already an achievement, since global protein motions tend to be masked by random diffusion for short simulations, an issue which is addressed in detail e.g. by Hess^{599, 600}.

NORMAL MODE ANALYSIS (NMA) AS A COMPLEMENT TO PCA

Regarding the online NMA analysis, only the six lowest frequency normal modes out of 246 calculated ones were considered for further analysis. Of these, mode 9 was best comparable for HSV-1 TK (dT), (DHBT), and (NMeDHBT), and showed distinct normalized squared atomic displacements.

As can be seen from Figure 17, in the LID domain encompassing residues 216 to 226, this displacement decreases in the order HSV-1 TK (DHBT) > HSV-1 TK (NMeDHBT) > HSV-1 TK (dT), reflecting exactly the order of increasing affinity of the three different substrates bound to HSV-1 TK (Table 12). Although it is very speculative, the low propensity of the LID belonging to HSV-1 TK (dT) to undergo low frequency vibrations could be a reason for the high affinity of dT towards HSV-1 TK compared to other ligands.

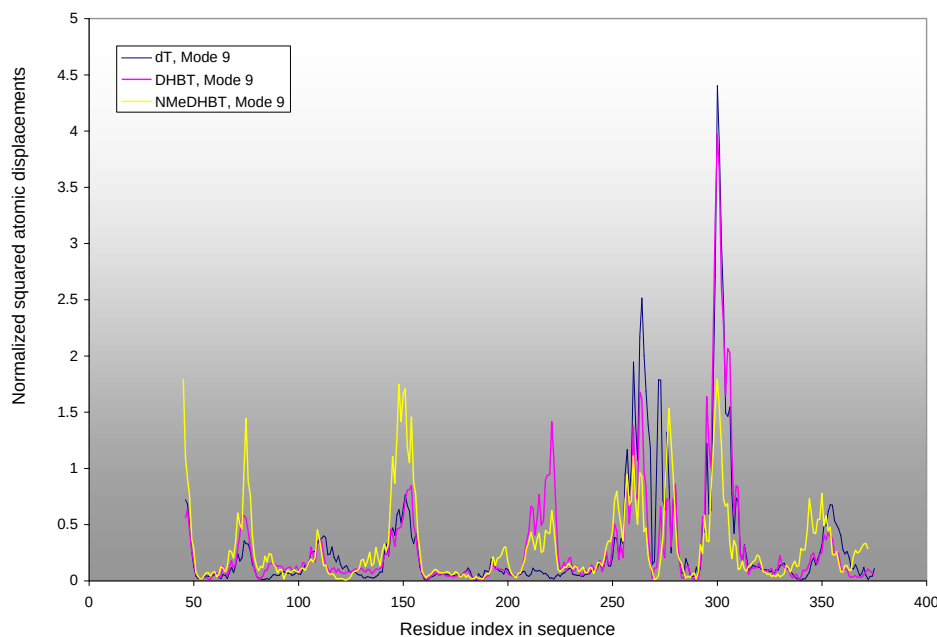


Figure 17: Superposition of the 3rd lowest frequency mode (mode 9) for HSV-1 TK (dT), (DHBT), and (NMeDHBT).

1.3.2 OPEN SUBSTRATE BOUND CONFORMATION

As stated in the methods part, for obtaining an open (accessible), still substrate HSV-1 TK bound conformation, either a “pseudo” apo structure (from which dT was removed) or the holo structure (HSV-1 TK/dT) were used with different pulling scenarios. The “pseudo” apo structure SMDs showed a linear force increase over time, with extremely high forces in the order of 8000 pN after 200 ps, but without any LID opening. For the complex structure, an opening is only observed with Gly206 and Gly254 as anchors. In the other cases, probably too many atoms are kept restrained to allow for a significant LID movement to happen. The crystallographic B-factors and the PCA results of the closed holo structure also point in this direction, as the LID domain and a flexible segment around residue 270 undergo considerable movements. Importantly, the five H-bonds with dT are kept during the whole holo HSV-1 TK SMD run with Gly206 and Gly254 as anchors. This is especially of relevance for Glu225, as this residue is part of the LID domain which is forced to open. Thus, loop opening by SMD was successful for the HSV-1 TK holo structure when using Gly206 and Gly254 as anchors.

To assess whether the obtained open conformation is reasonable and can be used for the substrate pulling experiments, the obtained HSV-1 TK open LID conformation was compared with the open adenylate kinase (AdK), of which a crystal structure is known (PDB ID: 4AKE). The β -sheet core of both overlapped well and both LID domains pointed in the same direction, corroborating that the loop movement simulated for HSV-1 TK is likely, although the AdK LID is much larger. The gradual LID opening of HSV-1 TK in the presence of dT over 200 ps is depicted in Figure 18. A 1 ns CMD confirmed the stability of the open conformation (white loop in Figure 18): The RMSD per residue for the LID domain was less than 1 Å.

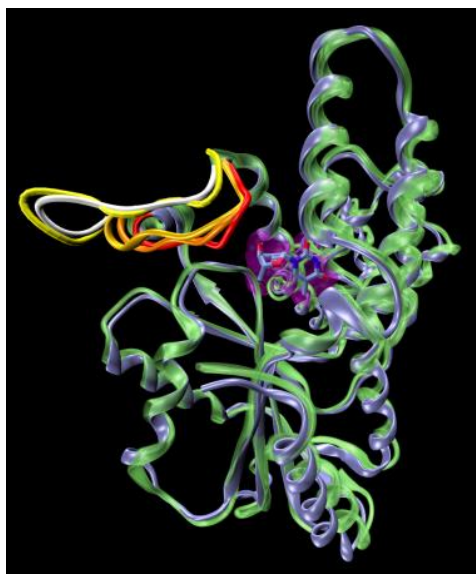


Figure 18: Opening of the LID domain of HSV-1 TK, which is responsible for ATP binding and induced fit movements during phosphorylation. Different conformations are depicted: its crystal structure (red) and conformations at 50 (orange), 100 (light orange), 150 (yellow), and 200 ps (white). The rest of the HSV-1 TK structure is depicted in blue for the starting conformation and in green for the 200 ps conformation. dT is shown with cyan carbons within a transparent violet VdW surface.

1.3.3 SUBSTRATE EXIT ROUTES IDENTIFIED BY RAMD

The substrate exit routes were successfully identified for both HSV-1 TK and hTK1 using RAMD. The only parameters of Vashisth et al.⁵⁸⁵ leading to dT exit from HSV-1 TK within 50 ps in 18% of the cases were $f_0 = 20$ kcal/mol*Å, $m = 10$, and $r_{\min} = 0.006$ Å. For hTK1, taking into account its fairly narrow active site, the altered parameters included a maximal exit time of 100 ps and $f_0 = 75$ kcal/mol*Å. Provided the exit took twice as long as for HSV-1 TK, the substrate egress happened in all cases. Given that the substrate-binding site in HSV-1 TK is considerably larger than in hTK1, the possible exit routes were considerably more heterogeneous. Among the nine identified exit pathways for dT in HSV-1 TK, the six straightest pathways are shown in Figure 19a. The final exit direction corresponds to the orange one in this Figure, as it is most close to the phosphate transfer direction. For hTK1, the analogous direction corresponds to the orange dots in Figure 19b.

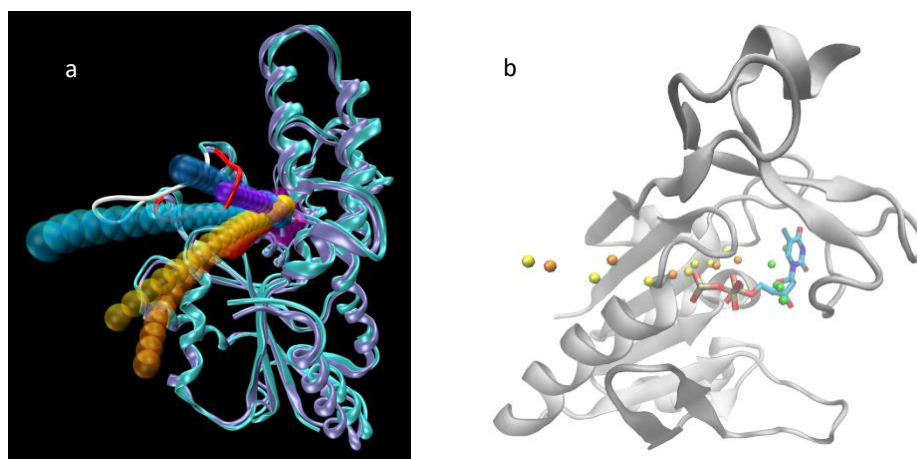


Figure 19: The straightest exit pathways identified in HSV-1 TK (A) and hTK1 (B) with RAMD shown as trajectories of the N3 atom of dT in (a) or dTTP in (b). In the crystal structures, the hydrogen attached to this nitrogen is involved in an H-bond with Gln125 in HSV-1 TK (A) or Val172 in hTK1 (B). In HSV-1 TK (a), the open LID is depicted in white. For comparison, the closed LID is indicated in red. The corresponding three exit directions in hTK1 (b) show two directions in accordance with the location of the ATP binding site (yellow and orange) and one roughly perpendicular direction (green).

1.3.4 SUBSTRATE UNBINDING BY SMD

Different pulling speeds were explored to obtain the longest pulling time possible, and a less widespread accumulated work distribution for the repetitions of each ligand pulling. With a pulling speed of 0.1 Å/ps, any substrate analogue exited the ligand binding pocket on average within 32 ps. Repeating each ligand pulling many times was necessary due to the probabilistic nature of the work distributions. 20 repetitions were chosen here and their median calculated. The median was preferred to the mean as the mean is known to be artificially biased towards extreme data points, whereas the median is not. Thus, the median provides a better estimate of the central tendency, even when not considering the extremes. Only the 10 curves closest to the median were retained for relating their accumulated work values to the experimentally determined affinities or extrapolating affinities.

Picturing dT unbinding

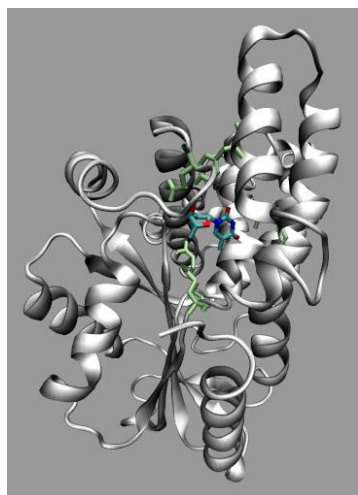
To illustrate the dT unbinding of HSV-1 TK with a median-like work profile, Figure 20 shows a time series of snapshots of the full enzyme, Figure 21a a close-up view on the active site's starting conformation, Figure 21b the conformation after 10 ps, and Figure 21c the serial rupture of the H-bonds involved in fixing dT over 16 ps. In contrast to most H-bonds, the H-bond between Glu225 of the LID domain and dT is only broken after 15 ps, as illustrated in Figure 21c. Thus, for the first 15.5 ps, Glu225 is moving together with the ligand.

Correlation of the substrate-pulling work with experimentally determined affinity data and extrapolation of the affinity for potential substrate analogues

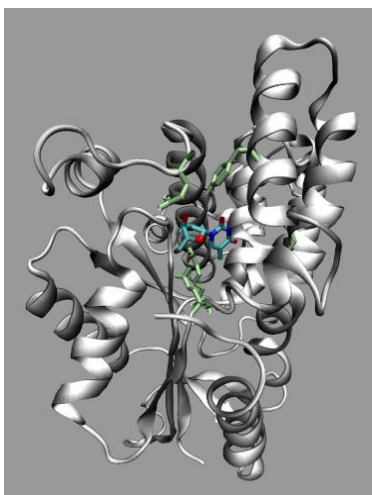
Figure 22a and b represent the central results of this chapter. The calculated substrate extraction work correlated well with experimental affinity data (Figure 22a). The R^2 for a linear regression at the time point 32 ps (the mean time of complete substrate exit) was 0.98 (in Figure 22a and b), not considering the experimental errors indicated in Table 12. These errors vary between $\pm 1.5\%$ (for HPT) and $\pm 25\%$ (for dT).

The SMD parameters yield a “steep enough” correlation to easily distinguish the work values associated with each substrate. A slower SMD pulling velocity would e.g. have led to a flatter work-affinity profile. The standard deviations for the work values increase with simulation time, ranging at 32 ps from 38.2 for dT to 110.8 for HPT.

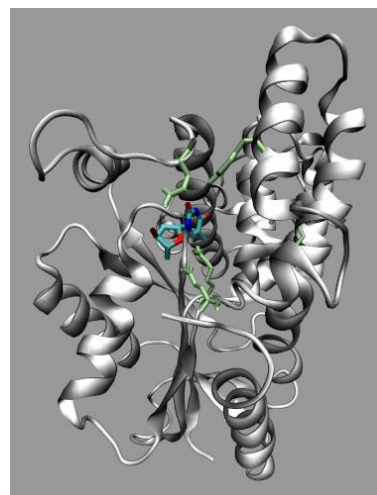
With respect to the extrapolation of unknown affinities, substrates having a double bond in the acyclic part show better K_i values than the corresponding compounds bearing a single bond (e.g. SK-7,8 versus DHBT). Comparing the predicted binding affinity of fluorinated compounds (FHBT, NMeFHBT, and SMA2) with the K_i of the corresponding non-fluorinated compounds (DHBT, NMeDHBT, and SK-7,8), fluorinated compounds are predicted to have a lower affinity than their non-fluorinated counterparts (except for the pair FHBT-DHBT).



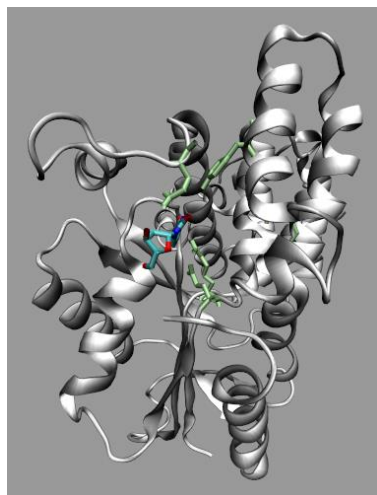
a) Starting conformation



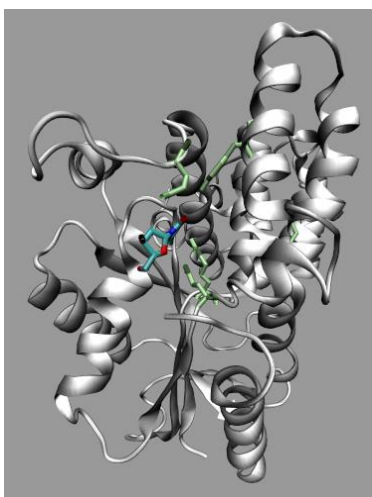
b) After 6 ps



c) After 12 ps



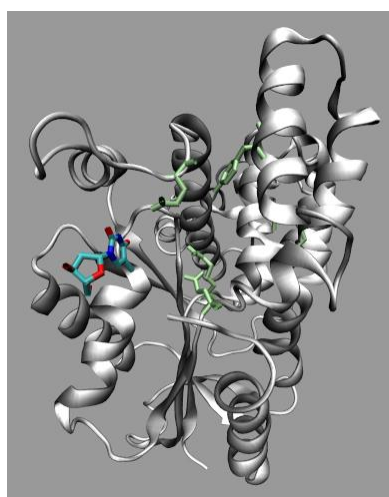
d) After 15 ps



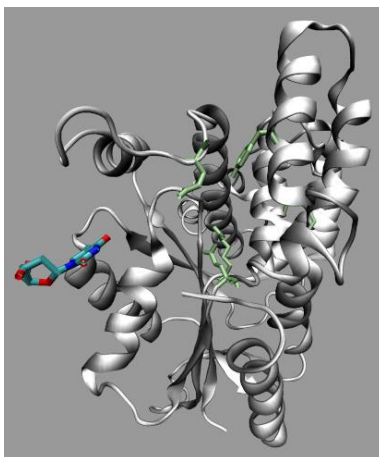
e) After 15.2 ps



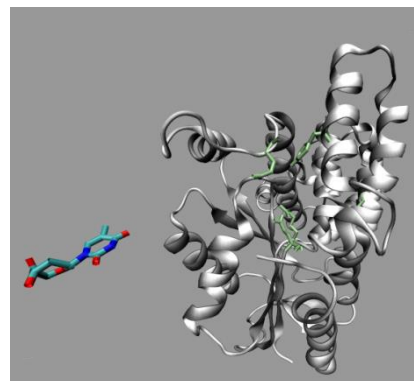
f) After 16 ps



g) After 20 ps



h) After 24 ps



i) After 32 ps

(Figure legend on the next page)

Figure 20 (continued): Nine representative snapshots for the dT exit pathway from HSV-1 TK. dT has cyan carbon atoms and the H-bonding residues in the active site are depicted in light green.

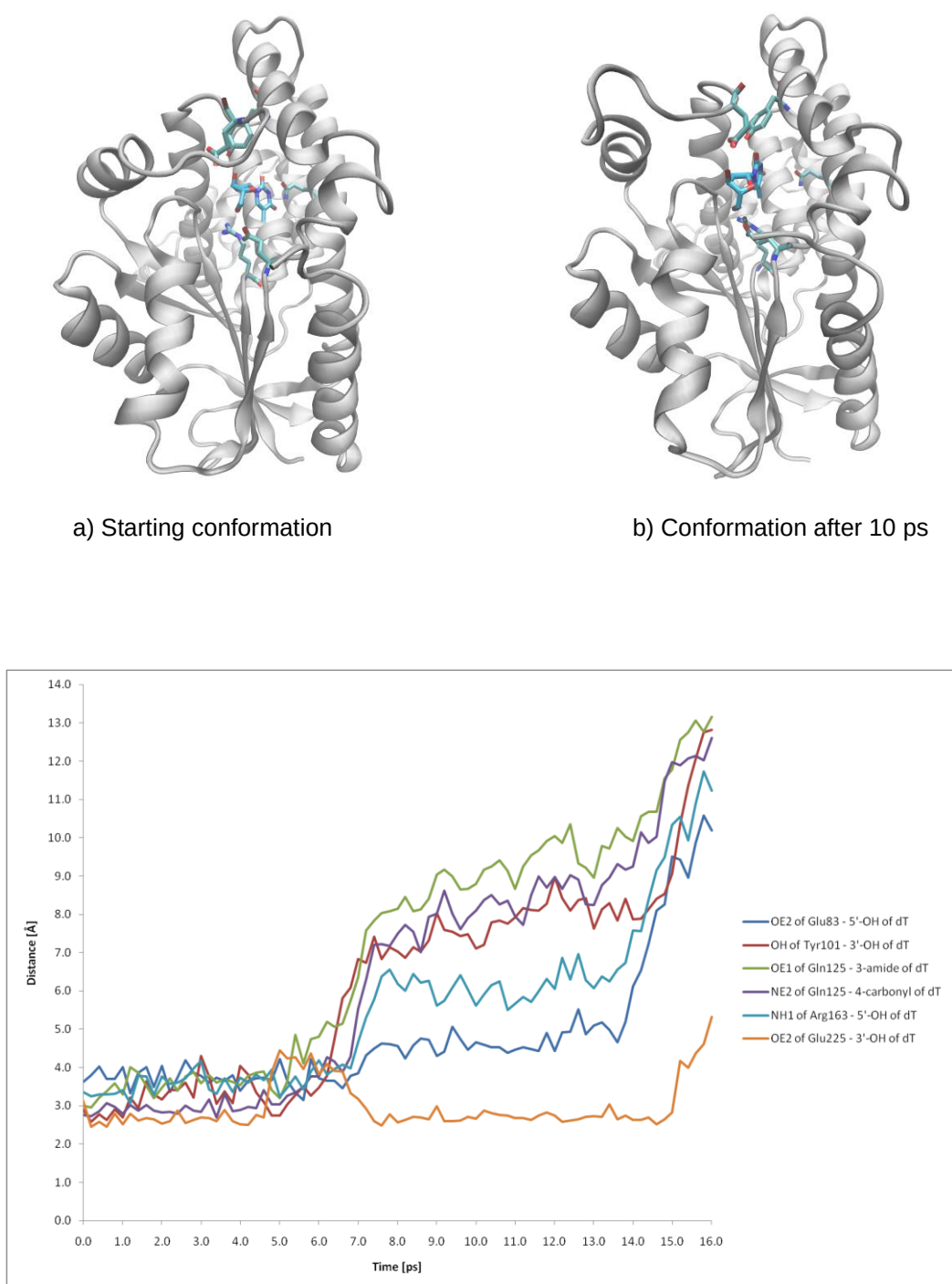


Figure 21a, b, and c: Close-up view on dT exit from HSV-1 TK and the associated H-bond evolution. Figure 21a depicts the starting conformation of the complex, Figure 21b the conformation after 10 ps of SMD, and Figure 21c the H-bond evolution over 16 ps. H-bonds are considered to be broken above 3.6 Å. While most H-bonds are relatively rapidly broken, the last H-bond with the LID domain persists until 15.5 ps.

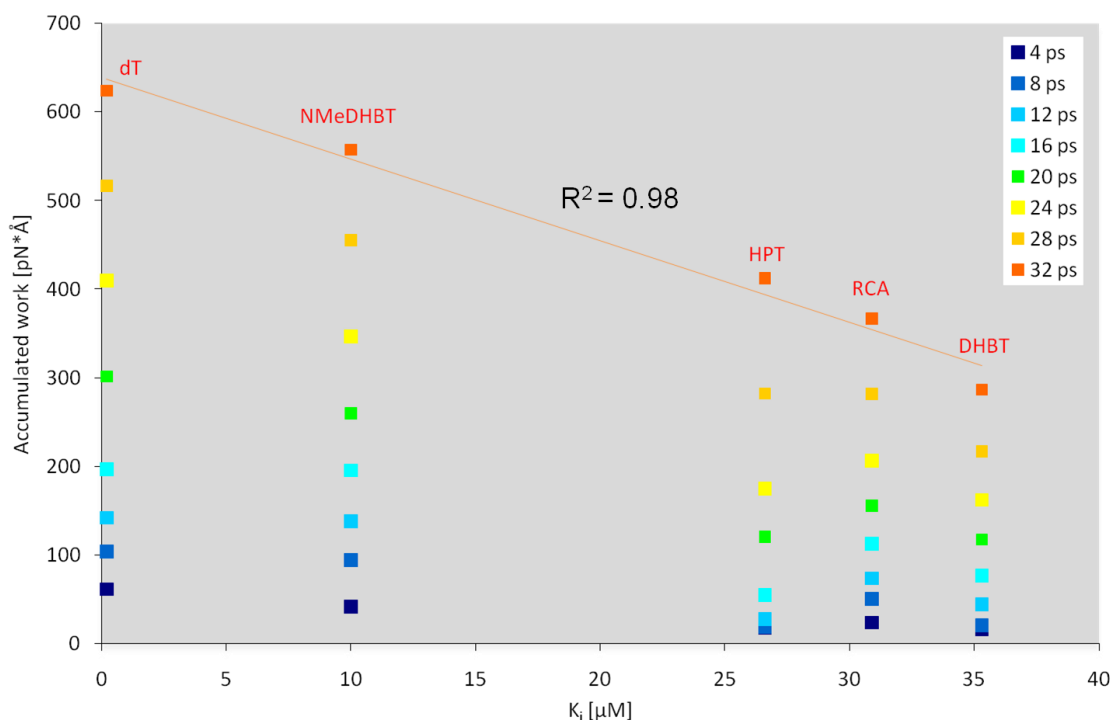
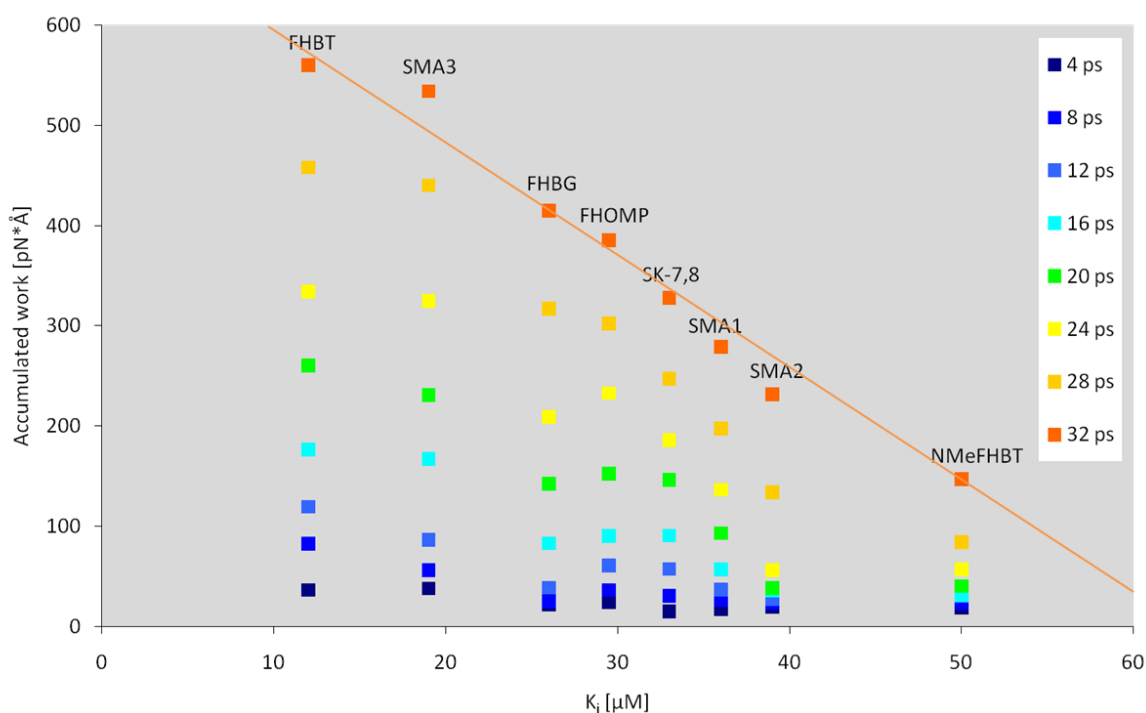
a) Correlation for substrates with known K_i valuesb) Correlation for known K_i values and extrapolation of the K_i values for potential HSV-1 TK substrates

Figure 22a and b: Correlation between the accumulated substrate pulling work [$\text{pN}\cdot\text{\AA}$] and the K_i values [μM] for HSV-1 TK pyrimidine substrates. The time evolution of the accumulated work values for each substrate is shown as a gradient ranging from 4 to 32 ps. Substrates with known experimental K_i values are labelled in red and the ones with a K_i predicted by SMD in black. Generally, the longer the simulation time, the better the work-binding affinity correlation for the known substrates. The R^2 of a linear regression line at 32 ps (mean substrate exit time) is equal to 0.98 in (a) and 0.96 in (b). The standard deviations are omitted in the graphs for the sake of clarity, but mentioned in the text.

Even though for pyrimidine substrates, the applied SMD method was able to distinguish affinities differing by 4.3 μM for known substrates, several limitations of the method were noted and are briefly explained in the following:

- For the 5-substituted uracil derivatives with their particular binding modes owed to the volume of the bases and the substituents at position 5 requiring a mainly non-polar environment, the pulling work profiles were exceptionally flat, meaning that the pulling direction rapidly disrupts the interactions at this position. The here applied SMD protocol is thus not suitable for BTd, BVD, IDU, AHU, and 5IU (Tables 12 and 13).
- Similarly, the 5-substituted uracil derivatives, NCT showed a flat work profile. Compared to SCT, the crystal structure of HSV-1 TK with NCT is neither H-bonding to Tyr101 nor to Glu225⁵⁹⁶. Especially the absent H-bond to Glu225 of the LID domain is thought to cause an easy substrate exit. The work profiles for SCT were in some cases flat and in some cases comparable to the ones of dT, which would rather correspond to its measured K_i value of 4.1 μM ⁴⁷⁷. It is assumed that for the methanocarbathymidines NCT and SCT, the pulling direction was not optimal to fully account for the hydrophobic interactions between HSV-1 TK and methanocarbathymidine during substrate exit.
- The same SMD protocols were also applied to four purine substrates of HSV-1 TK (aciclovir⁶⁰¹, ganciclovir⁴⁸⁶, penciclovir⁴⁸⁶, and the non-substrate inhibitor 9-(4-hydroxybutyl)-N2-phenylguanine⁶⁰¹). No reliable correlation can be drawn yet, because only three substrates do have known K_i values (i.e. 417 μM ⁶⁰² or $200 \pm 50 \mu\text{M}$ ⁴⁷⁸ for aciclovir, 47.6 μM for ganciclovir⁶⁰², and 1.5 μM for penciclovir⁶⁰³) and it is not clear which of the two K_i values should be taken for aciclovir.
- Considering the SMDs on hTK1 performed with dT, dTMP, dTDP, and dTTP, no clear correlation was achieved. This is most probably because all these ligands feature the same base and sugar. By using the same pulling direction, affinity predictions can only be made for thymidine analogues with substitutions in the base or the sugar moiety.

1.4 CONCLUSION

In conclusion, it can be stated that both HSV-1 TK and hTK1 are stable over the 5 ns CMDs, with RMSDs within the range of crystal resolution and showing the same tendency as the crystallographic B-factors. The PCAs were successful in retrieving main motions. Importantly, the PCA results for the three holo HSV-1 TKs show that the LID (residues 219 - 226) motion is opposed to the movement of another flexible segment (around residue 270) in the x- and y-directions, while these domains move rather as one unit in the z-direction. This points at a hinge-like movement of the LID, which could be encoded in the HSV-1 TK structure. Such a movement is known for AdK^{535, 540, 543, 604}, but is shown here for the first time for HSV-1 TK. To fully confirm this statement, the apo HSV-1 TK should be analyzed as well.

The shown SMDs on five different pyrimidines bound to HSV-1 TK are able to distinguish binding affinity differences (K_i values) in the order of 4.3 μM , which corresponds to a difference in ΔG of 0.5 kJ/mol). The excellent correlation of the substrate extraction work with experimentally determined affinity data (R^2 value of 0.98), will potentially allow extrapolating binding affinities of (fluorinated) substrate analogues in the future. Considering HSV-1 TK and purines, the approach is valid, but still more data are needed. The method is generally applicable to potential substrate analogues with modifications in the base and the sugar, e.g. pyrimidines and purines in HSV-1 TK. The SMDs on pyrimidines in hTK1 have shown that the same affinity prediction method is probably applicable to compounds with modifications in the base and the sugar, but not to thymidines which are only phosphorylated to different degrees. For both enzymes, several known substrates need to be at hand to be able to establish the pulling work-affinity correlation.

The approach presents a novelty in the sense that SMDs have so far only been applied to distinguishing active from inactive compounds in an enzyme, but not to a more precise affinity prediction. The developed *in silico* affinity prediction method for small ligands binding to HSV-1 TK is a relative approach and needs calibration with compounds of which the K_i values are known. It is suitable for developing and optimizing PET tracer design prior to synthesis.

Although simulation timescales accessible by SMD are much shorter than the experimentally observed ones and that unbinding pathways are assumed to be linear, which might not correspond to the situation encountered *in vivo*, improvements are underway to develop similar nonequilibrium dynamics-based methods to account for multiple substrate or inhibitor exit routes or curved egress pathways. Recently, Yang et al.⁵⁷⁹ have proposed an SMD method based on adaptive direction adjustments. They found that the metyrapone dissociation from cytochrome P450 3A4 required less dissociation time, a smaller rupture force and displayed a lower energy barrier than with conventional SMD. The pulling direction out of the deeply buried active site was optimized with a genetic algorithm.

One could argue that the median out of 20 SMDs as used in this work is not accurate enough to predict affinity. To statistically fine-tune the work-affinity correlation, future improvements could include the calculation of at least 100 SMDs per confirmed or potential substrate. A PMF calculated based on these data would then be even more precise than a simple median.

It is well known that SMD processes are nonequilibrium processes of which the PMF can be constructed, with the error bound being proportional to the irreversible work done in the simulated system⁶⁰⁵. The two exact equations for nonequilibrium processes are Jarzynski's equality (JE)^{563, 606} and the Crooks equation⁵⁹⁵. While the former is used to calculate a PMF along a reaction coordinate, the latter yields only the free energy differences between the final and the initial states⁶⁰⁷. According to the Crooks equation, the energy difference between two states is given by $\frac{P_F(+\beta W)}{P_R(-\beta W)} = \exp(\beta(W - \Delta A))$, where W is the external work performed from state 0 to 1, β the inverse temperature of the heat reservoir with which the system is equilibrated, P_F and P_R are probability distributions (F standing for forward and R for reverse) and ΔA is the energy difference between the two states. The more general JE is obtained by integrating the Crooks equation: $\langle \exp(-\beta W) \rangle_F = \exp(-\beta \Delta A)$. It relates the work performed during a nonequilibrium process to the free energy difference between two equilibrium states. Brackets denote averages over several simulations. Foundations and applications of the JE are discussed in⁶⁰⁸. It has been noted that applying the JE to reconstruct PMFs from a small number of nonequilibrium unidirectional SMD paths is hindered by the lack of sampling of extreme rare paths⁶⁰⁹. Because of these convergence problems, the JE is generally used via the so-called cumulant expansion^{563, 610, 611}, an example is e.g. the free energy calculation based on 10 SMD trajectories of decalanine stretching⁵⁶⁴. In the here presented case, it is considered that the 20 SMDs done per pyrimidine are not yet sufficient to reconstruct a PMF via the cumulant expansion.

Otherwise, the free energy difference could be recovered using both forward (here: pulling out) and reverse (here: pulling in) medians. Indeed, Bennett has found that it is possible to use the informations contained in both forward and reverse processes together to estimate the free energy difference in a significantly better way than by using just one direction⁶¹². However, here it is somewhat difficult to recover the PMF from path-ensemble averages using both directions, as for the HSV-1 TK (dT) system employed as a test case, the accumulated work values of the reverse reaction were roughly half of the forward reaction and it was not possible to automate the pushing in so as to stop dT once it had reached its crystal-like binding mode. In fact, the calculation in one direction generally overestimates the free energy, while in the other direction it underestimates it⁶¹³. The system's state and evolution are thus path dependent. This phenomenon is called hysteresis. Also, the free energy errors are not the same, because the two directions do not sample the same degrees of freedom. Thus, it is not straightforward to take the mean of the forward and reverse SMDs as an estimate of free energy difference for the TK systems.

1.5 ACKNOWLEDGEMENTS

I am thankful to Prof. Dr. Simon Ametamey from the Institute of Radiopharmaceutical Science at the Swiss Federal Institute of Technology Zurich and his former and present PhDs and Post-Docs for insights into the development of PET tracers. Dr. L. Pernot is acknowledged for having solved the crystal structure of HSV-1 TK in complex with NMeDHBT. Dr. F. Colizzi (University of Bologna, IT), A. Horn (University of Erlangen, DE), Dr. S. Ahmed, and A. Cristiani are kindly acknowledged for insightful discussions. H. Vashisth (Drexel University, USA) is thanked for having provided his RAMD and SMD template scripts. C. Durand is thanked for his help regarding the PCAs and programming with the program R.

1.6 REFERENCES

1. Colizzi, F., Perozzo, R., Scapozza, L., Recanatini, M. & Cavalli, A. Single-molecule pulling simulations can discern active from inactive enzyme inhibitors. *J Am Chem Soc* **132**, 7361-71 (2010).
2. Jorgensen, W. L. Drug discovery: Pulled from a protein's embrace. *Nature* **466**, 42-3 (2010).
3. Cohen, A., Barankiewicz, J., Lederman, H. M. & Gelfand, E. W. Purine and pyrimidine metabolism in human T lymphocytes. Regulation of deoxyribonucleotide metabolism. *J Biol Chem* **258**, 12334-40 (1983).
4. Arner, E. S. & Eriksson, S. Mammalian deoxyribonucleoside kinases. *Pharmacol Ther* **67**, 155-86 (1995).
5. Sutterluety, H. & Seiser, C. Thymidine inhibits the growth-arrest-specific degradation of thymidine kinase protein in transfected L fibroblasts. *J Mol Biol* **265**, 153-60 (1997).
6. Lee, L. S. & Cheng, Y. Human deoxythymidine kinase II: substrate specificity and kinetic behavior of the cytoplasmic and mitochondrial isozymes derived from blast cells of acute myelocytic leukemia. *Biochemistry* **15**, 3686-90 (1976).
7. Ikeda, S., Chakravarty, R. & Ives, D. H. Multisubstrate analogs for deoxynucleoside kinases. Triphosphate end products and synthetic bisubstrate analogs exhibit identical modes of binding and are useful probes for distinguishing kinetic mechanisms. *J Biol Chem* **261**, 15836-43 (1986).
8. Kim, M. Y. & Ives, D. H. Human deoxycytidine kinase: kinetic mechanism and end product regulation. *Biochemistry* **28**, 9043-7 (1989).
9. Birringer, M. S. et al. Structure of a type II thymidine kinase with bound dTTP. *FEBS Lett* **579**, 1376-82 (2005).
10. Eriksson, S., Kierdaszuk, B., Munch-Petersen, B., Oberg, B. & Johansson, N. G. Comparison of the substrate specificities of human thymidine kinase 1 and 2 and deoxycytidine kinase toward antiviral and cytostatic nucleoside analogs. *Biochem Biophys Res Commun* **176**, 586-92 (1991).
11. Munch-Petersen, B., Cloos, L., Tyrsted, G. & Eriksson, S. Diverging substrate specificity of pure human thymidine kinases 1 and 2 against antiviral dideoxynucleosides. *J Biol Chem* **266**, 9032-8 (1991).
12. Munch-Petersen, B., Tyrsted, G., Cloos, L., Beck, R. A. & Eger, K. Different affinity of the two forms of human cytosolic thymidine kinase towards pyrimidine analogs. *Biochim Biophys Acta* **1250**, 158-62 (1995).
13. Eriksson, S., Munch-Petersen, B., Johansson, K. & Eklund, H. Structure and function of cellular deoxyribonucleoside kinases. *Cell Mol Life Sci* **59**, 1327-46 (2002).
14. Celen, S. et al. Synthesis and evaluation of a (99m)Tc-MAMA-propyl-thymidine complex as a potential probe for in vivo visualization of tumor cell proliferation with SPECT. *Nucl Med Biol* **34**, 283-91 (2007).
15. Al-Madhoun, A. S. et al. Synthesis of a small library of 3-(carboranylalkyl)thymidines and their biological evaluation as substrates for human thymidine kinases 1 and 2. *J Med Chem* **45**, 4018-28 (2002).
16. Al-Madhoun, A. S., Johnsamuel, J., Barth, R. F., Tjarks, W. & Eriksson, S. Evaluation of human thymidine kinase 1 substrates as new candidates for boron neutron capture therapy. *Cancer Res* **64**, 6280-6 (2004).
17. Johnsamuel, J. et al. Synthesis of ethyleneoxide modified 3-carboranyl thymidine analogues and evaluation of their biochemical, physicochemical, and structural properties. *Bioorg Med Chem* **12**, 4769-81 (2004).
18. Byun, Y. et al. Synthesis and biological evaluation of neutral and zwitterionic 3-carboranyl thymidine analogues for boron neutron capture therapy. *J Med Chem* **48**, 1188-98 (2005).
19. Yan, J. et al. Synthesis and biological evaluation of 3'-carboranyl thymidine analogues. *Bioorg Med Chem Lett* **12**, 2209-12 (2002).
20. Barth, R. F. et al. Boron-containing nucleosides as potential delivery agents for neutron capture therapy of brain tumors. *Cancer Res* **64**, 6287-95 (2004).
21. Byun, Y. et al. The synthesis and biochemical evaluation of thymidine analogues substituted with nido carborane at the N-3 position. *Appl Radiat Isot* **61**, 1125-30 (2004).
22. Desbouis, D., Struthers, H., Spiwok, V., Kuster, T. & Schibli, R. Synthesis, in vitro, and in silico evaluation of organometallic technetium and rhenium thymidine complexes with retained substrate activity toward human thymidine kinase type 1. *J Med Chem* **51**, 6689-98 (2008).

23. Toyohara, J. et al. Alkyl-fluorinated thymidine derivatives for imaging cell proliferation I. The in vitro evaluation of some alkyl-fluorinated thymidine derivatives. *Nucl Med Biol* **33**, 751-64 (2006).
24. Mitsuya, H. et al. 3'-Azido-3'-deoxythymidine (BW A509U): an antiviral agent that inhibits the infectivity and cytopathic effect of human T-lymphotropic virus type III/lymphadenopathy-associated virus in vitro. *Proc Natl Acad Sci U S A* **82**, 7096-100 (1985).
25. Balzarini, J., de Clercq, E., Ayusawa, D. & Seno, T. Murine mammary FM3A carcinoma cells transformed with the herpes simplex virus type 1 thymidine kinase gene are highly sensitive to the growth-inhibitory properties of (E)-5-(2-bromovinyl)-2'-deoxyuridine and related compounds. *FEBS Lett* **185**, 95-100 (1985).
26. Balzarini, J., Bohman, C. & De Clercq, E. Differential mechanism of cytostatic effect of (E)-5-(2-bromovinyl)-2'-deoxyuridine, 9-(1,3-dihydroxy-2-propoxymethyl)guanine, and other antiherpetic drugs on tumor cells transfected by the thymidine kinase gene of herpes simplex virus type 1 or type 2. *J Biol Chem* **268**, 6332-7 (1993).
27. Balzarini, J., Bohman, C., Walker, R. T. & de Clercq, E. Comparative cytostatic activity of different antiherpetic drugs against herpes simplex virus thymidine kinase gene-transfected tumor cells. *Mol Pharmacol* **45**, 1253-8 (1994).
28. Culver, K. W. et al. In vivo gene transfer with retroviral vector-producer cells for treatment of experimental brain tumors. *Science* **256**, 1550-2 (1992).
29. Qasim, W., Gaspar, H. B. & Thrasher, A. J. T cell suicide gene therapy to aid haematopoietic stem cell transplantation. *Curr Gene Ther* **5**, 121-32 (2005).
30. Denny, W. A. Prodrugs for Gene-Directed Enzyme-Prodrug Therapy (Suicide Gene Therapy). *J Biomed Biotechnol* **2003**, 48-70 (2003).
31. Denny, W. A. Tumor-activated prodrugs--a new approach to cancer therapy. *Cancer Invest* **22**, 604-19 (2004).
32. Tiberghien, P. et al. Administration of herpes simplex-thymidine kinase-expressing donor T cells with a T-cell-depleted allogeneic marrow graft. *Blood* **97**, 63-72 (2001).
33. Markham, A. & Faulds, D. Ganciclovir. An update of its therapeutic use in cytomegalovirus infection. *Drugs* **48**, 455-84 (1994).
34. Sherley, J. L. & Kelly, T. J. Regulation of human thymidine kinase during the cell cycle. *J Biol Chem* **263**, 8350-8 (1988).
35. Dittmann, H. et al. Early changes in [18F]FLT uptake after chemotherapy: an experimental study. *Eur J Nucl Med Mol Imaging* **29**, 1462-9 (2002).
36. Cobben, D. C. et al. 3'-18F-fluoro-3'-deoxy-L-thymidine: a new tracer for staging metastatic melanoma? *J Nucl Med* **44**, 1927-32 (2003).
37. Been, L. B. et al. [18F]FLT-PET in oncology: current status and opportunities. *Eur J Nucl Med Mol Imaging* **31**, 1659-72 (2004).
38. Blasberg, R. G. & Tjuvajev, J. G. Molecular-genetic imaging: current and future perspectives. *J Clin Invest* **111**, 1620-9 (2003).
39. Herschman, H. R. et al. Seeing is believing: non-invasive, quantitative and repetitive imaging of reporter gene expression in living animals, using positron emission tomography. *J Neurosci Res* **59**, 699-705 (2000).
40. Tjuvajev, J. G. et al. Comparison of radiolabeled nucleoside probes (FIAU, FHBG, and FHPG) for PET imaging of HSV-1 TK gene expression. *J Nucl Med* **43**, 1072-83 (2002).
41. Qiao, J. et al. Tumor-specific transcriptional targeting of suicide gene therapy. *Gene Ther* **9**, 168-75 (2002).
42. Tjuvajev, J. G. et al. Imaging adenoviral-mediated herpes virus thymidine kinase gene transfer and expression in vivo. *Cancer Res* **59**, 5186-93 (1999).
43. Hackman, T. et al. Imaging expression of cytosine deaminase-herpes virus thymidine kinase fusion gene (CD/TK) expression with [124I]FIAU and PET. *Mol Imaging* **1**, 36-42 (2002).
44. Bennett, J. J. et al. Positron emission tomography imaging for herpes virus infection: Implications for oncolytic viral treatments of cancer. *Nat Med* **7**, 859-63 (2001).
45. Jacobs, A. et al. Positron emission tomography-based imaging of transgene expression mediated by replication-conditional, oncolytic herpes simplex virus type 1 mutant vectors in vivo. *Cancer Res* **61**, 2983-95 (2001).
46. Iyer, M. et al. 8-[18F]Fluoropenciclovir: an improved reporter probe for imaging HSV-1 TK reporter gene expression in vivo using PET. *J Nucl Med* **42**, 96-105 (2001).
47. Cai, H. et al. Preparation and biological evaluation of 2-amino-6-[18F]fluoro-9-(4-hydroxy-3-hydroxy-methylbutyl) purine (6-[18F]FPCV) as a novel PET probe for imaging HSV-1 TK reporter gene expression. *Nucl Med Biol* **34**, 717-25 (2007).

48. Johayem, A. et al. Synthesis and characterization of a C6 nucleoside analogue for the in vivo imaging of the gene expression of herpes simplex virus type-1 thymidine kinase (HSV-1 TK). *Chem Biodivers* **3**, 274-83 (2006).
49. Serganova, I., Ponomarev, V. & Blasberg, R. Human reporter genes: potential use in clinical studies. *Nucl Med Biol* **34**, 791-807 (2007).
50. Degreve, B., De Clercq, E. & Balzarini, J. Bystander effect of purine nucleoside analogues in HSV-1 tk suicide gene therapy is superior to that of pyrimidine nucleoside analogues. *Gene Ther* **6**, 162-70 (1999).
51. Gambhir, S. S. et al. Imaging transgene expression with radionuclide imaging technologies. *Neoplasia* **2**, 118-38 (2000).
52. Schelling, P. et al. Biochemical and structural characterization of (South)-methanocarbothymidine that specifically inhibits growth of herpes simplex virus type 1 thymidine kinase-transduced osteosarcoma cells. *J Biol Chem* **279**, 32832-8 (2004).
53. Pilger, B. D. et al. Substrate diversity of herpes simplex virus thymidine kinase. Impact Of the kinematics of the enzyme. *J Biol Chem* **274**, 31967-73 (1999).
54. Russ, P. et al. Synthesis and biological evaluation of 5-substituted derivatives of the potent antiherpes agent (north)-methanocarbothymine. *J Med Chem* **46**, 5045-54 (2003).
55. Kussmann-Gerber, S., Kuonen, O., Folkers, G., Pilger, B. D. & Scapozza, L. Drug resistance of herpes simplex virus type 1--structural considerations at the molecular level of the thymidine kinase. *Eur J Biochem* **255**, 472-81 (1998).
56. Brown, D. G. et al. Crystal structures of the thymidine kinase from herpes simplex virus type-1 in complex with deoxythymidine and ganciclovir. *Nat Struct Biol* **2**, 876-81 (1995).
57. Wild, K., Bohner, T., Aubry, A., Folkers, G. & Schulz, G. E. The three-dimensional structure of thymidine kinase from herpes simplex virus type 1. *FEBS Lett* **368**, 289-92 (1995).
58. Schulz, G. E., Muller, C. W. & Diederichs, K. Induced-fit movements in adenylate kinases. *J Mol Biol* **213**, 627-30 (1990).
59. Dreusicke, D. & Schulz, G. E. The glycine-rich loop of adenylate kinase forms a giant anion hole. *FEBS Lett* **208**, 301-4 (1986).
60. Wild, K., Bohner, T., Folkers, G. & Schulz, G. E. The structures of thymidine kinase from herpes simplex virus type 1 in complex with substrates and a substrate analogue. *Protein Sci* **6**, 2097-106 (1997).
61. Champness, J. N. et al. Exploring the active site of herpes simplex virus type-1 thymidine kinase by X-ray crystallography of complexes with aciclovir and other ligands. *Proteins* **32**, 350-61 (1998).
62. Perozzo, R., Jelesarov, I., Bosshard, H. R., Folkers, G. & Scapozza, L. Compulsory order of substrate binding to herpes simplex virus type 1 thymidine kinase. A calorimetric study. *J Biol Chem* **275**, 16139-45 (2000).
63. Muller, C. W., Schlauderer, G. J., Reinstein, J. & Schulz, G. E. Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding. *Structure* **4**, 147-56 (1996).
64. Balasubramaniam, N. K., Veerisetty, V. & Gentry, G. A. Herpesviral deoxythymidine kinases contain a site analogous to the phosphoryl-binding arginine-rich region of porcine adenylate kinase; comparison of secondary structure predictions and conservation. *J Gen Virol* **71** (Pt 12), 2979-87 (1990).
65. Remond, M., Sheldrick, P., Lebreton, F. & Foulon, T. Sequence of the canine herpesvirus thymidine kinase gene: taxon-preferred amino acid residues in the alphaherpesviral thymidine kinases. *Virus Res* **39**, 341-54 (1995).
66. Alber, F., Kuonen, O., Scapozza, L., Folkers, G. & Carloni, P. Density functional studies on herpes simplex virus type 1 thymidine kinase-substrate interactions: the role of Tyr-172 and Met-128 in thymine fixation. *Proteins* **31**, 453-9 (1998).
67. Sulpizi, M., Schelling, P., Folkers, G., Carloni, P. & Scapozza, L. The rational of catalytic activity of herpes simplex virus thymidine kinase. a combined biochemical and quantum chemical study. *J Biol Chem* **276**, 21692-7 (2001).
68. Degreve, B., Esnouf, R., De Clercq, E. & Balzarini, J. Selective abolishment of pyrimidine nucleoside kinase activity of herpes simplex virus type 1 thymidine kinase by mutation of alanine-167 to tyrosine. *Mol Pharmacol* **58**, 1326-32 (2000).
69. Graham, D., Larder, B. A. & Inglis, M. M. Evidence that the 'active centre' of the herpes simplex virus thymidine kinase involves an interaction between three distinct regions of the polypeptide. *J Gen Virol* **67** (Pt 4), 753-8 (1986).

70. Chacko, A. M., Qu, W. & Kung, H. F. Synthesis of 5-fluoroalkylated pyrimidine nucleosides via Negishi cross-coupling. *J Org Chem* **73**, 4874-81 (2008).
71. Chacko, A. M., Qu, W. & Kung, H. F. Synthesis and in vitro evaluation of 5-[(18F)]fluoroalkyl pyrimidine nucleosides for molecular imaging of herpes simplex virus type 1 thymidine kinase reporter gene expression. *J Med Chem* **51**, 5690-701 (2008).
72. Raic-Malic, S. et al. Synthesis, 18F-radiolabelling and biological evaluations of C-6 alkylated pyrimidine nucleoside analogues. *Nucleosides Nucleotides Nucleic Acids* **23**, 1707-21 (2004).
73. Martic, M. PhD thesis: Development of new nucleoside analogues as PET imaging agents for monitoring gene expression. *Swiss Federal Institute of Technology, Zurich ETH diss. no. 17632* (2008).
74. Welin, M. et al. Structures of thymidine kinase 1 of human and mycoplasmic origin. *Proc Natl Acad Sci U S A* **101**, 17970-5 (2004).
75. El Omari, K., Solaroli, N., Karlsson, A., Balzarini, J. & Stammers, D. K. Structure of vaccinia virus thymidine kinase in complex with dTTP: insights for drug design. *BMC Struct Biol* **6**, 22 (2006).
76. Segura-Pena, D., Lutz, S., Monnerjahn, C., Konrad, M. & Lavie, A. Binding of ATP to TK1-like enzymes is associated with a conformational change in the quaternary structure. *J Mol Biol* **369**, 129-41 (2007).
77. Kuzin, A. P. et al. X-ray structure of *Clostridium acetobutylicum* thymidine kinase with ADP. Northeast Structural Genomics Target CAR26. (2006).
78. Kosinska, U. et al. Structural studies of thymidine kinases from *Bacillus anthracis* and *Bacillus cereus* provide insights into quaternary structure and conformational changes upon substrate binding. *Febs J* **274**, 727-37 (2007).
79. Newton, A., Mackay, J. & Crossley, M. The N-terminal zinc finger of the erythroid transcription factor GATA-1 binds GATC motifs in DNA. *J Biol Chem* **276**, 35794-801 (2001).
80. Gilles, A. M. et al. Zinc, a structural component of adenylate kinases from gram-positive bacteria. *J Bacteriol* **176**, 520-3 (1994).
81. Segura-Pena, D. et al. Quaternary structure change as a mechanism for the regulation of thymidine kinase 1-like enzymes. *Structure* **15**, 1555-66 (2007).
82. Byun, Y. et al. Preparation and biological evaluation of 10B-enriched 3-[5-{2-(2,3-dihydroxyprop-1-yl)-o-carboran-1-yl}pentan-1-yl]thymidine (N5-2OH), a new boron delivery agent for boron neutron capture therapy of brain tumors. *J Med Chem* **49**, 5513-23 (2006).
83. Frederiksen, H., Berenstein, D. & Munch-Petersen, B. Effect of valine 106 on structure-function relation of cytosolic human thymidine kinase. Kinetic properties and oligomerization pattern of nine substitution mutants of V106. *Eur J Biochem* **271**, 2248-56 (2004).
84. Fogli, M. V. PhD thesis: Mutagenesis studies on human thymidine kinase 1: Toward an enzyme with broader substrate acceptance. *Swiss Federal Institute of Technology, Zurich ETH diss. no. 16241* (2005).
85. Feynman, R. "Lectures on Physics". *Addison-Wesley* **1**, 3-6 (1963).
86. Karplus, M. & Kuriyan, J. Molecular dynamics and protein function. *Proc Natl Acad Sci U S A* **102**, 6679-85 (2005).
87. Freddolino, P. L., Liu, F., Gruebele, M. & Schulten, K. Ten-microsecond molecular dynamics simulation of a fast-folding WW domain. *Biophys J* **94**, L75-7 (2008).
88. Mura, C. & McCammon, J. A. Molecular dynamics of a kappaB DNA element: base flipping via cross-strand intercalative stacking in a microsecond-scale simulation. *Nucleic Acids Res* **36**, 4941-55 (2008).
89. Amadei, A., Linssen, A. B. & Berendsen, H. J. Essential dynamics of proteins. *Proteins* **17**, 412-25 (1993).
90. Berendsen, H. J. & Hayward, S. Collective protein dynamics in relation to function. *Curr Opin Struct Biol* **10**, 165-9 (2000).
91. de Groot, B. L., van Aalten, D. M., Amadei, A. & Berendsen, H. J. The consistency of large concerted motions in proteins in molecular dynamics simulations. *Biophys J* **71**, 1707-13 (1996).
92. de Groot, B. L., Hayward, S., van Aalten, D. M., Amadei, A. & Berendsen, H. J. Domain motions in bacteriophage T4 lysozyme: a comparison between molecular dynamics and crystallographic data. *Proteins* **31**, 116-27 (1998).
93. Amadei, A., Ceruso, M. A. & Di Nola, A. On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations. *Proteins* **36**, 419-24 (1999).

94. Lange, O. F. & Grubmuller, H. Full correlation analysis of conformational protein dynamics. *Proteins* **70**, 1294-312 (2008).
95. Brooks, B. & Karplus, M. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc Natl Acad Sci U S A* **80**, 6571-5 (1983).
96. Levitt, M., Sander, C. & Stern, P. S. Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. *J Mol Biol* **181**, 423-47 (1985).
97. Ma, J. Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure* **13**, 373-80 (2005).
98. Zheng, W. & Doniach, S. A comparative study of motor-protein motions by using a simple elastic-network model. *Proc Natl Acad Sci U S A* **100**, 13253-8 (2003).
99. Vonrhein, C., Schlauderer, G. J. & Schulz, G. E. Movie of the structural changes during a catalytic cycle of nucleoside monophosphate kinases. *Structure* **3**, 483-90 (1995).
100. Ostermann, N. et al. Insights into the phosphoryltransfer mechanism of human thymidylate kinase gained from crystal structures of enzyme complexes along the reaction coordinate. *Structure* **8**, 629-42 (2000).
101. Sekulic, N., Shuvalova, L., Spangenberg, O., Konrad, M. & Lavie, A. Structural characterization of the closed conformation of mouse guanylate kinase. *J Biol Chem* **277**, 30236-43 (2002).
102. Zhang, H. J., Sheng, X. R., Niu, W. D., Pan, X. M. & Zhou, J. M. Evidence for at least two native forms of rabbit muscle adenylate kinase in equilibrium in aqueous solution. *J Biol Chem* **273**, 7448-56 (1998).
103. Kern, D., Eisenmesser, E. Z. & Wolf-Watz, M. Enzyme dynamics during catalysis measured by NMR spectroscopy. *Methods Enzymol* **394**, 507-24 (2005).
104. Shapiro, Y. E. & Meirovitch, E. Activation energy of catalysis-related domain motion in E. coli adenylate kinase. *J Phys Chem B* **110**, 11519-24 (2006).
105. Shapiro, Y. E. et al. Domain flexibility in ligand-free and inhibitor-bound Escherichia coli adenylate kinase based on a mode-coupling analysis of 15N spin relaxation. *Biochemistry* **41**, 6271-81 (2002).
106. Tugarinov, V., Shapiro, Y. E., Liang, Z., Freed, J. H. & Meirovitch, E. A novel view of domain flexibility in E. coli adenylate kinase based on structural mode-coupling (15N) NMR relaxation. *J Mol Biol* **315**, 155-70 (2002).
107. Kubitzki, M. B. & de Groot, B. L. The atomistic mechanism of conformational transition in adenylate kinase: a TEE-REX molecular dynamics study. *Structure* **16**, 1175-82 (2008).
108. Han, Y., Li, X. & Pan, X. Native states of adenylate kinase are two active sub-ensembles. *FEBS Lett* **528**, 161-5 (2002).
109. Eisenmesser, E. Z. et al. Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **438**, 117-21 (2005).
110. Tobi, D. & Bahar, I. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc Natl Acad Sci U S A* **102**, 18908-13 (2005).
111. Henzler-Wildman, K. A. et al. Intrinsic motions along an enzymatic reaction trajectory. *Nature* **450**, 838-44 (2007).
112. Snow, C., Qi, G. & Hayward, S. Essential dynamics sampling study of adenylate kinase: comparison to citrate synthase and implication for the hinge and shear mechanisms of domain motions. *Proteins* **67**, 325-37 (2007).
113. Arora, K. & Brooks, C. L., 3rd. Large-scale allosteric conformational transitions of adenylate kinase appear to involve a population-shift mechanism. *Proc Natl Acad Sci U S A* **104**, 18496-501 (2007).
114. Pontiggia, F., Zen, A. & Micheletti, C. Small- and large-scale conformational changes of adenylate kinase: a molecular dynamics study of the subdomain motion and mechanics. *Biophys J* **95**, 5901-12 (2008).
115. Hanson, J. A. et al. Illuminating the mechanistic roles of enzyme conformational dynamics. *Proc Natl Acad Sci U S A* **104**, 18055-60 (2007).
116. Lu, Q. & Wang, J. Single molecule conformational dynamics of adenylate kinase: energy landscape, structural correlations, and transition state ensembles. *J Am Chem Soc* **130**, 4772-83 (2008).
117. Bae, E. & Phillips, G. N., Jr. Roles of static and dynamic domains in stability and catalysis of adenylate kinase. *Proc Natl Acad Sci U S A* **103**, 2132-7 (2006).
118. Beach, H., Cole, R., Gill, M. L. & Loria, J. P. Conservation of mus-ms enzyme motions in the apo- and substrate-mimicked state. *J Am Chem Soc* **127**, 9167-76 (2005).

119. Aden, J. & Wolf-Watz, M. NMR identification of transient complexes critical to adenylate kinase catalysis. *J Am Chem Soc* **129**, 14003-12 (2007).
120. Henzler-Wildman, K. A. et al. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **450**, 913-6 (2007).
121. Cukier, R. I. Apo adenylate kinase encodes its holo form: a principal component and varimax analysis. *J Phys Chem B* **113**, 1662-72 (2009).
122. Ma, B., Kumar, S., Tsai, C. J. & Nussinov, R. Folding funnels and binding mechanisms. *Protein Eng* **12**, 713-20 (1999).
123. Lange, O. F. & Grubmuller, H. Can principal components yield a dimension reduced description of protein dynamics on long time scales? *J Phys Chem B* **110**, 22842-52 (2006).
124. Kamikubo, H. et al. Characterization of the solution structure of the M intermediate of photoactive yellow protein using high-angle solution x-ray scattering. *Biophys J* **92**, 3633-42 (2007).
125. Liu, T., Whitten, S. T. & Hilser, V. J. Functional residues serve a dominant role in mediating the cooperativity of the protein ensemble. *Proc Natl Acad Sci U S A* **104**, 4347-52 (2007).
126. Lou, H. & Cukier, R. I. Molecular dynamics of apo-adenylate kinase: a distance replica exchange method for the free energy of conformational fluctuations. *J Phys Chem B* **110**, 24121-37 (2006).
127. Cui, Q., Li, G., Ma, J. & Karplus, M. A normal mode analysis of structural plasticity in the biomolecular motor F(1)-ATPase. *J Mol Biol* **340**, 345-72 (2004).
128. Krebs, W. G. et al. Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic. *Proteins* **48**, 682-95 (2002).
129. James, L. C. & Tawfik, D. S. Conformational diversity and protein evolution--a 60-year-old hypothesis revisited. *Trends Biochem Sci* **28**, 361-8 (2003).
130. Kumar, S., Ma, B., Tsai, C. J., Sinha, N. & Nussinov, R. Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci* **9**, 10-9 (2000).
131. Goh, C. S., Milburn, D. & Gerstein, M. Conformational changes associated with protein-protein interactions. *Curr Opin Struct Biol* **14**, 104-9 (2004).
132. Bahar, I., Chennubhotla, C. & Tobi, D. Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. *Curr Opin Struct Biol* **17**, 633-40 (2007).
133. Cavasotto, C. N. & Orry, A. J. Ligand docking and structure-based virtual screening in drug discovery. *Curr Top Med Chem* **7**, 1006-14 (2007).
134. Lou, H. & Cukier, R. I. Molecular dynamics of apo-adenylate kinase: a principal component analysis. *J Phys Chem B* **110**, 12796-808 (2006).
135. Nicolay, S. & Sanejouand, Y. H. Functional modes of proteins are among the most robust. *Phys Rev Lett* **96**, 078104 (2006).
136. Ikeguchi, M., Ueno, J., Sato, M. & Kidera, A. Protein structural change upon ligand binding: linear response theory. *Phys Rev Lett* **94**, 078102 (2005).
137. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc Natl Acad Sci U S A* **99**, 12562-6 (2002).
138. Laio, A., Rodriguez-Forte, A., Gervasio, F. L., Ceccarelli, M. & Parrinello, M. Assessing the accuracy of metadynamics. *J Phys Chem B* **109**, 6714-21 (2005).
139. Alonso, H., Bliznyuk, A. A. & Gready, J. E. Combining docking and molecular dynamic simulations in drug design. *Med Res Rev* **26**, 531-68 (2006).
140. Jarzynski, C. Nonequilibrium equality for free energy difference. *Phys Rev Lett* **78**, 2690-2693 (1997).
141. Park, S., Khalili-Araghi, F., Tajkhorshid, E. & Schulten, K. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *J Chem Phys* **119**, 3559-3566 (2003).
142. Izrailev, S. et al. Steered molecular dynamics. *Lecture Notes in Computational Science and Engineering; Computational Molecular Dynamics: Challenges, Methods, Ideas; Springer Verlag, New York* **4**, 39-65 (1998).
143. Binnig, G., Quate, C. F. & Gerber, C. Atomic force microscope. *Phys Rev Lett* **56**, 930-933 (1986).
144. Sotomayor, M. & Schulten, K. Single-molecule experiments in vitro and in silico. *Science* **316**, 1144-8 (2007).
145. Izrailev, S., Stepaniants, S., Balsera, M., Oono, Y. & Schulten, K. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys J* **72**, 1568-81 (1997).

146. Isralewitz, B., Izrailev, S. & Schulten, K. Binding pathway of retinal to bacterio-opsin: a prediction by molecular dynamics simulations. *Biophys J* **73**, 2972-9 (1997).
147. Wriggers, W. & Schulten, K. Nucleotide-dependent movements of the kinesin motor domain predicted by simulated annealing. *Biophys J* **75**, 646-61 (1998).
148. Li, W. et al. Possible pathway(s) of metyrapone egress from the active site of cytochrome P450 3A4: a molecular dynamics simulation. *Drug Metab Dispos* **35**, 689-96 (2007).
149. Jensen, M. O., Yin, Y., Tajkhorshid, E. & Schulten, K. Sugar transport across lactose permease probed by steered molecular dynamics. *Biophys J* **93**, 92-102 (2007).
150. Kosztin, D., Izrailev, S. & Schulten, K. Unbinding of retinoic acid from its receptor studied by steered molecular dynamics. *Biophys J* **76**, 188-97 (1999).
151. Niu, C. et al. Dynamic mechanism of E2020 binding to acetylcholinesterase: a steered molecular dynamics simulation. *J Phys Chem B* **109**, 23730-8 (2005).
152. Xu, Y. et al. How does huperzine A enter and leave the binding gorge of acetylcholinesterase? Steered molecular dynamics simulations. *J Am Chem Soc* **125**, 11340-9 (2003).
153. Lu, H., Isralewitz, B., Krammer, A., Vogel, V. & Schulten, K. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys J* **75**, 662-71 (1998).
154. Isralewitz, B., Gao, M. & Schulten, K. Steered molecular dynamics and mechanical functions of proteins. *Curr Opin Struct Biol* **11**, 224-30 (2001).
155. Gao, M., Lu, H. & Schulten, K. Simulated refolding of stretched titin immunoglobulin domains. *Biophys J* **81**, 2268-77 (2001).
156. Zhang, J., King, C. A., Dalby, K. & Ren, P. Conformational preference of ChaK1 binding peptides: a molecular dynamics study. *PMC Biophys* **3**, 2 (2010).
157. Yang, K., Liu, X., Wang, X. & Jiang, H. A steered molecular dynamics method with adaptive direction adjustments. *Biochem Biophys Res Commun* **379**, 494-8 (2009).
158. Pongprayoon, P., Beckstein, O., Wee, C. L. & Sansom, M. S. Simulations of anion transport through OprP reveal the molecular basis for high affinity and selectivity for phosphate. *Proc Natl Acad Sci U S A* **106**, 21614-8 (2009).
159. Ludemann, S. K., Lounnas, V. & Wade, R. C. How do substrates enter and products exit the buried active site of cytochrome P450cam? 2. Steered molecular dynamics and adiabatic mapping of substrate pathways. *J Mol Biol* **303**, 813-30 (2000).
160. Winn, P. J., Ludemann, S. K., Gauges, R., Lounnas, V. & Wade, R. C. Comparison of the dynamics of substrate access channels in three cytochrome P450s reveals different opening mechanisms and a novel functional role for a buried arginine. *Proc Natl Acad Sci U S A* **99**, 5361-6 (2002).
161. Schleinkofer, K., Sudarko, Winn, P. J., Ludemann, S. K. & Wade, R. C. Do mammalian cytochrome P450s show multiple ligand access pathways and ligand channelling? *EMBO Rep* **6**, 584-9 (2005).
162. Wang, T. & Duan, Y. Chromophore channeling in the G-protein coupled receptor rhodopsin. *J Am Chem Soc* **129**, 6970-1 (2007).
163. Vashisth, H. & Abrams, C. F. Ligand escape pathways and (un)binding free energy calculations for the hexameric insulin-phenol complex. *Biophys J* **95**, 4193-204 (2008).
164. Perakyla, M. Ligand unbinding pathways from the vitamin D receptor studied by molecular dynamics simulations. *Eur Biophys J* **38**, 185-98 (2009).
165. Gervasio, F. L., Laio, A. & Parrinello, M. Flexible docking in solution using metadynamics. *J Am Chem Soc* **127**, 2600-7 (2005).
166. Masetti, M., Cavalli, A., Recanatini, M. & Gervasio, F. L. Exploring complex protein-ligand recognition mechanisms with coarse metadynamics. *J Phys Chem B* **113**, 4807-16 (2009).
167. Berman, H. M. et al. The Protein Data Bank. *Nucleic Acids Res* **28**, 235-42 (2000).
168. Berman, H. M. et al. The Protein Data Bank. *Acta Crystallogr D Biol Crystallogr* **58**, 899-907 (2002).
169. Case, D. A. et al. The Amber biomolecular simulation programs. *J Comput Chem* **26**, 1668-88 (2005).
170. Homeyer, N., Horn, A. H., Lanig, H. & Sticht, H. AMBER force-field parameters for phosphorylated amino acids in different protonation states: phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine. *J Mol Model* **12**, 281-9 (2006).
171. Walker, R. C. PhD thesis: The Development of a QM/MM Based Linear Response Method and its Application to Proteins. *Imperial College London* (2003).
172. Phillips, J. C. et al. Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781-802 (2005).

173. Seeber, M., Cecchini, M., Rao, F., Settanni, G. & Caflisch, A. Wordom: a program for efficient analysis of molecular dynamics simulations. *Bioinformatics* **23**, 2625-7 (2007).
174. Marco, E. & Gago, F. Overcoming the inadequacies or limitations of experimental structures as drug targets by using computational modeling tools and molecular dynamics simulations. *ChemMedChem* **2**, 1388-401 (2007).
175. Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* **267**, 727-48 (1997).
176. Kellenberger, E., Rodrigo, J., Muller, P. & Rognan, D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins* **57**, 225-42 (2004).
177. Pospisil, P., Kuoni, T., Scapozza, L. & Folkers, G. Methodology and problems of protein-ligand docking: case study of dihydroorotate dehydrogenase, thymidine kinase, and phosphodiesterase 4. *J Recept Signal Transduct Res* **22**, 141-54 (2002).
178. Crooks, G. E. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* **60**, 2721-6 (1999).
179. Protá, A. et al. Kinetics and crystal structure of the wild-type and the engineered Y101F mutant of Herpes simplex virus type 1 thymidine kinase interacting with (North)-methanocarba-thymidine. *Biochemistry* **39**, 9597-603 (2000).
180. De Winter, H. & Herdewijn, P. Understanding the binding of 5-substituted 2'-deoxyuridine substrates to thymidine kinase of herpes simplex virus type-1. *J Med Chem* **39**, 4727-37 (1996).
181. Wurth, C. et al. The effect of substrate binding on the conformation and structural stability of Herpes simplex virus type 1 thymidine kinase. *Protein Sci* **10**, 63-73 (2001).
182. Gohlke, H. & Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew Chem Int Ed Engl* **41**, 2644-76 (2002).
183. Hess, B. Similarities between principal components of protein dynamics and random diffusion. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* **62**, 8438-48 (2000).
184. Hess, B. Convergence of sampling in protein simulations. *Phys Rev E Stat Nonlin Soft Matter Phys* **65**, 031910 (2002).
185. Bennett, M. S. et al. Structure to 1.9 Å resolution of a complex with herpes simplex virus type-1 thymidine kinase of a novel, non-substrate inhibitor: X-ray crystallographic comparison with binding of aciclovir. *FEBS Lett* **443**, 121-5 (1999).
186. Kokoris, M. S., Sabo, P., Adman, E. T. & Black, M. E. Enhancement of tumor ablation by a selected HSV-1 thymidine kinase mutant. *Gene Ther* **6**, 1415-26 (1999).
187. Schelling, P., Folkers, G. & Scapozza, L. A spectrophotometric assay for quantitative determination of *k_{cat}* of herpes simplex virus type 1 thymidine kinase substrates. *Anal Biochem* **295**, 82-7 (2001).
188. Maragakis, P. & Karplus, M. Large amplitude conformational change in proteins explored with a plastic network model: adenylate kinase. *J Mol Biol* **352**, 807-22 (2005).
189. Balsera, M., Stepaniants, S., Izrailev, S., Oono, Y. & Schulten, K. Reconstructing potential energy functions from simulated force-induced unbinding processes. *Biophys J* **73**, 1281-7 (1997).
190. Jarzynski, C. Rare events and the convergence of exponentially averaged work values. *Phys Rev E Stat Nonlin Soft Matter Phys* **73**, 046105 (2006).
191. Chelli, R., Marsili, S. & Procacci, P. Calculation of the potential of mean force from nonequilibrium measurements via maximum likelihood estimators. *Phys Rev E Stat Nonlin Soft Matter Phys* **77**, 031104 (2008).
192. Jarzynski, C. Nonequilibrium work relations: foundations and applications. *Eur Phys J B* **64**, 331-340 (2008).
193. Kosztin, I., Barz, B. & Janosi, L. Calculating potentials of mean force and diffusion coefficients from nonequilibrium processes without Jarzynski's equality. *J Chem Phys* **124**, 64106 (2006).
194. Jensen, M. O., Park, S., Tajkhorshid, E. & Schulten, K. Energetics of glycerol conduction through aquaglyceroporin GlpF. *Proc Natl Acad Sci U S A* **99**, 6731-6 (2002).
195. Hummer, G. & Szabo, A. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proc Natl Acad Sci U S A* **98**, 3658-61 (2001).
196. Bennett, C. H. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J Comput Phys* **22**, 245-268 (1976).
197. Lu, N., Kofke, D. A. & Woolf, T. B. Improving the efficiency and reliability of free energy perturbation calculations using overlap sampling methods. *J Comput Chem* **25**, 28-39 (2004).

CHAPTER 2 – STABILIZATION OF THERAPEUTIC MONOCLONAL ANTIBODIES USING SMALL MOLECULE AGGREGATION BREAKERS

ABSTRACT

In the context of stabilizing therapeutic monoclonal antibodies, this work deals with the *in silico* discovery and *in vitro* validation of small molecule aggregation breakers and is based on observations described in a previous patent application (US 9155745 B2). According to this patent, the corticosteroids dexamethasone phosphate and betamethasone phosphate stabilize the antibody bevacizumab by reducing its propensity to form dimers or reverting already formed aggregates into an essentially monomeric state. Using homology modelling, a monomer and a dimer aggregation model of bevacizumab were built. Bevacizumab is a humanized monoclonal antibody formed by two Fab regions responsible for its activity and one Fc region derived from immunoglobulin G1 (IgG1). The crystal contacts-based aggregation model revealed a unique contact region between the Fc of one monomer and one Fab of the adjacent monomer, which is conserved in marketed IgG1 antibodies like trastuzumab (Herceptin®) and rituximab (MabThera®). Different steroids were docked onto the monomer Fc region, checked for their clashing propensity in the aggregation model and experimentally validated by Asymmetric Field-Flow Fractionation coupled to Multi-Angle Light Scattering (AFFF-MALS). A ligand-based search for similar aggregation breaking molecules yielded adenosine 5'-monophosphate (AMP) and was confirmed by docking and experimental validation to be an aggregation breaker. Because of the conserved primary aggregation region in IgG1 antibodies, this molecule is predicted to be broadly applicable to current antibody aggregation issues.

2.1 BACKGROUND

The work on antibody aggregation breakers presented in this chapter was carried out in collaboration with the Laboratory of Pharmaceutics, Biopharmaceuticals, and Pharmaceutical Technology of the University of Geneva (Prof. Dr. Robert Gurny and Marieke Veurink). Here, a detailed view of the *in silico* scientific background of two new patents (WO 2011/121559 A2 and WO 2011/121560 A2) on stabilized intact antibody formulations, related methods and uses thereof is given; particularly, methods of stabilizing an intact antibody in a liquid carrier using small molecular weight compounds.

To put this work in a broader research context, first current challenges in the field of therapeutic antibodies are described, focusing on formulation stability and not on other pharmacokinetic issues, as well as ways to improve antibody stability. The starting points of the project, the working hypothesis and the molecular modelling study are then introduced. Next, the information value of the model based on experimental validation is assessed. Other low molecular weight compounds having an aggregation breaking effect are subsequently presented and a first proof-of-principle with a molecule commonly found in the human body is provided. To conclude, potential future directions of the project are given.

2.1.1 STABILITY: A MAJOR CHALLENGE IN THE FIELD OF THERAPEUTIC ANTIBODIES

Therapeutic antibodies are currently the fastest growing area of biopharmaceuticals with an average yearly market growth rate of 35% since 2001⁶¹⁴. Especially monoclonal antibodies are forecast to outpace the market growth rates of small molecules, therapeutic proteins, and vaccines⁶¹⁵. The recent development of chimeric and fully-humanized monoclonal antibodies has spawned an unprecedented interest in using these molecules as therapeutic agents since they can specifically target molecules implicated in disease, thus essentially side-stepping the secondary effects that may be associated with conventional drug therapies. Recent progress in recombinant gene technology has enabled the large-scale production of physiologically active proteins such as monoclonal antibodies for diagnostic and therapeutic applications.

The provision of stable therapeutic protein formulations, particularly stable antibody formulations, is challenging. Physical and chemical instability of antibodies in aqueous media is a complex function of solution conditions and temperature. Antibodies are, for example, susceptible to deamidation, isomerisation, oxidation, proteolysis, aggregation, and other covalent modifications. Degradation of

antibody formulations due to aggregation phenomena is a considerable issue. Not only does the formation of aggregates lead to a reduction in antibody activity, thereby reducing the efficacy of the protein drug, but may also result in potential clinical side-effects or toxicity, since aggregates can increase the immunogenicity of the antibody drug ⁶¹⁶. Antibody aggregation is also a source of batch-to-batch variations in the antibody production chain, and its control leads to regulatory and quality control burdens, with their associated costs. Further, the propensity of antibodies to aggregate adversely affects the stability of therapeutic antibody formulations upon storage, including their shelf-life, and their administration time once removed from optimum storage conditions.

Antibody stability depends on protein concentration, buffer concentration, salt concentration, or agitation. Stabilizing antibodies is problematic since antibodies are extremely sensitive to environmental conditions. These conditions render aggregation and degradation difficult to predict, notably because each antibody may have a specific and characteristic stability profile.

2.1.2 WAYS TO IMPROVE ANTIBODY STABILITY

Several approaches have been investigated to attempt to improve antibody stability. These include approaches based on the addition of “stabilizing” agents to a solution containing immunoglobulin (Ig) and attempts to engender single amino acid mutations at the site(s) implicated in the formation of aggregates on the Ig molecules. Examples of species investigated as “stabilizing” agents in prior patented attempts to improve stability of Ig in solution include polysorbate-based surfactants (GB2175906), amino acids (EP0025275, WO2005/049078), polyethers (EP0018609), glycerin, albumin, and dextran sulphate (US4,808,705). However, their success has been limited. One reason is that these “stabilizing” agents are directed at optimizing the environment in which the Ig is contained, and not specifically at interfering with the mechanism of interaction of Ig molecules regarding the formation of aggregates. This approach has also limitations regarding the quantity of stabilizing agent(s) that may be required to produce a positive effect. Such quantities may have other detrimental effects on Ig molecules, such as protein unfolding (e.g. for surfactants), or on the suitability and safety of the “stabilized” preparations for clinical administration. Single amino acid mutations in Igs could provide a method of specifically targeting sites implicated in aggregation, but such an approach necessarily modifies the structure of the Ig, and this may affect both its clinical efficacy, and its immunogenicity in the recipient.

Although many different approaches have been proposed, and certain methods have been incorporated into antibody formulations, aggregation is still an issue. There is to date no available single, effective, and widely applicable solution to the aggregation of Igs used for clinical applications. It becomes clear from the above challenges that effective methods for inhibiting and/or reducing antibody aggregations are needed. Despite the previous investigations on preventing antibody aggregation, the precise nature of the antibody-antibody binding and contact surfaces of aggregates remains unclear. A better understanding of monoclonal antibodies at the structural level can help to address aggregation issues affecting production, formulation, and stability, such as the loss of biological activity, solubility, and even increased immunogenicity. This better understanding could then lead to improved stability and shelf-life.

2.1.3 IMMUNOGLOBULIN STRUCTURE

To gain understanding of monoclonal antibodies at structural level, an introduction of the structure of IgG1 is provided, which is representative of therapeutic antibodies. IgG1 is composed of 4 polypeptide chains, two identically long light (L) and two identically long heavy (H) chains, interconnected by non-covalent interactions and disulfide bridges, so that each H is linked to an L and both Hs are linked.

The full human IgG1 has an overall Y shape and displays functionally distinct fragments: Fab (Fragment, antigen binding), Fv (Fragment, variable; the tip part of the Fab) and Fc (Fragment, crystallisable; found to crystallize readily). After a partial digestion of the IgG by proteases, two Fabs and one Fc are obtained. Each of IgGs' polypeptide chains consists of a variable (V) amino terminus and a constant (C) region. The H chains are composed of one variable (VH) and three constant

domains each (CH1 to 3), while the L chains contain one variable (VL) and one constant (CL) domain each (Figure 23).

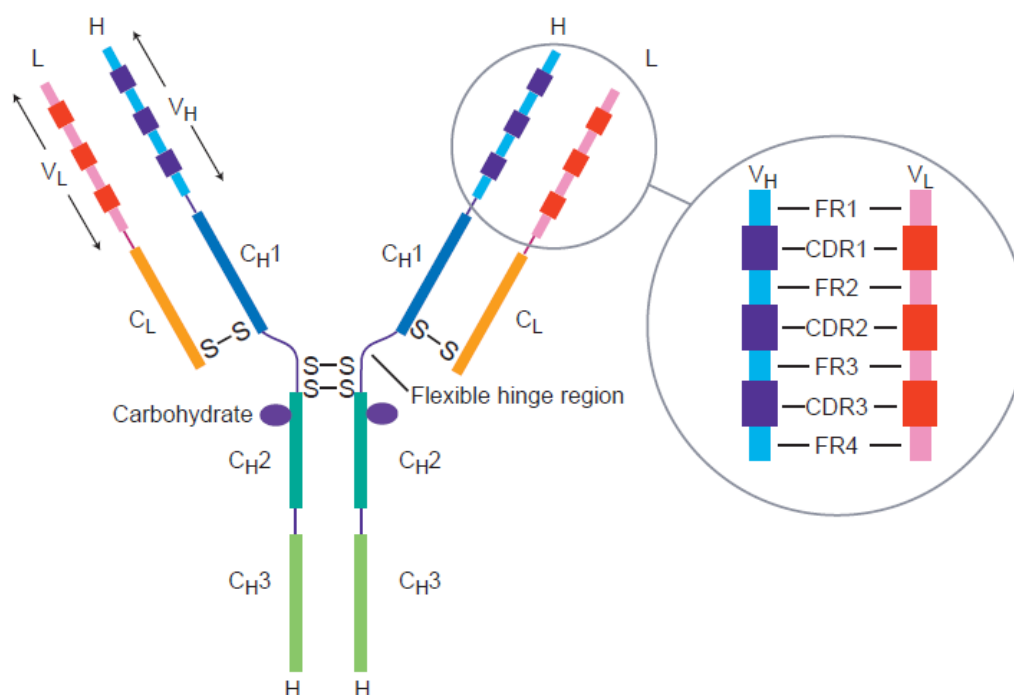


Figure 23: Key features of the protein structure of human IgG1 taken from ⁶¹⁷. Light (L) and heavy (H) chains are covalently linked by disulphide bonds (shown as S–S) between cysteines. The variable domains of the heavy chain (VH) and the light chain (VL) are involved in antigen binding. The heavy chain constant region consists of three constant domains (CH1, CH2, and CH3) and the (flexible) hinges. The light chain also displays a constant domain (CL). The variable regions of the heavy and light chains comprise four framework regions (FRs; FR1, FR2, FR3, and FR4) and three complementarity-determining regions (CDRs; CDR1, CDR2, and CDR3). The carbohydrates indicate the positions at which IgG1 is glycosylated to form a functional antibody.

Each Fab consists of one VL, one VH, one CL and one CH1. The tip part of a Fab, the Fv, is formed by VL and VH to yield identical antigen binding sites at each top of the two arms of the Y, allowing for a much more stable fixation of an antigen than with just one arm. Each Fv consists of a framework (FR) and a hypervariable (HV) region, the latter displaying an antigen binding site with complementarity determining regions (CDRs). Differing in length and sequence between different antibodies, CDRs are responsible of the recognition specificity and binding affinity towards the antigens. The CDRs are named CDRL1-3 and CDRH1-3 according to their position in the L or H chain. They are involved in antigen binding and form a paratope, as opposed to the complementary region on the antigen called epitope. Furthermore, the four FRs of each full antibody provide with their β -sheets a stabilizing framework for the CDRs.

The C-terminal Hs (paired CH2s and CH3s) forming the trunk of the Y, which is called Fc, modulate immune cell activity. The flexible regions between the tips and the trunk are called hinges. Both the Fc and the hinges determine the Ig isotype, while the overall structure is similar among isotypes. The immune modulatory role of the Fc has two components: (i) its binding to specific proteins elicits an appropriate immune response for a given antigen and (ii) its association with various cell receptors and complement proteins mediates antigen destruction.

All Ig domains consist of two antiparallel β -sheets, linked by a disulphide bond, and folding into a β -barrel or –sandwich, which is called the Ig fold.

2.1.4 AVASTIN® (BEVACIZUMAB) AND LUCENTIS® (RANIBIZUMAB)

To date, most therapeutic monoclonal antibodies in clinical use are of the immunoglobulin G (IgG) type. One example is bevacizumab (Avastin®), a recombinant humanized monoclonal IgG1 antibody binding to and inhibiting the tumor angiogenesis activity of the human vascular endothelial growth factor (VEGF). It has a molecular weight of 149 kDa⁶¹⁸ and contains the human FRs as well as the CDRs of a murine antibody that bind to VEGF. It is commercialized by Genentech and the sequence of its Fabs has been published⁶¹⁹. Avastin® recognizes and blocks VEGF-A⁶²⁰ and inhibits thereby endothelial cell proliferation and new blood vessel formation. According to the DrugBank (<http://www.drugbank.ca>), known targets of Avastin® include the VEGF-A, the high affinity immunoglobulin γ Fc receptor I, some low affinity immunoglobulin γ Fc region receptors and different complement subcomponents.

In cancer treatment, Avastin® acts as an antiangiogenesis, antineoplastic, and immunomodulating agent, and is often used in combination with chemotherapy and/or mitotic inhibitors (taxanes). Therapeutic indications include metastatic colorectal cancer, non-squamous non-small cell lung cancer, metastatic HER2-negative breast cancer, glioblastoma, and metastatic renal cell carcinoma (<http://www.avastin.com/avastin/patient/>). A rather new indication for Avastin® has been identified in ophthalmology⁶²¹: Promising treatment results were obtained with the intravitreal off-label use of Avastin® in patients with neovascular Age-related Macular Degeneration (AMD). Macular oedemas are usually treated with the steroid triamcinolone acetonide, but with an unpredictable treatment success. Inhibition of VEGF could be more effectively, safely, and cheaply achieved with Avastin®. Initially, Avastin® was believed to be unable to cross the retina, leading to the creation of the Fab analogue of Avastin®, Lucentis® (Ranibizumab). Vision improvements are similar in both Avastin® and Lucentis®, but intravitreal injections with Lucentis® are up to 20 times more expensive.

Reports of Rosenfeld et al.⁶²²⁻⁶²⁴ on the off-label use of bevacizumab for the treatment of AMD led the basis for the present work. They noted that 1.0 mg of intravitreal bevacizumab resulted in the stabilization of visual acuity and resolution in subretinal fluid. With respect to intravitreal combination therapy, several clinical studies investigated a possible synergy by combining bevacizumab with dexamethasone⁶²⁵ or triamcinolone acetonide^{626, 627}, both of which are well-known anti-inflammatory steroid drugs. The influence of dexamethasone disodium phosphate or triamcinolone acetonide on antibody stability was studied monitoring the monomeric state of the antibody.

The starting points of the here presented project are summarized in Table 15.

Table 15: Summary of the starting points.

Drug	Monoclonal antibody (mAb)	Effect(s)
None	Ranibizumab (Fabs only)	Stable, almost no aggregation
None	Bevacizumab (full-length mAb)	Aggregation
Dexamethasone phosphate		Inhibition of bevacizumab aggregation
Triamcinolone acetonide		No effect on bevacizumab aggregation

The different behaviour of ranibizumab and bevacizumab with respect to aggregation suggests that the Fc region is important. The starting points further led to the question why dexamethasone phosphate “stabilizes” bevacizumab while triamcinolone acetonide does not. Thus, the following working hypothesis was set: The mechanism by which dexamethasone phosphate inhibits bevacizumab aggregation is linked to the Fc region and driven by the presence of the phosphate group on dexamethasone phosphate.

2.2 METHODS

2.2.1 ASYMMETRIC FIELD-FLOW FRACTIONATION COUPLED TO MULTI-ANGLE LIGHT SCATTERING (AFFF-MALS)

To precisely quantify and distinguish different monoclonal antibody species (monomers, dimers and higher order aggregates), Asymmetric Field-Flow Fractionation coupled to Multi-Angle Light Scattering (AFFF-MALS) was used. Fraunhofer and Winter give an overview of principles and applications of field-flow fractionation⁶²⁸.

All aggregation assessments in the presence or absence of steroids were carried out as described in the following for dexamethasone phosphate. The aggregation of bevacizumab was investigated by AFFF with and without dexamethasone phosphate. The molecular weight of each monomer and the differences were determined together with the percentage of monomers with and without dexamethasone phosphate. The isoelectric point of bevacizumab was around pH 8.0 and measurements were carried out at pH 7.0. After 5 weeks, the monomer stabilization at three different temperatures was compared to the control.

The percentages of monomers of the commercial formulation of bevacizumab (Avastin®, Roche Pharma, Reinach, Switzerland) were obtained based on the following protocol: Avastin® was dialyzed overnight into isotonic buffers to remove all excipients present in the commercial product and to change the pH. A 50 mM phosphate buffer pH 7.0 was used. The buffer choice was based on a pH range and a buffer capacity that are physiologically tolerated and acceptable for the stability of antibodies. After dialysis, the bevacizumab preparation with a concentration of 25 mg/mL was stored for 7 days at 40°C at pH 7.0 to stress the antibody and to induce the formation of aggregates. A first sample of bevacizumab was separated (to test the aggregation of bevacizumab only). To another sample of bevacizumab at a concentration of 19 mg/mL, AMP powder (purity 99%, Acros Organics) was added in three different concentrations, obtaining the following molar ratios: (i) bevacizumab: AMP 1:153, (ii) bevacizumab: AMP 1:15.3 and (iii) bevacizumab: AMP 1:1.53. All samples were stored at 40°C for 28 days. Samples were analyzed directly after preparation (to) and after 7, 14, and 28 days. The aggregation state of the antibodies was measured by MALS after separation by AFFF. The concentration of bevacizumab was determined by UV (ultra-violet) spectroscopy at 280 nm, based upon an extinction coefficient of 1.7 cm²/mg. Data were collected and analysed with the Astra software (Wyatt Technology Europe GmbH, Dernbach, Germany). The aggregation state was expressed as the percentage of monomers versus time. Further control experiments on the stability of bevacizumab alone were carried out: Concentration effect (5, 10, 18, and 25 mg/mL in 50 mM phosphate buffer pH 6.2), pH effect (pH 5.0, pH 6.2, and pH 7.0) and storage temperature (4°C, 25°C, and 40°C for 35 days).

Furthermore, the short-term (over 24 h) protective effect of AMP on the commercial product Avastin® was investigated using AFFF-MALS as well. Avastin® was stored for 24h at 40°C before adding AMP in three different ratios (1:1, 1:10, and 1:100).

The experimental part of the work has been performed by M. Veurink and will be presented in detail in her thesis.

2.2.2 MOLECULAR MODELING

The molecular modelling study comprises seven steps:

- 1) The construction of the 3D monomer homology model of bevacizumab,
- 2) the generation of the 3D (dimer) aggregation model,
- 3) molecular docking of dexamethasone phosphate and triamcinolone acetonide on the bevacizumab monomer to explore the molecular recognition pattern characterizing the small molecule – monoclonal antibody interaction,
- 4) the comparison with the bevacizumab dimer to find a putative mechanism of action for differentiating between dexamethasone phosphate and triamcinolone acetonide,

- 5) extending this mechanism to other steroids,
- 6) ligand-based screening for known drugs with a similar mechanism using the program ROCS, and
- 7) molecular docking for assessing the aggregation breaking potential of the compounds selected in the previous step.

Before step 3, the experimental results for the two docked steroids were already known. After step 5, some other steroids were tested by AFFF-MALS to confirm the predictions made by docking. Following step 7, a selection of the suggested compounds underwent experimental validation.

2.2.2.1 CONSTRUCTING THE 3D MONOMER HOMOLGY MODEL OF BEVACIZUMAB

Comparative modelling or homology modelling uses experimentally determined protein structures to predict the conformation of other proteins with similar amino acid sequences⁶²⁹, based on the finding that small sequence changes normally result in small 3D structure changes^{630, 631}. It is the only approach by which the 3D protein structure can be predicted with an accuracy comparable to that of low-resolution experimental structures⁶³². Based on the relation between the divergence of sequence and structure in proteins⁶³³, a protein structure is predicted as target based on a sequentially related, “homologous” known 3D structure (template)⁶³⁴. When the protein of interest shares at least 30% sequence identity with a template, they are assumed to have a similar 3D structure.

Four steps can be discerned in homology modelling: (i) template identification, (ii) alignment, (iii) model building, and (iv) model validation. Step (ii) is the most important and is achieved with sequence alignment algorithms featuring a scoring matrix and penalties for introducing gaps into the alignment, which have to be tested on a case-by-case basis⁶³⁵. The interconnected steps of homology modelling are depicted in a schematic way as gears in Figure 24.

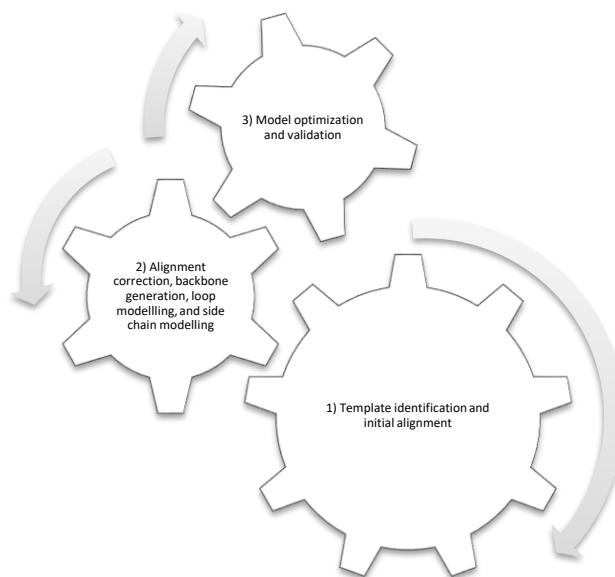


Figure 24: The interconnected steps of homology modelling depicted as gears

Template identification and alignment

Bevacizumab is a humanized antibody formed by a Fab region responsible for its activity and an Fc region derived from IgG1. The Protein Data Bank (PDB)^{10, 11} (<http://www.rcsb.org>) was searched for suitable crystal structures (as of the 17th of April 2010) to be used as templates. Three structures were

found: The bevacizumab Fab moieties in complex with the human VEGF⁶³⁶ (PDB ID: 1BJ1) and two full length antibodies, one being human (PDB ID: 1HZH)⁶³⁷ and one being murine (PDB ID: 1IGY). Because bevacizumab is a humanized antibody, the human antibody should be taken as template. However, the only full-length human IgG1 crystallized so far (PDB ID: 1HZH), has an unusual crystal symmetry (H3₂) and strongly distorted orientations of the Fabs with respect to the axis of the Fc region. In contrast, the intact murine IgG1k is a template with a more usual overall 3D structure, resembling a distorted Y. This shape is due to an angle between the two Fab axes of 115°. The resolution of this structure is with 3.2 Å acceptable. To humanize its hinge-Fc region, the residues were mutated so as to match the sequence of the structure 1HZH. All sequence alignments were carried out with ClustalW^{27,28} in the BiologyWorkbench 3.2 (<http://seqtool.sdsc.edu/CGI/BW.cgi>).

3D model building

The initial 3D model of bevacizumab resulted from the following modelling steps that were carried out in Sybyl 8.0:

- 1) Structural superposition of the Fabs of IgG1 (1IGY) with the ones of bevacizumab (crystallized with VEGF, 1BJ1),
- 2) replacement of the IgG1 Fabs by the ones of bevacizumab,
- 3) connection the Fabs of bevacizumab with the hinge-Fc of IgG1 (1IGY), and
- 4) humanization of the hinge-Fc region using the sequence of 1HZH.

The full process is summarized in Figure 25.

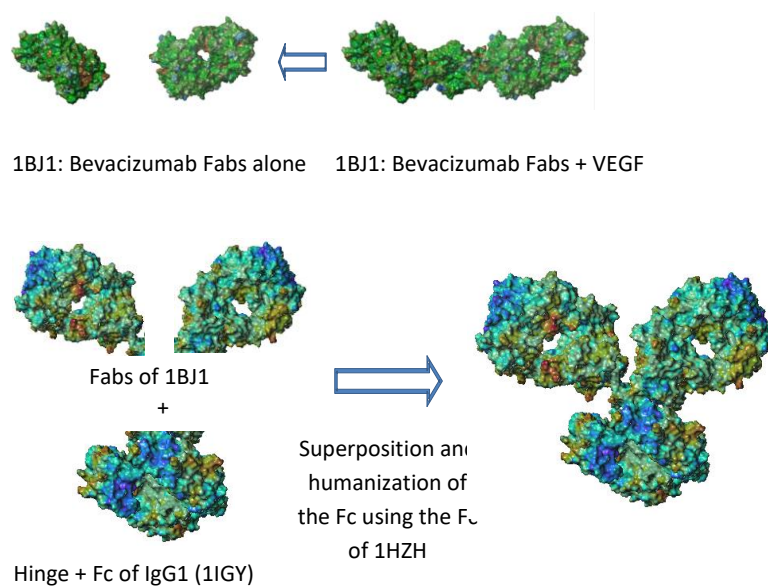


Figure 25: Building the initial 3D model. The two Fabs of the bevacizumab crystallized with the vascular endothelial growth factor VEGF (1BJ1) are superposed onto the full IgG1 antibody (1IGY). The Fabs are replaced by the Fabs of bevacizumab and linked to the hinge and the Fc region of 1IGY. The Fc region is then humanized using the sequence of the human antibody 1HZH.

The connecting region of the model was submitted to energy minimization using Sybyl 8.0 default parameters and keeping the disulfide bridges intact.

Model validation

The quality of the resulting model was assessed using Procheck. Upon Ramachandran plot analysis (in Procheck) of the amino acid conformations using a resolution mean between the crystal structures of both the Fc (resolution of IgG1: 3.2 Å) and the Fabs (resolution of 1BJ1: 2.4 Å), critical side chains were corrected for distortion in Sybyl 8.0 and the refinement procedure (manual adjustments and minimization) repeated until reaching conformations comparable to the input crystal structures.

2.2.2.2 GENERATING THE 3D (DIMER) AGGREGATION MODEL

A 3D aggregation model was built to identify the bevacizumab-bevacizumab interface. The basic idea was that crystal contacts represent privileged protein-protein interfaces. Based on the crystal symmetry of the IgG1 structure, from which the Fc was taken for the bevacizumab model, putative crystal contacts between two full length bevacizumabs were defined using SwissPDBViewer 4.0.1²⁹. The crystallographic symmetry ($P2_1$) of the IgG1 crystal structure was applied to obtain a layer translated along the unit cell and recreate crystal contacts. The first translation as the only close crystal contact being in the order of 4 Å was saved in pdb format and the bevacizumab model overlaid according to the C α atom positions in Sybyl 8.0.

2.2.2.3 MOLECULAR DOCKING

Docking dexamethasone phosphate and triamcinolone acetonide onto the bevacizumab monomer

Dexamethasone phosphate and triamcinolone acetonide were docked onto the bevacizumab monomer to see where they were binding preferentially, whether they were binding to the aggregation zone and, if so, how the observed binding propensity could be related to properties of the small molecule and the binding site, respectively. In general words, docking aims at predicting the binding mode of a ligand in a target by minimizing a scoring function accounting for the interaction between a protein and a ligand. In this thesis, docking is either used i) for understanding a protein-ligand binding mechanism, ii) for finding new compounds or iii) for finding potential targets. Approach i) is applied in this chapter.

The small molecule setup for docking was done in Sybyl 8.0. Hydrogens were added to the small molecular weight molecules, each phosphate was left unprotonated and Gasteiger-Hückel charges were added. The resulting molecules were minimized using 100 steps of the default Powell minimization protocol of Sybyl 8.0. Systematic docking with standard parameters was then performed with FlexX 3.1.3 (Biosolveit GmbH) all over the bevacizumab antibody surface (monomer: Fabs, hinge and Fc), previously divided into several segments 10 Å around each Arg and Lys. 10 poses per molecule and docking site were kept, as this number revealed to be sufficient to distinguish aggregation breakers from non-aggregation breakers. The docking poses were scored with the FlexX scoring function and evaluation was carried out by analyzing the attributed bevacizumab-small molecule interaction scores, visually inspecting all poses and retaining the ones projecting from the bevacizumab surface.

Molecular docking of the drugs retained in the ligand-based screening

The molecules resulting from the ligand-based similarity search, one analogue of these molecules, and commercially available mono- and disaccharides were docked using the same protocol, but by restricting docking to the region involved in dimerisation.

Comparison with the bevacizumab dimer

The number of docking poses successful in interfering with the bevacizumab-bevacizumab interaction interface (out of 10 solutions) was a key selection/analysis criterion for the small molecular weight compounds. Molecules with more than half of their solutions fulfilling the protein-protein interference criterion were considered to be included in experimental testing.

2.2.2.4 MECHANISM OF ACTION AND ITS VALIDATION

Based on the binding propensities of both molecules at the bevacizumab-bevacizumab interface and the physicochemical properties of both the ligands and the binding site on one monomer, a putative mechanism of action was postulated.

This mechanism was subsequently verified by docking other steroids, in order to find out whether they bind in the antibody-antibody contact region. For the exhaustive docking all over the antibody surface, the docking parameters were the same as described in 2.2.2.3. All the docked steroids are summarized in Table 16 and include true binders (experimentally confirmed, positive controls), a putative non-binder, and non-binding (experimentally confirmed, negative controls) compounds.

2.2.2.5 LIGAND-BASED SCREENING: SIMILARITY SEARCH/SCAFFOLD HOPPING

With the aim of finding a real adjuvant able to break the primary event of bevacizumab dimerisation, other low molecular weight compounds potentially also having an aggregation inhibiting effect on bevacizumab were identified by a similarity search/scaffold hopping approach. A similarity search usually relies on similar physicochemical properties (pharmacophores) or shape. Scaffold hopping deals with finding isofunctional, but structurally dissimilar molecular entities³⁰⁻³³. Ideal screening methods that perform successful scaffold hops would not only find a maximum number, but also a maximally diverse set of active compounds from a given chemical structure³⁴.

ROCS 2.4.2 (OpenEye Scientific Software, Windows version) was used for fast shape comparisons, with the assumption that molecules do have similar shape if their volumes overlay well. Any volume mismatch is treated as a measure of dissimilarity.

Similarity searches were done using one conformer of betamethasone phosphate, dexamethasone phosphate and triamcinolone phosphate each. A multi mol2 file containing the three molecules was defined as query in ROCS, together with a sd file containing the small molecules. Almost 4800 FDA-approved DrugBank molecules (downloaded on the 5th of November 2009) were used as database. Results were visualized and browsed in Vida 4.0.0 (OpenEye Scientific Software) for negatively charged groups accessible for binding to Lys445 of the bevacizumab monomer model. The selection criteria were the following: (i) at least two negative charges (preferentially a phosphate group), (ii) none of them included in a ring, (iii) high ranks (out of 500), and (iv) not only being composed of one ring and a negatively charged entity. Planar molecules meeting these criteria were discarded, as they would not have been able to properly adapt to the bevacizumab surface around Lys 445. A second round of similarity searches was performed with one hit on the same input database.

Furthermore, the phosphorus-containing subset of the ZINC database was screened with the three above mentioned steroid phosphates and one similarity search hit from the DrugBank screen. This online database is openly accessible at <http://zinc.docking.org> and contains commercially available compounds for virtual screening³⁵.

2.2.3 MARKUSH STRUCTURE REPRESENTATION

To adequately describe and protect the compounds in two patents on the findings resulting from this work, the common Markush structure representation was used, which is an enumeration of possible structures based on different substitutions. Substitutions may comprise substituent variations, positional variations (e.g. ortho, meta, para), frequency variations (n, multiple occurrences), homology variations (generically described groups, e.g. alkyl 1-4 C) or atom variations. Named after Eugene Markush³⁶, Markush structures represent a scaffold as most stringent substructure query of a set of compounds³⁷. The total number of molecules represented in a Markush structure is given by: $(L * F)^S$, where L is the number of linkers, F the number of functional groups, and S the number of substitutions on the Markush structure.

2.3 RESULTS AND DISCUSSION

2.3.1 *IN VITRO* STARTING POINTS FOR THE MODELLING STUDY

The main *in vitro* findings motivating this modelling study were that dexamethasone phosphate stabilizes bevacizumab, while triamcinolone acetonide does not, and that the bevacizumab-bevacizumab aggregation seems to be dependent on the Fc region of bevacizumab.

Figure 26a shows that dexamethasone disodium phosphate “stabilizes” bevacizumab *in vitro* at pH 7.0 over a broad temperature range, with the most pronounced effect at 40°C. In contrast, as can be seen from Figure 26b, triamcinolone acetonide does not inhibit the aggregation of bevacizumab.

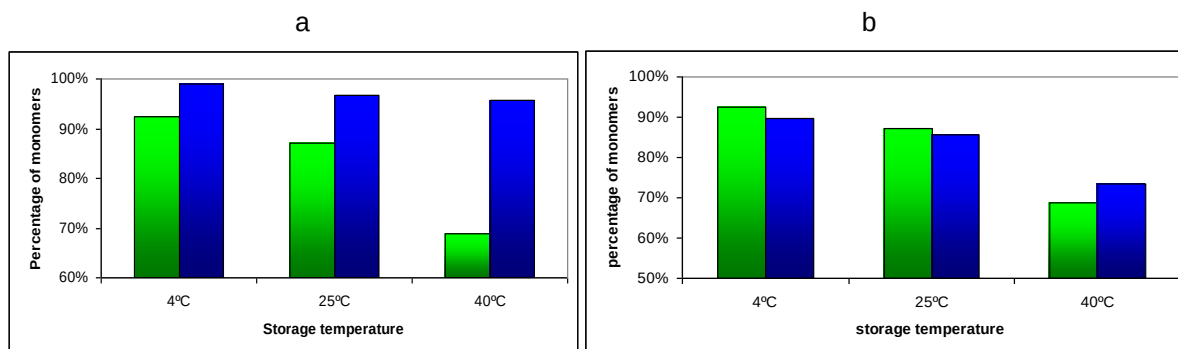


Figure 26: The effect of steroids on bevacizumab stabilization after storage during 35 days at three different temperatures in PBS buffer pH 7.0. a) Stabilizing effect of dexamethasone disodium phosphate with bevacizumab and b) no stabilization of bevacizumab by triamcinolone acetonide. In both Figures, the molar ratio bevacizumab:steroid of 1:153 ($n = 2$) is depicted in blue and bevacizumab only ($n = 2$) in green. n stands for the number of replicas. This Figure was kindly provided by M. Veurink.

Furthermore, the percentages of monomers of the marketed products Lucentis® and Avastin® have been compared. Lucentis® is composed of the Fabs of Avastin®, while Avastin® is a full length intact antibody. In Figure 27, the percentage of monomers for Lucentis® decreases by 1% over 35 days, while 10% are lost for Avastin® over the same period of time. Therefore, aggregation is not an issue for Lucentis®. In contrast, it clearly is an issue for Avastin®, which possesses an Fc region contrarily to Lucentis®.

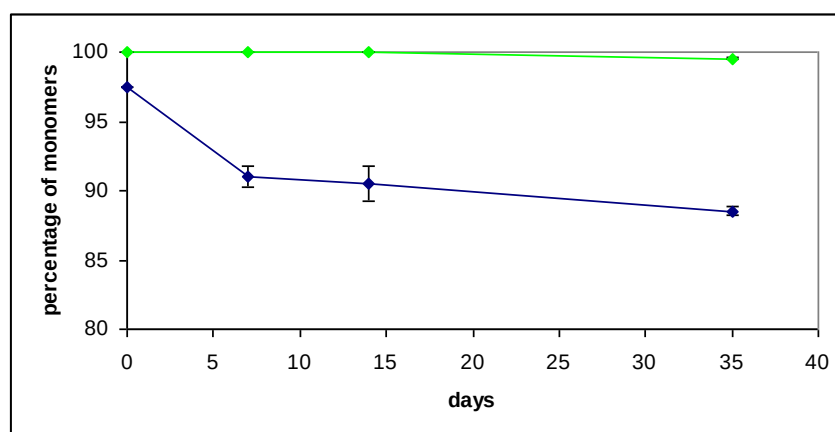


Figure 27: Comparison of the time course of the percentage of monomers of the two commercial formulations of Avastin® (dark blue) and Lucentis® (light green), $n=3$ (the mean is represented by “diamonds” and the standard deviations by bars), storage at 40°C. This Figure was kindly provided by M. Veurink.

2.3.2 BEVACIZUMAB MONOMER AND AGGREGATION MODELS

The sequence alignments used for creating the Fabs of the monomer model revealed that the IgG1 template Fabs are homologous to the bevacizumab Fabs (Figure 28 for the light chain alignment and Figure 29 for the heavy chain alignment). For the light chain alignment, a sequence identity of 58% and a homology of 86% were obtained, whereas the identity was 53% and the homology 80% in the heavy chain, respectively. The well distributed conservation facilitated the modelling.

```

KCAHTVS. KSMMSVGERVTLTCKASENVVTYVS WYQQKPEQS PKLLIYGASNRY 1IGY_A_PDB
DIQMTQSPSSLSASVGRVTITCSASQDISNYLNWYQQKPKGAPKVLIVFTSSLH 1BJ1_J_PDB
-1-mT-Sp-SIS-SVGdRVTiTC-AS--i--YI-WYQQKP---PKvLIY--S--y consensus

TGVPRDFTGSGSATDFTLTISSTQAEDLADYHCQGQGYSPYTFGGGTKLEIKRAD 1IGY_A_PDB
SGVPSRFSGSGSGTDFTLTISSTQPEDFATVYCCQYSTVPWTFGGGTKVEIKRTV 1BJ1_J_PDB
sGVP-RFSGSGSGTDFTLTISSTQaED-A-YyC-Q--t-PwTFG-GTKvEIKR-- consensus

AAPVVSIFPPSSEQLTSGGASVVCFLNNFYPKDINVKKKIDGSRQNGVLNSWTD 1IGY_A_PDB
AAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTE 1BJ1_J_PDB
AAPsV-IFPPS-EQL-SG-ASVVC-LNNFYPre--V-WKvD-----S-Te consensus

QDSKDYSTYSSTLTTLTKDEYERHNSYTCETHKTSSTSPVVKSFNRNEC 1IGY_A_PDB
QDSKDYSTYSSTLTTLTKADYERHKVYACEVTHQGLSSPVTKSFNRGEC 1BJ1_J_PDB
QDSKDYSTYSSTLTTLTK-dYEkH--Y-CE-TH---sSPv-KSFNR-EC consensus

```

Figure 28: Sequence alignment of the light chains of both IgG1 (PDB ID: 1IGY, chain A) and bevacizumab (PDB ID: 1BJ1, chain J) performed with the ClustalW tool^{27,28} of the BiologyWorkbench and coloured with TexShade³⁸. Complete residue matches are highlighted on a black, similar residues on a light grey, and non-matching residues on a white background. The consensus sequence for a full match (upper case letters) and a conserved amino acid property (lower case letters) between the two compared sequences are indicated as well.

```

VKIQESGAELARPGASVKMSCKASGYTFTTYYTIHWIKRPGQGLEWIGYINPSS 1IGY_B_PDB
EVQLVESGGGLVQPGGSLRLSCAASGYTFTTNYGMNWVRQAPCKGLEWVGWINTYT 1BJ1_K_PDB
eV-L-ESGg-L--PGgSlrlSC-ASGYTFT-Y-mnWvrQ-PG-GLEWvGwInt-t consensus

VYTNVYQRFKDKATLTDRSSNTANIHLSSLTSDSAVYYCVRE...GE...V 1IGY_B_PDB
GEPTVAADFRRRTFSLDTSKSTAYLQMNSLRAEDTAVYYCAKYPHYGSSHWYF 1BJ1_K_PDB
--t-Y---FK-r-T-s-D-S--TA-lqm-SL--eDtAVYYC-k-phyyG-shwy- consensus

PYWGQGTITVTVSSAKTTPPSVYPLAPGSAAQTNMVTLGCLVKGYFPEPVTVSWN 1IGY_B_PDB
DVWGQGTITVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWN 1BJ1_K_PDB
d-WGQGT-ITVTVSSA-T-gPSVfPLAP-S---s-----LGCLVK-YFPEPVTVsWN consensus

SGSLSGGVHTFPAVLQSSGLYSLSSVTVPSSTPRPSETVTCNVAHPASSTKVDKK 1IGY_B_PDB
SGALTSGVHTFPAVLQSSGLYSLSSVTVPSSTLGTQTIYICNVNHKPSNTKVDKK 1BJ1_K_PDB
SG-LtSGVHTFPAVLQSSs-LYsLSS-VTVPSSTs-gt-T--CNV-HkaS-TKVDKK consensus

IVPRDGGCKPCICTVPEVSSVFIFPPKPKDILLITVTPKVTGVVVDISKDDPEVQ 1IGY_B_PDB
VEPKSCDKHTHT 1BJ1_K_PDB
v-Pk-C-k--tictvpevssvfifppkpkdtllitvtpkvtcgvvvdiskddpevq consensus

```

Beginning of hinge and Fc region

Figure 29: Sequence alignment of the heavy chains of both IgG1 (PDB ID: 1IGY, chain B) and bevacizumab (PDB ID: 1BJ1, chain K) using the same programs and colour codes as in Figure 28. Inserts are depicted with a dark grey background.

Similarly, the murine Fc (PDB ID: 1IGY) had a sequence identity of 58% and a homology of 73% with the human Fc of IgG1.

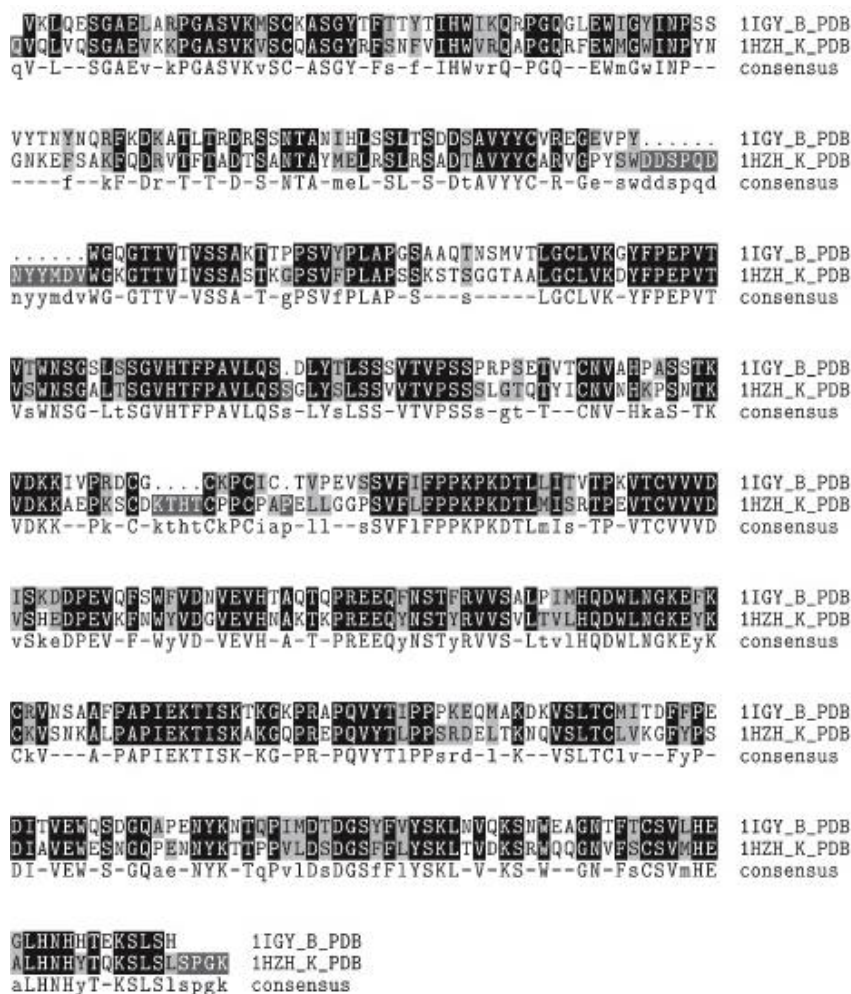


Figure 30: Sequence alignment of the Fc region of the murine (PDB ID: 1IGY, chain B) and human IgG1 (PDB ID: 1HZH, chain K) using the same programs and colour codes as in Figure 28.

The final monomer model consisting of the bevacizumab Fabs and the humanized hinge-Fc had Ramachandran plot values comparable to the input template structures: 74.9% of the residues in the model were found in most favoured regions of the Ramachandran plot, 21.4% in additionally allowed regions, 2.2% in generously allowed regions, and 1.5% in disallowed areas. The aggregation model, obtained by replicating the monomer model according to the crystal symmetry (P21) of the murine IgG1 (PDB ID: 1IGY), is depicted with an electrostatic potential surface in Figure 31a. Notably, one close contact area (least distance in the order of 4 Å) between the two monomers can be discerned. Zooms on this contact region are provided in Figure 31b and Figure 31c.

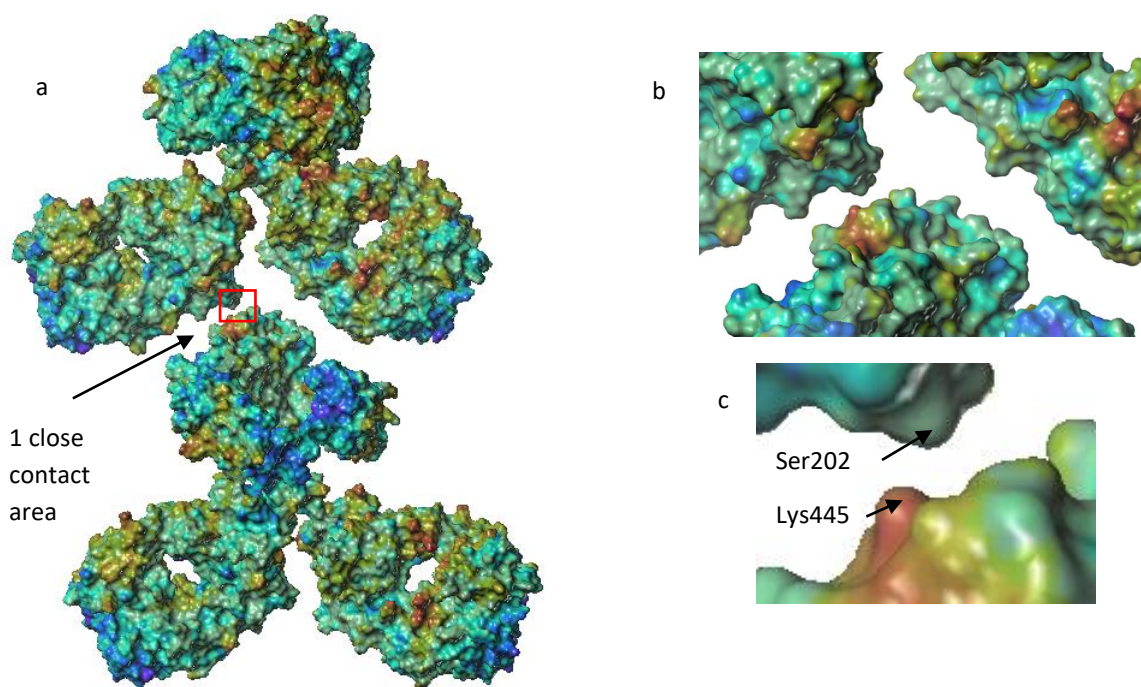


Figure 31: The 3D aggregation model (a) of bevacizumab and two zooms on the contact region (b and c). The aggregation model was obtained by replicating the monomer model according to the crystal symmetry ($P2_1$) of the template IgG1. In the closest zoom depicted in (c), the least distance in the contact region between Lys445 and Ser202 is in the order of 4 Å. The colouring gradient of the electrostatic potential surface goes from blue for negatively charged to red for positively charged atoms.

Upon displaying the bevacizumab model and its translation, the bottom of one Fc comes to lie in between the two Fab regions of the other bevacizumab. A slight distortion is noted with respect to the orientation of both Fcs, which comes from the crystal structure of IgG1.

Figure 31c shows that the contact region is around the only close crystal contact represented by the interaction of Ser202, belonging to one Fab arm of one bevacizumab, with Lys445, which is part of the Fc of the second bevacizumab. The contact points Lys445 and Ser202 are fully conserved in IgG1s. The amino acids lining the contact area include for the Fab region His198 to Pro204 and for the Fc region Met381 to Val386 and Val443 to Glu449.

Comparing the murine (PDB ID: 1IGY) with the human Fc (PDB ID: 1HZH), the residues surrounding and including Lys445 are conserved (Figure 31a, b, and c). The non-conserved amino acids are not involved in intra-antibody interactions and are thus believed to affect the prediction outcome only marginally for potential aggregation breakers.

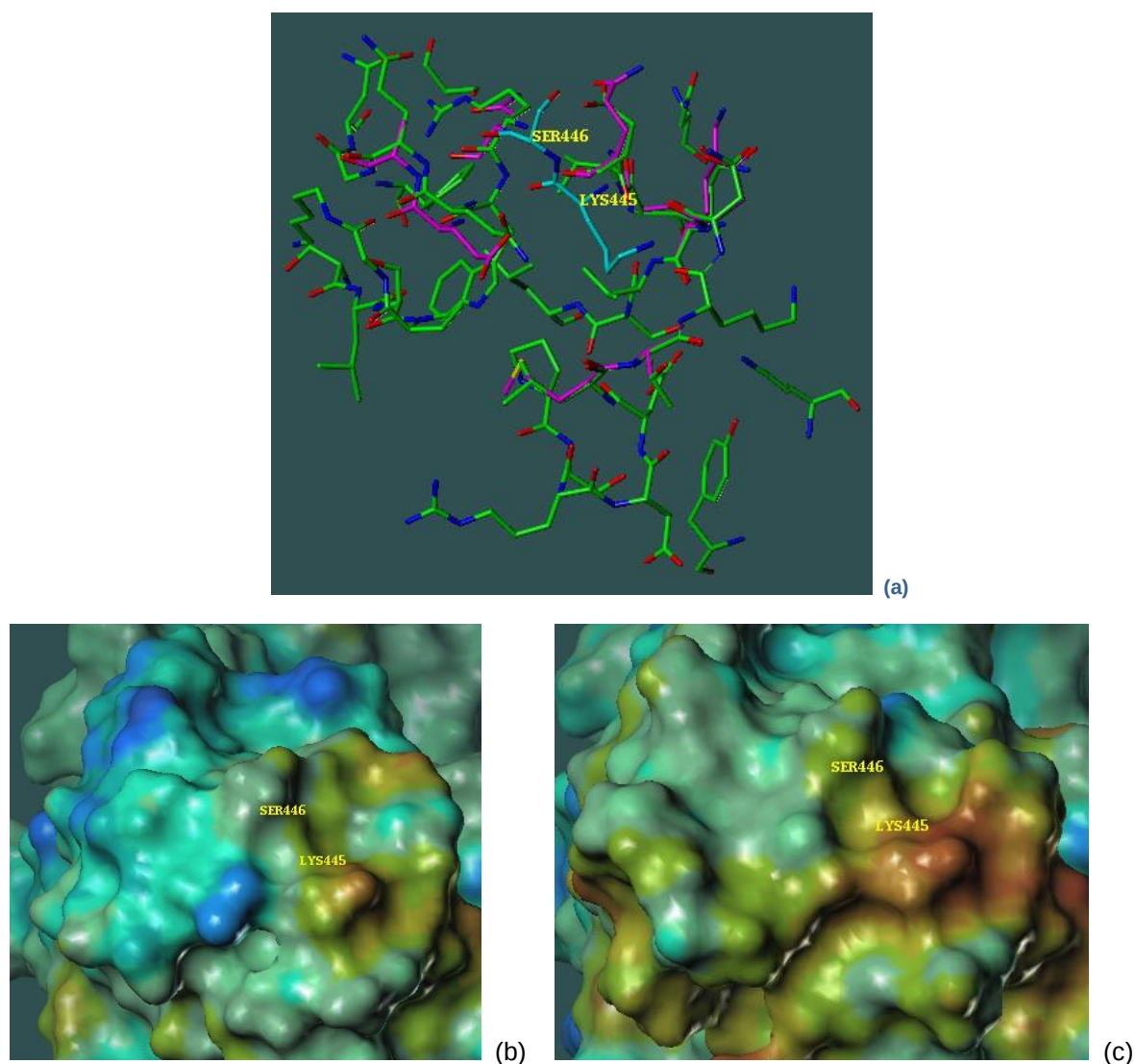


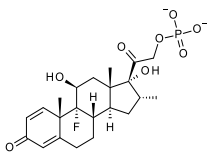
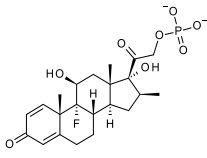
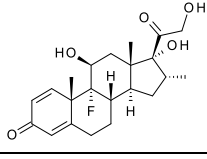
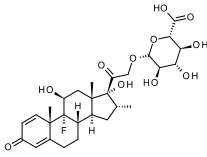
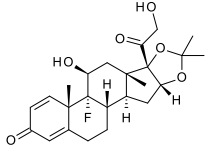
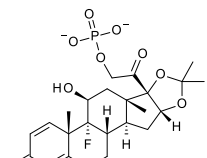
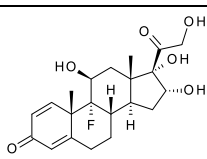
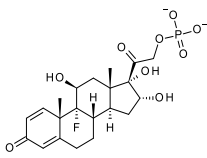
Figure 32: Superposition of the Fc aggregation zone of the mouse IgG1 (PDB ID: 1IGY) with the corresponding human IgG1 (PDB ID: 1HZH). (a) Amino acid view of the superposition. Both Lys445 and the nearby Ser446 (shown with cyan carbon atoms) are conserved and mainly surrounded by conserved amino acids. The amino acids displayed with green carbons correspond to the humanized bevacizumab model. Only a few residues (shown with magenta carbons) are different in the murine crystal structure. The electrostatic surface of the Fc crystal contact region around Lys445 as found in the murine Fc is drawn in (b) and the human Fc in (c). (c) was used as the Fc of the bevacizumab model.

2.3.3 DOCKING AND SIMILARITY SEARCH RESULTS

The following Tables summarize the docking results for the steroids considered (Table 16), the potential aggregation breakers identified by docking of similarity search hits (Table 17), and the mono- and disaccharide phosphates (Table 18).

Concerning the docking of the steroids (Table 16) all over the monomer antibody model, highest docking scores (reflecting the antibody-small molecule interaction energy) were obtained for phosphate-containing steroids binding to Lys445 at the aggregation interface. Interactions with other Lys were less favourable, and the docking poses to different Arg received even lower scores.

Table 16: Steroids considered. n/a = not available. * could only be tested as Kenacort formulation.

	Common name	Structure	Reason for inclusion	Predicted in this work to inhibit aggregation	Experimentally inhibiting aggregation
Dexamethasone series	Dexamethasone phosphate		Known positive control and experimentally confirmed aggregation breaker	yes	yes
	Betamethasone phosphate		Predicted and experimentally confirmed aggregation breaker	yes	yes
	Dexamethasone		Predicted non-breakers	no	n/a; not soluble
	Dexamethasone glucuronide			no	n/a; soluble
Triamcinolone series	Triamcinolone acetonide (Kenacort*)		Known negative control and experimentally confirmed non-breaker	no	no
	Triamcinolone acetonide phosphate (Kenacort soluble)		Predicted and experimentally confirmed non-breaker	no	no
	Triamcinolone		Predicted non-breaker	no	n/a
	Triamcinolone phosphate		Predicted breaker	yes	n/a; not commercially available

Summarizing Table 16, with respect to the steroids investigated, the *in silico* predictions match the available *in vitro* results. In the following, some representative Figures of four steroid dockings are shown.

For dexamethasone phosphate, an experimentally confirmed aggregation breaker (2.2.1), a strong electrostatic interaction between the phosphate of the ligand and the Lys 445 side chain is noted (Figure 33a).

To provide an order of the strength of this interaction, an exposed salt bridge is known to be as strong as a neutral H-bond (5 ± 1 kJ/mol³⁹)⁴⁰. Additional H-bonds between residues adjacent to Lys445 and the steroid phosphate further stabilize it in the binding pocket. The moiety of dexamethasone phosphate clashing with the second monomer in the aggregation model, indicated schematically in Figure 33a, is clearly visible in Figure 33c. Figure 33b shows how dexamethasone phosphate is positioned on the monomer.

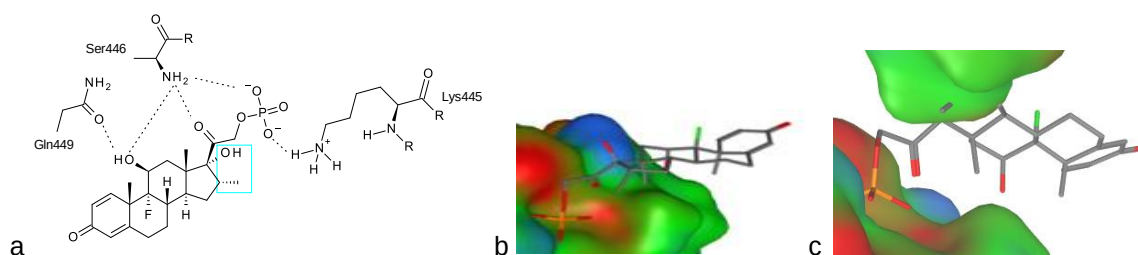


Figure 33: Representative docking pose of dexamethasone phosphate docked onto the monomer. For the sake of better visibility, the most representative, but not necessarily the best-scored out of the 10 docking poses is shown. For each docked compound, first its interactions with the bevacizumab monomer are drawn in a schematic way (a), where the dotted lines represent H-bonds and the cyan box the parts of the steroid that clash with the second monomer of the aggregation model (c). Second, the same docking pose is shown as docked onto the monomer Fc region (b). Third, still the same pose is displayed in the aggregation model (c). In both (b) and (c), the carbon atoms of the steroid are drawn in grey and the electrostatic surface gradient goes from red for positively charged to blue for negatively charged atoms (MOE 2010, Chemical Computing Group).

These findings contrast with triamcinolone acetone, which does not inhibit aggregation experimentally. Indeed, this molecule did not adopt the protein-protein interaction interfering orientation upon docking and the most populated one is neither located at the interface nor binding to Lys445 (Figure 34a and b).

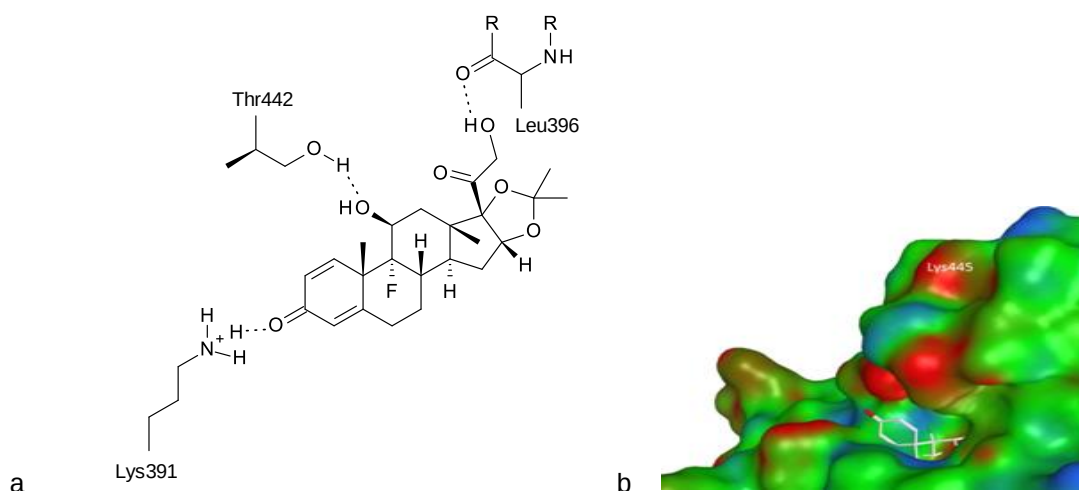


Figure 34: Representative docking pose of triamcinolone acetone docked onto the monomer. Both the schematic interaction drawing (a) and the display on the monomer (b) show that triamcinolone acetone (white carbon atoms) binds to a pocket far away from Lys445.

Given these *in silico* results, the question why dexamethasone phosphate inhibits aggregation (stabilizes the monomer of bevacizumab), while triamcinolone acetonide does not, can be answered. Dexamethasone phosphate is able to interact with Lys445 via its phosphate, masking the interaction interface between the two monomers and thus hindering the dimer formation. Triamcinolone acetonide in turn lacks the phosphate important for anchoring, adopting many putative binding orientations that do not interfere with the protein-protein interaction interface.

To see whether the chosen docking protocol can distinguish aggregation breakers from non-breakers in a predictive way, further steroids were docked, predictions made, and the compounds experimentally tested (Table 16).

Betametasone phosphate was predicted to be a bevacizumab-bevacizumab interaction breaker. This statement was based on the docking onto the 3D monomer, the interaction scores obtained, and the binding propensity of the 10 poses in the aggregation model.

Figure 35c unambiguously illustrates that betamethasone phosphate is clashing into the surface of the upper antibody. Comparing the binding mode of the two aggregation breakers dexamethasone phosphate and betamethasone phosphate, both phosphate groups interact with Ser446 and clashing with Ser202 occurs via the same small molecule part.

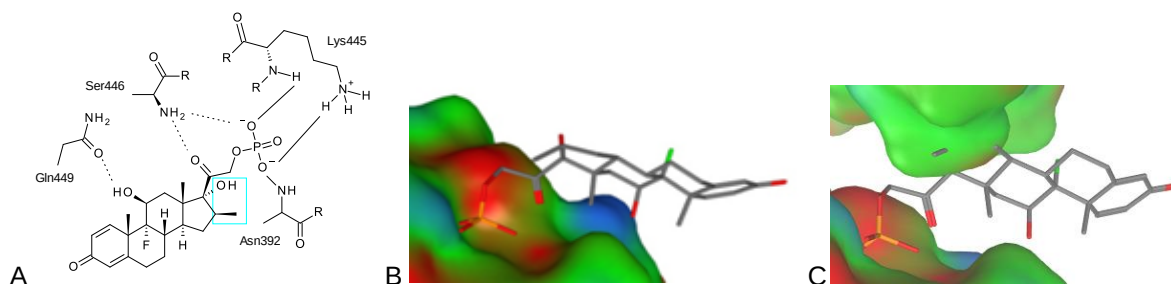


Figure 35: Representative docking pose of betamethasone phosphate docked onto the monomer. Its orientation is remarkably similar to the one of dexamethasone phosphate with respect to H-bonds and the moiety interfering with the second antibody in the aggregation model. (a) shows the schematic interactions with the monomer, (b) the docking onto the monomer, and (c) the display of the docking pose in the aggregation model.

Conversely, triamcinolone acetonide phosphate is predicted not to be a bevacizumab-bevacizumab interaction breaker (Figure 36b).

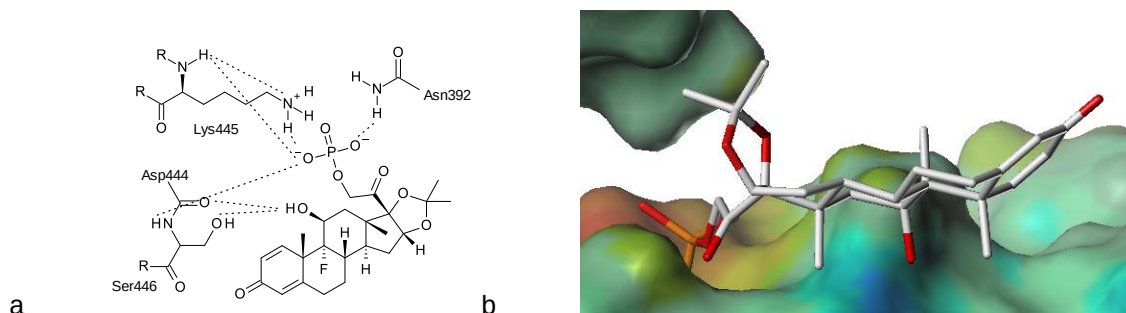


Figure 36: Representative docking solution of triamcinolone acetonide phosphate docked onto the monomer. Schematic view of the interactions (a) and display of the docking pose in the aggregation model (b).

Even though triamcinolone acetonide phosphate features a phosphate group, the acetonide moiety restricts its conformational flexibility. Lys445 is still able to interact with the phosphate, but the steroid part is located parallel to the Fc surface (Figure 36b). Furthermore, the acetonide is pointing towards the other monomer, but is not voluminous enough to interfere with it.

In general, compounds predicted to inhibit dimer formation mostly docked to Lys445 and Ser446 and were positioned in such a way as to interfere with the adjacent antibody, while compounds predicted not to be inhibitors of dimer formation did not at all or not frequently adopt such a protein-protein interaction interfering orientation.

Both the crystal contacts and the docking results suggest that aggregation occurs through the interaction of the Fc region with one Fab region of the adjacent bevacizumab. The presence of a phosphate group on the molecule is a prerequisite for inhibiting aggregation, which occurs due to the interaction between the ligand's phosphate group and a Lys in the Fc region of the antibody. A negatively charged functional group binding to the Fc region is favourable because it can form an electrostatic interaction with the (at physiological pH) positively charged terminal amide of Lys445. Compared to other functional groups like carboxyl, phosphate has the advantage of bearing two negative charges. However, a flexible bending of the phosphate with respect to the rest of the molecule is needed. Such a flexibility is e.g. guaranteed in dexamethasone phosphate, but impossible in triamcinolone acetonide phosphate due to the acetonide moiety.

Using betamethasone phosphate or dexamethasone phosphate as queries for the ligand-based screening on the DrugBank screening, identical hit lists in terms of Tanimoto coefficients were obtained. Similar, but not identical coefficients were found for triamcinolone phosphate as a query. Thus, a slightly different list of hits was produced by this search. Taking the three queries together, fludarabine, fosphenytoin, AMP, tenofovir and cidofovir were identified as hits. A second screening was performed with the three steroid phosphates on the phosphorus-containing portion of the Zinc database (<http://zinc.docking.org>). The Zinc database search did not yield molecules containing phosphate, but dioxido-oxo-phosphane also bearing two negative charges on oxygens. With the DrugBank hit AMP, the DrugBank was then screened with ROCS, yielding as new hits ATP, tiludronate, amifostine and pyridoxal phosphate in decreasing rank order. In a similar way to the steroids, AMP was also used as input for searching the phosphorus-containing subset of the Zinc database. In analogy to AMP and ATP, ADP was added to all the similarity search hits shown in Table 17.

All of these hits were docked onto the Fc contact region of the monomer. The second monomer of the aggregation model was then displayed and the molecules were selected or rejected based on their propensity of interfering with dimer formation while binding to one monomer. To check whether the sugar-phosphate moiety was sufficient for binding to Lys445 and clashing into the adjacent antibody or if an additional spacer attached to the sugar moiety would be needed, commercially available mono- and disaccharides were considered as well (Table 18).

Among all the molecules shown in Tables 17 and 18, the most striking molecule identified was AMP, which is part of many biochemical pathways in the human body and used as food supplement. The FDA classifies it as GRAS (Generally Recognized As Safe). In the series AMP-ADP-ATP, the predicted propensity to inhibit aggregation decreased with the increasing number of phosphates. AMP was foreseen to best inhibit aggregation, which has been experimentally confirmed (2.3.6), followed by ADP. ATP could not be docked in a clashing orientation due to its many degrees of freedom.

Together with the DrugBank molecules fludarabine, tenofovir, cidofovir, tiludronate, and amifostin, the Zinc database compounds depicted in Table 17 constitute a panel of molecules that could be experimentally tested for their aggregation breaking propensity, but which cannot easily be implemented in the formulation of Avastin® antibodies, either because they are drugs with known side-effects or because they are not yet clinically approved (in the case of the Zinc database compounds). One Zinc database compound was particularly interesting because of a double attachment at the Fc region via dioxido-oxo-phosphane and CO_2^- (Table 17). Fosphenytoin and pyridoxal phosphate could be discarded because the former was only seldom clashing, and the latter was rather too small to interfere with aggregation.

To assess the role of the ribose spacer, different mono- and disaccharide phosphates (Table 18) were docked onto the monomer and displayed in the aggregation model. Among the hexoses investigated,

the only potential aggregation breaker is α -D-galactose-1-phosphate, based on the number and orientations of docking solutions. Besides, disaccharide phosphates are candidate inhibitors due to their size and shape. Based on these results, the nutritional supplements AMP, ADP, α -D-galactose-1-phosphate, α -lactose-1-phosphate, and α -D(+)maltose-1-phosphate constitute a promising panel of (potential) aggregation breakers.

Table 17: Potential aggregation breakers identified by docking of similarity search hits

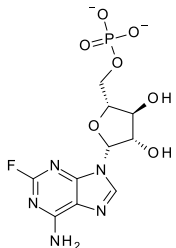
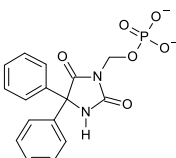
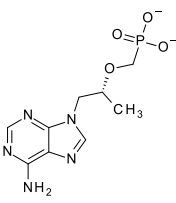
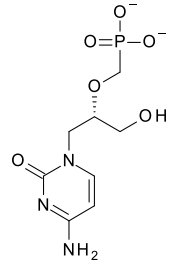
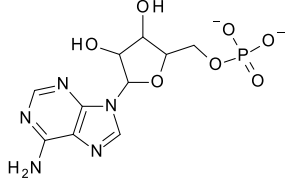
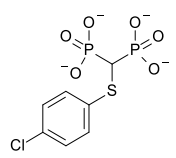
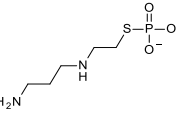
Name	Structure	Query molecule(s)	Prediction based on docking
Fludarabine		Steroid phosphates	Aggregation breaker
Fosphenytoin		Steroids phosphates	Rather not an aggregation breaker, seldom clashing
Tenofovir		Steroid phosphates and AMP	Aggregation breaker
Cidofovir		Steroid phosphates and AMP	Aggregation breaker
AMP		Steroid phosphates	Aggregation breaker
Tiludronate		AMP	Aggregation breaker
Amifostine		AMP	Aggregation breaker without sugar part

Table 17 (continued): Potential aggregation breakers identified by docking of similarity search hits.

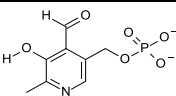
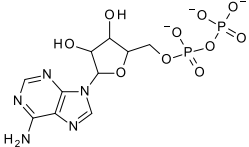
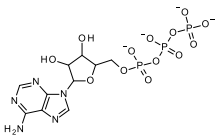
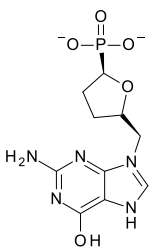
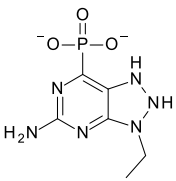
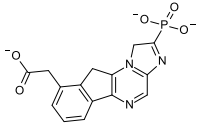
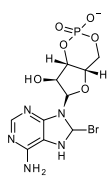
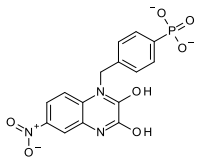
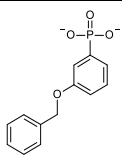
Name	Structure	Query molecule(s)	Prediction based on docking
Pyridoxal phosphate		AMP	Not an aggregation breaker, too small
ADP		Added in analogy to AMP and ATP	Aggregation breaker
ATP		AMP	Too many degrees of freedom in the triphosphate part to be an efficient aggregation breaker
Zinc01903563		Steroid phosphates	Aggregation breaker
Zinc01614527		Steroid phosphates	Aggregation breaker
Zinc01909478		Steroid phosphates	Aggregation breaker with two attachment points (CO ₂ ⁻ and PO ₃ ⁻) on the Fc of bevacizumab
Zinc01692858		Steroid phosphates	Probably not an aggregation breaker, because it has just one negative charge, despite a suitable shape and volume
Zinc01547758		Steroid phosphates	Aggregation breaker with one attachment point (PO ₃ ⁻) on the Fc of bevacizumab
ZINC01870279		AMP	Aggregation breaker

Table 17 (continued): Potential aggregation breakers identified by docking of similarity search hits.

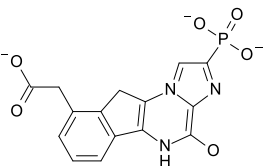
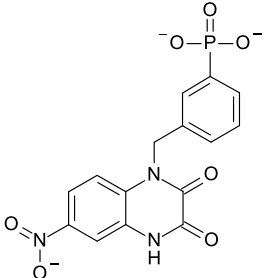
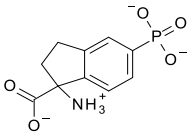
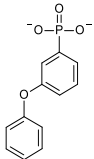
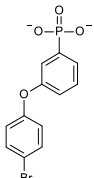
Name	Structure	Query molecule(s)	Prediction based on docking
ZINC01909478		AMP	Aggregation breaker
ZINC01547758		AMP	Aggregation breaker
ZINC02379522		AMP	Aggregation breaker
ZINC01870276		AMP	Aggregation breaker
ZINC01870305		AMP	Aggregation breaker

Table 18: Docked mono- and disaccharide phosphates.

Mono- and disaccharide-phosphates (commercially available as salts at Sigma-Aldrich)							
Name		Structure	Predicted to inhibit aggregation?	Name		Components	Predicted to inhibit aggregation?
Monosaccharides	Pentoses	ribose phosphate	no, very few docking poses clashing	Disaccharides	α -lactose-1-phosphate	glucose-phosphate + galactose	yes
		arabinose phosphate					
	Hexoses	α -D-glucose-6-phosphate			α -D(+) maltose-1-phosphate	glucose-phosphate + glucose	
		D-fructose-6-phosphate					
		L-fucose-1-phosphate			sucrose-phosphate (not commercially available)	glucose-phosphate + fructose	
		α -D-galactose-1-phosphate					
			Yes				

2.3.4 EXPERIMENTAL CONFIRMATION OF THE *IN SILICO* PREDICTIONS

The *in silico* predictions made were then submitted to experimental validation in order to verify the working hypothesis and mechanism of action generated by molecular modelling. Indeed, betamethasone phosphate stabilizes bevacizumab similarly to dexamethasone phosphate in a concentration-dependent manner (Figure 37).

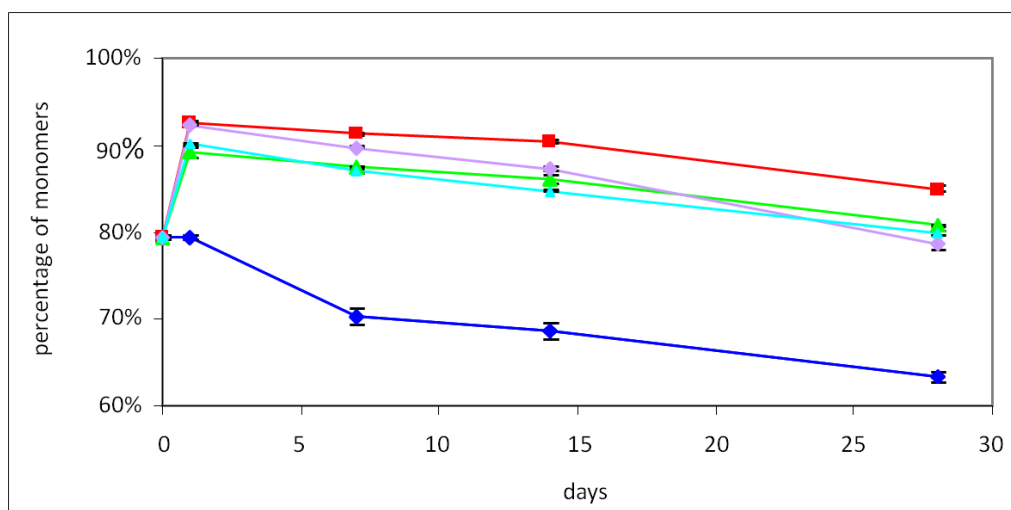


Figure 37: Betamethasone phosphate “stabilizes” bevacizumab *in vitro* similarly to dexamethasone phosphate. The colour code is the following: Bevacizumab formulated with dexamethasone disodium phosphate at a ratio of 1:153 (red) or 1:15.3 (green), with betamethasone disodium phosphate at a ratio of 1:153 (violet), 1:15.3 (light blue) or bevacizumab only as a negative control (dark blue). All bevacizumab samples in PBS buffer pH 7.0 were pre-stressed by storing them for 1 week at 40°C before adding the low molecular weight compounds. All measures were repeated three times and the corresponding means (data points) and standard deviations are indicated. This Figure was kindly provided by M. Veurink.

In contrast, triamcinolone acetonide phosphate does not enhance the percentage of bevacizumab monomers (Figure 38).

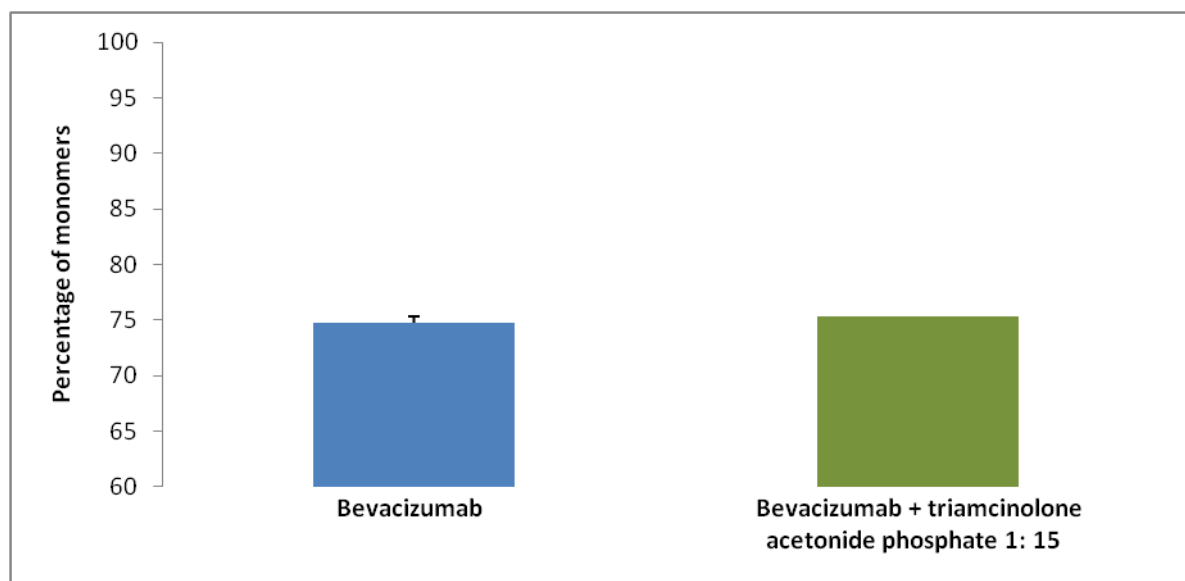


Figure 38: Triamcinolone acetonide phosphate does not stabilize bevacizumab: No stabilizing effect of bevacizumab stored at 40°C at pH 7.0 over 35 days (blue bar) could be observed. Bevacizumab with triamcinolone acetonide phosphate at a molar ratio of 1:15 is represented by the green bar. This Figure was kindly provided by M. Veurink.

The experimental results with the aggregation breaker betamethasone phosphate and the non-breaker triamcinolone acetonide phosphate confirm the modelling predictions and show that the *in silico* procedure together with the postulated underlying mechanism of action are able to distinguish breakers from non-breakers in a predictive way.

2.3.5 RATIONALE FOR THE CHOICE OF AMP

As described in 2.3.3, AMP has been found by similarity screening on over 4800 approved drug molecules using ROCS. Because it is a ubiquitous metabolite in the human body and the FDA considers it to be a GRAS compound, its potential use as an adjuvant is particularly interesting. Figure 39 illustrates the docking results for AMP. It binds perfectly and preferably in the region of Lys445 on the bevacizumab monomer and interferes strongly with the formation of the aggregation dimer, as measured by the number and orientation of the poses interfering with dimer formation.

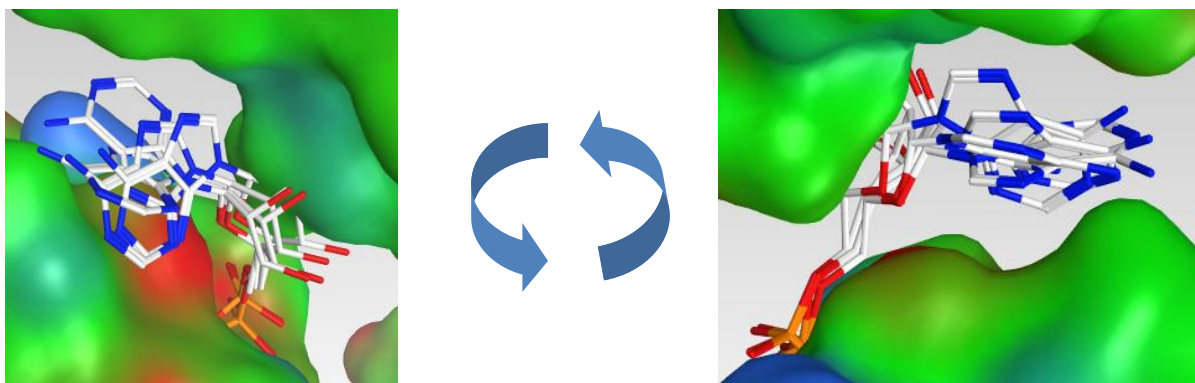


Figure 39: 2 different views of AMP docking poses, turned by approximately 180°.

2.3.6 PROOF-OF-CONCEPT FOR AMP

To prove that AMP is able to effectively stabilize bevacizumab *in vitro*, the concentration-dependent effect of AMP on the percentage of bevacizumab monomers was assessed over 28 days of storage at 40° at pH 7.0. Figure 40 evidences that the association of the dialyzed bevacizumab with AMP causes a remarkable stabilization of the antibody in comparison with the sample of bevacizumab only.

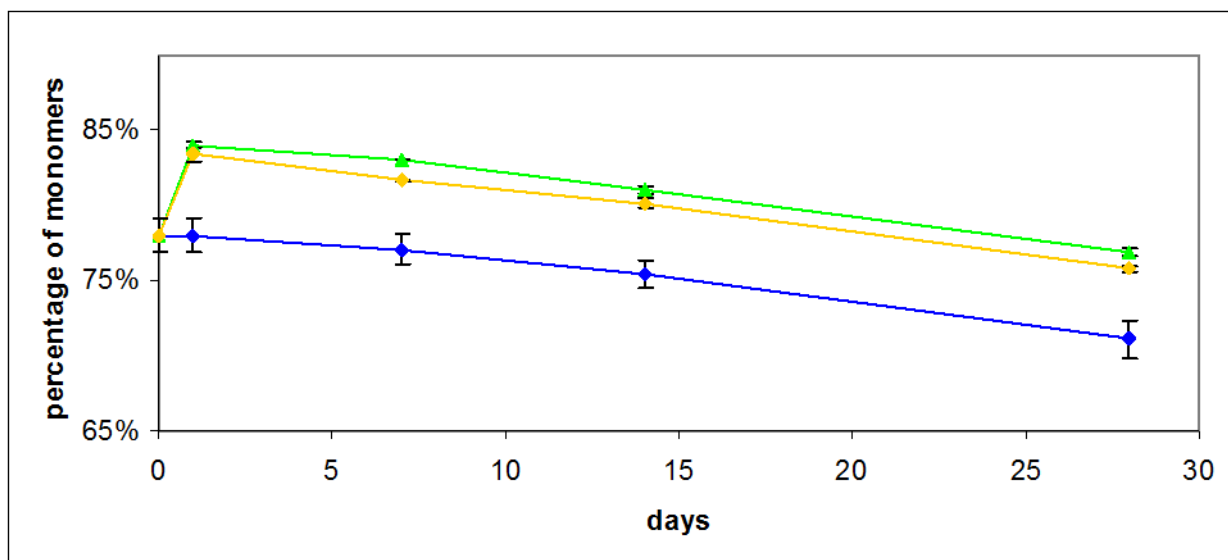


Figure 40: AMP stabilizes bevacizumab *in vitro*. Green stands for a 1:1 ratio, yellow for a 1:10 ratio, and blue for bevacizumab only. This Figure was kindly provided by M. Veurink.

Furthermore, the short-term protective effect of AMP in respect of aggregation on the commercial product Avastin®, i.e. the formulation administered to the patients, was investigated. The difference between the control and AMP 1:1 or 1:10 is statistically significant (assessed by a Kruskal-Wallis test, data not shown), while this is not the case when comparing the control with AMP 1:100 (Figure 41).

Thus, AMP has at least a short-term anti-aggregation effect on Avastin®, increasing the percentage of monomers by 3-4% to 96-97% depending on the ratio of Avastin® to AMP used. Besides, 31.8% of the preaggregated bevacizumab can be recovered.

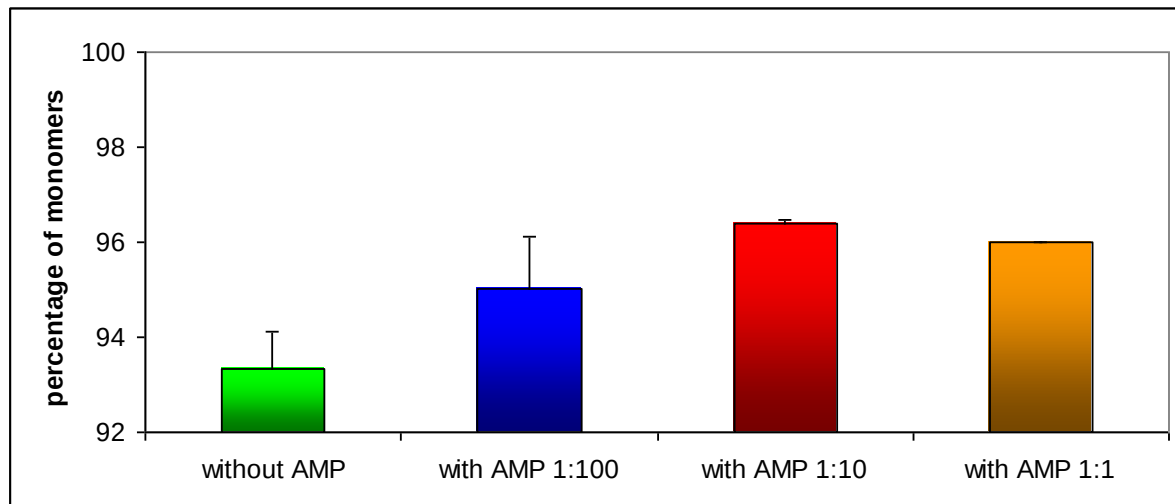


Figure 41: Short term protective effect of AMP on the commercial product Avastin®. Avastin® was stored for 24h at 40°C before adding AMP. The percentage of Avastin® monomers without AMP is shown in light green, the one with Avastin® and AMP at 1:100 in dark blue, the one with AMP at 1:10 in red and the one with AMP at an equimolar ratio in orange. Standard deviations are also shown. This Figure was kindly provided by M. Veurink.

This short-term protective effect of AMP is in line with the accelerated stability test (Figure 40).

2.3.7 MARKUSH STRUCTURE AROUND AMP IN FORM OF A STRUCTURE-ACTIVITY RELATIONSHIP AROUND AMP

To investigate whether there were other compounds that could work similarly to AMP and for developing a Markush structure, a structure-activity relationship was derived from the known experimental data and the compounds identified by docking and ligand-based similarity searches (Tables 16, 17, and 18). A “base-spacer-phosphate” principle was identified, with the following criteria for the base, the spacer, and the phosphate:

- (i) Base or bulky group: clashes into a Fab,
- (ii) Spacer: bridges at least partially the 4 Å which define the minimal distance in the Fc-Fab interface, and
- (iii) Phosphate: functions as an anchor in the Fc region via its interaction with the positively charged Lys445

Regarding (ii), the sugar spacer was used as a « platform to tailor diversity »⁴¹. From a synthesis point of view, the monosaccharide D-glucose, which is the most abundant monosaccharide in nature, can be seen as a pentavalent chiral scaffold⁴², on which the 5 chiral OH can be diversified by esterification.

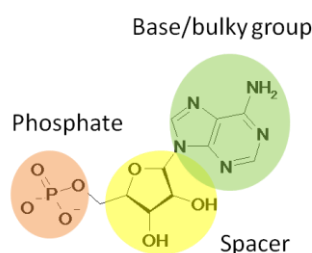
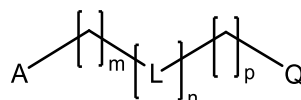


Figure 42: The “base-spacer-phosphate” principle applied to AMP.

With the principle in Figure 42, the docking results for the steroids (2.3.3) can be re-evaluated. For some docking poses of betamethasone phosphate, the spacer hydroxyl closest to the phosphate can eventually take over the binding to Lys445 instead of the phosphate. The role of this hydroxyl is thus either to form a H-bond with a close-by binding site residue (in the case of the representative docking poses for dexamethasone phosphate or betamethasone phosphate) or to bind directly to Lys445 (in the case of some betamethasone phosphate poses). However, it can be noted for triamcinolone acetonide phosphate, which despite its phosphate group does not break aggregation, that a spacer should neither be too rigid nor too bulky for the molecule to be able to clash into the antibody.



This principle led to a first Markush structure of the type $A-[CH_2]_m-L-[CH_2]_n-[CH_2]_p-Q$, wherein n equals 0 or 1, m and p are each independently 0 or 1; A is a negatively charged anchor moiety, for example selected from a carboxyl, phosphate, phosphonate, phosphinate, phosphorothioate, sulfate, or sulfonate moiety. A may preferably be selected from a phosphonate moiety, a phosphate moiety, or a bioisostere thereof; L is an optional linker group, wherein, when present, L is a C_{1-6} alkyl, C_{1-6} carbonyl, C_{1-6} ether, optionally substituted by one or more group(s) independently selected from C_{1-6} alkyl, hydroxyl, C_{1-6} alkoxy, ketone, halo or carboxyl group, or a substituted 5- or 6-membered alicyclic, heteroalicyclic, aromatic or heteroaromatic group containing from 0 to 3 heteroatoms selected from N, O and S, optionally substituted by one or more group(s) independently selected from C_{1-6} alkyl, hydroxyl, C_{1-6} alkoxy, ketone, halo or carboxyl group; Q is a cyclic moiety selected from an optionally substituted alicyclic, heteroalicyclic, aromatic or heteroaromatic moiety comprising 1 to 5 five- or six-membered rings, and 0 to 5 heteroatoms selected from N, O, and S optionally substituted by one or more group(s) independently selected from C_{1-6} alkyl, hydroxyl, C_{1-6} alkoxy, ketone, aldehyde, carboxyl, amine, nitro, thio or halo group, or a pharmaceutically acceptable salt thereof.

A more explicit Markush structure around AMP is shown in Figure 43, wherein R_1 is a nucleobase; R_2 is H or OR_4 , wherein R_4 is H or a C_{1-4} alkyl group; R_3 is H or OR_5 , wherein R_5 is H or a C_{1-4} alkyl group; and n is an integral from 1-3, or a pharmaceutically acceptable salt thereof. The nucleobase R_1 may be selected from the group consisting of adenine, guanine, thymine, uracil, xanthine, ethanoadenine, inosine, orotidine, or cytosine.

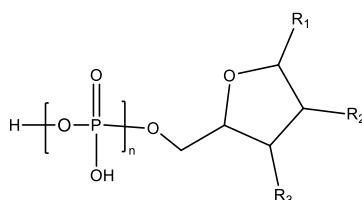


Figure 43: The Markush structure around AMP, with n , R_1 , R_2 , and R_3 explained in the text preceding this Figure.

Further details can be found in the patents WO 2011/121559 A2 and WO 2011/121560 A2.

2.3.8 APPLICABILITY OF AMP TO OTHER MONOCLONAL ANTIBODIES

To gain insight into the applicability of AMP to other monoclonal antibodies, the sequence of the mouse (PDB ID: 1IGY) and the human IgG1 Fc (PDB ID: 1HZH) are compared to the sequences of crystallized Fc regions. The C-terminal Fc regions are highly homologous to the human IgG1 used for the monomer and aggregation models (Figure 44). No statement can be made about the N-terminal Fc regions, as apart from 1IGY and 1HZH, they have mostly not been resolved. Notably, the contact residue Lys445 is fully conserved in all sequences, including five IgG1 and one IgG2a. The residues surrounding Lys445 are also located in a relatively well conserved amino acid patch (Figure 30). Most importantly, marketed antineoplastic monoclonal antibodies like rituximab (MabThera®, Hoffmann-La Roche) and trastuzumab (Herceptin®, Genentech) also contain this Lys and the conserved patch is also present. Rituximab is a chimeric antibody with antineoplastic properties used for the treatment of B-cell non-Hodgkin's lymphoma, and trastuzumab is a humanized anti-HER2 monoclonal antibody for the treatment of HER2-positive metastatic breast cancer.

In addition, regarding the docking on the whole bevacizumab monomer surface, the highest Fc-small molecule interaction scores and best outward-pointing poses were obtained on docking around Lys 445. Both the conservation and the preferred docking at Lys445 evidence that this contact residue is a specific binding site.

Lys445 should however be compared to aggregation-prone regions on the Fc and Fabs of IgG1 mentioned by Chennamsetty et al.⁴³ (two patents: WO2009/15513 & WO2009/15518). To obtain less aggregation than for an undisclosed wild-type antibody, they mutated hydrophobic residues to hydrophilic ones. This could be successfully achieved for non-CDR mutants, as demonstrated by a turbidity assay, HPLC, and differential scanning calorimetry, with a monomer increase of 2-5% (maximally 96% of monomers)⁴⁴. The main rationale for this approach was that the hinge of antibodies is known to be highly solvent-accessible and non-polar, suggesting hydrophobic interactions to be important driving forces for binding in this region⁴⁵.

Using a homology model based on separate Fab and human Fc crystal structures, they performed the first full antibody simulation in explicit solvent, which gave rise to a spatial aggregation propensity analysis. Previous studies on *in silico* aggregation used either small proteins or peptides (reviewed in⁴⁶) or molecular dynamics on the Fab only^{47,48}. The propensity values of Chennamsetty et al.^{43,44} reflect the effectively dynamically exposed hydrophobicity of a certain patch on the protein surface. Lys445, the amino acid of interest in this study, was not spotted. Hence, they did not mutate it for a better stability of the antibody monomer. One preceding residue, Leu443, was however identified as aggregation prone motif⁴³, but was not mutated either. This Leu is conserved among IgGs (IgG1-4). Looking at the sequence alignment provided by Chennamsetty et al.⁴³ (Figure 44), the last residue corresponds to the Lys identified here (Lys 445). It is conserved among IgGs as well.

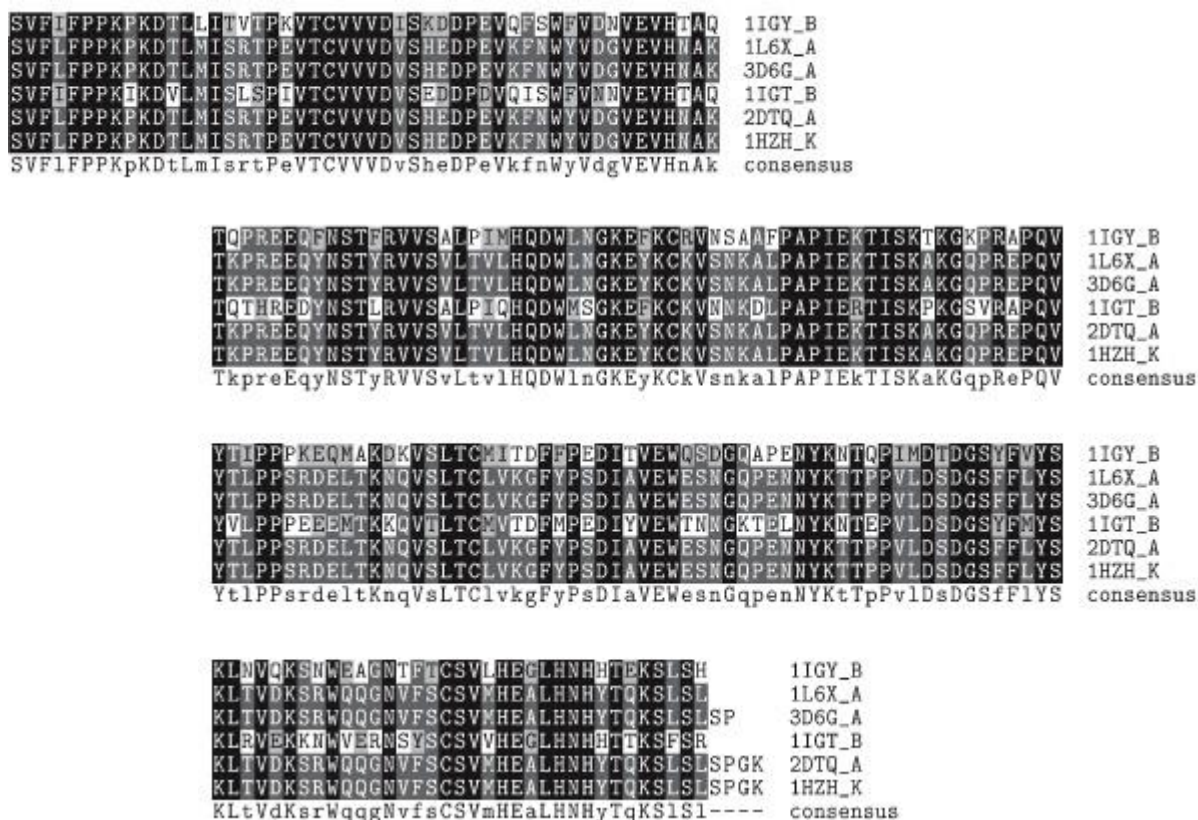


Figure 44: Sequence alignment of the murine Fc of IgG1 (PDB ID: 1IGY, chain B), the corresponding part of rituximab (an IgG1, PDB ID: 1L6X, chain A), trastuzumab (an IgG1, PDB ID: 3D6G, chain A), an IgG2a (PDB ID: 1IGT, chain B), an IgG1 (PDB ID: 2DTQ, chain A), and the human Fc of IgG1 (PDB ID: 1HZH, chain K). The same colouring as in Figure 28 applies. The sequence alignment starts at the first residue which is common to all sequences. Lys445 (red arrow) is conserved in all sequences.

	..A...	loop	...B....	loop	..C...C'loop..D....	loop	...E... loop	...F... loop	...G....	
	351	361	371	381	391	401	411	421	431	441
IgG1	PQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTT	PPVLDSDGGS	FLYSKLTVDKSRWQQGNVFC	SCVMHEALHNHYTQKSLS	SPGK					
IgG2	PQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTT	PPVLDSDGGS	FLYSKLTVDKSRWQQGNVFC	SCVMHEALHNHYTQKSLS	SPGK					
IgG4	PQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTT	PPVLDSDGGS	FLYSKLTVDKSRWQQGNVFC	SCVMHEALHNHYTQKSLS	SPGK					
IgG3	PQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTT	PPVLDSDGGS	FLYSKLTVDKSRWQQGNVFC	SCVMHEALHNHYTQKSLS	SPGK					

Figure 45: Sequence comparison of different IgG subclasses: The aggregation prone motifs identified are marked in red and the sequence differences between different IgGs in yellow. The red arrow marks the lysine residue Lys445. In this thesis, it was identified to be the anchor point for small molecular weight antibody aggregation breakers. The Figure was adapted from Chennamsetty et al. ⁴³.

However, with respect to the conservation of Lys445 among all IgGs (Figure 45), it is important to specify that only IgG1s and one IgG2a Fc have been crystallized so far. Because the crystal symmetries and thus the potential aggregation zones of IgG3 and IgG4 are unknown, the here presented findings cannot be extrapolated to these IgGs at the moment.

2.4 CONCLUSIONS

In this chapter, it was shown how a clinical observation can lead to unravelling molecular recognition with the aim of stabilizing therapeutic monoclonal antibody formulations. Based on a combined *in silico-in vitro* approach, involving predictions on the aggregation breaking propensity of small molecules and successful experimental validation, a mechanistic explanation on the antibody stabilizing effect of steroid drugs could be given. Ligand-based screening followed by docking onto the modelled monomeric antibody bevacizumab revealed molecules with a “phosphate-linker-base” pattern to clash into the adjacent monomer, i.e. to interfere with dimer formation. AMP revealed to be an equimolarly active aggregation breaker stabilizing the fully formulated therapeutic monoclonal antibody Avastin® (bevacizumab) by increasing the percentage of monomers by 3-4% to 96-97% compared to the antibody only and recovering 31.8% of the preaggregated bevacizumab. A notable advantage of AMP for inclusion in formulation and production is that it is classified as a Generally Recognized As Safe (GRAS) compound by the FDA.

The crystal symmetry-based aggregation model of bevacizumab showed only one contact region with a distinct conserved Lys, to which phosphate-containing molecules could easily bind due to a strong electrostatic interaction. This Lys is conserved in therapeutic monoclonal IgG1 antibodies like trastuzumab (Herceptin®) or rituximab (MabThera®). Therefore, if the primary event of aggregation modelled in this work is the most important, the food supplement AMP could be used to tackle a broad range of antibody aggregation problems.

2.5 ACKNOWLEDGEMENTS

I would like to thank Prof. Dr. R. Gurny and M. Veurink of the the Laboratory of Pharmaceutics, Biopharmaceutics, and Pharmaceutical Technology of the University of Geneva for the productive and nice collaboration.

2.6 REFERENCES

1. Aggarwal, S. What's fueling the biotech engine? *Nat Biotechnol* **25**, 1097-104 (2007).
2. Pharmavita sector analysis - Monoclonal Antibodies: 2009 Update. *Thomson Pharma* Reference code: DMHC2579 (2009).
3. Demeule, B., Gurny, R. & Arvinte, T. Where disease pathogenesis meets protein formulation: renal deposition of immunoglobulin aggregates. *Eur J Pharm Biopharm* **62**, 121-30 (2006).
4. Kalsi, J., Ravirajan, C. T., Rahman, A. & Isenberg, D. A. Structure-function analysis and the molecular origins of anti-DNA antibodies in systemic lupus erythematosus. *Expert Rev Mol Med* 1999, 1-28 (1999).
5. Vermeer, A. W. & Norde, W. The thermal stability of immunoglobulin: unfolding and aggregation of a multi-domain protein. *Biophys J* **78**, 394-404 (2000).
6. Presta, L. G. et al. Humanization of an anti-vascular endothelial growth factor monoclonal antibody for the therapy of solid tumors and other disorders. *Cancer Res* **57**, 4593-9 (1997).
7. Los, M., Roodhart, J. M. & Voest, E. E. Target practice: lessons from phase III trials with bevacizumab and vatalanib in the treatment of advanced colorectal cancer. *Oncologist* **12**, 443-50 (2007).
8. Pavlou, F. Is Avastin the future? *Ophthalmology Times Europe* **1** (2005).
9. Rosenfeld, P. J. et al. Maximum tolerated dose of a humanized anti-vascular endothelial growth factor antibody fragment for treating neovascular age-related macular degeneration. *Ophthalmology* **112**, 1048-53 (2005).
10. Rosenfeld, P. J., Moshfeghi, A. A. & Puliafito, C. A. Optical coherence tomography findings after an intravitreal injection of bevacizumab (avastin) for neovascular age-related macular degeneration. *Ophthalmic Surg Lasers Imaging* **36**, 331-5 (2005).
11. Rosenfeld, P. J., Fung, A. E. & Puliafito, C. A. Optical coherence tomography findings after an intravitreal injection of bevacizumab (avastin) for macular edema from central retinal vein occlusion. *Ophthalmic Surg Lasers Imaging* **36**, 336-9 (2005).
12. Augustin, A. J., Puls, S. & Offermann, I. Triple therapy for choroidal neovascularization due to age-related macular degeneration: verteporfin PDT, bevacizumab, and dexamethasone. *Retina* **27**, 133-40 (2007).
13. Ahmadieh, H. et al. Single-session photodynamic therapy combined with intravitreal bevacizumab and triamcinolone for neovascular age-related macular degeneration. *BMC Ophthalmol* **7**, 10 (2007).
14. Colucciello, M. Intravitreal bevacizumab and triamcinolone acetate combination therapy for exudative neovascular age-related macular degeneration: short-term optical coherence tomography results. *J Ocul Pharmacol Ther* **24**, 15-24 (2008).
15. Fraunhofer, W. & Winter, G. The use of asymmetrical flow field-flow fractionation in pharmaceuticals and biopharmaceuticals. *Eur J Pharm Biopharm* **58**, 369-83 (2004).
16. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**, 779-815 (1993).
17. Hubbard, T. J. & Blundell, T. L. Comparison of solvent-inaccessible cores of homologous proteins: definitions useful for protein modelling. *Protein Eng* **1**, 159-71 (1987).
18. Lesk, A. M., Levitt, M. & Chothia, C. Alignment of the amino acid sequences of distantly related proteins using variable gap penalties. *Protein Eng* **1**, 77-8 (1986).
19. Sanchez, R. & Sali, A. Advances in comparative protein-structure modelling. *Curr Opin Struct Biol* **7**, 206-14 (1997).
20. Chothia, C. & Lesk, A. M. The relation between the divergence of sequence and structure in proteins. *Embo J* **5**, 823-6 (1986).
21. Johnson, M. S., Srinivasan, N., Sowdhamini, R. & Blundell, T. L. Knowledge-based protein modeling. *Crit Rev Biochem Mol Biol* **29**, 1-68 (1994).
22. Vogt, G., Etzold, T. & Argos, P. An assessment of amino acid exchange matrices in aligning protein sequences: the twilight zone revisited. *J Mol Biol* **249**, 816-31 (1995).
23. Berman, H. M. et al. The Protein Data Bank. *Nucleic Acids Res* **28**, 235-42 (2000).
24. Berman, H. M. et al. The Protein Data Bank. *Acta Crystallogr D Biol Crystallogr* **58**, 899-907 (2002).
25. Muller, Y. A., Christinger, H. W., Keyt, B. A. & de Vos, A. M. The crystal structure of vascular endothelial growth factor (VEGF) refined to 1.93 Å resolution: multiple copy flexibility and receptor binding. *Structure* **5**, 1325-38 (1997).
26. Saphire, E. O. et al. Crystal structure of a neutralizing human IgG against HIV-1: a template for vaccine design. *Science* **293**, 1155-9 (2001).

27. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673-80 (1994).
28. Larkin, M. A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-8 (2007).
29. Guex, N. & Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714-23 (1997).
30. Schneider, G., Neidhart, W., Giller, T. & Schmid, G. "Scaffold-Hopping" by Topological Pharmacophore Search: A Contribution to Virtual Screening. *Angew Chem Int Ed Engl* **38**, 2894-2896 (1999).
31. Schneider, G. et al. Virtual Screening for Bioactive Molecules by Evolutionary De Novo Design. Special thanks to Neil R. Taylor for his help in preparation of the manuscript. *Angew Chem Int Ed Engl* **39**, 4130-4133 (2000).
32. Boehm, H. J., Flohr, A. & Stahl, M. Scaffold hopping. *Drug Discov Today* **1**, 217-224 (2004).
33. Jenkins, J. L., Glick, M. & Davies, J. W. A 3D similarity method for scaffold hopping from known drugs or natural ligands to new chemotypes. *J Med Chem* **47**, 6144-59 (2004).
34. Renner, S. & Schneider, G. Scaffold-hopping potential of ligand-based similarity concepts. *ChemMedChem* **1**, 181-5 (2006).
35. Irwin, J. J. & Shoichet, B. K. ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model* **45**, 177-82 (2005).
36. Markush, E. A. Pyrazolone Dye and Process of Making the Salt. *US Patent 1506316* (1924).
37. Schuffenhauer, A. et al. Molecular diversity management strategies for building and enhancement of diverse and focused lead discovery compound screening collections. *Comb Chem High Throughput Screen* **7**, 771-81 (2004).
38. Beitz, E. TEXshade: shading and labeling of multiple sequence alignments using LATEX2 epsilon. *Bioinformatics* **16**, 135-9 (2000).
39. Schneider, H.-J., Schiestel, T. & Zimmermann, P. Importance of two buried salt bridges in the stability and folding pathway of barnase. *JACS* **114**, 7698-7703 (1992).
40. Tissot, A. C., Vuilleumier, S. & Fersht, A. R. Importance of two buried salt bridges in the stability and folding pathway of barnase. *Biochemistry* **35**, 6786-94 (1996).
41. Opatz, T. et al. D-glucose as a multivalent chiral scaffold for combinatorial chemistry. *Carbohydr Res* **337**, 2089-110 (2002).
42. Le, G. T. et al. Molecular diversity through sugar scaffolds. *Drug Discov Today* **8**, 701-9 (2003).
43. Chennamsetty, N., Helk, B., Voynov, V., Kayser, V. & Trout, B. L. Aggregation-prone motifs in human immunoglobulin G. *J Mol Biol* **391**, 404-13 (2009).
44. Chennamsetty, N., Voynov, V., Kayser, V., Helk, B. & Trout, B. L. Design of therapeutic proteins with enhanced stability. *Proc Natl Acad Sci U S A* **106**, 11937-42 (2009).
45. DeLano, W. L., Ultsch, M. H., de Vos, A. M. & Wells, J. A. Convergent solutions to binding at a protein-protein interface. *Science* **287**, 1279-83 (2000).
46. Cellmer, T., Bratko, D., Prausnitz, J. M. & Blanch, H. W. Protein aggregation in silico. *Trends Biotechnol* **25**, 254-61 (2007).
47. Noon, W. H., Kong, Y. & Ma, J. Molecular dynamics analysis of a buckyball-antibody complex. *Proc Natl Acad Sci U S A* **99 Suppl 2**, 6466-70 (2002).
48. Sinha, N. & Smith-Gill, S. J. Molecular dynamics simulation of a high-affinity antibody-protein complex: the binding site is a mosaic of locally flexible and preorganized rigid regions. *Cell Biochem Biophys* **43**, 253-73 (2005).

CHAPTER 3 - NADPH OXIDASES

ABSTRACT

With the aim of complementing research efforts on small molecular weight and peptide inhibitors of different NADPH oxidases (NOXes), which are ongoing at the Centre Médical Universitaire of the University of Geneva and a Geneva-based start-up (GenKyoTex), different molecular modelling studies were conducted as part of this thesis.

In the main part of this chapter, a crystallized protein-protein interaction interface of importance in the assembly and activation of the phagocyte NOX2 and so far not targeted by known inhibitors is assessed:

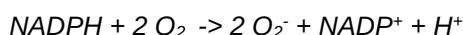
- 43 potential small molecular weight inhibitors were identified by combining a pharmacophore search with structure-based virtual screening (in collaboration with GenKyoTex).
- Using four different docking programs (Dock, FlexX, GOLD, and Surflex), a potential binding orientation was suggested for the experimental hit celastrol. Subsequent intrinsic fluorescence measurements and native MS confirmed that celastrol is inhibiting the targeted protein-protein interaction (in collaboration with Dr. V. Jaquet and Prof. Dr. K.-H. Krause, Department of Pathology and Immunology, Centre Médical Universitaire, University of Geneva).
- In an attempt to improve the binding affinities for peptidomimetics targeting the same protein-protein interface, a GRID-based interaction analysis was carried out to spot favourable substitutions of natural with non-natural amino acids (in collaboration with Dr. H. Gaertner/Prof. Dr. O. Hartley (synthesis) and Dr. V. Jaquet (binding affinity measurements)). The *in vitro* predicted peptides had IC₅₀ values comparable to the native peptide. Besides, a putative binding mode of the retropeptide was predicted.

3.1 INTRODUCTION TO THE BIOLOGICAL CONTEXT

3.1.1 NOX2, THE PROTOTYPE NADPH OXIDASE

At the front-line of immunological host defence, the innate immune system is activated by professional phagocytes comprising neutrophils, monocytes, macrophages, and eosinophils. Here, the focus is put on the oxygen-dependent defence mechanism of phagocytes, while other oxygen-independent functions like chemotaxis, phagocytosis, degranulation, release of lytic enzymes, and bactericidal peptides (reviewed in e.g. ¹) are left aside. Phagocytes engulf a microbe and subsequently produce superoxide (O₂⁻) as microbicide, which can be further transformed into other oxygen- or nitrogen-based reactive oxygen species (ROS). Besides ROS, neutrophil proteases are also involved in microbial killing ²⁻⁴. ROS are produced by NADPH oxidases, but also as by-products of the metabolism of mitochondria, peroxysomes, and cytochromes. They also interact with various host proteins, lipids and nucleic acids to promote pro-inflammatory events critical in the pathogenesis of e.g. Alzheimer's disease or atherosclerosis. The popular free radical theory of ageing ^{5,6} postulates that ageing is either *per se* a process involving ROS or that such processes are involved in degenerative diseases associated with ageing.

The phagocyte NADPH oxidase (NOX2) oxidizes O₂ while reducing NADPH according to:



The human body provides the reducing equivalents (NADPH) via the hexose monophosphate shunt, which is an alternative to glycolysis and the Krebs cycle for glucose metabolism. The catalytic core of the enzyme is the flavocytochrome b₅₅₈ (Figure 46), named after its UV absorbance maximum at about 558 nm. Electrons are carried from NADPH via FAD and heme b to O₂ ⁷, with the molar ratio of FAD to heme b being 1:2. In line with the role of FAD as part of the electron transport chain, FAD enhances enzymatic activity ⁸ and 5-deaza-FAD inhibits it ⁹. The flavocytochrome b₅₅₈ depicted in Figure 46 is a heterodimeric membrane constituent composed of the glycoprotein NOX2 and the smaller protein p22^{phox} (phox for phagocyte oxidase).

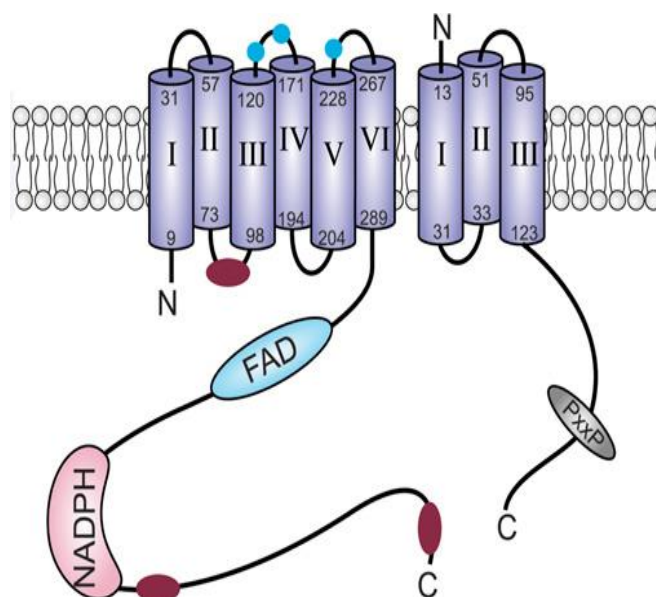


Figure 46: The flavocytochrome b_{558} . The predicted transmembrane helices of NOX2 (I to VI) and $p22^{phox}$ (I to III) are shown in violet, the NADPH binding site in pink, the FAD binding site in cyan, regions that are thought to interact with $p47^{phox}$ in the active state in red and glycosylation sites as cyan dots. The consensus motif PxxP (P = Pro and x = any amino acid) in the cytoplasmic region of $p22^{phox}$, which is known to interact with $p47^{phox}$, is indicated in grey. The Figure was taken from Groemping and Rittinger ¹⁰.

In the absence of a crystal structure of NOX2, whose apo NADPH binding site has been solved in 2010 only, hydropathy plots were used to predict transmembrane domains ^{7,11}. Phage display library screening served to define extracellular domains ¹²⁻¹⁴, antibody mapping revealed the C-terminus to be located in the cytoplasm ^{13,15,16} and sequencing followed by antibody mapping disclosed the cytoplasmic N-terminus ^{17,18}. As depicted in Figure 46, NOX2 has six transmembrane domains, and both the N- and C-termini are facing the cytoplasm. The high glycosylation of human NOX2 with N-acetyl glucosamine, galactose, fructose, mannose, and glucose ¹⁹ stabilizes the enzyme.

Cytosolic components recruited to the membrane upon activation include $p40^{phox}$, $p47^{phox}$, $p67^{phox}$, and the GTP-binding proteins Rac 1 or 2. Clinically, one or several mutations in the genes encoding NOX2, $p22^{phox}$, $p47^{phox}$ or $p67^{phox}$ lead to recurrent and life-threatening forms of Chronic Granulomatous Disease (CGD). CGD patients suffer from bacterial and fungal infections that do not cause disease in healthy individuals. Mutations in NOX2 concern mostly the level of assembly or activation, which is mediated by multiple SH3 interactions and phosphorylations. CGD is most probably the first genetic pathology that has been linked to an impaired protein-protein interaction ²⁰.

In NOX2, the most prominent mutation is Pro156Gln ²¹ in the Pro-rich region (PRR) of the membrane-associated $p22^{phox}$, rendering the interaction with the cytosolic $p47^{phox}$ impossible ²⁰. Based on further studies on the mutations Pro157Gln and Pro160Gln in this PRR, which selectively inhibited the association of NOX2 with $p47^{phox}$ and $p67^{phox}$ ²², it was found that $p22^{phox}$ may play a role in NOX2 stabilization and/or the recruitment of regulatory subunits.

3.1.2 ISOFORMS OF NOX2

Isoforms of NOX2 are known since the year 2000 ²³. To date, seven homologues are known, NOX1-5 and DUOX1-2 ²⁴, with molecular weights ranging from 65 kDa (NOX1-4) to 180 kDa (DUOXes). Their properties with respect to tissue distribution, expression, physiological function, regulatory subunits, secondary structure, ligands, and activation are summarized in Table 19. Occurring in non-phagocytic cells, they are involved in cell signalling or O_2 -sensing ²⁵. Lower levels of ROS are seen in NOX2 homologues ²⁴. Further details on the molecular composition and regulation of the NOX family can be found in Krause et al. ²⁶ and Sumimoto et al. ²⁷.

Table 19: Characteristics of the seven homologues of the NOX/DUOX family. PRR stands for Pro-rich region.

	NOX1	NOX2	NOX3	NOX4	NOX5	DUOXes
Tissue distribution ²⁶	Mostly colon ^{23,28}	Phagocytes	Inner ear	Kidney, endothelium, smooth muscles, heart, pancreas, placenta, skeletal muscle, ovary, testis, fibroblasts, astrocytes, osteoclasts	Lymphoid tissues, testis ²⁹	DUOX1: thyroid, respiratory epithelia DUOX2: thyroid, gastrointestinal glandular epithelia
Expression ²⁶	Sometimes inducible ²⁸	Neutrophil granulocytes, monocytes, macrophages, eosinophiles	Abundant in the auditory system ^{30,31}	Abundant in kidney cortex ^{32,33} , widest expressed Nox; distinct pharmacology and pattern of ROS generation, potentially inducible ³⁴		
Physiological function ²⁶	Host defence, eventually blood pressure regulation	Host defence	Otoconia morphogenesis	Possibly oxygen sensing	Possibly signalling	Thyroid hormone synthesis, possibly also epithelial host defence ³⁵
Activation	By regulatory subunits ^{22,36,37}			Constitutively active ^{22,36,37}	Ca ²⁺ -dependent ^{29,38,39}	
Regulatory subunits	NOXO1, NOXA1, p22 ^{phox} required ²²	p47 ^{phox} , p67 ^{phox} , Rac, p22 ^{phox} required ²²	NOXO1, PRR of p22 ^{phox} required under certain conditions ²²	None, but PRR of p22 ^{phox} required ²²	None, PRR of p22 ^{phox} not required ²²	Nothing known
Predicted secondary structure	6 transmembrane α -helices ²⁴				N-terminal additional Ca ²⁺ binding domain ^{29,38}	Similar to NOX5, additional N-terminal peroxidase domain ⁴⁰
Ligands	2 hemes in transmembrane region , 1 FAD + 1 NADPH in C-terminal ²⁴				2 hemes, 1 FAD, 1 NADPH, Ca ²⁺	

Sequence and structural comparison of NOX isoforms

Due to the recent publication of the crystal structure of the apo NOX2 NADPH-binding domain in the PDB (PDB ID: 3A1F), the secondary structure elements can be added to published multiple sequence alignments of NOX1 to NOX5. Such a multiple sequence alignment is depicted in Figure 47.

Two main statements can be made: (i) The amino acids involved in NADPH binding belong to the most conserved regions and (ii) NOX5 is longer than the other NOXes, presenting different inserts, particularly in the region between the pyrophosphate and the ribose binding residues.

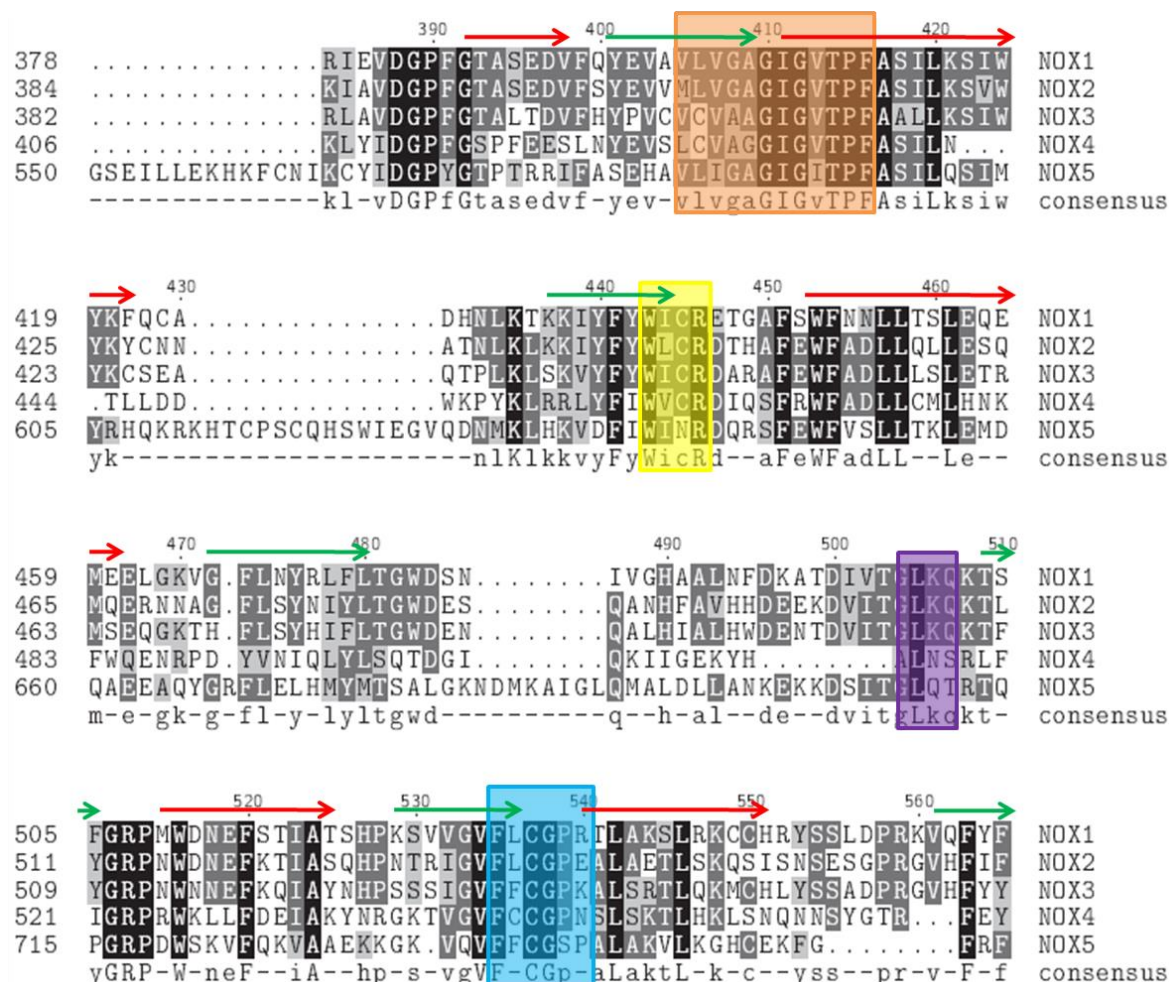


Figure 47: ClustalW⁴¹ multiple sequence alignment of NOX1 to NOX5. Each sequence numbering is provided on the left, each sequence name on the right. The residue ruler corresponds to the sequence numbering of NOX2, the prototype NOX. Fully conserved residues at equivalent positions are depicted as capital letters in the consensus sequence and are highlighted using a black background in each NOX. Amino acids that are conserved in three out of five cases are shown on a dark grey background. Residues conserving the amino acid type only are shown on a light grey background. The arrows above NOX1 stand for secondary structure elements as found in the apo crystal structure of NOX2 (PDB ID: 3A1F), with α -helices in red and β -sheets in green. The parts putatively involved in NADPH binding⁴² are indicated as coloured boxes: the pyrophosphate interacting part in orange, the ribose one in yellow, the adenine one in violet, and the nicotinamide one in cyan. The raw multiple sequence alignment was graphically refined in TexShade⁴³.

The crystal structure of the apo NADPH-binding domain of NOX2 (PDB ID: 3A1F) displays five β -strands forming a β -sheet and five surrounding α -helices. From a structural point of view, it is known that the NOXes share an N-terminal hydrophobic membrane-spanning domain, with an additional α -helix in NOX5, and a C-terminal flavoprotein domain with weak homology to ferredoxin NADPH-reductases (FNRs).

Evolution

A study by Kawahara et al. ⁴⁴ on 68 highly conserved residues, 14 of which are known to cause X-chromosome linked CDG when mutated, is informative with respect to the NOX/DUOX evolution. Apparently, Ca²⁺-regulated orthologues (NOX5, DUOX1 and 2) diverged early. NOX4 has a common “ancestor” with NOX1 to 3, given that NOX1 to 4 require p22^{phox} for activity/stability ^{22,45}. Subunit-regulated NOXes (NOX1 to 3) appeared first in fungi and amoeba. Furthermore, NOX3, which is involved in gravity perception, emerged as a land adaptation of vertebrates.

When comparing the evolutionary substitution rates, NOX2, p47^{phox}, p67^{phox}, and DUOXes are better conserved in vertebrates than other NOXes and regulatory subunits. In fact, duplication of primordial p47^{phox} and p67^{phox} genes occurred in vertebrates, with the duplicated branches evolving into the respective homologues NOXO1 (NOX organizer protein 1) and NOXA1 (NOX activator protein 1) ⁴⁶.

Interestingly and of importance for targeting the p47^{phox}-p22^{phox} interaction, evolutionary very highly conserved residues in the tandem SH3 domain of p47^{phox} include Trp193, Phe209, and Trp263, which are essential for the p47^{phox}-dependent activation of human NOX2 ^{47,48}. Also, the p47^{phox}-interacting PRR in p22^{phox} is highly conserved among NOX2 orthologues (i.e. NOX2 homologues in other species) ⁴⁶.

Relevance of the NOXes in health and disease

The phagocyte NOX2 is quiescent in resting cells, but activated upon exposure to bacteria or chemical stimuli. As already mentioned for CGD, deregulation of this isoform heavily compromises the immune system of a patient due to its decreased capacity to fight off bacteria or fungi.

Other isoforms are found in cancer: NOX1 has oncogenic properties when overexpressed ²⁸ and NOX4 is frequently expressed in cancer cells ⁴⁹.

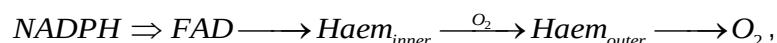
NOXes also have a role in the cardiovascular system (reviewed in ⁵⁰), as cardiovascular pathophysiology is influenced by increased oxidative stress ⁵¹. In the normal endothelium, vascular NOX have anticoagulant and anti-inflammatory properties, promoting vasodilatation by e.g. nitric oxide (NO). However, the superoxides generated by NOX2 and NOX4 contribute to the impairment of vasodilatation by inactivating NO ⁵². Furthermore, NOX is known to regulate other oxidases such as eNOS ⁵³ and xanthine oxidase ⁵⁴. In early disease stages, oxidative stress affects cellular compartments and single enzymes, while the cellular redox potential is only influenced in advanced disease stages ⁵⁵.

The central nervous system (CNS) accounts for over 20% of the oxygen consumed by the body. It produces large quantities of ROS and is sensitive to oxidative stress particularly affecting polyunsaturated fatty acids in membranes. The inhibition of NADPH oxidase-dependent ROS generation in the CNS could improve conditions of patients suffering from ischemic stroke, Alzheimer's disease, Parkinson's disease or HIV dementia.

Means to act upstream of the NOXes include statins, angiotensin-converting enzyme (ACE) inhibitors, and angiotensin receptor antagonists ⁵⁵. While statins inhibit the expression and activity of NOX1 ⁵⁶ and NOX2 ⁵⁷, as well as the isoprenylation of Rac1 required for the attachment to the flavocytochrome and thus enzyme activation, angiotensin receptor blockers and ACE inhibitors inhibit NOX activation and expression ⁵⁸. In this chapter, different direct ways of NOX inhibition are discussed.

3.1.3 ELECTRON AND PROTON TRANSFER

It is postulated ⁵⁹ that upon NOX activation, conformational changes give access to the NADPH binding site ⁶⁰, modifying the relative orientation of the redox centres and/or inducing a spin change in heme ⁶¹. The NOX electron transport chain can be presented as:



with $\Rightarrow$ representing a two electron transport and $\longrightarrow$ a one electron transport. In total, two electrons are delivered to molecular oxygen. It appears that the first step from NADPH to reduced FAD is the rate limiting step of the reaction ⁵⁹. The third step between the two non-equivalent hemes ⁶² with different redox potentials ⁶³ is energetically unfavourable and requires the final electron acceptor oxygen to be present. The electron flow seems to be controlled between NADPH and FAD (from the inactive to the intermediate state) and between FAD and heme (from the intermediate to the active state) ⁶⁴⁻⁶⁶. The intermediate state is rapidly formed from the inactive one, while the formation of the fully active one is slower ⁶⁷. Lipid and cytosolic factors possibly influence the equilibrium, strongly favouring the transition of the inactive towards the active flavocytochrome ⁶⁸.

The activation of NOX2 causes the vacuole pH to rise to about 8.0, after which it slowly falls ⁶⁹. Given that the enzyme generates electrons, a charge compensation is required for its continuous functioning: The protons extruded likely correspond to the molar quantity of O_2^- generated ^{70,71}. Without such a compensation in the form of a built-in H^+ channel, massive cellular acidification and cell damage would happen ⁷². This channel is formed by membrane-spanning His, which coordinate the two hemes of the redox centre, and is activated upon heme release ⁷³. By increasing the mobility of these His and the accessibility of their imidazoles to the solvent pH, electron and proton transport might be coupled during activation ⁷⁴. Evidence for such a coupling was e.g. found by semi-empirical calculations of electron tunnelling currents in bovine heart cytochrome c ⁷⁵ (catalyzing the reduction from O_2 to H_2O), with amino acids involved in electron tunnelling showing very little evolutionary variability.

In accordance with the conservation of residues involved in such reactions, the four His that coordinate the two hemes in a pairwise manner are conserved across the NOXes ⁷⁶ in many species ⁷⁷, but also ferric reductases like the *S. cerevisiae* FRE1 ^{78,79}. Initially, these residues were discovered by site-directed mutagenesis ^{78,79} demonstrating that His101 and His115 in helix 3, as well as His209 and His222 in helix 5 serve as non-covalent axial heme ligands residing within the hydrophobic N-terminus of NOX2 ⁸⁰. Because heme can be hexacoordinated and His imidazoles provide two interaction sites per heme, two different heme binding models exist. Both of them are supported by a substantial body of evidence reviewed in ⁸¹. According to the shared heme model, one heme is coordinated by NOX2 and one shared between NOX2 and the invariant His of p22^{phox} ⁸¹. In the non-shared heme model, both hemes are coordinated by NOX2. It seems as if the only His in p22^{phox}, His94, is not required for heme binding ⁸², even if a His94Arg mutation in a CGD patient led to undetectable b₅₅₈ levels ⁸³. Regarding enzymes binding heme, His imidazole is by far the most common proximal heme ligand, while a Cys thiolate is involved in e.g. cytochromes ^{84,85}, which metabolize diverse xenobiotics such as drug molecules, environmental compounds and pollutants ⁸⁶. Interestingly, by mutating a heme binding residue, it is even possible to shift enzyme function. As an example, Vatsis et al. showed how the cytochrome P450 2B4 loses its oxygenase function, retains NADPH oxidase activity, and gains a peroxidase function upon replacing Cys436 by a His ⁸⁵.

3.1.4 ACTIVATION OF NADPH OXIDASES

The activation of NADPH oxidases is a complex process regulated by phosphorylation, translocation, and multiple conformational changes. For NOX2 to be active, the cytosolic components (p40^{phox}, p47^{phox}, p67^{phox}, and Rac1 or 2) need to translocate to the membrane-bound flavocytochrome composed of the subunits NOX2 and p22^{phox} (Figure 48).

p47^{phox} and p67^{phox}, which are both absent in CGD patients, contain each two SH3 domains ⁸⁷ and are shown in the following to be crucial players in the assembly required for activation.

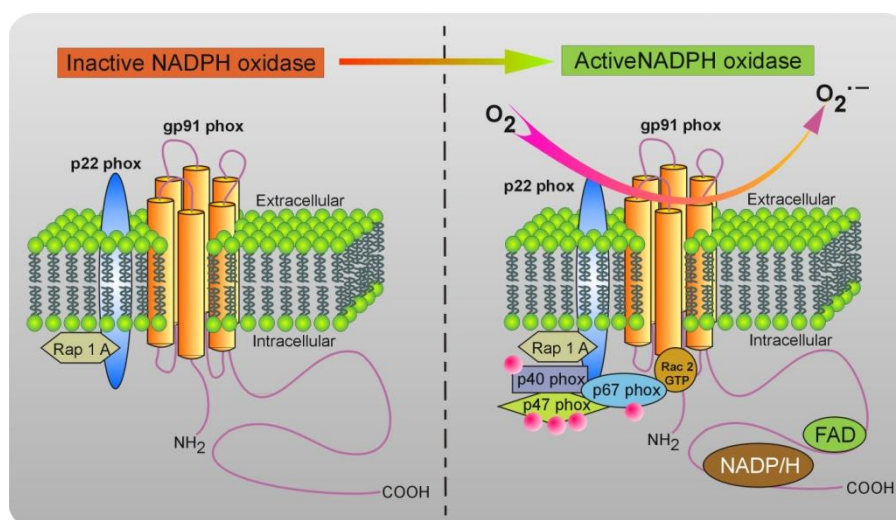


Figure 48: The activation of the NADPH oxidase requires the cytosolic components p40^{phox}, p47^{phox}, p67^{phox}, and Rac 1 or 2 to be recruited to the membrane into which gp91^{phox} (NOX2) and p22^{phox} are inserted. The pink dots on the cytosolic subunits stand for multiple phosphorylations required for translocation to happen. The Ras-related protein 1A (Rap 1A) is thought to stabilize the flavocytochrome. The Figure was taken from <http://www.GenKyoTex.com/index.php?rubID=5>.

Unravelling activation

Early NMR studies with a C-terminal peptide of NOX2 bound to p47^{phox} suggested it to bind in an extended, mostly rigid conformation⁸⁸. Shortly later, a conformational change was hypothesized because *in vitro* activators of NADPH oxidase and phosphorylation by protein kinase C induced a dose-dependent decrease in intrinsic Trp fluorescence of p47^{phox}⁸⁹. Fluorescence studies on full length, N- and C-terminal p47^{phox} revealed the C-terminal part to undergo a conformational change during activation⁹⁰. Indeed, small-angle X-ray scattering revealed an extended organization of the resting p47^{phox}⁹¹. Further evidence arose with the resolution and comparison of the crystal structures representing the resting (PDB IDs: 1NG2⁹², 1UEC⁹³) and the p22^{phox}-bound p47^{phox} state (PDB IDs: 1OV3 (dimer)⁹², 1WLP (monomer)⁴⁸). The inactive and the active conformation are compared in Figure 49.

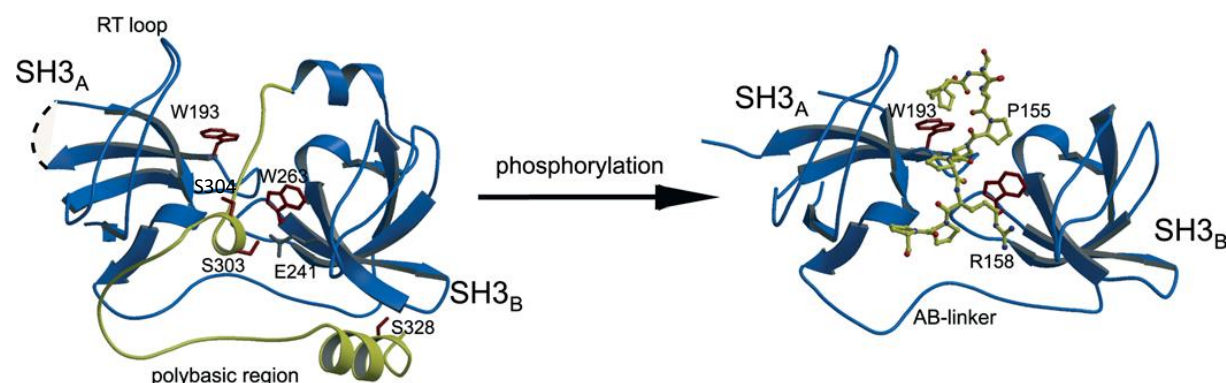


Figure 49: Regulation of p47^{phox} activity: In the monomer of the auto-inhibited p47^{phox} conformation on the left (PDB ID: 1NG2), the blue tandem SH3 domains are masked from binding by the yellow polybasic region. The core of this region interacts with Trp193 and Trp263 (red). Phosphorylated Ser residues leading to the active state by a conformational change are shown in red as well (Ser303, Ser304, and Ser328). On the right, the same tandem SH3 domain is bound to the yellow C-terminal p22^{phox} peptide (residues 151-160, PDB ID: 1OV3). The Figure was adapted from Groemping et al.¹⁰.

In the inactive structure, the sequence Gly-Ala-Pro-Pro-Arg, which is involved in the polybasic/autoinhibitory region (PBR/AIR), forms a left-handed polyproline type II (PPII) helix (detailed and compared to PPI in Table 20) and interacts simultaneously with each SH3 domain⁹⁴.

Autoinhibition is also a common feature of inactive kinases, where the juxtambrane (e.g. in c-KIT ⁹⁵) or a myristoyl group (e.g. in c-Abl ⁹⁶) exert this function.

Table 20: For polyprolines, the two stable secondary structures are helices of either the polyproline type I (PPI) or type II (PPII).

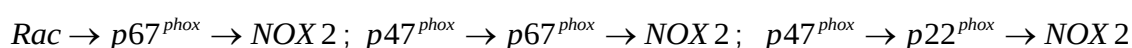
Type	Conformation	Å per turn	Residues per turn	Direction of turns	Peptide bonds	Backbone dihedrals	Occurrence
PPI	Compact	5.6	3.3	right-handed	all cis	-75, 160, and 0°	rare in a biological context
PPII	Extended	9.3	3	left-handed	all trans	-75, 145, and 180°	common in folded proteins and unfolded polypeptides

Phosphorylation

Phosphorylation of p47^{phox} has extensively been shown to be essential for activation ^{25,97-102}. In total, 11 phosphorylation sites are involved in activation. Ser359 and Ser370 initiate activation and precede the phosphorylation of other Ser residues. The autophosphorylation sites of p47^{phox} lie in non-catalytic, cytoplasmic regions. In general, short-term activation happens through phosphorylation of subunits, while long term activation is triggered by enhanced subunit expression. Candidate enzymes for p47^{phox} phosphorylation include many kinases ¹⁰³⁻¹¹⁰. Nevertheless, phosphorylation of certain residues may also lead to deactivation of p47^{phox} ¹¹¹. A common Ser/Thr kinase casein kinase 2 (CK2) inhibitor for example potentiated NADPH oxidase activity, pointing to the role of CK2 as regulatory kinase of NADPH oxidase deactivation ¹¹².

Today's view of activation

Today, p47^{phox} is seen as “positive effector” ^{113,114} or “organizer” of activation, while more importance is given to either direct Rac binding ⁶⁶ or binding of p67^{phox}, the “activator”, to NOX2 ¹¹⁵. The p47^{phox}-p22^{phox} interaction and Rac serve as anchor to p67^{phox}, which does not have a membrane attachment signal, enabling its interaction with NOX2 ^{116,117} (Figure 48). But the NADPH oxidase can also be activated *in vitro* by p67^{phox} and Rac only ^{118,119}. Mizrahi et al. ¹¹⁶ summarize conditions under which p47^{phox} is obsolete for activation ¹¹⁸⁻¹²¹ and propose a “propagated wave” model essentially telling that conformational changes occurring in one component of the NADPH oxidase are propagated to the next, with possible sequences being:



Further support of the importance of p67^{phox} is provided by the Leu505Arg mutation in a cell-free NOX2 system ⁴², a complementary 3D-model of the apo C-terminal NOX2 ¹²² and the corresponding crystal structure (PDB ID: 3a1f). Leu505 occupies a strategic position at the entry of the NADPH binding site at the end of an α -helical loop (residues 484-504), which is a p67^{phox} binding site. Mutating this Leu into an Arg abolishes the stabilization of this loop by p67^{phox} and the NOX2 substrate NADPH cannot enter anymore. Additionally, a monoclonal antibody binding to residues 498-506 seems to inhibit initial electron transfer ¹⁶, corroborating the role of the α -helical loop as gate towards NADPH ⁴².

3.2 TARGETING SH3 DOMAINS

Many SH3-polyproline interactions occur between the cytosolic triad p47^{phox}, p67^{phox}, and p40^{phox}. Given the crucial role of SH3 domains in NOX assembly, their ligand preferences, structures, and functions are reviewed. It is assumed that SH3 domains are the most abundant molecular recognition modules of the proteome ¹²³. According to a publication in 2007, there are over 800 SH3 domains present in mammalian genomes ¹²⁴. Dalgarno et al. provide in-depth information on these domains in drug design ¹²⁵.

3.2.1 FACE-TO-FACE WITH SH3 DOMAINS

With a size of 55 to 70 amino acids, these non-catalytic domains mediating protein-protein interactions are quite small ¹²⁶. They are composed of a 5-stranded β -sandwich, in which a kink in the second of the five β -strands delimits the two roughly perpendicular β -sheets. A 5-residue long 3_{10} -helix separates the fourth from the fifth strand (Figure 50).

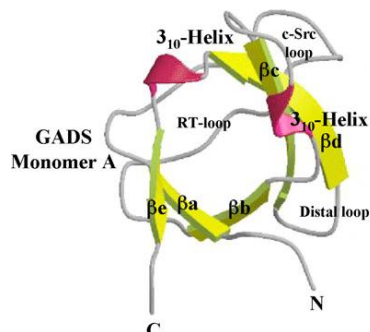


Figure 50: Example of a SH3 domain taken from ¹²⁴.

The first SH3 domain was discovered in the tyrosine kinase of v-Src, an oncogene, and was shown to interact transiently with other modulatory domains and the catalytic site. In Src, SH3 disruption leads to the activation of the tyrosine kinase. SH3 is thought to directly or indirectly bind substrates and to transport them to the kinase domain for phosphorylation. In this way, SH3 domains can influence enzymatic activity and interactions with the cytoskeleton or membrane. Based on these functions, SH3-containing pharmaceutical targets are found predominantly in cancer, immunological, and neuronal diseases.

SH3 domains are well-known to bind peptides having a conserved Pro-rich core adopting a PPII conformation, displaying weak interactions ($K_d = 5\text{--}100\ \mu\text{M}$) with a highly aromatic surface binding site, and low selectivity. Binding strength is selectively improved by surrounding amino acids, which increase the size and complexity of the peptides. There are three relatively shallow binding subpockets ¹²⁷: the acidic specificity pocket, and two pockets accommodating a dipeptide composed of Pro and a hydrophobic residue (Ala, Ile, Leu, Pro or Val) ¹³⁰. The first pocket is held in place by two loops ¹²⁸, the Arg-Thr (RT) (Figure 50) and the n- or c-Src loop (Figure 50), which assist in the creation of the binding surface ¹²⁹ and can be exploited to gain affinity. The three mentioned sites form grooves separated by surface ridges. Positively charged Lys or Arg residues provide additional electrostatic anchoring in the specificity pocket and orient the ligand in the binding site ^{128,131,132}. Longer peptides or proteins in contact with a SH3-domain may benefit from core-remote, less homologous contacts to achieve a better binding.

The versatility of SH3 domains is beautifully illustrated by NOX2. The tandem SH3 domains of p47^{phox} bind the polybasic region (PBR, Figure 49). The PBR has only binding affinity to both SH3s ($K_d = 1.5\ \mu\text{M}$), with multiple Arg and Lys decreasing the affinity by a factor of 20 when omitted ⁹². The specificity of the SH3 interactions of p47^{phox} was studied with mutations of critical residues. De Mendez et al. found that the first SH3 is responsible for the p47^{phox}-p22^{phox} interaction and that the second SH3 interacts with the N-terminal p67^{phox} ^{133,134}.

3.2.2 REASONS FOR THE VERSATILITY OF SH3 DOMAINS

SH3 domains are intrinsically promiscuous because they can bind two different peptide classes. The distinction between these two classes is based on the sequence orientation (N- or C-terminal first) of SH3-binding peptides, and gives rise to two “reverse” ¹³¹ binding modes. As an example, peptides with a class I (plus) orientation binding to Src display an Arg in the N-terminal as opposed to peptides with a class II (minus) orientation with an Arg located at the C-terminus. More generally, they have an XPpXP motif in common (where X is any amino acid, P a Pro contacting the SH3, and p a Pro promoting the PPII helix formation). The presence of SH3 domains allows thus accessing different proteins in a multi-protein complex.

3.2.3 REQUIREMENTS FOR SH3 DOMAIN-BINDING LINEAR PEPTIDES

With these comments in mind, requirements for micromolar linear peptides binding to SH3 can be summarized according to Dalgarno in the following way: (i) A high Pro content to promote the formation of the PPII helix, (ii) complementarity between flanking non-core residues and non-hotspot SH3 portions, (iii) and optimized SH3-peptide interactions within the hotspot/Pro-rich core ¹²⁵. A tool for predicting *in silico* (prior to wet-lab experiments) whether a peptide potentially binds to SH3 domains has been developed: The neuronal network-based SH3-Hunter (<http://cbm.bio.uniroma2.it/SH3-Hunter/>) ¹³⁵. It is however restricted to natural amino acids.

3.2.4 IMPROVING PEPTIDE SPECIFICITY

Once micromolar affinity is attained, how can SH3 ligand-binding specificity be improved? Based on an analysis of the structural requirements for ligand specificity in SH3 domains and mutagenesis of the SH3 belonging to the Fyn tyrosine kinase, Fernández-Ballester et al. identified the orientation of a Trp (the “Trp switch”) to condition peptide class specificity ¹³⁶. With a single Tyr132Ile mutation in Fyn, a switch in the peptide orientation from class I to class II can be obtained. Aromatic residues at this position (e.g. Tyr), allow for the motion of Trp and promote promiscuous binding, while β -branched or long aliphatic amino acids (e.g. Ile or Met) block the Trp orientation and promote class II binding. Thus, the Trp switch contributes essentially to the versatility and promiscuity of SH3 domains.

3.2.5 SELECTIVITY AND SPECIFICITY IN PROTEIN-PROTEIN AND PROTEIN-SMALL MOLECULE INTERACTIONS

Both peptide (3.5.2) and small molecules NOX inhibitors (3.2.4) are known to date. A general strategy to improve selectivity is based on the fact that the pre-organization of a PPII helix can be mimicked by a rigid moiety. The approach consists in adding a carbon bridge between the Pro rings using a 5,6-spirolactam or an even more rigid tricyclic spirolactam (Figure 51) ¹³⁷.

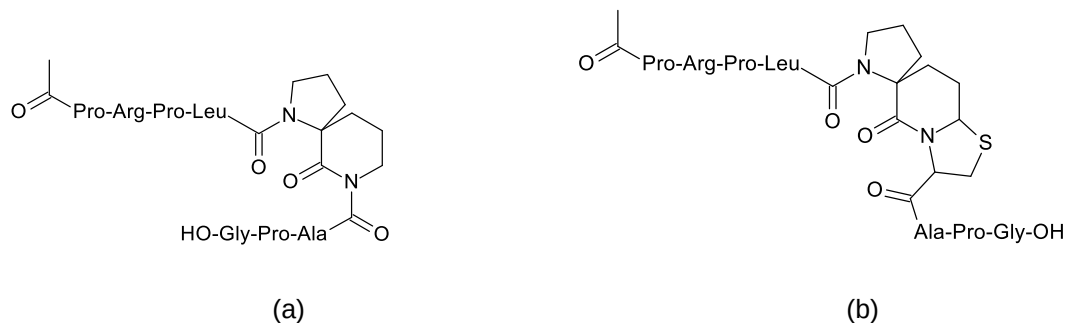


Figure 51: Mimicking the pre-organization of a PPII helix by 5,6-spirolactams (a) or tricyclic spirolactams (b).

Although the overwhelming majority of ligands for SH3 domains are peptide-based, these domains are druggable by small molecule inhibitors, as has been demonstrated by Inglis et al. ¹³⁸ by an *in silico* screen in the mouse Tec kinase SH3. The binding site was identified with NMR chemical shift perturbation, SARs, and site-directed mutagenesis as the one where the Pro-rich peptide sequences usually bind. The most potent heterocycle, a 2-aminoquinolone with a K_d of 125 μ M successfully competed against a peptide in a peptide displacement assay based on fluorescence polarization. The largest NMR ^1H (H-N) chemical shift was observed for the Trp215 side chain, which functions as Trp switch (3.2.4). Substitutions of 2-aminoquinolines at position 5, 6 and 7 had similar, better and lower affinity compared to unsubstituted 2-aminoquinolines, respectively ¹³⁹. For a 6-substituted-N-benzylated-2-aminoquinoline, some evidence for the formation of additional contacts at other regions of the SH3 domain was found, which renders aminoquinolones promising for targeting SH3 domains in general.

3.3 THE IMPORTANCE OF PROLINE

One of the main amino acids implicated in protein-protein interactions is Pro. The intermolecular interaction between a Pro-rich sequence and a protein is often achieved via phosphorylation or conformational adjustments. Two extensive reviews ^{140,141} report on Pro-rich motifs interacting with Pro-rich domains. Pro-rich motifs with specificity-determining “aromatic cradles” ¹⁴², with the distances and angles between the side chains determining the Pro-binding site shape ^{142,143}, are recognized by six distinct protein families adopting two different binding modes: binding mode 1 is represented by the SH3 ^{144,145}, the EVH1 (Ena/Vasp Homology domain 1) ¹⁴⁶⁻¹⁴⁸, and the single-domain profilins ^{149,150}, whereas binding mode 2 is found in the WW (containing 2 conserved Trps) ^{143,151}, the GYF ¹⁵² (containing a conserved Gly, Tyr, and Phe), and the UEV (Ubiquitin E2 Variant) ^{153,154} domains. In binding mode 1, the peptide PxxP forms an umbrella-like structure around the exposed H-bond donor Tyr or Trp side chain and both forward and reverse peptide orientations can be observed ¹⁴⁰. In binding mode 2, binding of the domain happens on the rear of the PPII helix ¹⁴⁰. Many crystal structures of SH3 domains in complex with Pro-rich peptides exist ^{131,132,144,155-161}.

The unique characteristics of Pro are conferred by its cyclised side chain conformation and the bulk of the N-substituent, yielding a backbone ϕ angle around -65° ¹⁶². A PPII helix has three residues per turn, with the Pro residues on the same, protein-interacting side. The helix itself has an amphipathic character, with the carbonyls being electron-rich, rigid H-bond donors, and the Pro residues being hydrophobic. Provided these examples, the PPII helix seems to fit the steric and hydrogen-bonding pattern of the SH3 peptide-binding surface best ¹³¹.

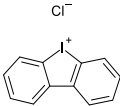
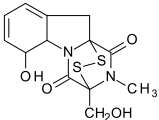
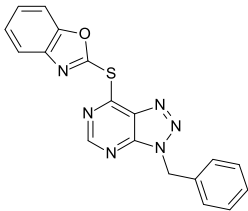
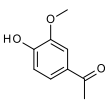
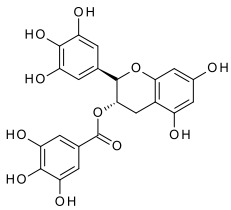
Attempts to stabilize Pro PPII helices have been published. 4(R)-fluoroPro enhances and 4(S)-fluoroPro decreases PPII helix stability ¹⁶³. The underlying reason is that depending on the ring puckering, either the trans (4-(R)-fluoroPro) or the cis (4-(S)-fluoroPro) amide bond is stabilized. More recently, the incorporation of N-alkyl-substituted amino acids mimicking Pro residues has been shown to improve binding affinity ¹⁶⁴.

3.4 SMALL MOLECULE NOX INHIBITORS

Non-peptide inhibitors have generally a better stability and *in vivo* administration profile than peptide inhibitors. A systematic approach to the discovery of truly selective NOX inhibitors is presented by Jaquet et al., together with a review on small molecule inhibitors with respect to structures, activities, and *in vitro/in vivo* specificities ¹⁶⁵. Most NOX inhibitors act on NOX2 and lack specificity. Nevertheless, they are useful tools to investigate e.g. on electron transfer ¹⁶⁶, such as diphenyleneiodinium (DPI), quinacrine or 5-deaza-FAD (flavoprotein inhibitors), bipyridyl or benzimidazole ¹⁶⁷ (heme ligand) or ADP (a NADPH analogue). The specificity of these over 350 ⁵⁹ inhibitors is also hampered by the wide tissue distribution of NOX2 together with other NOX isoforms, their effect on upstream pathways and their action on both NOXes and other targets (e.g. other flavin-dependent enzymes) ¹⁶⁵. Indirect, i.e. upstream inhibitors of NOX are angiotensin receptor 1 blockers like valsartan ¹⁶⁸ and nebivolol ¹⁶⁹. To be able to distinguish real NOX inhibitors from upstream effectors *in vitro*, a whole series of assays are required, which are described in detail in ¹⁶⁵ (i.e. dose-dependent inhibition of ROS production in intact cells, cell viability, antioxidant capacity, inhibition of oxygen consumption, dose-dependent inhibition of ROS in cell-free assays and potential off-target effects e.g. via phosphorylation by kinases).

Inhibitors are either direct or indirect inhibitors of the NOX catalytic subunit. Direct inhibitors with most of the above cited assays include aryl iodine derivatives ^{170,171}, thiol-modifying compounds ¹⁷²⁻¹⁷⁴, molecules of plant or fungal origin ¹⁷⁵⁻¹⁷⁷ or endogenous compounds (i.e. produced by the human body) ^{169,178,179}. Examples of indirect inhibitors can be found in the literature ¹⁸⁰⁻¹⁸⁴. For many of the molecules published so far by companies ^{165,185-193}, biological activity data are not fully available to the public. A very general overview of the diversity of NOX inhibitors is provided in Table 21.

Table 21: The diversity of direct and indirect NOX inhibitors. A question mark is put for the inhibitor EGCG because it has not been proven yet that it is an indirect inhibitor.

Structure					
Name	DPI 170	Gliotoxin	VAS2780	Apocynin	EGCG
Origin	Aryliodonium derivative	Fungal origin	Vasopharm (proprietary compound)	Plant origin	Plant origin, frequent in green tea, "nutriceutical"
Inhibitor type	Direct inhibitors			Indirect inhibitor	Indirect inhibitor(?)

3.4.1 (-)-EPIGALLOCATECHIN GALLATE

The molecule (-)-epigallocatechin gallate (EGCG) drawn in Table 21 has a potential as "nutriceutical". The reasons for it together with an unpublished potential mechanism of action are explained in the following. EGCG is the most abundant biologically active polyphenol in tea, the globally most consumed beverage. The antioxidant properties of tea are well-documented¹⁹⁴⁻¹⁹⁷. EGCG is mainly found in the hot, water-soluble fraction of unfermented green tea leaves, but is instable due to auto-oxidation and epimerization¹⁹⁸. It is known that (-)-epicatechin (EC) scavenges O_2^- ¹⁹⁹⁻²⁰¹, while its 3'- or 4'-O-methylated metabolites prevent O_2^- generation through inhibition of the endothelial NADPH oxidase activity. The 3'-isomer is twice as potent as the 4'-isomer, and both are stronger inhibitors than apocynin²⁰¹. Contrarily to the scavenger EC, 3'-, 4'-O-methylated EC, and apocynin all contain a mono-methylated catechin, which seems to be crucial for NOX inhibition. Similar inhibitory properties can be obtained with O-methylated-EGCG²⁰². Interestingly, when comparing EC with EGCG, the additional gallate in EGCG causes also NOX inhibition²⁰². Thus, both mono-methylation and/or a gallate substituent on EC render the molecule a NOX inhibitor.

3.4.2 CELASTROL

Celastrol has been found to inhibit NOX activity in a cellular screen (*in vitro* pulldown, $IC_{50} = 20 \mu M$) of a NINDS (National Institute of Neurological Disorders and Stroke) library of FDA-approved compounds. This screen has been performed at the Department of Immunology and Pathology of the University of Geneva.

With a chemical formula of $C_{29}H_{38}O_4$, celastrol's molecular weight is in the order of 450.6 Da. It features two H-bond acceptors and two H-bond donors. As a quinone-methide triterpene, celastrol is composed of a phenolic quinone, a triterpene and a carboxyl part, respectively (Figure 52).

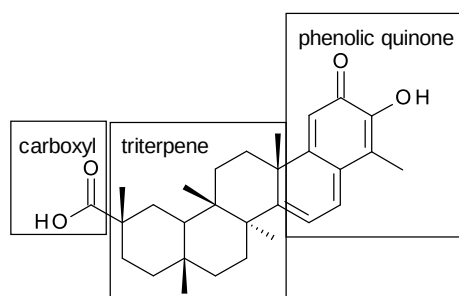


Figure 52: Substructures of celastrol, a quinone methide triterpene.

Celastrol is a constituent of *Tripterygium wilfordii* extracts, which are used in traditional Chinese medicine as anti-inflammatory remedy. With its well-known beneficial pharmacological properties in oxidative stress-related pathologies, including a strong neuroprotective effect in a mouse model of amyotrophic lateral sclerosis ²⁰³, it is worth pursuing the investigation on celastrol's NOX inhibition. Furthermore, this compound is a protein-protein interaction breaker, as it acts as proteasome inhibitor not only leading to the accumulation of proteasome target proteins, but also degrading heat shock protein 90 (Hsp90) client proteins. As an example, it inhibits the Hsp90-Cdc37 complex, exhibiting thereby *in vitro* and *in vivo* activity against pancreatic cancer ²⁰⁴. Initial binding hints in this study came from docking studies, which were then however refined with binding free energy calculations based upon molecular dynamics trajectories. The celastrol conformation with the most favourable binding energy in Hsp90 had a ΔG of -18.8 kcal/mol (not considering the entropy), with the phenolic quinone in a polar groove, and the hydroxyl as well as the carbonyl pointing in a polar charged pocket forming two H-bonds. One residue implied in H-bonding also formed a π - π stacking interaction with celastrol. These general features are also found when docking celastrol into p47^{phox} (3.7.3). In the study cited, celastrol induced a major conformational change disrupting an essential H-bond between Hsp90 and its cochaperone.

3.5 PEPTIDOMIMETICS

Peptides which include non-natural amino acids are called peptidomimetics. They are developed with the aim of increasing the binding affinity by enlarging the contact repertoire with the target protein. Amino acid mimetic structures of therapeutical relevance include Arg mimetics binding to serine proteases and integrin receptor antagonists used for the treatment of blood coagulation disorders ²⁰⁵ as well as Pro-Arg mimetics that are selective thrombin inhibitors ²⁰⁶.

Peptides are generally characterized by bad pharmacokinetic properties. To increase their bioavailability, they have to be coupled to amino acids facilitating the import into cells (e.g. TAT) ²⁰⁷. Peptidomimetics are "depeptized" by getting rid of the peptide bonds to make them less susceptible to enzymatic degradation.

3.5.1 PEPTIDOMIMETIC STRATEGIES ADDRESSING SH3 DOMAINS

Peptidomimetic strategies addressing SH3 domains focus on the core binding sequence PxxP. It is well known that oligomers containing Pro analogues are able to adopt a PPII helix ^{208,209} and that a Glu close in sequence to PPII helices is common and promotes helix formation. Peptides longer than four amino acids also frequently contain Gln, Arg, Ala, Leu, Ser, Asp, and His ²¹⁰. As an example of a stabilizing mutation, Ala can be replaced by α -aminobutyric acid (AIB; Tables 28, 32, and 34), which has a higher rigidity and a more helical character. Substituting Pro by N-substituted Gly is possible ²¹¹. As a general rule, the replacement of Pro by such non-natural amino acids results in low K_d values (as low as 30 nM), with a gain of three orders of magnitude in selectivity for different SH3 domains ^{211,212}. The affinity improvements are achieved through hydrophobic contacts not present with Pro alone. Also, non-peptide substituents binding to the specificity pocket or the terminal pockets have been used ^{130,213}, where the peptide part served as anchor and the non-peptidic part as "umbrella" covering the surface and improving binding affinity.

3.5.2 CURRENT KNOWLEDGE ON PEPTIDES ADDRESSING THE NOXES

Cell-permeable NOX inhibiting peptides include gp91-dstat, Rac, and PR-39 peptides. They have been reviewed by Selemidis et al.²¹⁴. Only PR-39 is discussed further, being the most similar to the peptidomimetics designed here. Like other antimicrobials adopting a PPII helix, PR-39 is able to cross cell membranes. It is a Pro- and Arg-rich (PR) peptide of 39 amino acids in length, binds to the cytosolic p47^{phox} SH3 domains, inhibits the interaction with p22^{phox}, and blocks thereby NOX assembly²¹⁵. PR-39 has also been reported to stimulate angiogenesis²¹⁶, to inhibit inflammatory responses²¹⁷⁻²²⁰ by specifically blocking the degradation of some proteins, and to be an anti-apoptotic factor²²¹.

Zooming on the p47^{phox}-p22^{phox} interface, the question is whether a “protein recognition code” for p47^{phox} exists, i.e. a finite set of precise rules governing this interaction²²². Prior to the knowledge from crystal structures, p22^{phox} GST fusion proteins of variable length were used to tackle the importance of the Pro156 residue²⁰. A strong binding to p47^{phox} was observed with residues 145 to 170 of p22^{phox}, with the 10-amino acid Pro-rich stretch (Pro-Pro-Ser-Asn-Pro-Pro-Arg-Pro-Pro) resembling a peptide binding to the tyrosine kinase Abl (Ala-Pro-Thr-Met-Pro-Pro-Pro-Leu-Pro-Pro)²²³. Using the 10-residue peptide only instead of residues 145 to 170, a weaker binding was observed. This was explained by an altered 3D structure or the absence of an interaction further stabilizing the interactions with the tandem SH3 domain. A complete loss of binding was observed with the Pro156Gln mutation in this peptide, demonstrating that the interaction between Pro156 and p47^{phox} is direct. As can be seen from Figure 53, this “anchoring” Pro is crucial for maintaining the complementarity-determining geometry with the tandem SH3 domain in the crystal structure.

In the wild type peptide, residues 154 to 158 in p22^{phox} form a PPII helix characterized by a triangular prism-like structure: While the basal plane formed by residues Asn154, Pro156, and Pro157 contacts the N-SH3, the basal plane formed by Pro155, Pro156, and Arg158 contacts the C-SH3; Pro156 is located at the apex of the prism and interacts simultaneously with N-SH3 and C-SH3²²⁴. Due to this interaction pattern and its location in the middle of the PPII helix, a particular role in complex formation and stabilization of Pro156 is probable. Mutating it is sufficient to interrupt the p22^{phox}-p47^{phox} interaction. Comparable observations have been made for similarly exposed PPII helices²²⁵. The remaining Pro residues in the peptide are thought to further stabilize the helical structure, in accordance with Yu et al.¹⁴⁴, and to be responsible for the small loss in conformational entropy upon binding, reflected in an entropic stabilization of ca. 1 kcal/mol per residue^{226,227}.

A 1 ns molecular dynamics on one of the p47^{phox}-p22^{phox} crystal structures (PDB ID: 1OV3,⁹²) provided a convenient way of studying the effect of H-bonding residue mutations on the H-bonding patterns and complex stability. In their study, Watanabe et al.²²⁴ investigated on residues stabilizing the n-Src loops in the N-SH3 and the C-SH3 as well as forming H-bonds to Asn154 and Pro156 of p22^{phox}. The Gly262Ser mutation in p47^{phox} did not dramatically change the H-bonding distance between Trp263 of p47^{phox} and Pro156 of p22^{phox}. In contrast, Gly192Ser in p47^{phox} led to an H-bond distance increase between Trp193 (p47^{phox}) and Asn154 (p22^{phox}) compared to the wild type. These two results are in agreement with ITC measurements⁹². Also for the Pro156Gln mutation in p22^{phox}, the relative distance between Trp263 (p47^{phox}) and Gln156 (p22^{phox}) was higher. The monitored interaction energy in the wild type p22^{phox} with N-SH3 increased proportionally to simulation time, but decreased for the three mutants, while it remained stable in all cases for the tandem SH3 and the C-SH3. Thus, the authors concluded that inactivation of NADPH may be caused by a lack of interaction between the N-SH3 and p22^{phox}.

3.5.3 INSIGHTS FROM PRIOR PEPTIDE DOCKING OF P22^{PHOX} AND ITS ANALOGUES

Prior to this thesis, p22^{phox} and its analogues were docked into the tandem SH3 domain of p47^{phox} with the program EADOCK²²⁸ at the University of Lausanne (Switzerland), with the aim of predicting amino acid mutations that would be possible in p22^{phox} from a steric point of view. EADOCK placed the native sequence within 1 Å heavy atom RMSD to the crystallized structure, which is a considerable achievement for protein-protein docking. The amino acid mutations that were suggested are depicted in Table 22. The hydrophobic pocket around Pro151-Pro152 could be filled with different hydrophobic residues. Regarding the triproline region (Pro156-Pro157-Pro158), the Pro155 subpocket accepted a wide range of hydrophobic residues, but also Cys and His, whereas the Pro156 subpocket could only accommodate Ala or Cys, and the Pro157 subpocket was too narrow for substitutions by natural

amino acids. Unsurprisingly, the specificity pocket around Arg158 could only harbour the similarly shaped, positively charged amino acids Arg and Lys.

Table 22: Amino acid mutations identified by peptide docking with EADOCK. The sequence numbering is indicated in bold and the wild-type p22^{phox} sequence is underlined.

150 <u>Gln</u>	151 <u>Pro</u> Ile Leu Ala	152 <u>Pro</u> Met Ile Ala Val	153 <u>Ser</u>	154 <u>Asn</u>	155 <u>Pro</u> Ile Met Leu Cys His Ala	156 <u>Pro</u> Ala Cys	157 <u>Pro</u>	158 <u>Arg</u> Lys	159 <u>Pro</u>	160 <u>Pro</u> Ala
--------------------------	---	--	--------------------------	--------------------------	--	--	--------------------------	---------------------------------	--------------------------	---------------------------------

As a continuation of the EADOCK work, the work presented here aims at ameliorating the IC₅₀ values of p22^{phox} analogues with non-natural amino acid substitutions. Such substitutions are geared towards improving the pharmacokinetic properties. The work on the p47^{phox}-p22^{phox} interface presents a novelty in the sense that none of the known NOX2 inhibitors targets this interface. The features of this interface are presented in the following.

3.6 TARGETING THE P47^{PHOX}-P22^{PHOX} INTERACTION

3.6.1 INSIGHTS FROM CRYSTAL STRUCTURES

As mentioned in 3.1.4, both p47^{phox} and p22^{phox} are required for activation. Several crystal structures of either the autoinhibited p47^{phox} tandem SH3 domains or the activated p47^{phox}-p22^{phox} complex exist (Table 23).

Table 23: Crystal structures of the autoinhibited p47^{phox} tandem SH3 domains or the activated p47^{phox}-p22^{phox} complex.

PDB ID	Ref.	State	Resolution [Å]	Residues
1NG2	92	autoinhibited p47 ^{phox}	1.7	p47 ^{phox} : 156-340
1UEC	93		1.82	p47 ^{phox} : 151-340
1WLP	48	active p47 ^{phox} -p22 ^{phox} complex, monomer	solution NMR structure	p47 ^{phox} : 149-168; p22 ^{phox} : 144-168
1OV3	92	active p47 ^{phox} -p22 ^{phox} complex, crystallographic dimer	1.8	p47 ^{phox} : 150-283/152-283; p22 ^{phox} : 150-160 (two times)

Comparing the active state structures there is one monomer and one dimer structure, both encompassing the PRR of p22^{phox} (residues 151-160). The main PRR-p47^{phox} interactions are depicted in Figure 53. The monomer structure, including also the regions of p22^{phox} outside the PRR, was particularly useful for seeing that (i) the central portion of the p22^{phox} peptide forms a PPII helix sandwiched between the N- and C-terminal SH3s and (ii) the C-terminal p22^{phox} peptide adopts a short α -helical conformation (residues 161-164), further anchoring the peptide in the N-terminal SH3, increasing binding affinity over 10 times ⁴⁸. The Ala163Glu/Arg164Glu double mutation destabilizes this α -helix ⁴⁸.

Figure 53 shows the H-bonds and hydrophobic interactions between p22^{phox} and the N- and C-terminal portion of p47^{phox}. Particularly, p47^{phox} displays two hydrophobic grooves in which diprolines bind (Pro151-Pro152 and Pro159-Pro160) and one specificity pocket encompassing the triproline motif (Pro155-Pro156-Pro157) and Arg158.

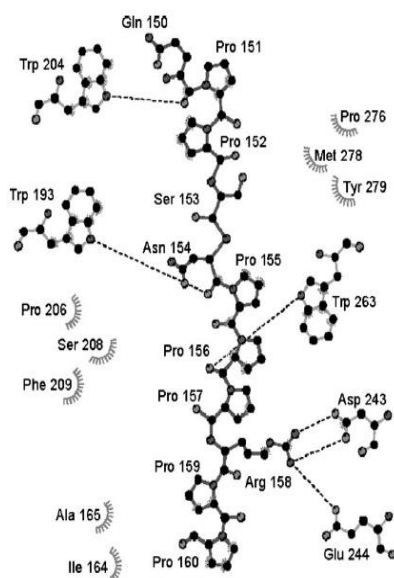


Figure 53: H-bonds (dotted lines) and hydrophobic interactions (“cilia-like”) of the PRR of p22^{phox} (Pro151-Pro160) with the N-SH3 (on the left) and the C-SH3 (on the right) of p47^{phox}. The Figure was taken from ²²⁴.

From the exhaustive listing of the different p47^{phox}-p22^{phox} interaction types in Table 24, it can be seen that the N-terminal SH3 (N-SH3) participates in more interactions than the C-terminal SH3 (C-SH3), rendering it more important for p22^{phox} stabilization, but that the central Pro156 in p22^{phox} interacts with both domains (two hydrophobic interactions with the N-SH3 and the C-SH3 as well as one H-bond with the C-SH3).

Table 24: Different p47^{phox}-p22^{phox} interaction types (H-bonds and hydrophobic interactions) and residues involved.

Type	N-SH3	p22 ^{phox}	Type	C-SH3	p22 ^{phox}
Hydrophobic	Val186, Phe195, Trp204	Lys149-Pro152	Hydrophobic	Pro276, Met278, Tyr279	Pro151-Ser153
H-bond	Trp204	Gln150	H-bond	Trp263	Pro156
H-bond	Trp193	Asn154	Hydrophobic	Tyr237, Trp263, Tyr274, Pro276, Met278, Tyr279	Pro155, Pro156, Arg158
Hydrophobic	Ile164, Ala165, Tyr167, Trp193, Pro206, Ser208, Phe209	Pro156, Pro157, Pro159, Pro160	H-bond	Asp243, Glu244	Arg158

The two SH3 domains of p47^{phox} are related by a twofold pseudosymmetry ^{127,131}. Nobuhisa et al. demonstrated that the deletion of the linker connecting the two SH3 domains (AB-linker in Figure 49) resulted not only in a defective p22^{phox} binding, but also in a loss of activity ²²⁹. In the same publication, they also showed by alanine scanning mutagenesis that Pro152, Pro156, and Arg158 in the PRR of p22^{phox} ^{20,230,231} are indispensable for interacting with p47^{phox}. All three residues make hydrophobic contacts, and Pro156 as well as Arg158 are additionally involved in H-bonds. Similarly, mutating Trp263 to Ala in the C-terminal SH3 of p47^{phox} resulted in a significant loss of the ability to support NADPH oxidase activation ^{47,134,232} due to the loss of an H-bond and hydrophobic interactions.

Thus, Pro152, Pro156, and Arg158⁹² are essential for binding, while the other prolines (Pro155, Pro157, and Pro160) play a rather modest role²²⁹. On the p47^{phox} side, residues 151 to 284, containing both SH3 domains, are required for translocation and activation¹³³.

Generally, SH3-PRR interactions are weak with K_d s in the micromolar range^{123,129,141,233}. In the case presented here however, specificity and high affinity are conferred by the cooperative binding to a tandem SH3 domain²²⁹. In this manner, also the tail-to-tail attachment²³⁴ of p67^{phox} on p47^{phox}, which induces electron flow, is stabilized^{66,235,236}. Protein-protein interactions like the ones mentioned confine local redox signalling^{237,238}. Notably, the PRRs of p22^{phox} (residues 149-162 and 176-195) and p47^{phox} (residues 70-83, 338-351, and 358-371) display the same consensus sequence (XPXXPPPP ψ XP, with X standing for non-conserved, P for Pro, and ψ for hydrophobic) like Abl SH3-binding proteins and display also the PPRP (R for Arg) motif of proteins binding to the SH3 domain of the regulatory p85 subunit of phosphatidylinositol 3-kinase^{144,230}.

The “Trp switch” concept introduced earlier (3.2.4) for SH3s holds also true for the p47^{phox}-p22^{phox} interaction. Trp263 of the C-SH3 plays a key role in p22^{phox} recognition. In the autoinhibited state, it forms an H-bond to Gly297 of p47^{phox} and in the active state to Asn154 of p22^{phox}.

Based on this information, the p47^{phox}-p22^{phox} interface seems to be targetable via the hot spot represented by the diproline (Pro151-Pro152), the triproline (Pro155-Pro156-Pro157), and Arg158.

3.7 METHODS

To find inhibitors of the p47^{phox}-p22^{phox} interaction, both target-based and ligand-based VS were performed using commercially available small molecule libraries. The target-based screening was done on the p47^{phox} interface, while the ligand-based screening focused on a pharmacophore search with the diproline Pro151-Pro152 or the triproline Pro155-Pro156-Pro157 of p22^{phox} as template. Next, potential binding modes of the experimental screening hit celastrol were suggested based on docking. Peptidomimetics of p22^{phox} were then addressed with a GRID-based search and a hypothetical binding mode of the retropeptide of p22^{phox} in p47^{phox} was found.

3.7.1 STRUCTURE-BASED VIRTUAL SCREENING

Before docking a large compound database into a target, accurate chemical library filtering²³⁹ needs to be carried out. Using the standard settings of Filter 2.0 (OpenEye, Inc.), the Specs database containing 192423 compounds (release of June 2006) was filtered to discard non lead-like molecules. The differences between drug- and lead-like properties of molecules are listed in Table 25.

Table 25: Drug-like versus lead-like properties of molecules. Drug-likeness can be defined using Lipinski's “rule of 5”²⁴⁰, a rule of thumb defining upper thresholds of physicochemical properties as multiples of five and improvements thereof^{241 242}. Criteria for lead-like²⁴⁴ compounds are more stringent than for drug-likes.

Physicochemical properties	Drug-like	Lead-like
H-bond donors	≤ 5	≤ 5
H-bond acceptors	≤ 10	≤ 9
Molecular weight	≤ 500 Da	≤ 460 Da
logP	≤ 5	$-4 \leq \log P \leq 4.2$

The remaining entries were split, yielding 23611 molecules with low molecular weight (150-440 Da) and 21 with high molecular weight (440-1000 Da). This was done to be able to distinguish the docking behaviour of these two classes. Hydrogens and charges were assigned with Quacpacs (OpenEye, Inc.) and the 3D structures were obtained from 2D structures with Corina (Molecular Networks GmbH). For each filtered database entry, only one tautomer/conformer was generated, because adding multiple conformers or tautomers has been shown to increase the false positive rate (compounds which are mistakenly treated as active molecules)²⁴⁶.

After the ligand database setup, the p47^{phox}-based screening involved a calibration phase, a “rough” screening with the docking program Dock, and a refined screening of the 1000 best interaction-energy ranked compounds with FlexX to narrow the conformational search. On the FlexX poses, consensus scoring was done with the CScore module of Sybyl. If a docking solution scored among the first 20% of all solutions examined with five scoring functions, a consensus score of 5 was attributed. The scores obtained with each scoring function as well as the consensus scores were extracted together with the compound names to an Excel sheet.

For both DOCK 4.03 and FlexX, a protocol was developed and validated with the diproline (151-152) and the triproline (155-157) of p22^{phox}.

Docking evaluation criteria

The docked hits from the ligand-based and the structure-based VS were evaluated based on the ranking of the calculated interaction energy of the five scoring functions used for consensus scoring, or its normalization (per atom number), the consensus scores, and visual inspection. Both the diproline and triproline best-docked poses achieved a consensus score of 5 in each target- and ligand-based screening, indicating that the docking parameters chosen were reliable. With respect to visual inspection, the similarity to the interactions found in the p47^{phox}-p22^{phox} crystal structure (PDB ID: 1WLP) was checked, as the chances of finding competitive inhibitors are higher if at least some interactions are conserved. Additional favourable contacts not seen in the crystal structure were taken into account as well, while solutions with unfavourable VdW interactions with p47^{phox} were discarded. Only solutions with a consensus score ≥ 3 which were well placed in the binding site were kept.

3.7.2 LIGAND-BASED VIRTUAL SCREENING

The filtered Specs library was split into four equally sized parts for a pharmacophore search. A standard pharmacophore query was carried out in Catalyst (Accelrys) with the diproline or triproline motif of p22^{phox}, respectively. The dipeptide pharmacophore had to be discarded as being too small. Compounds in accordance with the triproline pharmacophore hypothesis were then treated as in the structure-based procedure, i.e. they underwent docking with Dock, ranking according to the interaction energy, docking of the 1000 best-ranked compounds with FlexX, and consensus scoring.

3.7.3 CELASTROL DOCKING

To assess whether celastrol could bind to different NOX isoforms and inhibit them, a combined *in vitro-in silico* study was performed, with the *in vitro* part carried out mainly by Dr. V. Jaquet at the Department of Immunology and Pathology of the University of Geneva. The methods for the *in vitro* part can be found in ²⁴⁷. For the *in silico* part, docking was performed in the tandem SH3 domain with the programs Surflex ²⁴⁸, GOLD ²⁴⁹, FlexX ²⁵⁰⁻²⁵², and DOCK ²⁵³⁻²⁵⁵. Except for the binding site definition, standard parameters were used. Calibration of each docking program with respect to the binding site definition was done by docking the triproline region of p22^{phox} into p47^{phox} and comparing it to the corresponding crystallized region, retaining only the docking programs able to dock the triproline within a C α atom RMSD ≤ 2 Å.

3.7.4 PEPTIDOMIMETICS DESIGN USING GRID

The libraries containing Pro, N-substituted Gly, and other amino acid substitutions considered for the peptidomimetics project are listed in Tables 26 to 28.

Table 26: Library of substituted Pro for the design of peptidomimetics.

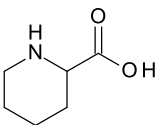
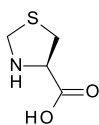
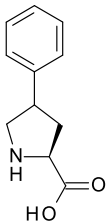
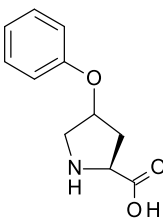
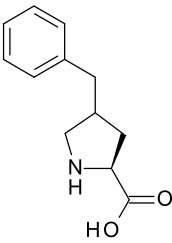
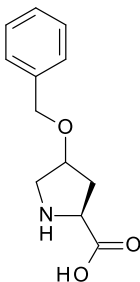
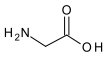
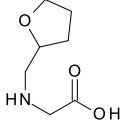
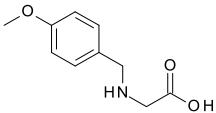
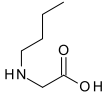
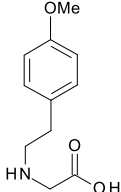
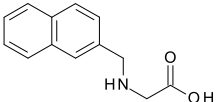
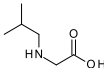
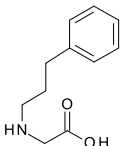
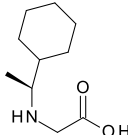
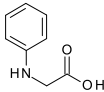
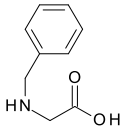
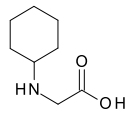
Pro	Structure	Pro analogue	Structure
Pipecolic acid (Pip)		ThioPro (PrS)	
4-phenylPro (PhP)		4-phenoxy Pro (PhOP)	
4-benzylPro (BzIP)		4-benzoxy Pro (BzOP)	

Table 27: Library of N-substituted Gly for the design of peptidomimetics (continues on next page).

N-substituted Gly	Structure	N-substituted Gly	Structure	N-substituted Gly	Structure
Gly (G)		N-(tetrahydrofurfuryl)-Gly (NfG)		N-(4-methoxybenzyl)-Gly (4mG)	
N-(butyl)-Gly (Nbu)		N-((R)-(+)-1-(4-methoxyphenyl)ethyl)-Gly (4fG)		N-(naphthylmethyl)-Gly (NmG)	
N-(isobutyl)-Gly (Ibu)		N-(phenyl-propyl)-Gly (NoG)		N(1-R-cyclohexylethyl) Gly (Nyg)	
N-(phenyl)-Gly (NpG)		N-(benzyl)-Gly (NzG)		N-(cyclohexyl)-Gly (NcG)	

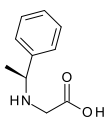
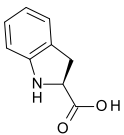
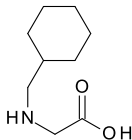
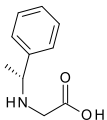
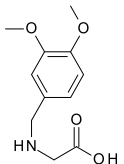
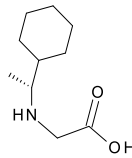
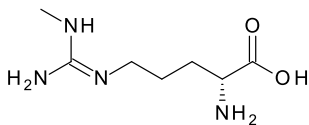
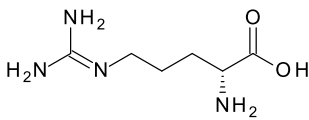
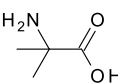
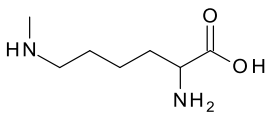
N-(R(+)) α -methyl, benzyl)-Gly (NrG)		L-indoline-2-carboxylic acid (Ind)		N-(cyclohexylmethyl)-Gly (NhG)	
N-(S(-)) α -methyl, benzyl)-Gly (NsG)		N-(veratryl)-Gly (NvG)		N(1-S-cyclohexylethyl) Gly (NxG)	

Table 28: Other amino acid derivatives considered.

Name	Structure
N-Me-Arg (MeR)	
Homo-Arg (hR)	
α -amino isobutyric acid (AIB)	
Me-Lys (MeK)	

A GRID search was performed on the p47^{phox} surface and a 3D GRID search box was placed around the binding site of interest. This box had the following dimensions (in Å): [-31, 37] for x, [-20, 19] for y, and [-27, 23] for z, corresponding to the atomic positions of OE2 of Glu241B and C α of Phe195B of p47^{phox} for x, C α of Gly172B and OH of Tyr274B of p47^{phox} for y as well as OG of Ser189B and OE2 of Glu241B for z. The types of probes used for obtaining the GRID maps were probes reflecting H-bonding properties (neutral flat NH₂, the alkyl OH, H₂O or the sp² carbonyl oxygen) or hydrophobic probes (CH₃ or a dry probe). Among these probes, H₂O is the smallest probe.

Retropeptide

To obtain ideas on further peptidomimetics, it was assessed whether a retropeptide of p22^{phox} could fit into p47^{phox}. Compared to the native sequence, its retropeptide has a reversed sequence order. The wild type (residues 149 to 168 of p22^{phox}) and the retropeptide were aligned manually so as to match the Pro residues. Features considered while trying to match the Pro residues of the retropeptide with the native p22^{phox} sequence included: side chain polarity (non-polar/polar), side chain acidity or basicity (acidic/neutral/basic), and the hydropathy index (according to Kyte and Doolittle ²⁵⁶). Based on the manual alignment, the wild type residues were mutated into the corresponding retropeptide residues using Sybyl.

3.8 RESULTS AND DISCUSSION

3.8.1 HITS DERIVED FROM THE STRUCTURE- AND THE LIGAND-BASED SCREENING

Using the whole p47^{phox} interface as active site, a clear preference for proline-binding pockets of p47^{phox} was observed for all ligands. With a more restrictive active site definition covering the diproline or triproline region of p22^{phox}, the best interaction scores with p47^{phox} were found to be located in the triproline region. As can be seen from Figure 54, triproline docking (b) was more accurate in reproducing the respective crystal structure conformation than diproline docking (a).

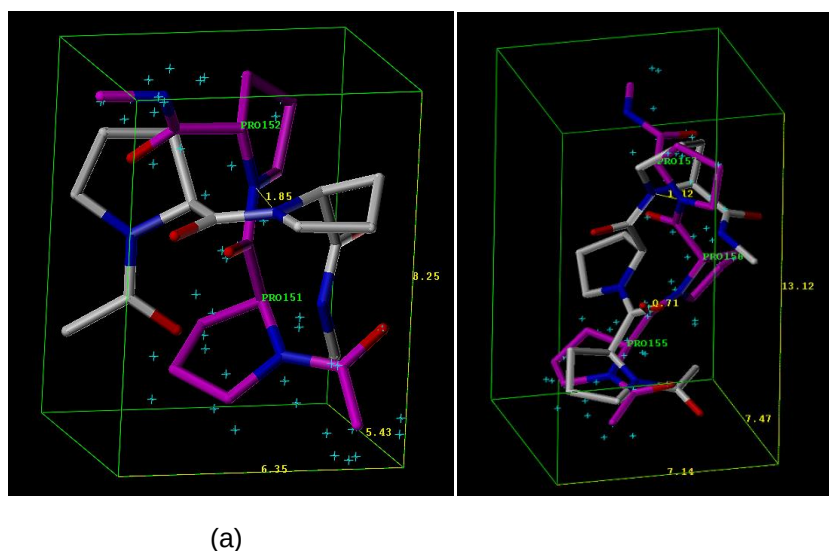
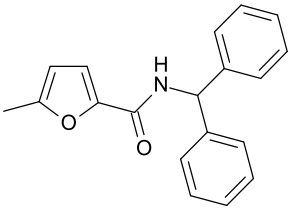
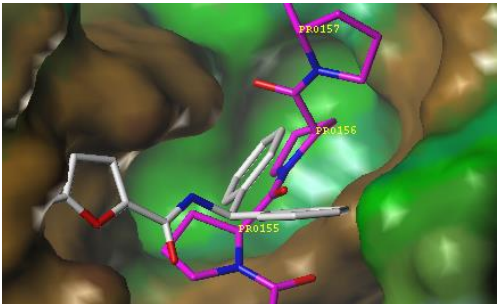
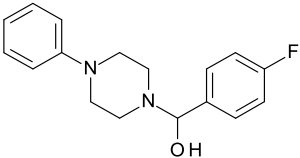
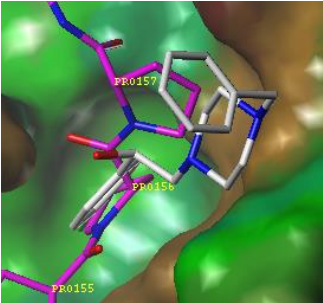
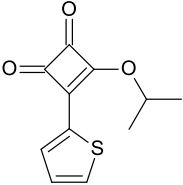
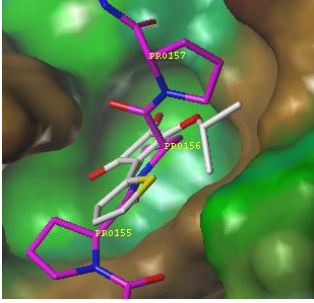
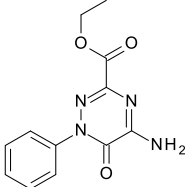
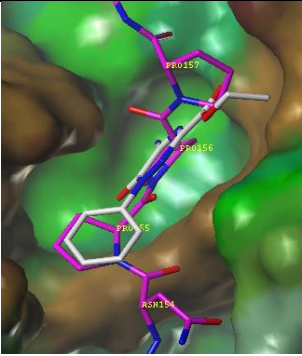


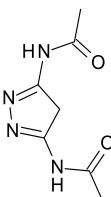
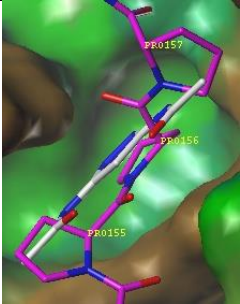
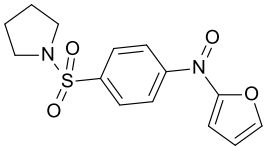
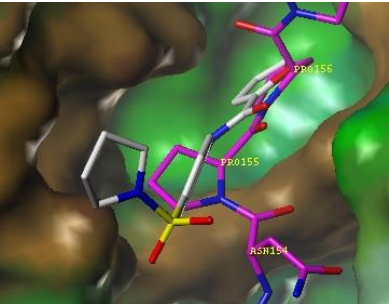
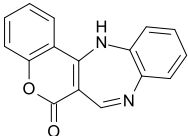
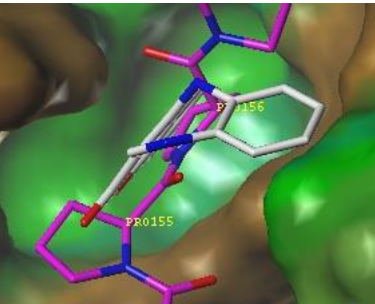
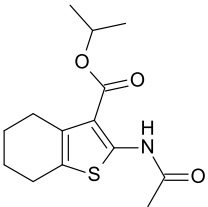
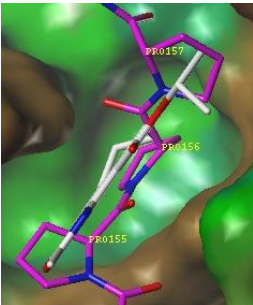
Figure 54: Comparison of the best interaction energy-ranked DOCK poses of diproline (a) and triproline (b) (carbons in white) with the respective crystal structures (carbons in magenta). The DOCK binding site box dimensions are indicated in Å.

The low molecular weight structure- and ligand-based screening with 23611 compounds allowed selecting 43 compounds representing 8 chemical classes. A representative of each class is listed in Table 29.

The representatives depicted in Table 29 overlay all relatively well with the central Pro156. “Butterfly” compounds include e.g. N-benzhydryl-5-methyl-furan-2-carboxamide, N,N-dibenzylisonicotinamide, and 4-[phenyl(phenylsulfanyl)acetyl]morpholine. Here, they are called “butterfly” compounds because they share a central polar part and two lateral hydrophobic wings, which are well covering the triproline pocket surface. Together with the benzodiazepines which are pointing deeper in the pocket surrounding Pro156, they are structurally interesting.

Table 29: Low molecular weight compound classes identified by protein- and/or ligand-based virtual screening. The formula of a representative of each class is depicted and the IUPAC name is provided together with the corresponding docking pose. In the picture of the docking pose, the carbon atoms are shown in magenta for p22^{phox} and in white for the docked compound, respectively. * Selected using the ligand-based virtual screening. ** Selected using the structure-based virtual screening. (continued on the next two pages)

Compound class	Docking pose	Characteristics
"Butterfly" compound **  N-benzhydryl-5-methyl-furan-2-carboxamide Other butterfly compounds: N,N-dibenzylisonicotinamide 4-[phenyl(phenylsulfany)acetyl]morpholine		Y-shaped molecules docked into the triproline region
Statin-like **  (4-fluorophenyl)-(4-phenyl-piperazin-1-yl)methanol		mimicking Pro156
4-membered ring **  3-isopropoxy-4-(2-thienyl)cyclobut-3-ene-1,2-dione		resonance stabilized; cyclobuten mimicking the ring of Pro156
Triazine **  ethyl 5-amino-6-oxo-1-phenyl-1,2,4-triazine-3-carboxylate		triazine mimicking Pro156 and phenyl mimicking Pro155

Azoles: Pyrazoles, Thiazoles, Thio- and Oxodiazoles, Triazoles, Tetrazoles, Imidazoles. Pyrazole * and **  N-(5-acetamido-4H-pyrazol-3-yl)acetamide		mimicking Pro156; from a synthesis point of view an easily modifiable symmetric pyrazole
Furamide **  N-(4-pyrrolidin-1-ylsulfonylphenyl) furan-2-amine oxide		mimicking Pro156
Benzodiazepines **  13H-chromeno[4,3-b][1,5]benzo- diazepin-6-one		mimicking Pro156; bent part exploring Pro156 surrounding pocket
Tetrahydrobenzoates**  isopropyl 2-acetamido-4,5,6,7- tetrahydrobenzothiophene-3-carboxylate		resembling the tetrahydroindoles of GenKyoTex ¹⁹²

The butterfly compounds resemble known non-nucleoside HIV-1 reverse transcriptase inhibitors (NNRTIs). NNRTIs largely prefer such a conformation to a planar one ²⁵⁷. Examples of NNRTIs having a “butterfly” arrangement are drawn in Figure 55.

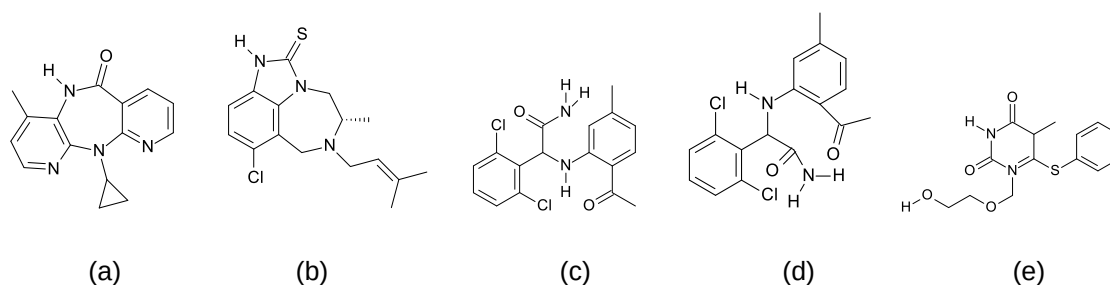


Figure 55: Non-nucleoside reverse transcriptase inhibitors (NNRTIs): Nevirapine (a), TIBO (tivrapipe) (b), α -APA (loviridine) (c), loviride (d), and 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine (HEPT) (e).

Nowadays, it is well established that the anti-HIV-1 activity of a “butterfly” compound depends upon the conformational rigidity and upon the orientation of the wings relative to the body²⁵⁸⁻²⁶⁰. Similarly, the wings of the butterfly compounds identified here were separated by amide groups of defined distance and angle, providing a defined directionality to the p47^{phox}-p22^{phox} inhibition.

The 4-membered ring compound of another compound class in Table 29 is similar to squaric acid, a precursor of tyrosine phosphatase inhibitors. Statins are hypolipidemic agents used to lower cholesterol levels in people at risk for cardiovascular diseases and molecules like the one depicted (a β -amino alcohol) are known to inhibit NMDA receptors. Benzodiazepines are tranquilizers acting on GABA_A receptors. Interestingly, the azoles and the tetrahydroindoles resemble pyrazolopyridines or tetrahydroindoles patented by GenKyoTex^{191,192}.

Ubiquitous binders were also present in the 1000 best energy-ranked solutions. They usually had a medium consensus score of 3 and many H-bonding interactions. The example shown in Figure 56 was ranked twentyfifth out of 1000 and displayed four H-bond acceptors, one H-bond donor and two aromatic rings.

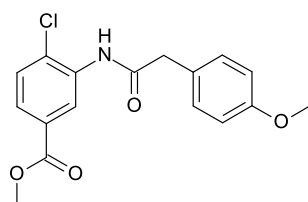


Figure 56: Example of a ubiquitous binder.

On the high molecular weight side, the five following compounds were retained (Table 30) out of the 12 screened.

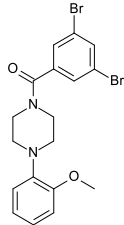
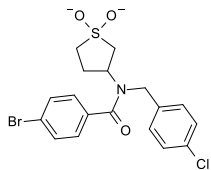
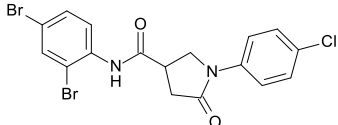
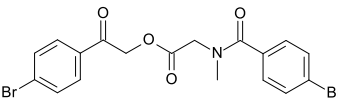
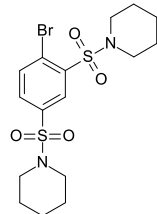
Taking the low and the high molecular weight hits together, 48 highly diverse, commercially available compounds were suggested for *in vitro* experimental testing.

The compound classes identified were similar to already tested active molecules of GenKyoTex but did not present radically new compound classes. For strategic reasons of the company, the 48 compounds have not been tested because priority has been given to other compounds. Generally, the molecular weight of the here identified compounds was lower compared to the GenKyoTex compounds, which was due to the choice of the ligand database and the need to leave room for improvement towards drug-like molecules.

With respect to peptidomimetics, spirolactams (including di- and tricyclic ones) could be a viable strategy. For intellectual property reasons, the 5,6-spirolactams from Cadus Pharmaceuticals or the conformationally constrained β -turn templates and modulators of SH3 domains (US patent 6638941)

should e.g. be excluded. In the filtered Specs database used for screening, 88 spiro compounds were present, however, none was retained among the 1000 best energy ranked compounds.

Table 30: High molecular weight hits.

High molecular weight hits		
Consensus score	IUPAC name	Formula
5	1-(3,5-dibromobenzoyl)-4-(2-methoxyphenyl)piperazine	
4	4-bromo-N-(4-chlorobenzyl)-N-(1,1-dioxidotetrahydro-3-thienyl)benzamide	
	1-(4-chlorophenyl)-N-(2,4-dibromophenyl)-5-oxopyrrolidine-3-carboxamide	
3	2-(4-bromophenyl)-2-oxoethyl [(4-bromobenzoyl)(methyl)amino]acetate	
3	1-[[2-bromo-5-(1-piperidinylsulfonyl)phenyl]sulfonyl]piperidine	

3.8.2 CELASTROL: BINDING MODE AND MECHANISM OF ACTION

The celastrol study is divided into a wet-lab and an *in silico* part. The experimental results of the NOX isoform inhibition by celastrol are displayed in Figure 57 and show that celastrol rapidly and completely inhibits NOX1, NOX2, NOX4, and NOX5 with IC₅₀ values in the micromolar range. Based on the IC₅₀ values and the Hill slopes (steepness of the curves), multiple sites of action are probable for the isoforms NOX1 and NOX2, whose activation is dependent on cytosolic subunits. In other words, the sigmoidal curve that evidences multiple sites of action indicates a competition with p22^{phox} for binding. For the constitutively active NOX4 and the Ca²⁺-activated NOX5, provided the higher IC₅₀ and the flatter slope, a single site of action is suggested.

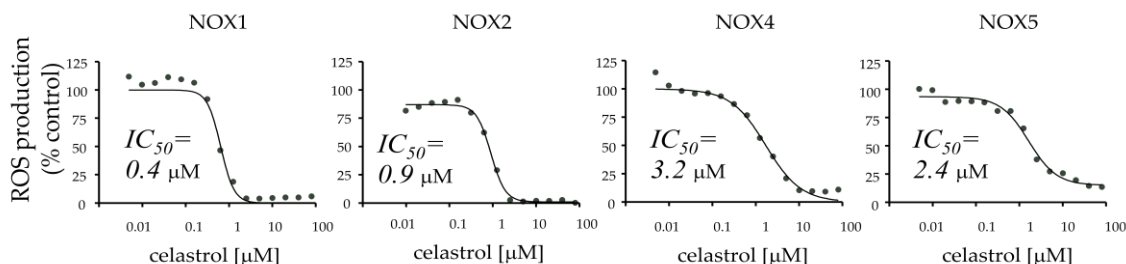


Figure 57: NOX isoform inhibition by celastrol. This Figure was kindly provided by Dr. V. Jaquet ²⁴⁷.

The oximeter study showed that celastrol blocked the NOX activity-dependent oxygen consumption and H_2O_2 detection (Figure 58). H_2O_2 is produced from O_2^- via the enzyme superoxide dismutase. Using an oximeter allows the distinction between antioxidant effects (using Trolox, a cell-permeable, water-soluble derivative of vitamin E) and true NOX2 inhibition (using DPI or celastrol).

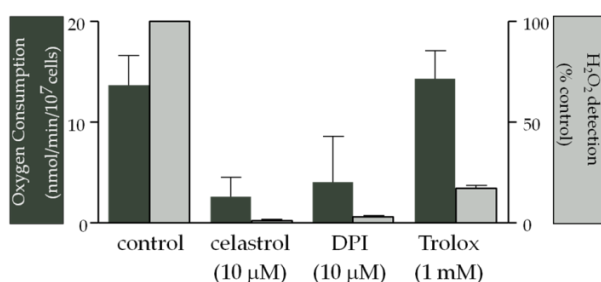


Figure 58: Celastrol blocks the oxygen consumption and H_2O_2 detection. This Figure was kindly provided by Dr. V. Jaquet.

The *in silico* part aimed at predicting the binding orientation of celastrol with four docking programs (DOCK, FlexX, GOLD, and Surflex). During the validation step, the triproline of $\text{p}22^{\text{phox}}$ was docked into the tandem SH3 interface of $\text{p}47^{\text{phox}}$. GOLD and Surflex showed even better RMSDs than FlexX and DOCK with respect to triproline docking and were therefore chosen for this part of the work.

Only the celastrol poses with the highest interaction energy in Surflex and GOLD were kept. Overall, the 10 generated Gold poses found were the same, with heavy atom RMSDs from the best-scored pose in the order of 2.0 Å. Considering the 10 generated Surflex poses, they showed roughly the same orientation of the carboxyl part as the GOLD solutions, but had the phenolic quinone placed more closely to the binding pocket surface. Both the best-scored GOLD and Surflex poses are involved in H-bonding via the carboxyl part and hydrophobic interactions via the triterpene part of celastrol, while the phenolic quinone covers the surface of the tandem $\text{p}47^{\text{phox}}$ binding pocket (Figure 59a and b).

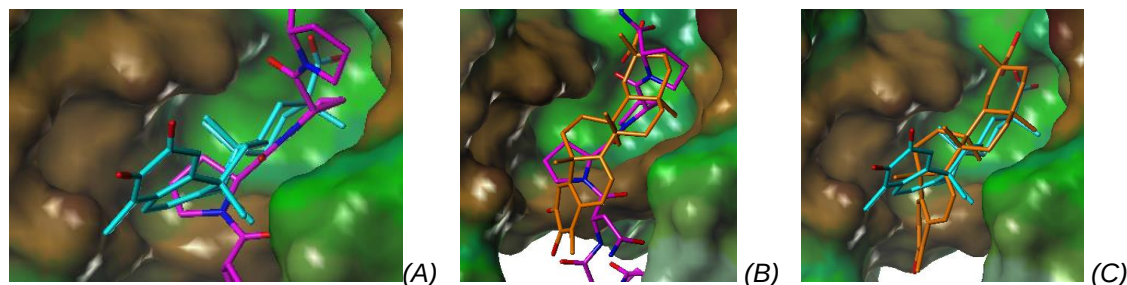


Figure 59: Proposed binding orientation of celastrol. The carbon atoms of the GOLD docking pose (a) are drawn in cyan and the ones of the Surflex pose (b) in orange. The crystallographic orientation of $\text{p}22^{\text{phox}}$ is indicated with magenta carbon atoms. (c) is the superposition of the docking poses of (a) and (b). The $\text{p}47^{\text{phox}}$ interface is coloured according to its lipophilicity (brown = lipophilic; green = less lipophilic).

The H-bond interactions of the carboxyl part involve in both cases Asp261. Pro156 of p22^{phox} is well mimicked by both poses and the triterpene parts of celastrol form hydrophobic interactions with Trp193 and Pro206 of the N-SH3 (on the right side in Figure 53) and with Trp 263, Pro276, Met278 and Tyr 279 of the C-SH3 (on the left side in Figure 53). Looking at the superposition of the two docking poses generated by the two different programs in Figure 59c, it can be stated that the H-bonding carboxyl orientation is conserved and the triterpene mimics the triproline region of the p22^{phox} wild type peptide, while the phenolic quinone is either covering the binding site surface or pointing away from it. Thus, the docking poses identified represent rather variants of one binding mode, here termed binding orientation, than two radically different binding modes.

Thus, the competitive inhibition observed for NOX1 and NOX2 and the docking into the tandem SH3 domains of p47^{phox} are both in favour of celastrol being able to bind to the organizer subunit (NOXO1 in the case of NOX1 and p47^{phox} in the case of NOX2). Meanwhile, it has been shown that celastrol interacts with recombinant p47^{phox} (residues 1 to 342) using both intrinsic fluorescence and native MS²⁴⁷. Intrinsic tryptophan fluorescence is commonly used to monitor protein conformational changes²⁶¹ or ligand binding²⁶². Indeed, celastrol quenched p47^{phox} fluorescence, reflecting thereby an interaction between the protein and the ligand (Figure 60).

Direct interaction in native conditions was detected by MS, where binding occurred with a molar ratio of 1:3 (p47^{phox} : celastrol) and a dose-dependent effect revealed the interaction to be specific.

A further indication that NOX1 or 2 are targets of celastrol is that this compound acts on ERK, MAPK, NF-kappaB, TNF-alpha, IL-1, caspase-3, nitric oxide, and potassium channels, which are all potential downstream targets of NOX-dependent ROS generation. Because celastrol displays however toxicity at concentrations fairly close to the EC₅₀ and inhibits hERG potassium channels, it is incompatible with a chronic use in humans. Nevertheless, it is useful as *in vitro* molecular probe for further studies on NOX.

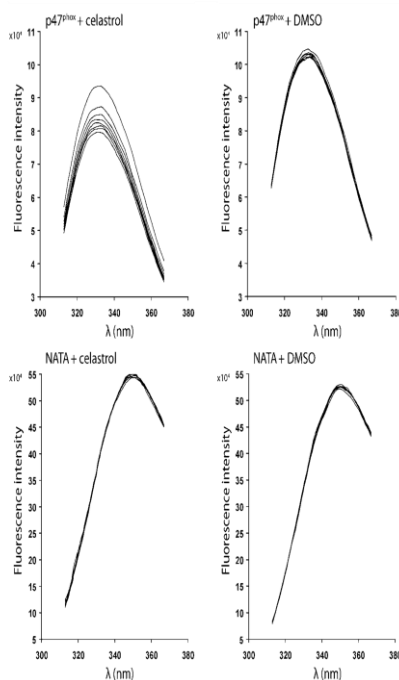


Figure 60: Celastrol quenches p47^{phox} fluorescence²⁴⁷. Recombinant p47^{phox} (residues 1 to 342) was treated with celastrol (molar ratio of 1:9) and the binding upon time was analyzed in a scanning fluorimeter using an excitation wavelength of 295 nm. The background signal from buffer alone was subtracted from the spectra shown. The upper left panel shows p47^{phox} with celastrol, the upper right one p47^{phox} with 1.2% of DMSO, the lower left one N-acetyl-tryptophanamide (NATA) with celastrol, and the lower right one NATA with 1.2% of DMSO. For each of the panels, the measurements were taken after 0, 2.5, 7.5, 10, 12.5, 15, 17.5, and 20 min. This Figure was kindly provided by Dr. V. Jaquet.

3.8.3 PEPTIDOMIMETICS

HIV-1 protease inhibitors with a peptidomimetic character, but sharing the butterfly-like arrangement of some compounds identified in structure-based virtual screening (3.8.1), were assessed without success as starting points for developing a peptidomimetic inhibitor of p47^{phox}-p22^{phox} because they do not fit into the binding site. For this reason, the protease inhibitor-like peptidomimetics were not further investigated in this thesis.

Starting points of the peptidomimetics project presented in this thesis

The peptidomimetics project presented here started from a series of already synthesised peptides for which most IC₅₀ values were already determined (Table 32). None of the peptides had a significantly better IC₅₀ than the wild type p22^{phox} sequence (IC₅₀ = 2 µM).

The questions arising from these results were the following:

- (i) Why was there no improvement in the p22^{phox} analogues as compared to the wild type peptide? and
- (ii) Is there enough space to improve the binding by using non-natural, commercially available amino acids?

GRID search

Subsequently, an *in silico* GRID search²⁶³ approach was chosen to point out affinity improving substitutions for the design of further p22^{phox} peptidomimetics. The underlying assumption was that p22^{phox} had favourable interactions with p47^{phox}. The search was performed on p47^{phox} (PDB ID: 1WLP) only to identify energetically favourable spots which could be exploited for gaining affinity and modifying the peptidic character as compared to the wild type p22^{phox}. Briefly, a probe with defined physicochemical properties was rolled over the surface of p47^{phox} and the interaction energy between the probe and the surface was calculated. Then, contour levels at appropriate energy levels were displayed together with the investigated surface. Contours at negative energy levels delineated regions of attraction between a probe and the protein²⁶³ and were found particularly at known peptide binding clefts. The characteristics of the GRID probes running over p47^{phox} only are summarized in Table 31.

Table 31: Characteristics of the GRID probes running over p47^{phox} only.

Type of probe	Probe name in GRID	Most negative interaction energy [kcal/mol]
H-bond donor	neutral flat NH ₂	-14.0
H-bond donor/acceptor	alkyl OH	-11.5
H-bond acceptor/smallest probe	H ₂ O	-12.1
H-bond acceptor	sp ² carbonyl oxygen	-8.9
Hydrophobic	CH ₃	-4.7
Hydrophobic	dry	-2.5

Table 32 provides an overview of the already synthesised peptides at the beginning of this project.

Table 32: Overview of the already synthesised peptides at the beginning of this peptidomimetics project. The full names of the abbreviations for the non-natural amino acids are provided in Tables 26 to 28 together with their chemical structure. ND = not determined. These data were kindly provided by Dr. H. Gaertner.

p22^{phox} analogues

Sequence numbering		149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168		% of inhibition at 10 μM	IC ₅₀ (μM)
<u>Controls</u>																								
p22 ^{phox} wild type		K	Q	P	P	S	N	P	P	P	R	P	P	A	E	A	R	K	K	P	S			2
p22 ^{phox} wild type (149-168)	Ac	K	Q	P	P	S	N	P	P	P	R	P	P	A	E	A	R	K	K	P	S	NH ₂	98	2
Cys-p22 ^{phox} wild type (149-168)-Cys oxidized	G	K	Q	P	P	S	N	P	P	P	R	P	P	A	E	A	R	K	K	P	S	G	98	2.6
p22 ^{phox} wild type (149-166)	Ac	K	Q	P	P	S	N	P	P	P	R	P	P	A	E	A	R	K	K	NH ₂			93	2.5
p22 ^{phox} wild type (151-166)		Ac		P	P	S	N	P	P	P	R	P	P	A	E	A	R	K	K	NH ₂			92	2
<u>Substitutions at Pro152</u>																								
[N-Naphm-Gly] ¹⁵² p22 ^{phox} (151-166)		Ac		P	NmG	S	N	P	P	P	R	P	P	A	E	A	R	K	K	NH ₂			95	1.9
[N-mpe-Gly] ¹⁵² p22 ^{phox} (151-166)		Ac		P	4fG	S	N	P	P	P	R	P	P	A	E	A	R	K	K	NH ₂			95	1.9
[N-Chm-Gly] ¹⁵² p22 ^{phox} (151-166)		Ac		P	NhG	S	N	P	P	P	R	P	P	A	E	A	R	K	K	NH ₂			95	1.8
[N-R-Che-Gly] ¹⁵² p22 ^{phox} (151-166)		Ac		P	Nyg	S	N	P	P	P	R	P	P	A	E	A	R	K	K	NH ₂			93	2.6
[N-Bzl-Gly] ¹⁵² p22 ^{phox} (149-168)	Ac	K	Q	P	NzG	S	N	P	P	P	R	P	P	A	E	A	R	K	K	P	S	NH ₂	85	3.5
[N-4mob-Gly] ¹⁵² p22 ^{phox} (151-166)		Ac		P	4mG	S	N	P	P	P	R	P	P	A	E	A	R	K	K	NH ₂			85	2.6
[N-Spe-Gly] ¹⁵² p22 ^{phox} (151-166)		Ac		P	NsG	S	N	P	P	P	R	P	P	A	E	A	R	K	K	NH ₂			85	2.4
[Pip] ¹⁵² p22 ^{phox} (149-168)	Ac	K	Q	P	Pip	S	N	P	P	P	R	P	P	A	E	A	R	K	K	P	S	NH ₂	85	3.9
[N-Rpe-Gly] ¹⁵² p22 ^{phox} (151-166)		Ac		P	NrG	S	N	P	P	P	R	P	P	A	E	A	R	K	K	NH ₂			81	4.8
[N-ver-Gly] ¹⁵² p22 ^{phox} (149-168)	Ac	K	Q	P	NvG	S	N	P	P	P	R	P	P	A	E	A	R	K	K	P	S	NH ₂	75	6.3
[N-S-Che-Gly] ¹⁵² p22 ^{phox} (151-166)		Ac		P	Nxg	S	N	P	P	P	R	P	P	A	E	A	R	K	K	NH ₂			74	4.2
[N-Ch-Gly] ¹⁵² p22 ^{phox} (151-166)		Ac		P	NcG	S	N	P	P	P	R	P	P	A	E	A	R	K	K	NH ₂			74	3.7
[PrS] ¹⁵² p22 ^{phox} (149-168)	Ac	K	Q	P	PrS	S	N	P	P	P	R	P	P	A	E	A	R	K	K	P	S	NH ₂	71	7.2
[N-Pheprop-Gly] ¹⁵² p22 ^{phox} (149-168)	Ac	K	Q	P	NoG	S	N	P	P	P	R	P	P	A	E	A	R	K	K	P	S	NH ₂	69	4.3
[N-ibut-Gly] ¹⁵² p22 ^{phox} (149-168)	Ac	K	Q	P	Ibu	S	N	P	P	P	R	P	P	A	E	A	R	K	K	P	S	NH ₂	69	ND
[N-Furf-Gly] ¹⁵² p22 ^{phox} (149-168)	Ac	K	Q	P	NfG	S	N	P	P	P	R	P	P	A	E	A	R	K	K	P	S	NH ₂	60	ND
[N-But-Gly] ¹⁵² p22 ^{phox} (149-168)	Ac	K	Q	P	Nbu	S	N	P	P	P	R	P	P	A	E	A	R	K	K	P	S	NH ₂	50	ND
[N-Phe-Gly] ¹⁵² p22 ^{phox} (149-168)	Ac	K	Q	P	NpG	S	N	P	P	P	R	P	P	A	E	A	R	K	K	P	S	NH ₂	39	ND

Chapter 3

sequence numbering		149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168		% of inhibition at 10 μ M	IC ₅₀ (μ M)
p22 ^{phox} wild type		K	Q	P	P	S	N	P	P	P	R	P	P	A	E	A	R	K	K	P	S			2
<u>Substitutions at Pro152</u>																								
<u>(continued)</u>																								
[Gly] ¹⁵² p22 ^{phox} (149-168)	Ac	K	Q	P	G	S	N	P	P	P	R	P	P	A	E	A	R	K	K	P	S	NH ₂	36	ND
[Gly] ¹⁵² p22 ^{phox} (149-168)	Ac	K	Q	P	Ind	S	N	P	P	P	R	P	P	A	E	A	R	K	K	P	S	NH ₂	27	ND
[N-4mob-Gly] ¹⁵² p22 ^{phox} (151-166)			Ac	P	4mG	S	N	P	P	P	R	P	P	A	E	A	R	K	K		NH ₂		19	ND
<u>Substitutions at Pro156</u>																								
[N-Phe-Gly] ¹⁵⁶ p22 ^{phox} (151-166)			Ac	P	P	S	N	P	NpG	P	R	P	P	A	E	A	R	K	K		NH ₂		49	ND
[N-Spe-Gly] ¹⁵⁶ p22 ^{phox} (151-166)			Ac	P	P	S	N	P	NsG	P	R	P	P	A	E	A	R	K	K		NH ₂		34	ND
[N-Bzl-Gly] ¹⁵⁶ p22 ^{phox} (151-166)			Ac	P	P	S	N	P	NzG	P	R	P	P	A	E	A	R	K	K		NH ₂		20	ND
[N-R-Che-Gly] ¹⁵⁶ p22 ^{phox} (151-166)			Ac	P	P	S	N	P	Nyg	P	R	P	P	A	E	A	R	K	K		NH ₂		11	ND
[N-Rpe-Gly] ¹⁵⁶ p22 ^{phox} (151-166)			Ac	P	P	S	N	P	NrG	P	R	P	P	A	E	A	R	K	K		NH ₂		11	ND
[N-Ch-Gly] ¹⁵⁶ p22 ^{phox} (151-166)			Ac	P	P	S	N	P	NcG	P	R	P	P	A	E	A	R	K	K		NH ₂		9	ND
[N-Chm-Gly] ¹⁵⁶ p22 ^{phox} (151-166)			Ac	P	P	S	N	P	NhG	P	R	P	P	A	E	A	R	K	K		NH ₂		9	ND
[N-S-Che-Gly] ¹⁵⁶ p22 ^{phox} (151-166)			Ac	P	P	S	N	P	Nxg	P	R	P	P	A	E	A	R	K	K		NH ₂		-2	ND
<u>Substitutions at Arg158</u>																								
[N-Me-Arg] ¹⁵⁸ p22 ^{phox} (149-168)	Ac	K	Q	P	P	S	N	P	P	P	MeR	P	P	A	E	A	R	K	K	P	S	NH ₂	95	3
[hArg] ¹⁵⁸ p22 ^{phox} (151-166)			Ac	P	P	S	N	P	P	P	hR	P	P	A	E	A	R	K	K		NH ₂		95	2.3
[A] ¹⁵⁸ p22 ^{phox} (149-168)	Ac	K	Q	P	P	S	N	P	P	P	A	P	P	A	E	A	R	K	K	P	S	NH ₂	0	ND
<u>Aib substitutions</u>																								
[Aib] ¹⁶³ p22 ^{phox} (149-168)	Ac	K	Q	P	P	S	N	P	P	P	R	P	P	A	E	Aib	R	K	K	P	S	NH ₂	97	2.5
[Aib] ¹⁶¹ p22 ^{phox} (151-166)			Ac	P	P	S	N	P	P	P	R	P	P	Aib	E	A	R	K	K		NH ₂		81	6.1
<u>Other derivatives</u>																								
p22 ^{phox} -A (149-166)	Ac	N	Q	P	P	S	N	P	P	P	R	P	P	D	Q	T	Q	G	K		NH ₂	28	ND	
p22 ^{phox} -B (149-166)	Ac	T	Q	P	P	S	N	P	P	P	R	P	P	P	E	M	V	Q	K		NH ₂	33	ND	
p22 ^{phox} -C (149-166)	Ac	N	Q	P	P	S	N	P	P	P	R	P	P	P	T	D	M	S	K		NH ₂	51	ND	
p22 ^{phox} -D (149-166)	Ac	N	Q	P	P	S	N	P	P	P	R	P	P	P	Y	G	E	G	K		NH ₂	36	ND	
p22 ^{phox} -E (149-166)	Ac	N	Q	P	P	S	N	P	P	P	R	P	P	Y	A	D	Q	D	K		NH ₂	41	ND	
p22 ^{phox} -F (149-166)	Ac	N	Q	P	P	S	N	P	P	P	R	P	P	P	V	S	N	E	K		NH ₂	36	ND	
p22 ^{phox} -G (149-166)	Ac	N	Q	P	P	S	N	P	P	P	R	P	P	Q	S	C	L	Q	K		NH ₂	55	ND	
p22 ^{phox} -H (149-166)	Ac	N	Q	P	P	S	N	P	P	P	R	P	P	G	Q	G	C	C	K		NH ₂	ND	ND	

GRID search on the p47^{phox} only

In the following, substitution possibilities using non-natural amino acids are discussed in sequence order. It is also assessed whether the experimental results can be explained by the GRID analysis. Figure 61 shows the hydrophobic GRID maps at the p47^{phox} interface, which cover particularly well the di- and triproline regions of p22^{phox}.

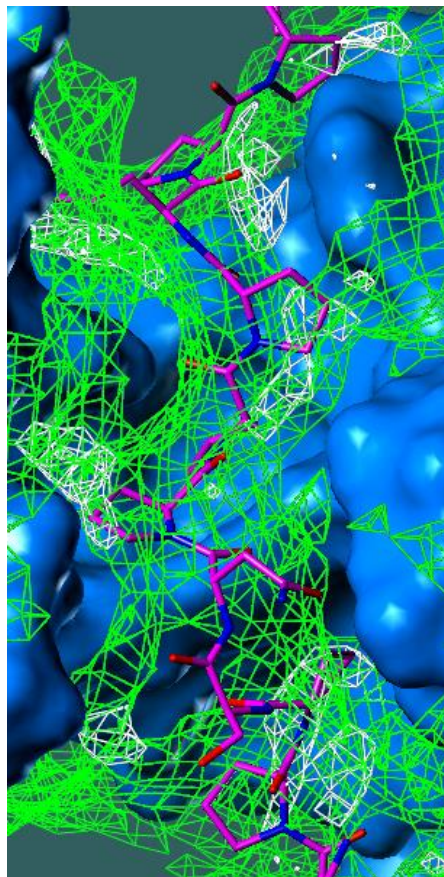


Figure 61: Grid maps of p47^{phox} only. p47^{phox} is shown as blue surface and p22^{phox} with magenta carbon atoms. The hydrophobic GRID interaction fields are CH₃ (green) and the dry probe (white). The former is displayed at an energy contour level of -1.0 kcal/mol and the latter at -0.5 kcal/mol. This means that GRID points with more favourable energies (less or equal to the stated numbers) are visible.

When looking at such a GRID map, the available space for substitutions on p22^{phox} needs to be analyzed first. The energy contour of the GRID is chosen such as to cover the hydrophobic parts of p22^{phox} and to reveal connected empty areas which could be reached by substitutions. Displaying the same GRID probe map at different contour levels allows for identifying energy minima which could be occupied by a suitable substitution to favourably contribute to binding affinity. Thus, after having assessed the location, the quality and the means of reaching the available space, suggestions for p22^{phox} peptidomimetics can be made. GRID serves therefore as a tool for rationalizing changes which are meaningful for the design of peptidomimetics.

Table 33 summarizes the spots that were identified by displaying the different GRID maps and gives a general overview of potential substitutions.

Table 33: Grid map of $p47^{phox}$ according to the probe type used.

Type of probe	$p47^{phox}$ only	
neutral flat NH_2	Pro156	Large hotspots suggesting substitutions at position 4 or 5 of the Pro ring
alkyl OH	Pro156	Substitution at position 5 of the Pro ring with – $(CH_2)_2OH$
	Arg158	Replacing one terminal NH_2 by OH
H_2O	Pro155-Pro156-Pro157	Many hotspots
sp^2 carbonyl	All	No hotspot of interest
CH_3	Pro151-Pro152	Large hotspot overlapping with the Pro152 ring
	Pro155-Pro156	Large hotspot overlapping with the Pro156 ring
dry	Pro151-Pro152	Two hotspots, one of them overlapping with the Pro152 ring
	Pro155-Pro156	Several small hotspots, one slightly larger at position 4 of the Pro155 ring
	Arg158	Several hotspots, one huge

Pro151-Pro152

Figure 62 shows that hydrophobic substitutions at the diproline region of $p22^{phox}$ are favourable. This is in agreement with the IC_{50} values of hydrophobic N-Gly analogues in Table 32.

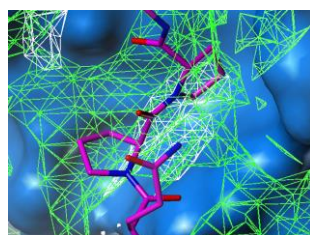


Figure 62: CH_3 (methyl) and dry (white) GRID maps around the diproline region (Pro151-Pro152) of $p22^{phox}$. The colouring is as in Figure 61.

Pro155

Considering the first Pro of the triproline region, Pro155, and the 2 hydrophobic GRID maps, thioPro, pipecolic acid, phenylPro, benzylPro or phenoxyPro can take advantage of the available space (Figure 63 a-d). In Figure 63c and d, a combined dry/methyl patch can be reached. The subsequent measurement of the four IC_{50} values, which range from 4.0 to 5.3 μM (Table 34), confirmed these GRID-based suggestions.

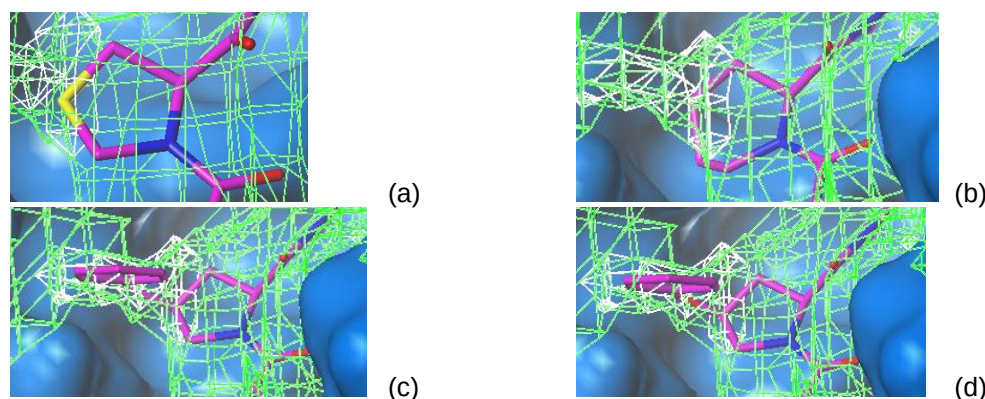


Figure 63: Favourable substitutions at Pro155 involve e.g. thioPro (a), pipecolic acid (b), benzylPro (c) and phenoxyPro (d). In (c) and (d), a combined dry/methyl patch is reached.

Pro156

With regard to the second triproline residue, Pro156, possible substitutions include thioPro and pipecolic acid (Figure 64 a and b). Additionally, N-phenyl Gly points advantageously inside the shallow pocket below the peptide backbone (Figure 65c).

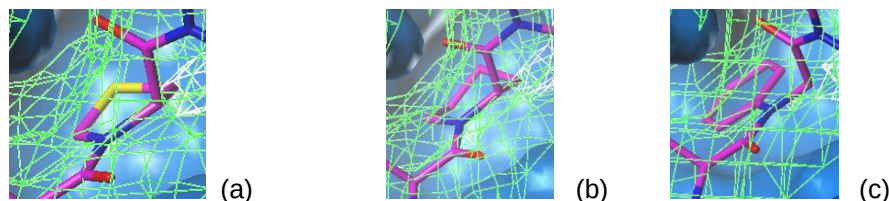


Figure 64: Examples of substitutions at Pro156: ThioPro (a), pipecolic acid (b), and N-phenyl Gly (c)

Pro157

Regarding the third triproline residue, Pro157, the same substitutions as for Pro155 can be suggested based on the space availability in the pocket. ThioPro and pipecolic acid are drawn as examples in Figure 65.

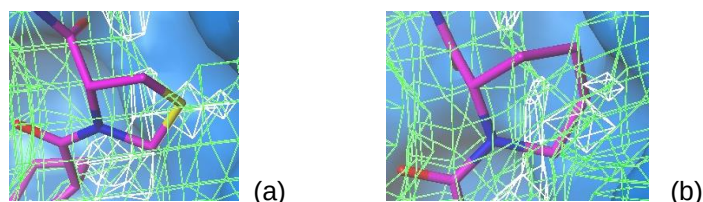


Figure 65: Examples of substitutions at Pro157: ThioPro (a) and pipecolic acid (b)

Arg158

The side chain NH_2 and NH groups of Arg158 are involved in H-bonding interactions with p47^{phox} (Figure 53). All these groups are located within the NH_2 GRID map shown in Figure 66. Replacing this amino acid with homoArg (hR) maintains the three H-bonds, while only one H-bond is kept with the Me-Arg (MeR) mutation. According to Table 32, one H-bond is still sufficient for obtaining an IC_{50} value of 3 μM . Provided the bottom of the Arg158 binding pocket is hydrophobic, the $(\text{CH}_2)_3$ of the side chain of Arg is appropriate and should be maintained. Based on this analysis, it is not astonishing that the Arg158Ala mutation resulted in 0% inhibition at 10 μM (Table 32).

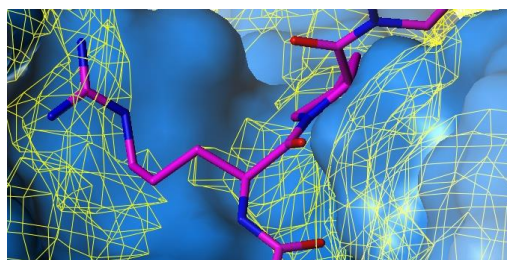


Figure 66: Arg158 of p22^{phox} and the NH_2 GRID map of p47^{phox} in yellow.

The residues Ala161 to Ala163 are the last ones that are interacting with p47^{phox} . Mutations at these residues have almost no effect on the IC_{50} values (Tables 32 and 34).

Double or triple mutations

To gain specificity, double or triple mutations could be envisaged such as any combination of the above-mentioned Pro156 or Pro155 mutations, any combination of the cited Pro155 and Pro156 mutations, any combination of the presented Pro156 and Pro157 mutations, a triple mutation like Pro155Pip-Pro156PrS-Pro157PrS, or a Pro155/Pro156 or Pro156/Pro157 double mutation together with a mutation at the position of Arg158.

Conclusion on the GRID searches

The backbone of the wild type peptide seems to fit perfectly into the interface space. While the diproline region (Pro151-Pro152) was tolerant with respect to N-Gly substitutions, a ring system with a similar volume as Pro had to be kept in the triproline region (Pro155-Pro156-Pro157) for not losing activity, except for N-phenyl-Gly at Pro156, which is exploiting the cavity further. The Gly substitutions at Pro152 were feasible because they did not significantly influence the peptide backbone conformation. Favourable substitutions at Arg158 had to maintain at least one H-bond with p47^{phox}.

With the results obtained, the two questions that have been asked earlier can be answered.

- (i) Why was there no improvement in the p22^{phox} analogues (Table 32) as compared to the wild type? The substitutions of Table 32 do not introduce major new favourable interactions with p47^{phox}.
- (ii) Is there enough space to improve the binding mode by using non-natural, commercially available amino acids? There was indeed enough space, given that some interesting potential interaction points were found by GRID analysis (Table 33).

Measured IC₅₀ values for the suggested peptides

The IC₅₀ values for the control peptides (PR-39, NOX2 (149-168), NOX2 (149-166), Cys-NOX2 (149-168)-Cys oxidized, NOX2 (151-166), and NOX2 (151-166) scrambled) are displayed together with newly synthesised peptides suggested based on the GRID analysis in Table 34. From the IC₅₀ values obtained, it can be concluded that Pro can be moderately substituted. This is already a considerable achievement considering the rigid character of Pro. Nevertheless, this has globally neither a positive nor a negative effect on the IC₅₀ values. However, more non-natural amino acids could be introduced for obtaining peptidomimetics.

The C-terminal substitutions at position 163, 164, and 165 listed in Table 34 were not a result of the GRID analysis but were synthesised based on suggestions of Prof. Dr. O. Hartley.

For the di- and tri-substitutions, a synergistic, affinity improving effect was hypothesised based on the GRID interaction fields, but these peptides could not be successfully synthesised.

Table 34: Summary of synthesised and tested p22^{phox} analogues with mutations at the triproline and C-terminal region. The newly synthesised peptide names are marked in grey and the ones suggested by the GRID analysis are underlined.

p22 ^{phox} analogues		Sequence numbering	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	IC ₅₀ (μM)
<u>Controls</u>																									
PR-39																									0.6
p22 ^{phox} (149-168)			2	Q	P	P	S	N	P	P	P	R	P	P	A	E	A	R	K	K	P	S			2
Cys- p22 ^{phox} (149-168)-Cys oxidized		C	2.6	Q	P	P	S	N	P	P	P	R	P	P	A	E	A	R	K	K	P	S	C		2.6
p22 ^{phox} (149-166)			2.5	Q	P	P	S	N	P	P	P	R	P	P	A	E	A	R	K	K					2.5
p22 ^{phox} (151-166)			2.7		P	P	S	N	P	P	P	R	P	P	A	E	A	R	K	K					2.7
p22 ^{phox} (151-166) scrambled			ND		P	P	R	K	N	P	R	P	P	A	P	E	S	P	A	K					ND
<u>Substitutions at Pro155</u>																									
<u>[Pip]¹⁵⁵ p22^{phox} (151-166)</u>			4.3		P	P	S	N	Pip	P	P	R	P	P	A	E	A	R	K	K					4.3
<u>[PrS]¹⁵⁵ p22^{phox} (151-166)</u>			4.4		P	P	S	N	PrS	P	P	R	P	P	A	E	A	R	K	K					4.4
<u>[BzlP]¹⁵⁵ p22^{phox} (151-166)</u>			4.8		P	P	S	N	BzlP	P	P	R	P	P	A	E	A	R	K	K					4.8
<u>[PhOP]¹⁵⁵ p22^{phox} (151-166)</u>			5.3		P	P	S	N	phOP	P	P	R	P	P	A	E	A	R	K	K					5.3
<u>Substitutions at Pro156</u>																									
[N-Rpe-Gly] ¹⁵⁶ p22 ^{phox} (151-166)			4.8		P	P	S	N	P	NrG	P	R	P	P	A	E	A	R	K	K					4.8
<u>[PrS]¹⁵⁶ p22^{phox} (151-166)</u>			5.4		P	P	S	N	P	PrS	P	R	P	P	A	E	A	R	K	K					5.4
<u>[Pip]¹⁵⁶ p22^{phox} (151-166)</u>			8.3		P	P	S	N	P	Pip	P	R	P	P	A	E	A	R	K	K					8.3
<u>[N-Phe-Gly]¹⁵⁶ p22^{phox} (151-166)</u>			ND		P	P	S	N	P	NpG	P	R	P	P	A	E	A	R	K	K					ND
[N-Spe-Gly] ¹⁵⁶ p22 ^{phox} (151-166)			ND		P	P	S	N	P	NsG	P	R	P	P	A	E	A	R	K	K					ND
[N-Bzl-Gly] ¹⁵⁶ p22 ^{phox} (151-166)			ND		P	P	S	N	P	NzG	P	R	P	P	A	E	A	R	K	K					ND
[N-R-Che-Gly] ¹⁵⁶ p22 ^{phox} (151-166)			ND		P	P	S	N	P	Nyg	P	R	P	P	A	E	A	R	K	K					ND
[N-Rpe-Gly] ¹⁵⁶ p22 ^{phox} (151-166)			ND		P	P	S	N	P	NrG	P	R	P	P	A	E	A	R	K	K					ND
[N-Ch-Gly] ¹⁵⁶ p22 ^{phox} (151-166)			ND		P	P	S	N	P	NcG	P	R	P	P	A	E	A	R	K	K					ND
[N-Chm-Gly] ¹⁵⁶ p22 ^{phox} (151-166)			ND		P	P	S	N	P	NhG	P	R	P	P	A	E	A	R	K	K					ND
[N-S-Che-Gly] ¹⁵⁶ p22 ^{phox} (151-166)			ND		P	P	S	N	P	Nxg	P	R	P	P	A	E	A	R	K	K					ND

Chapter 3

Sequence numbering	148	IC ₅₀ (μM)	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	IC ₅₀ (μM)
<u>Substitutions at Pro157</u>																							
[PrS] ¹⁵⁷ p22 ^{phox} (151-166)		4.9		P	P	S	N	P	P	PrS	R	P	P	A	E	A	R	K	K				4.9
[Pip] ¹⁵⁷ p22 ^{phox} (151-166)		missing		P	P	S	N	P	P	Pip	R	P	P	A	E	A	R	K	K				missing
<u>Di- and trisubstitutions</u>																							
[Pip] ¹⁵⁵ [PrS] ¹⁵⁶ p22 ^{phox} (151-166)		not sucessfully synthesised		P	P	S	N	Pip	PrS	P	R	P	P	A	E	A	R	K	K				not sucessfully synthesised
[Pip] ¹⁵⁵ [PrS] ¹⁵⁷ p22 ^{phox} (151-166)		not sucessfully synthesised		P	P	S	N	Pip	P	PrS	R	P	P	A	E	A	R	K	K				not sucessfully synthesised
[PrS] ¹⁵⁶ [PrS] ¹⁵⁷ p22 ^{phox} (151-166)		not sucessfully synthesised		P	P	S	N	P	PrS	PrS	R	P	P	A	E	A	R	K	K				not sucessfully synthesised
[Pip] ¹⁵⁵ [PrS] ¹⁵⁶ [PrS] ¹⁵⁷ p22 ^{phox} (151-166)		not sucessfully synthesised		P	P	S	N	Pip	PrS	PrS	R	P	P	A	E	A	R	K	K				not sucessfully synthesised
[Pip] ¹⁵⁵ [hR] ¹⁵⁸ p22 ^{phox} (151-166)		not sucessfully synthesised		P	P	S	N	Pip	P	P	hR	P	P	A	E	A	R	K	K				not sucessfully synthesised
[PrS] ¹⁵⁶ [hR] ¹⁵⁸ p22 ^{phox} (151-166)		not sucessfully synthesised		P	P	S	N	P	PrS	P	hR	P	P	A	E	A	R	K	K				not sucessfully synthesised
[PrS] ¹⁵⁷ [hR] ¹⁵⁸ p22 ^{phox} (151-166)		not sucessfully synthesised		P	P	S	N	P	P	PrS	hR	P	P	A	E	A	R	K	K				not sucessfully synthesised
<u>C-terminal substitutions</u>																							
[Aib] ¹⁶³ p22 ^{phox} (151-166)		6.1		P	P	S	N	P	P	P	R	P	P	A	E	Aib	R	K	K				6.1
[Har] ¹⁶⁴ p22 ^{phox} (151-166)		3.6		P	P	S	N	P	P	P	R	P	P	A	E	A	hR	K	K				3.6
[MeK] ¹⁶⁵ p22 ^{phox} (151-166)		6.7		P	P	S	N	P	P	P	R	P	P	A	E	A	R	MeK	K				6.7

IS IT WORTH TO SYNTHETIZE A RETROPEPTIDE OF P22^{PHOX}?

Based on the results in Tables 32 and 34, new ideas on peptidomimetics targeting the tandem SH3 domain of p47^{phox} had to be found. The retropeptide of p22^{phox} could be a starting point for new peptidomimetics. Provided the p22^{phox} peptide contains several “proline only” stretches, it was interesting to predict how the retropeptide of p22^{phox} could bind to the p47^{phox} interface.

In peptide design, each L-peptide has three analogue sequences built from either L- or D-amino acids:

- (i) The inversed-peptide has the same sequence, but only D-enantiomers.
- (ii) The retropeptide has a reversed sequence order, but only L-amino acids.
- (iii) The retro-inversed peptide has a reversed sequence order and only D-enantiomers.

With the sequence comparison method chosen here, no distinction between retro- and retro-inversed peptides can be made. Features considered while trying to match the Pro residues of the retropeptide with the native p22^{phox} sequence (residues 149 to 168) included: side chain polarity (non-polar/polar), side chain acidity or basicity (acidic/neutral/basic) and the hydropathy index (according to Kyte and Doolittle²⁵⁶: the larger the number, the more hydrophobic the amino acid).

When trying to align p22^{phox} with its retropeptide *in cerebro*, four different alignment cases can be imagined:

- Case A: five Pro aligned, triprolines matched, partial overlap of diproline regions;
- Case B & C: four prolines aligned, partial matching of the wild type triproline with one diproline of the retropeptide; and
- Case D: six prolines aligned, diproline match, and triproline alignment lost.

Table 35 summarizes the matches and mismatches among the present features (prolines, side chain polarity, side chain acidity or basicity, hydropathy, and sum). The sum of the matches and mismatches of the side chain polarity, side chain acidity/basicity and hydropathy, respectively, is not constant from one case to the other. The reason for this is that the main aligned features were the prolines, and several residues before the first Pro and after the last Pro of p22^{phox} were not aligned.

Table 35: Number and total of matches (M) and mismatches (MM) between the retropeptide sequence and the wild type p22^{phox} peptide. Considered features include prolines, the side chain polarity, the side chain acidity or basicity and hydropathy.

Features	Prolines		Side chain polarity		Side chain acidity or basicity		Hydropathy		Sum	
	M	MM	M	MM	M	MM	M	MM	M	MM
Case A	5	3	10	3	9	3	6	7	30	16
B	4	4	9	6	9	7	4	12	26	29
C	4	4	12	6	9	7	6	9	31	26
D	6	2	9	3	9	3	6	6	30	14

Case D shows a high number of matches, the lowest number of mismatches, and the most prolines aligned:

Retropeptide



Wild type p22^{phox} (149-168)



The retropeptide positioned according to the alignment of case D (Table 35) in the p47^{phox} interface should have the highest binding affinity of the four cases cited. This would confirm that a high Pro conservation is a key feature for correct binding to happen. Also, the position of the Pro156 residue, which is important for the anchoring in both SH3 domains, would be conserved. When mutating the wild type p22^{phox} peptide into the retropeptide of case D, the residues still fit well in the binding pocket. The Lys149Glu mutation is nevertheless expected to lead to repulsion, even if in the crystal structure, the flexible side chains point towards the solvent. Furthermore, the H-bonding interactions mediated by Arg158 are completely lost, which leads to a destabilisation of the peptide as seen for the

Arg158Ala substitution (Table 32) if additional hydrophobic interactions elsewhere in the retropeptide cannot compensate for it.

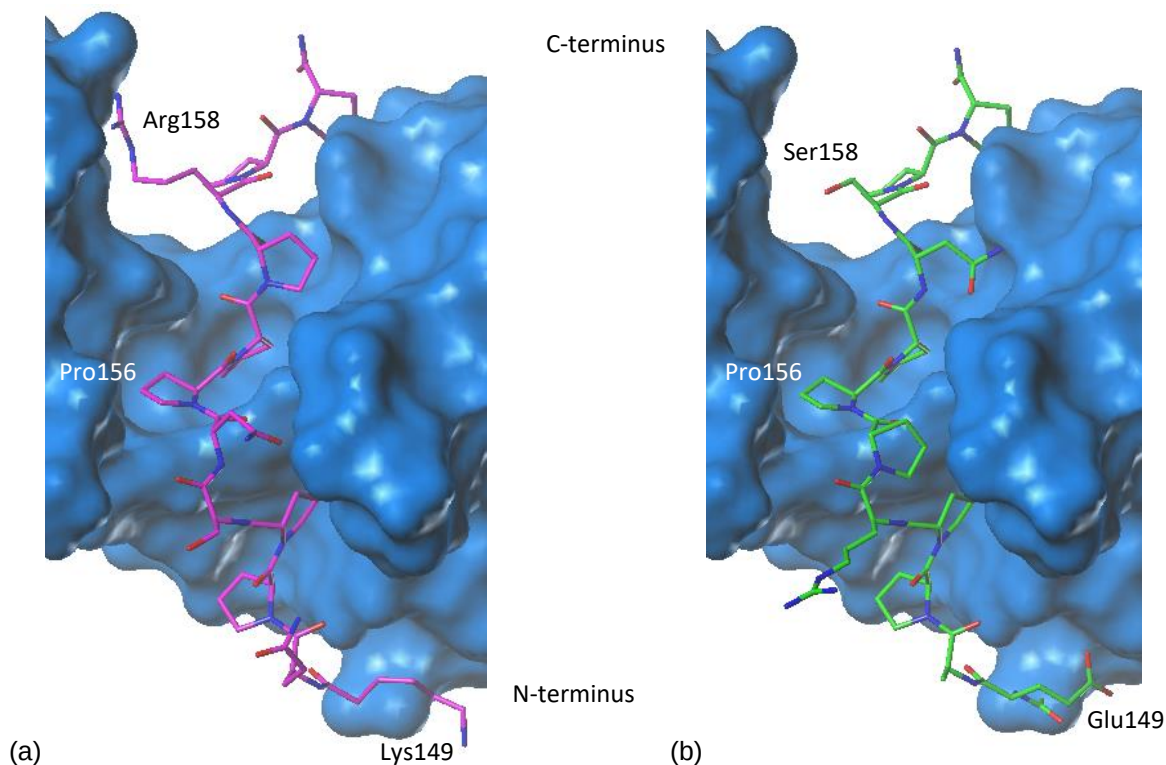


Figure 67: The wild type $p22^{phox}$ (a) and its retropeptide (b) in $p47^{phox}$. The binding mode of the retropeptide was obtained by simply mutating each residue.

Considering a retroinverso peptide, the inverted arrangement of the side chains in the D-enantiomers could potentially invert the importance of each SH3 for binding, which is a rather unfavourable starting point for new peptidomimetics.

3.9 CONCLUSIONS

Various molecular modelling techniques presented in this chapter contributed to a better understanding of the characteristics needed for NADPH oxidase 2 (NOX2) $p47^{phox}$ - $p22^{phox}$ protein-protein interaction breakers. The combined structure- and ligand-based VS strategy yielded a panel of potential small molecular weight inhibitors, some of which are similar to pyrazolopyridines¹⁹¹ and tetrahydroindoles¹⁹² already patented by GenKyoTex. For celastrol, which was identified to be an inhibitor by experimental medium throughput screening, docking could identify a potential binding orientation in $p47^{phox}$. The binding has recently been confirmed by intrinsic fluorescence measurements²⁴⁷. Using molecular interaction fields, favourable substitutions of the wild type $p22^{phox}$ peptide with commercially available, non-natural amino could be suggested. Although none of these interactions had a significantly better IC_{50} value than the wild type peptide, first steps towards peptidomimetics could be taken. This is already a considerable achievement, given that such an analysis does not take dynamic aspects into account. With respect to the retropeptide of $p22^{phox}$, a potential binding mode could be identified, which has probably a lower affinity towards $p47^{phox}$ due to the loss of H-bonds.

3.10 ACKNOWLEDGEMENTS

Dr. P. Page and Dr. M. Orchard of GenKyoTex are acknowledged for insights into the work of medicinal chemists in a start-up company. Prof. Dr. O. Hartley and Dr. H. Gaertner (Structural Biology and Bioinformatics, Centre Médical Universitaire, University of Geneva) are thanked for having considered a part of the suggested peptidomimetics for synthesis. Dr. V. Jaquet (Department of Pathology & Immunology, Centre Médical Universitaire, University of Geneva) is thanked for the

inhibition kinetics on the peptidomimetics and the exhaustive experimental work on celastrol. Finally, Prof. Dr. K.-H. Krause (Department of Pathology & Immunology, Centre Médical Universitaire, University of Geneva) is acknowledged for insightful discussions on NADPH oxidases and his involvement in all parts of the work presented in this chapter.

3.11 REFERENCES

1. Witko-Sarsat, V., Rieu, P., Descamps-Latscha, B., Lesavre, P. & Halbwachs-Mecarelli, L. Neutrophils: molecules, functions and pathophysiological aspects. *Lab Invest* **80**, 617-53 (2000).
2. Reeves, E. P. et al. Killing activity of neutrophils is mediated through activation of proteases by K⁺ flux. *Nature* **416**, 291-7 (2002).
3. Tkalecic, J. et al. Impaired immunity and enhanced resistance to endotoxin in the absence of neutrophil elastase and cathepsin G. *Immunity* **12**, 201-10 (2000).
4. Ahluwalia, J. et al. The large-conductance Ca²⁺-activated K⁺ channel is essential for innate immunity. *Nature* **427**, 853-8 (2004).
5. Harman, D. Aging: a theory based on free radical and radiation chemistry. *J Gerontol* **11**, 298-300 (1956).
6. Harman, D. The biologic clock: the mitochondria? *J Am Geriatr Soc* **20**, 145-7 (1972).
7. Segal, A. W. et al. Cytochrome b-245 is a flavocytochrome containing FAD and the NADPH-binding site of the microbicidal oxidase of phagocytes. *Biochem J* **284** (Pt 3), 781-8 (1992).
8. Babior, B. M. & Kipnes, R. S. Superoxide-forming enzyme from human neutrophils: evidence for a flavin requirement. *Blood* **50**, 517-24 (1977).
9. Light, D. R., Walsh, C., O'Callaghan, A. M., Goetzl, E. J. & Tauber, A. I. Characteristics of the cofactor requirements for the superoxide-generating NADPH oxidase of human polymorphonuclear leukocytes. *Biochemistry* **20**, 1468-76 (1981).
10. Groemping, Y. & Rittinger, K. Activation and assembly of the NADPH oxidase: a structural perspective. *Biochem J* **386**, 401-16 (2005).
11. Henderson, L. M., Banting, G. & Chappell, J. B. The arachidonate-activable, NADPH oxidase-associated H⁺ channel. Evidence that gp91-phox functions as an essential part of the channel. *J Biol Chem* **270**, 5909-16 (1995).
12. Nakamura, M., Murakami, M., Koga, T., Tanaka, Y. & Minakami, S. Monoclonal antibody 7D5 raised to cytochrome b558 of human neutrophils: immunocytochemical detection of the antigen in peripheral phagocytes of normal subjects, patients with chronic granulomatous disease, and their carrier mothers. *Blood* **69**, 1404-8 (1987).
13. Imajoh-Ohmi, S., Tokita, K., Ochiai, H., Nakamura, M. & Kanegasaki, S. Topology of cytochrome b558 in neutrophil membrane analyzed by anti-peptide antibodies and proteolysis. *J Biol Chem* **267**, 180-4 (1992).
14. Burritt, J. B. et al. Phage display epitope mapping of human neutrophil flavocytochrome b558. Identification of two juxtaposed extracellular domains. *J Biol Chem* **276**, 2053-61 (2001).
15. Rotrosen, D. et al. Evidence for a functional cytoplasmic domain of phagocyte oxidase cytochrome b558. *J Biol Chem* **265**, 8745-50 (1990).
16. Burritt, J. B. et al. Functional epitope on human neutrophil flavocytochrome b558. *J Immunol* **170**, 6082-9 (2003).
17. Teahan, C., Rowe, P., Parker, P., Totty, N. & Segal, A. W. The X-linked chronic granulomatous disease gene codes for the beta-chain of cytochrome b-245. *Nature* **327**, 720-1 (1987).
18. Paclet, M. H., Henderson, L. M., Campion, Y., Morel, F. & Dagher, M. C. Localization of Nox2 N-terminus using polyclonal antipeptide antibodies. *Biochem J* **382**, 981-6 (2004).
19. Harper, A. M., Chaplin, M. F. & Segal, A. W. Cytochrome b-245 from human neutrophils is a glycoprotein. *Biochem J* **227**, 783-8 (1985).
20. Sumimoto, H. et al. Role of Src homology 3 domains in assembly and activation of the phagocyte NADPH oxidase. *Proc Natl Acad Sci U S A* **91**, 5345-9 (1994).
21. Dinanuer, M. C. et al. Point mutation in the cytoplasmic domain of the neutrophil p22-phox cytochrome b subunit is associated with a nonfunctional NADPH oxidase and chronic granulomatous disease. *Proc Natl Acad Sci U S A* **88**, 11231-5 (1991).
22. Kawahara, T., Ritsick, D., Cheng, G. & Lambeth, J. D. Point mutations in the proline-rich region of p22phox are dominant inhibitors of Nox1- and Nox2-dependent reactive oxygen generation. *J Biol Chem* **280**, 31859-69 (2005).

23. Banfi, B. et al. A mammalian H⁺ channel generated through alternative splicing of the NADPH oxidase homolog NOH-1. *Science* **287**, 138-42 (2000).
24. Lambeth, J. D. NOX enzymes and the biology of reactive oxygen. *Nat Rev Immunol* **4**, 181-9 (2004).
25. Vignais, P. V. The superoxide-generating NADPH oxidase: structural aspects and activation mechanism. *Cell Mol Life Sci* **59**, 1428-59 (2002).
26. Krause, K. H. Tissue distribution and putative physiological function of NOX family NADPH oxidases. *Jpn J Infect Dis* **57**, S28-9 (2004).
27. Sumimoto, H., Miyano, K. & Takeya, R. Molecular composition and regulation of the Nox family NAD(P)H oxidases. *Biochem Biophys Res Commun* **338**, 677-86 (2005).
28. Suh, Y. A. et al. Cell transformation by the superoxide-generating oxidase Mox1. *Nature* **401**, 79-82 (1999).
29. Banfi, B. et al. A Ca²⁺-activated NADPH oxidase in testis, spleen, and lymph nodes. *J Biol Chem* **276**, 37594-601 (2001).
30. Banfi, B. et al. NOX3, a superoxide-generating NADPH oxidase of the inner ear. *J Biol Chem* **279**, 46065-72 (2004).
31. Paffenholz, R. et al. Vestibular defects in head-tilt mice result from mutations in Nox3, encoding an NADPH oxidase. *Genes Dev* **18**, 486-91 (2004).
32. Shiose, A. et al. A novel superoxide-producing NAD(P)H oxidase in kidney. *J Biol Chem* **276**, 1417-23 (2001).
33. Geiszt, M., Kopp, J. B., Varnai, P. & Leto, T. L. Identification of renox, an NAD(P)H oxidase in kidney. *Proc Natl Acad Sci U S A* **97**, 8010-4 (2000).
34. Serrander, L. et al. NOX4 activity is determined by mRNA levels and reveals a unique pattern of ROS generation. *Biochem J* **406**, 105-14 (2007).
35. Geiszt, M., Witta, J., Baffi, J., Lekstrom, K. & Leto, T. L. Dual oxidases represent novel hydrogen peroxide sources supporting mucosal surface host defense. *Faseb J* **17**, 1502-4 (2003).
36. Babior, B. M., Lambeth, J. D. & Nauseef, W. The neutrophil NADPH oxidase. *Arch Biochem Biophys* **397**, 342-4 (2002).
37. Cheng, G., Ritsick, D. & Lambeth, J. D. Nox3 regulation by NOXO1, p47phox, and p67phox. *J Biol Chem* **279**, 34250-5 (2004).
38. Banfi, B. et al. Mechanism of Ca²⁺ activation of the NADPH oxidase 5 (NOX5). *J Biol Chem* **279**, 18583-91 (2004).
39. De Deken, X. et al. Cloning of two human thyroid cDNAs encoding new members of the NADPH oxidase family. *J Biol Chem* **275**, 23227-33 (2000).
40. Edens, W. A. et al. Tyrosine cross-linking of extracellular matrix is catalyzed by Duox, a multidomain oxidase/peoxidase with homology to the phagocyte oxidase subunit gp91phox. *J Cell Biol* **154**, 879-91 (2001).
41. Larkin, M. A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-8 (2007).
42. Li, X. J. et al. Leu505 of Nox2 is crucial for optimal p67phox-dependent activation of the flavocytochrome b558 during phagocytic NADPH oxidase assembly. *J Leukoc Biol* **81**, 238-49 (2007).
43. Beitz, E. TEXshade: shading and labeling of multiple sequence alignments using LATEX2 epsilon. *Bioinformatics* **16**, 135-9 (2000).
44. Kawahara, T., Quinn, M. T. & Lambeth, J. D. Molecular evolution of the reactive oxygen-generating NADPH oxidase (Nox/Duox) family of enzymes. *BMC Evol Biol* **7**, 109 (2007).
45. Ambasta, R. K. et al. Direct interaction of the novel Nox proteins with p22phox is required for the formation of a functionally active NADPH oxidase. *J Biol Chem* **279**, 45935-41 (2004).
46. Kawahara, T. & Lambeth, J. D. Molecular evolution of Phox-related regulatory subunits for NADPH oxidase enzymes. *BMC Evol Biol* **7**, 178 (2007).
47. Hata, K., Ito, T., Takeshige, K. & Sumimoto, H. Anionic amphiphile-independent activation of the phagocyte NADPH oxidase in a cell-free system by p47phox and p67phox, both in C terminally truncated forms. Implication for regulatory Src homology 3 domain-mediated interactions. *J Biol Chem* **273**, 4232-6 (1998).
48. Ogura, K. et al. NMR solution structure of the tandem Src homology 3 domains of p47phox complexed with a p22phox-derived proline-rich peptide. *J Biol Chem* **281**, 3660-8 (2006).
49. Cheng, G., Cao, Z., Xu, X., van Meir, E. G. & Lambeth, J. D. Homologs of gp91phox: cloning and tissue expression of Nox3, Nox4, and Nox5. *Gene* **269**, 131-40 (2001).
50. Cave, A. C. et al. NADPH oxidases in cardiovascular health and disease. *Antioxid Redox Signal* **8**, 691-728 (2006).

51. Keaney, J. F., Jr. et al. Obesity and systemic oxidative stress: clinical correlates of oxidative stress in the Framingham Study. *Arterioscler Thromb Vasc Biol* **23**, 434-9 (2003).
52. Ray, R. & Shah, A. M. NADPH oxidase and endothelial cell function. *Clin Sci (Lond)* **109**, 217-26 (2005).
53. Landmesser, U. et al. Oxidation of tetrahydrobiopterin leads to uncoupling of endothelial cell nitric oxide synthase in hypertension. *J Clin Invest* **111**, 1201-9 (2003).
54. McNally, J. S. et al. Role of xanthine oxidoreductase and NAD(P)H oxidase in endothelial superoxide production in response to oscillatory shear stress. *Am J Physiol Heart Circ Physiol* **285**, H2290-7 (2003).
55. Guzik, T. J. & Harrison, D. G. Vascular NADPH oxidases as drug targets for novel antioxidant strategies. *Drug Discov Today* **11**, 524-33 (2006).
56. Wassmann, S. et al. Cellular antioxidant effects of atorvastatin in vitro and in vivo. *Arterioscler Thromb Vasc Biol* **22**, 300-5 (2002).
57. Rueckschloss, U., Galle, J., Holtz, J., Zerkowski, H. R. & Morawietz, H. Induction of NAD(P)H oxidase by oxidized low-density lipoprotein in human endothelial cells: antioxidative potential of hydroxymethylglutaryl coenzyme A reductase inhibitor therapy. *Circulation* **104**, 1767-72 (2001).
58. Rajagopalan, S. & Harrison, D. G. Reversing endothelial dysfunction with ACE inhibitors. A new trend. *Circulation* **94**, 240-3 (1996).
59. Cross, A. R. & Segal, A. W. The NADPH oxidase of professional phagocytes--prototype of the NOX electron transport chain systems. *Biochim Biophys Acta* **1657**, 1-22 (2004).
60. Ponting, C. P. Novel domains in NADPH oxidase subunits, sorting nexins, and PtdIns 3-kinases: binding partners of SH3 domains? *Protein Sci* **5**, 2353-7 (1996).
61. Doussiere, J., Bouzidi, F., Poinas, A., Gaillard, J. & Vignais, P. V. Kinetic study of the activation of the neutrophil NADPH oxidase by arachidonic acid. Antagonistic effects of arachidonic acid and phenylarsine oxide. *Biochemistry* **38**, 16394-406 (1999).
62. Nisimoto, Y., Otsuka-Murakami, H. & Lambeth, D. J. Reconstitution of flavin-depleted neutrophil flavocytochrome b558 with 8-mercapto-FAD and characterization of the flavin-reconstituted enzyme. *J Biol Chem* **270**, 16428-34 (1995).
63. Cross, A. R., Rae, J. & Curnutte, J. T. Cytochrome b-245 of the neutrophil superoxide-generating system contains two nonidentical hemes. Potentiometric studies of a mutant form of gp91phox. *J Biol Chem* **270**, 17075-7 (1995).
64. Cross, A. R., Yarchover, J. L. & Curnutte, J. T. The superoxide-generating system of human neutrophils possesses a novel diaphorase activity. Evidence for distinct regulation of electron flow within NADPH oxidase by p67-phox and p47-phox. *J Biol Chem* **269**, 21448-54 (1994).
65. Cross, A. R. & Curnutte, J. T. The cytosolic activating factors p47phox and p67phox have distinct roles in the regulation of electron flow in NADPH oxidase. *J Biol Chem* **270**, 6543-8 (1995).
66. Diebold, B. A. & Bokoch, G. M. Molecular basis for Rac2 regulation of phagocyte NADPH oxidase. *Nat Immunol* **2**, 211-5 (2001).
67. Cross, A. R., Erickson, R. W. & Curnutte, J. T. Simultaneous presence of p47(phox) and flavocytochrome b-245 are required for the activation of NADPH oxidase by anionic amphiphiles. Evidence for an intermediate state of oxidase activation. *J Biol Chem* **274**, 15519-25 (1999).
68. Koshkin, V. & Pick, E. Superoxide production by cytochrome b559. Mechanism of cytosol-independent activation. *FEBS Lett* **338**, 285-9 (1994).
69. Segal, A. W., Geisow, M., Garcia, R., Harper, A. & Miller, R. The respiratory burst of phagocytic cells is associated with a rise in vacuolar pH. *Nature* **290**, 406-9 (1981).
70. van Zwieten, R., Wever, R., Hamers, M. N., Weening, R. S. & Roos, D. Extracellular proton release by stimulated neutrophils. *J Clin Invest* **68**, 310-3 (1981).
71. Takanaka, K. & O'Brien, P. J. Proton release associated with respiratory burst of polymorphonuclear leukocytes. *J Biochem* **103**, 656-60 (1988).
72. Banfi, B. et al. A novel H(+) conductance in eosinophils: unique characteristics and absence in chronic granulomatous disease. *J Exp Med* **190**, 183-94 (1999).
73. Maturana, A. et al. Heme histidine ligands within gp91(phox) modulate proton conduction by the phagocyte NADPH oxidase. *J Biol Chem* **276**, 30277-84 (2001).
74. Maturana, A., Krause, K. H. & Demarex, N. NOX family NADPH oxidases: do they have built-in proton channels? *J Gen Physiol* **120**, 781-6 (2002).
75. Medvedev, D. M., Daizadeh, I. & Stuchebrukhov, A. A. Electron transfer tunneling pathways in bovine heart cytochrome c oxidase. *JACS* **122**, 6571-6582 (2000).

76. Lambeth, J. D., Cheng, G., Arnold, R. S. & Edens, W. A. Novel homologs of gp91phox. *Trends Biochem Sci* **25**, 459-61 (2000).
77. Davis, A. R., Mascolo, P. L., Bunger, P. L., Sipes, K. M. & Quinn, M. T. Cloning and sequencing of the bovine flavocytochrome b subunit proteins, gp91-phox and p22-phox: comparison with other known flavocytochrome b sequences. *J Leukoc Biol* **64**, 114-23 (1998).
78. Shatwell, K. P., Dancis, A., Cross, A. R., Klausner, R. D. & Segal, A. W. The FRE1 ferric reductase of *Saccharomyces cerevisiae* is a cytochrome b similar to that of NADPH oxidase. *J Biol Chem* **271**, 14240-4 (1996).
79. Finegold, A. A., Shatwell, K. P., Segal, A. W., Klausner, R. D. & Dancis, A. Intramembrane bis-heme motif for transmembrane electron transport conserved in a yeast iron reductase and the human NADPH oxidase. *J Biol Chem* **271**, 31021-4 (1996).
80. Biberstine-Kinkade, K. J. et al. Heme-ligating histidines in flavocytochrome b(558): identification of specific histidines in gp91(phox). *J Biol Chem* **276**, 31105-12 (2001).
81. Quinn, M. T. & Gauss, K. A. Structure and regulation of the neutrophil respiratory burst oxidase: comparison with nonphagocyte oxidases. *J Leukoc Biol* **76**, 760-81 (2004).
82. Biberstine-Kinkade, K. J. et al. Mutagenesis of p22(phox) histidine 94. A histidine in this position is not required for flavocytochrome b558 function. *J Biol Chem* **277**, 30368-74 (2002).
83. Cross, A. R., Noack, D., Rae, J., Curnutte, J. T. & Heyworth, P. G. Hematologically important mutations: the autosomal recessive forms of chronic granulomatous disease (first update). *Blood Cells Mol Dis* **26**, 561-5 (2000).
84. Scott, E. E. et al. An open conformation of mammalian cytochrome P450 2B4 at 1.6-Å resolution. *Proc Natl Acad Sci U S A* **100**, 13196-201 (2003).
85. Vatsis, K. P., Peng, H. M. & Coon, M. J. Abolition of oxygenase function, retention of NADPH oxidase activity, and emergence of peroxidase activity upon replacement of the axial cysteine-436 ligand by histidine in cytochrome P450 2B4. *Arch Biochem Biophys* **434**, 128-38 (2005).
86. Anzenbacher, P. & Anzenbacherova, E. Cytochromes P450 and metabolism of xenobiotics. *Cell Mol Life Sci* **58**, 737-47 (2001).
87. Leto, T. L. et al. Cloning of a 67-kD neutrophil oxidase factor with similarity to a noncatalytic region of p60c-src. *Science* **248**, 727-30 (1990).
88. Adams, E. R. et al. Interaction of human neutrophil flavocytochrome b with cytosolic proteins: transferred-NOESY NMR studies of a gp91phox C-terminal peptide bound to p47phox. *Biochem J* **325** (Pt 1), 249-57 (1997).
89. Park, H. S. & Park, J. W. Conformational changes of the leukocyte NADPH oxidase subunit p47(phox) during activation studied through its intrinsic fluorescence. *Biochim Biophys Acta* **1387**, 406-14 (1998).
90. Lee, J. H., Lee, K. S., Chung, T. & Park, J. C-terminal region of the cytosolic subunit p47(phox) is a primary target of conformational change during the activation of leukocyte NADPH oxidase. *Biochimie* **82**, 727-32 (2000).
91. Durand, D. et al. Small-angle X-ray scattering reveals an extended organization for the autoinhibitory resting state of the p47(phox) modular protein. *Biochemistry* **45**, 7185-93 (2006).
92. Groemping, Y., Lapouge, K., Smerdon, S. J. & Rittinger, K. Molecular basis of phosphorylation-induced activation of the NADPH oxidase. *Cell* **113**, 343-55 (2003).
93. Yuzawa, S. et al. Solution structure of the tandem Src homology 3 domains of p47phox in an autoinhibited form. *J Biol Chem* **279**, 29752-60 (2004).
94. Yuzawa, S. et al. A molecular mechanism for autoinhibition of the tandem SH3 domains of p47phox, the regulatory subunit of the phagocyte NADPH oxidase. *Genes Cells* **9**, 443-56 (2004).
95. Mol, C. D. et al. Structural basis for the autoinhibition and STI-571 inhibition of c-Kit tyrosine kinase. *J Biol Chem* **279**, 31655-63 (2004).
96. Nagar, B. et al. Structural basis for the autoinhibition of c-Abl tyrosine kinase. *Cell* **112**, 859-71 (2003).
97. Clark, R. A., Volpp, B. D., Leidal, K. G. & Nauseef, W. M. Two cytosolic components of the human neutrophil respiratory burst oxidase translocate to the plasma membrane during cell activation. *J Clin Invest* **85**, 714-21 (1990).
98. Rotrosen, D. & Leto, T. L. Phosphorylation of neutrophil 47-kDa cytosolic oxidase factor. Translocation to membrane is associated with distinct phosphorylation events. *J Biol Chem* **265**, 19910-5 (1990).

99. el Benna, J., Faust, L. P. & Babior, B. M. The phosphorylation of the respiratory burst oxidase component p47phox during neutrophil activation. Phosphorylation of sites recognized by protein kinase C and by proline-directed kinases. *J Biol Chem* **269**, 23431-6 (1994).
100. DeLeo, F. R. & Quinn, M. T. Assembly of the phagocyte NADPH oxidase: molecular interaction of oxidase proteins. *J Leukoc Biol* **60**, 677-91 (1996).
101. Leusen, J. H., Verhoeven, A. J. & Roos, D. Interactions between the components of the human NADPH oxidase: intrigues in the phox family. *J Lab Clin Med* **128**, 461-76 (1996).
102. Nauseef, W. M. The NADPH-dependent oxidase of phagocytes. *Proc Assoc Am Physicians* **111**, 373-82 (1999).
103. Didichenko, S. A., Tilton, B., Hemmings, B. A., Ballmer-Hofer, K. & Thelen, M. Constitutive activation of protein kinase B and phosphorylation of p47phox by a membrane-targeted phosphoinositide 3-kinase. *Curr Biol* **6**, 1271-8 (1996).
104. Lal, A. S., Parker, P. J. & Segal, A. W. Characterization and partial purification of a novel neutrophil membrane-associated kinase capable of phosphorylating the respiratory burst component p47phox. *Biochem J* **338** (Pt 2), 359-66 (1999).
105. Nixon, J. B. & McPhail, L. C. Protein kinase C (PKC) isoforms translocate to Triton-insoluble fractions in stimulated human neutrophils: correlation of conventional PKC with activation of NADPH oxidase. *J Immunol* **163**, 4574-82 (1999).
106. Dang, P. M., Fontayne, A., Hakim, J., El Benna, J. & Perianin, A. Protein kinase C zeta phosphorylates a subset of selective sites of the NADPH oxidase component p47phox and participates in formyl peptide-mediated neutrophil respiratory burst. *J Immunol* **166**, 1206-13 (2001).
107. Korchak, H. M. & Kilpatrick, L. E. Roles for beta II-protein kinase C and RACK1 in positive and negative signaling for superoxide anion generation in differentiated HL60 cells. *J Biol Chem* **276**, 8910-7 (2001).
108. Chen, Q. et al. Akt phosphorylates p47phox and mediates respiratory burst activity in human neutrophils. *J Immunol* **170**, 5302-8 (2003).
109. Hoyal, C. R. et al. Modulation of p47PHOX activity by site-specific phosphorylation: Akt-dependent activation of the NADPH oxidase. *Proc Natl Acad Sci U S A* **100**, 5130-5 (2003).
110. Hirsch, E. et al. Central role for G protein-coupled phosphoinositide 3-kinase gamma in inflammation. *Science* **287**, 1049-53 (2000).
111. Park, H. S. & Park, J. W. Fluorescent labeling of the leukocyte NADPH oxidase subunit p47(phox): evidence for amphiphile-induced conformational changes. *Arch Biochem Biophys* **360**, 165-72 (1998).
112. Park, H. S. et al. Phosphorylation of the leucocyte NADPH oxidase subunit p47(phox) by casein kinase 2: conformation-dependent phosphorylation and modulation of oxidase activity. *Biochem J* **358**, 783-90 (2001).
113. Paclet, M. H., Coleman, A. W., Vergnaud, S. & Morel, F. P67-phox-mediated NADPH oxidase assembly: imaging of cytochrome b558 liposomes by atomic force microscopy. *Biochemistry* **39**, 9302-10 (2000).
114. Vergnaud, S., Paclet, M. H., El Benna, J., Pocardalo, M. A. & Morel, F. Complementation of NADPH oxidase in p67-phox-deficient CGD patients p67-phox/p40-phox interaction. *Eur J Biochem* **267**, 1059-67 (2000).
115. Han, C. H., Freeman, J. L., Lee, T., Motalebi, S. A. & Lambeth, J. D. Regulation of the neutrophil respiratory burst oxidase. Identification of an activation domain in p67(phox). *J Biol Chem* **273**, 16663-8 (1998).
116. Mizrahi, A. et al. Assembly of the phagocyte NADPH oxidase complex: chimeric constructs derived from the cytosolic components as tools for exploring structure-function relationships. *J Leukoc Biol* **79**, 881-95 (2006).
117. Kreck, M. L., Freeman, J. L., Abo, A. & Lambeth, J. D. Membrane association of Rac is required for high activity of the respiratory burst oxidase. *Biochemistry* **35**, 15683-92 (1996).
118. Freeman, J. L. & Lambeth, J. D. NADPH oxidase activity is independent of p47phox in vitro. *J Biol Chem* **271**, 22578-82 (1996).
119. Koshkin, V., Lotan, O. & Pick, E. The cytosolic component p47(phox) is not a sine qua non participant in the activation of NADPH oxidase but is required for optimal superoxide production. *J Biol Chem* **271**, 30326-9 (1996).
120. Gorzalczany, Y., Sigal, N., Itan, M., Lotan, O. & Pick, E. Targeting of Rac1 to the phagocyte membrane is sufficient for the induction of NADPH oxidase assembly. *J Biol Chem* **275**, 40073-81 (2000).

121. Gorzalczany, Y., Alloul, N., Sigal, N., Weinbaum, C. & Pick, E. A prenylated p67phox-Rac1 chimera elicits NADPH-dependent superoxide production by phagocyte membranes in the absence of an activator and of p47phox: conversion of a pagan NADPH oxidase to monotheism. *J Biol Chem* **277**, 18605-10 (2002).
122. Taylor, R. M. et al. Analysis of human phagocyte flavocytochrome b(558) by mass spectrometry. *J Biol Chem* **281**, 37045-56 (2006).
123. Mayer, B. J. SH3 domains: complexity in moderation. *J Cell Sci* **114**, 1253-63 (2001).
124. Dimasi, N. Crystal structure of the C-terminal SH3 domain of the adaptor protein GADS in complex with SLP-76 motif peptide reveals a unique SH3-SH3 interaction. *Int J Biochem Cell Biol* **39**, 109-23 (2007).
125. Dalgarno, D. C., Botfield, M. C. & Rickles, R. J. SH3 domains and drug design: ligands, structure, and biological function. *Biopolymers* **43**, 383-400 (1997).
126. Pawson, T. Protein modules and signalling networks. *Nature* **373**, 573-80 (1995).
127. Lim, W. A., Richards, F. M. & Fox, R. O. Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains. *Nature* **372**, 375-9 (1994).
128. Feng, S., Kasahara, C., Rickles, R. J. & Schreiber, S. L. Specific interactions outside the proline-rich core of two classes of Src homology 3 ligands. *Proc Natl Acad Sci U S A* **92**, 12408-15 (1995).
129. Musacchio, A. How SH3 domains recognize proline. *Adv Protein Chem* **61**, 211-68 (2002).
130. Morken, J. P., Kapoor, T. M., Feng, S., Shirai, F. & Schreiber, S. L. Exploring the leucine-proline binding pocket of the Src SH3 domain using structure-based, split-pool synthesis and affinity-based selection. *JACS* **120**, 30-36 (1998).
131. Feng, S., Chen, J. K., Yu, H., Simon, J. A. & Schreiber, S. L. Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions. *Science* **266**, 1241-7 (1994).
132. Wu, X. et al. Structural basis for the specific interaction of lysine-containing proline-rich peptides with the N-terminal SH3 domain of c-Crk. *Structure* **3**, 215-26 (1995).
133. de Mendez, I., Adams, A. G., Sokolic, R. A., Malech, H. L. & Leto, T. L. Multiple SH3 domain interactions regulate NADPH oxidase assembly in whole cells. *Embo J* **15**, 1211-20 (1996).
134. de Mendez, I., Homayounpour, N. & Leto, T. L. Specificity of p47phox SH3 domain interactions in NADPH oxidase assembly and activation. *Mol Cell Biol* **17**, 2177-85 (1997).
135. Ferraro, E., Peluso, D., Via, A., Ausiello, G. & Helmer-Citterich, M. SH3-Hunter: discovery of SH3 domain interaction sites in proteins. *Nucleic Acids Res* **35**, W451-4 (2007).
136. Fernandez-Ballester, G., Blanes-Mira, C. & Serrano, L. The tryptophan switch: changing ligand-binding specificity from type I to type II in SH3 domains. *J Mol Biol* **335**, 619-29 (2004).
137. Witter, D. J., Famiglietti, S. J., Cambier, J. C. & Castelano, A. L. Design and synthesis of SH3 domain binding ligands: modifications of the consensus sequence XPPXP. *Bioorg Med Chem Lett* **8**, 3137-42 (1998).
138. Inglis, S. R. et al. Identification and specificity studies of small-molecule ligands for SH3 protein domains. *J Med Chem* **47**, 5405-17 (2004).
139. Inglis, S. et al. Synthesis of 5-, 6- and 7-substituted-2-aminoquinolines as SH3 domain ligands. *Org Biomol Chem* **3**, 2543-57 (2005).
140. Ball, L. J., Kuhne, R., Schneider-Mergener, J. & Oschkinat, H. Recognition of proline-rich motifs by protein-protein-interaction domains. *Angew Chem Int Ed Engl* **44**, 2852-69 (2005).
141. Kay, B. K., Williamson, M. P. & Sudol, M. The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *Faseb J* **14**, 231-41 (2000).
142. Huang, X. et al. Structure of a WW domain containing fragment of dystrophin in complex with beta-dystroglycan. *Nat Struct Biol* **7**, 634-8 (2000).
143. Macias, M. J. et al. Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide. *Nature* **382**, 646-9 (1996).
144. Yu, H. et al. Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell* **76**, 933-45 (1994).
145. Sparks, A. B. et al. Distinct ligand preferences of Src homology 3 domains from Src, Yes, Abl, Cortactin, p53bp2, PLCgamma, Crk, and Grb2. *Proc Natl Acad Sci U S A* **93**, 1540-4 (1996).
146. Reinhard, M. et al. The proline-rich focal adhesion and microfilament protein VASP is a ligand for profilins. *Embo J* **14**, 1583-9 (1995).
147. Reinhard, M., Jouvenal, K., Tripiet, D. & Walter, U. Identification, purification, and characterization of a zyxin-related protein that binds the focal adhesion and microfilament

- protein VASP (vasodilator-stimulated phosphoprotein). *Proc Natl Acad Sci U S A* **92**, 7956-60 (1995).
148. Reinhard, M., Rudiger, M., Jockusch, B. M. & Walter, U. VASP interaction with vinculin: a recurring theme of interactions with proline-rich motifs. *FEBS Lett* **399**, 103-7 (1996).
 149. Schutt, C. E., Myslik, J. C., Rozycki, M. D., Goonesekere, N. C. & Lindberg, U. The structure of crystalline profilin-beta-actin. *Nature* **365**, 810-6 (1993).
 150. Mahoney, N. M., Janmey, P. A. & Almo, S. C. Structure of the profilin-poly-L-proline complex involved in morphogenesis and cytoskeletal regulation. *Nat Struct Biol* **4**, 953-60 (1997).
 151. Chen, H. I. & Sudol, M. The WW domain of Yes-associated protein binds a proline-rich ligand that differs from the consensus established for Src homology 3-binding modules. *Proc Natl Acad Sci U S A* **92**, 7819-23 (1995).
 152. Nishizawa, K., Freund, C., Li, J., Wagner, G. & Reinherz, E. L. Identification of a proline-binding motif regulating CD2-triggered T lymphocyte activation. *Proc Natl Acad Sci U S A* **95**, 14897-902 (1998).
 153. Pornillos, O., Alam, S. L., Davis, D. R. & Sundquist, W. I. Structure of the Tsg101 UEV domain in complex with the PTAP motif of the HIV-1 p6 protein. *Nat Struct Biol* **9**, 812-7 (2002).
 154. Pornillos, O. et al. Structure and functional interactions of the Tsg101 UEV domain. *Embo J* **21**, 2397-406 (2002).
 155. Musacchio, A., Saraste, M. & Wilmanns, M. High-resolution crystal structures of tyrosine kinase SH3 domains complexed with proline-rich peptides. *Nat Struct Biol* **1**, 546-51 (1994).
 156. Feng, S., Kapoor, T. M., Shirai, F., Combs, A. P. & Schreiber, S. L. Molecular basis for the binding of SH3 ligands with non-peptide elements identified by combinatorial synthesis. *Chem Biol* **3**, 661-70 (1996).
 157. Renzoni, D. A. et al. Structural and thermodynamic characterization of the interaction of the SH3 domain from Fyn with the proline-rich binding site on the p85 subunit of PI3-kinase. *Biochemistry* **35**, 15646-53 (1996).
 158. Pisabarro, M. T., Serrano, L. & Wilmanns, M. Crystal structure of the abl-SH3 domain complexed with a designed high-affinity peptide ligand: implications for SH3-ligand interactions. *J Mol Biol* **281**, 513-21 (1998).
 159. Vidal, M. et al. Molecular and cellular analysis of Grb2 SH3 domain mutants: interaction with Sos and dynamin. *J Mol Biol* **290**, 717-30 (1999).
 160. Wittekind, M. et al. Solution structure of the Grb2 N-terminal SH3 domain complexed with a ten-residue peptide derived from SOS: direct refinement against NOEs, J-couplings and ¹H and ¹³C chemical shifts. *J Mol Biol* **267**, 933-52 (1997).
 161. Ghose, R., Shekhtman, A., Goger, M. J., Ji, H. & Cowburn, D. A novel, specific interaction involving the Csk SH3 domain and its natural ligand. *Nat Struct Biol* **8**, 998-1004 (2001).
 162. MacArthur, M. W. & Thornton, J. M. Influence of proline residues on protein conformation. *J Mol Biol* **218**, 397-412 (1991).
 163. Holmgren, S. K., Bretscher, L. E., Taylor, K. M. & Raines, R. T. A hyperstable collagen mimic. *Chem Biol* **6**, 63-70 (1999).
 164. Ruzza, P. et al. Introduction of N-alkyl residues in proline-rich peptides: effect on SH3 binding affinity and peptide conformation. *Adv Exp Med Biol* **611**, 65-6 (2009).
 165. Jaquet, V., Scapozza, L., Clark, R. A., Krause, K. H. & Lambeth, J. D. Small-molecule NOX inhibitors: ROS-generating NADPH oxidases as therapeutic targets. *Antioxid Redox Signal* **11**, 2535-52 (2009).
 166. Ellis, J. A., Mayer, S. J. & Jones, O. T. The effect of the NADPH oxidase inhibitor diphenyleneiodonium on aerobic and anaerobic microbicidal activities of human neutrophils. *Biochem J* **251**, 887-91 (1988).
 167. Poinas, A., Gaillard, J., Vignais, P. & Doussiere, J. Exploration of the diaphorase activity of neutrophil NADPH oxidase. *Eur J Biochem* **269**, 1243-52 (2002).
 168. Yamamoto, E. et al. Excess salt causes cerebral neuronal apoptosis and inflammation in stroke-prone hypertensive rats through angiotensin II-induced NADPH oxidase activation. *Stroke* **39**, 3049-56 (2008).
 169. Oelze, M. et al. Nebivolol inhibits superoxide formation by NADPH oxidase and endothelial dysfunction in angiotensin II-treated rats. *Hypertension* **48**, 677-84 (2006).
 170. Hancock, J. T. & Jones, O. T. The inhibition by diphenyleneiodonium and its analogues of superoxide generation by macrophages. *Biochem J* **242**, 103-7 (1987).
 171. O'Donnell, B. V., Tew, D. G., Jones, O. T. & England, P. J. Studies on the inhibitory mechanism of iodonium compounds with special reference to neutrophil NADPH oxidase. *Biochem J* **290** (Pt 1), 41-9 (1993).

172. Cross, A. R. Inhibitors of the leukocyte superoxide generating oxidase: mechanisms of action and methods for their elucidation. *Free Radic Biol Med* **8**, 71-93 (1990).
173. Le Cabec, V. & Maridonneau-Parini, I. Complete and reversible inhibition of NADPH oxidase in human neutrophils by phenylarsine oxide at a step distal to membrane translocation of the enzyme subunits. *J Biol Chem* **270**, 2067-73 (1995).
174. Doussiere, J., Poinas, A., Blais, C. & Vignais, P. V. Phenylarsine oxide as an inhibitor of the activation of the neutrophil NADPH oxidase--identification of the beta subunit of the flavocytochrome b component of the NADPH oxidase as a target site for phenylarsine oxide by photoaffinity labeling and photoinactivation. *Eur J Biochem* **251**, 649-58 (1998).
175. Lassegue, B. & Griendling, K. K. Mycophenolic acid is a new Nox2 inhibitor. *Hypertension* **49**, 25-6 (2007).
176. Waring, P., Eichner, R. D. & Mullbacher, A. The chemistry and biology of the immunomodulating agent gliotoxin and related epipolythiodioxopiperazines. *Med Res Rev* **8**, 499-524 (1988).
177. Tsunawaki, S., Yoshida, L. S., Nishida, S., Kobayashi, T. & Shimoyama, T. Fungal metabolite gliotoxin inhibits assembly of the human respiratory burst NADPH oxidase. *Infect Immun* **72**, 3373-82 (2004).
178. Decoursey, T. E. & Ligeti, E. Regulation and termination of NADPH oxidase activity. *Cell Mol Life Sci* **62**, 2173-93 (2005).
179. de Groot, A. A., Mathy, M. J., van Zwieten, P. A. & Peters, S. L. Antioxidant activity of nebivolol in the rat aorta. *J Cardiovasc Pharmacol* **43**, 148-53 (2004).
180. Stolk, J., Hiltermann, T. J., Dijkman, J. H. & Verhoeven, A. J. Characteristics of the inhibition of NADPH oxidase activation in neutrophils by apocynin, a methoxy-substituted catechol. *Am J Respir Cell Mol Biol* **11**, 95-102 (1994).
181. Johnson, D. K. et al. Inhibition of NADPH oxidase activation in endothelial cells by ortho-methoxy-substituted catechols. *Endothelium* **9**, 191-203 (2002).
182. Barbieri, S. S. et al. Apocynin prevents cyclooxygenase 2 expression in human monocytes through NADPH oxidase and glutathione redox-dependent mechanisms. *Free Radic Biol Med* **37**, 156-65 (2004).
183. Simons, J. M., Hart, B. A., Ip Vai Ching, T. R., Van Dijk, H. & Labadie, R. P. Metabolic activation of natural phenols into selective oxidative burst agonists by activated human neutrophils. *Free Radic Biol Med* **8**, 251-8 (1990).
184. Gajewska, G., Ganowiak, Z., Grzybowski, J., Radecki, A. & Wrzesniowska, K. Analysis of an industrial smoke preparation Part VI; acute toxicity, bactericidal, and antioxidative activities of some phenolic fraction components of industrial smoke preparation. *Bromat. Chem. Toksykol.* **XIV**, 301-305 (1981).
185. Cayatte, A. J. et al. S17834, a new inhibitor of cell adhesion and atherosclerosis that targets nadph oxidase. *Arterioscler Thromb Vasc Biol* **21**, 1577-84 (2001).
186. Kalinowski, L., Dobrucki, I. T. & Malinski, T. Race-specific differences in endothelial function: predisposition of African Americans to vascular diseases. *Circulation* **109**, 2511-7 (2004).
187. Zang, M. et al. Polyphenols stimulate AMP-activated protein kinase, lower lipids, and inhibit accelerated atherosclerosis in diabetic LDL receptor-deficient mice. *Diabetes* **55**, 2180-91 (2006).
188. Yamamoto, T. & Yamada, K. Specific NAD(P)H oxidase inhibitor. WO 2004/089412. *Osaka-shi, Japan. Mitsubishi Pharma Corporation.* (2006).
189. Stielow, C. et al. Novel Nox inhibitor of oxLDL-induced reactive oxygen species formation in human endothelial cells. *Biochem Biophys Res Commun* **344**, 200-5 (2006).
190. Droge, W. Free radicals in the physiological control of cell function. *Physiol Rev* **82**, 47-95 (2002).
191. Page, P., Orchard, M., Fioraso-Cartier, L. & Mottironi, B. Pyrazolo pyridine derivatives as NADPH oxidase inhibitors. WO 2008/113856 A1. *Genkyotex (CH)* (2008).
192. Page, P., Orchard, M., Fioraso-Cartier, L. & Mottironi, B. Tetrahydroindole derivatives as NADPH oxidase inhibitors. WO 2008/116926 A1. *Genkyotex (CH)* (2008).
193. Seno, K., Nishi, K., Matsuo, Y. & Fujishita, T. Pyrazolo [1,5-A] pyrimidine derivative and NAD(P)H oxidase inhibitor containing the same. US 2006/0089362. *Osaka, Japan. Shionogi & Co, Ltd.* (2006).
194. Higdon, J. V. & Frei, B. Tea catechins and polyphenols: health effects, metabolism, and antioxidant functions. *Crit Rev Food Sci Nutr* **43**, 89-143 (2003).
195. Nakagawa, T. & Yokozawa, T. Direct scavenging of nitric oxide and superoxide by green tea. *Food Chem Toxicol* **40**, 1745-50 (2002).

196. Brown, M. D. Green tea (*Camellia sinensis*) extract and its possible role in the prevention of cancer. *Altern Med Rev* **4**, 360-70 (1999).
197. Wang, H. K. The therapeutic potential of flavonoids. *Expert Opin Investig Drugs* **9**, 2103-19 (2000).
198. Sang, S., Lee, M. J., Hou, Z., Ho, C. T. & Yang, C. S. Stability of tea polyphenol (-)-epigallocatechin-3-gallate and formation of dimers and epimers under common experimental conditions. *J Agric Food Chem* **53**, 9478-84 (2005).
199. Cos, P. et al. Structure-activity relationship and classification of flavonoids as inhibitors of xanthine oxidase and superoxide scavengers. *J Nat Prod* **61**, 71-6 (1998).
200. Nagao, A., Seki, M. & Kobayashi, H. Inhibition of xanthine oxidase by flavonoids. *Biosci Biotechnol Biochem* **63**, 1787-90 (1999).
201. Steffen, Y., Schewe, T. & Sies, H. (-)-Epicatechin elevates nitric oxide in endothelial cells via inhibition of NADPH oxidase. *Biochem Biophys Res Commun* **359**, 828-33 (2007).
202. Nishikawa, H., Wakano, K. & Kitani, S. Inhibition of NADPH oxidase subunits translocation by tea catechin EGCG in mast cell. *Biochem Biophys Res Commun* **362**, 504-9 (2007).
203. Kiaei, M. et al. Celastrol blocks neuronal cell death and extends life in transgenic mouse model of amyotrophic lateral sclerosis. *Neurodegener Dis* **2**, 246-54 (2005).
204. Zhang, T. et al. A novel Hsp90 inhibitor to disrupt Hsp90/Cdc37 complex against pancreatic cancer cells. *Mol Cancer Ther* **7**, 162-70 (2008).
205. Masic, L. P. Arginine mimetic structures in biologically active antagonists and inhibitors. *Curr Med Chem* **13**, 3627-48 (2006).
206. Danilewicz, J. C. et al. Design of selective thrombin inhibitors based on the (R)-Phe-Pro-Arg sequence. *J Med Chem* **45**, 2432-53 (2002).
207. Schwartz, J. J. & Zhang, S. Peptide-mediated cellular delivery. *Curr Opin Mol Ther* **2**, 162-7 (2000).
208. Zhang, R., Brownnewell, F. & Madalengoitia, J. S. Pseudo-A(1,3) strain as a key conformational control element in the design of poly-L-proline type II peptide mimics. *JACS* **120**, 3894-3902 (1998).
209. Zhang, R. & Madalengoitia, J. S. Design, synthesis and evaluation of poly-L-proline type-II peptide mimics based on the 3-azabicyclo[3.1.0]hexane system. *J Org Chem* **64**, 330-331 (1999).
210. Adzhubei, A. A. & Sternberg, M. J. Left-handed polyproline II helices commonly occur in globular proteins. *J Mol Biol* **229**, 472-93 (1993).
211. Nguyen, J. T., Turck, C. W., Cohen, F. E., Zuckermann, R. N. & Lim, W. A. Exploiting the basis of proline recognition by SH3 and WW domains: design of N-substituted inhibitors. *Science* **282**, 2088-92 (1998).
212. Nguyen, J. T. et al. Improving SH3 domain ligand selectivity using a non natural scaffold. *Chem Biol* **7**, 463-73 (2000).
213. Kapoor, T. M., Hamilton Andreotti, A. & Schreiber, S. L. Exploring the specificity pockets of two homologous SH3 domains using structure-based, split-pool synthesis and affinity-based selection. *JACS* **120**, 23-29 (1998).
214. Selemidis, S., Sobey, C. G., Wingler, K., Schmidt, H. H. & Drummond, G. R. NADPH oxidases in the vasculature: molecular features, roles in disease and pharmacological inhibition. *Pharmacol Ther* **120**, 254-91 (2008).
215. Shi, J., Ross, C. R., Leto, T. L. & Blecha, F. PR-39, a proline-rich antibacterial peptide that inhibits phagocyte NADPH oxidase activity by binding to Src homology 3 domains of p47 phox. *Proc Natl Acad Sci U S A* **93**, 6014-8 (1996).
216. Li, J. et al. PR39, a peptide regulator of angiogenesis. *Nat Med* **6**, 49-55 (2000).
217. Korthuis, R. J., Gute, D. C., Blecha, F. & Ross, C. R. PR-39, a proline/arginine-rich antimicrobial peptide, prevents postischemic microvascular dysfunction. *Am J Physiol* **277**, H1007-13 (1999).
218. Hoffmeyer, M. R., Scalia, R., Ross, C. R., Jones, S. P. & Lefer, D. J. PR-39, a potent neutrophil inhibitor, attenuates myocardial ischemia-reperfusion injury in mice. *Am J Physiol Heart Circ Physiol* **279**, H2824-8 (2000).
219. Bao, J. et al. PR-39 and PR-11 peptides inhibit ischemia-reperfusion injury by blocking proteasome-mediated I kappa B alpha degradation. *Am J Physiol Heart Circ Physiol* **281**, H2612-8 (2001).
220. Ikeda, Y., Young, L. H., Scalia, R., Ross, C. R. & Lefer, A. M. PR-39, a proline/arginine-rich antimicrobial peptide, exerts cardioprotective effects in myocardial ischemia-reperfusion. *Cardiovasc Res* **49**, 69-77 (2001).

221. Wu, J. et al. PR39 inhibits apoptosis in hypoxic endothelial cells: role of inhibitor apoptosis protein-2. *Circulation* **109**, 1660-7 (2004).
222. Sudol, M. From Src Homology domains to other signaling modules: proposal of the 'protein recognition code'. *Oncogene* **17**, 1469-74 (1998).
223. Ren, R., Mayer, B. J., Cicchetti, P. & Baltimore, D. Identification of a ten-amino acid proline-rich SH3 binding site. *Science* **259**, 1157-61 (1993).
224. Watanabe, Y. et al. Molecular dynamics study on the ligand recognition by tandem SH3 domains of p47phox, regulating NADPH oxidase activity. *Comput Biol Chem* **30**, 303-12 (2006).
225. Creamer, T. P. Left-handed polyproline II helix formation is (very) locally driven. *Proteins* **33**, 218-26 (1998).
226. Williamson, M. P. The structure and function of proline-rich regions in proteins. *Biochem J* **297** (Pt 2), 249-60 (1994).
227. Petrella, E. C., Machesky, L. M., Kaiser, D. A. & Pollard, T. D. Structural requirements and thermodynamics of the interaction of proline peptides with profilin. *Biochemistry* **35**, 16535-43 (1996).
228. Grosdidier, A., Zoete, V. & Michielin, O. EADock: docking of small molecules into protein active sites with a multiobjective evolutionary optimization. *Proteins* **67**, 1010-25 (2007).
229. Nobuhisa, I. et al. Activation of the superoxide-producing phagocyte NADPH oxidase requires co-operation between the tandem SH3 domains of p47phox in recognition of a polyproline type II helix and an adjacent alpha-helix of p22phox. *Biochem J* **396**, 183-92 (2006).
230. Leto, T. L., Adams, A. G. & de Mendez, I. Assembly of the phagocyte NADPH oxidase: binding of Src homology 3 domains to proline-rich targets. *Proc Natl Acad Sci U S A* **91**, 10650-4 (1994).
231. Leusen, J. H. et al. 156Pro-->Gln substitution in the light chain of cytochrome b558 of the human NADPH oxidase (p22-phox) leads to defective translocation of the cytosolic proteins p47-phox and p67-phox. *J Exp Med* **180**, 2329-34 (1994).
232. Hata, K., Takeshige, K. & Sumimoto, H. Roles for proline-rich regions of p47phox and p67phox in the phagocyte NADPH oxidase activation in vitro. *Biochem Biophys Res Commun* **241**, 226-31 (1997).
233. Zarrinpar, A., Bhattacharyya, R. P. & Lim, W. A. The structure and function of proline recognition domains. *Sci STKE* **2003**, RE8 (2003).
234. Kami, K., Takeya, R., Sumimoto, H. & Kohda, D. Diverse recognition of non-PxxP peptide ligands by the SH3 domains from p67(phox), Grb2 and Pex13p. *Embo J* **21**, 4268-76 (2002).
235. Nisimoto, Y., Motalebi, S., Han, C. H. & Lambeth, J. D. The p67(phox) activation domain regulates electron flow from NADPH to flavin in flavocytochrome b(558). *J Biol Chem* **274**, 22999-3005 (1999).
236. Dang, P. M., Cross, A. R., Quinn, M. T. & Babior, B. M. Assembly of the neutrophil respiratory burst oxidase: a direct interaction between p67PHOX and cytochrome b558 II. *Proc Natl Acad Sci U S A* **99**, 4262-5 (2002).
237. Xu, Y. C. et al. Involvement of TRAF4 in oxidative activation of c-Jun N-terminal kinase. *J Biol Chem* **277**, 28051-7 (2002).
238. Li, J. M., Fan, L. M., Christie, M. R. & Shah, A. M. Acute tumor necrosis factor alpha signaling via NADPH oxidase in microvascular endothelial cells: role of p47phox phosphorylation and binding to TRAF4. *Mol Cell Biol* **25**, 2320-30 (2005).
239. Charifson, P. S. & Walters, W. P. Filtering databases and chemical libraries. *Mol Divers* **5**, 185-97 (2002).
240. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* **46**, 3-26 (2001).
241. Ghose, A. K., Viswanadhan, V. N. & Wendoloski, J. J. A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *J Comb Chem* **1**, 55-68 (1999).
242. Oprea, T. I. Property distribution of drug-related chemical databases. *J Comput Aided Mol Des* **14**, 251-64 (2000).
243. Clark, D. E. & Pickett, S. D. Computational methods for the prediction of 'drug-likeness'. *Drug Discov Today* **5**, 49-58 (2000).
244. Hann, M. M. & Oprea, T. I. Pursuing the leadlikeness concept in pharmaceutical research. *Curr Opin Chem Biol* **8**, 255-63 (2004).

245. Olah, M. M., Bologa, C. G. & Oprea, T. I. Strategies for compound selection. *Curr Drug Discov Technol* **1**, 211-20 (2004).
246. Knox, A. J., Meegan, M. J., Carta, G. & Lloyd, D. G. Considerations in compound database preparation--"hidden" impact on virtual screening results. *J Chem Inf Model* **45**, 1908-19 (2005).
247. Jaquet, V. et al. Nox NADPH oxidase isoforms are inhibited by celastrol with a dual mode of action. *to be submitted* (2010).
248. Jain, A. N. Ligand-based structural hypotheses for virtual screening. *J Med Chem* **47**, 947-61 (2004).
249. Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* **267**, 727-48 (1997).
250. Rarey, M., Kramer, B., Lengauer, T. & Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J Mol Biol* **261**, 470-89 (1996).
251. Rarey, M., Kramer, B. & Lengauer, T. The particle concept: placing discrete water molecules during protein-ligand docking predictions. *Proteins* **34**, 17-28 (1999).
252. Rarey, M., Kramer, B. & Lengauer, T. Docking of hydrophobic ligands with interaction-based matching algorithms. *Bioinformatics* **15**, 243-50 (1999).
253. Ewing, T. J., Makino, S., Skillman, A. G. & Kuntz, I. D. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J Comput Aided Mol Des* **15**, 411-28 (2001).
254. Moustakas, D. T. et al. Development and validation of a modular, extensible docking program: DOCK 5. *J Comput Aided Mol Des* **20**, 601-19 (2006).
255. Lang, P. T. et al. DOCK 6: combining techniques to model RNA-small molecule complexes. *Rna* **15**, 1219-30 (2009).
256. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105-32 (1982).
257. Parreira, R. L. T., Abrahão, O. & Galembeck, S. E. Conformational preferences of non-nucleoside HIV-1 reverse transcriptase inhibitors. *Tetrahedron* **57**, 3243-3253 (2001).
258. De Clercq, E. What can be Expected from Non-nucleoside Reverse Transcriptase Inhibitors (NNRTIs) in the Treatment of Human Immunodeficiency Virus Type 1 (HIV-1) Infections? *Rev Med Virol* **6**, 97-117 (1996).
259. Karapetyan, H., Tamazyan, R., Martirosyan, A., Hovhannesian, V. & Gasparyan, S. 1-Substituted derivatives of 2-aryl-5-oxopyrrolidine-2-carboxylic acid. *Acta Crystallogr C* **58**, o399-401 (2002).
260. Tamazyan, R. et al. 1,2,5-Substituted derivatives of 2-phenylpyrrolidine. *Acta Crystallogr C* **60**, o390-2 (2004).
261. Swain, S. D., Helgerson, S. L., Davis, A. R., Nelson, L. K. & Quinn, M. T. Analysis of activation-induced conformational changes in p47phox using tryptophan fluorescence spectroscopy. *J Biol Chem* **272**, 29502-10 (1997).
262. Jault, J. M. et al. The HPr kinase from *Bacillus subtilis* is a homo-oligomeric enzyme which exhibits strong positive cooperativity for nucleotide and fructose 1,6-bisphosphate binding. *J Biol Chem* **275**, 1773-80 (2000).
263. Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem* **28**, 849-57 (1985).

CHAPTER 4 - *IN SILICO* SCREENING FOR TARGETS OF PHENAZOPYRIDINE: STATE OF THE ART, VALIDATION AND OPTIMIZATION OF SCORING EVALUATION/TARGET SELECTION

ABSTRACT

Based on experimental data suggesting that phenazopyridine induces the homogeneous differentiation of embryonic CGR8 stem cells into neuronal cells ¹ and the need for target finding, structure-activity relationships (SARs) and inverse *in silico* screening were applied as complements to functional inhibition tests and competitive binding assays. Besides a discussion on the strengths and weaknesses of inverse screening, a typical setup/running procedure, as well as results and strategies for optimizing the selection are presented.

Two different *in silico* protocol validations were performed using GOLD 4.0 ²:

- (i) An unbiased validation: 12 ligands were chosen randomly and their corresponding targets seeded into a database of 188 decoy proteins. Of the ligands docked into their known targets, 58% displayed root mean square deviations (RMSDs) ≤ 2 Å compared to the experimentally measured binding orientation and 75% were found in the first 3.5% of the interaction energy-based ranking, discriminating between “true” and supposedly “false” targets.
- (ii) A knowledge-based validation: Using well-studied targets with ligands of different specificity degrees, the focus was on docking/scoring accuracy and enrichment. A threshold at 10% of the ranking was defined to eliminate unspecific binders.

Subsequently, phenazopyridine was docked into over 5000 protein crystal structures. Docking the corresponding co-crystallized ligand into each protein served as a control. Selection criteria included (i) the interaction energy-based ranking of phenazopyridine, (ii) the control ligand RMSD, (iii) shared key target interactions between phenazopyridine and the control as well as (iv) publications reporting on the biological context. Functional inhibition tests and competitive binding assays served as experimental validation.

By applying three of the four selection criteria to *in silico* screening, 80% of the experimentally identified targets were retrieved, providing a rough selectivity profile of phenazopyridine.

INTRODUCTION

The basic principle of inverse screening

Classical structure-based virtual screening relies on *in silico* binding predictions of a wealth of compounds to a given protein target and aims at fishing out (specific) ligands. It is a forward, target-driven docking approach, where the target space serves as filter. An overview on successes and gaps in docking and scoring can be found in Leach et al. ³. Forward *in silico* screening is opposed to inverse screening, which is ligand-driven and aims at identifying a specific target for a given ligand. An intermediate category can be termed interaction- or complex-driven and is based on fingerprints which encode the protein-ligand interactions in bit strings (e.g. applied to profiling different classes of kinase inhibitors with respect to one target ⁴).

Here, *in silico* target prediction for target discovery is treated as a complementary approach to the work- and time-intensive wet-lab studies such as chemical proteomics. Based on genomic or protein databases and followed up by experimental validation, target finding can be more efficiently achieved, as has been shown by several studies ⁵⁻¹⁰.

Studies on target identification with a small molecular weight compound are either prospective ^{9,10} or retrospective ^{6,8,11}. These studies show that inverse screening cannot only be applied to *de novo* target discovery ^{9,11}, but also to the re-examination of the target space of already established inhibitors (multi-target mechanisms of action) ^{8,10,11}, target de-orphanisation or toxicity/side-effect prediction ⁶. The inverse screening technique is seen as part of a rational approach maximizing the likelihood for achieving target-selective inhibition with minimal *in vivo* side effects. More specifically, this rational

approach consists in (i) target selection based on gene expression, tissue distribution, and disease relevance, (ii) target assessment using the 3D target binding site with respect to druggability and selectivity as well as (iii) target validation evidencing a functional role *in vitro*¹². It is also connected to approaches relating protein pharmacology to ligand similarity¹³⁻¹⁶.

Docking programs for inverse in silico screening

The interest in target discovery has so far led to several modifications of docking programs commonly used in virtual screening campaigns: INVDOCK⁷, FlexX¹⁷, and GOLD⁸. Their basic principles, success stories and weaknesses are shortly presented hereafter and reasons for the choice of GOLD in this work are given.

The very first reverse docking algorithm to be developed was INVDOCK, which was derived from the shape-based DOCK program¹⁸. It renders ligands as abstract spheres. A successful application was the correct prediction of 83% of known toxicity and side-effects of drugs⁶.

In the inverse FlexX approach, the protein is incrementally constructed around the ligand. A comparison of forward and reverse docking on 10 relatively flexible receptor-ligand complexes from the Cambridge Structural Database (CSD) showed that the accuracy was comparable, but that the reverse docking was generally faster¹⁷. Its enhanced speed was due to the restricted conformational search around the ligand compared to the search in a ligand binding pocket.

GOLD^{2,19,20} uses a genetic algorithm. Such algorithms are known to perform well for generating a large number of low-energy conformers. The Goldscore^{19,21} predicts binding energies as a sum of H-bonds derived from donor/acceptor pairs and geometries, internal VdW energy, and steric interactions between the ligand and the protein^{2,19}.

The choice of GOLD for the work discussed here has several reasons: INVDOCK has a less efficient docking/scoring method than other currently used docking programs, the inverse FlexX approach was not available in-house, while GOLD was shown to be highly accurate for a variety of targets⁸ and was indeed at hand. GOLD selects the degrees of freedom in the binding site that correspond to reorientations of H-bond donors and acceptors². In a study of Kontoyianni et al. comparing five often used docking programs, GOLD performed best in reproducing experimentally observed binding modes²². In yet another study comparing GOLD to two other docking programs with respect to pharmaceutically relevant targets, it was found to have a performance that is rather binding site-dependent and poorer for predominantly hydrophobic interactions²³. GOLD thus has a tendency to favour hydrophilic interactions. With respect to the side chain flexibility accounted for in GOLD, it cannot be considered in an inverse screening of a large database, as this would require the user to prepare each target individually. However, some degrees of protein flexibility are accounted for here by the presence of the same protein crystallized with various ligands in the target database.

Target database requirements and web-based inverse screening

Given that *in silico* screening strongly depends upon the amount and quality of information available on the system under investigation²⁵, a well-defined collection of 3D coordinates²⁴ is required. The ideal case from a drug discovery perspective would be a large, up-to-date, publicly (online) available target database with pathology-related information containing wild-type and mutant high resolution X-ray or NMR protein structures, homology models, and selected MD snapshots, together with reproducible experimental affinity data for a range of ligands. Work in this direction is ongoing with databases covering both PDB structures and experimental affinities like the AffinDB²⁶, the BindingDB²⁷⁻²⁹ and the BindingMOAD^{30,31}.

With the launch of TarFisDock (Target Fishing Docking: <http://www.dddc.ac.cn/TarFisDock/>)¹¹, a web-server for inverse docking has become available to the scientific community. The Potential Drug Target Database (PDTD: <http://www.dddc.ac.cn/pdtd/>)³² constitutes its corresponding target database, which is highly enriched in enzymes. The protein-ligand complexes included originate from the Therapeutic Target Database (TTD: known therapeutic targets, diseases, and drugs)³³, the DrugBank (bio-/cheminformatic repository of drug data with the associated target information)^{34,35}, and Thomson Pharma (diseases, biological functions, and associated signalling pathways). Validation of the TarFisDock approach was performed by a search for putative targets for vitamin E and 4-OH tamoxifen¹¹. This approach inherits the limitations of its Dock-based docking algorithm and has a

limited number of available known and potential targets (841 targets represented by 1207 structural entries). Other target databases applicable to inverse screening for navigating the structural genome include a screening collection of druggable binding sites derived from the PDB (sc-PDB⁸) and the human GPCR database³⁶, both being useful tools for simultaneous potency and selectivity assessments. Hereafter, the sc-PDB (<http://bioinfo-pharma.u-strasbg.fr/scPDB/>)⁸ is used, because it was specifically developed to identify putative binding sites for screening purposes, emphasize molecular similarities in cavities, and establish links between ligands and active site features. Furthermore, it can be used to benchmark docking, predict targets, compare the chemical similarity of binding sites or link the ligand with the target space⁸.

Successful examples of inverse screening

Successful applications of the concept of *in silico* screening to target identification are cited in the following. Inverse screening has been applied to kinase inhibitor targets¹⁰ to tackle the broad promiscuity of inhibitors causing off-target side effects. Particularly, the phosphoinositide-dependent kinase 1 (PDK1) was identified as a target for an indirubin derivative. This finding was experimentally supported by an *in vitro* kinase assay, the intracellular inhibition of phosphorylation of PDK1 substrates, and the inhibition of endothelial cell migration (known to be highly dependent on PDK1). Besides, two inhibitors of human secreted phospholipase A2 with μM affinities have been discovered by inverse screening on the sc-PDB⁸ as the most likely targets of a series of compounds sharing the same molecular scaffold⁹.

Limitations of inverse screening

Obviously, inverse screening inherits the limitations of the docking tool considered. Considerable advances have been made in the field, but docking accuracy (the prediction of binding orientation), scoring (the prediction of absolute binding free energy, the most ambitious and unfulfilled goal in computer-aided drug design³) and ranking (the discrimination of active from random compounds or targets) still depend on the protein target and physicochemical characteristics of the protein-ligand interactions^{8,37}.

Many docking programs are capable of successfully reproducing crystallographic binding modes, but have difficulties or fail with attributing a score reflecting experimental affinities³⁸, even if some studies show a correlation between docking scores and binding affinities³⁹⁻⁴⁵. Head-to-head comparisons of docking programs and/or scoring functions⁴⁶⁻⁴⁹ and comparative studies involving more than two programs^{22,23,37,50-60} are relatively frequently found in the literature, illustrating that the choice of the best docking-scoring method is still rather difficult and needs appropriate assessment.

Importantly, (i) targets can only be found if they have been previously included in the considered target database and (ii) inverse screening should only be applied to molecules which are assumed to be relatively specific. Some specificity is needed as broad binders easily insert themselves into non-specific cavities and carry the risk of generating biologically irrelevant target associations.

It has to be clearly stated that inverse screening is *per se* unable to find targets which are not to a large extent complementary to the query ligand with respect to its protonation and tautomeric state. Moreover, the sc-PDB target database used here is not sensitive to significant conformational changes, multiple druggable binding sites or interspecies residue variability⁴⁹. Lastly, despite the facts that water is rather often essential in protein-ligand interactions, that explicit water mediated interactions are cited in the literature^{56,61-64}, and that the effect of water inclusion on docking has been widely studied⁶⁵⁻⁷⁵, considering explicit water molecules^{69,73,74,76} is almost impossible in inverse screening. As an example, crystallographic waters can be retained in GOLD⁶³, but only for one target at a time. It is difficult to predict which water molecules are important for a yet unknown protein-ligand complex.

This work particularly addresses improvements of the target scoring evaluation by using a ranking threshold and multiple selection criteria.

The quest for targets of phenazopyridine: Biological context

Small molecules are a relatively recent tool to direct cell fate⁷⁷⁻⁸⁴. They are advantageously binding to intracellular targets, entailing a profound modification of cell signalling¹. They can be seen as the

“missing link” between gene expression, protein translation and expression, allowing for rapid, reversible modulation of the individual macromolecular function⁸⁵. Indeed, they are increasingly used as tools for e.g. getting a grip on stem cell commitment. Examples include TWS119 (a pyrrolopyrimidine), which induces neuron formation from embryonic stem cells by modulation of glycogen synthase kinase 3 β ⁸⁶ and Cardiogenol C (a pyrimidine), identified from a heterocycles library, triggering cardiac muscle cell formation from embryonic stem cells⁸⁷.

The focus in this study is on phenazopyridine, a well-known local analgesic in urinary tract infections, usually used in combination with an antibiotic. It was identified in a screening for small molecules enhancing neuronal differentiation¹. The detection of such enhancers among 1040 FDA-approved drugs relied on a dual luciferase reporter system introduced by a lentivector into a murine embryonic CGR8 stem cell line. Phenazopyridine was found to selectively induce the differentiation of these stem cells into neuronal ones and a multi-target mechanism of action was suspected with respect to regulating the balance between self-renewal and differentiation of stem cells¹. Altering this balance could be responsible for major phenotypic changes. Phenazopyridine was further analyzed in human embryonic cells and was shown to: (i) enhance neuronal differentiation (more neuronal markers and neuronal cell subtypes), (ii) increase cell survival, (iii) decrease the amount of non-neuronal and undifferentiated cells, and (iv) synchronize the cellular differentiation state¹.

With respect to the influence of such a small molecule, it is important to note that neurogenesis strongly depends upon the timely expression and/or repression of transcription factors and growth factors. Each development stage can be characterized by specific markers. Moreover, a coordinated interplay between activation and inhibition of signalling pathways is crucial for proper neurogenesis to be initiated⁸⁸. Well-known proteins involved in mammalian stem cell differentiation involve metabotropic glutamate receptors (mGluRs) and kinases. Both are further detailed in the following.

mGluRs modulate glutamatergic neurotransmission and can be divided into three groups according to sequence identity, pharmacology, and signalling. Only the group I (mGluR1 and mGluR5) depending upon phospholipase C and regulating neuronal excitability⁸⁹ is further discussed. mGluR5 is a modulator of synaptic plasticity, ion channel activity, and excitotoxicity⁹⁰. The molecular basis of both group I receptors was investigated by mutational analysis and homology modelling⁹⁰⁻⁹². Apo as well as holo crystal structures of mGluR1 were resolved⁹³⁻⁹⁵. A broader overview on mGluR5 structure and pharmacology is given by Kew and Kemp⁹⁶.

With regard to stem cell differentiation, mouse embryonic stem cells in the undifferentiated state express only mGluR5, while the main transcription factors for embryonic stem cell renewal and alkaline phosphatase, a marker for the undifferentiated state, are reduced by the addition of 2-methyl-6-(phenylethynyl)-pyridine (MPEP) or antisense knockdown⁹⁷. Upon differentiation of these cells into embryoid bodies, both mGluR5 mRNA and protein are lost, while mGluR4 expression is started and enhanced in the presence of the mGluR4 agonist L-2-amino-4-phosphobutanoate (L-AP4)⁹⁸. Thus, differentiation changes the mGluR expression towards mGluR4, underlining the previous statement on the importance of mGluR for activation and inhibition in neurogenesis.

Table 36 on page 185 lists small molecules that revealed to be useful as tools for investigating the function of mGluR4 and/or mGluR5 and the findings associated with each molecule. Phenazopyridine's effect on either of these receptors was unknown at the beginning of the *in silico* study.

Summarizing, the unequal expression of the two metabotropic glutamate receptor subtypes influences cell commitment^{97,98}. Differentiation is triggered after mGluR4 expression and activation. Conversely, activation of the mGluR5 maintains stem cell renewal. Moreover, the two receptors are not triggering the same signalling pathways⁹⁹. Thus, pharmacological evidence links mGluR5 inhibition to the modulation of embryonic stem cell fate^{97,100}.

Effect of phenazopyridine on mGluR4 and mGluR5

Negative allosteric modulators of mGluR5 may affect neuronal differentiation through off-target effects such as positive allosteric modulation of mGluR4 and NMDA antagonism^{101,102}. As already stated, mGluR4 is not expressed in undifferentiated embryonic stem cells, but starts to be expressed upon

differentiation⁹⁸ and its activation enhances neuronal differentiation of mouse embryonic stem cells¹⁰⁰.

Experimentally, neither negative allosteric modulators of mGluR4 nor a positive allosteric modulator or NMDA blockers had any effect on the dual luciferase ratio by which the neuronal differentiation was assessed (personal communication of Dr. D. M. Suter). In contrast, mGluR5 inhibitors enhanced neuronal differentiation at concentrations in the micromolar range (Figure 71). Conversely, in HEK cells transfected with mGluR5, phenazopyridine did not act as an antagonist, even at a concentration of 30 μ M (Figure 68). Thus, phenazopyridine does not act via inhibition of mGluR5.

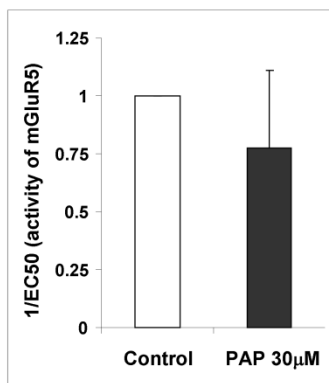


Figure 68: Phenazopyridine has no antagonist activity on mGluR5. HEK cells transiently transfected with a rat mGluR5 construct were assayed for a quisqualate-induced rise of intracellular Ca^{2+} concentrations, with or without phenazopyridine (PAP) treatment. Values are shown as the inverse of the EC_{50} . The Figure was kindly provided by Dr. D. M. Suter.

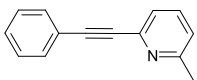
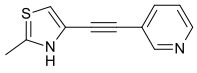
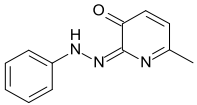
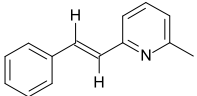
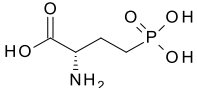
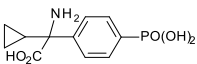
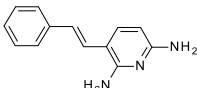
4.1 METHODS

The *in silico* methods of this chapter can be divided into two parts: (i) SARs for close phenazopyridine analogues and compounds acting on the metabotropic glutamate receptor 5 (mGluR5) and (ii) the inverse screening procedure. The SARs were conducted prior to the experimental evidence that phenazopyridine had no antagonist activity on mGluR5 (Figure 68).

To identify the parts of phenazopyridine that are essential for triggering neuronal differentiation of embryonic stem cells, a panel of commercially available analogues and several negative allosteric modulators of mGluR5 with modifications in the aza moiety were selected (Y. Westermaier and Dr. D. M. Suter) and their effect on neuronal differentiation was analyzed via a dual luciferase assay on CGR8 cells (Dr. D. M. Suter). Detailed SARs are given for phenazopyridine analogues and compounds acting on mGluR5 in the results subchapter 4.2.1.

For inverse screening purposes, the sc-PDB⁸ was used as a target and cognate ligand database. In its 2008 version, it contained more than 5000 redundant 3D binding sites extracted from the PDB. Active sites and ligands were downloaded from <http://bioinfo-pharma.u-strasbg.fr/scPDB/> in mol2 format, split, and further targets were added based on the biological context or the controls required. Some proteins fitting into the biological context, but absent from the sc-PDB because of a resolution worse than 2.5 Å, were manually added: the human tau protein kinase 1, the glycogen synthase kinase-3 β , the mitogen-activated protein (MAP) kinase kinase 1, the p21 activated kinase 4, MNK2, AKT2 (phenazopyridine can be superimposed to some extent with the cognate ligands), and some PIM1 structures. The downloaded active sites devoid of water, cofactors, and ligands, corresponding to the residues 6.5 Å around the main co-crystallized ligand, and cognate ligand files were already correctly protonated. The manually added proteins and their cognate ligands were prepared similarly. Additional ligands included phenazopyridine and PD98059, a MEK1 kinase inhibitor with a distinct kinase binding motif (Figure 69). These ligands were prepared accordingly. Based on the SAR analysis, an activity screening on a kinase panel was performed at the University of Dundee and revealed that phenazopyridine lowered MEK1 activity by 25% (data kindly provided by J. Bain).

Table 36: Small molecules as tools for investigating on the function of metabotropic glutamate receptors 4 (mGluR4) and 5 (mGluR5). NMDA stands for N-methyl D-aspartate.

(Ant)agonist	Short name	Structure	Small molecule effects
Agonist, allosteric regulator of mGluR5	MPEP: 2-methyl-6-(phenylethynyl)-pyridine		- Activates mGluR5, which is required for long-term potentiation ¹⁰³ and critical for aversive and spatial learning ¹⁰³. - NMDA receptor antagonism results in decreased neuronal toxicity ¹⁰¹ and reduced traumatic neural injury ¹⁰⁴. Its neuroprotective effect is probably due to the inhibition of the NMDA receptor, and not the mGluR5 ¹⁰⁵.
Antagonists, negative allosteric regulators of mGluR5	MTEP: 3-[(2-methyl-1,3-thiazol-4-yl)-ethynyl]pyridine		Its neuroprotective effect is probably due to the inhibition of the NMDA receptor, and not the mGluR5 ¹⁰⁵ .
	SIB-1757: (2E)-6-methyl-2-(phenylhydrazono)pyridin-3-one		Human mGluR5 antagonist; IC ₅₀ of 0.37 μM ¹⁰⁶
	SIB-1893: (E)-2-methyl-6-(2-phenylethenyl)-pyridine		NMDA receptor antagonism results in decreased neuronal toxicity ¹⁰¹ and reduced traumatic neural injury ¹⁰⁴ . IC ₅₀ of 0.29 μM ¹⁰⁶
Full agonist of mGluR4	L-AP4: L-2-amino-4-phosphobutanoate		mGluR4 expression is started and enhanced in the presence of L-AP4 ⁹⁸ .
Antagonist of mGluR4	CPPG: 2-amino-2-cyclopropyl-2-(p-tolyl) acetic acid		Competitively blocks the action of SIB-1893 and MPEP on mGluR4 ¹⁰²
Unknown effect	Phenazopyridine		Structure similar to MPEP

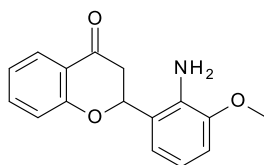


Figure 69: Structure of the MEK1 kinase inhibitor PD98059.

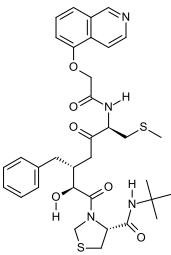
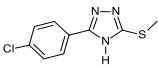
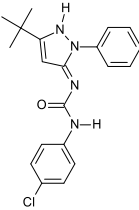
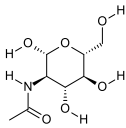
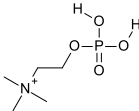
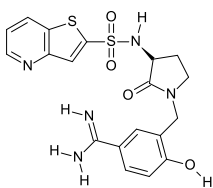
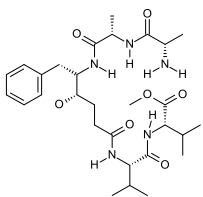
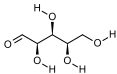
Prior to docking with GOLD 4.0 (Cambridge Crystallographic Data Centre) on a 32-bit processor, configuration files were generated automatically by in-house c-shell scripts, allowing for the flipping of planar trigonal nitrogens between cis and trans in the ligand and early termination if the three top poses of the cognate ligand were within 1.5 Å from the crystallized binding mode.

Two different validations were carried out: (i) an unbiased one, where 12 different crystallized ligands, 12 corresponding targets, and 188 decoy proteins were randomly chosen out of the sc-PDB database, and (ii) a knowledge-driven one, where specific, less specific, and promiscuous ligands already used in other studies^{8,49} were considered. For (i) and (ii), it was investigated whether the crystal structure could be reproduced within a RMSD of 2 Å and whether the native ligand poses were first or high-ranked.

The ligands considered in (i) included: a human estrogen receptor α ligand (AIH, PDB ID: 1XP1); the cofactor flavin mononucleotide (FMN, PDB ID: 1BKJ); α -D-glucose 1,6-bisphosphate (G16, PDB ID: 1O08); a HIV-1 protease inhibitor (KNI, PDB ID: 1HPX); a thrombin inhibitor (L03, PDB ID: 1WBG); a P38 α MAP kinase inhibitor (L10, PDB ID: 1W82); N-acetyl-D-glucosamine (NAG, PDB ID: 1M3Y); a thyroid hormone receptor β thyromimetic (PFA, PDB ID: 1N46); phosphocholine (POC, PDB ID: 1H8P); a factor Xa inhibitor (PR1, PDB ID: 1F0T); a HIV-1 protease inhibitor (PSI, PDB ID: 1AAQ); and the linear form of D-xylose (XLS, PDB ID: 9XIM) (Table 37).

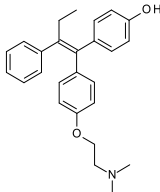
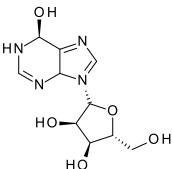
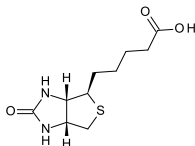
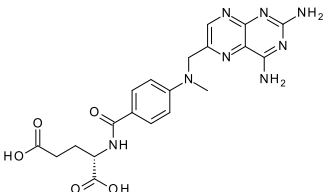
Table 37: Ligands considered for the unbiased validation (continued on the next page).

Acronym	Ligand IUPAC or common name	PDB ID	Ligand structure
AIH	(2S,3R)-2-[4-[2-[(3R,4R)-3,4-dimethylpyrrolidin-1-yl]ethoxy]phenyl]-3-(4-hydroxyphenyl)-2,3-dihydro-1,4-benzoxathiin-6-ol	1XP1	
FMN	Flavin mononucleotide	1BKJ	
G16	Alpha-D-glucose 1,6-bisphosphate	1O08	

KNI	KNI-272	1HPX	
L03	3-(4-chlorophenyl)-5-(methylsulfanyl)-4H-1,2,4-triazole	1WBG	
L10	1-[(3E)-5-tert-butyl-2-phenyl-1,2-dihydro-3H-pyrazol-3-ylidene]-3-(4-chlorophenyl)urea	1W82	
NAG	N-acetyl-D-glucosamine	1M3Y	
POC	Phosphocholine	1H8P	
PR1	4-hydroxy-3-((3S)-2-oxo-3-[(thieno[3,2-b]pyridin-2-ylsulfonyl)amino]pyrrolidin-1-yl)methyl)benzenecarboximidamide	1F0T	
PSI	(2S,5S,8R,10S,11S,14R,17R)-8,17-diamino-11-benzyl-10-hydroxy-2-[(1S)-3-hydroxy-1-methylpropyl]-14-methyl-5-(1-methylethyl)-4,7,13,16-tetraoxo-3,6,12,15-tetraazaoctadecan-1-oic acid	1AAQ	
XLS	Linear form of D-xylose	9XIM	

Because the rate of targets to identify depends on the specificity and promiscuity of protein-ligand interactions and on the effectiveness of the scoring function, four compounds served as controls of the knowledge-driven validation (ii). These compounds have already been used in other studies ^{8,49}: OHT (4-OH tamoxifen), PRH (6-OH-1,6-dihydropurine nucleoside), BTN (biotin), and MXT (methotrexate) (Table 38).

Table 38: The four control compound names, abbreviations, structures, and specificities (from highly specific (+++++) to promiscuous (+)).

Name	4-OH tamoxifen	6-OH-1,6-dihydropurine nucleoside	Biotin	Methotrexate
Abbreviation	OHT	PRH	BTN	MTX
Structure				
Specificity	+++++	+++	++	+

Further ligands used included a series of Herpes Simplex Virus 1 Thymidine Kinase (HSV-1 TK) substrates: the pyrimidines deoxythymidine (THM), (North)-methanocarbothymidine (TMC), 6-hydroxypropylthymidine (HPT), 6-[4-(hydroxymethyl)-5-methyl-2,6-dioxohexahydropyrimidin-5-yl]methyl-5-methylpyrimidine-2,4(1H,3H)-dione (RCA), 6-[3-hydroxy-2-(hydroxymethyl)propyl]-5-methyl-2,4(1H,3H)-pyrimidinedione (CCV), 5-bromothienyldeoxyuridine (BTD), 5-iodouracil anhydrohexitol nucleoside (AHU), 5-iododeoxyuridine (ID2), 5-bromovinyldeoxyuridine (BVD), (South)-methanocarbothymidine (SCT) (Table 39a), and the purines 9-(4-hydroxybutyl)-N2-phenylguanine (BPG), ganciclovir (GA2), penciclovir (PE2), and aciclovir (AC2) (Table 39b).

Table 39a: Pyrimidine substrates of the HSV-1 TK used for the knowledge-driven validation

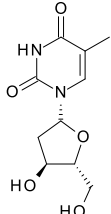
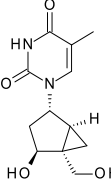
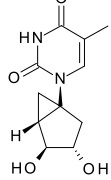
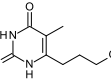
Acronym	Ligand IUPAC or common name	PDB ID	Ligand structure
THM	Deoxythymidine	1P7C	
TMC	(North)-methanocarbothymidine	1E2K	
SCT	(South)-methanocarbothymidine	1OF1	
HPT	6-hydroxypropylthymidine	1E2M	

Table 39a (continued): Pyrimidine substrates of the HSV-1 TK used for the knowledge-driven validation

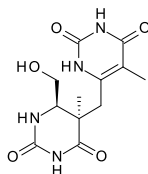
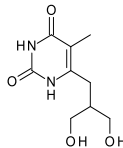
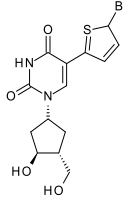
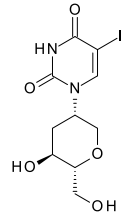
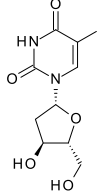
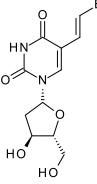
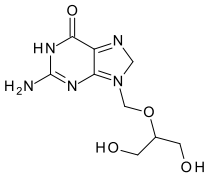
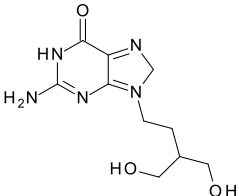
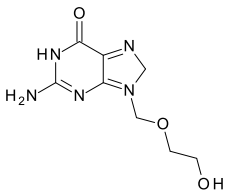
Acronym	Ligand IUPAC or common name	PDB ID	Ligand structure
RCA	6-ff-5-methylpyrimidine-2,4(1H,3H)-dione	1E2N	
CCV	6-[3-hydroxy-2-(hydroxymethyl)propyl]-5-methyl-2,4(1H,3H)-pyrimidinedione	1E2P	
BTD	5-bromothiényldeoxyuridine	1K14	
AHU	5-iodouracil anhydrohexitol nucleoside	1K16	
ID2	5-iododeoxyuridine	1K17	
BVD	5-bromovinyldeoxyuridine	1K18	

Table 39b: Purine substrates of the HSV-1 TK used for the knowledge-driven validation

Acronym	Ligand IUPAC or common name	PDB ID	Ligand structure
GA2	Ganciclovir	1KI2	
PE2	Penciclovir	1KI3	
AC2	Aciclovir	2KI5	

The scoring function used for evaluating the protein-ligand interactions is called Goldscore, and sometimes also fitness function, in accordance with the underlying genetic algorithm. It is composed of four terms including the protein-ligand hydrogen bond energy, the protein-ligand VdW energy, the ligand's internal VdW energy, and the ligand's torsional strain energy. All 5279 proteins were ranked according to the phenazopyridine-based fitness score. Their PDB headers were then searched online in the PDB via a customized Excel macro for keywords related to neurodegenerative diseases, neurotransmitter functions or the like. Additional information on the enzyme classification (EC) number, the three-letter ligand acronym, the source organism, and the PDB keywords were extracted as well. For each of these hits, the heavy atom RMSD between the best-scored pose for the cognate ligand and its crystal pose were then calculated using the smartrms utility of GOLD 4.0 and the two poses were visually compared using Sybyl 8.0. Moreover, the best pose of phenazopyridine was overlaid with the crystal pose of the crystallized ligand and the matching of the donor/acceptor functions involved in H-bonding patterns was checked.

Experimental validation of kinases

The kinases identified by inverse screening were tested with several wet-lab experiments to confirm the selectivity profile of putative targets. The methods for determining kinase inhibitor profiles^{107,108} are well established. A kinase activity screening containing mainly Ser/Thr kinases was performed at the University of Dundee to confirm the hypothesis that kinases are targets. This screening was done according to a previously described procedure^{107,108}. Furthermore, a kinome scan was performed by Ambit Biosciences (<http://www.ambitbio.com/>). The IC₅₀ value of phenazopyridine was determined on Abl at the University of Milano-Bicocca as detailed in Gunby et al.¹⁰⁹.

4.2 RESULTS AND DISCUSSION

4.2.1 STRUCTURE-ACTIVITY RELATIONSHIPS

The effects of phenazopyridine and its analogues on CGR8 cells are summarized in terms of a structure-activity relationship (SAR) and extended to negative allosteric modulators of the mGluR5.

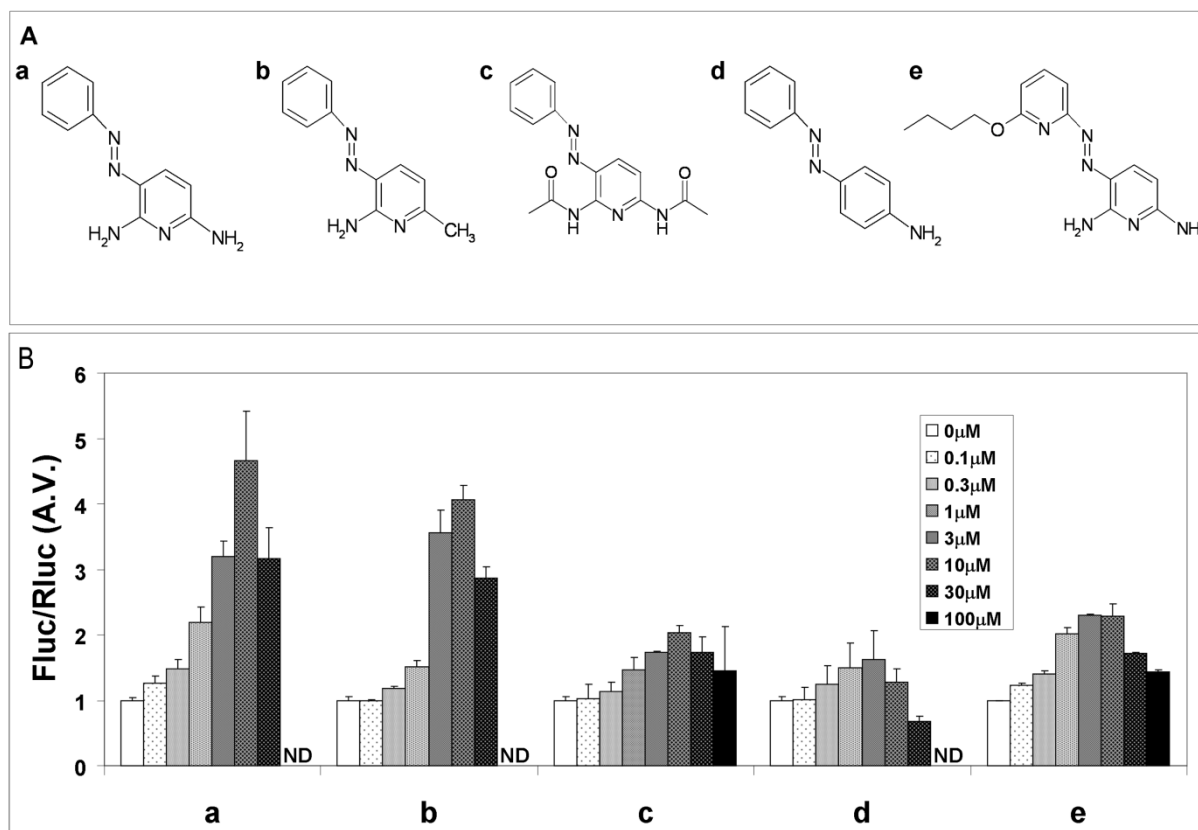


Figure 70: The pyridine-2,6-diamine ring is essential for activity. A: Chemical structure of phenazopyridine (a) and related molecules (b-e). B: Activity of phenazopyridine and related molecules on the Firefly luciferase/Renilla luciferase ratio (Fluc/Rluc) of the CGR8 dual luciferase cell line. A.V.: Arbitrary values. ND: Not determined. The Figure was kindly provided by Dr. D. M. Suter.

Compared to phenazopyridine (compound a in Figure 70), the activity of the molecule was not substantially altered by introducing a methyl group at position 6 of the pyridine ring (compound b in Figure 70). In contrast, replacing both the 2- and 6-amino by acetamide (compound c in Figure 70) led to a substantial decrease in activity. The substitution by acetamide maintains the H-bond donor but adds an acidic character and is bulkier compared to a primary amino group. A loss of activity was also observed when the pyridine nitrogen, a H-bond acceptor, was substituted by an aromatic carbon and one amine on the ring was removed by changing the pyridine-2,6-diamine substituent to aniline (compound d in Figure 70). In order to investigate the pocket size occupied by phenazopyridine in its target, an analogue displaying a butoxy group attached to the 3-phenyl ring was tested (compound e in Figure 70). As a decreased activity on neuronal differentiation was obtained for this molecule, it can be concluded that the pocket accommodating phenazopyridine is quite restricted. In summary, based on the results obtained for the analogues, it is hypothesized that in the case of phenazopyridine itself, the 2,6-diaminopyridine moiety is interacting with its target via H-bonds, in particular via the pyridine nitrogen, while the 3-phenyl group is taking part in hydrophobic interactions. Both types of interactions could define the specific binding of phenazopyridine to its target.

To exclude that the differences in activity may be due to differences in cell penetration, physicochemical characteristics were analysed using the molecular modelling suite Sybyl 7.3. It was

found that the compounds are similar in size and have very similar clogP values, ranging from 1.8 to 3.6. Such a narrow range of clogP values suggests both very similar hydrophobic characters and propensities to enter cells. Despite the fact that the polar surface area or the flexibility is not considered for this cell penetration assessment, it can be assumed that the observed activity differences are not due to differences related to cell entry.

Several negative allosteric modulators of mGluR5 with modifications in the aza moiety enhance neuronal differentiation similarly to phenazopyridine (Figure 71).

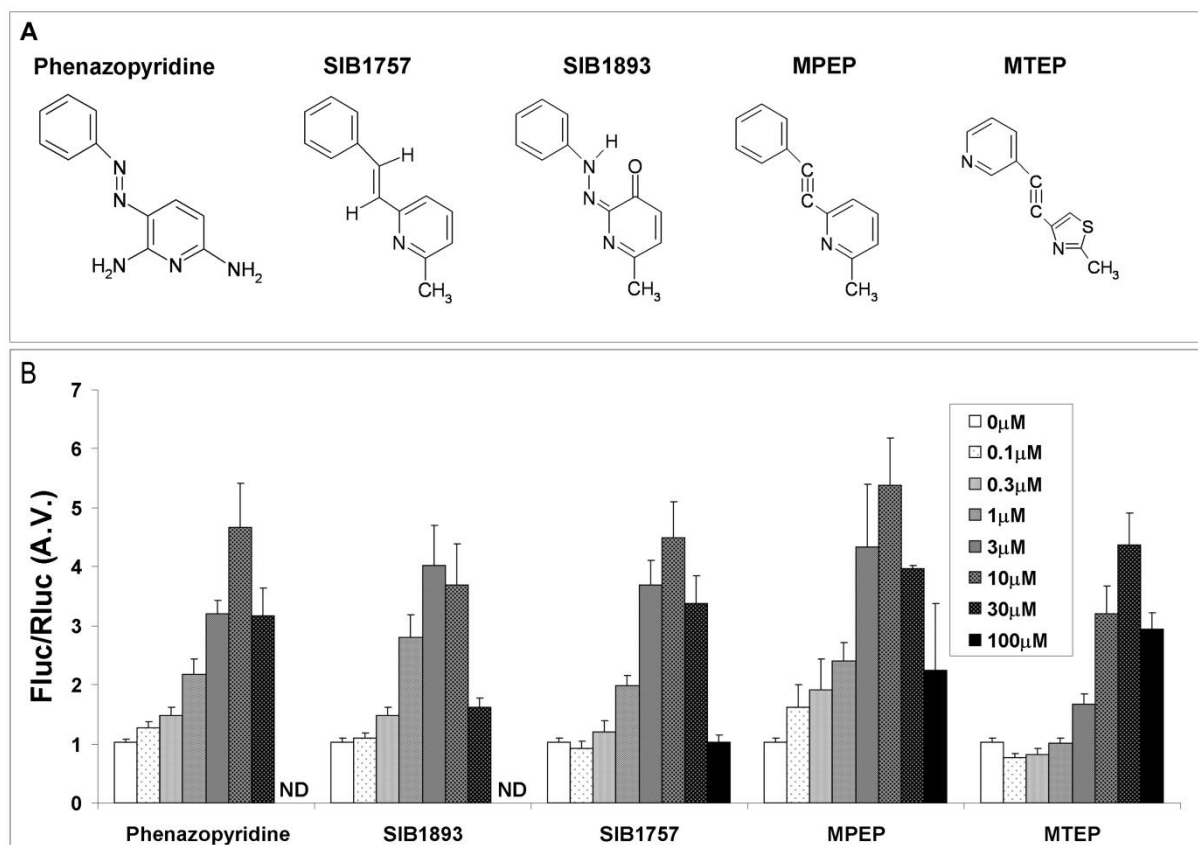


Figure 71: Negative allosteric modulators of mGluR5 enhance neuronal differentiation of mouse embryonic stem cells. **A:** Chemical structure of phenazopyridine and negative allosteric modulators of mGluR5. **B:** Activity of phenazopyridine and negative allosteric modulators of mGluR5 on the Firefly luciferase/Renilla luciferase ratio (Fluc/Rluc) of the CGR8 dual luciferase cell line. A.V.: Arbitrary values. ND: Not determined. The Figure was kindly provided by Dr. D. M. Suter.

Given the results depicted in Figure 71, the statements made based on Figure 70 must be reformulated. Provided that the allosteric mGluR5 modulators do not have a kinase pattern as the pyridine-2,6-diamine of phenazopyridine, the SAR results, the luciferase measurements (Figure 71) and the data on HEK cells (Figure 68) all show that phenazopyridine does not act via mGluR5 allosteric modulation.

4.2.2 RESULTS OF THE UNBIASED VALIDATION

The 12 protein-ligand complexes chosen randomly for the unbiased validation were representative, as they ranged from highly complementary protein-ligand interactions (L10) to narrow binding pockets with highly flexible ligands (PSI), covering the whole range of hydrophilic to hydrophobic binding pockets. Considering the RMSDs for the lowest energies, i.e. the most stable docking poses of the 12 ligands (Table 37), the crystal pose was reproduced with a RMSD ≤ 2 Å in 7 out of 12 cases. In the

remaining cases, the crystal pose was not found because of a narrow binding site and/or highly flexible ligands. With respect to the energy score-based ranking, 42% of the native ligand poses were first-ranked with an outstanding shape complementarity between the ligand and the surrounding protein. Considering the overall ranking, 75% of the docking poses of each cognate ligand ranked in the top 3.5% based on their energy score (Figure 72). The remaining 25% were either characterized by binding sites with composite areas having a narrow connection between them, being very compact or deeply buried.

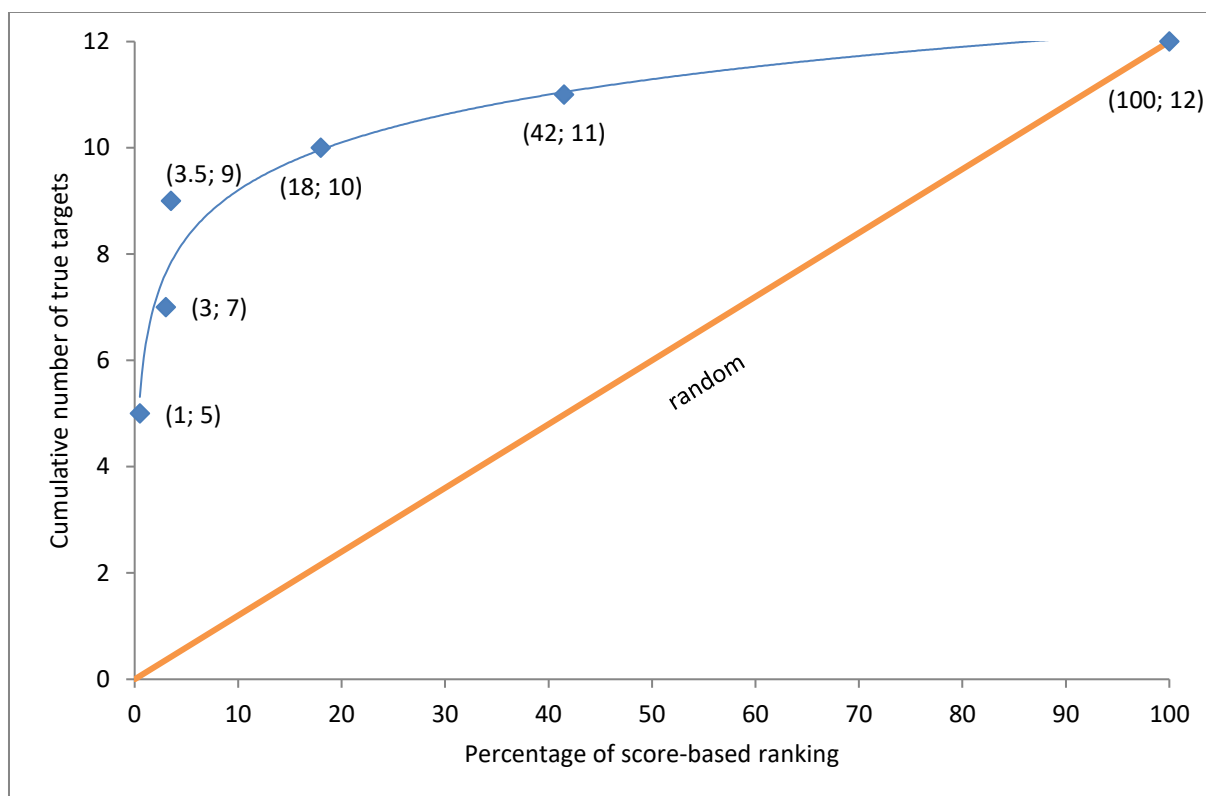


Figure 72: The majority of the cognate docking poses (9 out of 12, i.e. 75%) are found within the first 3.5% of the ranking. The graph depicts the cumulative number of true targets against the percentage of score-based ranking. The blue diamonds represent the cognate docking poses and the orange line the results that would have been obtained for random protein-ligand docking.

In summary, a good docking accuracy could be achieved for most of the ligands and with respect to the ranking, the present protocol allows to enrich the target library to be tested experimentally by a factor of 21 (considering the first 3.5% of the ranking).

4.2.3 RESULTS OF THE KNOWLEDGE-DRIVEN VALIDATION

The set of specific and promiscuous ligands already used by Paul et al.⁸ is examined first, with the specificity decreasing in the order OHT, PRH, BTN, and MTX (Table 38). As mentioned earlier, the sc-PDB target database contains one or more holo structures of a target. According to Carlson¹¹⁰, this might account for some protein flexibility.

Estrogen receptors are considered easy docking targets, because their active sites are well shielded from solvent and do not rearrange dramatically upon binding different steroid hormones^{37,111,112}. In the OHT docking, the three estrogen receptors of the sc-PDB were ranked first, second, and seventh and corresponding RMSDs of 1.3, 0.6, and 1.1 Å were obtained, respectively. True OHT targets were thus ranked in the first 0.1%, while Li et al. found 30% of the true OHT targets in the first 2% of the ranking using TarFisDock¹¹. In contrast, according to this publication, glutathione transferase ranked in the top 2% and was confirmed experimentally to be a target, while here it is found in the top 7.2%. This is still an acceptable result, given that the two docking algorithms are considerably different. Owing to the

fact that docking of ligands into estrogen receptors has been thoroughly studied in the past ^{8,37,49}, and does not represent a challenge for docking anymore, it is not pursued here.

Docking of PRH into his target adenosine deaminase yielded the ranks 1 and 18 (with RMSDs of 0.2 and 0.4 Å, respectively). The biotin-binding proteins in the BTN docking were ranked 1st, 2nd, 5th, 7th, and 15th (with RMSDs of 0.4, 0.4, 0.3, 0.4, and 0.7 Å, respectively). Unsurprisingly, the protein crystal structures containing the unspecific ligand MTX were found in a broad range of ranks: 3rd, 34th, 54th, 126th, and 1101st (with RMSDs of 1.1, 2.0, 0.4, 1.8, and 8.6 Å, respectively). These crystal structures represent four oxidoreductases, e.g. dihydrofolate reductase or pteridine reductase, and a methyltransferase.

Considering the ranks and RMSDs, the above-mentioned specificity decrease (Table 38) could be reliably reproduced with this docking protocol. Furthermore, it helped to establish a ranking threshold of 10% to distinguish rather specific from promiscuous ligands.

The HSV-1 TK was further investigated. It is known to be a notoriously difficult docking target, as it harbours a water-accessible binding site, whose residues can undergo induced fit movements (Gln125 for purines; Tyr132 and Arg222 for pyrimidines). Also, the number of water-directed H-bonds depends upon the purine- or pyrimidine-substrate bound, which mostly have an affinity in the micromolar range. A comparison of DOCK, FlexX, and GOLD showed that the latter provided the best docking accuracy ³⁷ for substrates of this target.

Similarly to ³⁷, results were analyzed in terms of (i) docking accuracy (RMSD), (ii) scoring accuracy (prediction of binding free energy), (iii) discrimination of active from random compounds as well as (iv) hit rates and enrichment factors among the top scorers. Generally, to obtain good enrichments, a docking program needs to produce reliable binding modes ⁵⁸, which can be enhanced by considering flexible side chains ¹¹³ instead of rigid proteins. Given multiple HSV-1 TK holo structures with a variety of purine- and pyrimidine-based substrates bound are considered in this work, some side flexibility is already taken into account.

i) Docking accuracy: The docking accuracy in terms of the RMSD of 14 substrates (10 pyrimidine- and 4 purine-based ones, Table 39a and b) obtained upon docking into their corresponding crystal structure ranged from 0.5 (THM) to 0.85 Å (RCA) for pyrimidines and from 0.48 (BPG) to 1.60 Å (GA2) for purines. Comparing these results with the ones of Bissantz et al. ³⁷, where some of the same compounds were docked into a reference crystal structure (PDB ID: 1KIM), the following statements can be made:

- 6 substrates out of 14 (pyrimidines: THM, TMC, and CCV; purines: GA2, PE2, and AC2) perform better when docked into their crystal structure than into the reference structure used in ³⁷, which is due to the omission of water molecules mediating protein-ligand interactions in the latter. With this respect, it is important to note that the success of crossdocking critically depends upon structural similarity between the protein active sites, also in terms of structural waters.
- Three pyrimidines (HPT, AHU and ID2) docked equally well (0.1-0.3 Å deviation in RMSD).
- The remaining five structures could not be compared with the mentioned study, as they were not yet crystallized at the time that study was published.
- The purines GA2, PE2, and AC2 display the highest overall RMSD (1.1-1.6 Å), reflecting the high flexibility of the substitution at position 9 of guanine.

ii) Scoring accuracy: Considering the scoring accuracy, it was noted in a comparative study that the combination GOLD/Goldscore yielded the best docking/scoring scheme ³⁷. Comparing the RMSDs with the ranks, the larger purines are comparatively poorer docked but better ranked (except for the purine BPG), while pyrimidines are better docked but poorer ranked. In this light, it is not astonishing that no correlation between the scores and the experimentally determined affinity was found for the 14 substrates.

iii) Discrimination of active from random compounds: Generally, scoring functions perform well in recovering the crystal binding pose but poor when it comes to distinguishing actives from inactives, leading to many false positive high-scorers resulting in lower enrichment rates for virtual screening ³⁸. Figure 73 depicts the score-based ranking for deoxythymidine (dT) with the crystal structures crystallized with dT (true targets), other pyrimidines or purines. True targets of dT all cluster in the first

10% of the ranking. The percentages of coverage of known dT targets, i.e. the ratios in percentage between the number of true targets recovered by inverse screening at a defined top scoring fraction and the total number of true target entries in the sc-PDB, are the following: 100% for the top 10% of the ranking, 50% for the top 2% and 16.7% for the top 1%, which is far superior to random screening. With respect to other pyrimidine and purine targets, the clear ranking outlier corresponds to HSV-1 TK crystallized with aciclovir, while the second worst is HSV-1 TK in complex with ganciclovir, both being purines. This is in accordance with Kellenberger et al.⁵⁵, who also found that GOLD provides better ranks for pyrimidines than for purines.

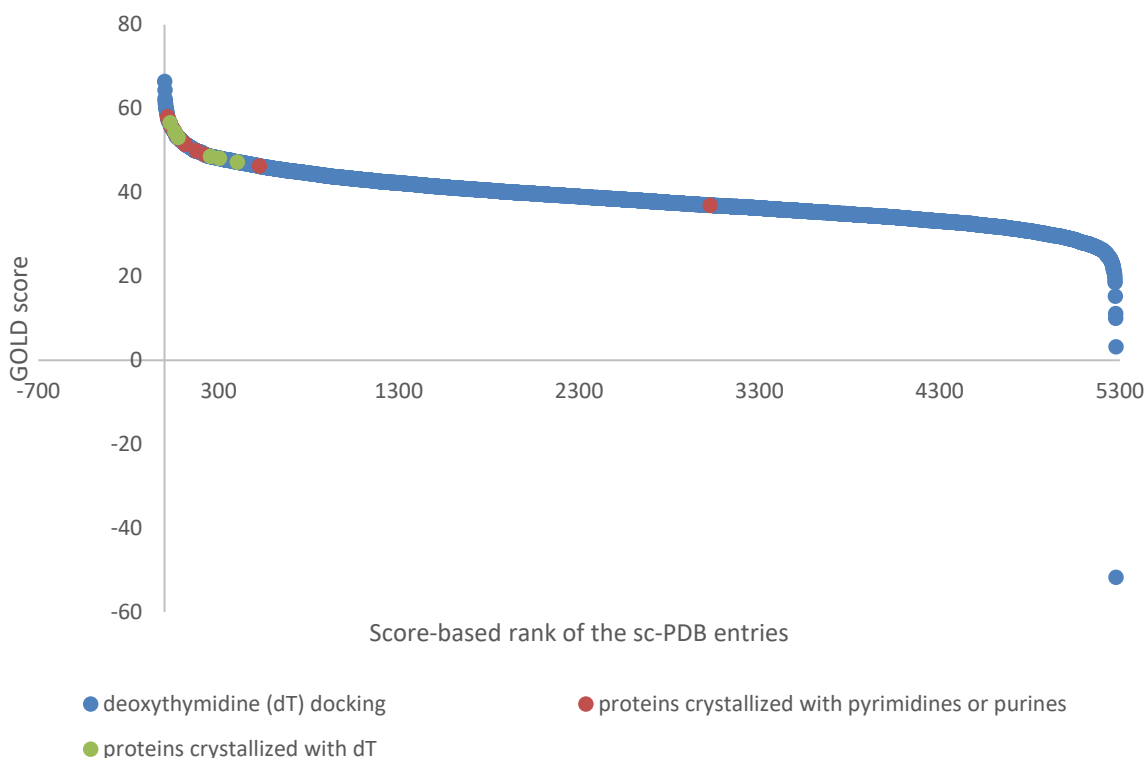


Figure 73: Deoxythymidine (dT) docking into the target database: The score-based ranking of dT with respect to all proteins is indicated in blue. For comparison, crystal structures being crystallized with either pyrimidines, purines (brown) or dT (green) are indicated.

iv) Hit rate and enrichment factor among the top scorers: Considering the hit rate over all pyrimidines and purines, the percentage of true positives (TPs) in the top 3.5% of the ranking attained 70%. According to Pearlman and Charifson¹¹⁴, the enrichment factor is given by:

$$EF = \frac{Hits_{sampled} / N_{sampled}}{Hits_{total} / N_{total}} \text{ and corresponds here (considering the top 3.5\% of the ranking) to}$$

$$EF = \frac{14/184}{20/5279} = 20.$$

This is comparable to the library focusing for the unbiased validation, where a factor of 21 was obtained.

Summarizing the knowledge-driven validation, the following statements can be made about the score-based ranking:

- (i) All best solutions display remarkably high fitness values.
- (ii) For specific ligands, true targets are quite sharply separated from other proteins by a scoring gap.
- (iii) The targets of MTX, a promiscuous ligand, are broadly spread over the score-based ranking.

The control docking of the cognate ligand revealed systematic docking limitations in line with Kellenberger et al.⁵⁵, particularly insufficient conformational sampling for highly flexible ligands and a high ratio of H-bond donors/acceptors to carbons for D-xylose isomerase.

4.2.4 ESTABLISHING CRITERIA FOR TARGET PRIORITISATION

To be able to choose potential targets for phenazopyridine, different target selection criteria were defined. Perhaps the most classical definition of docking success is the RMSD < 2 Å criterion. Docking poses not meeting this criterion are commonly termed docking failures^{115,116}, which originate either from incomplete sampling or inaccuracies in the scoring function¹¹⁷. However, the RMSD is flawed as a success measure: (i) Poses with good RMSDs can form different interactions than the ones seen in the experimental X-ray or NMR structure; (ii) the RMSD might be poor despite a correct alignment of essential features in the binding site (almost symmetrical molecules); and (iii) a terminal functional group of a ligand could be completely solvent-exposed¹¹⁸. An alternative is to calculate the relative displacement error¹¹⁹, where the influence of large RMSDs is down-weighted¹¹⁸.

Considering scoring, it should be kept in mind that for a small molecule¹²⁰, scoring functions do not only generate relevant target-ligand associations but yield a continuum of binding energies, out of which putative targets should be isolated. Selectivity in such a context means that the energy gap between the lowest docked structure and the lowest energy misdocked structure is high¹¹⁶. It can be achieved with a good scoring function limiting the number of both false positives (FPs) and false negatives (FNs)¹²¹. FPs may occur because uncertainties in the crystal structure with respect to protein side chains distort scores in favour of a secondary or transitional binding mode that is slightly higher in energy. Consensus scoring is a means to improve the ratio of valid hits to FPs¹²².

Given the difficulties in scoring, an empirical cut-off based on the knowledge-based validation can be introduced as a first solution. Here, a cut-off of 10% is applied.

Furthermore, when comparing *in silico* with *in vitro/in vivo* experimental results, four degrees of matching have to be taken into account: TP (true positive, full match), FP (false positive, putative *in silico* target with no experimental confirmation), FN (false negative, experimentally confirmed target not detected *in silico*), and TN (true negative, neither an *in silico* nor an *in vitro/in vivo* target). Dealing with TPs and TNs is the easiest. FPs can partially be explained if the experimental test on the putative *in silico* target could not be done, if ligand aggregation happened or if induced fit mechanisms render the binding of an investigated ligand particularly difficult. Reasons for FN are related to the difficulties to crystallize very flexible or membrane proteins and therefore to the lack of this 3D information in the target database. The remaining FP and FN are a matter of concern for the reliability of both *in silico* and wet-lab experiments.

Reasons why binders are poorly ranked include (i) correct docking but suboptimal ligand-protein contacts, (ii) the omission of water molecules and ions, (iii) ionization state uncertainties, and (iv) insufficient sampling or (v) most importantly, failures/weaknesses¹²³ of the scoring function. True scoring failures include cases for which docking predictions within 2 Å from the X-ray pose exist but which are never top ranked.

There are two ways of assessing whether true actives are effectively prioritized by inverse virtual screening: (i) enrichment based on the GOLD fitness function or (ii) receiver-operating characteristic (ROC) scores. The latter performs better than the former if a property is sufficiently sampled in a dataset⁴⁹. With ROC curves, originating from radar and sonar detection during Second World War, one can explore the relationships between sensitivity and specificity, with sensitivity being the TP rate and specificity being the TN rate and both adopting values between 0 and 1 (ideal case)¹²⁴. The area under the curve of ROC plots¹²⁵ derived from scoring list can be used to depict sensitivity and

specificity. In the following, only (i) is taken into account, because the experimental data at hand are considered too scarce for (ii).

Enrichment assessments for reverse docking still lag behind the ones for traditional forward virtual screening (e.g. ⁵⁸), but one study by Kellenberger et al. ²⁴ discusses the methods and problems related to target ranking. Enrichment enhancement by supervised scoring, e.g. applied to optimizing scoring functions against a given target protein ¹²⁶, cannot be applied to inverse screening. Kontoyanni et al. ⁵³ highlighted for classical virtual screening that the number of FPs can be lowered with conformers close to bioactive ones. Phenazopyridine is a quite rigid molecule, which was prepared as a conformer with a minimized energy.

Target prioritization has to happen on a multi-criteria basis, allowing potential links to be established regardless of the width of the biological context. The target selection presented here was based on an extension of the criteria defined by Muller et al. ⁹ and Bissantz et al. ³⁷. A target was selected if four of the following criteria were fulfilled, with criterion (iii) being mandatory:

- (i) 50% of the sc-PDB entries for a particular target were scored among the top 2% scoring entries,
- (ii) the average fitness score for all entries of the corresponding target was above 50,
- (iii) the keyword search in the PDB and the literature search revealed a link with neurogenesis, and
- (iv) the docking accuracy, i.e. the RMSD to the cognate ligand was preferably below 2 Å.

With respect to (i), Bursulaya et al. ⁵⁴ defined that for being specific in comparing library screens, a percentage of the scanned database should be used that contains approximately half of the original ligands.

4.2.5 TARGET PREDICTIONS

The most frequently high-scored potential targets include kinases, acetylcholine esterase and two oxidoreductases. Kinases are further pursued as main targets of interest because the kinase motif of phenazopyridine is already known from the SAR study. The results for acetylcholine esterase are presented briefly following the kinases. The involvement of oxidoreductases in the biological context treated here is unknown. Therefore, these target hits were not further pursued.

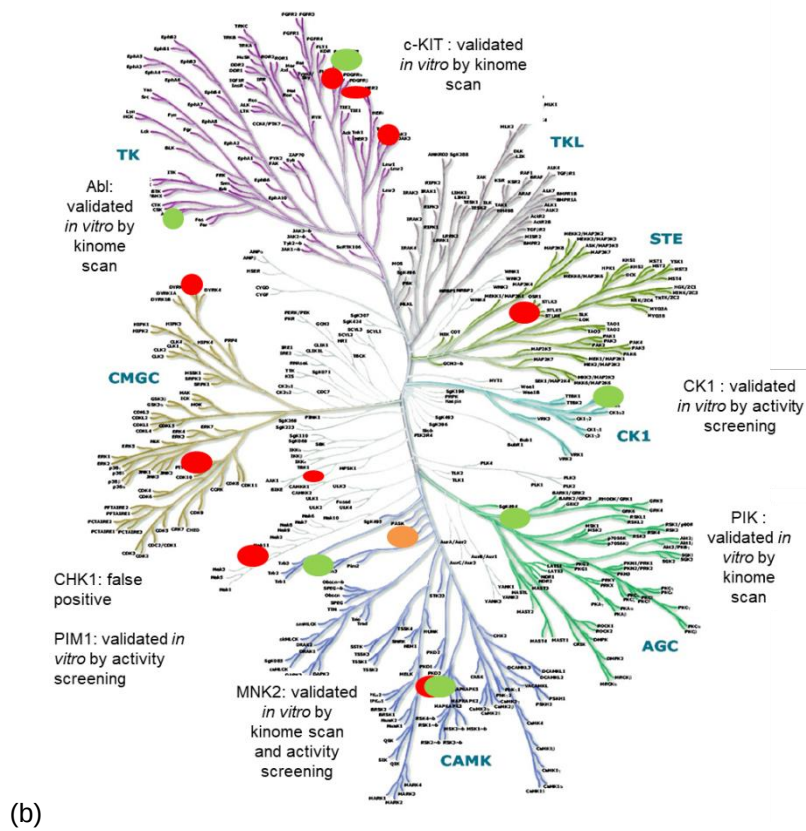
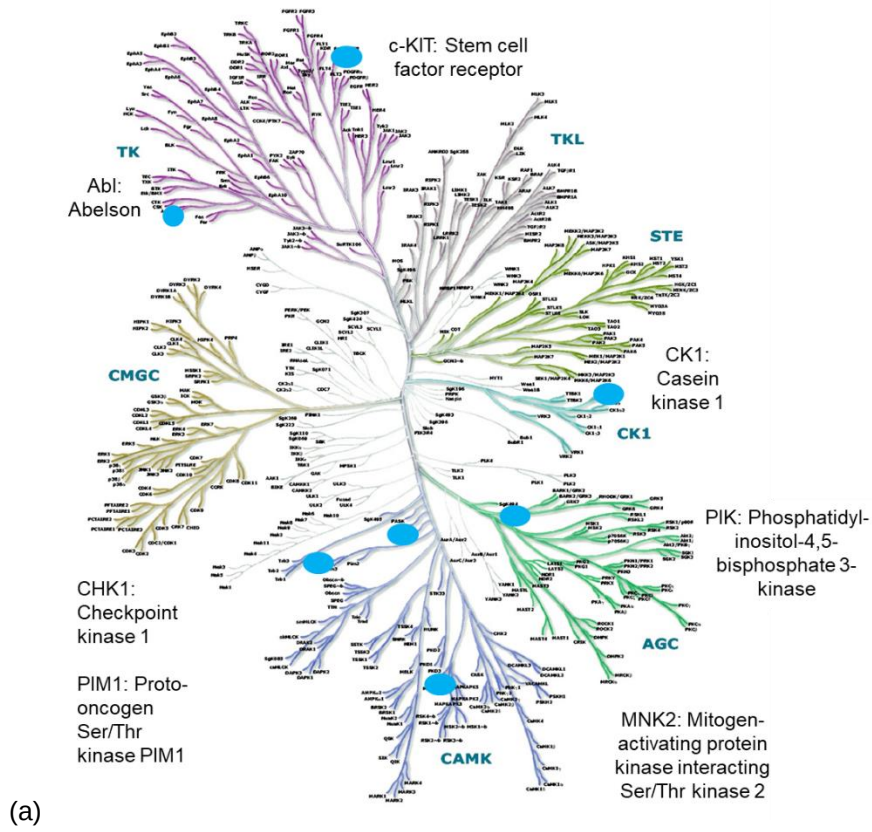
Hereafter, the kinases identified by inverse screening are presented together with their experimental validation. Table 40 depicts the kinase targets included in the inverse screening that have been experimentally tested by two different activity tests and a kinome scan. The lower the experimental percentage of activity is, the higher is the inhibitory effect of phenazopyridine.

In the following, each kinase present in Table 40 and acetylcholinesterase is discussed separately. Their potential as targets of phenazopyridine is also assessed. Unsurprisingly, phenazopyridine adopts conformations that resemble the cognate ligands in that the same plane is adopted for the ring systems.

Figure 74 depicts the kinases of Table 40 in the human kinome phylogenetic tree ¹²⁷, which comprises 518 kinases organized in seven main groups: (A) shows the predicted targets and (B) the results of the experimental validation. These kinases are located in clearly distinct areas of the tree, which possibly means that phenazopyridine acts in a selective manner.

Table 40: Names and acronyms of kinase targets included in the inverse screening and experimentally tested. "Rank (%)" means that a potential target is found in the top N% of the phenazopyridine score-based ranking. wt stands for wild type protein and n/d for not determined.

Name (acronym)	PDB ID	Rank (%)	RMSD of cognate ligand [Å]	Kinome scan (% of control at 10 µM)	Dundee activity test (% of remaining activity at 10 µM)	Milano activity test
Proto-oncogen Ser/Thr kinase PIM1 (PIM1)	2O65 and others	0.8	3.5	100	49	n/d
Checkpoint kinase 1 (CHK1)	2CGU and others	3.6	2.8	100	n/d	n/d
Abelson (Abl)	1FPU and others	4.5	1.6	48	n/d	Phenazopyridine is inactive at 100 µM on the wt
Stem cell factor receptor (c-KIT)	1PKG	6.0	5.4	41 (wt), 32 (Val559Asp, best mutant)	n/d	n/d
phosphatidylinositol-4,5- bisphosphate 3-kinase γ (PIK3CG)	2A5U and others	6.1	2.3	27	n/d	n/d
Mitogen-activating protein kinase interacting Ser/Thr kinase (MNK2)	2HW7	30.3	0.6	6.6	20	n/d
Casein kinase 1 isoform γ 3 (CK1)	2IZU	74.6	0.4	44	n/d	n/d

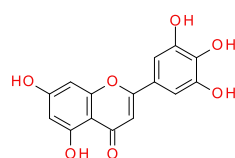


(Figure legend on the next page)

Figure 74: The human kinome tree ¹²⁷ and (a) the predicted and (b) validated *in vitro* targets, one potential target and one false positive target of phenazopyridine. Predicted targets are marked with a cyan dot. Matches between *in silico* prediction and *in vitro* validation are shown in green, the *in silico* false positive in orange and the targets identified by the kinome scan but not the *in silico* screening in red. The human kinome encompasses 518 kinases and can be divided into seven main groups: AGC (containing the PKA, PKG, and PKC families), CAMK (calcium/calmodulin-dependent protein kinases), CK1 (casein kinase 1), CMGC (containing the CDK, MAPK, GSK3, and CLK families), STE (homologues of yeast Sterile 7, Sterile 11, and Sterile 20 kinases), TK (tyrosine kinase), and TKL (tyrosine kinase-like). The full name corresponding to each kinase abbreviation in the kinome tree can be found in ¹²⁷.

Proto-oncogen Ser/Thr kinase PIM1 (PIM1)

Most of the established selection criteria do apply to PIM1. The average phenazopyridine fitness score for the 10 PIM1 entries is above 50. The keyword and literature search show a clear involvement in neurogenesis. Based on the score-based ranking, a relaxed RMSD and 49% of remaining activity, PIM1 is identified as target of phenazopyridine with a ranking in the first 0.8%. The RMSD criterion of RMSD > 2 Å had to be made less stringent because it was realized that some cognate ligands with many H-bond donor/acceptor functions, like the pentahydroxyflavone depicted below (RMSD = 3.5 Å), had rather large RMSDs. It is also noteworthy that the activity assay data (49% of remaining activity) could not be reproduced by the competitive binding assay, where no effect at all was detected with phenazopyridine. This difference is quite often observed in general and is due to conformational issues, i.e. the fact that the former accounts for enzyme dynamics, while the latter considers a static protein.



IC₅₀ = 650 nm (BindingDB)

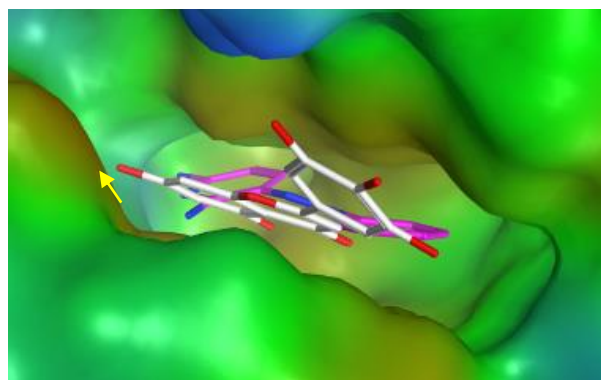


Figure 75: Pim1 kinase in complex with 5,7-dihydroxy-2-(3,4,5-trihydroxyphenyl)-4H-chromen-4-one (PDB ID: 2O65) and docked phenazopyridine. The carbon atoms are coloured in white for the cognate ligand and in magenta for phenazopyridine. The H-bond of the cognate ligand with the target is indicated by a yellow arrow. Here, and in the following Figures, the IC₅₀ value refers to the native ligand.

Another PIM1 in complex with AMP (PDB ID: 1YXU) is found in the first 3.6 %, followed by one in complex with (3E)-3-[(4-OH-phenyl)imino]-1H-indol-2(3H)-one (PDB ID: 1YXX) in the first 7.5%. Seven other PIM1 are much worse ranked because water molecules which are involved in PIM1 inhibitor binding ¹²⁸ have been omitted. Comparing the PIM1 structures from the sc-PDB, the ones meanwhile deposited in the PDB, and the newer ones from the Structural Genomics Consortium (<http://www.sgc.ox.ac.uk/>), a first insight into the propensity of structural adaptation of this enzyme upon ligand binding can be gained.

Up to date, several crystal structures of PIM1 have been solved, from which it became evident that their hinge region between the N- and C-terminal lobe resembles the one of phosphatidylinositol 3-kinase closely ¹²⁹, which is another putative target of phenazopyridine that is presented in this chapter. However, they feature a unique proline (Pro123) in the hinge region and an insertion of one or two residues right afterwards, conferring selectivity to PIM1. Interestingly, substituted pyridones featuring a similar planarity as phenazopyridine were identified previously as inhibitors of the enzyme ¹²⁸.

With respect to neuronal differentiation, PIM1 and PIM3 are known to be involved in cell proliferation and survival ¹³⁰, particularly the maintenance of self-renewal in cultured mouse stem cells ¹³⁰. Their inhibition could thus contribute to the elimination of undifferentiated cells. Additionally, they also play a role in the differentiation of the hematopoietic cell lineage ¹³¹ and chromatin dynamics ¹³². Later on in development, PIM1 expression is induced by long-term potentiation (LTP) stimulation in the hippocampus and required for the consolidation of enduring LTP in the brain. Furthermore, its level of expression and, consequently, its capacity to phosphorylate target proteins in dendritic and nuclear compartments of stimulated neurons might be a determining factor for the establishment of long-lasting changes in synaptic strength. Also, post-mitotic PIM1 expression seems to be part of the functional plasticity of the brain. Indeed, PIM1 mRNA expression levels are differentially regulated by neuronal activity ¹³³. PIM1 expression is regulated by growth factors, cytokines, and mitogens via the JAK/STAT pathway.

Based on these findings, PIM1 can be considered a true target of phenazopyridine.

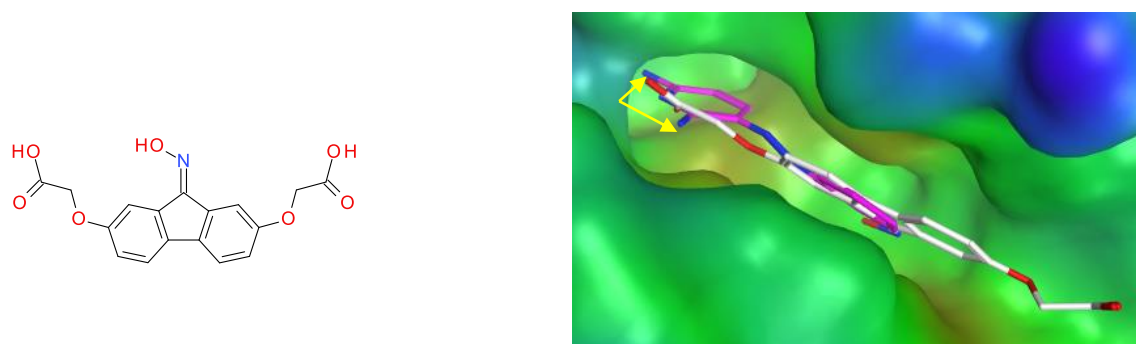
Checkpoint kinase 1 (CHK1)

CHK1 displays a phenazopyridine score-based ranking in the high 3.6% with a fitness score above 50 and a RMSD of 2.8 Å for its cognate ligand.

Besides its role in cell survival and checkpoint control ¹³⁴, CHK1 coordinates cell cycle progression ¹³⁵, preserves genome integrity ¹³⁵, and is required for mammalian homologous recombination repair ¹³⁶. In normal cells, DNA damage causes cell cycle arrest through the tumor suppression protein p53 and activation of CHK1, which allows for repair. CHK1 phosphorylates p53 at multiple DNA damage-inducible sites ¹³⁷ and regulates acetylation ¹³⁸. In tumor cells that have an impaired p53 function, a defect common to 50–70% of all cancers, CHK1 inhibition results in the abrogation of the S-phase and premature cell cycle progression into mitosis where cell death occurs through mitotic catastrophe and apoptotic pathways ¹³⁹. CHK1 is necessary for an intra-S phase checkpoint, ensuring that the activation of late replication origins is blocked and arrested replication fork integrity is maintained when DNA synthesis is inhibited ¹⁴⁰. It both mediates S- and G2-phase arrests through Cdc25A degradation in response to DNA-damaging agents ¹⁴¹ and negatively regulates the mitotic function of Cdc25A phosphatase through 14-3-3 binding ¹⁴². In embryonic cortical neurons treated with a DNA-damaging agent, Cdc25A levels rise so as to inactivate CHK1 and promoting neuronal death ¹⁴³.

CHK1 was found to be inactive by the kinome scan, although its role in checkpoint control and cell survival is mentioned in the literature ¹³⁴. Totally, phenazopyridine was docked into 13 CHK1 proteins, with heterogeneous ranking percentages ranging from 3.6% to 94.3%. Based on this finding and the kinome scan indicating that CHK1 kept its full activity upon phenazopyridine treatment, this protein is treated as a FP.

As seen in the next Figure, the cognate ligand and phenazopyridine roughly share the same plane, which was also observed for the closely related CHK2.

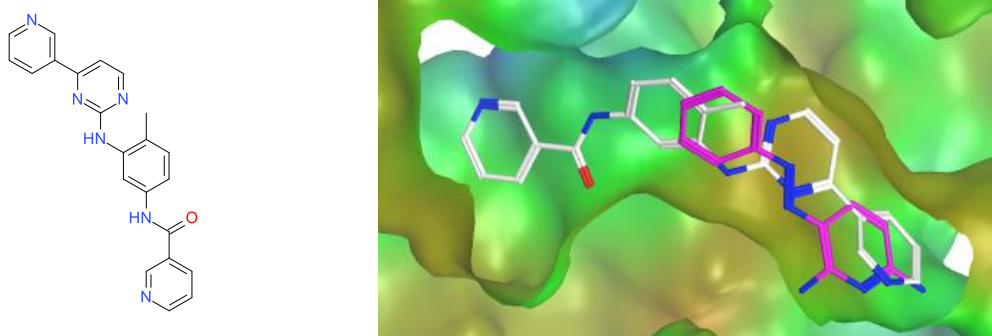


IC₅₀ = 13.4 μM (BindingDB and BindingMOAD)

Figure 76 Checkpoint kinase 1 with 2,2'-[9-(hydroxyimino)-9H-fluorene-2,7-diyl]bis(oxy)} diacetic acid (PDB ID: 2CGU) and docked phenazopyridine. The same colouring as in Figure 75 is applied. H-bonds of the cognate ligand with the target are indicated by yellow arrows.

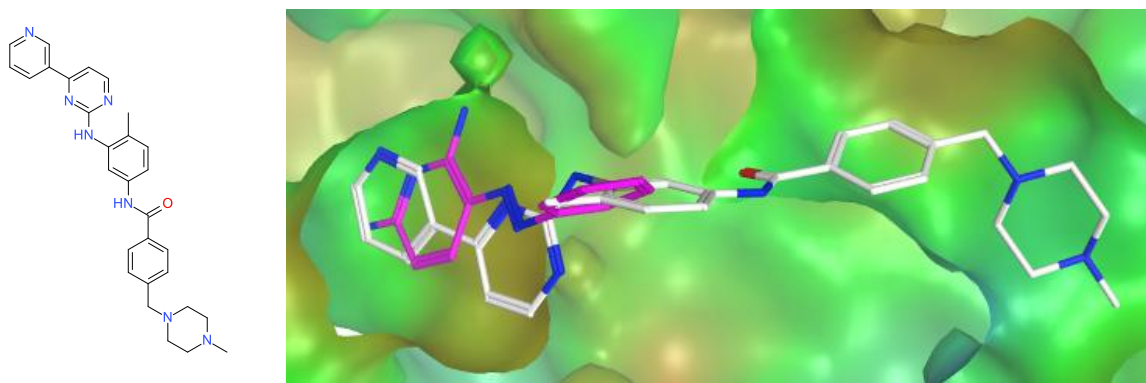
Abelson kinase (Abl)

In the phenazopyridine-based ranking, the first inactive conformation of Abl ranks among the top 4.5% and has a fitness score above 50. Its docked cognate ligand, an imatinib analogue lacking the methylpiperazine moiety (Figure 77), which is known to increase solubility¹⁴⁴, has a RMSD of 1.6 Å. The second inactive conformation ranks among the first 8.1%, had a RMSD of 0.8 Å, and a fitness score also higher than 50 is obtained. The imatinib (cognate ligand) (Figure 78) docking is with a RMSD of 0.8 Å successful.



$K_i = 37$ nm (BindingMOAD)

Figure 77: Abelson kinase in complex with N-[4-methyl-3-[[4-(3-pyridinyl)-2-pyrimidinyl]amino]phenyl]-3-pyridinecarboxamide (PDB ID: 1FPU) and docked phenazopyridine. The same colouring as in Figure 75 is applied. The protein is z-clipped for better visibility of the ligand.



$K_i = 13$ nm (BindingDB)

Figure 78: Abelson kinase in complex with imatinib (PDB ID: 2HYY) and docked phenazopyridine. The same colouring as in Figure 75 is applied. The protein is z-clipped for better visibility of the ligand.

Remarkably, with respect to the seven Abl structures in the screening database, there is a clear score gap between the inactive (better ranked) and active conformation of Abl, suggesting that if phenazopyridine binds, it does it preferentially to the inactive conformation of Abl.

An experimental activity of 48% was measured for the unphosphorylated His396Pro mutant of Abl in the kinome scan, while phenazopyridine had no effect on the wild type at 100 µM and showed a 35% inhibition at 500 µM (Figure 79).

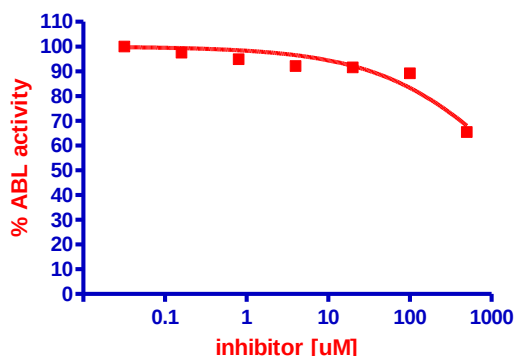


Figure 79: Wild type Abelson is not inhibited by phenazopyridine up to 100 μM and causes 35% inhibition at 500 μM . The Figure was kindly provided by Dr. L. Mologni from the University of Milano-Bicocca, Italy.

The discrepancy between the kinome scan and the activity assay is potentially because phenazopyridine binds to Abl but does not inhibit it (at least at concentrations below 500 μM). An analogy to GNF-2 (a pyrimidine benzamide) binding to the myristoyl binding site ($\text{IC}_{50} = 267 \text{ nM}$ ¹⁴⁵) can be seen. GNF2 competes with myristate for binding at the N-terminal myristate-binding pocket of Abl ¹⁴⁶. GNF2 binding induces or stabilizes the inactive Abl conformation and is non-competitive with respect to ATP ¹⁴⁷. In this sense, it is a representative of allosteric kinase inhibitors discovered in 2006 ¹⁴⁵, which neither target the active nor the inactive ATP binding pocket. A recent comprehensive review on both ATP- and myristate-competitive Abl inhibitors is provided by Fabbro et al. ¹⁴⁸.

Considering the biological context, Abl is involved in neural morphogenesis ¹⁴⁹, and cell-cell adhesive junctions ¹⁵⁰ in both the central nervous system and the epidermis ¹⁵¹.

Based on conflicting experimental data, Abl is a potential target, but further experimental studies are needed.

Stem cell factor receptor (c-KIT)

The kinase c-KIT ranking in the first 6.0% of the score-based ranking and had a fitness score above 50. The RMSD of the cognate ligand ADP was with 5.3 Å rather high, and not a reliable selection criterion. The high RMSD is not astonishing given the flexibility of its side chains. Its wild type activity was measured to be 41% (kinome scan) upon treatment with phenazopyridine.

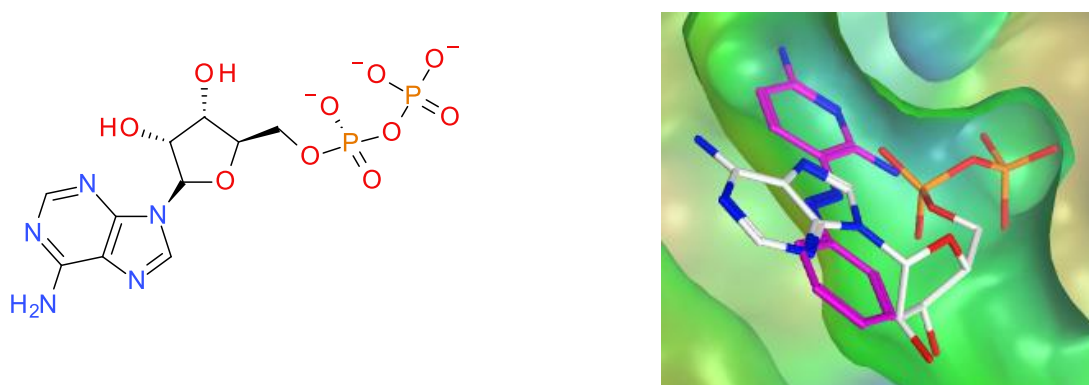


Figure 80: c-KIT in complex with its cognate ligand ADP (PDB ID: 1PKG) and docked phenazopyridine. The same colouring as in Figure 75 is applied. The protein is z-clipped for better visibility of the ligand.

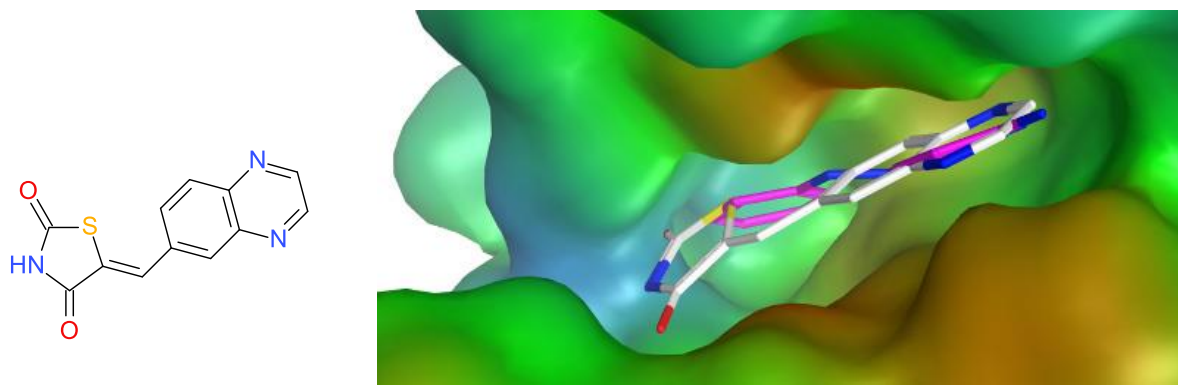
Upon binding its ligand, the stem cell factor (SCF, a growth and differentiation factor for hematopoietic progenitor cells), c-KIT forms an active dimer that autophosphorylates itself and activates a signalling

cascade that induces cell growth. Disease-causing human mutations that activate the SCF-independent constitutive expression of c-KIT are found in acute myelogenous leukemia, human mast cell disease, and gastrointestinal stromal tumors. c-KIT takes over pivotal functions in promoting the development, survival, and proliferation of hematopoietic stem cells, neuronal crest-derived cells, and germ cells. Upon SCF stimulation, the c-KIT receptor associates with and/or phosphorylates other signal transduction kinases, phosphatases, signal transducers, and activators of transcription ¹⁵².

Based on literature evidence, the score-based ranking and the kinome scan data, c-KIT is seen as a true positive hit.

Phosphatidylinositol-4,5-bisphosphate 3-kinase γ (PIK3CG)

PIK3CG is found to be ranked in the first 6.1% of the score-based ranking with a fitness score above 50, a RMSD of 2.3 Å of the cognate ligand, and a residual activity of 27% (kinome scan).



IC₅₀ = 8 nM (BindingDB)

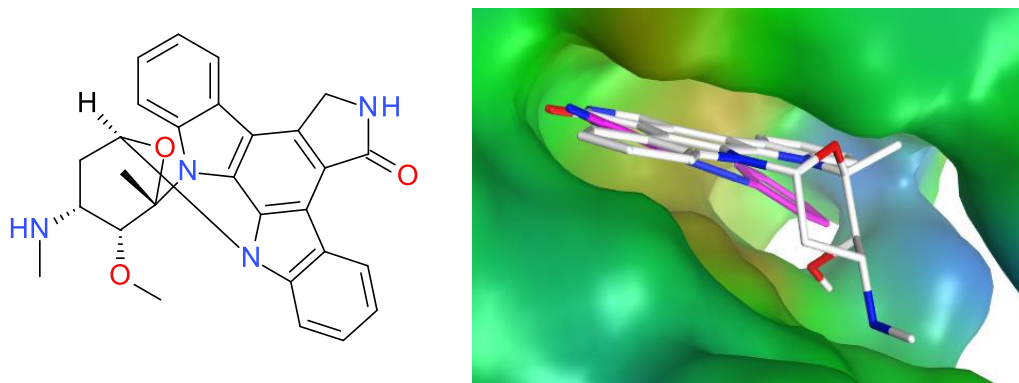
Figure 81: Phosphatidylinositol-4,5-bisphosphate 3-kinase γ in complex with its cognate ligand (5E)-5-(quinoxalin-6-ylmethylene)-1,3-thiazolidine-2,4-dione (PDB ID: 2A5U) and docked phenazopyridine. The same colouring as in Figure 75 is applied. The protein is z-clipped for better visibility of the ligand.

PIK3CG is involved in the regulation of both murine embryonic stem cell pluripotency and proliferation ¹⁵³. Besides, PIK3CG signalling in hypothalamic neurons contributes to the regulation of glucose homeostasis ¹⁵⁴.

Provided this evidence, it is considered as a potential target.

Mitogen-activating protein kinase interacting Ser/Thr kinase 2 (MNK2)

MNK2 shares a sequence homology of 70% with PIM1 but provided the huge active site, phenazopyridine has many possibilities to be accommodated, leading to a rank above the threshold of 10% and a fitness score being with 46.28 slightly below 50. There is only one MNK2 present in the target database. The low RMSD for the large cognate ligand staurosporin of 0.6 Å is remarkable. Both the kinome scan and the Dundee screening reveal MNK2 to be well inhibited by phenazopyridine (6.6% and 20% of remaining activity at 10 μ M, respectively).



$K_d = 17 - 22$ nm (BindingDB)

Figure 82: MNK2 with staurosporine (PDB ID: 2HW7) and docked phenazopyridine. The same colouring as in Figure 75 is applied. The protein is z-clipped for better visibility of the ligand.

In the ranking according to the MEK1 inhibitor PD98059, MNK2 appears in the first 0.1%, hinting at a possible use of this compound as MNK2 inhibitor.

MNKs are activated via phosphorylation by the growth factor-stimulated Ras/extracellular signal-regulated kinase pathway and the stress-induced p38 pathway. Upon traumatic brain injury, mRNA translation in dendrites is stimulated via the mTOR pathway and eIF4E, a rate-limiting mRNA translation factor, phosphorylated by MNK1¹⁵⁵. The kinase activity of MNKs, working eventually through phosphorylation of eIF4E, may serve to limit cap-dependent translation (enhanced by PI3K) under physiological conditions¹⁵⁶. No direct link to neurogenesis was found in the literature.

MNK2 is assumed to be a target. Promising experimental results have also been obtained for the homologue MNK1, of which no holo crystal structure exists to date.

Casein kinase 1 (CK1)

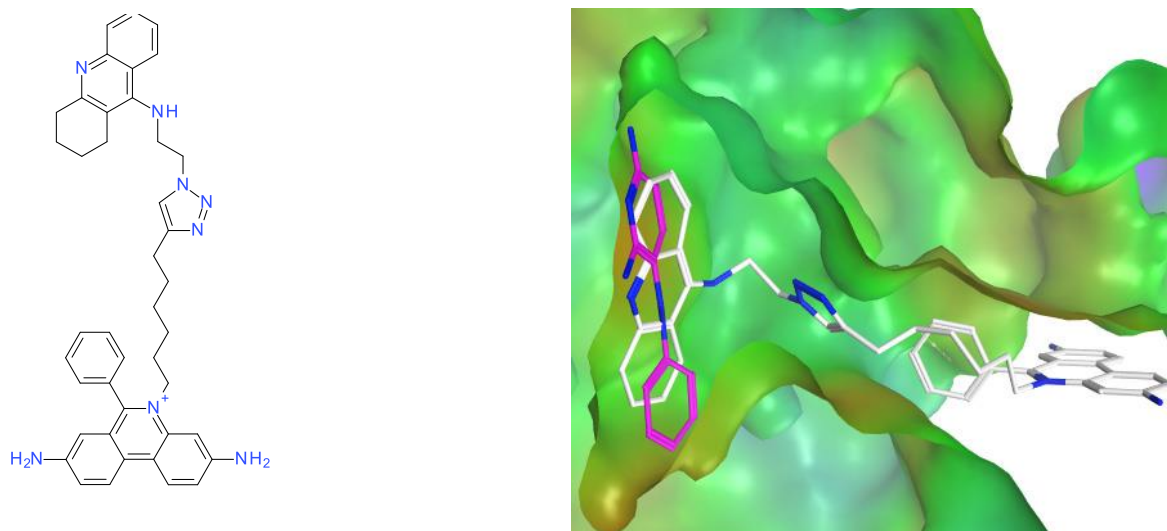
The RMSD of the cognate ligand 2-({6-[(3-chlorophenyl)amino]-9-isopropyl-9H-purin-2-yl}amino)-3-methylbutan-1-ol of 0.4 Å is excellent and the kinome scan reveals an activity of 44% upon phenazopyridine treatment. However, with a score-based ranking threshold of 10% and a fitness score of 40.15, CSK1 would not have been identified.

Casein kinases are involved in cell differentiation, particularly hematopoietic cell differentiation¹⁵⁷, proliferation, apoptosis, circadian rhythms^{158,159}, and membrane transport. In vertebrates, the Wnt signaling pathway regulates axis formation, organ development as well as cellular proliferation, morphology, motility, and fate¹⁶⁰. Dpr acts as a molecular switch, inhibiting Wnt signalling when unphosphorylated but promoting Wnt signaling when phosphorylated by casein kinase I δ/ϵ ¹⁶¹. Its upregulation in dorsal root ganglia and spinal cord after mouse spinal nerve injury contributes to neuropathic pain¹⁶².

CK1 is thus considered as a true target.

Target awaiting experimental confirmation: Acetylcholinesterase (AChE)

AChE awaits experimental confirmation. The highest scored AChE ranked in the top 0.1% with a RMSD of 1.8 Å. Figure 83 shows the gorge and binding pocket of *Torpedo californica* acetylcholinesterase with the crystal pose of the cognate ligand and the docking pose of phenazopyridine, roughly aligned with the tricyclic moiety in the pocket interior.



$K_i = 0.01$ nm (BindingMOAD)

Figure 83: Acetylcholine esterase from *Torpedo californica* with its cognate ligand 3,8-diamino-6-phenyl-5-[6-[1-[2-[(1,2,3,4-tetrahydro-9-acridinyl)amino]ethyl]-1H-1,2,3-triazol-4-yl]hexyl]-phenanthridinium (PDB ID: 1Q84) and docked phenazopyridine.

AChE is expressed during murine embryonic stem cell differentiation¹⁶³. Yang et al. showed it to play an important role in the procession of neuroblastoma cell apoptosis, which makes a link between AChE and neuronal apoptosis in Alzheimer disease probable¹⁶⁴.

Target prediction success

Target prediction success is defined as the number of predicted targets divided by the number of experimentally confirmed targets. In the search for targets of phenazopyridine, the success rate was 80% when strictly only considering the targets in the 10% of the phenazopyridine-based ranking and 71% to 83% when considering all the TPs out of 7 the kinases listed in Table 41 (not including or including Abl).

Table 41: Overview of the predicted and/or experimentally validated targets. + stands for validated *in vitro*, (+) for potential target and – for false positive.

Predicted <i>in silico</i>	Validated <i>in vitro</i>
PIM1 kinase	+
Stem cell factor receptor	+
Phosphatidylinositol-4,5-bisphosphate 3-kinase γ	+
Mitogen-activating protein kinase interacting Ser/Thr kinase 2	+
Casein kinase 1	+
Abelson (Abl)	(+)
Checkpoint kinase 1	-

The confirmed targets include: PIM1 kinase, the stem cell factor receptor c-kit, phosphatidylinositol-4,5-bisphosphate 3-kinase γ , mitogen-activating protein kinase interacting Ser/Thr kinase 2, and casein kinase 1. Further investigations are needed on Abelson kinase. The checkpoint kinase 1 can be ruled out as a target.

As a comparison, another approach yielded 56% for an inverse screening with eight common drugs ⁶. The success rate depends heavily on the biological question asked, the docking-scoring combination, the completeness of the target database, the complexity of the query molecule used, and the quality and amount of corresponding experimental data at hand.

Considering the proteins manually added to the sc-PDB, human tau protein kinase and glycogen synthase kinase-3 were badly ranked in the phenazopyridine score-based ranking. A MAPK ranked in the first 13% and one of the two AKT2 structures ranked in the first 2.4%, but these were not tested experimentally. Additionally, many targets crystallized with flavonoids were well-ranked, reflecting experimental findings by Suter et al. (personal communication) that flavonoids also induce neuronal differentiation, although to a lower extent than phenazopyridine. It is likely that also here kinases are involved, as flavonoids work mainly via (unspecific) kinase inhibition ^{107,165}. From a docking point of view, the conjugated cycles of flavonoids can be matched with phenazopyridine and vice versa. Furthermore, the flavonoid hydroxyls mimic the NH₂ groups of phenazopyridine.

4.3 EXTENSION OF THE ANALYSIS

Knowing the limitation of inverse screening that only 3D structures present in the target database can be retrieved, and based on the obtained results, the analysis could be extended by:

- Clustering cavities on the basis of phenazopyridine binding patterns, similarly to Snyder et al. ¹⁶⁶
- Consensus scoring ^{122,167}: Allows by-passing some limitations of single scoring functions ³⁷, as combining two or three scoring functions has already shown to improve the success rate ⁵⁰. It has not yet been applied to inverse docking. Consensus scoring is not useful if it cannot effectively eliminate the FPs obtained by single scorings ¹⁶⁸.
- Post-docking fingerprinting devised to alter scores in such a way as to better match the biological reality, better quantify docking success than the commonly used RMSD value, and further enrich true actives ^{120,169}. Here, post-docking filtering for a biological context was applied instead to account for the biological context of neuronal differentiation. It revealed to be better for target spotting than narrowing the set of targets upfront docking (as in the TarFisDock approach ¹¹, using e.g. the functional categories “synaptic”, “neuroeffector”, “junctional sites”, and “central nervous system”).
- A measure of the distance between the centre of mass of the protein and the cognate ligand or between the centre of mass of the phenazopyridine and the cognate ligand pose could be implemented as a new decision parameter for cases where the RMSD is not sufficient.

Based on the here established targets of phenazopyridine, a better coverage of the target space by improving target finding for non-crystallized targets or targets not included in the sc-PDB could involve a search on the sequence level by the following means:

- SPASM (<http://xray.bmc.uu.se/usf/spasm.html>) ¹⁷⁰ for finding homologues with similar binding site arrangements,
- the ProSite (<http://www.expasy.ch/prosite/>) ¹⁷¹ database for a scan with 2D fingerprints of the active site residues or
- the PINTS server ¹⁷² (<http://www.russell.embl-heidelberg.de/pints/>) for finding similar pattern in less related or non-related proteins that do not necessarily have the same fold, followed by docking phenazopyridine in the corresponding crystal structures or homology models.

Finally, detecting allosteric inhibition effects could be of interest in future inverse screenings. For identifying allosteric sites, a cavity finder tool (as e. g. implemented in Sybyl (Tripos, Inc.), MOE (Chemical Computing Group, Inc.) or the Schrödinger Program Suite (Schrödinger, Inc.)) would have to be used.

4.4 EXPERIMENTAL VALIDATION OF KINASES

The pyridine-2,6-diamine moiety of phenazopyridine is a typical kinase inhibition motif ¹⁷³. With respect to such a motif, it is well known that compounds acting on a given protein family are rarely far apart from each other in chemical space ¹⁷⁴. Indeed, kinases were retrieved as hits by inverse screening.

Therefore, the effect of phenazopyridine on the activity of a panel of 70 kinases was investigated. Interestingly, phenazopyridine was found to inhibit the activity of several kinases, whereas the analogue lacking the pyridine-2,6-diamine part did not affect neuronal differentiation (compound d in Figure 70) and did not have any kinase inhibitory activity (Figure 84). Thus, kinase inhibition is likely to be involved in the effects of phenazopyridine described in this study.

Besides, different kinases are known to be key-players in differentiation. MAP kinases have been shown to be dispensable for the embryonic stem cell renewal and cell cycle but are involved in the modulation of embryonic stem cell lineage commitment ¹⁷⁵. Provided that phenazopyridine is an analgesic acting against urinary tract pain and irritation, anti-inflammatory properties of kinases like p38 MAPK ¹⁷⁶ or JNK ¹⁷⁷ and pathways related to them deserve attention as well.

4.5 CONCLUSIONS

Inverse screening has become a promising computational tool to find the probable target proteins for active compounds ^{7-9,11,178}. In one of the cited cases, targets were confirmed not only by binding assays, but also by crystal structures ¹⁷⁸. Inverse screening can also provide clues on new uses for old drugs ^{179,180}.

By combining *in silico* target screening with a literature search and experimental data, the importance of kinases was corroborated by a knowledge-driven approach and weaknesses of the procedure identified. The experimentally tested kinases represent 402 kinases of the total kinome (518 kinases). Considering the percentage of TPs found, i.e. 1.5%, only one or two kinases can be expected to be further targets. Importantly, a phenazopyridine analogue devoid of the presumable kinase inhibition motif lost the effect on neuronal differentiation. Thus, the *in silico/in vitro* results suggest that kinase inhibition participates in the effect of phenazopyridine on neuronal differentiation of embryonic stem cells. Given the profound and complex effects of phenazopyridine, it is likely that the concomitant inhibition of several kinases is important, together with other enzymes as the yet to be tested acetylcholinesterases. Knowing that kinase and phosphatase activities are linked, phenazopyridine could also indirectly modulate phosphatases by acting on kinases.

Inverse screening was useful for lowering the number of targets related to the biological context. Considering only the top 10% of the phenazopyridine score-based ranking, a success rate of 80% when compared to all available experimental hits was achieved, which is far superior to random picking. Provided a curated database like the sc-PDB, appropriate post-docking treatment, and experimental results, a rough selectivity profile of phenazopyridine could be identified.

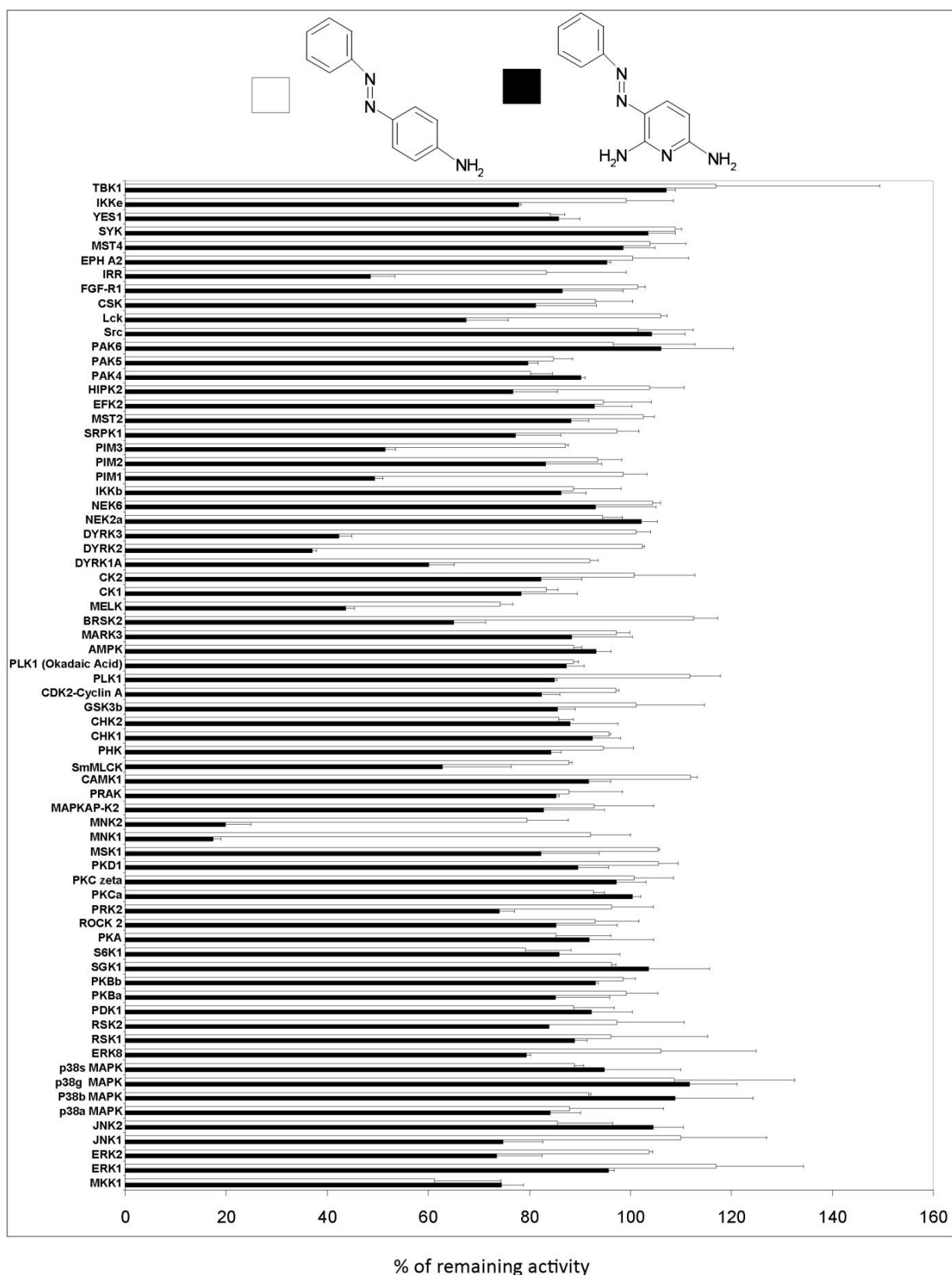


Figure 84: Phenazopyridine is a kinase inhibitor. The effects of phenazopyridine (black bars, compound a in Figure 70) and of one inactive analogue (white bars, compound d in Figure 70) were tested on a panel of 70 human recombinant kinases. The full name of each abbreviated kinase is given in the annex. Values are presented as residual kinase activity. The data were kindly provided by Sir P. Cohen and J. Bain from the University of Dundee (UK). Dr. D.M. Suter is acknowledged for having provided this Figure.

4.5 ACKNOWLEDGEMENTS

Dr. D. M. Suter and Prof. Dr. K.-H. Krause (Department of Pathology and Immunology at the Centre Médical of the University of Geneva) are thanked for inspiring discussions on target discovery for phenazopyridine and for providing their experimental data. Furthermore, J. Bain and Sir Prof. P. Cohen (Sir James Black Centre, College of Life Sciences, University of Dundee, UK), are acknowledged for having tested phenazopyridine on their kinase panel and Dr. L. Mologni (University of Milano-Bicocca, IT) for having tested phenazopyridine on Abelson. Sincere thanks go to Dr. S. Ahmed for many helpful discussions, D. Vitel for having performed his diploma work on inverse screening, and C. Durand for having written a macro in Excel facilitating the import of target-relevant information from PDB headers.

4.6 REFERENCES

1. Suter, D. M., Preynat-Seauve, O., Tirefort, D., Feki, A. & Krause, K. H. Phenazopyridine induces and synchronizes neuronal differentiation of embryonic stem cells. *J Cell Mol Med* **13**, 3517-27 (2009).
2. Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* **267**, 727-48 (1997).
3. Leach, A. R., Shoichet, B. K. & Peishoff, C. E. Prediction of protein-ligand interactions. Docking and scoring: successes and gaps. *J Med Chem* **49**, 5851-5 (2006).
4. Singh, J., Deng, Z., Narale, G. & Chuaqui, C. Structural interaction fingerprints: a new approach to organizing, mining, analyzing, and designing protein-small molecule complexes. *Chem Biol Drug Des* **67**, 5-12 (2006).
5. Rockey, W. M. & Elcock, A. H. Rapid computational identification of the targets of protein kinase inhibitors. *J Med Chem* **48**, 4138-52 (2005).
6. Chen, Y. Z. & Ung, C. Y. Prediction of potential toxicity and side effect protein targets of a small molecule by a ligand-protein inverse docking approach. *J Mol Graph Model* **20**, 199-218 (2001).
7. Chen, Y. Z. & Zhi, D. G. Ligand-protein inverse docking and its potential use in the computer search of protein targets of a small molecule. *Proteins* **43**, 217-26 (2001).
8. Paul, N., Kellenberger, E., Bret, G., Muller, P. & Rognan, D. Recovering the true targets of specific ligands by virtual screening of the protein data bank. *Proteins* **54**, 671-80 (2004).
9. Muller, P. et al. In silico-guided target identification of a scaffold-focused library: 1,3,5-triazepan-2,6-diones as novel phospholipase A2 inhibitors. *J Med Chem* **49**, 6768-78 (2006).
10. Zahler, S. et al. Inverse in silico screening for identification of kinase inhibitor targets. *Chem Biol* **14**, 1207-14 (2007).
11. Li, H. et al. TarFisDock: a web server for identifying drug targets with docking approach. *Nucleic Acids Res* **34**, W219-24 (2006).
12. Schneider, M. A rational approach to maximize success rate in target discovery. *Arch Pharm (Weinheim)* **337**, 625-33 (2004).
13. Mestres, J., Martin-Couce, L., Gregori-Puigjane, E., Cases, M. & Boyer, S. Ligand-based approach to in silico pharmacology: nuclear receptor profiling. *J Chem Inf Model* **46**, 2725-36 (2006).
14. Kauvar, L. M. et al. Predicting ligand binding to proteins by affinity fingerprinting. *Chem Biol* **2**, 107-18 (1995).
15. Keiser, M. J. et al. Relating protein pharmacology by ligand chemistry. *Nat Biotechnol* **25**, 197-206 (2007).
16. Xia, X., Maliski, E. G., Gallant, P. & Rogers, D. Classification of kinase inhibitors using a Bayesian model. *J Med Chem* **47**, 4463-70 (2004).
17. Kamper, A., Apostolakis, J., Rarey, M., Marian, C. M. & Lengauer, T. Fully automated flexible docking of ligands into flexible synthetic receptors using forward and inverse docking strategies. *J Chem Inf Model* **46**, 903-11 (2006).
18. Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R. & Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J Mol Biol* **161**, 269-88 (1982).
19. Jones, G., Willett, P. & Glen, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J Mol Biol* **245**, 43-53 (1995).
20. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W. & Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins* **52**, 609-23 (2003).

21. Jones, S. & Thornton, J. M. Protein-protein interactions: a review of protein dimer structures. *Prog Biophys Mol Biol* **63**, 31-65 (1995).
22. Kontoyianni, M., McClellan, L. M. & Sokol, G. S. Evaluation of docking performance: comparative data on docking algorithms. *J Med Chem* **47**, 558-65 (2004).
23. Perola, E., Walters, W. P. & Charifson, P. S. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins* **56**, 235-49 (2004).
24. Kellenberger, E. et al. sc-PDB: an annotated database of druggable binding sites from the Protein Data Bank. *J Chem Inf Model* **46**, 717-27 (2006).
25. Klebe, G. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discov Today* **11**, 580-94 (2006).
26. Block, P., Sotriffer, C. A., Dramburg, I. & Klebe, G. AffinDB: a freely accessible database of affinities for protein-ligand complexes from the PDB. *Nucleic Acids Res* **34**, D522-6 (2006).
27. Chen, X., Liu, M. & Gilson, M. K. BindingDB: a web-accessible molecular recognition database. *Comb Chem High Throughput Screen* **4**, 719-25 (2001).
28. Chen, X., Lin, Y., Liu, M. & Gilson, M. K. The Binding Database: data management and interface design. *Bioinformatics* **18**, 130-9 (2002).
29. Liu, T., Lin, Y., Wen, X., Jorissen, R. N. & Gilson, M. K. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res* **35**, D198-201 (2007).
30. Hu, L., Benson, M. L., Smith, R. D., Lerner, M. G. & Carlson, H. A. Binding MOAD (Mother Of All Databases). *Proteins* **60**, 333-40 (2005).
31. Benson, M. L. et al. Binding MOAD, a high-quality protein-ligand database. *Nucleic Acids Res* **36**, D674-8 (2008).
32. Gao, Z. et al. PDTD: a web-accessible protein database for drug target identification. *BMC Bioinformatics* **9**, 104 (2008).
33. Chen, X., Ji, Z. L. & Chen, Y. Z. TTD: Therapeutic Target Database. *Nucleic Acids Res* **30**, 412-5 (2002).
34. Wishart, D. S. et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res* **34**, D668-72 (2006).
35. Wishart, D. S. et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res* **36**, D901-6 (2008).
36. Rognan, D. Development and virtual screening of target libraries. *J Physiol Paris* **99**, 232-44 (2006).
37. Bissantz, C., Folkers, G. & Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J Med Chem* **43**, 4759-67 (2000).
38. Warren, G. L. et al. A critical assessment of docking programs and scoring functions. *J Med Chem* **49**, 5912-31 (2006).
39. Esposito, E. X., Baran, K., Kelly, K. & Madura, J. D. Docking of sulfonamides to carbonic anhydrase II and IV. *J Mol Graph Model* **18**, 283-9, 307-8 (2000).
40. Tao, P. & Lai, L. Protein ligand docking based on empirical method for binding affinity estimation. *J Comput Aided Mol Des* **15**, 429-46 (2001).
41. Terp, G. E., Johansen, B. N., Christensen, I. T. & Jorgensen, F. S. A new concept for multidimensional selection of ligand conformations (MultiSelect) and multidimensional scoring (MultiScore) of protein-ligand binding affinities. *J Med Chem* **44**, 2333-43 (2001).
42. Gohlke, H. & Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew Chem Int Ed Engl* **41**, 2644-76 (2002).
43. Wang, R., Lai, L. & Wang, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J Comput Aided Mol Des* **16**, 11-26 (2002).
44. Olsen, L., Pettersson, I., Hemmingsen, L., Adolph, H. W. & Jorgensen, F. S. Docking and scoring of metallo-beta-lactamases inhibitors. *J Comput Aided Mol Des* **18**, 287-302 (2004).
45. Wang, R., Lu, Y., Fang, X. & Wang, S. An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein-ligand complexes. *J Chem Inf Comput Sci* **44**, 2114-25 (2004).
46. Ha, S., Andreani, R., Robbins, A. & Muegge, I. Evaluation of docking/scoring approaches: a comparative study based on MMP3 inhibitors. *J Comput Aided Mol Des* **14**, 435-48 (2000).
47. Perez, C. & Ortiz, A. R. Evaluation of docking functions for protein-ligand docking. *J Med Chem* **44**, 3768-85 (2001).

48. Schulz-Gasch, T. & Stahl, M. Binding site characteristics in structure-based virtual screening: evaluation of current docking tools. *J Mol Model* **9**, 47-57 (2003).
49. Kellenberger, E., Foata, N. & Rognan, D. Ranking targets in structure-based virtual screening of three-dimensional protein libraries: methods and problems. *J Chem Inf Model* **48**, 1014-25 (2008).
50. Wang, R., Lu, Y. & Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J Med Chem* **46**, 2287-303 (2003).
51. Ferrara, P., Gohlke, H., Price, D. J., Klebe, G. & Brooks, C. L., 3rd. Assessing scoring functions for protein-ligand interactions. *J Med Chem* **47**, 3032-47 (2004).
52. Kroemer, R. T. et al. Assessment of docking poses: interactions-based accuracy classification (IBAC) versus crystal structure deviations. *J Chem Inf Comput Sci* **44**, 871-81 (2004).
53. Kontoyianni, M., Sokol, G. S. & McClellan, L. M. Evaluation of library ranking efficacy in virtual screening. *J Comput Chem* **26**, 11-22 (2005).
54. Bursulaya, B. D., Totrov, M., Abagyan, R. & Brooks, C. L., 3rd. Comparative study of several algorithms for flexible ligand docking. *J Comput Aided Mol Des* **17**, 755-63 (2003).
55. Kellenberger, E., Rodrigo, J., Muller, P. & Rognan, D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins* **57**, 225-42 (2004).
56. Friesner, R. A. et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* **47**, 1739-49 (2004).
57. Halgren, T. A. et al. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem* **47**, 1750-9 (2004).
58. Verdonk, M. L. et al. Virtual screening using protein-ligand docking: avoiding artificial enrichment. *J Chem Inf Comput Sci* **44**, 793-806 (2004).
59. Cummings, M. D., DesJarlais, R. L., Gibbs, A. C., Mohan, V. & Jaeger, E. P. Comparison of automated docking programs as virtual screening tools. *J Med Chem* **48**, 962-76 (2005).
60. Mozziconacci, J. C. et al. Optimization and validation of a docking-scoring protocol; application to virtual screening for COX-2 inhibitors. *J Med Chem* **48**, 1055-68 (2005).
61. Rarey, M., Kramer, B. & Lengauer, T. The particle concept: placing discrete water molecules during protein-ligand docking predictions. *Proteins* **34**, 17-28 (1999).
62. Osterberg, F., Morris, G. M., Sanner, M. F., Olson, A. J. & Goodsell, D. S. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins* **46**, 34-40 (2002).
63. Verdonk, M. L. et al. Modeling water molecules in protein-ligand docking using GOLD. *J Med Chem* **48**, 6504-15 (2005).
64. Schnecke, V. & Kuhn, L. A. Virtual screening with solvation and ligand-induced complementarity. *Perspectives in Drug Discovery and Design* **20**, 171-190 (2000).
65. Minke, W. E., Diller, D. J., Hol, W. G. & Verlinde, C. L. The role of waters in docking strategies with incremental flexibility for carbohydrate derivatives: heat-labile enterotoxin, a multivalent test case. *J Med Chem* **42**, 1778-88 (1999).
66. Rao, M. S. & Olson, A. J. Modelling of factor Xa-inhibitor complexes: a computational flexible docking approach. *Proteins* **34**, 173-83 (1999).
67. Birch, L., Murray, C. W., Hartshorn, M. J., Tickle, I. J. & Verdonk, M. L. Sensitivity of molecular docking to induced fit effects in influenza virus neuraminidase. *J Comput Aided Mol Des* **16**, 855-69 (2002).
68. Nissink, J. W. et al. A new test set for validating predictions of protein-ligand interaction. *Proteins* **49**, 457-71 (2002).
69. Pospisil, P., Kuoni, T., Scapozza, L. & Folkers, G. Methodology and problems of protein-ligand docking: case study of dihydroorotate dehydrogenase, thymidine kinase, and phosphodiesterase 4. *J Recept Signal Transduct Res* **22**, 141-54 (2002).
70. Pickett, S. D. et al. Discovery of novel low molecular weight inhibitors of IMPDH via virtual needle screening. *Bioorg Med Chem Lett* **13**, 1691-4 (2003).
71. Floriano, W. B., Vaidehi, N., Zamanakos, G. & Goddard, W. A., 3rd. HierVLS hierarchical docking protocol for virtual ligand screening of large-molecule databases. *J Med Chem* **47**, 56-71 (2004).
72. Bellocchi, D., Macchiarulo, A., Costantino, G. & Pellicciari, R. Docking studies on PARP-1 inhibitors: insights into the role of a binding pocket water molecule. *Bioorg Med Chem* **13**, 1151-7 (2005).
73. de Graaf, C., Pospisil, P., Pos, W., Folkers, G. & Vermeulen, N. P. Binding mode prediction of cytochrome p450 and thymidine kinase protein-ligand complexes by consideration of water and rescoring in automated docking. *J Med Chem* **48**, 2308-18 (2005).

74. de Graaf, C. et al. Catalytic site prediction and virtual screening of cytochrome P450 2D6 substrates by consideration of water and rescoring in automated docking. *J Med Chem* **49**, 2417-30 (2006).
75. Aparna, V., Rambabu, G., Panigrahi, S. K., Sarma, J. A. & Desiraju, G. R. Virtual screening of 4-anilinoquinazoline analogues as EGFR kinase inhibitors: importance of hydrogen bonds in the evaluation of poses and scoring functions. *J Chem Inf Model* **45**, 725-38 (2005).
76. McConkey, B. J., Sobolev, V. & Edelman, M. The performance of current methods in ligand-protein docking. *Current Science* **83**, 845-856 (2002).
77. Ding, S. et al. Synthetic small molecules that control stem cell fate. *Proc Natl Acad Sci U S A* **100**, 7632-7 (2003).
78. Takahashi, T. et al. Ascorbic acid enhances differentiation of embryonic stem cells into cardiac myocytes. *Circulation* **107**, 1912-6 (2003).
79. Chen, S., Zhang, Q., Wu, X., Schultz, P. G. & Ding, S. Dedifferentiation of lineage-committed cells by a small molecule. *J Am Chem Soc* **126**, 410-1 (2004).
80. Chen, S. et al. Self-renewal of embryonic stem cells by a small molecule. *Proc Natl Acad Sci U S A* **103**, 17266-71 (2006).
81. Chen, S., Hilcove, S. & Ding, S. Exploring stem cell biology with small molecules. *Mol Biosyst* **2**, 18-24 (2006).
82. Sachinidis, A. et al. Identification of small signalling molecules promoting cardiac-specific differentiation of mouse embryonic stem cells. *Cell Physiol Biochem* **18**, 303-14 (2006).
83. Chen, S. et al. Reversine increases the plasticity of lineage-committed mammalian cells. *Proc Natl Acad Sci U S A* **104**, 10482-7 (2007).
84. Desbordes, S. C. et al. High-throughput screening assay for the identification of compounds regulating self-renewal and differentiation in human embryonic stem cells. *Cell Stem Cell* **2**, 602-12 (2008).
85. Schreiber, S. L. Small molecules: the missing link in the central dogma. *Nat Chem Biol* **1**, 64-6 (2005).
86. Mol, C. D. et al. Structure of a c-kit product complex reveals the basis for kinase transactivation. *J Biol Chem* **278**, 31461-4 (2003).
87. Wu, X., Ding, S., Ding, Q., Gray, N. S. & Schultz, P. G. Small molecules that induce cardiomyogenesis in embryonic stem cells. *J Am Chem Soc* **126**, 1590-1 (2004).
88. Suter, D. M. & Krause, K. H. Neural commitment of embryonic stem cells: molecules, pathways and potential for cell therapy. *J Pathol* **215**, 355-68 (2008).
89. Ohnuma, T., Augood, S. J., Arai, H., McKenna, P. J. & Emson, P. C. Expression of the human excitatory amino acid transporter 2 and metabotropic glutamate receptors 3 and 5 in the prefrontal cortex from normal individuals and patients with schizophrenia. *Brain Res Mol Brain Res* **56**, 207-17 (1998).
90. Malherbe, P. et al. Mutational analysis and molecular modeling of the binding pocket of the metabotropic glutamate 5 receptor negative modulator 2-methyl-6-(phenylethynyl)-pyridine. *Mol Pharmacol* **64**, 823-32 (2003).
91. Hampson, D. R. et al. Probing the ligand-binding domain of the mGluR4 subtype of metabotropic glutamate receptor. *J Biol Chem* **274**, 33488-95 (1999).
92. Bessis, A. S. et al. Three-dimensional model of the extracellular domain of the type 4a metabotropic glutamate receptor: new insights into the activation process. *Protein Sci* **9**, 2200-9 (2000).
93. Kunishima, N. et al. Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor. *Nature* **407**, 971-7 (2000).
94. Rondard, P. et al. Coupling of agonist binding to effector domain activation in metabotropic glutamate-like receptors. *J Biol Chem* **281**, 24653-61 (2006).
95. Gerber, U., Gee, C. E. & Benquet, P. Metabotropic glutamate receptors: intracellular signaling pathways. *Curr Opin Pharmacol* **7**, 56-61 (2007).
96. Kew, J. N. & Kemp, J. A. Ionotropic and metabotropic glutamate receptor structure and pharmacology. *Psychopharmacology (Berl)* **179**, 4-29 (2005).
97. Cappuccio, I. et al. Endogenous activation of mGlu5 metabotropic glutamate receptors supports self-renewal of cultured mouse embryonic stem cells. *Neuropharmacology* **49 Suppl 1**, 196-205 (2005).
98. Cappuccio, I. et al. Context-dependent regulation of embryonic stem cell differentiation by mGlu4 metabotropic glutamate receptors. *Neuropharmacology* **51**, 606-11 (2006).
99. Melchiorri, D. et al. Metabotropic glutamate receptors in stem/progenitor cells. *Neuropharmacology* **53**, 473-80 (2007).

100. Sarichelou, I. et al. Metabotropic glutamate receptors regulate differentiation of embryonic stem cells into GABAergic neurons. *Cell Death Differ* **15**, 700-7 (2008).
101. O'Leary, D. M., Movsesyan, V., Vicini, S. & Faden, A. I. Selective mGluR5 antagonists MPEP and SIB-1893 decrease NMDA or glutamate-mediated neuronal toxicity through actions that reflect NMDA receptor antagonism. *Br J Pharmacol* **131**, 1429-37 (2000).
102. Mathiesen, J. M., Svendsen, N., Brauner-Osborne, H., Thomsen, C. & Ramirez, M. T. Positive allosteric modulation of the human metabotropic glutamate receptor 4 (hmGluR4) by SIB-1893 and MPEP. *Br J Pharmacol* **138**, 1026-30 (2003).
103. Simonyi, A., Schachtman, T. R. & Christoffersen, G. R. The role of metabotropic glutamate receptor 5 in learning and memory processes. *Drug News Perspect* **18**, 353-61 (2005).
104. Movsesyan, V. A. et al. mGluR5 antagonists 2-methyl-6-(phenylethynyl)-pyridine and (E)-2-methyl-6-(2-phenylethenyl)-pyridine reduce traumatic neuronal injury in vitro and in vivo by antagonizing N-methyl-D-aspartate receptors. *J Pharmacol Exp Ther* **296**, 41-7 (2001).
105. Lea, P. M. t., Movsesyan, V. A. & Faden, A. I. Neuroprotective activity of the mGluR5 antagonists MPEP and MTEP against acute excitotoxicity differs and does not reflect actions at mGluR5 receptors. *Br J Pharmacol* **145**, 527-34 (2005).
106. Varney, M. A. et al. SIB-1757 and SIB-1893: selective, noncompetitive antagonists of metabotropic glutamate receptor type 5. *J Pharmacol Exp Ther* **290**, 170-81 (1999).
107. Bain, J., McLauchlan, H., Elliott, M. & Cohen, P. The specificities of protein kinase inhibitors: an update. *Biochem J* **371**, 199-204 (2003).
108. Bain, J. et al. The selectivity of protein kinase inhibitors: a further update. *Biochem J* **408**, 297-315 (2007).
109. Gunby, R. H. et al. An enzyme-linked immunosorbent assay to screen for inhibitors of the oncogenic anaplastic lymphoma kinase. *Haematologica* **90**, 988-90 (2005).
110. Carlson, H. A. Protein flexibility is an important component of structure-based drug discovery. *Curr Pharm Des* **8**, 1571-8 (2002).
111. Brzozowski, A. M. et al. Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* **389**, 753-8 (1997).
112. Shiau, A. K. et al. The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* **95**, 927-37 (1998).
113. Kokh, D. B. & Wenzel, W. Flexible side chain models improve enrichment rates in in silico screening. *J Med Chem* **51**, 5919-31 (2008).
114. Pearlman, D. A. & Charifson, P. S. Improved scoring of ligand-protein interactions using OWFEG free energy grids. *J Med Chem* **44**, 502-11 (2001).
115. Ewing, T. J., Makino, S., Skillman, A. G. & Kuntz, I. D. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J Comput Aided Mol Des* **15**, 411-28 (2001).
116. Vieth, M., Hirst, J. D., Dominy, B. N., Daigler, H. & Brooks, C. L., 3rd. Assessing Search Strategies for Flexible Docking. *J Comput Chem* **19**, 1623-1631 (1998).
117. Verkhivker, G. M. et al. Deciphering common failures in molecular docking of ligand-protein complexes. *J Comput Aided Mol Des* **14**, 731-51 (2000).
118. Cole, J. C., Murray, C. W., Nissink, J. W., Taylor, R. D. & Taylor, R. Comparing protein-ligand docking programs is difficult. *Proteins* **60**, 325-32 (2005).
119. Abagyan, R. A. & Totrov, M. M. Contact area difference (CAD): a robust measure to evaluate accuracy of protein models. *J Mol Biol* **268**, 678-85 (1997).
120. Marcou, G. & Rognan, D. Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J Chem Inf Model* **47**, 195-207 (2007).
121. Brooijmans, N. & Kuntz, I. D. Molecular recognition and docking algorithms. *Annu Rev Biophys Biomol Struct* **32**, 335-73 (2003).
122. Clark, R. D., Strizhev, A., Leonard, J. M., Blake, J. F. & Matthew, J. B. Consensus scoring for ligand/protein interactions. *J Mol Graph Model* **20**, 281-95 (2002).
123. Graves, A. P., Brenk, R. & Shoichet, B. K. Decoys for docking. *J Med Chem* **48**, 3714-28 (2005).
124. Crichton, N. INFORMATION POINT: Receiver Operating Characteristic (ROC) curves. *J Clin Nursing* **11**, 136 (2002).
125. Triballeau, N., Acher, F., Brabet, I., Pin, J. P. & Bertrand, H. O. Virtual screening workflow development guided by the "receiver operating characteristic" curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. *J Med Chem* **48**, 2534-47 (2005).

126. Teramoto, R. & Fukunishi, H. Supervised scoring models with docked ligand conformations for structure-based virtual screening. *J Chem Inf Model* **47**, 1858-67 (2007).
127. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912-34 (2002).
128. Cheney, I. W. et al. Identification and structure-activity relationships of substituted pyridones as inhibitors of Pim-1 kinase. *Bioorg Med Chem Lett* **17**, 1679-83 (2007).
129. Jacobs, M. D. et al. Pim-1 ligand-bound structures reveal the mechanism of serine/threonine kinase inhibition by LY294002. *J Biol Chem* **280**, 13728-34 (2005).
130. Aksoy, I. et al. Self-renewal of murine embryonic stem cells is supported by the serine/threonine kinases Pim-1 and Pim-3. *Stem Cells* **25**, 2996-3004 (2007).
131. Hammerman, P. S., Fox, C. J., Birnbaum, M. J. & Thompson, C. B. Pim and Akt oncogenes are independent regulators of hematopoietic cell growth and survival. *Blood* **105**, 4477-83 (2005).
132. Koike, N., Maita, H., Taira, T., Ariga, H. & Iguchi-Ariga, S. M. Identification of heterochromatin protein 1 (HP1) as a phosphorylation target by Pim-1 kinase and the effect of phosphorylation on the transcriptional repression function of HP1(1). *FEBS Lett* **467**, 17-21 (2000).
133. Konietzko, U. et al. Pim kinase expression is induced by LTP stimulation and required for the consolidation of enduring LTP. *Embo J* **18**, 3359-69 (1999).
134. Takagaki, K. et al. Role of Chk1 and Chk2 in Ara-C-induced differentiation of human leukemia K562 cells. *Genes Cells* **10**, 97-106 (2005).
135. Sorensen, C. S. et al. Chk1 regulates the S phase checkpoint by coupling the physiological turnover and ionizing radiation-induced accelerated proteolysis of Cdc25A. *Cancer Cell* **3**, 247-58 (2003).
136. Sorensen, C. S. et al. The cell-cycle checkpoint kinase Chk1 is required for mammalian homologous recombination repair. *Nat Cell Biol* **7**, 195-201 (2005).
137. Shieh, S. Y., Ahn, J., Tamai, K., Taya, Y. & Prives, C. The human homologs of checkpoint kinases Chk1 and Cds1 (Chk2) phosphorylate p53 at multiple DNA damage-inducible sites. *Genes Dev* **14**, 289-300 (2000).
138. Ou, Y. H., Chung, P. H., Sun, T. P. & Shieh, S. Y. p53 C-terminal phosphorylation by CHK1 and CHK2 participates in the regulation of DNA-damage-induced C-terminal acetylation. *Mol Biol Cell* **16**, 1684-95 (2005).
139. Fraley, M. E. et al. 3-(Indol-2-yl)indazoles as Chk1 kinase inhibitors: Optimization of potency and selectivity via substitution at C6. *Bioorg Med Chem Lett* **16**, 6049-53 (2006).
140. Feijoo, C. et al. Activation of mammalian Chk1 during DNA replication arrest: a role for Chk1 in the intra-S phase checkpoint monitoring replication origin firing. *J Cell Biol* **154**, 913-23 (2001).
141. Xiao, Z. et al. Chk1 mediates S and G2 arrests through Cdc25A degradation in response to DNA-damaging agents. *J Biol Chem* **278**, 21767-73 (2003).
142. Chen, M. S., Ryan, C. E. & Piwnicka-Worms, H. Chk1 kinase negatively regulates mitotic function of Cdc25A phosphatase through 14-3-3 binding. *Mol Cell Biol* **23**, 7488-97 (2003).
143. Zhang, Y. et al. The Chk1/Cdc25A pathway as activators of the cell cycle in neuronal death induced by camptothecin. *J Neurosci* **26**, 8819-28 (2006).
144. Schindler, T. et al. Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science* **289**, 1938-42 (2000).
145. Adrian, F. J. et al. Allosteric inhibitors of Bcr-abl-dependent cell proliferation. *Nat Chem Biol* **2**, 95-102 (2006).
146. Choi, Y. et al. N-myristoylated c-Abl tyrosine kinase localizes to the endoplasmic reticulum upon binding to an allosteric inhibitor. *J Biol Chem* **284**, 29005-14 (2009).
147. Ohren, J. F. & Sebolt-Leopold, J. S. Inhibitors of Bcr-abl... breaking new ground again. *Nat Chem Biol* **2**, 63-4 (2006).
148. Fabbro, D. et al. Inhibitors of the Abl kinase directed at either the ATP- or myristate-binding site. *Biochim Biophys Acta* **1804**, 454-62 (2010).
149. Fogerty, F. J. et al. Dominant effects of the bcr-abl oncogene on Drosophila morphogenesis. *Oncogene* **18**, 219-32 (1999).
150. Bennett, R. L. & Hoffmann, F. M. Increased levels of the Drosophila Abelson tyrosine kinase in nerves and muscles: subcellular localization and mutant phenotypes imply a role in cell-cell interactions. *Development* **116**, 953-66 (1992).
151. Loureiro, J. & Peifer, M. Roles of Armadillo, a Drosophila catenin, during central nervous system development. *Curr Biol* **8**, 622-32 (1998).

152. Taniguchi, Y., London, R., Schinkmann, K., Jiang, S. & Avraham, H. The receptor protein tyrosine phosphatase, PTP-RO, is upregulated during megakaryocyte differentiation and is associated with the c-Kit receptor. *Blood* **94**, 539-49 (1999).
153. Kingham, E. & Welham, M. Distinct roles for isoforms of the catalytic subunit of class-IA PI3K in the regulation of behaviour of murine embryonic stem cells. *J Cell Sci* **122**, 2311-21 (2009).
154. Hill, J. W. et al. Phosphatidyl inositol 3-kinase signaling in hypothalamic proopiomelanocortin neurons contributes to the regulation of glucose homeostasis. *Endocrinology* **150**, 4874-82 (2009).
155. Chen, S. et al. Alterations in mammalian target of rapamycin signaling pathways after traumatic brain injury. *J Cereb Blood Flow Metab* **27**, 939-49 (2007).
156. Knauf, U., Tschopp, C. & Gram, H. Negative regulation of protein translation by mitogen-activated protein kinase-interacting kinases 1 and 2. *Mol Cell Biol* **21**, 5500-11 (2001).
157. Okamura, A. et al. Involvement of casein kinase I epsilon in cytokine-induced granulocytic differentiation. *Blood* **103**, 2997-3004 (2004).
158. Keesler, G. A. et al. Phosphorylation and destabilization of human period 1 clock protein by human casein kinase I epsilon. *Neuroreport* **11**, 951-5 (2000).
159. Walton, K. M. et al. Selective inhibition of casein kinase 1 epsilon minimally alters circadian clock period. *J Pharmacol Exp Ther* **330**, 430-9 (2009).
160. Hino, S., Michiue, T., Asashima, M. & Kikuchi, A. Casein kinase I epsilon enhances the binding of Dvl-1 to Frat-1 and is essential for Wnt-3a-induced accumulation of beta-catenin. *J Biol Chem* **278**, 14066-73 (2003).
161. Teran, E., Branscomb, A. D. & Seeling, J. M. Dpr Acts as a molecular switch, inhibiting Wnt signaling when unphosphorylated, but promoting Wnt signaling when phosphorylated by casein kinase I delta/epsilon. *PLoS One* **4**, e5522 (2009).
162. Sakurai, E. et al. Upregulation of casein kinase 1 epsilon in dorsal root ganglia and spinal cord after mouse spinal nerve injury contributes to neuropathic pain. *Mol Pain* **5**, 74 (2009).
163. Sperling, L. E. et al. Characterisation of cholinesterase expression during murine embryonic stem cell differentiation. *Chem Biol Interact* **175**, 156-60 (2008).
164. Yang, L., He, H. Y. & Zhang, X. J. Increased expression of intranuclear AChE involved in apoptosis of SK-N-SH cells. *Neurosci Res* **42**, 261-8 (2002).
165. Malemud, C. J. Inhibitors of stress-activated protein/mitogen-activated protein kinase pathways. *Curr Opin Pharmacol* **7**, 339-43 (2007).
166. Snyder, K. A., Feldman, H. J., Dumontier, M., Salama, J. J. & Hogue, C. W. Domain-based small molecule binding site annotation. *BMC Bioinformatics* **7**, 152 (2006).
167. Charifson, P. S., Corkery, J. J., Murcko, M. A. & Walters, W. P. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J Med Chem* **42**, 5100-9 (1999).
168. Xing, L., Hodgkin, E., Liu, Q. & Sedlock, D. Evaluation and application of multiple scoring functions for a virtual screening experiment. *J Comput Aided Mol Des* **18**, 333-44 (2004).
169. Mpamhanga, C. P., Chen, B., McLay, I. M. & Willett, P. Knowledge-based interaction fingerprint scoring: a simple method for improving the effectiveness of fast scoring functions. *J Chem Inf Model* **46**, 686-98 (2006).
170. Kleywegt, G. J. Recognition of spatial motifs in protein structures. *J Mol Biol* **285**, 1887-97 (1999).
171. Hulo, N. et al. The PROSITE database. *Nucleic Acids Res* **34**, D227-30 (2006).
172. Stark, A., Sunyaev, S. & Russell, R. B. A model for statistical significance of local similarities in structure. *J Mol Biol* **326**, 1307-16 (2003).
173. Lin, R. e. a. 3-Acyl-2,6-diaminopyridines as cyclin-dependent kinase inhibitors: synthesis and biological evaluation. *Bioorg Med Chem Lett* **15**, 2221-2224 (2005).
174. Lipinski, C. & Hopkins, A. Navigating chemical space for biology and medicine. *Nature* **432**, 855-61 (2004).
175. Binetruy, B., Heasley, L., Bost, F., Caron, L. & Aouadi, M. Concise review: regulation of embryonic stem cell lineage commitment by mitogen-activated protein kinases. *Stem Cells* **25**, 1090-5 (2007).
176. Schindler, J. F., Monahan, J. B. & Smith, W. G. p38 pathway kinases as anti-inflammatory drug targets. *J Dent Res* **86**, 800-11 (2007).
177. Bubici, C., Papa, S., Pham, C. G., Zazzeroni, F. & Franzoso, G. The NF-kappaB-mediated control of ROS and JNK signaling. *Histol Histopathol* **21**, 69-80 (2006).

178. Cai, J. et al. Peptide deformylase is a potential target for anti-*Helicobacter pylori* drugs: reverse docking, enzymatic assay, and X-ray crystallography validation. *Protein Sci* **15**, 2071-81 (2006).
179. O'Connor, K. A. & Roth, B. L. Finding new tricks for old drugs: an efficient route for public-sector drug discovery. *Nat Rev Drug Discov* **4**, 1005-14 (2005).
180. Chong, C. R. & Sullivan, D. J., Jr. New uses for old drugs. *Nature* **448**, 645-6 (2007).

CHAPTER 5 - CONCLUSIONS

Different computational and cheminformatics approaches were used in this thesis to tackle issues related to binding affinity prediction, protein-protein interactions, drug design and target finding.

In chapter 1, differentiating μM substrate affinities for thymidine kinases could be achieved using steered molecular dynamics.

Combining homology modelling, docking and shape-based similarity searches in chapter 2, AMP was identified among other small molecules as a potential aggregation breaker for monoclonal antibodies of the IgG1 type, a finding that could be validated experimentally.

With the aim of generating new ideas on small molecule- or peptide-based inhibitors of the NADPH oxidase 2, a functionally important protein-protein interface was addressed in chapter 3 by virtual screening, a binding orientation for an experimental hit was predicted, and peptidomimetics were designed based on molecular interaction fields.

In chapter 4, inverse virtual screening was successfully applied to the identification of new targets for a small molecule.

However, given all these different projects, are the herein used methods really related to “*in silico* pharmacology” as stated in the main title of this thesis (adapted from ¹ and ²)?

Provided they are all connected to understanding molecular recognition by establishing structure-activity relationships and putative mechanisms of action, as in classical *in vitro/in vivo* pharmacology, they are *in silico* pharmacology approaches. The work performed in this thesis corresponds to the definition of *in silico* pharmacology given by Elkins et al. ², stating that this field includes finding new antagonists or agonists for a target in the absence or presence of a target, identifying the biological pathways, and determining the selectivity of drugs when depicted in a network of targets. This thesis provided hypotheses for experimental verifications. These verifications were in most cases carried out and were found to support the molecular modelling findings.

Known disadvantages of such *in silico* approaches concern e.g. (i) the often insufficient sampling of the flexibility of proteins, (ii) the omission of all but the most prominent small molecule tautomers as well as protomers, and (iii) promiscuous binders (e.g. methotrexate in chapter 4). In spite of these limitations, valuable hypotheses were generated by molecular modelling.

Iterating molecular modelling and experimental validation has proven to be beneficial for the outcome of each chapter and should generally be applied as often as possible, also to improve *in silico* pharmacology *per se*.

References

1. Elkins, S., Mestres, J. & Testa, B. In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. *Br J Pharmacol* **152**, 9-20 (2007).
2. Elkins, S., Mestres, J. & Testa, B. In silico pharmacology for drug discovery: applications to targets and beyond. *Br J Pharmacol* **152**, 21-37 (2007).

ACKNOWLEDGEMENTS

I wish to express my gratitude to my supervisor Prof. Dr. Leonardo Scapozza, who allowed me to get insight into many different fields of computer-aided drug design. He has been present as a critical mind questioning my scientific understanding and encouraging me to actively participate in scientific meetings.

The jury members are acknowledged for having accepted to read and evaluate my thesis.

The collaborations are acknowledged at the end of each chapter.

“Merci vielmol” Shaheen for supervising my first steps in the astonishing world of molecular modelling, your friendship, and the SuperTux and PacMan sessions. Thank you, Anja, Pernilla, and Lucile for the collaboration on thymidine kinases and the good time spent together. Agata, thank you very much for the nice lunch breaks and for sharing your knowledge. Thank you, Andrea, Loris, and Fabian for your valuable help in cluster-related issues.

Many thanks go to the present and former group members of the Scapozza group for the group retreats with tons of pasta, and their support during my PhD thesis: Agata, Alessandra, Andrea B., Andrea C., Anna, Anja, Carole, Enrico, Gaby, Ivana, Majdeline, Marieke, Leonardo L., Loris, Lucile, Ralitz, Remo, Roberta, Patricia, Pernilla, Sabine, Sema, Shaheen, Surekha, Thomas, William, and all the internship and diploma students, especially Denis. Thank you also to Sachin and Jing for having been my office mates.

I acknowledge Serge Darbellay for having hosted me in his family and for being more stubborn than myself when it comes to physiotherapy, Sónia Cardoso for the nice accommodation in Carouge, and especially Felice and Ann Lanfranchi-Collier as well as Eva for a memorable writing time on their veranda in Bernex.

A huge thank you goes to Iris, Droll, Claudia and Kay, Pernilla and Simon (+ Emma), Agata and Fabio, Anja and Edgar, Antonio and Catia, Abbas, Shahram and Heidy, Vernonika and Anton, Annette and Christoph, Alessia and Luca, Hye and Markus, Markus and Dominik, Thuy, Éric, and the Navarre family, for their friendship and the weekends and holidays all over Europe. A hug to Virginie for great holidays in Charente-Maritime and Odile for snowshoe walking and her home-made “tartes Tatin” in Savoie.

Last but not least, I would like to thank my my parents, my sister Flavia, and Christophe, for their support, motivation, and love. I am incredibly grateful to my parents who allowed me to study whatever I liked and to Flavia for bringing me back to the non-scientific reality from time to time. Thank you, Christophe, for your unbreakable patience, for having explained and re-explained principal component analysis and Excel macros, and for all the time we spent together.

And to all whom I have forgotten to mention...

ABBREVIATIONS USED IN THIS THESIS

Abbreviation	Description
[¹⁸ F]FLT	[¹⁸ F]-fluoro-3'-deoxy-3'-fluorothymidine
¹ H- ¹⁵ N HSQC NMR	Heteronuclear single quantum correlation nuclear magnetic resonance, indicating the backbone N-H bonds
3D	3-dimensional
ACE	Angiotensin converting enzyme
AChE	Acetylcholinesterase
AdK	Adenylate kinase
ADMET	Adsorption, distribution, metabolism, excretion, toxicity
ADP	Adenosine 5'-diphosphate
AFFF-MALS	Asymmetric field flow fractionation coupled to multi-angle light scattering
AFM	Atomic force microscopy
AIB	Aminobutyric acid
AKT2	A Ser/Thr protein kinase
Ala	Alanine (A): Nonpolar side chain, neutral
alpha-APA	Loviridine
AMBER	Assisted model building and energy refinement
AMP	Adenosine 5'-monophosphate
Arg	Arginine (R): Polar side chain, positive charge
Asn	Asparagine (N): Polar side chain, neutral
Asp	Aspartic acid (D): Polar side chain, negative charge
ATP	Adenosine 5'-triphosphate
AZT	3'-azido-3'-deoxythymidine
Bcl-X _L	B-cell lymphoma extra large
B-factor	A measure of the flexibility of a crystallographically resolved amino acid
BindingMOAD	Binding mother of all databases
Blast	Basic local alignment search tool
BRENDA	Braunschweig Enzyme Database
BTN	Biotin
c-Abl	Cellular Abelson murine leukemia viral oncogene homologue 1, a tyrosine kinase
Cal	Calorie, unit of energy, 1 cal = 4.2 J
CAPRI	Critical assessment of predictions of interactions
CD	Circular Dichroism
Cdc37	Co-chaperone of HSP90
CDK2	Cyclin dependent kinase 2
cDNA	Complementary DNA
CDR1 to 3	Complementarity determining regions 1 to 3
CGD	Chronic granulomatous disease
CGR8	A mouse embryonic stem cell line
CH1, 2 and 3	Constant domains 1, 2 and 3 of an antibody's heavy chain
CHK1	Checkpoint kinase 1
CK2	Casein kinase 2, a Ser/Thr kinase
c-kit	Stem cell factor receptor, a receptor tyrosine kinase
CL	Constant domain of an antibody's light chain
class I SH3	Src Homology 3 domain binding in a N- to C-terminal orientation
class II SH3	Src Homology 3 domain binding in a C- to N-terminal orientation
clogP	Calculated water-octanol partition coefficient
CMD	Classical Molecular Dynamics
CNS	Central nervous system
CoMFA	Comparative molecular field analysis, 3D QSAR
CoMSIA	Comparative molecular similarity index analysis, 3D QSAR
Cys	Cysteine (C): Nonpolar side chain, neutral
Cα	Carbon alpha
Da	Dalton, the atomic mass unit. 1 Da = 1.66 * 10 ⁻²⁴ g
DB	Database
deltaG	Free energy change
deltaH	Enthalpy change

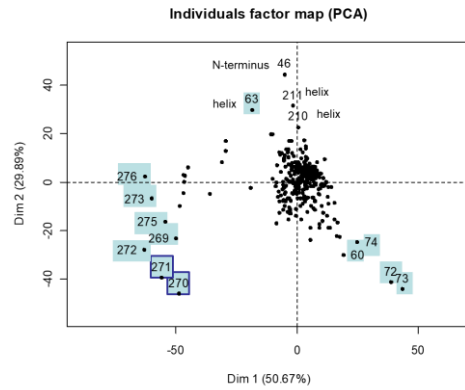
Abbreviation	Description
deltaS	Entropy change
DNA	Deoxyribonucleic acid
DPI	Diphenylene iodonium
DSC	Differential scanning calorimetry
dT	2'-deoxythymidine
dTDP	Deoxythymidine 5'-diphosphate
dTMP	Deoxythymidine 5'-monophosphate
dTTP	Deoxythymidine 5'-triphosphate
DUOX	Dual oxidase
EC	Enzyme classification
EC ₅₀	Half maximal effective concentration of e.g. a drug
ED	Euclidian distance
ELISA	Enzyme-linked immunosorbent assay
eNOS	Endothelial nitric oxide synthase
ERKs	Extracellular-signal-regulated kinases
EVH1 domain	Ena/Vasp homology domain 1
Fab	Fragment antigen binding: VL, VH, CL, CH1
FAD	Flavin adenine dinucleotide
Fasta	A DNA and protein sequence alignment format
Fc	Fragment, crystallisable: CH2, CH3
FDA	Food and Drug Administration (USA)
FEP	Free energy perturbation
fM	Femtomolar
FMN	Flavin mononucleotide
FN	False negative
FNR	Ferredoxin NADPH-reductase
FP	False positive
FR1, 2, 3 and 4	Framework regions 1 to 4 (antibody)
Fv	Fragment variable: FR, HV (antibody)
GAFF	Generalized Amber force field
GCV	Ganciclovir
Gln	Glutamine (Q): Polar side chain, neutral
Glu	Glutamic acid (E): Polar side chain, negative charge
Gly	Glycine (G): Nonpolar side chain, neutral
GOLD	Genetic optimization for ligand docking
GPCR	G protein-coupled receptor
GYF domain	Containing a conserved Gly, Tyr and Phe
H	Heavy chain: VH, CH1, CH2, CH3 (antibody)
Haddock	High ambiguity driven protein-protein docking
H-bond	Hydrogen bond
hdm2	Human negative regulator of the p53 tumor suppressor, also called mdm2
HEK cells	Human embryonic kidney cells
hERG channel	Human ether-à-go-go related gene potassium channel
His	Histidine (H): Polar side chain, mainly neutral
HIV	Human immunodeficiency virus
HPLC	High performance liquid chromatography
HSP90	Heat shock protein 90
HSV-1 TK	Herpes simplex virus 1 thymidine kinase
hTK1	Human Thymidine Kinase 1 (cytosolic)
hTK2	Human Thymidine Kinase 2 (mitochondrial)
HV	Hypervariable region containing CDRs (antibody)
IC ₅₀	Half maximal inhibitory concentration
IFREDA	ICM-flexible receptor docking algorithm
Ig	Immunoglobulin
IgG	Immunoglobulin G
IL-1	Interleukin 1
Ile	Isoleucine (I): Nonpolar side chain, neutral
iNOS	Inducible nitric oxide synthase
IR	Infrared

Abbreviation	Description
ITC	Isothermal titration calorimetry
J	Joule, unit of energy or work, $1 \text{ J} = 1 \text{ kg m}^2/\text{s}^2$
JE	Jarzynski's equality
JNK	c-JUN N-terminal kinase
K_a	Association constant
K_d	Dissociation constant
K_{eq}	Equilibrium constant
K_i	Binding affinity
k_{off}	Off-rate
k_{on}	On-rate
L	Light chain: VL, CL (antibody)
Leu	Leucine (L): Nonpolar side chain, neutral
LID	Flexible domain of HSV-1 TK involved in substrate binding
LIE	Linear interaction energy
Lys	Lysine (K): Polar side chain, positive charge
MAPK	Mitogen-activated protein kinase
MC	Monte Carlo
MD	Molecular dynamics
MEK1	Mitogen-activated protein kinase kinase 1
Met	Methionine (M): Nonpolar side chain, neutral
mGluR	Metabotropic glutamate receptor
MIF	Molecular interaction field
MM-GBSA	Molecular mechanics Generalized Born surface area
MM-PBSA	Molecular mechanics Poisson Boltzmann surface area
MNK2	Mitogen-activated protein kinase-interacting serine/threonine kinase 2
mol	Amount of substance containing as many elementary entities as there are atoms in 12 g of ^{12}C
MS	Mass spectrometry
MTX	Methotrexate
MW	Molecular weight
N	Newton, unit of force, $1 \text{ N} = 1 \text{ kg m s}^{-1}$
n/a	Not answered
ND	Not determined
NF- κ B	Nuclear factor kappa-light-chain-enhancer of activated B cells
NGF	Nerve growth factor
NMA	Normal mode analysis
NMDA	N-methyl-D-aspartic acid
NME	New molecular entity
NMR	Nuclear Magnetic Resonance
NNRTI	Non-nucleoside reverse transcriptase inhibitor
NO	Nitric oxide
NOX	NADPH oxidase
NOXA1	NOX activator protein 1 (homologue of p67 ^{phox} in NOX2)
NOXO1	NOX organizer protein 1 (homologue of p47 ^{phox} in NOX2)
NPT	Stable number of particles, pressure and temperature
OHT	4-OH tamoxifen
p22 ^{phox}	Small constituent of the membrane part of the NOX2 flavocytochrome
p40 ^{phox}	Cytosolic constituent of NOX2
p47 ^{phox}	Cytosolic constituent of NOX2
p53	Tumor suppressor, guardian of genomic integrity
p67 ^{phox}	Cytosolic constituent of NOX2
PAO	Phenylarsine oxide
PAP	Phenazopyridine
PBS	Phosphate buffer saline
PC	Principal component
PCA	Principal component analysis
PDB	Protein data bank
PDB ID	Protein data bank identifier
PDK1	Phosphoinositide-dependent kinase 1

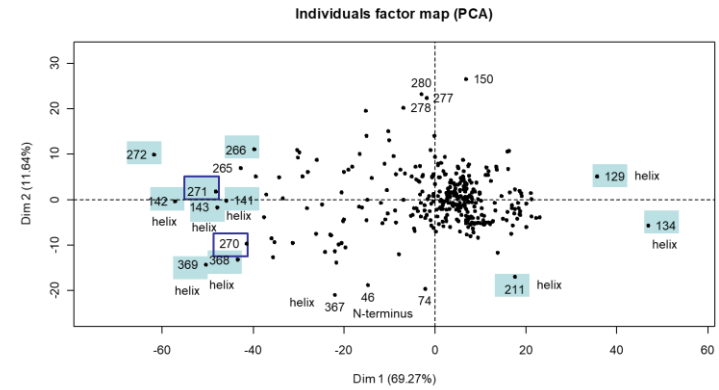
Abbreviation	Description
PET	Positron Emission Tomography
Phe	Phenylalanine (F): Nonpolar side chain, neutral
phox	Phagocyte oxidase
PIK3CG	Phosphatidylinositol-4,5-bisphosphate 3-kinase γ
PKC	Protein kinase C
PME	Particle Mesh Ewald
PMF	Potential of mean force
PPI	Polyproline helix I
PPII	Polyproline helix II
PRH	6-OH-1,6-dihydro purine nucleoside
PRM	Proline-rich motif
Pro	Proline (P): Nonpolar side chain, neutral
ProLINT	Affinity database of kinase and phosphatase ligands
PRR	Proline-rich region
QSAR	Quantitative structure-activity relationship
R	Gas constant, $8.314 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$
RAMD	Random acceleration molecular dynamics
Rap1A	Ras-related protein 1A
Ras	A superfamily of proteins involved in cellular signal transduction
RDD	Rational drug design
RECAP	Retrosynthetic combinatorial analysis procedure
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
RNA	Ribonucleic acid
ROC	Receiver operating curve
ROS	Reactive oxygen species
SAR	Structure-activity relationship
SASA	Solvent-accessible surface area
SCF	Stem cell factor
SCOTCH	Surface complementarity trace in complex history
sc-PDB	(Target) screening subset of the PDB
Ser	Serine (S): Polar side chain, neutral
SH3	Src homology 3 domain
SMD	Steered molecular dynamics
SPR	Surface plasmon resonance
TAT	Transactivator of transcription protein, containing a transfer sequence facilitating the crossing of cell membranes
TC	Tanimoto coefficient
Thr	Threonine (T): Polar side chain, neutral
TI	Thermodynamic integration
TK	Thymidine Kinase
TN	True negative
TNF- α	Tumor necrosis factor α
TP	True positive
Trp	Tryptophane (W): Nonpolar side chain, neutral
Tyr	Tyrosine (Y): Polar side chain, neutral
UEV domain	Ubiquitin E2 variant domain
UV	Ultraviolet
Val	Valine (V): Nonpolar side chain, neutral
VdW	Van der Waals
VEGF	Vascular endothelial growth factor
VH	Variable domain of the heavy chain
VL	Variable domain of the light chain
VS	Virtual screening
WW domain	Containing 2 conserved tryptophanes
X-ray	Röntgen crystallography

ANNEX

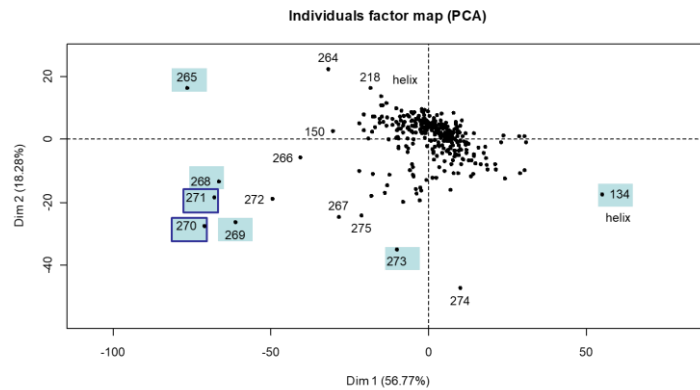
CHAPTER 1: PCA GRAPHS



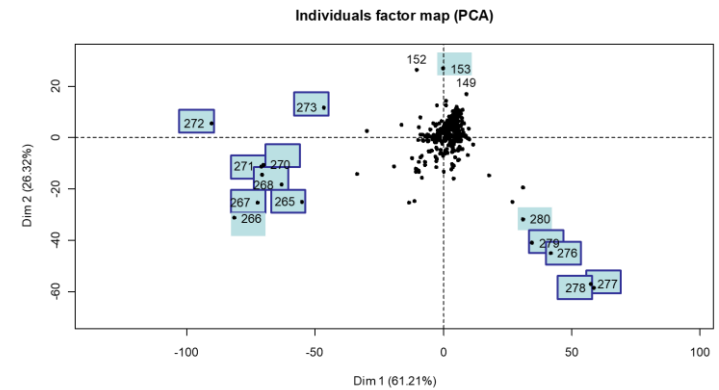
HSV-1 TK dT x, run 1



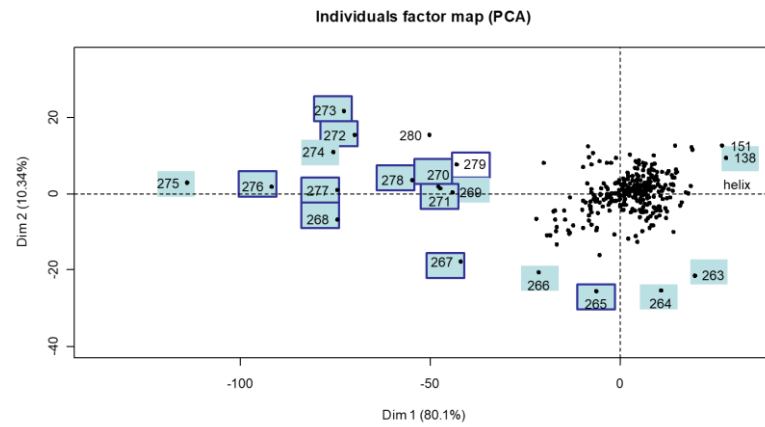
HSV-1 TK dT x, run 2



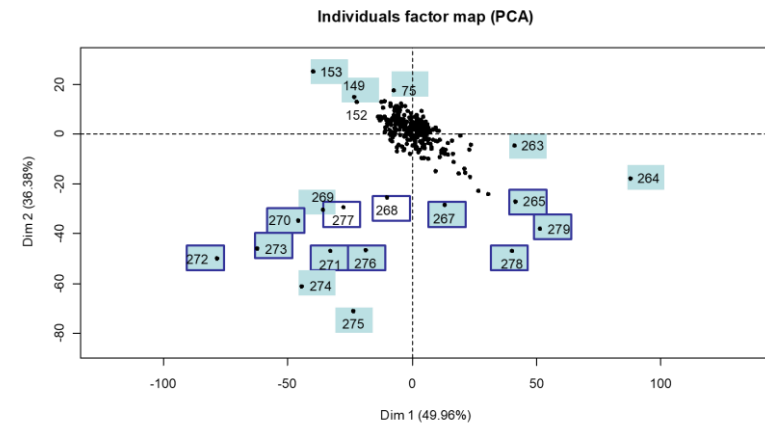
HSV-1 TK dT x, run 3



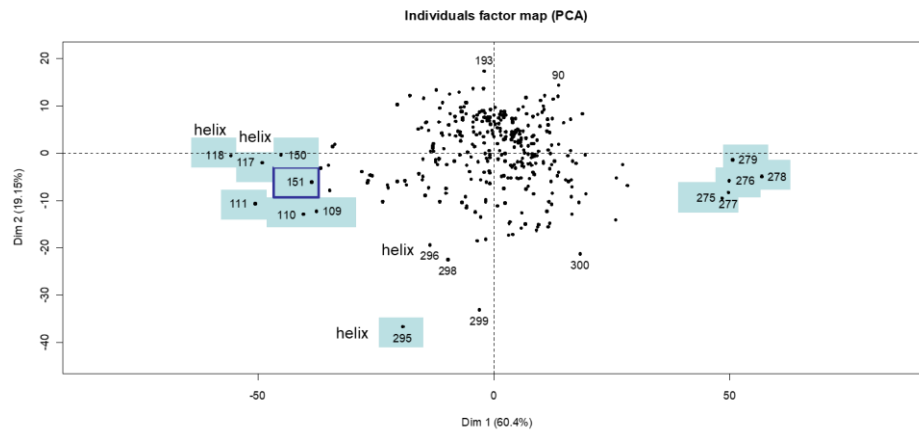
HSV-1 TK DHBT x, run 1



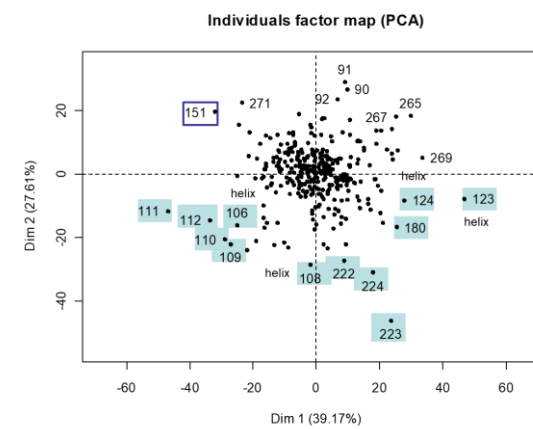
HSV-1 TK DHBT x, run 2



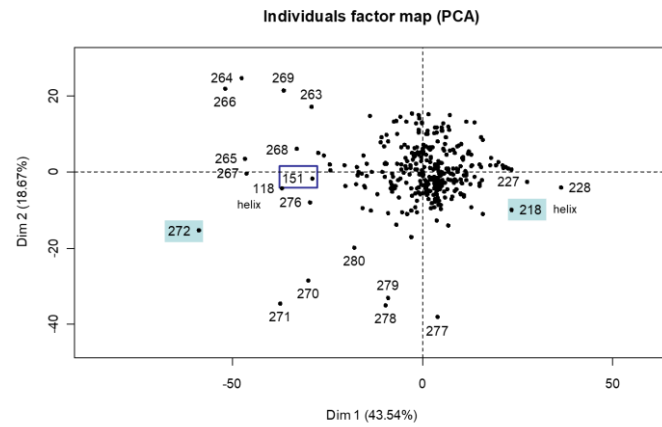
HSV-1 TK DHBT x, run 3



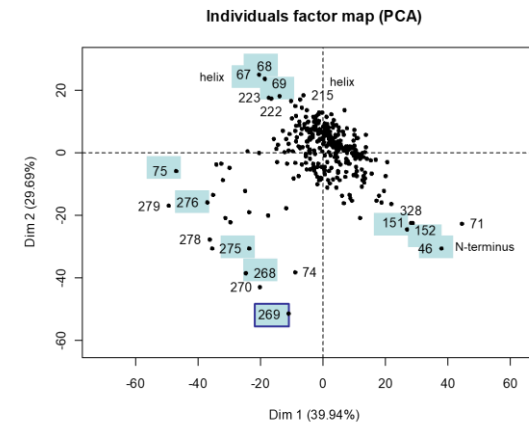
HSV-1 TK NMeDHBT x, run 1



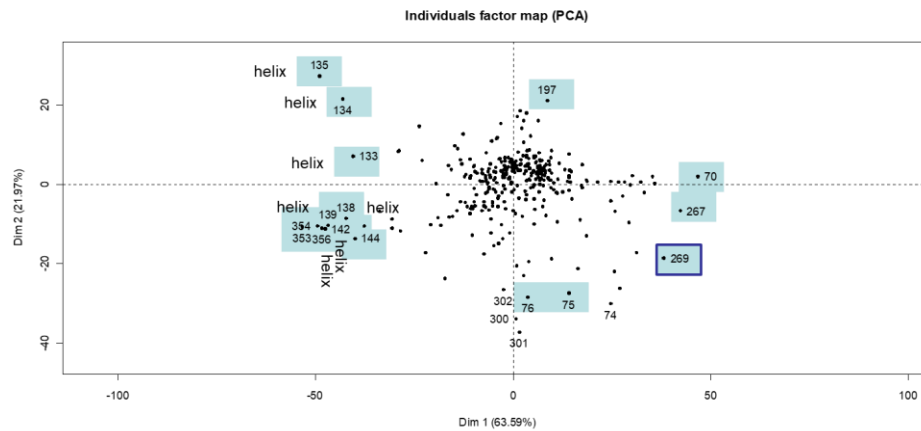
HSV-1 TK NMeDHBT x, run 2



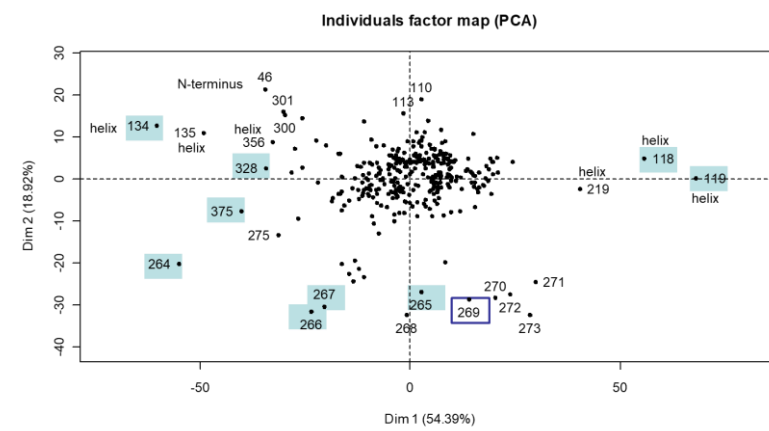
HSV-1 TK NMeDHT x, run 3



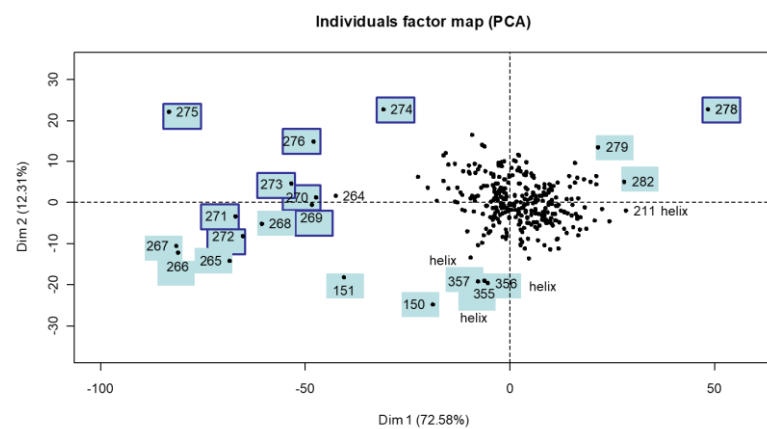
HSV-1 TK dT y, run 1



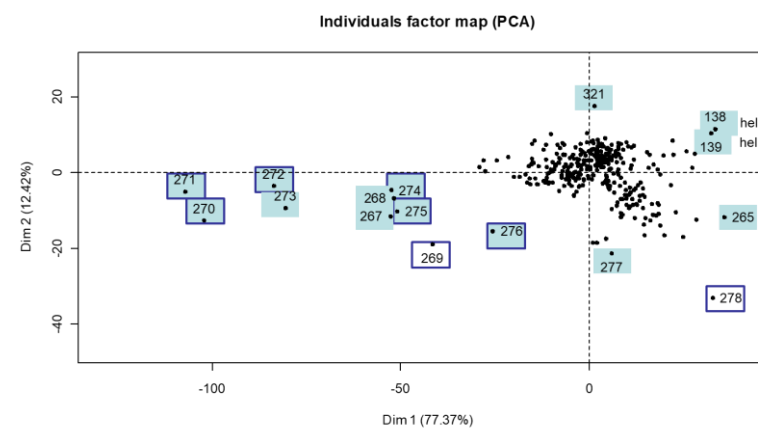
HSV-1 TK dT y, run 2



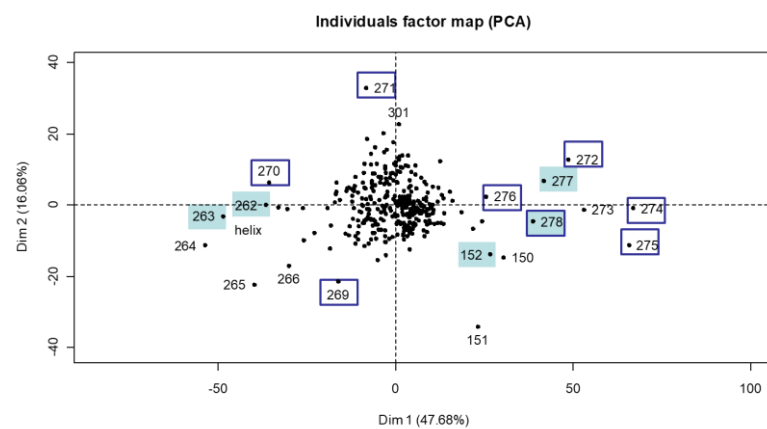
HSV-1 TK dT y, run 3



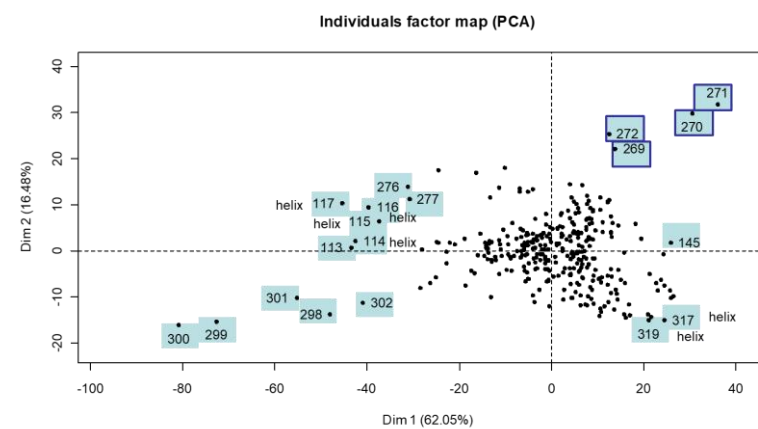
HSV-1 TK DHBT y, run 1



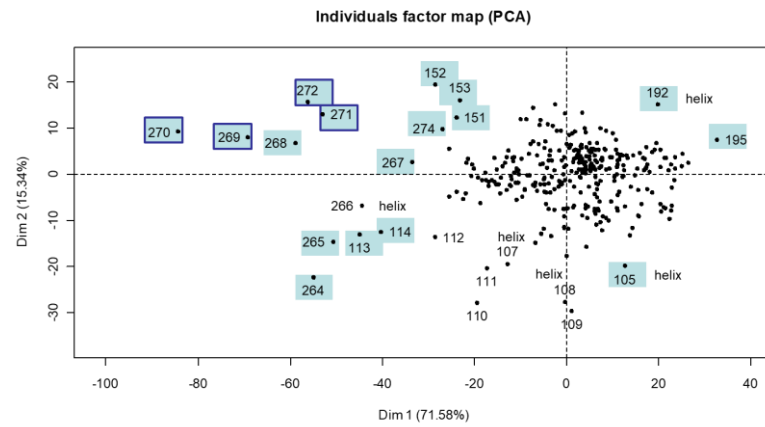
HSV-1 TK DHBT y, run 2



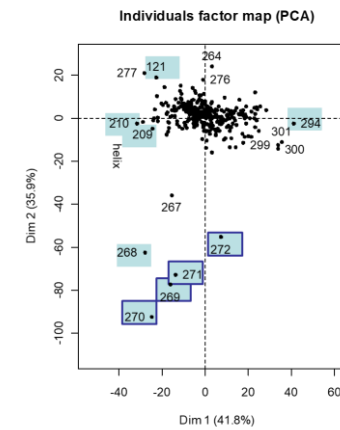
HSV-1 TK DHBT y, run 3



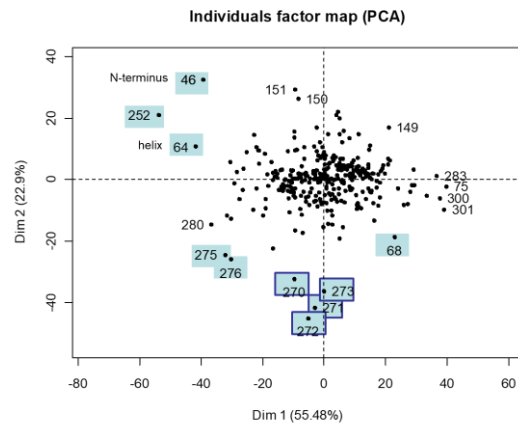
HSV-1 TK NMedHBT y, run 1



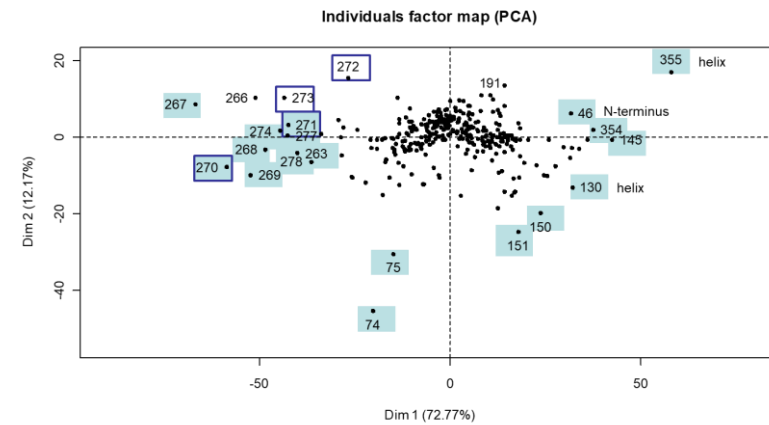
HSV-1 TK NMeDHBT y, run 2



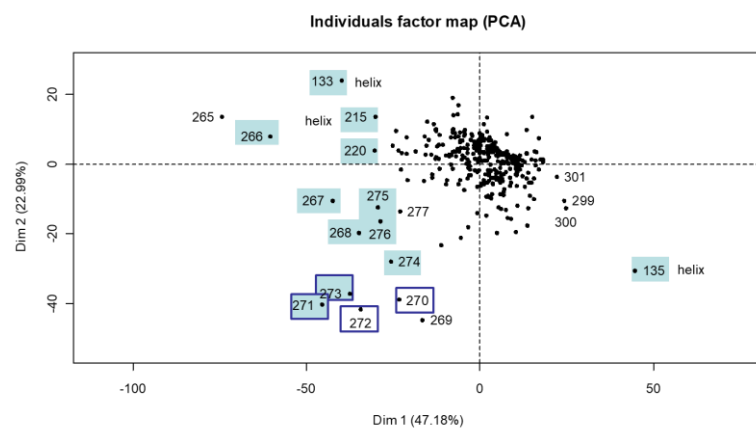
HSV-1 TK NMeDHBT y, run 3



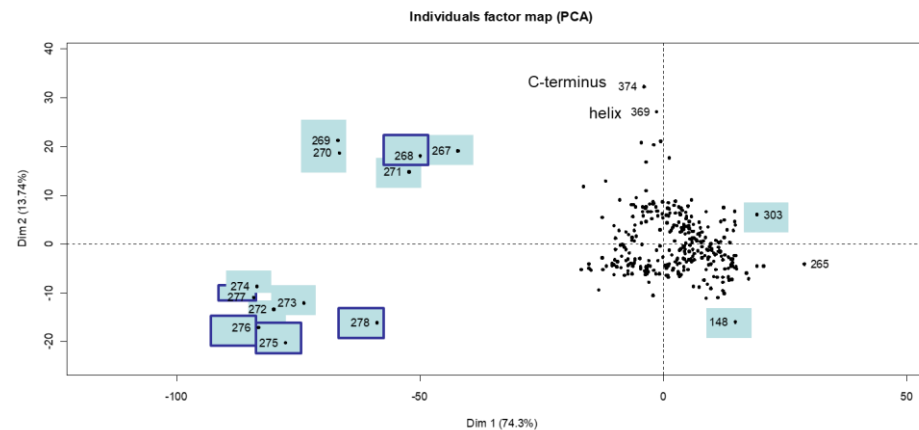
HSV-1 TK dT z, run 1



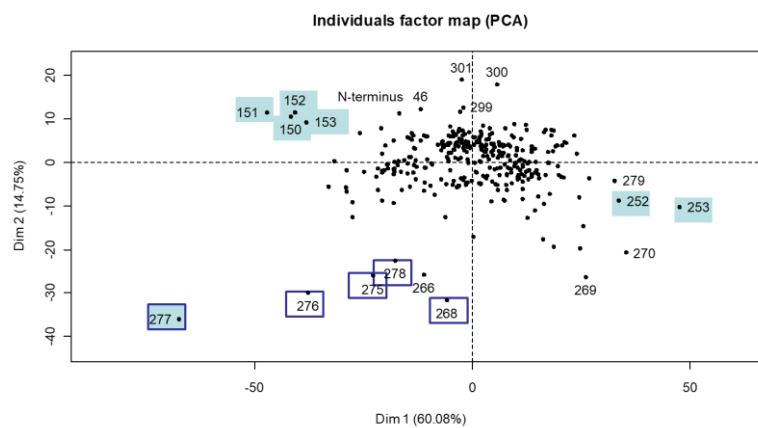
HSV-1 TK dT z, run 2



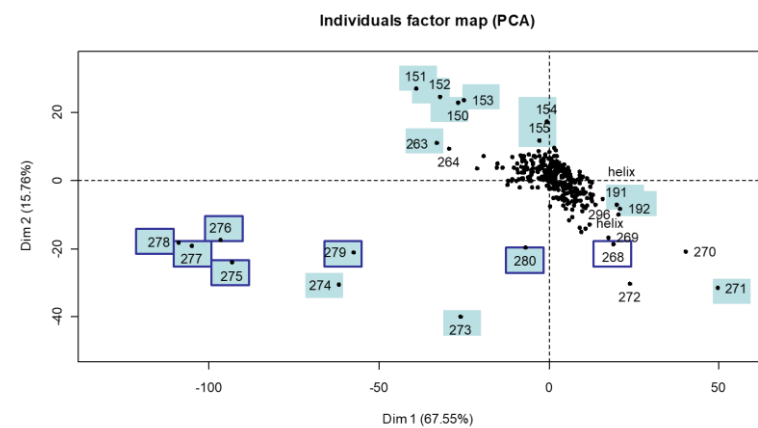
HSV-1 TK dT z, run 3



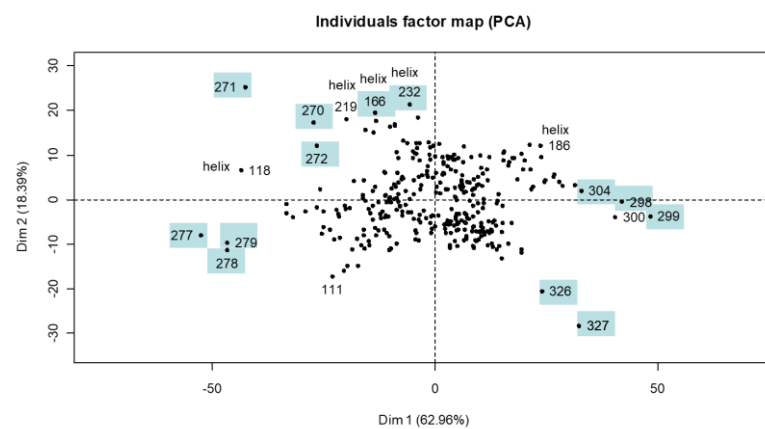
HSV-1 TK DHBT z, run 1



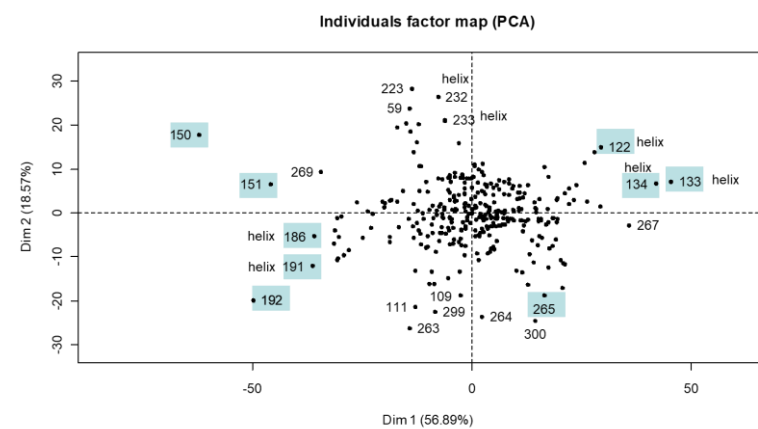
HSV-1 TK DHBT z, run 2



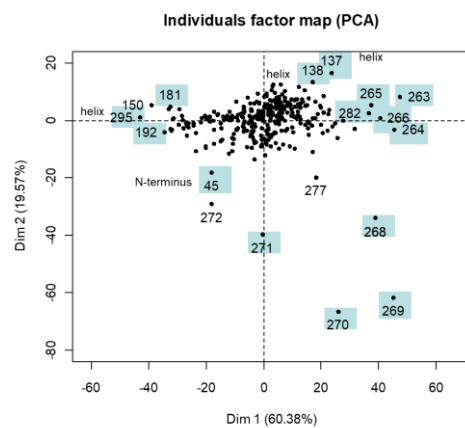
HSV-1 TK DHBT z, run 3



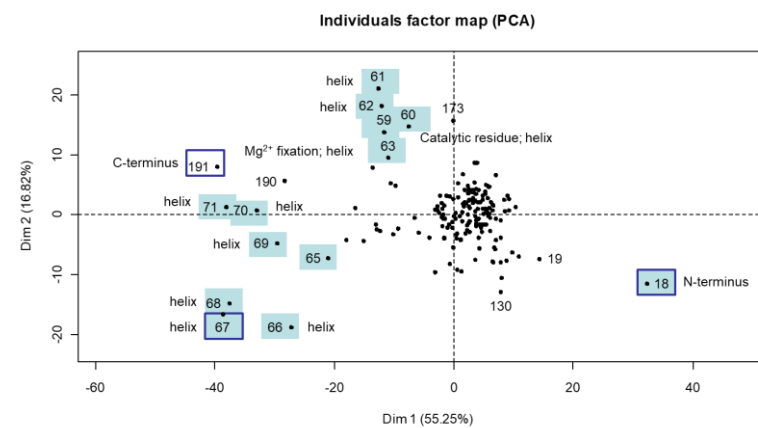
HSV-1 TK NMeDHBT z, run 1



HSV-1 TK NMeDHBT z, run 2



HSV-1 TK NMeDHBT z, run 3



hTK1 dTTP x

hTK1 dTTP y

hTK1 dTTP z

CHAPTER 4: KINASES TESTED (FIGURE 85)

Abbreviation	Full name
TBK1	Serine/threonine-protein kinase TBK1
IKKe	Inhibitor of nuclear factor kappa-B kinase subunit epsilon
YES1	Proto-oncogene tyrosine-protein kinase Yes
SYK	Tyrosine-protein kinase SYK
MST4	Serine/threonine-protein kinase MST4
EPH A2	Ephrin type-A receptor 2 precursor
IRR	Insulin receptor-related protein precursor
FGF-R1	Fibroblast growth factor receptor 1
CSK	Tyrosine-protein kinase CSK
Lck	Proto-oncogene tyrosine-protein kinase LCK
Src	Proto-oncogene tyrosine-protein kinase Src
PAK6	Serine/threonine-protein kinase PAK 6
PAK5	Serine/threonine-protein kinase PAK 5
PAK4	Serine/threonine-protein kinase
HIPK2	Homeodomain-interacting protein kinase 2
EFK2	Eukaryotic elongation factor 2 kinase
MST2	Serine/threonine-protein kinase 3
SRPK1	Serine/threonine-protein kinase SRPK1
PIM3	Serine/threonine-protein kinase Pim-3
PIM2	Serine/threonine-protein kinase Pim-2
PIM1	Proto-oncogene serine/threonine-protein kinase Pim-1
IKKb	Inhibitor of nuclear factor kappa-B kinase subunit beta
NEK6	Serine/threonine-protein kinase Nek6
NEK2a	Never in mitosis gene a-related kinase 2A splice variant NEK2A-T (Fragment)
DYRK3	Dual specificity tyrosine-phosphorylation-regulated kinase 3
DYRK2	Dual specificity tyrosine-phosphorylation-regulated kinase 2

Abbreviation	Full name
DYRK1A	Dual specificity tyrosine-phosphorylation-regulated kinase 1A, Protein kinase minibrain homolog, MNBH, HP86, Dual specificity YAK1-related kinase, hMNB
CK2	Casein kinase II subunit alpha, CK II
CK1	Casein kinase I isoform alpha-like, CKI-alpha-like
MELK	Maternal embryonic leucine zipper kinase
BRSK2	BR serine/threonine-protein kinase 2, Serine/threonine-protein kinase 29, Serine/threonine kinase SAD-A
MARK3	MAP/microtubule affinity-regulating kinase 3, Cdc25C-associated protein kinase 1, cTAK1, C-TAK1, Serine/threonine protein kinase p78, Ser/Thr protein kinase PAR-1, Protein kinase STK10
AMPK	5'-AMP-activated protein kinase subunit beta-1, AMPK beta-1 chain, AMPKb
PLK1 (Okadaic Acid)	Serine/threonine-protein kinase PLK1, Polo-like kinase 1, PLK-1, Serine/threonine-protein kinase 13, STPK13
PLK1	Serine/threonine-protein kinase PLK1, Polo-like kinase 1, PLK-1, Serine/threonine-protein kinase 13, STPK13
CDK2-Cyclin A	S-phase kinase-associated protein 1A, Cyclin A/CDK2-associated protein p19, p19A, p19skp1, RNA polymerase II elongation factor-like protein, Organ of Corti protein 2, OCP-II protein, OCP-2, Transcription elongation factor B, SIII).//S-phase kinase-associated protein 2, F-box protein Skp2, Cyclin A/CDK2-associated protein p45, p45skp2, F-box/LRR-repeat protein 1
GSK3b	Glycogen synthase kinase-3 beta
CHK2	Serine/threonine-protein kinase Chk2
CHK1	Serine/threonine-protein kinase Chk1
PHK	Phosphorylase b kinase regulatory subunit alpha
SmMLCK	Myosin light chain kinase 2
CAMK1	Calcium/calmodulin-dependent protein kinase type 1
PRAK	MAP kinase-activated protein kinase 5, MAPK-activated protein kinase 5, MAPKAP kinase 5, p38-regulated/activated protein kinase
MAPKAP-K2	MAP kinase-activated protein kinase 2, MAPK-activated protein kinase 2, MAPKAP kinase 2, MAPKAPK-2, MK2
MNK2	MAP kinase-interacting serine/threonine-protein kinase 2
MNK1	MAP kinase-interacting serine/threonine-protein kinase 1
MSK1	Ribosomal protein S6 kinase alpha-5, Nuclear mitogen-and stress-activated protein kinase 1, 90 kDa ribosomal protein S6 kinase 5, RSK-like protein kinase, RSKL
PKD1	Serine/threonine-protein kinase D1, nPKC-D1, Protein kinase D, Protein kinase C mu type, nPKC-mu
PKC zeta	Protein kinase C zeta type, nPKC-zeta
PKCa	Protein kinase C alpha type, PKC-alpha, PKC-A
PRK2	Serine/threonine-protein kinase N2, Protein kinase C-like 2, Protein-kinase C-related kinase 2

Abbreviation	Full name
ROCK2	Rho-associated protein kinase 2
PKA	cAMP-dependent protein kinase, alpha-catalytic subunit, PKA C-alpha
S6K1	Ribosomal protein S6 kinase beta-1, Ribosomal protein S6 kinase I, S6K, S6K1, 70 kDa ribosomal protein S6 kinase 1, p70 S6 kinase alpha, p70(S6K)-alpha, p70-S6K, P70S6K, p70-alpha
SGK1	Serine/threonine-protein kinase Sgk1, Serum/glucocorticoid-regulated kinase 1
PKBb	RAC-beta serine/threonine-protein kinase, RAC-PK-beta, Protein kinase Akt-2, Protein kinase B beta, PKB beta
PKBa	RAC-family serine/threonine-protein kinase
PDK1	Serine/threonine-protein kinase PDIK1L, PDLIM1-interacting kinase 1-like
RSK2	Ribosomal protein S6 kinase alpha-3, S6K-alpha 3, (90 kDa ribosomal protein S6 kinase 3, p90-RSK 3, Ribosomal S6 kinase 2, RSK-2, pp90RSK2, Insulin-stimulated protein kinase 1, ISPK-1, MAP kinase-activated protein kinase 1b, MAPKAPK1B
RSK1	Ribosomal protein S6 kinase alpha-1, MAP kinase-activated protein kinase 1a
ERK8	Mitogen-activated protein kinase 15, Extracellular signal-regulated kinase 8
p38a MAPK	Mitogen-activated protein kinase p38 alpha
p38g MAPK	Mitogen-activated protein kinase p38 gamma
p38b MAPK	Mitogen-activated protein kinase p38 beta
p38d MAPK	Mitogen-activated protein kinase p38 delta
JNK2	Mitogen-activated protein kinase 9
JNK1	Mitogen-activated protein kinase 8
ERK2	Mitogen-activated protein kinase 1
ERK1	Mitogen-activated protein kinase 3
MKK1	Dual specificity mitogen-activated protein kinase kinase 1