



Thèse

2023

Open Access

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

---

A novel tool for small molecule target identification in cytosolic and organellar sub-proteomes of Trypanosoma under *in vivo* conditions

---

Sahraoui, Suzanne Sherihan

#### How to cite

SAHRAOUI, Suzanne Sherihan. A novel tool for small molecule target identification in cytosolic and organellar sub-proteomes of Trypanosoma under *in vivo* conditions. Doctoral Thesis, 2023. doi: 10.13097/archive-ouverte/unige:171172

This publication URL: <https://archive-ouverte.unige.ch/unige:171172>

Publication DOI: [10.13097/archive-ouverte/unige:171172](https://doi.org/10.13097/archive-ouverte/unige:171172)

UNIVERSITÉ DE GENÈVE

FACULTÉ DES SCIENCES

Section des Sciences Pharmaceutiques

Département de Chimie/Biochimie pharmaceutique

Professeur Leonardo Scapozza

# **A novel tool for small molecule target identification in cytosolic and organellar sub-proteomes of *Trypanosoma* under *in vivo* conditions**

THÈSE

présentée aux Facultés de médecine et des sciences de l'Université de Genève  
pour obtenir le grade de Docteur ès sciences en sciences de la vie,  
mention Sciences pharmaceutiques

par

**Suzanne Sherihan Sahraoui**

de

Genève (GE)

Thèse N° 229

GENÈVE

Atelier de reproduction Repromail

2023



DOCTORAT ÈS SCIENCES EN SCIENCES DE LA VIE DES  
FACULTÉS DE MÉDECINE ET DES SCIENCES  
MENTION SCIENCES PHARMACEUTIQUES

**Thèse de Mme Suzanne Sherihan SAHRAOUI**

intitulée :

**« A novel tool for small molecule target identification in cytosolic and organellar sub-proteomes of Trypanosoma under *in vivo* conditions »**

Les Facultés de médecine et des sciences, sur le préavis de Monsieur Leonardo SCAPOZZA, Professeur ordinaire et directeur de thèse (Section des sciences pharmaceutiques), Monsieur Yogeshvar KALIA, Professeur ordinaire (Section des sciences pharmaceutiques), Monsieur Jean-Luc WOLFENDER, Professeur ordinaire (Section des sciences pharmaceutiques), Monsieur Mathieu BROCHET, Professeur associé (Département de microbiologie et médecine moléculaire) et Monsieur Pascal MÄSER, Professeur (Head of Parasite Chemotherapy, Swiss Tropical and Public Health Institute, Allschwil) autorisent l'impression de la présente thèse, sans exprimer d'opinion sur les propositions qui y sont énoncées.

Genève, le 31 juillet 2023

Thèse - 229 -

**Le Doyen**  
Faculté de médecine

**La Doyenne**  
Faculté des sciences

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".

*To my family...*

*Je dédie ma thèse à mon très cher et tendre père*

*Hassen Sakraoui et à ma défunte mère Ilham Sakraoui*

أهدي أطروحة الدكتوراه لأبي العزيز والعطاء

حسن الصحراوي وأمي الراحلة إلهام الصحراوي

*"Dieu ne joue pas au dés .... "*

— *Albert Einstein*

*"The present is theirs; the future, for which I really worked,  
is mine. "*

— *Nikola Tesla*

أشكر والدي...

# Acknowledgments

I am immensely grateful to Professor Leonardo Scapozza for giving me the opportunity to pursue my Ph.D. in his lab and for his exceptional guidance, support, and expertise throughout the entire thesis process. His insightful feedback and constructive criticism have been invaluable in shaping my research and writing. I deeply appreciate his patience and unwavering commitment to helping me to achieve my academic goal.

I would like to take this opportunity to express my sincere appreciation to PD. Dr. Remo Perozzo, my co-supervisor, for his incredible support, guidance, and expertise that have been invaluable to me during my master's thesis project. Dr. Perozzo's kindness, patience, and knowledge have been instrumental in my growth as a researcher and have helped me to navigate the complexities of my research with confidence. It was an immense honour for me to have been chosen to pursue a PhD. with him. His great expertise, mentorship, and guidance were invaluable, and he allowed me to conduct this research project, which was close to his heart. Sadly, Dr. Perozzo passed away in August 2019, but his legacy will continue to inspire me and future generations of students.

---

I would like to express my gratitude to all my professors who have contributed significantly to my academic journey. However, I would like to extend a special thank you to Professor Yogeshvar Kalia for his exceptional support, encouragement, and vast knowledge, which have been invaluable to me. Throughout my academic career, Professor Kalia has been an incredible mentor, always available to offer guidance and advice.

To sum up, I express my sincere gratitude to Professors Leonardo Scapozza and Yogeshvar Kalia, for their constant support and motivation. Their guidance has been pivotal in moulding my academic pursuits, and I am grateful to have had their mentorship.

I want to acknowledge the delightful ambiance in the FABIP Group and the unforgettable moments we shared, such as our group retreats in Annecy, Egypt, and Morocco. These were truly exceptional experiences, and I am immensely thankful to all my colleagues for contributing to such a fantastic journey.

I extend my profound appreciation to Aurélie Gouiller Olivier Petermann, Dr. Sébastien Tardy, and Dr. Oscar Vadas for their invaluable assistance that enabled me to complete my thesis project successfully. Their precious advice, generous support, and time spent explaining and demonstrating important techniques and methods were instrumental in my achievement.

I would like to express my gratitude to Prof. Pascal Maser and Dr. Marcel Kaiser for granting me the opportunity to undertake an internship at the Swiss Tropical Institute in Basel (Swiss TPH). This experience allowed me to gain valuable knowledge in the field of parasitology, and I was fortunate to have been exposed to an exceptional learning environment and training. I am also grateful to the remarkable parasitology team who provided me with excellent guidance and equipped me with the necessary skills for my professional growth.

With profound appreciation and gratitude, I dedicate this work to the most influential and significant person in my life, my beloved father. His steadfast support, encouragement, and love have been the cornerstone of my academic and personal achievements. Through every obstacle and challenge that I faced, my father was always there, providing me with invaluable guidance and motivation to persevere. His resolute belief in me has been a constant source of inspiration, fuelling my determination to pursue my dreams and reach new heights. It is his unwavering love and guidance that have helped shape the person I am today, and for that, I will always be grateful. From the bottom of my heart, I want to express my appreciation for his tireless efforts in supporting me and for instilling in me the values of hard work, perseverance, and dedication. I am forever indebted to him for the countless sacrifices he has made to ensure my success and well-being. Thank you, Dad, for your unyielding love, guidance, and support. This work is a testament to your impact on my life, and I am honored to share it with you.

I also want to pay tribute to my mother, who passed away 12 years ago due to multiple sclerosis. Her struggle with this debilitating disease gave me a deep motivation to pursue my career in research, develop new ideas, and improve my English. Her memory lives on in everything I do, and I am forever grateful for the love and sacrifices she made for our family.

I am incredibly blessed to have such amazing parents, and I cannot thank them enough for their love, support, and inspiration.

I would like to express my profound gratitude to my brothers, Nabil and Haykel, for their steadfast and resolute support throughout my life. They have been my pillars of strength, motivating me to pursue my dreams and always pushing me to be the best version of myself. Their unwavering dedication to my well-being has been a source of great comfort and inspiration to me, and I am truly fortunate to have them in my life.

Last but not least, I would like to express my gratitude to my twin sister, Firdaws Sahraoui. Our 29 years of existence together and our similar paths in studies have allowed us to be there for each other, train together, and most importantly, support each other. This year, we will defend our theses together, which will be the most significant moment for both of us and the pride of our parents. Despite our many arguments and moments of sadness, as well as joy, we have shared everything, and you will always be my better half, my twin sister for life.

I would like to extend my sincere thanks to Abir, my brother's wife, who has been an amazing mentor, a perfect big sister, and a best friend to me. Her wisdom, generosity, and compassion have been invaluable to me, and I feel incredibly lucky to have her in my life. Her valuable advice, kind-heartedness, and immense love have helped me in countless ways especially to be able to navigate some of the most challenging moments of my life, and I will always be grateful for her presence.

I would like to sincerely thank my other family members, Achraf, Inès, and Naima, for their constant support, inspiration, patient listening, and priceless counsel. Their guidance and assistance were essential, and I am grateful to them.

As I reflect on my academic journey, I am filled with gratitude for the three special people in my life: Radhia, Melissa, and Livia. More than friends, they are sisters to me. Their great support, kindness, and encouragement have been a constant source of strength and inspiration, guiding me through the highs and lows of this amazing trip. Their belief and confidence in me have pushed me to strive for excellence and have had an immeasurable impact on my life. I am incredibly blessed to have such amazing sisters, and I cannot thank them enough for being there for me every step of the way.

Radhia, you are more than a friend or colleague. You are my other twin, and I cannot express how much I appreciate your unfailing support, especially at times when I needed it most and

felt down. All the enjoyable times we had together while on vacation will always be treasured and wonderful memories.

I extend my heartfelt appreciation to my long-time friends, Mona, Wiem, and Lara for their amazing support and encouragement throughout my academic journey. We have created some truly unforgettable memories together, from sharing laughter and coffee during various class breaks to going on holiday together, and even preparing for the final federal exam. The memories that I have shared with them will always hold a special place in my heart. I am grateful for their friendship and the positive impact they have had on my life. I am deeply grateful for their friendship and the invaluable role they have played in my life.

I would like to express my sincere gratitude to my dear cousin Inès for introducing me to all the members of Sofia Plenitude Association. Their unwavering support throughout my thesis and during the "ma thèse en 180 secondes" MT180 competition was invaluable. Being a part of this association and getting to know such wonderful people has been an enriching experience. I am especially grateful for the opportunity to participate in the humanitarian trip to Palestine with my sisters from this association. It was an experience that I will cherish for the rest of my life.

Additionally, I would say that competing in the MT180 competition for the 2023 edition was one of the most fulfilling and unforgettable events of my life. Therefore, I want to express my sincere gratitude to all the members of MT180 for giving me the opportunity to participate in the contest, as well as for their invaluable support and guidance that helped me reach the finals in Geneva. I am truly grateful for their contributions for making this experience possible.

Lastly, I am deeply grateful to my friends and colleagues Ece, Sofia, and Solène for their kindness, exceptional support, and attentive listening throughout my academic journey. The wonderful moments we have shared together are truly cherished.

# Table of Contents

|                                                                                     |             |
|-------------------------------------------------------------------------------------|-------------|
| <b>ACKNOWLEDGMENTS</b>                                                              | <b>II</b>   |
| <b>TABLE OF CONTENTS</b>                                                            | <b>VI</b>   |
| <b>ABBREVIATION TABLE</b>                                                           | <b>VIII</b> |
| <b>SUMMARY</b>                                                                      | <b>X</b>    |
| <b>RÉSUMÉ</b>                                                                       | <b>XIII</b> |
| <b>PART 1: PARASITIC DISEASES AND TARGET DECONVOLUTION METHODS: AN INTRODUCTION</b> | <b>16</b>   |
| <b>.1 INTRODUCTION</b>                                                              | <b>17</b>   |
| <b>.2 PARASITIC DISEASES</b>                                                        | <b>19</b>   |
| .2.1 AFRICAN TRYPANOSOMIASIS                                                        | 19          |
| .2.2 LEISHMANIASIS                                                                  | 22          |
| .2.3 CHAGAS DISEASE                                                                 | 23          |
| <b>.3 TARGET DECONVOLUTION METHODS</b>                                              | <b>25</b>   |
| .3.1 GENETIC AND CHEMICAL PROTEOMIC-BASED STRATEGIES                                | 26          |
| .3.2 METABOLOMIC-BASED STRATEGIES                                                   | 32          |
| .3.3 COMPUTATIONAL-BASED STRATEGIES                                                 | 34          |
| <b>.4 SWOT ANALYSIS</b>                                                             | <b>39</b>   |
| .4.1 STRENGTHS                                                                      | 43          |
| .4.2 WEAKNESSES                                                                     | 45          |
| .4.3 OPPORTUNITIES                                                                  | 47          |
| .4.4 THREATS                                                                        | 50          |
| <b>.5 CONCLUSION</b>                                                                | <b>52</b>   |
| <b>.6 REFERENCES</b>                                                                | <b>54</b>   |
| <b>PART 2: AIM OF THE PROJECT</b>                                                   | <b>65</b>   |
| <b>.1 AIM OF THE PROJECT</b>                                                        | <b>66</b>   |
| <b>.2 REFERENCES</b>                                                                | <b>69</b>   |
| <b>PART 3: DEVELOPMENT OF A NEW TARGET DECONVOLUTION TOOL</b>                       | <b>71</b>   |
| <b>.1 INTRODUCTION</b>                                                              | <b>72</b>   |
| .1.1 <i>TRYPANOSOMA BRUCEI</i> PARASITES                                            | 72          |

|           |                                                                                                   |            |
|-----------|---------------------------------------------------------------------------------------------------|------------|
| .1.2      | A SMALL OVERVIEW OF OTHER NEGLECTED TROPICAL DISEASES CAUSED BY PARASITE OF THE SAME FAMILY       | 82         |
| .1.3      | CURRENT TARGET DECONVOLUTION METHODS AS STARTING TECHNOLOGIES FOR THE DEVELOPMENT OF THE NEW TOOL | 84         |
| .1.4      | CHALLENGES OF THE PROJECT                                                                         | 89         |
| <b>.2</b> | <b>MATERIALS AND METHODS</b>                                                                      | <b>93</b>  |
| .2.1      | MATERIALS                                                                                         | 93         |
| .2.2      | METHODS                                                                                           | 94         |
| <b>.3</b> | <b>RESULTS AND DISCUSSION</b>                                                                     | <b>119</b> |
| .3.1      | THE LINKER STORY: THE CULMINATION POINT OF BG DERIVATIVES AND FUSION PROTEINS                     | 119        |
| .3.2      | GENERATION OF DIFFERENT PLASMIDS FOR PARASITE TRANSFECTION                                        | 125        |
| .3.3      | TRANSFECTION OF PLASMIDS OF INTERESTS                                                             | 131        |
| .3.4      | FLUORESCENCE MICROSCOPY                                                                           | 133        |
| .3.5      | GROWTH CURVES OF DIFFERENT STRAINS                                                                | 134        |
| .3.6      | TIME-DEPENDANT PROTEIN EXPRESSION                                                                 | 143        |
| .3.7      | BIOTINYLATION STUDIES                                                                             | 154        |
| .3.8      | MS DETECTION OF THE CYS145 MODIFICATION OF THE SNAP-TAG ACTIVE SITE BY BG-PEG <sub>4</sub> -MTX   | 187        |
| .3.9      | KEY FINDINGS FROM DHFR-TS BIOTINYLATION STUDIES ACROSS VARIOUS TOOLS                              | 192        |
| .3.10     | EARLY FINDINGS OF <i>Tb</i> AK BIOTINYLATION AND CAPTURE USING THE cMYC_BIRA*_GGG_SNAP-TAG TOOL   | 193        |
| .3.11     | PROOF OF CONCEPT OF THE SUCCESSFUL FUNCTIONING OF THE TOOL                                        | 195        |
| <b>.4</b> | <b>CONCLUSIONS &amp; OUTLOOKS</b>                                                                 | <b>198</b> |
| .4.1      | GENERAL CONCLUSION                                                                                | 198        |
| .4.2      | OUTLOOKS                                                                                          | 200        |
| <b>.5</b> | <b>REFERENCES</b>                                                                                 | <b>205</b> |
| <b>.6</b> | <b>APPENDIX</b>                                                                                   | <b>214</b> |
| <b>.7</b> | <b>POSTERS, TALKS, AND PUBLICATIONS</b>                                                           | <b>244</b> |
| .7.1      | TALKS                                                                                             | 244        |
| .7.2      | POSTERS                                                                                           | 244        |

# Abbreviation table

|                               |                                                                                      |
|-------------------------------|--------------------------------------------------------------------------------------|
| <b><i>T. brucei</i></b>       | <i>Trypanosoma brucei</i>                                                            |
| <b><i>Tbb</i></b>             | <i>Trypanosoma brucei brucei</i>                                                     |
| <b><i>Tbr</i></b>             | <i>Trypanosoma brucei rhodesiense</i>                                                |
| <b><i>Tbg</i></b>             | <i>Trypanosoma brucei gambiense</i>                                                  |
| <b><i>Tbb_WT</i></b>          | <i>Trypanosoma brucei brucei</i> wild-type NYSM strain                               |
| <b><i>Tbb_G0</i></b>          | <i>Trypanosoma brucei brucei</i> transfected with the cMyc_BirA*_GGG_SNAP-tag gene   |
| <b><i>Tbb_G1</i></b>          | <i>Trypanosoma brucei brucei</i> transfected with the cMyc_BirA*_GGGS_SNAP-tag gene  |
| <b><i>Tbb_G2</i></b>          | <i>Trypanosoma brucei brucei</i> transfected with the cMyc_BirA*_GGGGG_SNAP-tag gene |
| <b><i>Tbb_TG0</i></b>         | <i>Trypanosoma brucei brucei</i> transfected with the cMyc_TbirA*_GGG_SNAP-tag gene  |
| <b><i>Tbb_mTG0</i></b>        | <i>Trypanosoma brucei brucei</i> transfected with the cMyc_mTbirA*_GGG_SNAP-tag gene |
| <b><i>Tbb_PA</i></b>          | <i>Trypanosoma brucei brucei</i> transfected with the cMyc_BirA*_PAPAP_SNAP-tag gene |
| <b><i>Tbb_PTP</i></b>         | <i>Trypanosoma brucei brucei</i> transfected with the cMyc_BirA*_PTP_SNAP-tag gene   |
| <b><i>gDNA</i></b>            | genomic DNA                                                                          |
| <b>Linker G0</b>              | GGG (Glycine-Glycine-Serine) linker                                                  |
| <b>Linker G1</b>              | GGGS (Glycine-Glycine-Glycine-Serine) linker                                         |
| <b>Linker G2</b>              | GGGGG (Glycine-Glycine-Glycine-Glycine-Serine) linker                                |
| <b>Linker PA</b>              | PAPAP (Proline-Alanine-Proline-Alanine-Proline) linker                               |
| <b>MS</b>                     | Mass Spectrometry                                                                    |
| <b>PDT</b>                    | Population doubling time                                                             |
| <b>BG</b>                     | Benzylguanine                                                                        |
| <b>Mtx</b>                    | Methotrexate                                                                         |
| <b>BG-PEG<sub>4</sub>-Mtx</b> | O <sup>6</sup> -benzylguanine-PEG <sub>4</sub> -Methotrexate derivative              |

|                               |                                                                       |
|-------------------------------|-----------------------------------------------------------------------|
| <b>BG-PEG<sub>4</sub>-FDA</b> | O <sup>6</sup> -benzylguanine-PEG <sub>4</sub> -Fluorescein diacetate |
| <b>RT</b>                     | Room Temperature                                                      |
| <b>DHFR-TS</b>                | Bifunctional dihydrofolate reductase-thymidylate synthase             |
| <b><i>TbAK</i></b>            | Adenosine kinase of <i>Trypanosoma brucei</i>                         |

# Summary

Human African trypanosomiasis (HAT), also known as sleeping sickness, and animal trypanosomiasis, also known as Nagana plague, are pathologies caused by single-celled parasites of the genus *Trypanosoma* of the family Trypanosomatidae belonging to the class Kinetoplastea. The species infecting humans are *Trypanosoma brucei gambiense* (**Tbg**), responsible for the chronic form of sleeping sickness, which accounts for over 95-97% of reported cases, and *Trypanosoma brucei rhodesiense* (**Tbr**), causing the acute form, which accounts for only 3-5% of cases.

Despite the introduction of fexinidazole in 2019 and some other promising molecules, there is an urgent need to develop new therapies for the disease in the hopes of curing it. This is because the present HAT medications are obsolete, out-of-date, and have significant adverse effects. The medical crisis remains persistent given the emergence of drug resistance, administration restrictions that necessitate a parenteral route for severe forms, or the great difficulty in developing molecules that can cross the blood-brain barrier to treat phase 2, which is typically lethal and fatal for patient given the emergence of many neurological symptoms.

The discovery of current drugs used against Human African Trypanosomiasis has been made possible by using phenotypic screening measuring the effects of molecules on parasite growth and survival. Although this approach has been successful, the targeted proteins, the mechanism of action, and often the cause of the side effects remained to be discovered.

To address the question of deconvoluting the active compound's targets and to support drug design and lead optimization, scientists started to develop methods to identify cellular targets and elucidate the mechanism of action of lead compounds. Up to now, many target identification techniques have been developed and successfully applied. However, some drawbacks such as the loss of some parts of the proteome or strong background signals have been reported.

This project aims at developing a novel deconvolution tool able to overcome the limitations of existing techniques. To achieve this aim, we generate a novel system that combines SNAP-tag technology and BioID technology based on the BirA\* protein and can integrate and express

the fusion protein within *Trypanosoma*. SNAP-tag, which is based on a modified version of human O6-alkylguanine-DNA-alkyltransferase (hAGT), a DNA repair protein with the unique ability to form a covalent bond with its substrate (benzylguanine (BG)), will add the required selectivity to the new system. A promising molecule whose target we hope to discover will be linked to the BG part via a linker: PEG<sub>4</sub>, which consists of a succession of four ethylene glycol units. BirA\* protein is a mutant version of the *Escherichia coli* (*E. coli*) biotin ligase that generates the highly reactive intermediate biotinoyl-5'-AMP. This modified biotin ligase, having lost its affinity for the intermediate, promotes its release into the environment, enabling BirA\* to biotinylate any partner in its immediate vicinity in the presence of ATP and biotin. This functionality will be exploited to enhance sensitivity by biotinylation of putative therapeutic targets of interest and detection of protein-protein interactions. The resulting fusion protein will facilitate the capture of small molecule targets in the cytosol and organelles of the *Trypanosoma brucei* (*T. brucei*), thereby surpassing the limitations of the individual methods.

Various challenges and scientific questions related to the development of the method were addressed. Indeed, the design of a plasmid allowing homologous recombination and expression of a polycistronic gene in *T. brucei* represents the first challenge that was successfully met. The ability to obtain different stable cell lines correctly integrating the gene of interest was challenging and difficult but has been successfully achieved. As the functionality of the system is depending on the three-dimensional arrangement of the different proteins in the 3D space, various tools have been generated by modifying either the linker or the nature of the BirA\* protein within the fusion protein.

In this project, we were able to alter the properties of BirA\* by using its two mutants, TBirA\* and mTBirA\*, which belong to TurboID and miniTurbo systems, both of which are based on the BioID system, and then, in a subsequent step, the linker by adjusting either its length or the nature of the amino acids that compose it.

As a result, BirA\* and the linker made of either GGS (Glycine-Serine-Glycine) or PTP (Proline-Threonine-Proline) gave the most convincing results, which included the successful isolation of DHFR-TS, the methotrexate's target that had previously been joined to the BG-PEG<sub>4</sub> moiety to form the BG-PEG<sub>4</sub>-Mtx derivative.

Expression of the various tools showed no parasitic toxicity. Moreover, adequate overexpression under induction by the antibiotic tetracycline was successfully achieved. Parallel to these results, electroporation of parasites increases the chances of detecting the target of interest captured by western blot. This method, generally used to insert different plasmids and thus generate genetically modified strains, was exploited to force entry of the BG-PEG<sub>4</sub>-Mtx derivative into the parasite in the hope of better observing DHFR-TS. This experiment improved visualization of the target of interest while simultaneously demonstrating the non-lethality of electroporation for the parasites.

Summarizing, the following systems composed by BirA\* the linker GGS (Glycine-Serine-Glycine) and SNAP-tag (cMyc\_BirA\*\_GGS\_SNAP-tag) or the linker PTP (Proline-Threonine-Proline) (cMyc\_BirA\*\_PTP\_SNAP-tag) have demonstrated the most promising and convincing results for effectively isolating the targets of control molecules of interest, specifically bifunctional dihydrofolate reductase-thymidylate synthase (DHFR-TS) and Adenosine kinase of *T. brucei* (TbAK), the respective targets of methotrexate and the C1 molecule.

In conclusion, this effort has resulted in a promising target detection and capture technology. With further advances, a powerful, accurate and targeted system could be generated and employed in the future to capture and discover potential targets of promising compounds in *T. brucei* and in many other parasite species.

# Résumé

La Trypanosomiase Humaine Africaine (THA), également appelée maladie du sommeil, et la trypanosomiase animale, également appelée peste de Nagana, sont des pathologies causées par des parasites unicellulaires du genre *Trypanosoma* de la famille des Trypanosomatidae appartenant à la classe des Kinetoplastea. Les espèces infectant l'homme sont *Trypanosoma brucei gambiense* (**Tbg**), responsable de la forme chronique de la maladie du sommeil, qui représente plus de 95 à 97 % des cas signalés, et *Trypanosoma brucei rhodesiense* (**Tbr**), causant de la forme aiguë, qui ne représente que 3 à 5 % des cas.

Malgré l'introduction du fexinidazole en 2019 et de quelques autres molécules prometteuses, il est urgent de développer de nouvelles thérapies pour cette maladie dans l'espoir de la guérir. En effet, les médicaments actuels contre la THA sont obsolètes, dépassés et ont des effets indésirables importants. La crise médicale reste persistante compte tenu de l'émergence de résistances aux médicaments, des restrictions d'administration qui nécessitent une voie parentérale pour les formes sévères, ou encore de la grande difficulté à développer des molécules capables de traverser la barrière hémato-encéphalique pour traiter la phase 2, typiquement mortelle et fatale pour le patient compte tenu de l'émergence de nombreux symptômes neurologiques.

La découverte des médicaments actuellement utilisés contre la trypanosomiase humaine africaine a été rendue possible par l'utilisation d'un criblage phénotypique mesurant les effets des molécules sur la croissance et la survie du parasite. Bien que cette approche ait été couronnée de succès, les protéines ciblées, le mécanisme d'action et souvent la cause des effets secondaires restent à découvrir.

Pour répondre à la question de la déconvolution des cibles du composé actif et pour soutenir la conception de médicaments et l'optimisation des têtes de série, les scientifiques ont commencé à développer des méthodes pour identifier les cibles cellulaires et élucider le mécanisme d'action des têtes de série. Jusqu'à présent, de nombreuses techniques d'identification des cibles ont été développées et appliquées avec succès. Cependant, certains inconvénients tels que la perte de certaines parties du protéome ou des signaux de fond importants ont été signalés.

Ce projet vise à développer un nouvel outil de déconvolution capable de surmonter les limites des techniques existantes. Pour atteindre cet objectif, nous générons un nouveau système qui combine la technologie SNAP-tag et la technologie BioID basée sur la protéine BirA\* et qui peut intégrer et exprimer la protéine de fusion dans *Trypanosoma*. SNAP-tag, qui est basé sur une version modifiée de la O6-alkylguanine-DNA-alkyltransférase humaine (hAGT), une protéine de réparation de l'ADN ayant la capacité unique de former une liaison covalente avec son substrat (benzylguanine (BG)), ajoutera la sélectivité nécessaire au nouveau système. Une molécule prometteuse, dont nous espérons découvrir la cible, sera liée à la partie BG par l'intermédiaire d'un linker : PEG4, qui consiste en une succession de quatre unités d'éthylène glycol. La protéine BirA\* est une version mutante de la biotine ligase d'*Escherichia coli* (*E. coli*) qui génère l'intermédiaire hautement réactif biotinoyl-5'-AMP. Cette biotine ligase modifiée, ayant perdu son affinité pour l'intermédiaire, favorise sa libération dans l'environnement, ce qui permet à BirA\* de biotinyler n'importe quel partenaire dans son voisinage immédiat en présence d'ATP et de biotine. Cette fonctionnalité sera exploitée pour améliorer la sensibilité par biotinylation des cibles thérapeutiques putatives et la détection des interactions protéine-protéine. La protéine de fusion résultante facilitera la capture de petites molécules cibles dans le cytosol et les organites de *Trypanosoma brucei* (*T. brucei*), dépassant ainsi les limites des méthodes individuelles.

Différents défis et questions scientifiques liés au développement de la méthode ont été abordés. En effet, la conception d'un plasmide permettant la recombinaison homologe et l'expression d'un gène polycistronique chez *T. brucei* représente le premier défi qui a été relevé avec succès. La capacité d'obtenir différentes lignées cellulaires stables intégrant correctement le gène d'intérêt a été un défi et une difficulté, mais a été réalisée avec succès. Comme la fonctionnalité du système dépend de l'arrangement tridimensionnel des différentes protéines dans l'espace 3D, divers outils ont été générés en modifiant soit le linker, soit la nature de la protéine BirA\* dans la protéine de fusion.

Dans ce projet, nous avons pu modifier les propriétés de BirA\* en utilisant ses deux mutants, TBirA\* et mTBirA\*, qui appartiennent aux systèmes TurboID et miniTurbo, tous deux basés sur le système BioID, puis, dans une étape ultérieure, le linker en ajustant soit sa longueur, soit la nature des acides aminés qui le composent.

Ainsi, BirA\* et le linker composé soit de GGS (Glycine-Sérine-Glycine) soit de PTP (Proline-Thréonine-Proline) ont donné les résultats les plus probants, parmi lesquels l'isolement réussi du DHFR-TS, la cible du méthotrexate qui avait été précédemment jointe à la partie BG-PEG<sub>4</sub> pour former le dérivé BG-PEG<sub>4</sub>-Mtx.

L'expression des différents outils n'a montré aucune toxicité parasitaire. De plus, une surexpression adéquate sous induction par l'antibiotique tétracycline a été obtenue avec succès. Parallèlement à ces résultats, l'électroporation des parasites augmente les chances de détecter la cible d'intérêt capturée par western blot. Cette méthode, généralement utilisée pour insérer différents plasmides et ainsi générer des souches génétiquement modifiées, a été exploitée pour forcer l'entrée du dérivé BG-PEG<sub>4</sub>-Mtx dans le parasite dans l'espoir de mieux observer le DHFR-TS. Cette expérience a permis d'améliorer la visualisation de la cible d'intérêt tout en démontrant la non-létalité de l'électroporation pour les parasites.

En résumé, les systèmes suivants composés de BirA\*, du linker GGS (Glycine-Sérine-Glycine) et du SNAP-tag (cMyc\_BirA\*\_GGS\_SNAP-tag) ou du linker PTP (Proline-Thréonine-Proline) (cMyc\_BirA\*\_PTP\_SNAP-tag) ont démontré les résultats les plus prometteurs et les plus convaincants pour isoler efficacement les cibles des molécules de contrôle d'intérêt, en particulier la dihydrofolate réductase-thymidylate synthase bifonctionnelle (DHFR-TS) et l'adénosine kinase de *T. brucei* (*TbAK*), les cibles respectives du méthotrexate et de la molécule C1.

En conclusion, cet effort a abouti à une technologie prometteuse de détection et de capture de cibles. Avec des avancées supplémentaires, un système puissant, précis et ciblé pourrait être généré et employé à l'avenir pour capturer et découvrir des cibles potentielles de composés prometteurs chez *T. brucei* et également chez de nombreuses autres espèces de parasites.

**Part 1: Parasitic diseases and target  
deconvolution methods: an introduction**

# Parasitic diseases and target deconvolution methods: an introduction

Suzanne Sherihan Sahraoui <sup>1,2\*</sup> , Radhia EL Phil<sup>1,2\*</sup> , Sylvian Cretton<sup>1,2</sup>, Muriel Cuendet<sup>1,2</sup>, Stefano Muscat<sup>5</sup>, Gianvito Grasso<sup>5</sup>, Andrea Danani<sup>5</sup>, Sabina Beilstein<sup>3,4</sup>, Pascal Maeser<sup>3,4</sup>, Leonardo Scapozza<sup>1,2</sup>

<sup>1</sup> School of Pharmaceutical Sciences, University of Geneva, 1211 Geneva, Switzerland

Suzanne.Sahraoui@unige.ch, [Radhia.ElPhil@unige.ch](mailto:Radhia.ElPhil@unige.ch), [Muriel.Cuendet@unige.ch](mailto:Muriel.Cuendet@unige.ch), Leonardo.Scapozza@unige.ch

<sup>2</sup> Institute of Pharmaceutical Sciences of Western Switzerland, University of Geneva, Switzerland

<sup>3</sup> Swiss Tropical and Public Health Institute, 4051 Basel, Switzerland

[pascal.maeser@swisstph.ch](mailto:pascal.maeser@swisstph.ch)

<sup>4</sup> University of Basel, 4002 Basel, Switzerland

<sup>5</sup> IDSIA-Dalle Molle Institute for Artificial Intelligence, University of Applied Sciences and Arts of Southern Switzerland (SUPSI), 6962 Lugano, Switzerland

\* The authors contributed equally to the work

**\* Note that this first introduction represents part of the review paper that will be submitted and includes some extension specific for the thesis manuscript.**

## .1 Introduction

As of today, two different approaches are widely used in the search for new, more effective drugs: Target-based and Phenotype-based approaches (1, 2).

The whole-genome sequencing of multiple diverse organisms made the target-based strategy easier to implement (3). Indeed, this strategy became more popular as a result of improvements in genome sequencing technology, which made it possible to efficiently and inexpensively sequence the whole genome of a wide range of species (3-5). Nowadays, many databases containing the whole genome sequences of different types of parasites are available (e.g. the Trityp database) and with the help of evolving bioinformatics and molecular biology tools, potential therapeutic targets could be identified (3, 6-8).

With the progress of new genetic methods including genome sequencing, genetic recombination, Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas9, RNA interference (RNAi), and knock-out/knockdown of a particular gene in addition to increasingly powerful bioinformatics tools, the target-based strategy is, in general, becoming more and more attractive for the search of more effective drugs (2). Many successes of the target-based approach have been reported, including some tyrosine kinase inhibitors against

cancer, such as Gleevec (imatinib targeting BCR-Abl), Iressa (gefitinib targeting EGFR), and various antivirals notably Isentress (raltegravir targeting HIV integrase), Relenza (zanamivir targeting influenza virus neuraminidase)(1, 9-12), Paxlovid (nirmatrelvir targeting the main protease ( $M^{pro}$ ) of SARS-CoV-2 virus) (13), Lagevrio (molnupiravir, targeting the RNA-dependent RNA polymerase (RdRp) of the SARS-CoV-2 virus)(13, 14) or also Remdesivir which was revealed to firmly attach to the SARS-CoV-2 membrane protein (Mprotein), RNA-dependent RNA polymerase (RdRp), and major protease ( $M^{pro}$ ) (15). Furthermore, an analysis of 113 first-in-class drugs that were approved by the Food and Drug Administration (FDA) showed that 78 of them were identified and discovered using target-based strategies (16). Out of these, 33 were biologics and 45 were small molecules (16). Nevertheless, the requirement for target expression in an adequate *in-vitro* system is a major drawback of target-based drug development (17). Tests using *in-vitro* recombinant proteins or recombinant cells do not always reflect the complexity found in intact organisms (17-19). Another challenge of target-based drug discovery is that drugs may act on multiple targets (18). Consequently, the therapeutic impact observed may not always be solely attributable to the identified molecular target (18, 19).

To recapitulate the complexity found in an intact organism and with the huge progress in the development of tools such as induced pluripotent stem cells, organoids, gene editing techniques such as CRISPR-Cas, and high-content imaging methods there has been a resurgence of phenotypic drug discovery in the pharmaceutical industry. Phenotypic drug discovery is based on the evaluation of a given compound on a particular phenotype, be it cells, protozoa, organs, or animals. This approach makes it possible to identify drugs with an interesting pharmacological action without necessarily knowing the target or the mechanism of action (2, 20). Indeed, the whole-organism screening strategy has been widely used for the development of current antiprotozoal drugs (20-22). A majority of the antiprotozoal agents currently on the market were discovered by phenotypic screening, either through the repurposing of already-approved therapies or through testing libraries of novel compounds. The repurposing axe is based on the idea of using an already approved drug for a new disease or indication. The main advantage of this strategy is that it saves time and money, and thus avoids some of the tedious and complicated approval procedures (21, 23). Phenotypic

screening on a whole organism has also some limitations one of which is that the mechanism of action causing the compound efficacy remains to be elucidated.

Knowing the target, as well as the mechanism of action, could speed up the progress of lead optimization and drug candidates in clinical studies (20). It is estimated that almost half of the safety failures in clinical trials are related to the target of the drug (24). It would therefore be possible to decrease development costs if several targets and mechanisms of action were to be studied and evaluated (1).

In conclusion, although the majority of antiprotozoal drugs are derived from whole parasite screening, target-based strategies, used as complementary methods to phenotypic screening, could potentially accelerate the drug development process and thus limit the costs involved, especially for drugs that have a low return on investment and are therefore unattractive to the pharmaceutical industry due to the high costs of approval procedures (21).

## **.2 Parasitic diseases**

### **.2.1 African Trypanosomiasis**

The protozoan parasite *Trypanosoma brucei* (*T. brucei*) causes Human African trypanosomiasis (HAT) also known as sleeping sickness.

It is transmitted by the bite of an infected tsetse fly (25, 26). HAT is localized in sub-Saharan Africa and is caused by two subspecies of the parasite. On one side, *Trypanosoma brucei gambiense* is responsible for the chronic form of the disease and is mainly found in 24 countries in west and central Africa and it accounts for over 95% of cases that were reported according to the WHO (27). On the other side, *T. b. rhodesiense* which is present in 13 countries in eastern and southern Africa, is responsible for the acute form and represents about 5% of reported cases (25, 27, 28). The two forms of the disease can be fatal, especially if left untreated. Both forms are characterized by two different stages. The early hemolymphatic stage where patients have non-specific symptoms such as episodes of fevers lasting up to 7 days, headache, weakness, malaise, fatigue, arthralgia and myalgia, and the late encephalitic stage where the parasite invades the central nervous system. This is

characterized by more specific and damaging symptoms such as limb tremors, personality changes, hallucinations, delusions, poor coordination, numbness, and disruption of the sleep cycle which represents the leading symptom of that stage (27-29). This latter stage occurs rapidly (within a few months) in the disease caused by *T. b. rhodesiense* whereas it may take years to develop when patients are infected with *T. b. gambiense* (27, 29, 30).

Although the number of confirmed cases has decreased significantly, from 26'550 in 2000 to 980 in 2019 according to the latest data provided by the WHO, the available treatments for either the early or the late stage present severe side effects (Table 1). For example, melarsoprol is generally used to treat the second phase of the disease, but this drug has many side-effects, one of which is life-threatening: the post-treatment encephalopathy (PTRE) which is fatal in 5-10% of treated patients (27, 31).

The other molecules are used to treat either the first or the second stage according to the two different subspecies of *T. brucei* (Table 1).

Table 1: Drugs used to treat stages 1 or 2 of HAT and their side effects and resistance profiles

| Drug                                                         | Spectrum                                             | Stage | Main side effects                                                                                                                     | Resistance                                                                               | References           |
|--------------------------------------------------------------|------------------------------------------------------|-------|---------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|----------------------|
| Pentamidine                                                  | <i>T. b. gambiense</i>                               | I     | Hypo/Hyperglycemia, Hypotension, renal dysfunction (nephrotoxicity)                                                                   | No (failure remains rare)                                                                | (28, 29, 32, 33)     |
| Suramin                                                      | <i>T. b. rhodesiense</i>                             | I     | Allergic reactions (anaphylaxis), nephrotoxicity, bone marrow toxicity                                                                | Yes (Expression of a particular Variant Surface Glycoprotein called VSG <sup>sur</sup> ) | (28, 29, 32-34)      |
| Eflornithine                                                 | <i>T. b. gambiense</i>                               | II    | Bone marrow toxicity, seizures, hearing loss, alopecia, gastrointestinal effects                                                      | Yes (through loss of a putative amino acid transporter TbAAT6)                           | (28, 29, 32, 33, 35) |
| Nifurtimox (used as a combination therapy with eflornithine) | <i>T. b. gambiense</i>                               | II    | Gastrointestinal effects, neurologic effects (tremor, vertigo, ataxia...), less side effects in combination therapy with eflornithine | Yes (cross-resistance mechanism with other molecules)                                    | (28, 29, 32, 33, 36) |
| Melarsoprol                                                  | <i>T. b. gambiense</i> ,<br><i>T. b. rhodesiense</i> | II    | Reactive encephalopathy (PTRE), agranulocytosis, skin rashes, cardiotoxicity, peripheral neuropathy                                   | Yes (though mutations affecting an aquaglyceroporin (AQP2))                              | (28, 31, 33, 37)     |
| Fexinidazole                                                 | <i>T. b. gambiense</i>                               | I, II | Gastrointestinal effects (mainly nausea, and vomiting)                                                                                | No                                                                                       | (28, 33, 38-41)      |

The toxic and even fatal side effects (e.g., melarsoprol), and the emergence of resistance demonstrate the need to develop more effective and safer compounds to treat the different stages of HAT (31). While intensive recent research led fexinidazole to the market, the need for further efficacious and safe treatments to treat both stages with fewer side effects remains (39-41) in order to reach the WHO's goal of eliminating the transmission of *T. gambiense* HAT by 2030 (42).

## **.2.2 Leishmaniasis**

Leishmaniasis is a category of tropical illnesses caused by parasitic protozoans of the genus *Leishmania*. There are more than 20 species of *Leishmania* in the globe, and each one can produce diverse clinical symptoms, which are divided into three categories: cutaneous leishmaniasis (CL), mucocutaneous leishmaniasis (MCL), and visceral leishmaniasis (VL) (43).

CL is the most prevalent type of the disease, it is usually painless but can be painful and persist for years. The sores usually appear where the infected sandfly bit, such as the face, legs, and arms.

MCL is characterized by lesions that can partially or completely damage the mucous membranes of the nose, mouth, and throat cavities, resulting in patient disfigurement and social marginalization (44, 45).

VL is the most dangerous form of the disease, and it is characterized by fever, weight loss, liver and spleen enlargement, pancytopenia, and hypergammaglobulinemia, and it can be deadly if left untreated. Leishmaniasis mostly affects the poor populations of African, Asian, and Latin American countries (44). Chemotherapy is currently the mainstay of antileishmanial treatment (46).

Various medications are available (Table 2), with the choice of medicine varying depending on the type of illness and geographic area. However, the available medications present significant toxicity and little effectiveness in some endemic areas. Furthermore, therapeutic failure, patient recurrence, and fatalities leading to illness complications are common, necessitating the development of new, less harmful therapy approaches (44).

Table 2 summarizes the drugs used against leishmaniasis, their side effects and the status of the resistance.

**Table 2: Drugs used to treat VL and CL and their side effects and resistance profiles**

| Drug                     | Main side effects                                                                                                                      | Resistance                                                                                                                                                      | References       |
|--------------------------|----------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------|
| Pentavalent antimonials  | Headache, myalgia, nephrotoxicity, cardiotoxicity, hepatotoxicity, pancreatitis (47-51)                                                | Yes (multifactorial resistance with different pathways: Entry, metabolism and transport of the drug and also impairment of apoptosis in resistant strains) (52) | (47-52)          |
| Paromomycin              | Reversible ototoxicity, nephrotoxicity, liver enzymes alteration (47, 50, 51, 53)                                                      | No (in vitro resistance only)                                                                                                                                   | (47, 50, 51, 53) |
| Amphotericin B           | Nephrotoxicity, myocarditis, hypokalemia (47, 54)                                                                                      | Yes (rare cases of resistance. Decreased drug uptake following the decline of the affinity to ergosterol) (55)                                                  | (47, 53-55)      |
| Liposomal amphotericin B | Nephrotoxicity (however more rare than with amphotericin B) (47, 53)                                                                   |                                                                                                                                                                 |                  |
| Miltefosine              | Renal and gastrointestinal toxicity (47, 53)                                                                                           | Yes (due to its long half-time, it is subject to the emergence of resistance. Decrease of the uptake of the drug) (53, 56)                                      | (47, 53, 56)     |
| Pentamidine              | Myalgia, syncope, transient hypoglycaemia and hyperglycemia, anemia, leukopenia, cardiotoxicity, nephrotoxicity, pancreatitis (47, 54) | Yes (Decreased accumulation in the cytosol and in the mitochondrion)                                                                                            | (47, 54)         |
| Sitamaquine              | Nephrotoxicity, cyanosis secondary to methemoglobinemia (47, 57)                                                                       | No (in vitro resistance only)                                                                                                                                   | (47, 57)         |

### .2.3 Chagas Disease

Chagas disease also known as American trypanosomiasis is a tropical vector-borne disease caused by the protozoan parasite *Trypanosoma cruzi*. The illness is mainly transmitted to humans by triatomine bugs also called “kissing bugs”. However, Chagas disease can also be transmitted congenitally (from mother to child), through organ transplantation and blood transfusion and even through oral contamination (58, 59). Acute infection can be fatal, and cardiomyopathy affects 25%–30% of those who are infected (59, 60).

In tropical and subtropical nations, particularly in Latin America, it is a neglected illness linked to poverty. People who are impacted reside in rural and peri-urban regions, in substandard housing and in precarious socioeconomic situations. Chagas disease has recently extended to non-endemic countries in Europe, the United States, and Japan, making it a worldwide public health and medical issue (59, 61).

The nitrofurans were launched as an effective class of anti-*T. cruzi* drugs among which nifurtimox stood out for its higher performance. Benznidazole, a 2-nitroimidazole derivative, is also used in the treatment of Chagas disease. Benznidazole and nifurtimox which were both launched in the 70' are still currently WHO-recommended standards of treatment in clinical Chagas disease chemotherapy. Their exact mechanism of action is unknown, but both operate as prodrugs that must be activated by nitro-reductases found in *T. cruzi* in order to cause cytotoxicity (61). The effect of nifurtimox on *T. cruzi* involves redox cycling and radical species, which causes harm to the parasites, by decreasing intracellular thiol levels as well as inducing DNA damage and lipid peroxidation (61).

The treatment schedule for nifurtimox and benznidazole is lengthy, and several side effects might arise, impeding the treatment's success. Gastrointestinal symptoms (nausea, vomiting, and anorexia), signs of central nervous system toxicity (insomnia, irritability, and disorientation) as well as headache, rash, myalgia, arthralgia, dizziness or vertigo, and mood swings are all common side effects of nifurtimox. While allergic dermatitis, nausea, vomiting, anorexia, weight loss, sleeplessness, loss of taste, onycholysis, and dose-dependent peripheral sensitive neuropathy are all typical side effects of benznidazole. Neuropathy and bone marrow depression are also two rare significant side effects. Although both nitroheterocyclic medicines are effective in treating recent stages of infection, their advantages are limited in the chronic phase, with efficacy varying according to geographical location. In addition to the fact that treatment efficiency decreases with the time of the infection pharmacological adverse effects are more prevalent among older patients (61).

Therapeutic failures are frequent and are due to several factors such as genetics of the host and the parasite, treatment compliance and resistance phenomenon. Despite being the first in-line drugs, several cases of resistance due to benznidazole and nifurtimox have emerged

(62). Resistance mechanisms often involve a mutation in the NADH-dependent type-I nitroreductase (TcNTR) which is essential to activate the pro-drugs (63).

*T. cruzi* infection is still treated using medications that were first approved more than 50 years ago, their safety and effectiveness profiles are far from ideal, and are mostly determined by the stage of infection and the patients' age (64). Thus, there is a consensus in stating that sustained drug discovery efforts in Chagas disease area are urgently needed.

As demonstrated by the wealth of literature summarized above, the development of new drugs is urgently needed due to the emergence of resistance to existing therapeutics and the severe side effects of some of the molecules currently used (65).

Nowadays, the discovery of new antiparasitic compounds using phenotypic approaches at the hit and lead stages of drug development is usually followed by the determination of the molecular mechanism of action, the identification of its cellular target for further optimization, and full development as a drug. Indeed, the characterization of the mechanism of action is instrumental for the identification of on and off-targets and therefore facilitates the estimation of the compound's safety profile (18, 66-68). At a later stage, knowing the precise molecular mechanism of action (MoA) impacts positively further development by simplifying the regulatory process, thus limiting unnecessary costs and reducing development time (69). Deconvolution techniques have been developed for this purpose, and these methods as well as those based on metabolomics and computational methodologies are reviewed below.

### **.3 Target deconvolution methods**

The technique of pinpointing the precise target of a biologically active small molecule is known as target deconvolution. This is an important step that greatly supports the drug discovery process since it enables scientists to comprehend how a molecule affects a certain ailment and to create more efficient remedies (18).

Multiple target deconvolution techniques are utilized to determine the targets of small molecules that have proven promise in treating neglected tropical diseases (NTDs)(22, 70, 71). Biochemical assays, molecular modeling, and chemical validation through high-throughput compound library screening are some of these techniques (18).

Furthermore, genetic strategies including RNA interference (RNAi) or CRISPR-Cas9-mediated gene editing are also employed in target deconvolution for NTDs (71). These techniques allow promising therapeutic targets to be identified and molecules that warrant further research to be prioritized.

The upcoming section will provide a detailed overview of three primary deconvolution methods based on different strategies. These methods include Genetic and chemical proteomic-based strategies, Metabolomic-based strategies, and Computational-based strategies. By presenting the defining features and approaches of each method, this section will provide an in-depth understanding of their unique characteristics.

### **.3.1 Genetic and chemical proteomic-based strategies**

#### *.3.1.1 Genetic-based strategies*

##### *.3.1.1.1 RNA interference (RNAi)*

Andrew Fire and Craig Mello were recognized for discovering RNAi, for which they shared the 2006 Nobel Prize in Physiology or Medicine (72). They used double-stranded RNA to silence genes in the nematode *Caenorhabditis elegans* in a work they published in 1998 (73, 74). Since then, RNAi has been extensively used to evaluate the biological role of genes by downregulating their expression (75). Thereafter, small RNAi libraries were generated in order to identify potential targets, and through screening, two kinases of *T. brucei*, CRK12, and ERK8, were identified and appear to play a critical role in the parasite's regular proliferation. This study underscores the possibility of utilizing high-throughput analysis of RNAi library for target finding and deconvolution (76).

Besides the potential for target finding, I would also like to shortly point out that RNAi has emerged as a potential therapeutic strategy for various diseases, such as genetic disorders, viral infections such as human immunodeficiency virus (HIV), hepatitis B and C viruses (HBV and HCV), respiratory syncytial virus (RSV) (77), and specific types of cancer (78, 79). Indeed, several RNAi therapeutics received regulatory approval, including givosiran (brand name Givlaari), approved by the FDA in 2019 to treat acute porphyria (80); patisiran (brand name Onpattro), approved by the FDA in 2018 to address hereditary transthyretin-mediated

amyloidosis (hATTR) (81); inclisiran (brand name Leqvio), approved by the European Medicines Agency in 2020 to treat hypercholesterolemia in adults (82); and lumasiran (brand name Oxlumo), approved by the FDA in 2020 for the treatment of primary hyperoxaluria type 1 (PH1), a rare genetic disease (83). This demonstrates the great potential of RNAi as an effective therapeutic approach for the treatment of a wide range of human disorders in addition to the discovery of promising drug targets (84, 85).

#### .3.1.1.2 CRISPR-based genome editing

The recently identified defense mechanism against mobile genetic components brought in by certain invaders such as phages or invasive plasmids is present in several prokaryotic organisms. This defense system is made up of a special type of repetitive DNA sequences intercut with spacers that are highly variable sequences from phages or plasmids that constitute prokaryotes' immunological memories (77, 86). It also includes the endonuclease protein Cas, which can cut DNA at specific locations using a guide sequence in an RNA duplex (77). This natural defense system called CRISPR/Cas represents a cutting-edge tool that enables accurate and effective genome editing in a range of organisms, allowing researchers to add, remove, or alter particular genes in order to study their function (87, 88). This groundbreaking approach that revolutionized the field of molecular biology, holds new promise for the swift and precise elucidation of the function of any gene (74). This technique demonstrated a significant value in the field of drug target discovery (89). In one investigation, CRISPR-Cas9 screening identified cancer therapeutic targets (90). Six established pharmacological targets and 19 new dependencies were found after screening 192 chromatin regulatory domains in murine acute myeloid leukaemia cells (90). The usefulness of this method in the field of target deconvolution was furthered by another study that demonstrated how the combination of metabolomic analysis and CRISPR interference allowed a rapid and efficient understanding of pharmacological mechanisms of action (89, 91).

In conclusion, several genetic approaches, including siRNA and CRISPR-Cas9, are utilized in the discovery of novel targets, whereas knock-out or knock-in will be more prevalent at the validation stage.

### .3.1.1.3 Genome mining

Genome mining has the potential for discovering new genes but has not been yet used for target deconvolution. In contrast it has been successfully to decipher gene function and used in the frame of drug discovery. Indeed, one of the most important applications of genome mining, which relies on several bioinformatic tools, is the discovery of new genes responsible for the biosynthesis of new bioactive products, such as secondary metabolites thought to be useful molecules produced by some microorganisms, including *Streptomyces*. The majority of these metabolites have shown a key therapeutic value (92). Additionally, when this strategy is combined with genetic engineering, which permits combinatorial biosynthesis and therefore the introduction of structural alterations in molecules that are challenging to reach chemically, new and more effective antibiotics can be designed and generated (93). Furthermore, thousands of genomic sequences from several species, such as bacteria, parasites, and others, are now publicly available. Improvements in high-throughput genome sequencing methods and the volume of DNA data presently accessible led to the development of multiple *in silico* genome mining algorithms and technologies to assist in the discovery and characterization of novel drugs (94). These techniques were used, for instance, to map the genome of *Streptomyces*, which uncovered new silent secondary metabolite biosynthetic gene clusters (smBGCs) that are ready to link with their encoded natural products for the discovery of novel compounds (92).

There are other powerful genetic methods for deciphering gene function. These typically use knock-out and knock-in techniques to suppress, alter, or replace specific genes of interest. After manipulating the genetic sequence, researchers analyze the resulting phenotype, or physical traits and behaviors, to gain insight into the gene's role in biological processes (95). However, these techniques are often applied in research to explore the functions of certain genes, to understand their roles in biological processes, and to validate a potential target but they are not employed yet for target deconvolution.

TbAK is an example of a prospective target validated by the knock-out method. This key enzyme, important in the purine salvage pathway, is hyperactivated by 4-[5-(4-phenoxyphenyl)-2H-pyrazol-3-yl]morpholine and its derivatives, thus paving the way for potential new antitrypanosomal compounds against *T. b. rhodesiense* (96-99). In another

example, researchers used genetic knock-out of the tyrosine aminotransferase (TAT) gene to investigate and better understand its function and to evaluate if this enzyme might be classified as a possible therapeutic target (100). The study showed that the genetic knockout of TAT causes apoptosis and that the life and viability of parasites depend on this enzyme (100).

#### .3.1.1.4 Generation of resistant parasites

Laboratory selection of parasites for drug resistance represents a common strategy for determining the mode of drug action as well as predicting the risk of the development of drug resistance and identifying the various resistance mechanisms (101). The method involves identifying drug-resistant strains by two means: First, *in vivo* by giving sub-curative doses to infected rodents, which are the most closely analogous to humans in terms of circumstance. Second *in-vitro* directly on the strains in culture by using sub-lethal medication doses that can be progressively increased as the parasites become less sensitive in order to impose constant selection pressure (101).

#### .3.1.2 Chemical proteomic–based strategies

##### .3.1.2.1 Affinity-based approaches

**Affinity purification:** This method has been widely utilized to identify target proteins that bind small compounds of interest and to clarify their direct protein binding (102). The affinity purification method involves passing cell lysates or various protein extracts through solid support (column, resins, beads, etc.) pre-immobilized with a ligand of interest. Next, washing procedures are carried out to remove any unbound components that are irrelevant (103, 104). Then, an elution step using specific buffers containing or not the free ligand able to disrupt the immobilized ligand-protein of interest interaction by competition or denaturation is performed in order to release the retained target(s) of interest. Finally, mass spectrometry is usually used to identify the retrieved protein(s) of interest (102-104). These techniques have been employed successfully to determine various protein targets (102, 105). The following articles, among others, provide a great description of the approach (102-104).

One of the targets, such as the *TbAK* enzyme stated above, was discovered by chemical proteomics (affinity chromatography combined with mass spectrometry) and was then

genetically confirmed by the RNAi method by generating an adenosine kinase knock-down mutant of the parasite (98).

**Three-Hybrid Systems:** These systems represent powerful tools for studying small molecule-protein interactions *in vivo*. They have been developed in yeasts and bacteria and usually take two-hybrid systems as a starting point. The brilliant strategy behind two-hybrid systems came after witnessing that the DNA-binding domain and the domain with activating regions of the transactivator protein of the yeast *Saccharomyces cerevisiae*, GAL4, can be physically separated and interact indirectly through the interaction of a protein 'X' attached to the DNA-binding domain and a protein 'Y' attached to the domain with activating regions (106). The aim is then to detect the interaction between the two proteins 'X' and 'Y' (106)'.

To turn the two-hybrid system into a three-hybrid system, a third element had to be added which is a synthetic hybrid ligand composed of an anchor molecule linked to the molecule of interest. The protein 'X' becomes a specific receptor for the anchor molecule, this part of the system is constant and represents an anchor. The protein 'Y' becomes the unknown target to identify, taken from a cDNA library fused to the domain with activating regions. Following an interaction between 'Y' and the molecule of interest, the transcription of a reporter gene is activated (107).

This system has proven useful in the identification of drug targets. Using a Yeast three-hybrid approach, the mechanism of action of the anti-inflammatory drug sulfasalazine was studied. It inhibited the biosynthesis of the cofactor tetrahydrobiopterin (108). Another study identified the procollagen-lysine 2-oxoglutarate 5-dioxygenase 2 (PLOD2 or lysyl hydroxylase, LH2) as a novel target of ethinylestradiol (109). This method has not yet been used for the discovery of new targets for Sleeping sickness, Chagas disease, and Leishmaniasis.

Within the thesis, the tools developed further from the Three-Hybrid Systems principle will be presented and discussed in depth in part 3.1.3.

#### .3.1.2.2 Non-affinity-based approaches

The need to modify each drug individually (without losing bioactivity) limits current affinity-based target identification techniques, whereas indirect, non-affinity-based approaches rely on the drug's ability to induce specific biochemical or cellular readouts (110).

**Drug affinity-responsive target stability (DARTS):** The interaction between small molecules and target proteins is usually formed by ionic bonds, hydrogen bonds, van der Waals forces, and other intermolecular forces. The DARTS approach is based on the fact that drug binding is thought to stabilize target proteins making them less prone to protease degradation. It can detect drug-target interactions in cells by monitoring changes in the stability of proteins (110). When the protein binds to a small molecule, its stability usually increases compared to the apo form. The binding of a small molecule to target proteins can be captured by comparing their enzyme hydrolysates after separation by SDS-PAGE (111). The binding partners of the drug disulfiram, the factors IQGAP1 and MYH9 were identified using DARTS (112, 113).

**Proteomics:** This approach is a growing area of protein biochemistry and a high-throughput, systematic, and holistic science (114). Classical protein investigations have been of great help and provided rich information about individual proteins. However, regarding the complexity of biological systems, it does not take into consideration the broad network linked to the protein (115). Proteomics, in contrast, highlights the dynamic nature of proteins and allows studying protein-protein, as well as small molecule-protein interactions (115). In the proteomic area, it is important to discern two main objectives. The first one aims to do a “protein profiling” by measuring the abundance, modification, activity, localization, and interaction of all the proteins in a sample. It is particularly useful to characterize biomarkers in disease. Coupled with LC-MS/MS, it led to the characterization of two proteins, neuroserpin, and moesin, which present an interest in the understanding of sleeping sickness pathophysiology and diagnosis (116). The second objective is designated as “functional proteomics” and is broadly used in drug discovery during the step of target identification allowing the identification of putative substrates or putative interactions between proteins (115). A subtype of functional proteomics is chemical proteomics, the latter allows using a small synthetic molecule to seek for the mode of action and protein function (117). In the proteomic field, we also distinguish two aspects, the discovery-oriented referred to as “unbiased” and the system-oriented referred to as “targeted”. In the first category, a complex biological sample is analyzed by separating and identifying as many proteins as possible, using, usually, two-dimensional gel electrophoresis, in-gel digestion, and mass spectroscopy in

tandem coupled to liquid chromatography (LC-MS/MS) (115). Therefore, “unbiased” proteomics allows the identification of both known and unknown proteins. On the contrary, the second category is oriented, known proteins related by sequence, biological function or diagnostic potential are analyzed and quantified (115). Exploiting proteomics, new antigens and target proteins have been identified and whole proteomes of etiological agents are being investigated (118).

**Activity-based protein profiling (ABPP):** The ABPP, which consists of a reactive group (reactive electrophile) required for covalent attachment to the active site, a linker to direct particular probes to specific enzymes, and a labeling group (fluorophore, biotin, clickable group) for target labeling. This approach relies on the use of this probes that can be used to isolate targets and formally define the proteins that bind to them, depending on the activity of small molecules of interest. The foundation concerning the activity-based probes (ABPs) is the use of covalent inhibitors that can immediately mark the target proteins, as opposed to affinity-based probes, which use non-covalent inhibitors. *In-situ* formation of stable covalent complexes from non-covalent contacts is made possible by the photoreactive group's capacity to produce a highly reactive intermediate in response to UV light irradiation (119) (120). Protein concentration and proteomic screening of the target proteins can be carried out after stable probe-protein complexes are established *in-situ* (119).

### .3.2 Metabolomic-based strategies

Metabolomics aims to identify and quantify small biochemicals (including metabolic intermediates, and secondary metabolites) within a given biological sample. Measurements of the metabolome can be used to detect subtle changes, and the study of metabolic changes within parasites after drug perturbation provides tools for analyzing the mode of drug action. In addition, knowing the concentrations of metabolites in the various life cycle stages of the parasites will help in the building of *in silico* models of parasite metabolism that will further be able to predict drug targets and resistance mechanisms (121, 122). Metabolomics can be divided in two categories: untargeted metabolomics which consists of the scan of all detectable metabolites (in intra- and/or extra-cellular matrices), and targeted metabolomics (quantification of preselected metabolites). Mass spectrometry (MS) and nuclear magnetic resonance (NMR) are the most used analytical methods in metabolomics, and MS is generally

hyphenated to a separation technique such as gas chromatography (GC), liquid chromatography (LC) or less frequently capillary electrophoresis (CE).

One of the first studies using LC-MS based non-targeted metabolomics approach to highlight the drug targets of anti protozoan agents was a study of the combination of nifurtimox and eflornithine against *T. brucei*. After treatment with a sub-IC<sub>50</sub> level of eflornithine (20 µM), as expected ornithine was found to increase significantly, whereas putrescine was found to decrease significantly. Nifurtimox was shown to be converted to a trinitrile metabolite indicative of metabolic activation, as well as inducing changes in levels of metabolites involved in carbohydrate and nucleotide metabolism (123). The metabolism of fluorinated pyrimidines in *T. brucei* was analysed by LC-MS (124). The study provided information on the MoA of these analogs, as well as information on the reversibility of the enzymes in pyrimidine biosynthesis, providing evidence that a uridine phosphorylase was expressed in the bloodstream-form of the trypanosomes (125). A novel benzoxaborole derivative (AN5568) was selected as a lead compound for the treatment of HAT and is currently being investigated in phase III clinical trials in the Democratic Republic of the Congo and Guinea (126). An LC-MS metabolomics-based approach showed that the treatment of blood-stage trypanosomes with AN5568 led to significant perturbations in the metabolism of the parasite. In particular, elevated levels of metabolites involved in the metabolism of S-adenosyl-L-methionine, an essential methyl group donor, were found (127).

The first experiment of untargeted LC-MS metabolomics on *T. cruzi* exposed the parasites to benznidazole, and revealed that cellular thiols such as trypanothione, homotrypanothione, and cysteine were depleted during the treatment. Numerous metabolites of benznidazole were also detected, as well as covalent conjugates of the drug combined with low molecular weight thiols and some non-thiol metabolites. Metabolomics pointed out that benznidazole primarily affected thiol-containing molecules in *T. cruzi*, and this interference with thiol metabolism contributed to the MoA (128). In a recent study (129), metabolic signatures of myoblasts infected with *T. cruzi* after treatment with six different drugs (benznidazole, nifurtimox, Posaconazole and three experimental drugs) were obtained by direct infusion to mass spectrometry (DI-MS) and NMR. The fingerprinting approach resulted in three different

clusters. Two could be explained by already known MoA for benznidazole, nifurtimox, and posaconazole, whereas the three experimental drugs formed a separate cluster. The latter impact on the tricarboxylic acid cycle, but no specific pathways could be attributed to drug action, which might be caused by a high percentage of common metabolome between a eukaryotic host cell and a eukaryotic parasite. Nevertheless, the authors concluded that this metabolic fingerprinting approach to a complex host-cell parasite system is validated and can potentially be applied in the early stage of drug discovery and could help to prioritize early leads or reconfirmed hits for further development.

Several untargeted metabolomic studies were led on *Leishmania* during the last decade with various success rates. Drug targets and resistance of *Leishmania* parasites towards several well-established drugs (pentavalent antimonials, miltefosine, pentamidine) were studied (130-135). More recently, a tentative to elucidate the MoA of methyldehydrodieugenol B, a natural compound isolated from *Nectandra leucantha*, towards *Leishmania infantum* via an untargeted metabolomics approach was carried out. Due to the complexity of the parasite metabolism and the great diversity of altered metabolites, a multi-target mechanism was assigned to the drug candidate (136).

### **.3.3 Computational-based strategies**

Wet-lab methodologies can be combined with computer-aided drug design (CADD) to address target deconvolution (137). In detail, CADD techniques can describe an atomic level structure-activity relationship (SAR) optimizing the drug design process (138). CADD methodologies are divided into ligand-based drug design (LBDD), which consider the features of known ligand for a target of interest, and structure-based drug design (SBDD), which takes into account the structural information of the drug-target interaction. Most of these approaches are based on the concept of chemical similarity: similar molecules cause an analogous biological response, and on the other hand, similar proteins bind comparable ligands (139).

#### **.3.3.1 Ligand-based computational approaches for Target Prediction**

Regarding target deconvolution strategies, LBDD methods aim to predict the ligand activity for an ensemble of targets (139). The simplest LBDD strategy is based on the similarity rank of

known compounds assuming that the target tested of the most similar compounds is also the most likely target of the query compounds. The first step to compute a similarity index is the representation of the ligands under investigation by molecular descriptors or chemical fingerprints. The molecular descriptors can be grouped in one-dimensional, two-dimensional, and three-dimensional (140-144), based on the molecular information represented. Instead, chemical fingerprints model the molecular properties into a bit string, allowing a fast comparing among compounds (145). Consequentially, each compound can be compared and ranked using metric algorithms such as Tanimoto coefficient (146) which is one of the most popular. The above-mentioned methodologies were exploited by web server tools such as Polypharmacology Browser(147), SpiDER(148), SuperPred(149), HitPick(150), Swiss Target Prediction (STP)(151), Prediction of Activity Spectra for Biologically Active Substance (PASS)(152), Similarity Ensemble Approach (SEA)(153), MOst-Similar ligand-based Target inference approach (MOST)(154), Candidate Ligand Identification Program (CLIP)(155) and Chemical Similarity Network Analysis Pulldown (CSNAP), proposing protein targets based on the similarity of the compounds (139, 156). Target prediction can be performed by machine learning (ML) classification that requires an interaction threshold (e.g., pKi, pEC50) or by ML regression making quantitative binding affinity predictions (e.g., Ki). Based on the protein-ligand predicted bioactivity, multiple candidate proteins were ranked, and the most likely target was identified. These methods are named quantitative structure-activity relationship (QSAR) models and exploit ML algorithms such as Gaussian processes (GP), support vector machines (SVM)(157), artificial neural networks (ANN), Bayesian artificial neural networks (BRANN), decision trees (DT), random forests (RF)(158), Naïve Bayes (NB)(159), and k-nearest neighbour (kNN) clustering methods(160). Recently, D. A. Winkler summarized the main applications of artificial intelligence and machine learning drug discovery tools for neglected tropical disease (161).

### *.3.3.2 Structure-based computational approaches*

The SBDD strategies identify and optimize the ligand-receptor interactions by exploiting 3D protein structural information obtained from experimental structures available in the Protein DataBank (PDB, <https://www.rcsb.org>)(162), or by comparative protein structure model in the ModBase (<https://modbase.compbio.ucsf.edu>)(163). Structure-based methods suffer from

the lack of experimental 3D structures, and this is especially true in the case of trypanosomatids(145). This deficiency can be partly overcome by using homology modeling techniques that reconstruct the 3D target protein structure from its amino acid sequence. Homology modeling consists of three main steps: i) choosing an evolutionarily related template protein with experimental structure available ii) alignment of the template and target amino acid sequence ensuring a high amino acid sequence similarity, especially at the binding site iii) reconstruction of the secondary and tertiary target structure. The quality of the homology model is a critical aspect and must be verified before the application of SBDD methods. Generally, a sequence identity greater than 40% is considered acceptable (164). Modeller(165), Swiss-Model(166), I-TASSER(167), and PROCHECK(168) represent the most used tools to address the homology modeling issue. They allow the computation of quality assessment scores for the homology model based on geometric indicators (bond lengths, bond angles), Ramachandran plots, physics-based energy functions, and statistical potentials(169-173). Regarding a target deconvolution perspective, SBDD methods highlight the most likely targets for a query ligand or the most similar targets for a query target (174), also named reverse screening. Reverse screening methods can be divided into three major classes: pharmacophore screening, shape screening and reverse docking (175, 176). Shape and pharmacophore screening methods, in the absence of target structure, describe the binding pocket comparing the overall shape or main biochemical features of query ligand based on compounds databases with target information, allowing target identification (177-179). The most used shape similarity descriptor 2D is FingerPrint2D (FP2) which encodes many chemical features towards a bit vector such as extended-connectivity fingerprints (ECFPs) and molecular access system (MACCS). In this context, some of the most used shape-screening tools are ChemProt(180), ROCS(181), ChemMapper (182), and SEA (183).

Pharmacophore screening describes the functional features of molecules with target proteins in a spatial arrangement. The main pharmacophore model characteristics are the hydrophobic center, hydrogen bond donor vector, hydrogen bond acceptor vector, positively charged center, and negatively charged center. The combination of the pharmacophore features for a specific compounds-target protein ensemble gives the pharmacophore model. In the same way as shape screening, the success of pharmacophore screening depends on a reference

database with protein annotations. In the pharmacophore screening landscape, the most widely used web server is PharmMapper(184-186), which maps over 7000 pharmacophore templates and ligand databases extracted from all the targets in TargetBank, DrugBank(187), BindingDB(188), and Potential Drug Target Database (PDTD)(189) to determine the target protein based on query molecules.

Reverse docking predicts the binding pose and estimates the binding energy of ligand-receptor complexes requiring 3D protein structures to identify targets with the highest binding affinity of the query ligand(175). First, each docking tool requires target pre-processing to reach an efficient comparison, encoding the binding site or the entire target. The query molecules are docked for each protein target in the database. Finally, the docking poses are scored using binding energy functions. Generally, reverse docking is a more complex process than shape and pharmacophore screening that requires a large dataset of protein structures, the binding site recognition, and the pre-processing of the target and query molecules which depends on the docking tool and a score function to rank the predicted targets. The main reverse docking engines available in literature are TarFisDock(190), idTarget(191), VTS(192), VinaMPI(193) and IFPTarget(194). Likewise, web software was developed for reverse screening as SEA(183) (Keiser et al., 2007b), PharmMapper(184), and INVDOCK(195).

### *.3.3.3 Molecular dynamics*

In a target deconvolution perspective, an objective is elucidating the molecular mechanism of query compounds. In this context, molecular dynamics (MD) simulations, can predict and characterize the target-drug interactions (196). In drug discovery, MD simulations are useful tools regarding binding free energy estimation at the molecular level (197). In the last decades, several techniques were developed such as free energy perturbation (FEP)(198), umbrella sampling (199), funnel metadynamics (200), and thermodynamic integration (TI)(201). However, the above-mentioned methods have a high computational cost and hence an imbalance between accuracy and efficiency. In this context, endpoints methods such as molecular mechanics Poisson–Boltzmann surface area (MM/PBSA), molecular mechanics generalized Born surface area (MM/GBSA)(202, 203), and linear interaction energy (LIE)(204) represent valid alternatives to achieve a good compromise between accuracy and

computational cost (203). Recently, hybrid approaches involving MD simulations and ML techniques are being developed to optimize computational resources (205).

#### *.3.3.4 Databases for neglected diseases*

Target deconvolution strategies make extensive use of protein and ligand databases that describe ligand structure and pharmacophore or protein structures. Generally, each web server application of target deconvolution has its own database, mostly constructed from public databases such as ChEMBL, PubChem, DrugBank, BindingDB, KEGG, and PDB. Regarding NTDs, the special program for Research and Training in Tropical Diseases (TDR)(206) released the TDR targets database (<http://tdrtargets.org/>), an online open-access framework that contains chemogenomic information for both pathogens and non-pathogenic model organisms and toward a set of informatic tools that help the identification and prioritization of drug targets for neglected tropical diseases, such as HAT, Chagas Disease, and Leishmaniasis. Similarly, the Target-Pathogen database (207) (<http://target.sbg.qb.fcen.uba.ar/patho>) is an online resource focused on neglected tropical diseases that collects structural information of proteins such as function, metabolic role, off-targeting, structural properties including druggability, essentiality, and omic experiments, facilitating the identification of drug targets. Likewise, the Eukaryotic Pathogen, Vector and Host Informatics Resource (VEuPathDB, <https://veupathdb.org>)(208) supports more than 500 organisms, integrating data from public repositories. Furthermore, VEuPathDB integrates bioinformatics tools for analysis, data mining and standardized workflows such as OrthoMCL for predicting orthology.

To summarize, target deconvolution utilizes a range of methodologies, such as proteomics, metabolomics, and computational methods along with various affinity-based techniques. These methods can lead to the identification of the target(s) of a drug or compound within a complex system, without modifying the system itself. Contrarily, genetic approaches, except for RNAi, are more commonly used for the validation stage since they modify the system, which may add uncontrollable factors. Among this knock-out and knock-in procedures, which entail the deletion or addition of a DNA sequence of interest are widely used to study the impact on cellular or organizational function.

## **.4 SWOT analysis**

To aid gaining a better understanding of the impact of each technique and its significance in target deconvolution a comprehensive SWOT (Strengths, Weaknesses, Opportunities, Threats) analysis of each method, detailing their strengths, weaknesses, opportunities, and threats has been done and is reported in a summarized manner in Table 3 and discussed more exhaustingly in the next chapters.

Table 3: SWOT analysis of different deconvolution techniques and examples

| Deconvolution tools                                                                                                                                                                                                                                                   | Strengths                                                                                                                                                                                                                                                                                              | Weaknesses                                                                                                                                                                                                                                                                                                                                                                                                    | Opportunities                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | Threats                                                                                                                                                                                                                                                                     | Examples of identified or validated Targets                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Genetic approaches</b><br>Reverse genetics (RNAi, CRISPR-Cas9),                                                                                                                                                                                                    | <ul style="list-style-type: none"> <li>- Allow identification of suitable targets</li> <li>- Simple, inexpensive, and easy to execute in the lab (22, 209) (210)</li> </ul>                                                                                                                            | <ul style="list-style-type: none"> <li>- Suitable knock-down system (RNAi) must be available</li> <li>- the low or non-existent expression of certain orphan gene clusters (211)</li> </ul>                                                                                                                                                                                                                   | <ul style="list-style-type: none"> <li>- Can be used in conjunction with other techniques, such as bioinformatics tools, AI, robotics, HTS.</li> <li>- Validation of proteins involved in the development of resistance (e.g. <i>TbAAT6</i>, <i>TbAT1/P2</i> adenosine Transporter) (22) (35, 212)</li> <li>- Vast DNA databases have been generated and are freely accessible (211)</li> <li>- QSAR studies can be achieved to discover various interactions (22)</li> <li>- Allow validation of a target (Knock-out/ Knock-in)</li> </ul> | <ul style="list-style-type: none"> <li>- Not always suitable to predict the biotechnological potential of natural compounds (209).</li> <li>- May only be applied to known biosynthetic gene clusters (209)</li> </ul>                                                      | <ul style="list-style-type: none"> <li>- <i>T.b.rhodensiense</i> adenosine kinase (<i>TbrAK</i>) validated by RNAi (96)</li> <li>- <i>T.b. Glutathione Synthetase (TbGS)</i>: identified as a potential drug target (213)</li> <li>- <i>T.b. ornithine decarboxylase</i> (214, 215) (<i>TbODC</i> knockdown)</li> <li>- <i>TbAT1/P2</i> Adenosine Transporter, putative amino acid transporter <i>TbAAT6</i></li> <li>- <i>CRK12</i> and <i>ERK8</i>: two kinase targets in <i>T. brucei</i> (76)</li> </ul> |
| <b>Affinity chromatography</b><br>linked to the proteomic approach for targets identification (the molecule of interest is attached to a matrix and targets (DNA, RNA, proteins coming from crude parasite extract or a specific fraction) are eluted and identified) | <ul style="list-style-type: none"> <li>- Easy, fast, and reproducible method</li> <li>- Soluble targets may be easily isolated and then identified by mass spectrometry (22)</li> <li>- Different types of proteins can be studied (membrane proteins, enzymes, cytoskeletal proteins) (22)</li> </ul> | <ul style="list-style-type: none"> <li>- Irrelevant proteins may bind to drug-matrix unspecifically</li> <li>- Transient interactions and insoluble protein targets may be missed (216)</li> <li>- Target protein may be lost during elution due to non-covalent interactions (105)</li> <li>- In the case of prodrugs, the enzyme of metabolism is identified rather than the actual target (217)</li> </ul> | <ul style="list-style-type: none"> <li>- The isolated protein can be cloned and overexpressed in bacteria or eukaryotic cells (22)</li> <li>- Specific proteins found in particular organelles such as the kinetoplast or apicoplast can be purified for further investigations (e.g. <i>T. brucei</i> Tb927.2.6100 protein, tryptophanyl-tRNA synthetase (TrpRSapi) (218) (209)</li> <li>- Off-target proteins can be identified</li> </ul>                                                                                                | <ul style="list-style-type: none"> <li>- Limited specificity (104)</li> <li>- Ligand availability (219)</li> <li>- Difficulty in positioning the linker (104)</li> <li>- Complex mixture</li> <li>- Proteins denaturation (220)</li> <li>- High background noise</li> </ul> | <ul style="list-style-type: none"> <li>- <i>T. brucei</i> glycerol-3-phosphate dehydrogenase (221)</li> <li>- <i>T. brucei</i> lapatinib-binding protein kinases (<i>TbLBPKs</i>) (222).</li> <li>- <i>Alba-domain</i> Proteins of <i>T. brucei</i> (223)</li> </ul>                                                                                                                                                                                                                                         |

Table 3: Continuation of the table

| Deconvolution tools                                                                                                                                         | Strengths                                                                                                                                                                                                                                                                                                                                                       | Weaknesses                                                                                                   | Opportunities                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | Threats                                                                                                                                                              | Examples of identified or validated Target                                                                                                                                                                       |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Generation of resistant parasites</b><br>Generation of resistant strains by either exposing them to higher drug concentrations or through mutations (22) | - Identification of drug resistance mechanisms (22)                                                                                                                                                                                                                                                                                                             | - Difficulty to obtain resistant strains and lengthy process<br>- Non-target proteins may be discovered (22) | - Pinpointing the resistance mechanism to guide first-line therapy (35) (212)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | - Obtaining a compound-resistant strain is a time-consuming process                                                                                                  | - <b>TbAT1/P2 adenosine transporter</b><br>- Putative amino acid transporter <b>TbAAT6</b>                                                                                                                       |
| <b>Activity-based protein profiling (ABPP)</b>                                                                                                              | - Enable the swift identification of lead compounds with the potential to treat human ailments (119).<br>- Allow the investigation of the effects of enzymes and small compounds in natural conditions (119, 224)<br>- Enable the understanding of the mode of action and evaluation of potential side effects (119)<br>- Monitor the activity of enzymes (225) | - Ineffective crosslinking (119)<br>- ambiguous labeling (119)                                               | - A library can be used directly for phenotypic screening and target identification (225)<br>- Combining activity-based probes (ABPs) and a covalent inhibitor can facilitate target identification and enable investigation of the function and mechanism of the identified proteins (225-227).<br>- Addition of a clickable group on the ABPP can smoothen the passage from phenotypic screening to target identification (225)<br>- Understanding of the pharmacological effects and possible adverse events facilitated by the use of ABPP in conjunction with bioimaging studies (226)<br>- A strong technique for identifying proteins involved in disease (225) | - An active nucleophile site must be present in the enzyme's active site (226)<br>- Difficulty to discriminate between generic labeling and target involvement (119) | - <b>TgDJ-1, a crucial protein involved in the invasion process of the host cell by <i>Toxoplasma gondii</i> (68)</b><br>- <b>Dipeptidylpeptidase 1 (DPAP1) as a key player in malarial infection (225, 228)</b> |

Table 3: Continuation of the table

| Deconvolution tools       | Strengths                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | Weaknesses                                                                                                                                                                                                                                                                                                                                                                                                                        | Opportunities                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | Threats                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | Examples of identified or validated Target                                                                                                                                                                                                                                                                                                                                                              |
|---------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>In silico approach</i> | <ul style="list-style-type: none"> <li>- Drug candidates can be predicted</li> <li>- Prediction of drug-target interactions at atomistic resolution</li> <li>- Identification of relevant features of the protein-ligand biological activity by ML (139, 229)</li> <li>- Protein conformation flexibility is considered by MD simulations (139, 230).</li> <li>- The visualization of the molecular mechanism through MD simulations can enhance the efficiency and knowledge of the design process.</li> </ul> | <ul style="list-style-type: none"> <li>- Molecular docking algorithms neglect or considerably simplify the protein conformation flexibility</li> <li>- Advanced sampling methods (FEP, umbrella sampling, funnel metadynamics, TI, etc. ) have a high computational cost.</li> <li>- Low performance of the ligand-based target prediction if the number of known ligands is small or activity cliffs (156, 174, 231).</li> </ul> | <ul style="list-style-type: none"> <li>- Availability of libraries with structures, activity, and targets of small molecules in public databases</li> <li>- Investigation of the mechanism of action of hits discovered from phenotypic screens (68)</li> <li>- Availability of a large range of applications (232).</li> <li>- Prioritizing and narrowing down the screening process(233).</li> <li>- Learning existing features of known ligands and targets to predict drug-target interactions by ML, DL, and AI (233).</li> </ul> | <ul style="list-style-type: none"> <li>- Useless prediction if the drug-target affinity is weak (233)</li> <li>- Unavailability of 3D structures of some targets (233).</li> <li>- Proper construction of target structure databases, correct identification of binding pockets, computational efficiency, and proper normalization of docking scores(230) can be an issue in reverse docking - Low accuracy of the binding energy estimation by docking scores (139, 230).</li> </ul> | <ul style="list-style-type: none"> <li>- Lung-intrinsic M1-like-driven cytokine as a target in the pathogenesis of bronchopulmonary dysplasia (234)</li> </ul>                                                                                                                                                                                                                                          |
| <i>Metabolomics</i>       | <ul style="list-style-type: none"> <li>- Simple to execute (121)</li> <li>- Prioritization of leads and confirmation of hits during the early stage of drug discovery (129)</li> </ul>                                                                                                                                                                                                                                                                                                                          | <ul style="list-style-type: none"> <li>- Difficulty to validate some biomarkers (235)</li> <li>- The annotation of unknown compounds is a time-consuming and labor-intensive task (236).</li> <li>- Efficient data deconvolution is required (121).</li> </ul>                                                                                                                                                                    | <ul style="list-style-type: none"> <li>- Using mass spectrometry and NMR as complementary technologies (237)</li> <li>- A single analytical technique is not able to scan the entire metabolome, hence multiplatform assemblies are usually preferred (136)</li> <li>- Great possibilities since the size of the metabolome is not well defined (238)</li> </ul>                                                                                                                                                                       | <ul style="list-style-type: none"> <li>- The use of orthogonal techniques is mandatory (237).</li> <li>- Statistical data is key (239).</li> <li>- A high demand for analytical performance is required (136).</li> <li>- Multiple profiling experiments are required (238).</li> </ul>                                                                                                                                                                                                | <ul style="list-style-type: none"> <li>- A benzoxaborole derivative (AN5568) perturbed the metabolism of <i>T. brucei</i> (127)</li> <li>- Eflornithine increased ornithine and decreased putrescine (240)</li> <li>- Nifurtimox activated the metabolism and induced changes in levels of metabolites (240)</li> <li>- Benznidazole affected molecules with thiols in <i>T. cruzi</i> (129)</li> </ul> |

## **.4.1 STRENGTHS**

Many innovative and effective deconvolution methods have been developed over the past few years. These made possible to then study drug-binding to ascertain the activity of the studied drug on the target of interest, as well as the mechanism of action (MoA) via the overexpression of a protein of interest in a specific organism or cell. These facilitated our understanding of the MoA interactions between a drug and its target.

The primary and common advantages of all these strategies are that they are usually straightforward, affordable, and easy to implement in the lab (22, 209).

In the following section the key benefits of each approach, briefly summarized in Table 3, are further described and discussed.

**Genetic approaches:** By examining changes to the phenotypes of cells and organisms containing mutations in their genomes, reverse genetics helps scientists comprehend how genes function. Additionally, the major sequencing effort and the emergence of databases containing whole genomic sequences of many parasite strains, notably the TriTrypDB database, made a large amount of data publicly available (6). With the help of new computational methods, complete and accurate target identification, as well as the search for potential targets of known drugs, may be possible in addition to the identification of parasite proteins that may interact with known drugs (6). Furthermore, genome mining, an interesting tool for drug discovery, led to the identification of numerous microbial genomes that contain orphan (cryptic) gene clusters that encode enzymes involved in the synthesis of new secondary metabolites paving the way for the discovery of new putative active compounds against bacteria and parasites in the future (210, 241) (242). For instance, the discovery of coelichelin, a cryptic modular non-ribosomal peptide synthetase (NRPS) produced by *Streptomyces coelicolor* A3(2)'s genomic sequence (211).

Last but not least, the most significant benefit of the direct genetic approach is the investigation's objectivity, which does not necessitate making any hypotheses about the molecular origins of the phenotype under consideration. Indeed, direct genetics produced several novel and unforeseen breakthroughs as a result. An illustration of such is the identification of a Toll-like receptor (TLR)-4 as the lipopolysaccharide sensor (95, 243).

Finally, these genetic approaches have been widely used not only for the discovery but also for the validation of therapeutic targets through gene insertion or deletion (88, 244). An additional example of a target that has been validated using genetics approaches following identification by targeted deconvolution techniques is the *TbAK* (as previously described) which was pinpointed by chemical proteomics and validated genetically using RNAi as the intracellular target of 4-[5-(4-phenoxyphenyl)-2H-pyrazol-3-yl]-morpholine (96, 97, 245).

**Affinity chromatography:** Given its ease of use, affordability, and capacity to develop drug-binding assays based on their attachment to a suitable matrix, this approach is frequently utilized for identifying possible targets interacting with promising compounds. Additionally, these processes are fast and reproducible, generally easy to set up, and need little in the way of equipment (affinity column, resins, beads) (22). These techniques also facilitate the isolation and identification of soluble targets by mass spectrometry. As a result, multiple protein types can be studied (enzymes, and membrane proteins as well as some cytoskeleton proteins after adaptation of the extraction conditions) (22).

**Generation of resistant parasites:** Different proteins involved in the development of drug resistance can be determined and studied (22).

**Activity-based protein profiling:** The ability to use this technology to examine the actions of enzymes and small compounds in natural surroundings along with its great potential to allow the study of *in vivo* proteins represents its most significant benefits. Therefore, its implementation into the early stages of the pharmaceutical development process has been mentioned to be facilitate the discovery of new therapeutic targets (119, 224). Moreover, when it is paired with bioimaging, it creates a potent approach that offers a global and quantitative evaluation of the proteins that bioactive compounds bind to (119).

**Metabolomics:** Untargeted metabolomics can be applied in the early stage of drug discovery and is intended to help prioritize leads or reconfirmed hits for further development (129). Moreover, laboratory experiments in the metabolomics workflow are relatively simple. Drugs are added to cells, and cell extracts are analyzed using MS or NMR, in which the levels of each metabolite can be recorded (121).

***In silico* approach:** The strength of the *in silico* approaches is to be able to mine the structural, chemical, biological data provided by years of research towards finding the target of a given molecule. In this context, reverse docking, ligand-based approaches complemented by machine learning and molecular dynamics can be defined as strengths intrinsic strengths of the approach. Machine learning methods statistically analyse the correlation between chemical structures and protein-ligand interaction and have demonstrated their ability to identify and predict relevant features of the protein-ligand biological activity (139, 229).

Molecular dynamics simulations provide an accurate description of the protein conformation flexibility which usually is neglected in reverse docking (139, 230). Furthermore, observing a specific molecular mechanism can enhance the efficiency and knowledge of the drug design process, enabling intuitive breakthroughs beyond mere interpretation and offering comprehensive insights that surpass the scope of the specific application under examination.

## **.4.2 WEAKNESSES**

Deconvolution approaches offer many benefits (Table 3), but some of them have drawbacks as well. These limitations are also intrinsically linked to the complexity of the task. Indeed, the target of interest must be specifically expressed in the pathogen, necessary for parasite life but distinct from that of the mammalian host cell, in order to develop effective antiprotozoal medications for neglected tropical diseases (21). This complexity may be also turned into an opportunity as the differential activity between host and pathogen can be exploited for finding the relevant target(s).

In the following section the key weaknesses of each strategy, briefly summarized in Table 3, are further described and discussed.

**Genetic approaches:** While genetic techniques are continuously evolving and improving (209), a suitable knock-down system for target discovery must typically be available to execute reverse genetics. Reverse genetics has other weaknesses, including technical difficulties, particularly in organisms that are complicated to genetically modify, including humans, some animals, and multicellular plants that have polygenic characteristics along with massive genomes (246) (247). Additionally, considering the phenotype, if it is intricate or pleiotropic (impacting various traits), it may be challenging to pinpoint the particular role of a gene.

Undoubtedly, a single gene can have multiple purposes, and its deletion or alteration can have an impact on a variety of characteristics (248). Last but not least, if a protein is universally expressed, the phenotype brought on by gene disruption can be associated with almost any type of cell or tissue (249).

Genome mining has provided some strength for drug discovery but has limitations when the expression of the relevant orphan gene cluster is either absent or low (211). Moreover, the identification of biosynthetic gene clusters is limited to those that are already known (210, 241). Furthermore, there remains a challenge in connecting these gene clusters with specific biological functions (250). Finally, while genome mining is rapidly evolving and starts giving results for drug discovery it needs more refinement and improvement for being used as primary target deconvolution method.

**Affinity chromatography:** This strategy has certain drawbacks, including the possibility of non-specific binding of irrelevant proteins to the drug matrix. The potential identification of secondary targets, such as transporters or enzymes, instead of the intended target is a weakness but at the same time represent an opportunity to better understand the pharmacological profile of the compound. Furthermore, when using compounds that in reality are prodrugs activated in situ, the target may not always be the intended protein, but rather the metabolizing enzyme that transforms the prodrug into the desired molecule. In this situation, the appropriate compound should be used (217).

Moreover, as the technique requires the use of strong detergents to solubilize the proteins, transient interactions, and insoluble targets (such as trypanosome cytoskeleton proteins) may go unnoticed in addition to the potential loss of the target protein during elution (216). Besides, rigorous washing will lessen the ability to recognize protein complexes with which the target of interest is directly interacting, which could lead to the loss of crucial information (105). Also, the generation of immobilized affinity agents that preserve cellular activity is another difficult aspect of affinity purification procedures (105).

**Generation of resistant parasites:** The major weakness of this method is the possibility of directing the search to other proteins involved in drug metabolism or transport rather than to the actual target (22)

**Activity-based protein profiling:** Although ABPP technology led to enormous strides and substantial breakthroughs, it is still difficult to discern the distinction between nonspecific labeling and real drug-target interaction (119). Moreover, the underlying problems with inadequate cross-linking still provide a significant difficulty, notwithstanding the successful uses of this method (119).

**Metabolomics:** Efficient data deconvolution is required, and the annotation of unknowns compounds remains a bottleneck in untargeted metabolomics, as it is a time-consuming and labor-intensive task (236). Furthermore, the validation of biomarkers is still quite challenging since some studies suffer from small size and a design not sufficiently specific for this task (235).

**In silico approach:** Molecular docking has limitations in accurately predicting the binding affinity of small molecules to target proteins. The accuracy of the predicted binding poses can also be affected by the quality of the protein structure and the flexibility of the ligand and receptor. The ligand-based target prediction strategy suffers from some limits in case of targets with a small number of active ligands or activity cliffs that are ligands pairs of compounds with high structural similarity but unexpectedly high activity difference, affecting the generalization ability of machine learning models (156, 174, 231). Finally, enhanced sampling methods such as FEP, umbrella sampling, funnel metadynamics, TI can be computationally expensive and consequently large numbers of ligand-protein complexes cannot be evaluated.

### **.4.3 OPPORTUNITIES**

Deconvolution approaches might offer a variety of opportunities since they give valuable information on the binding of an interesting molecule to its putative target(s) to be validated further with chemical genetic approaches.

In this section the key opportunities of each approach, briefly summarized in Table 3, are described and discussed in more depth.

**Genetic approaches:** One of the main advantages of these techniques is that they can be used in conjunction with other methods to ascertain the various interactions between the target

and other proteins or between the target and its ligands, such as bioinformatic tools, artificial intelligence, robotics and HTS. Indeed, for instance, the combined application of these cutting-edge techniques allowed for the discovery of similarities between parasite strains treated with an ODC inhibitor (eflornithine) and those whose ODC was inactivated, proving that ODC would be a useful target for the development of new HAT treatments (215).

Furthermore, reverse genetics, which includes sequencing, enabled the identification of key enzymes and the validation of proteins that are responsible for the development of resistance (such as *TbAAT6*, *TbAT1/P2* adenosine transporter), as well as the identification of crucial biochemical pathways that differ from those of mammalian hosts and essential targets (22) (35, 212).

Moreover, large panels of DNA databases have been developed and are accessible to the public, allowing for the application of bioinformatics tools and novel approaches for the target deconvolution in particular via in silico approaches and the subsequent drug discovery. In respect to drug discovery, genome mining seeks to identify the genes encoding for novel bioactive molecules (211). Thanks to its association with synthetic biology, natural substances can be produced using altered microorganisms (210). As an illustration, a precursor of artemisinin is produced via fermentation and is then semi-synthetically modified to get artemisinin used against malaria (210, 251).

Lastly, these techniques provide an opportunity for conducting QSAR analyses on target isoforms in order to comprehend the interactions between the relevant chemical and the active site (22).

**Affinity chromatography:** This technique is simple, fast, and reliable that opens up a wide range of possibilities since the identified drug-interacting protein can be cloned and overexpressed in bacteria or eukaryotic cells, then purified for future investigations (22).

Additionally, certain parasite organelle structures, such as the kinetoplast (containing circular DNA in trypanosomatids) or the apicoplast (found in the majority of apicomplexan parasites), can be isolated in order to examine them and ascertain their fundamental constituents. For instance, the *T. brucei* Tb927.2.6100 protein has been identified as an essential protein

associated with the kinetoplast DNA or the tryptophanyl-tRNA synthetase (TrpRS<sup>api</sup>) found in the apicoplast of *Plasmodium falciparum* (209, 218).

Furthermore, the combination of affinity chromatography and mass spectrometry enables the identification of particular protein-protein interactions as well as the discovery of target proteins, even those with low abundance (102, 252, 253).

**Generation of resistant parasites:** The ability to identify the mechanism of resistance to direct first-line treatment is one of this method's greatest opportunities: for example, analysis of *T.b. gambiense* from patients relapsing after melarsoprol treatment revealed alterations in the *TbAT1* gene linked to the adenosine transporter 1, resulting in loss of drug uptake by the P2 purine transporter (212). As another example, deletion of the amino acid transporter gene *TbAAT6* in resistant strains led to resistance to eflornithine (35).

**Activity-based protein profiling:** The ABPP, which relies on covalently binding the putative target(s) and immediate target(s) labeling, facilitates the generation of databases that would enable the simultaneous application of phenotypic screening and target identification (direct linkage mediated by the addition of an alkyne or an azide forming the clickable group ) (68, 228, 254) (226). Moreover, the combining ABPP with bioimaging allows a better understanding of pharmaceutical effects and prospective risks (119). Furthermore, the combination of activity-based probes and a covalent inhibitor in the same experiment can be used to examine the function of the identified protein and understand the mechanism of action of the inhibitor (226, 227). For instance, the inhibitor WRR-086 that had previously been chosen and transformed into an ABPP enabled the identification of TgDJ-1, a crucial protein involved in the process of host cell invasion by *Toxoplasma gondii* (225, 226).

Another example where the opportunities of ABPP have been successfully used, is the discovery of dipeptidylpeptidase1 (DPAP1) as promising therapeutic target. DPAP\_1 was found to have a substantial role in malarial infection. This discovery was made using a broad-spectrum cathepsin-C specific probe and thus demonstrated the effectiveness of the ABPP method for target identification and discovery of disease-associated proteins (225, 228).

**Metabolomics:** The combination of mass spectrometry and NMR as complementary technologies allow the reproducible identification, quantification of metabolites, and stable

isotope tracing in vivo (237, 255). Moreover, a challenge, which represents both an exciting opportunity for chemical discovery and an intimidating analytical burden, is that the size of the metabolome is not well defined (238). Despite the inherent drawbacks of this approach, metabolomics is a key tool and will be used more readily in the identification of drug targets for anti-parasite drug discovery (121).

***In silico* approach:** Computational approaches for target prediction can have a large range of applications; they can be used for predicting the undesirable effects of drugs or their mechanism of action (232). It can also help in prioritizing certain drug targets and therefore narrowing down the validation process and accelerating the screening step (233). In order to forecast novel drug-target interactions, machine learning, deep learning, and artificial intelligence use models that learn the characteristics of existing medicines and their targets (233).

Methods for computer-aided drug design have advanced rapidly. Methods such as virtual screening of induced-fit docking (IFD), molecular docking, and MD simulations aided in determining protein function. In contrast to drug affinity-responsive target stability (DARTS), *in silico* approaches provide accurate combined information, making them an effective supplement to DARTS. Because of their virtual state, *in silico* approaches that calculate the binding details of a small molecule ligand may be inaccurate. DARTS, on the other hand, is typically performed in response to physiological states, such as those found in cells or tissues. In this context, DARTS and *in silico* methods are complementary (111), allowing the limitations of each method to be overcome.

#### **.4.4 THREATS**

Deconvolution approaches have many benefits, but there may still be some concerns. These methods rely on the discovery and confirmation of prospective therapeutic targets, which might not include all pertinent proteins for a given disease. Indeed, targets designed for streamlined cell-based assays may not always behave as expected in intricate organ systems and may not be transferable to humans (2). The drug sildenafil serves as an excellent illustration of the great complexity of human organisms. This medicine, which was initially intended to treat cardiovascular disease, has now been approved for use in treating erectile

dysfunction owing to an unanticipated pharmacological impact (2). Furthermore, understanding a target or mode of action might not be adequate to fully describe the situation. Drugs may have multiple targets, and their apparent efficacy is not always the result of the mechanisms they were previously thought to have (2). Selective serotonin reuptake inhibitors (SSRIs), which are used to treat depression, are an outstanding example. Indeed, the effect that was first anticipated was an increase in serotonin levels in synapses, but it was revealed to have a more intricate mode of action (256).

The management and processing of big data produced by the myriad of newly discovered targets and combinatorial chemistry, as well as the requirement for sophisticated AI-based machine learning tools, is a enormous challenge and thus a potential threat to all deconvolution methods if not accurately addressed (257).

Below some threats associated with each approach summarized in Table 3 are discussed:

**Genetic approaches:** Despite its many advantages, this strategy may not be able to anticipate the biotechnological potential of some compounds, such as natural products, and may only be relevant to identify biosynthetic gene clusters (209). Additionally, it might set some boundaries regarding the formulation of chemical structures (209).

**Affinity chromatography:** Determining where to bind the linker to the molecule of interest without impairing activity is a challenge and remains a threat along with its eventual difficulties of production and purification (219). It is essential to produce ligands that are readily attached to solid supports and are similar to the relevant bioactive compounds. The risk is that these ligands do not have the same selectivity or affinity for binding to the desired targets, which would impede the capture of the intended proteins (104). Moreover, employing an affinity column to purify a particular target can be challenging if the sample contains a complex protein mixture. Furthermore, proteins may occasionally become denatured during the purification process, rendering them unusable for later applications (220). Finally, affinity chromatography may produce significant background noise, which can make it challenging to identify and measure the target protein (103).

**Generation of resistant parasites:** A hurdle would be that some strains might require a long time to develop drug resistance.

**Activity-based protein profiling:** It can be challenging to tell the difference between generic labeling (the non-specific marking) and actual drug and target interaction (119). Also, in some cases, an active nucleophile site is necessary to be present in the active site of the enzyme to attach covalently the ABPP to target proteins (but could be overcome by adding a photo-reactive moiety) (226).

**Metabolomics:** The complex nature of biological samples, comprising metabolites from different chemical classes with distinct physicochemical properties (polarity, solubility, volatility), imposes a high demand on the analytical performance, and multiplatform assemblies are usually preferred (136). It is clear that no single profiling experiment can be truly comprehensive for even currently known metabolites (238). In addition, the use of orthogonal techniques is mandatory to provide valuable information for the identification of unknown metabolites (237). Correct treatment of statistical data is important for any metabolomic approach (239).

**In silico approach:** On the one hand, one of the primary concerns of SBDD methodologies is the requirement of high-quality experimental or theoretical structures. The 3D structures of the targets have to be available if one wants to use the molecular docking or MD simulations (233). On the other hand, LBDD methods have also some limitations and drawbacks that need to be considered. One of the main challenges is the requirement of a known ligand or a set of active compounds to derive a reliable model. Additionally, if the binding of a drug to its target is not strong enough, the prediction of the interaction can be useless (233). Several issues are related to reverse docking such as proper construction of target structure databases, correct identification of binding pockets, computational efficiency, and proper normalization of docking scores (230). Furthermore, the binding energy estimation by docking scores is usually of low accuracy (139, 230).

## .5 Conclusion

To summarize, the target(s) deconvolution methods used within the drug discovery process are crucial in advancing the discovery of novel and more potent therapies. Each tool provides distinct skills and insights that can be useful when combined. For example, the use of computational methods to predict ligand activity for newly isolated targets through affinity

chromatography or using activity-based protein profiling with proteomics to screen identified targets, demonstrates the complementary nature of each method and the significance of taking this into account when a new target is to be discovered.

Despite the many benefits of the target(s) deconvolution methods, several limitations, such as difficulty in isolating targets of interest with weak interactions, interference from structurally related proteins, and complex mechanisms of action remain. Thus, it is crucial to continue developing target deconvolution tools in the field that takes into account the *in vivo* situation. With the continuous advancement and widespread use of such tools, the likelihood of effective drug discovery will increase.

In line with this need, the thesis project aim was set to develop a new tool that will operate *in vivo* within the *T. brucei* parasite, and that will enable the precise identification of targets of interest and their interacting partners (see Part 2 below)

## .6 References

1. Swinney DC. Phenotypic vs. target-based drug discovery for first-in-class medicines. *Clinical pharmacology and therapeutics*. 2013;93(4):299-301.
2. Croston GE. The utility of target-based discovery. *Expert Opinion on Drug Discovery*. 2017;12(5):427-9.
3. Sanseau P. Impact of human genome sequencing for in silico target discovery. *Drug discovery today*. 2001;6(6):316-23.
4. Spreafico R, Soriaga LB, Grosse J, Virgin HW, Telenti A. Advances in Genomics for Drug Development. *Genes*. 2020;11(8).
5. Debouck C, Metcalf B. The Impact of Genomics on Drug Discovery. *Annual Review of Pharmacology and Toxicology*. 2000;40(1):193-208.
6. Myler PJ. Searching the Trityp genomes for drug targets. *Advances in experimental medicine and biology*. 2008;625:133-40.
7. Xia X. Bioinformatics and Drug Discovery. *Current topics in medicinal chemistry*. 2017;17(15):1709-26.
8. Paananen J, Fortino V. An omics perspective on drug target discovery platforms. *Briefings in Bioinformatics*. 2019;21(6):1937-53.
9. Elliott M. Zanamivir: From Drug Design to the Clinic. *Philosophical Transactions: Biological Sciences*. 2001;356(1416):1885-93.
10. Gundla R, Kazemi R, Sanam R, Muttineni R, Sarma JARP, Dayam R, et al. Discovery of Novel Small-Molecule Inhibitors of Human Epidermal Growth Factor Receptor-2: Combined Ligand and Target-Based Approach. *Journal of Medicinal Chemistry*. 2008;51(12):3367-77.
11. Mouscadet J-F, Tchertanov L. Raltegravir: molecular basis of its mechanism of action. *Eur J Med Res*. 2009;14 Suppl 3(Suppl 3):5-16.
12. Iqbal N, Iqbal N. Imatinib: a breakthrough of targeted therapy in cancer. *Chemother Res Pract*. 2014;2014:357027-.
13. Atmar RL, Finch N. New Perspectives on Antimicrobial Agents: Molnupiravir and Nirmatrelvir/Ritonavir for Treatment of COVID-19. *Antimicrobial Agents and Chemotherapy*. 2022;66(8):e02404-21.
14. Kabinger F, Stiller C, Schmitzová J, Dienemann C, Kokic G, Hillen HS, et al. Mechanism of molnupiravir-induced SARS-CoV-2 mutagenesis. *Nature Structural & Molecular Biology*. 2021;28(9):740-6.
15. Khan FI, Kang T, Ali H, Lai D. Remdesivir Strongly Binds to RNA-Dependent RNA Polymerase, Membrane Protein, and Main Protease of SARS-CoV-2: Indication From Molecular Modeling and Simulations. *Frontiers in pharmacology*. 2021;12:710778.
16. Eder J, Sedrani R, Wiesmann C. The discovery of first-in-class drugs: origins and evolution. *Nature reviews Drug discovery*. 2014;13(8):577-87.
17. Davis RL. Mechanism of Action and Target Identification: A Matter of Timing in Drug Discovery. *iScience*. 2020;23(9):101487.
18. Terstappen GC, Schlüpen C, Raggiaschi R, Gaviraghi G. Target deconvolution strategies in drug discovery. *Nature Reviews Drug Discovery*. 2007;6(11):891-903.
19. Butcher EC, Berg EL, Kunkel EJ. Systems biology in drug discovery. *Nature Biotechnology*. 2004;22(10):1253-9.
20. Swinney DC. Phenotypic vs. Target-Based Drug Discovery for First-in-Class Medicines. *Clinical Pharmacology & Therapeutics*. 2013;93(4):299-301.
21. Müller J, Hemphill A. Drug target identification in protozoan parasites. *Expert Opin Drug Discov*. 2016;11(8):815-24.
22. Müller J, Hemphill A. New approaches for the identification of drug targets in protozoan parasites. *International review of cell and molecular biology*. 2013;301:359-401.
23. Andrews KT, Fisher G, Skinner-Adams TS. Drug repurposing and human parasitic protozoan diseases. *International Journal for Parasitology: Drugs and Drug Resistance*. 2014;4(2):95-111.
24. Zhang X, Wu F, Yang N, Zhan X, Liao J, Mai S, et al. In silico Methods for Identification of Potential Therapeutic Targets. *Interdisciplinary Sciences: Computational Life Sciences*. 2022;14(2):285-310.
25. Kennedy PGE. Update on human African trypanosomiasis (sleeping sickness). *Journal of neurology*. 2019;266(9):2334-7.
26. Welburn SC, Maudlin I. Priorities for the elimination of sleeping sickness. *Advances in parasitology*. 2012;79:299-337.

27. WHO. Trypanosomiasis, human African (sleeping sickness). 2021.
28. Kennedy PGE, Rodgers J. Clinical and Neuropathogenetic Aspects of Human African Trypanosomiasis. *Frontiers in Immunology*. 2019;10(39).
29. Kennedy PGE. Human African trypanosomiasis of the CNS: current issues and challenges. *J Clin Invest*. 2004;113(4):496-504.
30. Kennedy PG. Clinical features, diagnosis, and treatment of human African trypanosomiasis (sleeping sickness). *The Lancet Neurology*. 2013;12(2):186-94.
31. Fairlamb AH, Horn D. Melarsoprol Resistance in African Trypanosomiasis. *Trends in Parasitology*. 2018;34(6):481-92.
32. Bouteille B, Buguet A. The detection and treatment of human African trypanosomiasis. *Res Rep Trop Med*. 2012;3:35-45.
33. Field MC, Horn D, Fairlamb AH, Ferguson MAJ, Gray DW, Read KD, et al. Anti-trypanosomatid drug discovery: an ongoing challenge and a continuing need. *Nat Rev Microbiol*. 2017;15(4):217-31.
34. Wiedemar N, Graf FE, Zwyrer M, Ndomba E, Kunz Renggli C, Cal M, et al. Beyond immune escape: a variant surface glycoprotein causes suramin resistance in *Trypanosoma brucei*. *Molecular Microbiology*. 2018;107(1):57-67.
35. Vincent IM, Creek D, Watson DG, Kamleh MA, Woods DJ, Wong PE, et al. A molecular mechanism for eflornithine resistance in African trypanosomes. *PLoS pathogens*. 2010;6(11):e1001204.
36. Wilkinson SR, Taylor MC, Horn D, Kelly JM, Cheeseman I. A mechanism for cross-resistance to nifurtimox and benznidazole in trypanosomes. *Proceedings of the National Academy of Sciences of the United States of America*. 2008;105(13):5022-7.
37. Barrett MP, Vincent IM, Burchmore RJ, Kazibwe AJ, Matovu E. Drug resistance in human African trypanosomiasis. *Future Microbiology*. 2011;6(9):1037-47.
38. WHO. Human African trypanosomiasis (sleeping sickness) 2021 [updated 2021. Available from: [https://www.who.int/health-topics/human-african-trypanosomiasis#tab=tab\\_1](https://www.who.int/health-topics/human-african-trypanosomiasis#tab=tab_1).
39. Fairlamb AH. Fexinidazole for the treatment of human African trypanosomiasis. *Drugs of today (Barcelona, Spain : 1998)*. 2019;55(11):705-12.
40. Neau P, Hänel H, Lameyre V, Strub-Wourgaft N, Kuykens L. Innovative Partnerships for the Elimination of Human African Trypanosomiasis and the Development of Fexinidazole. *Tropical medicine and infectious disease*. 2020;5(1):17.
41. Hidalgo J, Ortiz JF, Fabara SP, Eissa-Garcés A, Reddy D, Collins KD, et al. Efficacy and Toxicity of Fexinidazole and Nifurtimox Plus Eflornithine in the Treatment of African Trypanosomiasis: A Systematic Review. *Cureus*. 2021;13(8):e16881-e.
42. WHO. Sustained decline in sleeping sickness cases puts elimination within reach 2020 [updated 23.06.2020. Available from: <https://www.who.int/news/item/23-06-2020-sustained-decline-in-sleeping-sickness-cases-puts-elimination-within-reach>.
43. Grevelink SA, Lerner EA. Leishmaniasis. *J Am Acad Dermatol*. 1996;34(2 Pt 1):257-72.
44. Santana W, de Oliveira SSC, Ramos MH, Santos ALS, Dolabella SS, Souto EB, et al. Exploring Innovative Leishmaniasis Treatment: Drug Targets from Pre-Clinical to Clinical Findings. *Chem Biodivers*. 2021;18(9):e2100336.
45. Aronson NE, Joya CA. Cutaneous Leishmaniasis: Updates in Diagnosis and Management. *Infect Dis Clin North Am*. 2019;33(1):101-17.
46. van Griensven J, Diro E. Visceral leishmaniasis. *Infect Dis Clin North Am*. 2012;26(2):309-22.
47. Pradhan S, Schwartz RA, Patil A, Grabbe S, Goldust M. Treatment options for leishmaniasis. *Clinical and Experimental Dermatology*. 2022;47(3):516-21.
48. Marques SA, Merlotto MR, Ramos PM, Marques MEA. American tegumentary leishmaniasis: severe side effects of pentavalent antimonial in a patient with chronic renal failure. *An Bras Dermatol*. 2019;94(3):355-7.
49. Frézard F, Demicheli C, Ribeiro RR. Pentavalent Antimonials: New Perspectives for Old Drugs. *Molecules*. 2009;14(7):2317-36.
50. Gasser RA, Jr., Magill AJ, Oster CN, Franke ED, Grögl M, Berman JD. Pancreatitis Induced by Pentavalent Antimonial Agents During Treatment of Leishmaniasis. *Clinical Infectious Diseases*. 1994;18(1):83-90.
51. Sundar S, Sinha PR, Agrawal NK, Srivastava R, Rainey PM, Berman JD, et al. A cluster of cases of severe cardiotoxicity among kala-azar patients treated with a high-osmolality lot of sodium antimony gluconate. *The American journal of tropical medicine and hygiene*. 1998;59(1):139-43.

52. Thakur CP, Sinha GP, Pandey AK, Kumar N, Kumar P, Hassan SM, et al. Do the diminishing efficacy and increasing toxicity of sodium stibogluconate in the treatment of visceral leishmaniasis in Bihar, India, justify its continued use as a first-line drug? An observational study of 80 cases. *Annals of Tropical Medicine & Parasitology*. 1998;92(5):561-9.
53. Sundar S, Chakravarty J. An update on pharmacotherapy for leishmaniasis. *Expert Opinion on Pharmacotherapy*. 2015;16(2):237-52.
54. Sundar S, Singh A. Chemotherapeutics of visceral leishmaniasis: present and future developments. *Parasitology*. 2018;145(4):481-9.
55. Purkait B, Kumar A, Nandi N, Sardar Abul H, Das S, Kumar S, et al. Mechanism of Amphotericin B Resistance in Clinical Isolates of *Leishmania donovani*. *Antimicrobial Agents and Chemotherapy*. 2012;56(2):1031-41.
56. Ponte-Sucre A, Gamarro F, Dujardin JC, Barrett MP, López-Vélez R, García-Hernández R, et al. Drug resistance and treatment failure in leishmaniasis: A 21st century challenge. *PLoS Negl Trop Dis*. 2017;11(12):e0006052.
57. Chowdhury IJK, & Haque, M. F. . Visceral Leishmaniasis Treatment and Its Challenges. *South Asian Journal of Parasitology*. 2020;3(4):1-16.
58. Pérez-Molina JA, Molina I. Chagas disease. *Lancet*. 2018;391(10115):82-94.
59. Mills RM. Chagas Disease: Epidemiology and Barriers to Treatment. *The American Journal of Medicine*. 2020;133(11):1262-5.
60. Nguyen T, Waseem M. Chagas Disease. *StatPearls*. Treasure Island (FL): StatPearls Publishing Copyright © 2021, StatPearls Publishing LLC.; 2021.
61. Mazzeti AL, Capelari-Oliveira P, Bahia MT, Mosqueira VCF. Review on Experimental Treatment Strategies Against *Trypanosoma cruzi*. *J Exp Pharmacol*. 2021;13:409-32.
62. Petravicius PO, Costa-Martins AG, Silva MN, Reis-Cunha JL, Bartholomeu DC, Teixeira MMG, et al. Mapping benznidazole resistance in trypanosomatids and exploring evolutionary histories of nitroreductases and ABCG transporter protein sequences. *Acta Tropica*. 2019;200:105161.
63. Campos MC, Leon LL, Taylor MC, Kelly JM. Benznidazole-resistance in *Trypanosoma cruzi*: evidence that distinct mechanisms can act in concert. *Mol Biochem Parasitol*. 2014;193(1):17-9.
64. Pérez-Molina JA, Crespillo-Andújar C, Bosch-Nicolau P, Molina I. Trypanocidal treatment of Chagas disease. *Enfermedades infecciosas y microbiología clínica (English ed)*. 2021;39(9):458-70.
65. De Rycker M, Baragaña B, Duce SL, Gilbert IH. Challenges and recent progress in drug discovery for tropical diseases. *Nature*. 2018;559(7715):498-506.
66. Jung HJ, Kwon HJ. Target deconvolution of bioactive small molecules: The heart of chemical biology and drug discovery. *Archives of Pharmacol Research*. 2015;38(9):1627-41.
67. Wilkinson IVL, Terstappen GC, Russell AJ. Combining experimental strategies for successful target deconvolution. *Drug discovery today*. 2020;25(11):1998-2005
68. Lee J, Bogoy M. Target deconvolution techniques in modern phenotypic profiling. *Current opinion in chemical biology*. 2013;17(1):118-26.
69. Mitchel JT. FDA relations during drug development. *Dialogues Clin Neurosci*. 2000;2(3):213-7.
70. Müller J, Hemphill A. Chapter Seven - New Approaches for the Identification of Drug Targets in Protozoan Parasites. In: Jeon KW, editor. *International review of cell and molecular biology*. 301: Academic Press; 2013;301:359-401.
71. Ferreira LLG, de Moraes J, Andricopulo AD. Approaches to advance drug discovery for neglected tropical diseases. *Drug discovery today*. 2022;27(8):2278-87.
72. Zamore PD. RNA Interference: Big Applause for Silencing in Stockholm. *Cell*. 2006;127(6):1083-6.
73. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*. 1998;391(6669):806-11.
74. Boettcher M, McManus Michael T. Choosing the Right Tool for the Job: RNAi, TALEN, or CRISPR. *Molecular Cell*. 2015;58(4):575-85.
75. Han H. RNA Interference to Knock Down Gene Expression. In: DiStefano JK, editor. *Disease Gene Identification: Methods and Protocols*. New York, NY: Springer New York; 2018;1706:293-302.
76. Mackey ZB, Koupparis K, Nishino M, McKerrow JH. High-throughput analysis of an RNAi library identifies novel kinase targets in *Trypanosoma brucei*. *Chemical biology & drug design*. 2011;78(3):454-63.
77. Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science (New York, NY)*. 2014;346(6213):1258096.
78. Kim DH, Rossi JJ. RNAi mechanisms and applications. *BioTechniques*. 2008;44(5):613-6.

79. Leonard JN, Schaffer DV. Antiviral RNAi therapy: emerging approaches for hitting a moving target. *Gene Therapy*. 2006;13(6):532-40.
80. Scott LJ. Givosiran: First Approval. *Drugs*. 2020;80(3):335-9.
81. Hoy SM. Patisiran: First Global Approval. *Drugs*. 2018;78(15):1625-31.
82. Lamb YN. Inclisiran: First Approval. *Drugs*. 2021;81(3):389-95.
83. Scott LJ, Keam SJ. Lumasiran: First Approval. *Drugs*. 2021;81(2):277-82.
84. Kelleher AD, Cortez-Jugo C, Cavalieri F, Qu Y, Glanville AR, Caruso F, et al. RNAi therapeutics: an antiviral strategy for human infections. *Current Opinion in Pharmacology*. 2020;54:121-9.
85. Setten RL, Rossi JJ, Han S-p. The current state and future directions of RNAi-based therapeutics. *Nature Reviews Drug Discovery*. 2019;18(6):421-46.
86. Al-Attar S, Westra ER, Oost Jvd, Brouns SJJ. Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defense mechanism in prokaryotes. 2011;392(4):277-89.
87. Ma Y, Zhang L, Huang X. Genome modification by CRISPR/Cas9. *The FEBS Journal*. 2014;281(23):5186-93.
88. Asano Y, Yamashita K, Hasegawa A, Ogasawara T, Iriki H, Muramoto T. Knock-in and precise nucleotide substitution using near-PAMless engineered Cas9 variants in *Dictyostelium discoideum*. *Scientific Reports*. 2021;11(1):11163.
89. Yang JH. CRISPR(e)R drug discovery. *Nature Chemical Biology*. 2022;18(5):435-6.
90. Shi J, Wang E, Milazzo JP, Wang Z, Kinney JB, Vakoc CR. Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nature Biotechnology*. 2015;33(6):661-7.
91. Anglada-Giroto M, Handschin G, Ortmayr K, Campos AI, Gillet L, Manfredi P, et al. Author Correction: Combining CRISPRi and metabolomics for functional annotation of compound libraries. *Nat Chem Biol*. 2022;18(5):575.
92. Lee N, Hwang S, Kim J, Cho S, Palsson B, Cho B-K. Mini review: Genome mining approaches for the identification of secondary metabolite biosynthetic gene clusters in *Streptomyces*. *Computational and Structural Biotechnology Journal*. 2020;18:1548-56.
93. Olano C, Méndez C, Salas JA. Strategies for the Design and Discovery of Novel Antibiotics using Genetic Engineering and Genome Mining. In: Villa TG, Veiga-Crespo P, editors. *Antimicrobial Compounds: Current Strategies and New Alternatives*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2014. p. 1-25.
94. Ziemert N, Alanjary M, Weber T. The evolution of genome mining in microbes – a review. *Natural Product Reports*. 2016;33(8):988-1005.
95. Moresco EM, Li X, Beutler B. Going forward with genetics: recent technological advances and forward genetics in mice. *The American journal of pathology*. 2013;182(5):1462-73.
96. Graven P. Genetic and chemical validation of *Trypanosoma brucei* adenosine kinase, the intracellular target of 4-[5-(4-phenoxyphenyl)-2H-pyrazol-3-yl]-morpholine 2013.
97. Kuettel S, Greenwald J, Kostrewa D, Ahmed S, Scapozza L, Perozzo R. Crystal Structures of T. b. rhodesiense Adenosine Kinase Complexed with Inhibitor and Activator: Implications for Catalysis and Hyperactivation. *PLOS Neglected Tropical Diseases*. 2011;5(5):e1164.
98. Kuettel S, Mosimann M, Maser P, Kaiser M, Brun R, Scapozza L, et al. Adenosine Kinase of T. b. Rhodesiense identified as the putative target of 4-[5-(4-phenoxyphenyl)-2H-pyrazol-3-yl]morpholine using chemical proteomics. *PLoS Negl Trop Dis*. 2009;3(8):e506.
99. Kuettel S, Zambon A, Kaiser M, Brun R, Scapozza L, Perozzo R. Synthesis and evaluation of antiparasitic activities of new 4-[5-(4-phenoxyphenyl)-2H-pyrazol-3-yl]morpholine derivatives. *J Med Chem*. 2007;50(23):5833-9.
100. Sasidharan S, Saudagar P. Knockout of Tyrosine Aminotransferase Gene by Homologous Recombination Arrests Growth and Disrupts Redox Homeostasis in *Leishmania* Parasite. *Parasitology research*. 2022;121(11):3229-41.
101. Beilstein S, El Phil R, Sahraoui SS, Scapozza L, Kaiser M, Mäser P. Laboratory Selection of Trypanosomatid Pathogens for Drug Resistance. *Pharmaceuticals (Basel, Switzerland)*. 2022;15(2):135.
102. Schenone M, Dančík V, Wagner BK, Clemons PA. Target identification and mechanism of action in chemical biology and drug discovery. *Nature Chemical Biology*. 2013;9(4):232-40.
103. Saxena C, Higgs RE, Zhen E, Hale JE. Small-molecule affinity chromatography coupled mass spectrometry for drug target deconvolution. *Expert Opinion on Drug Discovery*. 2009;4(7):701-14.
104. Sleno L, Emili A. Proteomic methods for drug target discovery. *Current opinion in chemical biology*. 2008;12(1):46-54.

- 
105. Ito T, Ando H, Suzuki T, Ogura T, Hotta K, Imamura Y, et al. Identification of a primary target of thalidomide teratogenicity. *Science (New York, NY)*. 2010;327(5971):1345-50.
  106. Fields S, Song O-k. A novel genetic system to detect protein-protein interactions. *Nature*. 1989;340(6230):245-6.
  107. Licitra EJ, Liu JO. A three-hybrid system for detecting small ligand-protein receptor interactions. *Proceedings of the National Academy of Sciences of the United States of America*. 1996;93(23):12817-21.
  108. Chidley C, Haruki H, Pedersen MG, Muller E, Johnsson K. A yeast-based screen reveals that sulfasalazine inhibits tetrahydrobiopterin biosynthesis. *Nat Chem Biol*. 2011;7(6):375-83.
  109. Wang P, Klassmüller T, Karg CA, Kretschmer M, Zahler S, Braig S, et al. Using the yeast three-hybrid system for the identification of small molecule-protein interactions with the example of ethinylestradiol. *Biological Chemistry*. 2022;403(4):421-31.
  110. Lomenick B, Hao R, Jonai N, Chin RM, Aghajan M, Warburton S, et al. Target identification using drug affinity responsive target stability (DARTS). *Proceedings of the National Academy of Sciences*. 2009;106(51):21984-9.
  111. Ren Y-S, Li H-L, Piao X-H, Yang Z-Y, Wang S-M, Ge Y-W. Drug affinity responsive target stability (DARTS) accelerated small molecules target discovery: Principles and application. *Biochemical Pharmacology*. 2021;194:114798.
  112. Saxena C. Identification of protein binding partners of small molecules using label-free methods. *Expert Opinion on Drug Discovery*. 2016;11(10):1017-25.
  113. Robinson TJ, Pai M, Liu JC, Vizeacoumar F, Sun T, Egan SE, et al. High-throughput screen identifies disulfiram as a potential therapeutic for triple-negative breast cancer cells: interaction with IQ motif-containing factors. *Cell Cycle*. 2013;12(18):3013-24.
  114. Tanca A, Deligios M, Addis MF, Uzzau S. High throughput genomic and proteomic technologies in the fight against infectious diseases. *Journal of infection in developing countries*. 2013;7(3):182-90.
  115. MacBeath G. Protein microarrays and proteomics. *Nature Genetics*. 2002;32:526-32.
  116. Bonnet J, Garcia C, Leger T, Couquet M-P, Vignoles P, Vatunga G, et al. Proteome characterization in various biological fluids of *Trypanosoma brucei* gambiense-infected subjects. *Journal of Proteomics*. 2019;196:150-61.
  117. Wright MH, Sieber SA. Chemical proteomics approaches for identifying the cellular targets of natural products. *Natural Product Reports*. 2016;33(5):681-708.
  118. Srivastava S. Proteomics-Based Investigations of Neglected and Tropical Diseases. *Proteomics Clinical applications*. 2018;12(4):e1800076.
  119. Xu J, Li X, Ding K, Li Z. Applications of Activity-Based Protein Profiling (ABPP) and Bioimaging in Drug Discovery. *Chemistry – An Asian Journal*. 2020;15(1):34-41.
  120. Murale DP, Hong SC, Haque MM, Lee JS. Photo-affinity labeling (PAL) in chemical proteomics: a handy tool to investigate protein-protein interactions (PPIs). *Proteome science*. 2016;15:14.
  121. Vincent IM, Barrett MP. Metabolomic-based strategies for anti-parasite drug discovery. *J Biomol Screening*. 2015;20(1):44.
  122. Creek DJ, Anderson J, McConville MJ, Barrett MP. Metabolomic analysis of trypanosomatid protozoa. *Mol Biochem Parasitol*. 2012;181(2):73-84.
  123. Vincent IM, Creek DJ, Burgess K, Woods DJ, Burchmore RJS, Barrett MP. Untargeted metabolomics reveals a lack of synergy between nifurtimox and eflornithine against *Trypanosoma brucei*. *PLoS Negl Trop Dis*. 2012;6(5):e1618.
  124. Ali JAM, Creek DJ, Burgess K, Allison HC, Field MC, Maser P, et al. Pyrimidine salvage in *Trypanosoma brucei* bloodstream forms and the trypanocidal action of halogenated pyrimidines. *Mol Pharmacol*. 2013;83(2):439-53.
  125. Creek DJ, Nijagal B, Kim D-H, Rojas F, Matthews KR, Barrett MP. Metabolomics guides rational development of a simplified cell culture medium for drug screening against *Trypanosoma brucei*. *Antimicrob Agents Chemother*. 2013;57(6):2768-79.
  126. Betu Kumeso VK, Kalonji WM, Rembry S, Valverde Mordt O, Ngolo Tete D, Prêtre A, et al. Efficacy and safety of acoziborole in patients with human African trypanosomiasis caused by *Trypanosoma brucei* gambiense: a multicentre, open-label, single-arm, phase 2/3 trial. *The Lancet Infectious Diseases*. 2023;23(4):463-70.
  127. Steketee PC, Vincent IM, Achcar F, Giordani F, Kim D-H, Creek DJ, et al. Benzoxaborole treatment perturbs Sadenosyl-L-methionine metabolism in *Trypanosoma brucei*. *PLoS Neglected Trop Dis*. 2018;12(5):e0006450/1.

- 
128. Trochine A, Creek DJ, Faral-Tello P, Barrett MP, Robello C. Benznidazole biotransformation and multiple targets in *Trypanosoma cruzi* revealed by metabolomics. *PLoS Neglected Trop Dis*. 2014;8(5):e2844/1.
  129. Hennig K, Abi-Ghanem J, Bunescu A, Meniche X, Biliat E, Ouattara AD, et al. Metabolomics, lipidomics and proteomics profiling of myoblasts infected with *Trypanosoma cruzi* after treatment with different drugs against Chagas disease. *Metabolomics*. 2019;15(9):117.
  130. t'Kindt R, Scheltema RA, Jankevics A, Bruncker K, Rijal S, Dujardin J-C, et al. Metabolomics to unveil and understand phenotypic diversity between pathogen populations. *PLoS Neglected Trop Dis*. 2010;4(11):e904.
  131. Rojo D, Canuto GAB, Castilho-Martins EA, Tavares MFM, Barbas C, Lopez-Gonzalez A, et al. A Multiplatform metabolomic approach to the basis of antimonial action and resistance in *Leishmania infantum*. *PLoS One*. 2015;10(7):e0130675/1.
  132. Berg M, Garcia-Hernandez R, Cuypers B, Vanaerschot M, Manzano JI, Poveda JA, et al. Experimental resistance to drug combinations in *Leishmania donovani*: metabolic and phenotypic adaptations. *Antimicrob Agents Chemother*. 2015;59(4):2242-55.
  133. Berg M, Vanaerschot M, Jankevics A, Cuypers B, Maes I, Mukherjee S, et al. Metabolic adaptations of *Leishmania donovani* in relation to differentiation, drug resistance, and drug pressure. *Mol Microbiol*. 2013;90(2):428-42.
  134. Vincent IM, Weidt S, Rivas L, Burgess K, Smith TK, Ouellette M. Untargeted metabolomic analysis of miltefosine action in *Leishmania infantum* reveals changes to the internal lipid metabolism. *Int J Parasitol Drugs Drug Resist*. 2014;4(1):20-7.
  135. Canuto GAB, Castilho-Martins EA, Tavares M, Lopez-Gonzalez A, Rivas L, Barbas C. CE-ESI-MS metabolic fingerprinting of *Leishmania* resistance to antimony treatment. *Electrophoresis*. 2012;33(12):1901-10.
  136. Canuto GAB, Dorr F, Lago JHG, Tempone AG, Pinto E, Pimenta DC, et al. New insights into the mechanistic action of methyldehydrodieugenol B towards *Leishmania* (L.) *infantum* via a multiplatform based untargeted metabolomics approach. *Metabolomics*. 2017;13(5):1-14.
  137. Njogu PM, Guantai EM, Pavada E, Chibale K. Computer-Aided Drug Discovery Approaches against the Tropical Infectious Diseases Malaria, Tuberculosis, Trypanosomiasis, and Leishmaniasis. *ACS Infectious Diseases: American Chemical Society*. 2016;2(1):8-31.
  138. Yu W, MacKerell AD. Computer-Aided Drug Design Methods. *Methods in molecular biology* (Clifton, NJ). 2017;1520(1):85-106.
  139. Sydow D, Burggraaff L, Szengel A, Van Vlijmen HWT, Ijzerman AP, Van Westen GJP, et al. Advances and Challenges in Computational Target Prediction. *Journal of Chemical Information and Modeling*. 2019;59(5):1728-42.
  140. Byrne R, Schneider G. In Silico Target Prediction for Small Molecules. 2019;1888:273-309.
  141. Danishuddin, Khan AU. Descriptors and their selection methods in QSAR analysis: paradigm for drug design. *Drug discovery today*. 2016;21(8):1291-302.
  142. Hert Jrm, Willett P, Wilton DJ, Acklin P, Azzaoui K, Jacoby E, et al. Comparison of topological descriptors for similarity-based virtual screening using multiple bioactive reference structures. *Organic & Biomolecular Chemistry*. 2004;2(22):3256-.
  143. Nettles JH, Jenkins JL, Bender A, Deng Z, Davies JW, Glick M. Bridging Chemical and Biological Space: "Target Fishing" Using 2D and 3D Molecular Descriptors. *Journal of Medicinal Chemistry*. 2006;49(23):6802-10.
  144. Raymond JW, Willett P. Effectiveness of graph-based and fingerprint-based similarity measures for virtual screening of 2D chemical structure databases. *Journal of computer-aided molecular design*. 2002;16(1):59-71.
  145. Pereira CA, Sayé M, Reigada C, Silber AM, Labadie GR, Miranda MR, et al. Computational approaches for drug discovery against trypanosomatid-caused diseases. *Parasitology*. 2020;147(6):611-33.
  146. Bajusz D, Rácz A, Héberger K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *Journal of Cheminformatics*. 2015;7(1):1-13.
  147. Awale M, Reymond J-L. The polypharmacology browser: a web-based multi-fingerprint target prediction tool using ChEMBL bioactivity data. *Journal of Cheminformatics*. 2017;9(1):11.
  148. Reker D, Rodrigues T, Schneider P, Schneider G. Identifying the macromolecular targets of de novo-designed chemical entities through self-organizing map consensus. *Proceedings of the National Academy of Sciences*. 2014;111(11):4067-72.
  149. Nickel J, Gohlke B-O, Erehman J, Banerjee P, Rong WW, Goede A, et al. SuperPred: update on drug classification and target prediction. *Nucleic Acids Research*. 2014;42(W1):W26-W31.

- 
150. Liu X, Vogt I, Haque T, Campillos M. HitPick: a web server for hit identification and target prediction of chemical screenings. *Bioinformatics*. 2013;29(15):1910-2.
  151. Gfeller D, Grosdidier A, Wirth M, Daina A, Michielin O, Zoete V. SwissTargetPrediction: a web server for target prediction of bioactive small molecules. *Nucleic Acids Research*. 2014;42(W1):W32-W8.
  152. Lagunin A, Stepanchikova A, Filimonov D, Poroikov V. PASS: prediction of activity spectra for biologically active substances. *Bioinformatics*. 2000;16(8):747-8.
  153. Wang Z, Liang L, Yin Z, Lin J. Improving chemical similarity ensemble approach in target prediction. *Journal of Cheminformatics*. 2016;8(1):20.
  154. Huang T, Mi H, Lin Cy, Zhao L, Zhong LLD, Liu Fb, et al. MOST: Most-similar ligand based approach to target prediction. *BMC Bioinformatics*. 2017;18(1):1-11.
  155. Rhodes N, Willett P, Calvet A, Dunbar JB, Humblet C. CLIP: Similarity Searching of 3D Databases Using Clique Detection. *Journal of Chemical Information and Computer Sciences*. 2003;43(2):443-8.
  156. Agamah FE, Mazandu GK, Hassan R, Bope CD, Thomford NE, Ghansah A, et al. Computational/in silico methods in drug target and lead prediction. *Briefings in Bioinformatics*. 2020;21(5):1663-75.
  157. Burbidge R, Trotter M, Buxton B, Holden S. Drug design by machine learning: support vector machines for pharmaceutical data analysis. *Computers & Chemistry*. 2001;26(1):5-14.
  158. Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP, Feuston BP. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. *Journal of Chemical Information and Computer Sciences*. 2003;43(6):1947-58.
  159. Xia X, Maliski EG, Gallant P, Rogers D. Classification of Kinase Inhibitors Using a Bayesian Model. *Journal of Medicinal Chemistry*. 2004;47(18):4463-70.
  160. Ballester PJ. Machine Learning for Molecular Modelling in Drug Design. *Biomolecules*. 2019;9(6):216.
  161. Winkler DA. Use of Artificial Intelligence and Machine Learning for Discovery of Drugs for Neglected Tropical Diseases. *Frontiers in Chemistry*. 2021;9:1-15.
  162. Berman HM. The Protein Data Bank. *Nucleic Acids Research*. 2000;28(1):235-42.
  163. Pieper U, Webb BM, Dong GQ, Schneidman-Duhovny D, Fan H, Kim SJ, et al. ModBase, a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Research*. 2014;42(D1):D336-D46.
  164. Xiang Z. Advances in Homology Protein Structure Modeling. *Current Protein & Peptide Science*. 2006;7(3):217-27.
  165. Webb B, Sali A. Comparative protein structure modeling using MODELLER. *Current Protocols in Bioinformatics*. 2016;54:5.6.1–5.6.37.
  166. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*. 2018;46(W1):W296-W303.
  167. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*. 2010;5(4):725-38.
  168. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*. 1993;26(2):283-91.
  169. Bhattacharya A, Wunderlich Z, Monleon D, Tejero R, Montelione GT. Assessing model accuracy using the homology modeling automatically software. *Proteins: Structure, Function, and Bioinformatics*. 2007;70(1):105-18.
  170. DasGupta D, Kaushik R, Jayaram B. From Ramachandran Maps to Tertiary Structures of Proteins. *The Journal of Physical Chemistry B*. 2015;119(34):11136-45.
  171. di Luccio E, Koehl P. The H-factor as a novel quality metric for homology modeling. *Journal of Clinical Bioinformatics*. 2012;2(1):18.
  172. Eramian D, Shen M-y, Devos D, Melo F, Sali A, Marti-Renom MA. A composite score for predicting errors in protein structure models. *Protein Science*. 2006;15(7):1653-66.
  173. Shen M-y, Sali A. Statistical potential for assessment and prediction of protein structures. *Protein Science*. 2006;15(11):2507-24.
  174. Koutsoukas A, Simms B, Kirchmair J, Bond PJ, Whitmore AV, Zimmer S, et al. From in silico target prediction to multi-target drug design: Current databases, methods and applications. *Journal of Proteomics*. 2011;74(12):2554-74.
  175. Huang H, Zhang G, Zhou Y, Lin C, Chen S, Lin Y, et al. Reverse screening methods to search for the protein targets of chemopreventive compounds. *Frontiers in Chemistry: Frontiers Media S. A*. 2018;6:138.

- 
176. Rognan D. Structure-Based Approaches to Target Fishing and Ligand Profiling. *Molecular Informatics*. 2010;29(3):176-87.
  177. Chen Z, Li H-l, Zhang Q-j, Bao X-g, Yu K-q, Luo X-m, et al. Pharmacophore-based virtual screening versus docking-based virtual screening: a benchmark comparison against eight targets. *Acta Pharmacologica Sinica*. 2009;30(12):1694-708.
  178. Hawkins PCD, Skillman AG, Nicholls A. Comparison of Shape-Matching and Docking as Virtual Screening Tools. *Journal of Medicinal Chemistry*. 2007;50(1):74-82.
  179. Schuffenhauer A, Floersheim P, Acklin P, Jacoby E. Similarity Metrics for Ligands Reflecting the Similarity of the Target Proteins. *Journal of Chemical Information and Computer Sciences*. 2003;43(2):391-405.
  180. Kringelum J, Kjaerulff SK, Brunak S, Lund O, Oprea TI, Taboureau O. ChemProt-3.0: a global chemical biology diseases mapping. *Database*. 2016;2016:bav123.
  181. Rush TS, Grant JA, Mosyak L, Nicholls A. A Shape-Based 3-D Scaffold Hopping Method and Its Application to a Bacterial Protein-Protein Interaction. *Journal of Medicinal Chemistry*. 2005;48(5):1489-95.
  182. Gong J, Cai C, Liu X, Ku X, Jiang H, Gao D, et al. ChemMapper: a versatile web server for exploring pharmacology and chemical structure association based on molecular 3D similarity method. *Bioinformatics*. 2013;29(14):1827-9.
  183. Keiser MJ, Roth BL, Armbruster BN, Ernsberger P, Irwin JJ, Shoichet BK. Relating protein pharmacology by ligand chemistry. *Nature Biotechnology*. 2007;25(2):197-206.
  184. Liu X, Ouyang S, Yu B, Liu Y, Huang K, Gong J, et al. PharmMapper server: a web server for potential drug target identification using pharmacophore mapping approach. *Nucleic Acids Research*. 2010;38(suppl\_2):W609-W14.
  185. Wang X, Pan C, Gong J, Liu X, Li H. Enhancing the Enrichment of Pharmacophore-Based Target Prediction for the Polypharmacological Profiles of Drugs. *Journal of Chemical Information and Modeling*. 2016;56(6):1175-83.
  186. Wang X, Shen Y, Wang S, Li S, Zhang W, Liu X, et al. PharmMapper 2017 update: a web server for potential drug target identification with a comprehensive target pharmacophore database. *Nucleic Acids Research*. 2017;45(W1):W356-W60.
  187. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research*. 2018;46(D1):D1074-D82.
  188. Gilson MK, Liu T, Baitaluk M, Nicola G, Hwang L, Chong J. BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Research*. 2016;44(D1):D1045-D53.
  189. Gao Z, Li H, Zhang H, Liu X, Kang L, Luo X, et al. PDTD: a web-accessible protein database for drug target identification. *BMC Bioinformatics*. 2008;9(1):104.
  190. Li H, Gao Z, Kang L, Zhang H, Yang K, Yu K, et al. TarFisDock: a web server for identifying drug targets with docking approach. *Nucleic Acids Research*. 2006;34(Web Server):W219-W24.
  191. Wang J-C, Chu P-Y, Chen C-M, Lin J-H. idTarget: a web server for identifying protein targets of small chemical molecules with robust scoring functions and a divide-and-conquer docking approach. *Nucleic Acids Research*. 2012;40(W1):W393-W9.
  192. Santiago DN, Pevzner Y, Durand AA, Tran M, Scheerer RR, Daniel K, et al. Virtual Target Screening: Validation Using Kinase Inhibitors. *Journal of Chemical Information and Modeling*. 2012;52(8):2192-203.
  193. Ellingson SR, Smith JC, Baudry J. VinaMPI: Facilitating multiple receptor high-throughput virtual docking on high-performance computers. *Journal of Computational Chemistry*. 2013;34(25):2212-21.
  194. Li G-B, Yu Z-J, Liu S, Huang L-Y, Yang L-L, Lohans CT, et al. IFPTarget: A Customized Virtual Target Identification Method Based on Protein-Ligand Interaction Fingerprinting Analyses. *Journal of Chemical Information and Modeling*. 2017;57(7):1640-51.
  195. Chen YZ, Zhi DG. Ligand-protein inverse docking and its potential use in the computer search of protein targets of a small molecule. *Proteins: Structure, Function, and Genetics*. 2001;43(2):217-26.
  196. Muscat S, Grasso G, Scapozza L, Danani A. In silico investigation of cytochrome bc1 molecular inhibition mechanism against *Trypanosoma cruzi*. *PLoS Negl Trop Dis*. 2023;17(1):e0010545.
  197. Christ CD, Mark AE, van Gunsteren WF. Basic ingredients of free energy calculations: A review. *Journal of Computational Chemistry*. 2010;31(8):1569-82.
  198. Wang L, Wu Y, Deng Y, Kim B, Pierce L, Krilov G, et al. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *Journal of the American Chemical Society*. 2015;137(7):2695-703.

- 
199. Torrie GM, Valleau JP. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*. 1977;23(2):187-99.
  200. Limongelli V, Bonomi M, Parrinello M. Funnel metadynamics as accurate binding free-energy method. *Proceedings of the National Academy of Sciences*. 2013;110(16):6358-63.
  201. Straatsma TP, Berendsen HJC. Free energy of ionic hydration: Analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations. *The Journal of Chemical Physics*. 1988;89(9):5876-86.
  202. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, et al. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Accounts of Chemical Research*. 2000;33(12):889-97.
  203. Wang E, Sun H, Wang J, Wang Z, Liu H, Zhang JZH, et al. End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chemical Reviews*. 2019;119(16):9478-508.
  204. Åqvist J, Medina C, Samuelsson J-E. A new method for predicting binding affinity in computer-aided drug design. "Protein Engineering, Design and Selection". 1994;7(3):385-91.
  205. Terayama K, Iwata H, Araki M, Okuno Y, Tsuda K. Machine learning accelerates MD-based binding pose prediction between ligands and proteins. *Bioinformatics*. 2018;34(5):770-8.
  206. Urán Landaburu L, Berenstein AJ, Videla S, Maru P, Shanmugam D, Chernomoretz A, et al. TDR Targets 6: driving drug discovery for human pathogens through intensive chemogenomic data integration. *Nucleic Acids Research*. 2020;48(D1):D992-D1005.
  207. Sosa EJ, Burguener G, Lanzarotti E, Defelipe L, Radusky L, Pardo AM, et al. Target-Pathogen: a structural bioinformatic approach to prioritize drug targets in pathogens. *Nucleic Acids Research*. 2018;46(D1):D413-D8.
  208. Amos B, Aurrecochea C, Barba M, Barreto A, Basenko Evelina Y, Bažant W, et al. VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center. *Nucleic Acids Research*. 2022;50(D1):D898-D911.
  209. Pasaje CFA, Cheung V, Kennedy K, Lim EE, Baell JB, Griffin MDW, et al. Selective inhibition of apicoplast tryptophanyl-tRNA synthetase causes delayed death in *Plasmodium falciparum*. *Scientific reports*. 2016;6:27531.
  210. Albarano L, Esposito R, Ruocco N, Costantini M. Genome Mining as New Challenge in Natural Products Discovery. *Marine Drugs*. 2020;18(4):199.
  211. Corre C, Challis GL. Heavy Tools for Genome Mining. *Chemistry & Biology*. 2007;14(1):7-9.
  212. Kazibwe AJN, Nerima B, de Koning HP, Mäser P, Barrett MP, Matovu E. Genotypic Status of the TbAT1/P2 Adenosine Transporter of *Trypanosoma brucei* gambiense Isolates from Northwestern Uganda following Melarsoprol Withdrawal. *PLOS Neglected Tropical Diseases*. 2009;3(9):e523.
  213. Pratt C, Nguyen S, Phillips MA. Genetic Validation of *Trypanosoma brucei* Glutathione Synthetase as an Essential Enzyme. *Eukaryotic Cell*. 2014;13(5):614-24.
  214. Ghoda L, Phillips MA, Bass KE, Wang CC, Coffino P. Trypanosome ornithine decarboxylase is stable because it lacks sequences found in the carboxyl terminus of the mouse enzyme which target the latter for intracellular degradation. *Journal of Biological Chemistry*. 1990;265(20):11823-6.
  215. Xiao Y, McCloskey DE, Phillips MA. RNA interference-mediated silencing of ornithine decarboxylase and spermidine synthase genes in *Trypanosoma brucei* provides insight into regulation of polyamine biosynthesis. *Eukaryot Cell*. 2009;8(5):747-55.
  216. Morriswood B, Havlicek K, Demmel L, Yavuz S, Sealey-Cardona M, Vidilaseris K, et al. Novel bilobe components in *Trypanosoma brucei* identified using proximity-dependent biotinylation. *Eukaryot Cell*. 2013;12(2):356-67.
  217. Müller J, Hemphill A. Drug target identification in intracellular and extracellular protozoan parasites. *Current topics in medicinal chemistry*. 2011;11(16):2029-38.
  218. Beck K, Acestor N, Schulfer A, Anupama A, Carnes J, Panigrahi Aswini K, et al. *Trypanosoma brucei* Tb927.2.6100 Is an Essential Protein Associated with Kinetoplast DNA. *Eukaryotic Cell*. 2013;12(7):970-8.
  219. Fassina G, Ruvo M, Palombo G, Verdoliva A, Marino M. Novel ligands for the affinity-chromatographic purification of antibodies. *Journal of Biochemical and Biophysical Methods*. 2001;49(1):481-90.
  220. Urh M, Simpson D, Zhao K. Chapter 26 Affinity Chromatography: General Methods. In: Burgess RR, Deutscher MP, editors. *Methods in Enzymology*. 463: Academic Press; 2009;463:417-38.

- 
221. Denise H, Giroud C, Barrett MP, Baltz T. Affinity chromatography using trypanocidal arsenical drugs identifies a specific interaction between glycerol-3-phosphate dehydrogenase from *Trypanosoma brucei* and Cymelarsan. *European Journal of Biochemistry*. 1999;259(1-2):339-46.
  222. Katiyar S, Kufareva I, Behera R, Thomas SM, Ogata Y, Pollastri M, et al. Lapatinib-binding protein kinases in the African trypanosome: identification of cellular targets for kinase-directed chemical scaffolds. *PLoS One*. 2013;8(2):e56150.
  223. Mani J, Güttinger A, Schimanski B, Heller M, Acosta-Serrano A, Pescher P, et al. Alba-domain proteins of *Trypanosoma brucei* are cytoplasmic RNA-binding proteins that interact with the translation machinery. *PLoS One*. 2011;6(7):e22463.
  224. Moellering Raymond E, Cravatt Benjamin F. How Chemoproteomics Can Enable Drug Discovery and Development. *Chemistry & Biology*. 2012;19(1):11-22.
  225. Lee J, Bogyo M. Target deconvolution techniques in modern phenotypic profiling. *Curr Opin Chem Biol*. 2013;17(1):118-26.
  226. Hall CI, Reese ML, Weerapana E, Child MA, Bowyer PW, Albrow VE, et al. Chemical genetic screen identifies *Toxoplasma* DJ-1 as a regulator of parasite secretion, attachment, and invasion. *Proceedings of the National Academy of Sciences of the United States of America*. 2011;108(26):10568-73.
  227. Arastu-Kapur S, Ponder EL, Fonović UP, Yeoh S, Yuan F, Fonović M, et al. Identification of proteases that regulate erythrocyte rupture by the malaria parasite *Plasmodium falciparum*. *Nat Chem Biol*. 2008;4(3):203-13.
  228. Deu E, Leyva MJ, Albrow VE, Rice MJ, Ellman JA, Bogyo M. Functional studies of *Plasmodium falciparum* dipeptidyl aminopeptidase I using small molecule inhibitors and active site probes. *Chem Biol*. 2010;17(8):808-19.
  229. Mayr A, Klambauer G, Unterthiner T, Steijaert M, Wegner JK, Ceulemans H, et al. Large-scale comparison of machine learning methods for drug target prediction on ChEMBL. *Chemical Science*. 2018;9(24):5441-51.
  230. Lee A, Lee K, Kim D. Using reverse docking for target identification and its applications for drug discovery. *Expert Opin Drug Discov*. 2016;11(7):707-15.
  231. Cruz-Monteagudo M, Medina-Franco JL, Pérez-Castillo Y, Nicolotti O, Cordeiro MNDS, Borges F. Activity cliffs in drug discovery: Dr Jekyll or Mr Hyde? *Drug discovery today*. 2014;19(8):1069-80.
  232. Sydow D, Burggraaff L, Szengel A, van Vlijmen HWT, AP IJ, van Westen GJP, et al. Advances and Challenges in Computational Target Prediction. *J Chem Inf Model*. 2019;59(5):1728-42.
  233. Thafar M, Raies AB, Albaradei S, Essack M, Bajic VB. Comparison Study of Computational Prediction Tools for Drug-Target Binding Affinities. *Front Chem*. 2019;7:782.
  234. Hirani D, Alvira CM, Danopoulos S, Milla C, Donato M, Tian L, et al. Macrophage-derived IL-6 trans-signalling as a novel target in the pathogenesis of bronchopulmonary dysplasia. *Eur Respir J*. 2022;59(2):2002248.
  235. Naz S, Vallejo M, Garcia A, Barbas C. Method validation strategies involved in non-targeted metabolomics. *J Chromatogr A*. 2014;1353:99-105.
  236. Pezzatti J, Boccard J, Codesido S, Gagnebin Y, Joshi A, Picard D, et al. Implementation of liquid chromatography-high resolution mass spectrometry methods for untargeted metabolomic analyses of biological samples: A tutorial. *Anal Chim Acta*. 2020;1105:28-44.
  237. Fillet M, Frédérick M. The emergence of metabolomics as a key discipline in the drug discovery process. *Drug Discov Today Technol*. 2015;13:19-24.
  238. Sindelar M, Patti GJ. Chemical Discovery in the Era of Metabolomics. *J Am Chem Soc*. 2020;142(20):9097-105.
  239. Barnes S. Overview of Experimental Methods and Study Design in Metabolomics, and Statistical and Pathway Considerations. *Methods Mol Biol*. 2020;2104:1-10.
  240. Vincent IM, Creek DJ, Burgess K, Woods DJ, Burchmore RJ, Barrett MP. Untargeted metabolomics reveals a lack of synergy between nifurtimox and eflornithine against *Trypanosoma brucei*. *PLoS Negl Trop Dis*. 2012;6(5):e1618.
  241. Zerikly M, Challis GL. Strategies for the Discovery of New Natural Products by Genome Mining. *ChemBioChem*. 2009;10(4):625-33.
  242. Scheffler RJ, Colmer S, Tynan H, Demain AL, Gullo VP. Antimicrobials, drug discovery, and genome mining. *Applied Microbiology and Biotechnology*. 2013;97(3):969-78.
  243. Poltorak A, He X, Smirnova I, Liu MY, Van Huffel C, Du X, et al. Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science (New York, NY)*. 1998;282(5396):2085-8.

244. LePage DF, Conlon RA. Animal Models for Disease. In: Wang QK, editor. Cardiovascular Disease: Methods and Protocols Volume 2: Molecular Medicine. Totowa, NJ: Humana Press; 2007. p. 41-67.
245. Kuettel S, Mosimann M, Mäser P, Kaiser M, Brun R, Scapozza L, et al. Adenosine Kinase of T. b. Rhodesiense identified as the putative target of 4-[5-(4-phenoxyphenyl)-2H-pyrazol-3-yl]morpholine using chemical proteomics. PLoS neglected tropical diseases. 2009;3(8):e506.
246. Kondratyev NV, Alfimova MV, Golov AK, Golimbet VE. Bench Research Informed by GWAS Results. Cells. 2021;10(11):3184.
247. Chen H, Liu H, Peng X. Reverse genetics in virology: A double edged sword. Biosafety and Health. 2022;4(5):303-13.
248. Hughes KA, Leips J. Pleiotropy, constraint, and modularity in the evolution of life histories: insights from genomic analyses. Annals of the New York Academy of Sciences. 2017;1389(1):76-91.
249. Hardy S, Legagneux V, Audic Y, Paillard L. Reverse genetics in eukaryotes. Biology of the cell. 2010;102(10):561-80.
250. Wohlleben W, Mast Y, Stegmann E, Ziemert N. Antibiotic drug discovery. Microbial Biotechnology. 2016;9(5):541-8.
251. Paddon CJ, Keasling JD. Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. Nature Reviews Microbiology. 2014;12(5):355-67.
252. Ranish JA, Yi EC, Leslie DM, Purvine SO, Goodlett DR, Eng J, et al. The study of macromolecular complexes by quantitative proteomics. Nat Genet. 2003;33(3):349-55.
253. Blagoev B, Kratchmarova I, Ong SE, Nielsen M, Foster LJ, Mann M. A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. Nat Biotechnol. 2003;21(3):315-8.
254. Puri AW, Lupardus PJ, Deu E, Albrow VE, Garcia KC, Bogoy M, et al. Rational design of inhibitors and activity-based probes targeting Clostridium difficile virulence factor TcdB. Chem Biol. 2010;17(11):1201-11.
255. Alarcon-Barrera JC, Kostidis S, Ondo-Mendez A, Giera M. Recent advances in metabolomics analysis for early drug development. Drug discovery today. 2022;27(6):1763-73.
256. Walker FR. A critical review of the mechanism of action for the selective serotonin reuptake inhibitors: do these drugs possess anti-inflammatory properties and how relevant is this in the treatment of depression? Neuropharmacology. 2013;67:304-17.
257. Batool M, Ahmad B, Choi S. A Structure-Based Drug Discovery Paradigm. International journal of molecular sciences. 2019;20(11):2783.

## **Part 2: Aim of the project**

## **.1 Aim of the project**

As described in the review of Part 1 above, Human African trypanosomiasis (HAT), also known as sleeping sickness, belongs to the class of neglected tropical diseases together with leishmaniasis and Chagas disease. HAT is treated with drugs that are unfortunately often associated with numerous and sometimes very serious side effects such as encephalopathy syndrome, which is a rare but serious side effect of melarsoprol, or strong allergic reactions and nephrotoxicity (1, 2). Moreover, resistance to these treatments seems to be emerging, which poses a real problem for the effective treatment of the populations concerned (2).

These serious adverse effects are typically attributed to an incomplete understanding of the molecule's mode of action, its selectivity, or also unidentified proteins that the medications might target (3-6).

The discovery of many molecules to treat sleeping sickness has been made possible by the use of the phenotypic screening approach, which consists of testing libraries of new compounds on cell cultures to determine which molecule will lead to parasite death. However, neither the target proteins nor the mode of action is clarified by this strategy (4, 6-11).

To answer this unresolved query, scientists have begun to focus on identifying targets in cells (or on membranes) and concentrate on elucidating the mechanism of action of a compound of interest (4, 6, 8, 10, 11).

In the process of developing new drugs, it is crucial to comprehend a molecule's mode of action and potential therapeutic targets, as putative adverse effects could eventually be predicted and a better understanding of the function of the targeted protein and its pathways could be envisaged (4, 12-14).

Since then, many target identification techniques and approaches have been developed and successfully applied such as chemical proteomics techniques, including affinity purification using compound-immobilized beads, photoaffinity labeling, and also genomic metabolomics and in silico approaches (detailed in Part 1) (5, 8, 11, 14-23).

Affinity-based methods are widely used for target isolation and identification (15, 24). These techniques have been applied in our lab in the context of research on neglected tropical diseases, and they have made it possible to identify and validate the *T. brucei* adenosine kinase (*TbAK*) as a target of the compound CD12001 (4-[5-(4-phenoxyphenyl)-2H-pyrazol-3-yl]-morpholine)(25-28).

However, affinity-based techniques come with significant limitations. For instance, the inability to probe particular portions of the proteome because of its loss brought on by the intense lysis conditions employed to solubilize proteins. Additionally, in order to make proteins more soluble, it frequently takes the application of potent detergents, which makes it difficult to identify fleeting or weak protein-protein interactions (29). Another drawback to this approach is the strong background signal caused by non-specific interactions (24, 30).

Several new approaches have been developed to overcome the difficulties encountered with the affinity-based method, including activity-based protein profiling, a chemical proteomics approach in which small molecules with extremely reactive functionality bind covalently to the catalytic site of their targets (31). Unfortunately, because of its restriction to specific chemicals and enzymes, this approach has some flaws that limit its usefulness (31).

The approach known as yeast-two hybrid (Y2H), which has been established for the detection of protein-protein interactions in a living cell, is just one of the many techniques that have been further developed to examine the molecular targets and the many interactions (32, 33). Nonetheless, this approach has certain limitations, such as a high apparent incidence of false positives and false negatives (34, 35).

Another method that builds on the Y2H screen is the yeast-three hybrid (Y3H), which has been defined as beneficial for finding unusual proteins in cells (36-40). Despite the power of this system, one of its drawbacks would be directly related to its readout based on the activation of the reporter gene (e.g., LEU2, LacZ) in yeast cells. In this context, only soluble target proteins translocating to the yeast nucleus can be identified, while proteins located in organelles or membranes may not be detected (39).

Two ground-breaking strategies have recently been discovered to get beyond many of the drawbacks of the various methods discussed above. These two technologies, known as BioID and SNAP-tag, are explained in depth in Part 3 below.

Although both methods present numerous advantages, some limitations of each technique taken separately have been pointed out and outlined in Part 3 below.

The project's concept is based on working *in vivo* and the working hypothesis, that combining two orthogonal methods, BioID and SNAP-tag, could lead to overcoming some of the limitations of the single approaches. The choice of working *in vivo* with *Trypanosoma brucei brucei* (**Tbb**) was made based on its favourable characteristics, namely, its lack of human pathogenicity, its rapid growth, and its ease of manipulation. Indeed, **Tbb** subspecies mainly infect animals such as cattle but not humans owing to the existence in human serum of trypanosome lytic factors (TLF1 and TLF2) that induce lysis of this parasite subtype (41, 42). While the idea of fusing two proteins seems straight forward how to link them is challenging as the proper functioning of a fusion protein depends on the length and type of amino acids constituting the linker connecting the distinct domains of the fusion protein. Thus, the choice of an appropriate linker between the BirA\* and SNAP-tag domains needs to be explored during this study. Since linkers have been shown to establish communication between the various domains composing fusion proteins, they are now understood to conduct far more essential functions than just serving as connectors between the different components (43). Additionally, they preserve the biological activities of the different modules and maintain their proper interaction (44).

These working hypotheses led to the definition of the project overall aim that is to develop a new *in vivo* target deconvolution system issued by the combination of BioID and SNAP-tag technologies and able to subsequently identify small molecule target proteins not only in the cytosol but also in the organellar sub-proteomes of *T. brucei* hence transcending the limitations of the two procedures used individually and breaking the constraints of conventional target identification approaches.

Therefore, the deliverable of this project will be a sensitive target(s) identification tool appropriate for small molecules that interact with their target proteins with moderate to high

affinity to be used in the future for deconvoluting targets of lead compounds discovered through phenotypic screening in several projects of the group as well as those carried out in collaboration with DNDi.

## .2 References

1. Seixas J, Atouguia J, Josenando T, Vatunga G, Bilenge CMM, Lutumba P, et al. Clinical Study on the Melarsoprol-Related Encephalopathic Syndrome: Risk Factors and HLA Association. *Tropical medicine and infectious disease*. 2020;5(1):5.
2. Babokhov P, Sanyaolu AO, Oyibo WA, Fagbenro-Beyioku AF, Iriemenam NC. A current analysis of chemotherapy strategies for the treatment of human African trypanosomiasis. *Pathog Glob Health*. 2013;107(5):242-52.
3. Chatelain E, Ioset J-R. Phenotypic screening approaches for Chagas disease drug discovery. *Expert Opinion on Drug Discovery*. 2018;13(2):141-53.
4. Gilbert IH. Drug discovery for neglected diseases: molecular target-based and phenotypic approaches. *Journal of medicinal chemistry*. 2013;56(20):7719-26.
5. Swinney DC. Phenotypic vs. target-based drug discovery for first-in-class medicines. *Clinical pharmacology and therapeutics*. 2013;93(4):299-301.
6. Eder J, Sedrani R, Wiesmann C. The discovery of first-in-class drugs: origins and evolution. *Nature reviews Drug discovery*. 2014;13(8):577-87.
7. Aulner N, Danckaert A, Ihm J, Shum D, Shorte SL. Next-Generation Phenotypic Screening in Early Drug Discovery for Infectious Diseases. *Trends in Parasitology*. 2019;35(7):559-70.
8. Shabajee P, Gaudeau A, Legros C, Dorval T, Stéphan JP. [From high content screening to target deconvolution: New insights for phenotypic approaches]. *Medecine sciences : M/S*. 2021;37(3):249-57.
9. Moffat JG, Vincent F, Lee JA, Eder J, Prunotto M. Opportunities and challenges in phenotypic drug discovery: an industry perspective. *Nature Reviews Drug Discovery*. 2017;16(8):531-43.
10. Swinney DC. Phenotypic vs. Target-Based Drug Discovery for First-in-Class Medicines. *Clinical Pharmacology & Therapeutics*. 2013;93(4):299-301.
11. Zheng W, Thorne N, McKew JC. Phenotypic screens as a renewed approach for drug discovery. *Drug Discovery Today*. 2013;18(21):1067-73.
12. Galletti C, Aguirre-Plans J, Oliva B, Fernandez-Fuentes N. Prediction of Adverse Drug Reaction Linked to Protein Targets Using Network-Based Information and Machine Learning. *Frontiers in Bioinformatics*. 2022;2: 906644.
13. Croston GE. The utility of target-based discovery. *Expert Opin Drug Discov*. 2017;12(5):427-9.
14. Sioud M. Main Approaches to Target Discovery and Validation. *Methods Mol Biol*. 2007;360:1-12.
15. Kubota K, Funabashi M, Ogura Y. Target deconvolution from phenotype-based drug discovery by using chemical proteomics approaches. *Biochimica et biophysica acta Proteins and proteomics*. 2019;1867(1):22-7.
16. Lee J, Bogoy M. Target deconvolution techniques in modern phenotypic profiling. *Curr Opin Chem Biol*. 2013;17(1):118-26.
17. Müller J, Hemphill A. Drug target identification in protozoan parasites. *Expert Opin Drug Discov*. 2016;11(8):815-24.
18. Müller J, Hemphill A. Chapter Seven - New Approaches for the Identification of Drug Targets in Protozoan Parasites. *Int Rev Cell Mol Biol*. 2013;301:359-401.
19. Müller J, Hemphill A. Drug target identification in intracellular and extracellular protozoan parasites. *Current topics in medicinal chemistry*. 2011;11(16):2029-38.
20. Myler PJ. Searching the Tritryp genomes for drug targets. *Advances in experimental medicine and biology*. 2008;625:133-40.
21. Sioud M. Main approaches to target discovery and validation. *Methods Mol Biol*. 2007;360:1-12.

22. Werbovetz KA. Target-based drug discovery for malaria, leishmaniasis, and trypanosomiasis. *Current medicinal chemistry*. 2000;7(8):835-60.
23. Wilkinson IVL, Terstappen GC, Russell AJ. Combining experimental strategies for successful target deconvolution. *Drug Discovery Today*. 2020;25(11):1998-2005.
24. Liang S, Guobo S, Xu X, Xu Y, Wei Y. Affinity Purification Combined with Mass Spectrometry-Based Proteomic Strategy to Study Mammalian Protein Complex and Protein-Protein Interactions. *Current Proteomics*. 2009;6:25-31.
25. Graven P. Genetic and chemical validation of *Trypanosoma brucei* adenosine kinase, the intracellular target of 4-[5-(4-phenoxyphenyl)-2H-pyrazol-3-yl]-morpholine 2013.p. 43-45.
26. Kuettel S, Greenwald J, Kostrewa D, Ahmed S, Scapozza L, Perozzo R. Crystal Structures of *T. b. rhodesiense* Adenosine Kinase Complexed with Inhibitor and Activator: Implications for Catalysis and Hyperactivation. *PLoS neglected tropical diseases*. 2011;5(5):e1164.
27. Kuettel S, Mosimann M, Mäser P, Kaiser M, Brun R, Scapozza L, et al. Adenosine Kinase of *T. b. Rhodesiense* identified as the putative target of 4-[5-(4-phenoxyphenyl)-2H-pyrazol-3-yl]morpholine using chemical proteomics. *PLoS neglected tropical diseases*. 2009;3(8):e506.
28. Kuettel S, Zambon A, Kaiser M, Brun R, Scapozza L, Perozzo R. Synthesis and Evaluation of Antiparasitic Activities of New 4-[5-(4-Phenoxyphenyl)-2H-pyrazol-3-yl]morpholine Derivatives. *Journal of medicinal chemistry*. 2007;50(23):5833-9.
29. Roux KJ, Kim DI, Burke B. BioID: a screen for protein-protein interactions. *Curr Protoc Protein Sci*. 2013;74:19.23.1-19.23.14.
30. Morriswood B, Havlicek K, Demmel L, Yavuz S, Sealey-Cardona M, Vidilaseris K, et al. Novel bilobe components in *Trypanosoma brucei* identified using proximity-dependent biotinylation. *Eukaryotic cell*. 2013;12(2):356-67.
31. Navarrete M, Wilkins JA, Lao Y, Rush DN, Nickerson PW, Ho J. Activity-based Protein Profiling Approaches for Transplantation. *Transplantation*. 2019;103(9):1790-8.
32. Brückner A, Polge C, Lentze N, Auerbach D, Schlattner U. Yeast two-hybrid, a powerful tool for systems biology. *Int J Mol Sci*. 2009;10(6):2763-88.
33. Fields S, Song O. A novel genetic system to detect protein-protein interactions. *Nature*. 1989;340(6230):245-6.
34. Walhout AJM, Boulton SJ, Vidal M. Yeast two-hybrid systems and protein interaction mapping projects for yeast and worm. 2000;17(2):88-94.
35. Mehla J, Caufield JH, Sakhawalkar N, Uetz P. A Comparison of Two-Hybrid Approaches for Detecting Protein-Protein Interactions. *Methods Enzymol*. 2017;586:333-58.
36. Licitra EJ, Liu JO. A three-hybrid system for detecting small ligand-protein receptor interactions. *Proceedings of the National Academy of Sciences of the United States of America*. 1996;93(23):12817-21.
37. Griffith EC, Licitra EJ, Liu JO. Yeast three-hybrid system for detecting ligand-receptor interactions. *Methods Enzymol*. 2000;328:89-103.
38. Chidley C, Haruki H, Pedersen MG, Muller E, Johnsson K. A yeast-based screen reveals that sulfasalazine inhibits tetrahydrobiopterin biosynthesis. *Nature chemical biology*. 2011;7(6):375-83.
39. Cottier S, Mönig T, Wang Z, Svoboda J, Boland W, Kaiser M, et al. The yeast three-hybrid system as an experimental platform to identify proteins interacting with small signaling molecules in plant cells: potential and limitations. *Front Plant Sci*. 2011;2:101.
40. Fellay C. Detection of Drug-Protein Interactions with an Improved Yeast Three-Hybrid System: EPFL.
41. Vanhamme L, Pays E. The trypanosome lytic factor of human serum and the molecular basis of sleeping sickness. *Int J Parasitol*. 2004;34(8):887-98.
42. Raper J, Fung R, Ghiso J, Nussenzweig V, Tomlinson S. Characterization of a novel trypanosome lytic factor from human serum. *Infection and immunity*. 1999;67(4):1910-6.
43. Gokhale RS, Khosla C. Role of linkers in communication between protein modules. *Curr Opin Chem Biol*. 2000;4(1):22-7.
44. Chen X, Zaro JL, Shen W-C. Fusion protein linkers: property, design and functionality. *Adv Drug Deliv Rev*. 2013;65(10):1357-69.

## **Part 3: Development of a new target deconvolution tool**

\* Note that the part 3.1.3 is part of the review paper (Part 1) that will be submitted

## .1 Introduction

### .1.1 *Trypanosoma brucei* parasites

#### .1.1.1 Classification and morphology

Trypanosomes are hemoflagellate protozoan parasites of the class Kinetoplastea, order Trypanosomatida, family Trypanosomatidae, and genus *Trypanosoma* (1, 2). These organisms belong to the class of kinetoplasts because of the characteristic presence of a dense structure called a kinetoplast, which takes the form of an array of circular mitochondrial DNA or kinetoplast DNA (kDNA) molecules located in a specialized part of the mitochondrion (1, 3).

Trypanosomes including other protozoan parasites such as *Leishmania spp.* are also characterized by the presence of a single flagellum in addition to the kinetoplast (4, 5).

Trypanosomatids can adopt six main morphologies, described in the figure hereunder:

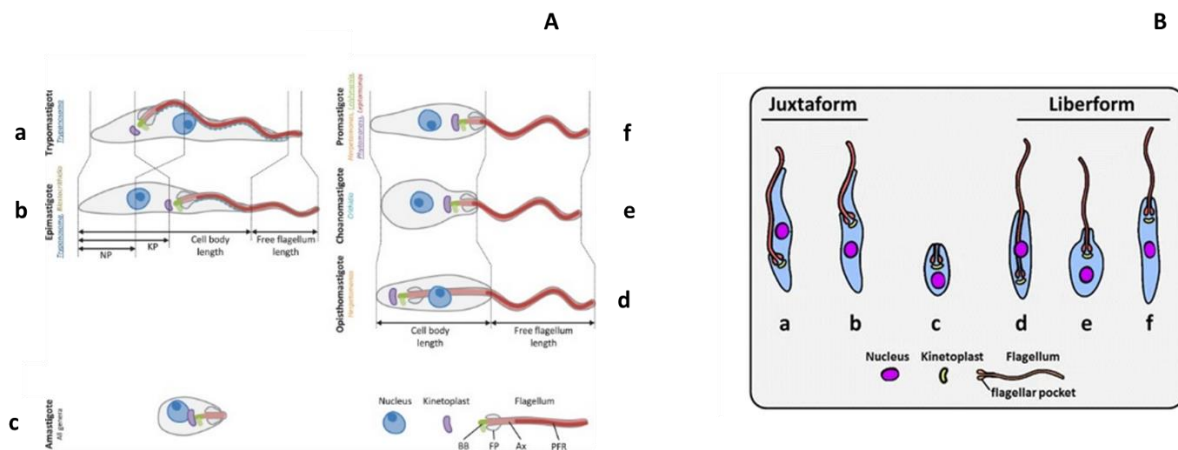


**Figure 1:** The six major morphologies adopted by trypanosomatids (image A taken from (6) and image B taken from (7))

The parasitic species *Trypanosoma brucei* (*T. brucei*) adopt only two forms: The trypomastigote (Figure 1A (a)) and epimastigote (Figure 1A (b)) morphologies (7).

Two morphologically indistinguishable subspecies of *T. brucei* cause sleeping sickness, each responsible for a particular form of the disease. The first subspecies *T. b. gambiense* (**Tbg**) is responsible for chronic African trypanosomiasis, a disease found mainly in West Africa, while the second subtype *T. b. rhodesiense* (**Tbr**) is responsible for acute African trypanosomiasis, a form of the illness primarily presents in East Africa. Yet, *T. brucei* has a third species known as

*T. b. brucei* (**Tbb**). This subtype of the parasite mainly infects cattle and is not pathogenic to humans (2, 8). Indeed, Human beings are naturally immune to this subtype owing to the presence of trypanosome lytic factors (TLFs) allowing their lysis (9, 10).

Trypanosome lytic factors (TLFs) are high-density lipoproteins that belong to human innate immunity (11, 12). These factors fall into two categories: TLF1 and TLF2. Both contain the apolipoprotein L1 (APOL1) which seemed to be important for trypanolytic activity (13, 14). Along with APOL1, they also share additional proteins including haptoglobin (Hp)-related protein and apolipoprotein A1 (APOA1) (Hpr). Apolipoprotein A2 (APOA2), paraoxonase, and hemoglobin are the key distinctions between TLF1 and TLF2, in addition to the fact that TLF1 is a smaller and more lipid-rich protein whereas TLF2 is larger and less lipid-rich. However, it has been noted that TLF2 is reportedly linked to immunoglobulin M (IgM), which accounts for its increased size (about 1000 kDa compared to about 500 kDa for TLF1) (10, 11, 13).

The other two subspecies of *T. brucei* (**Tbr** and **Tbg**) have developed resistance to TLFs, which explains their ability to infect humans and induce sleeping sickness (10, 12, 13).

The Serum Resistance-Associated protein (SRA), a shortened form of the Variant Surface Glycoprotein VSG that appears to interact with APOL1 and impede APOL1-mediated parasite lysis, mediates **Tbr** resistance to TLF (12, 15, 16). Hemoglobin (Hb) seems to enhance the activity of Hpr, which is thought to be involved in trypanosome lysis together with APOL1 in TLF1 (12, 14, 17-22). However, the Hpr/Hb combination in human serum would bind to the Hp/HbR receptor of **Tbb** (**TbbHpHbR**) which induces its lysis (17, 23, 24). Unlike **Tbr**, **Tbg** does not contain the SRA gene, and its resistance to TLFs seems to be multifactorial. Likewise, it has been demonstrated that a reduction in the expression of **TbgHpHbR** would cause this subtype of the parasite to escape from human innate immunity, explaining its pathogenicity and susceptibility to infection (17). In addition, an L210S substitution would inactivate the **TbgHpHbR** receptor conferring resistance (25). Another factor contributing to the defense mechanism would be the presence of a parasite-specific glycoprotein (TgsGP) preventing APOL1 toxicity (25-27).

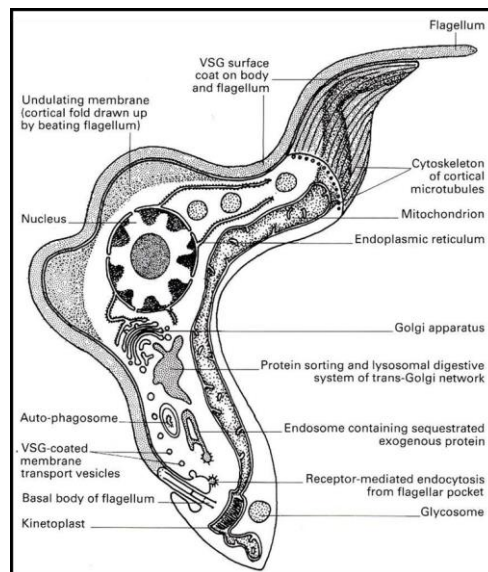


*Figure 2: General structure of Trypanosoma brucei (BSF) and its main components, image taken from (28, 29)*

As illustrated in Figure 2 above, *T. brucei* parasites have a very dense coat of glycoproteins called VSGs (Variant Surface Glycoproteins) covering the surface of their plasma membrane (30).

VSGs represent the main surface antigens. Indeed, when the parasite infects its host, it undergoes an antigenic variation during which the expression of the VSG is modified. This is achieved by drawing from a genomic repertoire of more than 1000 different genes. Each gene encodes a unique variant of the VSG, produced only once and encompassing the entire surface of the plasma membrane (30).

These VSGs are highly immunogenic, causing an immune response that allows the production of antibodies against the type of VSG expressed on the parasites' envelopes. However, some of them will evade the immune response by modifying the VSG expressed on the surface, which induces a re-increase in parasitemia. These fluctuations in parasitemia are a defining feature of HAT disease and are brought on by this antigenic variation, involving this cyclical expression of a new VSG resulting in the appearance of new mutants in the host blood (28, 30-33).

VSGs are characteristic of bloodstream forms. For procyclic variants, procyclins, which as VSGs, are glycosylphosphatidylinositol-anchored proteins, are the primary surface proteins of procyclic parasites (33-35). Unlike VSGs, procyclins are not required for procyclic cultures (33,

36). The expression of these VSGs takes place in the salivary gland of the tsetse fly, but this process remains poorly understood (37).

The three subspecies of *T. brucei* are morphologically indistinguishable. **Tbb** is lysed by TLFs as previously detailed but how to distinguish between **Tbr** and **Tbg** in order to adapt the treatment according to the targeted strain?

As mentioned above, one mechanism of action that allows **Tbr** parasites to escape TLFs present in human serum is the presence of serum resistance-associated protein (SRA) (15, 38). The gene encoding for this protein has, therefore, been taken as a diagnostic tool to be able to detect and validate the presence or absence of **Tbr** strains in the patient's blood (38).

One further distinctive feature of the parasite's morphology is the presence of a single flagellum which is essential for the mobility and thereby the propulsion of the cell. In addition, it showed to be an important organelle in morphogenesis as well as in cellular cytokinesis. Furthermore, it has been proven that the flagellar membrane hosts many factors involved in parasite virulence entailed in host-parasite signaling and interactions (39).

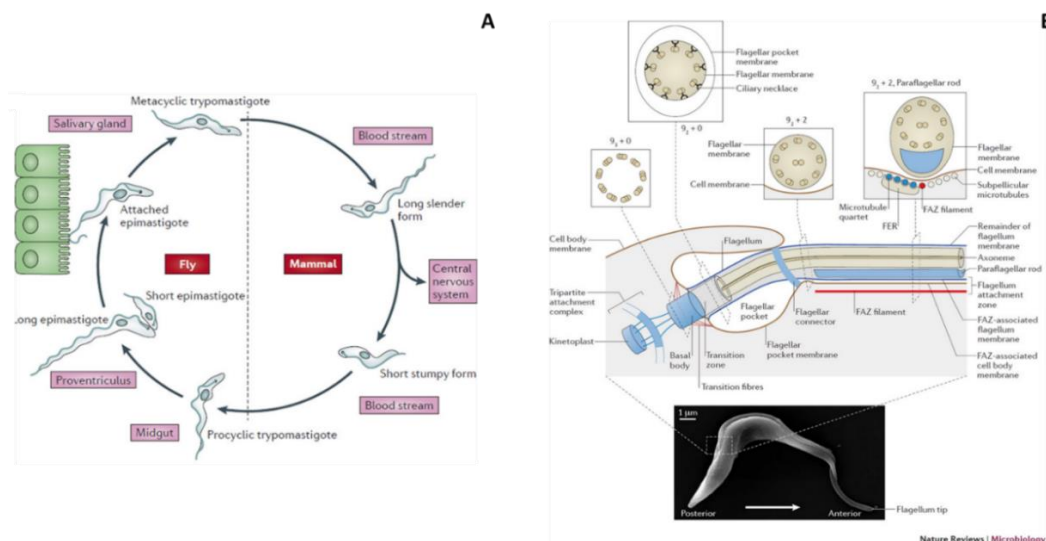


**Figure 3 :** A: Presence of the flagellum during the different stages of the life cycle of the parasite; B: details of the flagellum's composition (Images taken from (39))

The flagellum is always present in the parasite during each step of its life cycle, as depicted in Figure 3 above, and is necessary for its survival (39). This latter, as illustrated in part B of Figure 3, is composed of a central pair of microtubules forming the 9 + 2 axoneme which represents

the main source of motility of the parasite (39, 40). It emerges from the basal body (barrel-shaped in blue in Figure 3B), emerges from the cell body membrane at the flagellar pocket, and extends externally, forming this flagellar structure (39, 41).

#### *.1.1.2 Vector of Trypanosoma brucei: The tsetse fly*

The tsetse fly belonging to the genus *Glossina* is the vector that transmits the *T.brucei* parasites responsible for HAT disease (42). There are three main subgroups of tsetse flies classified according to the environment they inhabit: *Glossina palpalis* (*G.palpalis*) (riverine tsetse fly), *Glossina morsitans* (*G. morsitans*) (savannah tsetse fly ) and *Glossina fusca* (*G. fusca*) (forest-dwelling tsetse fly) (43). Although all tsetse fly species are capable of spreading the HAT, *G.palpalis* spp. and *G.fuscipes* spp. represent the main vectors engaged in HAT transmission (43).

The fusca subgroup of *Glossina* is generally implicated in the transmission of animal trypanosomiasis to cattle, while the palpalis subgroup, including mainly *G. palpalis*, is responsible for the transmission of the *T.b gambiense* (**Tbg**) subtype of the parasite in West and Central Africa (44). The *T.b rhodesiense* (**Tbr**) is transmitted by the morsitans subgroup, primarily *G. morsitans*, which transmits the disease in East Africa explaining the spread of its acute form in this geographical area (44).

*Glossina* species are characterized by tawny or brown-colored flies that vary in size depending on the subgroup (45). The largest flies, measuring between 9.5 and 14 mm, belong to the fusca subgroup (not including the proboscis). The species in the palpalis and moristans subgroups are smaller, with sizes ranging from 6.5 to 11 mm (excluding the proboscis) (45).

#### *.1.1.3 The genome of T. brucei*

The haploid nuclear genomic DNA of the *T. brucei* parasite is about 35 Mb in size. Its diploid DNA in the nucleus consists of 3 different types of chromosomes that differ in size. Indeed it consists of around 11 pairs of megabase chromosomes with sizes ranging from 1 Mb to 6 Mb, several chromosomes of 200 to 900 kb called intermediates, and about 100 linear minichromosomes of 50 to 150 kb (37).

The VSG genes are found on the subtelomeres of all different classes of *T. brucei* chromosomes (46).

Minichromosomes house the untranscribed base copies of VSG genes. Their expression sites have also been found on intermediate chromosomes, and their base copies and expression sites have been identified in megabase diploid chromosomes (47), which would also contain most of the transcribed genes (48-53).

Mitochondrial DNA, also known as kinetoplast DNA (kDNA), is localized in a peculiar structure of kinetoplastid protozoa, including trypanosomes, and this organelle is called kinetoplast (54, 55). The kDNA is a massive network made up of thousands of minicircles and several dozens of maxicircles (55-58). The kinetoplast is located near the flagellar basal body (55). Maxicircle DNAs encode rRNAs and certain proteins, including those belonging to respiratory complex subunits, and their transcripts require editing involving the insertion or deletion of uridylate or uridine residues at specific sites to form an open reading frame. As for minicircles, their DNA encodes guide RNAs that have the function of editing maxicircle DNA transcripts to produce functional mRNAs (55, 56, 58-60).

The mRNA processing in trypanosomatids differs from that of other eukaryotic groups, including fungi, plants, and mammals (61). Indeed, in kinetoplastids, gene transcription is polycistronic and requires a trans-splicing process, individual mature mRNAs are obtained by trans-splicing which consists of adding a short capped sequence at the 5' end, and this process is coupled with polyadenylation of the previous mRNA in order to stabilize and obtain the mature mRNA coding for proteins (61-65). These mechanisms allowed the emergence of treatments used against human and animal trypanosomiasis that target mRNA maturation, such as AN7973, a benzoxaborole-based compound that was found to rapidly inhibit mRNA trans-splicing (within one hour) (63).

#### *.1.1.4 Life-cycle of Trypanosoma brucei*

As mentioned in section 1.1.1, the *T. brucei* parasite can take on one of two morphologies depending on the infected host: The parasite adopts the trypomastigote shape when it is within its permanent mammalian host (human or animal). In this form, the kinetoplast and basal body are posterior to the nucleus whereas they are anterior when the parasite takes the

epimastigote morphology. The tsetse fly of the genus *Glossina* contracts blood trypomastigotes when it consumes a blood meal from an infected mammalian host, as seen in the illustration below that depicts the parasite's life cycle (Figure 4). The blood trypomastigotes are converted into procyclic forms in the infected fly's midgut. These multiply through binary fission. Trypomastigotes take on an epimastigote form after they exit the midgut and proliferate via binary fission in the insect's salivary glands. During the fly's blood meal on the mammalian host, these epimastigotes will later change into metacyclic trypomastigotes and be injected into the skin tissue of the latter. After passing from the lymphatic system to the bloodstream, these parasites assume the morphology of blood trypomastigotes that continue to replicate by binary fission and are transported to other body fluids such as cerebrospinal fluid or lymphatics. These different life stages of the *T. brucei* parasite are only extracellular, whereas the *Trypanosoma cruzi* (*T. cruzi*) parasite can adopt an intracellular phase (2).

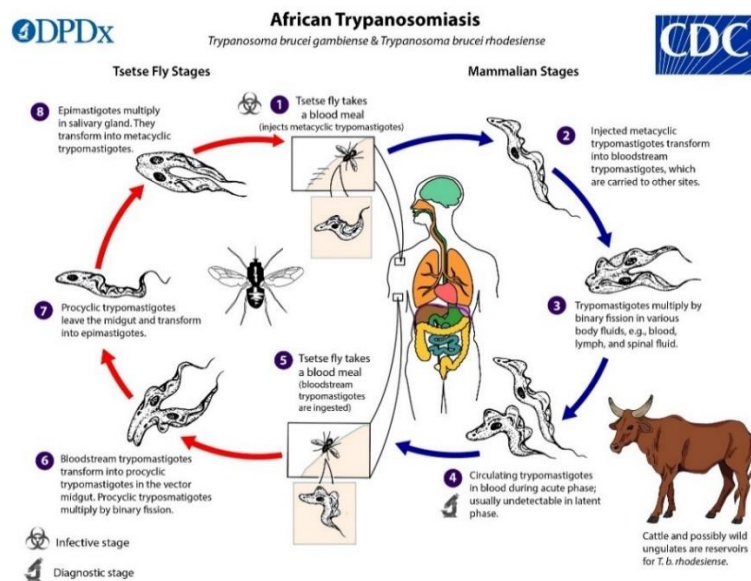


Figure 4: Life-cycle of African trypanosomes (Image taken from (2))

According to the infected host, *T. brucei*'s life cycle is divided into two primary stages, as indicated in Figure 4 above. When infecting either a human or a vector (the tsetse fly), the parasite will take either the procyclic form (PCF) or the bloodstream shape (BSF) (66).

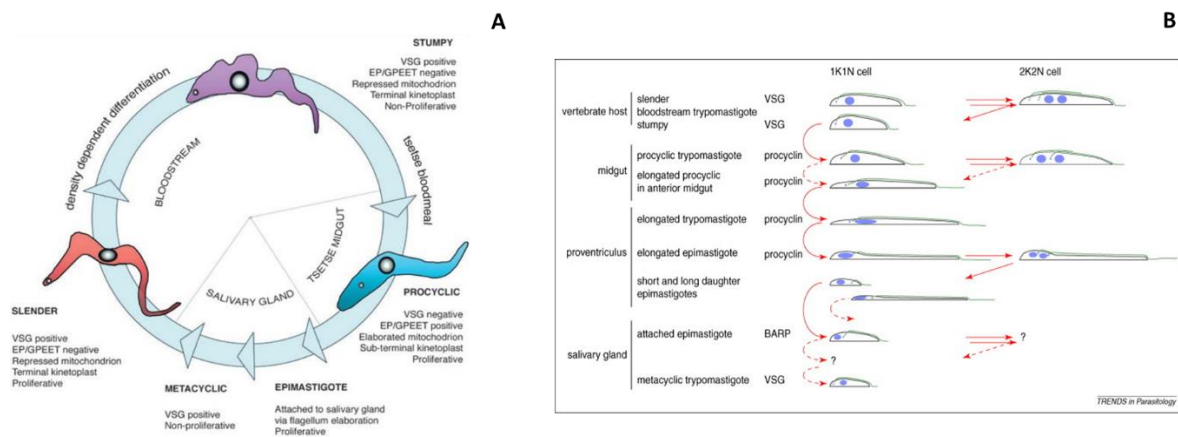


**Figure 5:** Different morphologies taken by *Trypanosoma brucei* parasite into the vector and mammalian host, image A taken from (66) and B taken from (67)

As the *T. brucei* grows in the bloodstream, its shape alters to take on a slender appearance (Figure 5 above). When the overgrowth is reached, the parasites switch to a non-proliferating stumpy form, which not only controls their number in circulation by preventing overgrowth and thereby extending their survival but also ensures coordination of the morphological changes that occur during the time the parasites are transmitted into the vector (tsetse fly) and upon reintroduction into the cell cycle since division arrest occurs in the G1 phase of the cellular cycle (66, 68). As mentioned earlier, and illustrated in the diagram above (image A), during the blood stage, the slender forms express VSGs glycoproteins allowing them to escape and bypass the immune system by the antigenic variation that enables the expression of a VSG distinct from the one that was previously expressed in response to the development of antibodies against the type of the glycoprotein that was previously expressed (69).

As outlined, the morphological transition into stumpy forms enables the transmission of parasites to tsetse flies when they consume a blood meal (66). Procyclic forms are then produced and these proliferate in the midgut of the fly. These forms also express surface glycoproteins which are the EP and GPEET procyclins (70). Following this, the parasites go from the midgut to the salivary glands, where they attach with their flagella and transform into a proliferating epimastigote (66). Finally, they transform into non-proliferating metacyclic forms that re-acquire the VSG coat, ready to be transmitted to a mammalian host (66).

Figure 6 below also effectively illustrates the parasite's ability to reorganize its internal organelles, including its kinetoplast, nucleus, and flagellum, according to the various stages of its life cycle (66, 67, 71).

As described in section 1.1 (part 1.1.1), various species of the kinetoplastids exhibit additional or distinct morphologies from those adopted by *T. brucei*. In addition to the trypomastigote and epimastigote stages that are shared with *T. brucei*, *T. cruzi* also exhibits the amastigote type that appears in the cytoplasm of an infected cell (67, 72). *Leishmania* species adopt the amastigote (proliferative aflagellate types found in mammalian host cells) and promastigote (proliferative flagellate types found in female sandflies) variations (67, 72-74).

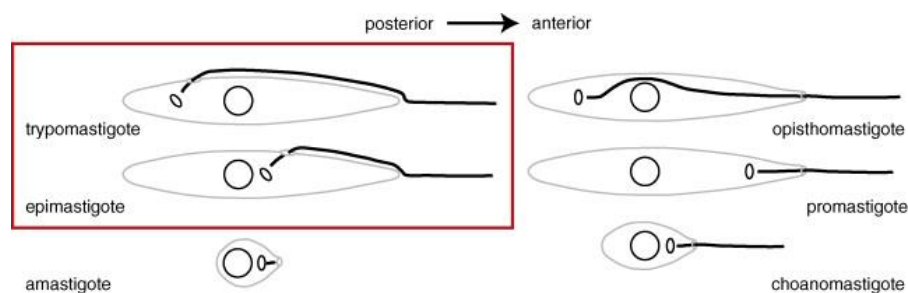


*Figure 6: Arrangement of organelles according to the different morphologies adopted by the kinetoplastid species (Image taken from (67))*

The only two morphologies taken by *T. brucei* are those highlighted in red in the figure above.

#### *.1.1.5 African Trypanosomiasis*

Human African Trypanosomiasis (HAT), often known as sleeping sickness, is provoked by **Tbr** (which is responsible for the acute form) and **Tbg** (which is responsible for the chronic form) and can be fatal if left untreated. These two forms are well described above in Part 1 (Review).

Animal trypanosomiasis also known as the Nagana pest is induced by **Tbb**. This latter infects animals and therefore livestock, which represents an enormous economic challenge and a considerable loss of income for sub-Saharan Africa, leading to a significant increase in poverty in the affected areas (30).

There are two distinct types of sleeping sickness (or HAT) which can be identified by two different stages, as was discussed in the review (Part 1): The hemolymphatic phase, also referred to as stage I, starts after a tsetse fly bite and is marked by the parasite's proliferation

in the blood and lymphatic systems as well as the body's internal organs. This phase is characterized by nonspecific symptoms such as headache, itchiness, joint pain, fever, weight loss, arthralgias, fatigue, or general malaise (75, 76).

The late phase (stage II), also known as the encephalopathic state, presents with more specific symptoms characterized by neurological signs such as tremors, numbness, poor coordination and general muscle weakness, confusion, and disruption of the sleep cycle, which is the main symptom of the disease and has led to its renaming: Sleeping sickness (76). A wide variety of mental and psychiatric symptoms, including anxiety, agitation, indifference, irritability, mania, and hallucinations were also documented (76). If neglected, this stage frequently results in coma and death (76-79).

#### *.1.1.6 Geographical distribution of the disease*

The **Tbg** subtype, which causes the chronic form of the disease distinguished by its extremely lengthy duration, is the most prevalent variety in West and Central Africa. Indeed, signs start to appear after a period of several months to several years (about 2 years) during which the parasites start to settle in the central nervous system (CNS) after crossing the blood-brain barrier, and usually, when symptoms start to develop, the patient is already in an advanced stage of the disease (76-79). The lack of symptoms during the initial phase suggestive of a **Tbg** parasite infection is explained by the low parasitemia of this subtype at the onset of the disease (8, 75, 79).

In eastern and southern Africa, the most common subtype is the **Tbr** parasite, which is responsible for the acute form of the disease (78, 80). Rhodesiense HAT causes a rapid onset of symptoms that can be seen within a few months or weeks of infection (80, 81). Then, the parasite quickly starts to invade the central nervous system. In contrast to HAT gambiense, HAT rhodesiense is a more acute version that often lasts a few weeks to months before death occurs if the disease is not treated swiftly (77, 78, 80, 81).

Figure 7 below shows the distribution of sleeping sickness according to *Trypanosoma brucei* subtype:

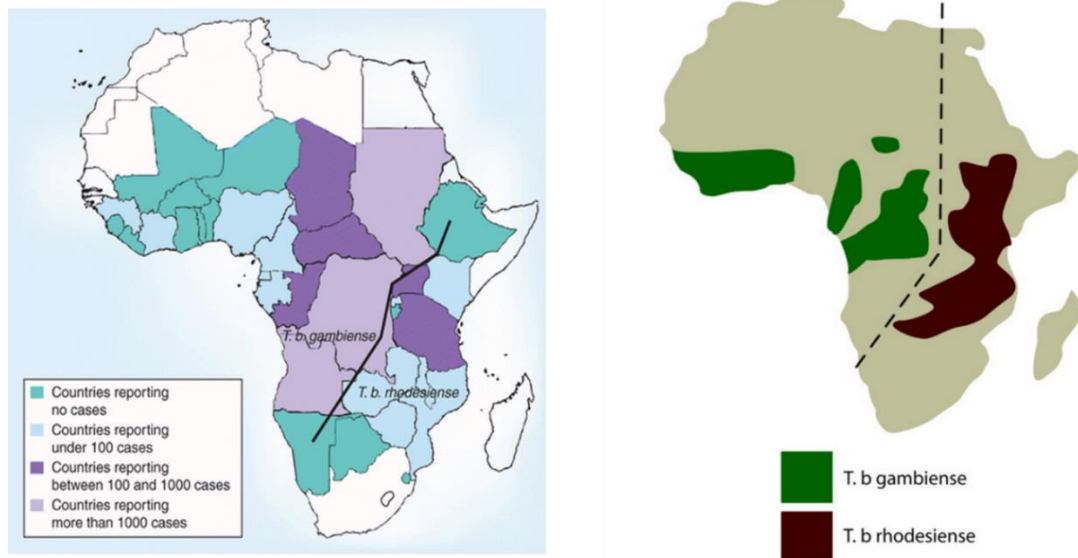


**Figure 7:** Distribution of sleeping sickness disease according to the *Trypanosoma brucei* subtype (Left image taken from (75) and right image taken from (76))

Gambiense HAT accounts for approximately 95-97% of all reported cases, while rhodesiense HAT represents only about 3-5% of cases (77, 80).

## .1.2 A small overview of other neglected tropical diseases caused by parasite of the same family

### .1.2.1 American Trypanosomiasis or Chagas disease

American Trypanosomiasis also called Chagas disease, is caused by *T. cruzi* which is another protozoan flagellate parasite belonging to the same family, class, order, and genus as *T. brucei* and is responsible for one of the major neglected tropical diseases. Unlike *T. brucei*, *T. cruzi* is able to alternate between the intracellular and extracellular stages (82). The infection of mammalian hosts including humans occurs through feces or urines of infected triatomine insects commonly known as “kissing bugs” (82-84). There are many other routes of transmission, including vector bites, transfusion of infected blood or organ transplants, consumption of food contaminated with vector excreta or secretions from infected mammalian hosts, also from laboratory exposure or an infected mother to her new-born during pregnancy or delivery (82, 85).

The parasites are encouraged to enter the body when the individual instinctively spills the bug's feces or urine into the bite, eyes, mouth, or skin sores. These bugs bite on exposed portions of the skin, such as the face, and then defecate or urinate nearby (86).

For more details about the disease, refer to Part 1 (the review).

#### *.1.2.2 Leishmaniasis*

Other parasites belonging to the same class, family, and order but differ in genus cause leishmaniasis. The disease is brought on by numerous species of the genus *Leishmania* and is spread to mammalian hosts, including humans, via the biting of infected female sandfly vectors, which feed on blood to create eggs (87, 88). The two major morphologies adopted by this parasite are the extracellular promastigote found in the digestive tract of the vector and the intracellular amastigote detected, after the phagocytosis phase of the promastigotes, in macrophages and other types of mononuclear phagocytic cells of the mammalian host (89, 90) (See Part 1 for more details).

Leishmaniasis is divided into three main forms: Visceral leishmaniasis (VL) also called *kala-azar*, fatal in over 95% of cases if left untreated, caused primarily by *Leishmania donovani* or *Leishmania infantum* (infecting also dogs and cats (91)), and leads to irregular fever outbreaks, weight loss, enlargement of the spleen and liver and anemia; Cutaneous leishmaniasis (CL), the most prevalent form of the sickness brought on by several species of *Leishmania* such as *Leishmania aethiopica*, *Leishmania major*, *Leishmania tropica*, *Leishmania mexicana*, and *Leishmania amazonensis*. This form is responsible for the appearance of ulcers or stigmata, leaving lifelong scars; Mucocutaneous leishmaniasis (ML) induced by certain species, including *Leishmania braziliensis*, *Leishmania panamensis*, and *Leishmania peruviana* (92-97), and characterized by destruction (partially or totally) of mucous membranes of the throat, nose, and mouth. Unfortunately, due to the existence of many diverse species, developing effective vaccines and treatments remains challenging (98-100).

These three major neglected tropical diseases have been described more in detail in the review (Part 1).

### **.1.3 Current target deconvolution methods as starting technologies for the development of the new tool**

Recognizing the importance of identifying a medicine's target to avoid potential side effects and understand its mechanism of action, several deconvolution techniques have been developed and extensively reviewed (refer to Part 1). These tools enable accurate identification and exploration of the desired target, simplifying the selection of potential hits and their optimisation towards producing the desired outcomes.

Revolutionary and highly successful target deconvolution approaches have been developed and widely used. These methods not only provide insight into the interactions between a molecule and its target(s) but have also facilitated the establishment of studies on the activity of the drug of interest in its target (101). Affinity chromatography, which is frequently used in conjunction with the proteomic approach for later target identification, is one of the most used target-based techniques (102). Indeed, this method, which consists of attaching the molecule of interest to a matrix in order to capture its target by elution (the target can be DNA, RNA, or proteins), is easy to handle, rapid, and reproducible (101, 103, 104).

Many additional techniques have also been found to be effective and beneficial, including metabolomic, computational, genetic, and genetic techniques, such as genome mining, DNA sequencing, and reverse genetics, which includes knockout, RNA interference (RNAi), overexpression of a gene of interest, and the method of generating resistant strains, which is widely used in parasitology to identify the proteins involved in resistance mechanisms (101, 103-105). In the review (Part 1), each of these techniques has been described in detail.

Another pair of cutting-edge approaches that have been developed and are widely employed are BioID and SNAP-tag technologies that are central approaches for this thesis and the aimed development of the novel deconvolution tool. Thus, for the thesis the details of these strategies including strengths, weaknesses and opportunities are described in chapter 3.1.3.1 and 3.1.3.2.

### *.1.3.1 BioID technology and its variants*

The BioID approach is a recent method based on a mutated form of the *E. coli* biotin ligase, called BirA\*, which produces a highly reactive intermediate called biotinoyl-5'-AMP. This mutated version of biotin ligase has lost its affinity for the intermediate, which leads to its promiscuous release into the surrounding environment. In the presence of ATP and biotin, biotinoyl-5'-AMP will be able to biotinylate all partners in its immediate vicinity (within a 10 nm radius (Figure 8 below) by forming an amide bond between the lysine residue of proteins and biotin (106, 107).

Numerous studies have already used the fusion of BirA\* to a protein of interest to show that the BioID approach is a potent and effective method for the detection of protein-protein interactions in particular, but also for the identification of novel components in the *T. brucei* cytoskeleton that are typically difficult to isolate due to their intractability and weak interactions (106-108).

The BioID method has several benefits, including the ability to use harsh lysis conditions, such as strong detergents, without losing information on interacting proteins owing to the high affinity between biotin and streptavidin ( $K_d$ ,  $10^{-14}$ M), the ability to detect transient and weak interactions, and because biotinylation occurs before cell lysis, BioID is not impacted by the solubility of interacting proteins in a test tube. Moreover, it may be used with a variety of cell types and species, including HEK-293, HeLa, parasites, and more (106-110). Since this method occurs in living cells, our understanding of cell structure and function could be improved. Even while this strategy has a lot to offer in terms of advantages, there are some drawbacks. Indeed, as a result of biotinylation's promiscuity, background signals may appear. Additionally, BirA\*'s small labeling radius (10 nm) may prevent intriguing targets from being biotinylated (108). The effectiveness of BioID also depends on the availability of primary amines (lysine residues), and BirA\* size (35 kDa) may affect the functions and characteristics of the fusion protein (111, 112). Pull-down or immune-precipitation techniques are typically used for the isolation of biotinylated proteins, given their simplicity of implementation. However, these techniques can be quite sensitive to the conditions present throughout the various cell lyses, particularly when isolating insoluble proteins, which demand more sophisticated procedures specifically

due to their significant interactions with multiple proteins in an overly complex environment. The utilization of intricate experimental designs would be required in this situation (110).

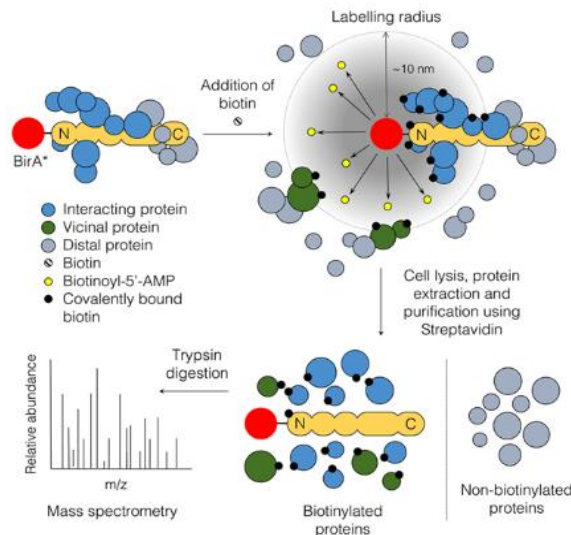


*Figure 8: Biotinylation of proteins via BirA\* (108)*

Along with BioID, APEX2 (27 kDa), a triple mutant of wild-type soybean ascorbate peroxidase (APX) obtained by structure-guided mutagenesis and screening, is another enzyme frequently used for enzymatically catalyzed proximal tagging. It is an improved second-generation version with greater labeling capacity than the original APEX. The presence of H<sub>2</sub>O<sub>2</sub> to initiate labeling is required for APEX2 to be functional (113-115).

One of the advantages of APEX2 is its speed, as proximal protein labeling occurs in 1 min or less (116, 117). However, in contrast to BioID labelling, it requires the use of H<sub>2</sub>O<sub>2</sub>, which can be toxic for living samples. Indeed, with BioID only biotin, which is not toxic, is required to begin the tagging, making the BioID approach one of the most popular methods in the field of proximity protein labeling with over 100 applications since its introduction in 2012 and cited in over 300 articles (116, 118, 119).

A smaller and improved version of BioID, called BioID2, was developed from a thermophilic bacterium *A. aeolicus*. The improved version has many advantages, such as greater sensitivity which requires less biotin supplementation, more selective targeting of fusion proteins, and better labeling of nearby partners. However, the labeling time and activity are still comparable to the original BioID, meaning that it still requires more than 16 hours of incubation (109, 116, 119).

The main limitation of BioID and BioID2 is their slow kinetics as they require a long incubation time (>16 hours) (116). Two new promiscuous ligases based on BioID technology have recently emerged. These new modified variants are called TurboID and miniTurbo (116, 119). Both were generated via BioID ligase-directed evolution, and they produced rapid and robust biotinylation after only 10 minutes of incubation with extra biotin, as opposed to 18–24 hours with the original BioID method (116, 119). Additionally, the miniTurboID protein, which is smaller (28 kDa) than BioID and TurboID (35 kDa), may be an alternative if their fusion protein configuration results in steric hindrance. MiniTurbo has 13 mutations with a loss in the N-terminal domain, while TurboID has 15 mutations compared to the wild type of BirA (116).

According to some research, TurboID can generate nearly as much biotinylated protein in 10 minutes as BioID2 or even the original BioID approach can in 18 hours. MiniTurbo is a good option when careful control of labeling at a specific moment is necessary since it was demonstrated in some experiments that the protein was 1.5 to 2 times less active than TurboID but displayed less labeling before the addition of exogenous biotin (116).

In addition, with TurboID, it has occasionally been shown that both protein instability, sustained biotinylation in the absence of exogenous biotin supply, and an increase in labeling radius contribute to a rather strong background (119). However, in some cases, TurboID demonstrated efficient and robust biotinylation in cellular sub-compartments such as in the lumen of the endoplasmic reticulum compared to the original BioID (119).

Studies comparing TurboID, miniTurbo, and the original BioID in each of the following organelles: the lumen of the endoplasmic reticulum, its membrane, the nucleus, and the mitochondrial matrix of HEK 293T cells showed a difference in their activity depending on the organelle. Indeed, In the lumen of the endoplasmic reticulum as well as in the mitochondrial matrix, TurboID-induced labeling demonstrated a greater signal after 10 min of incubation than BioID did after 18 h of incubation with extra biotin. Additionally, in all four compartments of the cell, TurboID was found to have a higher activity than miniTurbo. Therefore, these two technologies will be assessed as part of this thesis topic in order to compare their effectiveness with BioID (116).

Furthermore, it was revealed in the large-scale proteomic experiments that, despite TurboID's faster labeling speed, BioID was just as active in enriching proteins in the endoplasmic reticulum membrane facing the cytosol as compared to proteins in the cytosol. As a result, the labeling radius of these two technologies is comparable (116). Considering that the targeted proteins are primarily found in the parasite cytosol, this raises the interest level of testing TurboID and miniTurboID fused to SNAP during this project.

### *.1.3.2 SNAP-tag technology*

SNAP-tag is a revolutionary technology based on a modified version of human O<sub>6</sub>-alkylguanine-DNA-alkyltransferase (hAGT), a 20 kDa DNA repair protein that has the specificity of being covalently attached to its substrate O<sub>6</sub>-benzylguanine (**BG**), via a reactive cysteine residue (Cys145) of its active site (120-123) (Figure 9). The **BG** substrate can carry different markers including dyes, fluorophores such as fluorescein (BG-FL) or diacetylfluorescein (BG-FDA), guanine-conjugated beads, biotin, or even drugs such as methotrexate (BG-Mtx), which leads to a panel of BG derivatives (124-126).

Several techniques, including the yeast two-hybrid system, or fusions with fluorescent proteins, have certain shortcomings since they rely on non-covalent interactions between a small molecule and its target, making it very challenging to specifically label a protein *in vivo* in order to understand its function (123). Accordingly, for target identification and characterization in this study, SNAP-tag is a great system to use. Indeed, this technology is a potent tool for protein labeling and has shown many advantages over other tags, including the irreversible self-labeling of SNAP fusion proteins in the living cell prompted by covalent binding to its substrates, its high specificity, the avoidance of non-specific labeling induced by the low intrinsic reactivity and affinity of its substrates for other proteins or else many substrates have strong cellular permeability, allowing the tagging of intracellular proteins in living cells (125). There have been several reported limitations to this technique, such as the difficulty in tagging intracellular proteins attributable to the SNAP substrate's ability to pass through membranes when bulky groups are added (127). Furthermore, it can be challenging to evaluate protein dynamics over a limited time frame due to the labeling and washing processes, which can take up to an hour (127). Another drawback is the need for exogenous expression of the fusion protein which can affect cellular pathways (128).

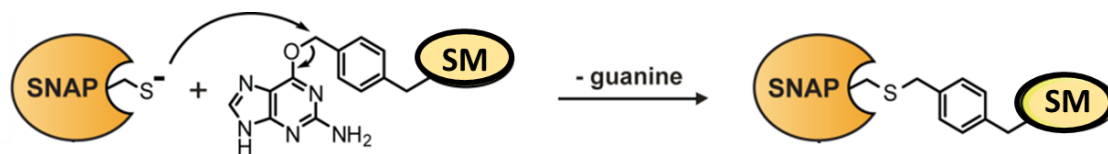


Figure 9: The SNAP-tag labeling technique (Figure modified from (130) )

The labeling reaction is shown in Figure 9 above. The BG carrying the molecule of interest that is chemically bound is covalently attached to SNAP through the reaction between the substituted benzyl group of the BG substrate and the cysteine of the active site, giving rise to the transfer of the compound to the protein (124-126, 130).

### .1.4 Challenges of the project

The ultimate goal of this thesis project, as outlined in Part 2, is to develop a highly sensitive and specific tool that can effectively capture the target of a molecule of interest *in vivo*.

This tool will be generated by joining two proteins, BirA\* and SNAP-tag, which have displayed extraordinary abilities, particularly in their capacity to capture the target of interest as well as any potential interference partners nearby (via the BirA\* domain) and to improve its selectivity of the interaction (through the SNAP-tag domain).

The development of this new system includes foreseeable challenges. Indeed, answering several scientific questions will be necessary for the project to reach its intended conclusion: In the course of this work, the cMyc\_BirA\*\_Linker\_SNAP-tag fusion protein (Figure 10 here under) will be overexpressed in the NYSM strain of the *Tbb* parasite derived from Lister 427 which is a cell line that simultaneously expresses the T7 RNA polymerase and the tetracycline repressor TetR (131).

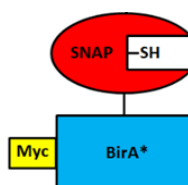


Figure 10: Protein of interest overexpressed in *Tbb*

The first set of questions will focus on the molecular biology part and are: Will the fusion protein be well expressed and if so, will it be toxic to the parasite? BirA\* is a large protein of 35.4 kDa (108), could its fusion with SNAP-tag alter the function of each domain of the fusion protein? Does the nature and length of the linker between BirA\* and SNAP-tag matter for the proper functioning of the biological activity of each module? What would be the optimum amino acid composition for an ideal linker? Furthermore, as mentioned in section 3.1.3.1, one of the limitations of BirA\* is its promiscuous biotinylation of all proteins within a radius of about 10 nm (107), as well as the potential for protein interactions in the cellular environment that might increase the background signals (132, 133). Will biotinylation of the desired target be accompanied by considerable background signals?

Based on the literature additional hurdles to overcome can be anticipated, particularly concerning the optimization of the pull-down protocol for biotinylated proteins, as the parasites are covered with a thick layer of glycoproteins (VGFs) as described previously (section 1.1) (30). Specific detergents and various adjustments will have to be established in order to be able to isolate and detect the proteins of interest without losing them during the different isolation processes.

The fusion protein will be *in vivo* tagged with synthesized BG derivatives after being overexpressed under tetracycline induction. The interaction between the small molecule (SM) substrate (Figure 11) and its target will bring it into the vicinity of BirA\* which should biotinylate the SM target upon the addition of biotin. The second series of scientific investigations relates to the project's chemical part: Does the target molecule already have a chemical functionalization that permits direct binding to BG derivatives or will more processing—perhaps functionalizing the drug candidate—be required to provide the desired result?

Will the BG derivatives be able to readily pass the parasite membrane once they have been properly produced and chemically characterized, including identification and stability testing under test conditions? If not, is there a way to force their crossing through the membranes?

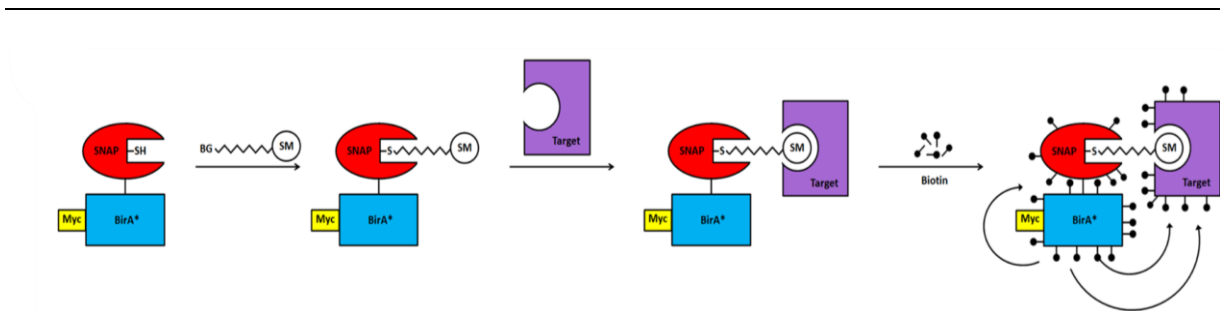


*Figure 11: the fusion protein labeled in vivo with BG derivative will bring the interacted target in proximity to the BirA\* for the biotinylation*

As noted, there will be a large number of scientific questions that need to be answered, and both the biology and chemistry sections will present how these various challenges and questions have been answered.

The isolation phase is the part that directly follows the labeling step: Owing to the extraordinary affinity between biotin and streptavidin ( $K_d = 10^{-14}\text{M}$ ), the isolation of biotinylated proteins will be performed by biotin-mediated affinity purification using streptavidin-coated Dynabeads as well described in both section 2.2.25 below and the article by Dr. Brooke Morriswood et al. (106). The proteins will be electrophoretically separated and identified by western blotting with a streptavidin fluorophore-bound to recognize biotinylated partners. This identification will be verified by LC-MS/MS (ESI-LC-MSMS) mass spectrometry after trypsin digestion, either directly on the beads or bands of interest excised from SDS-PAGE gels. The preparation of the samples and their analysis, which can be delicate and ambiguous given the complexity of the organism employed in this investigation, could also be noted as a hurdle.

In order to validate the method and provide a sensitive tool capable of detecting transient interactions, a well-characterized drug/target system with moderate to high affinity will be considered as a proof-of-concept system:

**C1/TbAK** (*Tb* adenosine kinase): C1 is a derivative of CD12001, a compound that was shown to be active against the parasite with a  $K_i$  of 1  $\mu\text{M}$  (134-136).

**Methotrexate (Mtx)/TbDHFR-TS** (*Tb* dihydrofolate reductase-thymidylate synthase): Mtx inhibits DHFR with high affinity ( $K_i = 0.1 \text{ nM}$ ) (137).

Structural formula of BG-PEG<sub>4</sub>-C1 and BG-PEG<sub>4</sub>-Mtx structures are in Figure 12 below.

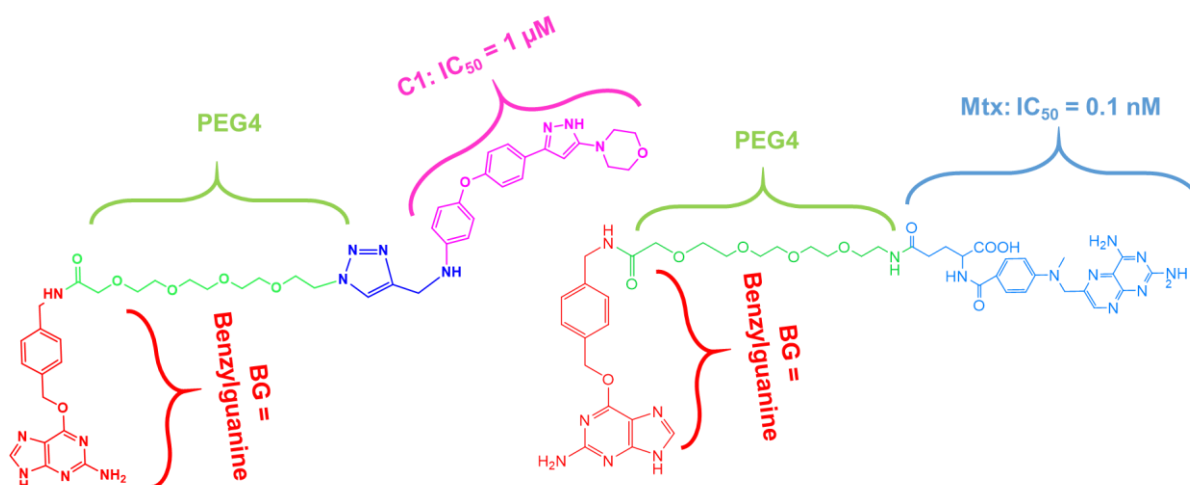


Figure 12: structures of the two BG-PEG4 derivatives: BG-PEG4-C1(left) and BG-PEG4-Mtx (right)

Another obstacle is the presence of heat shock proteins whose molecular weight may correspond to that of a target protein studied, such as DHFR-TS with a molecular weight of 56 kDa, close to that of the heat shock protein HSP60 (60 kDa) which is part of the chaperone system involved in the assembly and disassembly of different protein complexes, in the folding and unfolding as well as in the translocation of polypeptides (138).

Indeed, several heat shock proteins, also known as mitochondrial chaperones 70 and 40 (mtHsp70/mtHsp40), with molecular weights of 70 kDa and 40 kDa, respectively, were found to be not only abundant but also multifunctional and carry an important role in mitochondrial DNA replication and maintenance (139). Such chaperons may remain attached to the protein of interest and be revealed on the gel as a band confounding the results.

Many of the challenges presented in this chapter had to be addressed during this research to successfully achieve the ultimate goal of developing an efficient, sensitive, and selective *in vivo* target deconvolution tool.

## **.2 Materials and Methods**

### **.2.1 Materials**

SNAP-tag gene cloned into pUC57-Kan (GENEWIZ®), pLew100\_BirA\* plasmid, 3XHA-TurboID-NLS\_pcDNA3 plasmid, 3XHA-miniTurboID-NLS\_pcDNA3 plasmid, The strain was provided as a gift by Pascal Maser, Swiss Tropical and Public Health Institute (Basel, Switzerland), Q5® High-Fidelity 2X Master Mix (New England BioLabs<sub>Inc</sub>, Cat N° M0492S), T4 DNA Ligase (20'000 units, New England BioLabs<sub>Inc</sub>, Cat N° M0202S, NEB, USA), T4 buffer (10X, Cat N° M0202S, USA), FCS (Fetal Calf Serum; BioConcept; Cat N°: 2-01F10-I), FBS (Fetal Bovine Serum; Gibco; Cat N° 10270;), IMDM (1x) + GlutaMAX™ Supplement (Gibco; Cat N° 31980-022), 2-Mercaptoethanol (2ME)(Sigma Aldrich; Cat N° M6250), Penicillin/streptomycin (5000 Units/mL Pen, 5000 µg/mL Strep, Gibco, Cat N°: 15070-063 100 mL), G418 (Geneticin)(100 mg/mL, InvivoGen, Cat.# ant-gn-1), Hygromycin B Gold (100 mg/mL, InvivoGen, Cat.# ant-hg-1), PageRuler prestained protein ladder (ThermoFisher Scientific MA USA; Cat N°26616), TRIS-buffered saline (TBS, 10X), with 0.5% Tween 20 (Alfa Aesar by ThermoFisher Scientific, Cat N° J60448), IGEPAL® CA-630 (Sigma-Aldrich, Cat N°56741), BSA (Albumin Fraction V, AppliChem; A1391), Nonidet P-40 (NP-40, Fluka, Cat N° 74385) Red Anti-c-Myc Affinity Gel, (EZview™, Sigma-Aldrich, Cat N° E6654), western blot antibodies reference details below (section 2.2.23.2), PCR Thermocycler (Biometra Tone GmbH; serial N°:3729324; Germany), Centrifuge (Heraeus Megafuge 40, Thermo Scientific, Ref: 75004503), Microscope Slides (Superfrost Plus; Thermo Scientific; Ref J1800AMNZ; USA), Hoechst 33342 dye (Thermo Scientific), Mowiol 4-88 (Sigma Aldrich, Cat N°81381), ReadyProbes™ Tissue Autofluorescence Quenching kit (Invitrogen; Cat N°R37630), Sonicator (Bandelin Sonopuls; serial N° 519002 019), Filter ("rapid"-FILTERMAX 250-500, 0.2µm, Techno Plastic Products AG (TPP), Switzerland), Pierce™ BCA Protein Assay Kit (ThermoScientific, #23225 ), PageRuler™ prestained protein ladder 10 to 180 kDa (26616, ThermoFisher Scientific), PageRuler™ unstained protein ladder 10 to 200 kDa (26614, ThermoFisher Scientific), Ponceau S 5% (P3504, Sigma-Aldrich), Dynabeads™ M-280 Streptavidin (Invitrogen by Thermo Fisher Scientific, Cat No. 11205D), DynaMag™-2 (Invitrogen by ThermoFisher Scientific, Cat N°12321 D).

## .2.2 Methods

### .2.2.1 Generation of pLew100\_cMyc\_BirA\*\_GGG\_SNAP-tag plasmid:

The GGS\_SNAP-tag insert was synthesized by GENEWIZ®, the BamHI and XhoI restriction sites were incorporated into the gene of interest, and the gene was then cloned into the pUC57-kan plasmid. The SNAP-tag gene codons were optimized for expression in *Tbb* strains.

The insert was then digested with the restriction enzymes BamHI and XhoI and ligated into the expression vector pLew100\_cMyc\_BirA\* pre-digested with the same enzymes to obtain the final plasmid pLew100\_cMyc\_BirA\*\_GGG\_SNAP-tag encoding the first fusion protein cMyc\_BirA\*\_GGG\_SNAP-tag.

### .2.2.2 Generation of pLew100\_cMyc\_TBirA\*\_GGG\_SNAP-tag plasmid:

The cMyc\_TBirA\* gene encoding the cMyc\_TurboID system was amplified by PCR using as a template the plasmid 3XHA-TurboID-NLS\_pcDNA3 and incorporating through the fly-ends of the used primers (Primers oSS0025 TurboID FW and oSS0026 TurboID RV, present in the table below) the following restriction sites HindIII, XhoI, as well as the cMyc gene. The resulting insert (Figure 13 below) was then ligated into the expression plasmid pLew100\_cMyc\_BirA\*\_GGG\_SNAP-tag pre-digested with HindIII and XhoI (allowing removal of the BirA\* gene) to obtain the final plasmid pLew100\_cMyc\_TBirA\*\_GGG\_SNAP-tag encoding the second fusion protein cMyc\_TBirA\*\_GGG\_SNAP-tag. For details of the primers used to design the insert of interest, see Table 3 below.



*Figure 13: Design and production of the insert of interest HindIII\_cMyc\_TBirA\*\_XhoI*

**Table 1: PCR parameters used for the production of inserts of interest**

| Steps | T°C                                                                          | Time (h:min:s)            | Number of cycles  |
|-------|------------------------------------------------------------------------------|---------------------------|-------------------|
| 1     | 95                                                                           | 0:05:00                   |                   |
| 2     | 95                                                                           | 0:00:30                   | <b>Cycles 40x</b> |
| 3     | 55-77 (Annealing temperature: T <sub>m</sub> ) depending on the primers used | 0:00:30                   | <b>Cycles 40x</b> |
| 4     | 72 (Elongation temperature)                                                  | *0:01:00 (Extension time) | <b>Cycles 40x</b> |
| 5     | 72 (Final extension)                                                         | 0:05:00                   | <b>Cycles 40x</b> |
| 6     | 16                                                                           | ∞                         |                   |

\* To be adapted according to the DNA polymerase used. For Q5 DNA polymerase, the extension time is generally between 20 and 30 seconds per kb for complex or genomic samples and 10 seconds per kb for simple templates such as plasmids or complex templates of size < 1 kb.

### .2.2.3 Generation of pLew100\_cMyc\_mTBirA\*\_GGG\_SNAP-tag plasmid:

For the cMyc\_mTBirA\* gene encoding the cMyc\_miniTurbo system, the same method as mentioned above for the synthesis of the cMyc\_TBirA\* gene was used but this time using the 3XHA-miniTurboID-NLS\_pcDNA3 plasmid as a template and incorporating through the flying ends of the used primers (primers oSS0030 miniTurboID FW and oSS0031 TurboID RV, details in Table 4 below) the following restriction sites HindIII, XhoI, and the cMyc gene. The resulting insert (Figure 14 below) was then ligated into the expression plasmid pLew100\_cMyc\_BirA\*\_GGG\_SNAP-tag predigested with HindIII and XhoI to obtain the final plasmid pLew100\_cMyc\_mTBirA\*\_GGG\_SNAP-tag encoding the third fusion protein of interest cMyc\_mTBirA\*\_GGG\_SNAP-tag.



**Figure 14: Design and production of the insert of interest HindIII\_cMyc\_mTBirA\*\_XhoI**

### .2.2.4 Generation of pLew100\_cMyc\_GGGG\_BirA\*\_SNAP-tag plasmid:

The plasmid pLew100\_cMyc\_BirA\*\_GGG\_SNAP-tag (8373 bp) was used as a template to perform site-directed mutagenesis to insert an additional glycine into the GGS linker. To avoid any subsequent translation problems, a different codon coding for glycine was chosen. The

GGT codon was inserted before the first GGA codon to obtain the final **GGGS linker** (G: Glycine and S: Serine). The PCR parameters used, and the designated primers required to perform site-directed mutagenesis are described in the table below.

Q5 Master mix using Q5 High-Fidelity DNA Polymerase was used to perform the PCR.

**Table 2: PCR parameters used for site-directed mutagenesis**

| Steps | T°C                                                                | Time (h:min:s)            | Number of cycles  |
|-------|--------------------------------------------------------------------|---------------------------|-------------------|
| 1     | 95                                                                 | 0:01:00                   |                   |
| 2     | 95                                                                 | 0:01:00                   | <b>Cycles 18x</b> |
| 3     | <b>Tm:73.1 (GGGS), 74.8 (GGGGS)</b><br>(Tm: Annealing temperature) | 0:01:10                   | <b>Cycles 18x</b> |
| 4     | 72 (Elongation temperature)                                        | *0:06:00 (Extension time) | <b>Cycles 18x</b> |
| 5     | 72 (Final extension)                                               | 0:08:00                   | <b>Cycles 18x</b> |
| 6     | 4                                                                  | ∞                         |                   |

\* To be adapted according to the DNA polymerase used. For Q5 DNA polymerase, the extension time is generally between 20 and 30 seconds per kb for complex or genomic samples and 10 seconds per kb for simple templates such as plasmids or complex templates of size < 1 kb.

Immediately after PCR, the amplification reaction was chilled on ice for 2 min before adding 1µl of **DpnI** (20 U/µl, New England Biolabs) and 1µl of CutSmart 10X buffer (New England Biolabs) and incubating for 1 h at 37°C. DpnI will cleave the methylated DNA to separate the newly generated plasmid from the parental one.

After the incubation period, a transformation step with the reaction tubes was performed in competent NovaBlue *E. coli* bacteria (see section 2.2.10 below for details).

#### **.2.2.5 Generation of pLew100\_cMyc\_GGGGS\_BirA\*\_SNAP-tag plasmid:**

The method is the same as described in section 2.2.4 above. The only differences are the parental plasmid that was used as a template to perform the site-directed mutagenesis, which was the one generated by the first site-directed mutagenesis: pLew100\_cMyc\_GGGGS\_BirA\*\_SNAP-tag plasmid (8376 bp), the primers used and their hybridization temperature (see Table 3 below). The **Tm** was set at 74.8°C. The idea was to add

a glycine with a different codon (**GGC**) from the one next to it to obtain the **GGGGS** linker composed of four glycines and one serine.

**Table 3: Table of primers used for site-directed mutagenesis**

| Primer names                                                                | Sequences                                                           | Tm (°C) | Generated plasmids                       |
|-----------------------------------------------------------------------------|---------------------------------------------------------------------|---------|------------------------------------------|
| <b>oSS0035 fw</b>                                                           | 5'-GAA GCG CAG AGA AGC TCG AGG GTG GAG GTT CCA TGG ACA AGG ATT G-3' | 73.8    | <b>pLew100_cMyc_GGGS_BirA*_SNAP-tag</b>  |
| <b>oSS0036 rv</b>                                                           | 5'-CAA TCC TTG TCC ATG GAA CCT CCA CCC TCG AGC TTC TCT GCG CTT C-3' | 73.8    |                                          |
| <b>oSS0037 fw</b>                                                           | 5'-GAA GCG CAG AGA AGC TCG AGG GCG GTG GAG GTT CCA TGG ACA AG-3'    | 75.1    | <b>pLew100_cMyc_GGGGS_BirA*_SNAP-tag</b> |
| <b>oSS0038 rv</b>                                                           | 5'-CTT GTC CAT GGA ACC TCC ACC GCC CTC GAG CTT CTC TGC GCT TC-3'    | 75.1    |                                          |
| <u>Codons of the added amino acid Glycine (G); fw: forward; rv: reverse</u> |                                                                     |         |                                          |

For more details regarding the design of the GGGS and GGGGS linkers, see Appendix 4. The digestion and ligation steps required to generate all plasmids of interest are described in the sections below (2.2.8 and 2.2.9 respectively).

#### *.2.2.6 Generation of pLew100\_cMyc\_BirA\*\_PAPAP\_SNAP-tag plasmid:*

The XhoI\_ **PAPAP** \_SNAP\_BamHI gene (Figure 15 below) was generated by PCR amplification of the SNAP-tag portion of the cMyc\_BirA\_**GGG**\_SNAP-tag plasmid, using newly designed primers (details in Table 4 hereunder) allowing incorporation of the XhoI and BamHI restriction sites, and the **PAPAP** linker. The PCR parameters used were the same as described in Table 1 in section 2.2.2 except for the elongation time that was set to 2 min at 72 °C (instead of 1 min). The resulting insert (Figure 15 below) was then, ligated into the pre-digested expression plasmid pLew100\_cMyc\_BirA\*\_GGG\_SNAP-tag between XhoI and BamHI (where GGS\_SNAP was removed beforehand) to obtain the plasmid pLew100\_cMyc\_BirA\*\_**PAPAP**\_SNAP-tag encoding for the fusion protein cMyc\_BirA\*\_**PAPAP**\_SNAP-tag (details of the different steps in Appendix 5).

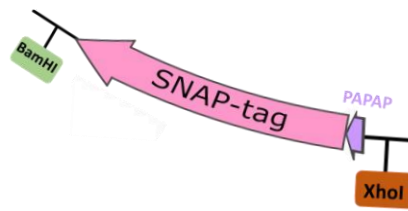


*Figure 15: Design and production of the insert of interest XhoI\_PAPAP\_SNAP-tag\_BamHI*

#### *.2.2.7 Generation of pLew100\_cMyc\_BirA\*\_PTP\_SNAP-tag plasmid:*

The XhoI\_PTP\_SNAP\_BamHI gene (Figure 16 below) was obtained by PCR amplification of the SNAP-tag portion of the original pLew100\_cMyc\_BirA\*\_GGSNAP-tag using the following primers oSS0040 SNAP\_BamHI rv and oSS0043 XhoI\_PTP\_SNAP fw (see Table 4 below for primer details), that allow the insertion of XhoI, BamHI restriction sites, and the PTP linker. All the following steps required to obtain the plasmid of interest encoding the cMyc\_BirA\*\_PTP\_SNAP-tag fusion protein are the same as those described in section 2.2.6 above (details of the different steps in Appendix 6).

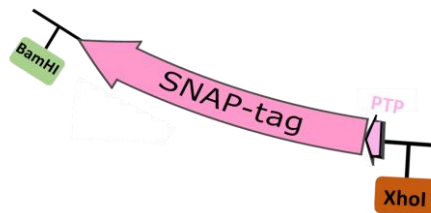


*Figure 16: Design and production of the insert of interest XhoI\_PTP\_SNAP-tag\_BamHI*

Table 4: Table of primers used for site-directed mutagenesis

| Primer's name                                           | Sequence                                                                                                      | Tm (°C) | Generated plasmid                                                                      |
|---------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|---------|----------------------------------------------------------------------------------------|
| oSS0025<br>TurboID FW                                   | 5'-<br>CCC <b>AAGCTT</b> CAT <b>ATGGAACAAAA</b><br><b>ACTCATCTCAGAAGAGGATCTC</b> aa<br>agacaatactgtgcctctg-3' | 74.0    | pLew100_cMyc_ <b>T</b> BirA*_GGG_SNAP-tag                                              |
| oSS0026<br>TurboID RV                                   | 5'-<br>CCG <b>CTCGAG</b> cttttcggcagaccgcaga<br>ctg-3'                                                        | 71.5    | pLew100_cMyc_ <b>T</b> BirA*_GGG_SNAP-tag                                              |
| oSS0030<br>miniTurboID<br>FW                            | 5'-<br>CCCAAGCTTCAT <b>ATGGAACAAAA</b><br><b>ACTCATCTCAGAAGAGGATCTC</b> at<br>cccgtgctgaacgctaaacag-3'        | 76.4    | pLew100_cMyc_ <b>mT</b> BirA*_GGG_SNAP-tag                                             |
| oSS0031<br>miniTurboID<br>RV                            | 5'-<br>CCG <b>CTCGAG</b> cttttcggcagaccgcaga<br>ctg -3'                                                       | 71.5    | pLew100_cMyc_ <b>mT</b> BirA*_GGG_SNAP-tag                                             |
| oSS0039<br>XhoI_PAPAP_<br>SNAP fw                       | 5'-<br>CCG <b>CTCGAG</b> <b>CCCGCTCCTGCACCC</b><br>atggacaaggattgcgagatgaag-3'                                | 77.0    | pLew100_cMyc_BirA*_ <b>PAPAP</b> _SNAP-tag                                             |
| oSS0040<br>SNAP_BamHI<br>rv                             | 5'-<br>CGC <b>GGATCC</b> tcacaattgacccaaaccg<br>ggcttgcccaagc-3'                                              | 74.8    | pLew100_cMyc_BirA*_ <b>PAPAP</b> _SNAP-tag<br>pLew100_cMyc_BirA*_ <b>PTP</b> _SNAP-tag |
| oSS0043<br>XhoI_PTP_<br>SNAP fw                         | 5'-<br>CCCCCTCGAG <b>CCCACACCG</b> ATGG<br>ACAAGGATTGCGAGATGAAG-3'                                            | 73.3    | pLew100_cMyc_BirA*_ <b>PTP</b> _SNAP-tag                                               |
| <b>XhoI</b> , <b>BamHI</b> , <b>HindIII</b> , cMyc gene |                                                                                                               |         |                                                                                        |

### *.2.2.8 Digestion:*

The digestion step was performed using the same restriction enzymes to digest the vector and insert of interest to create sticky ends.

For the generation of plasmid pLew100\_cMyc\_BirA\*\_**GGG**\_SNAP-tag, the restriction enzymes used to digest the pLew100\_cMyc\_BirA\* vector and the **GGG**\_SNAP-tag insert (incorporated into plasmid pUC57-kan) were **BamHI** (10 U/μL, Promega, USA) and **XhoI** (20 U/μL, New England Biolabs® (NEB), USA). CutSmart™ buffer (10X, NEB) was chosen because both restriction enzymes exhibited 100% activity in it.

For the production of the other plasmids of interest (pLew100\_cMyc\_BirA\*\_**PTP**\_SNAP-tag and pLew100\_cMyc\_BirA\*\_**PAPAP**\_SNAP-tag), the same digestion protocol as described above to produce the plasmid pLew100\_cMyc\_BirA\*\_**GGG**\_SNAP-tag was used.

For the generation of plasmids pLew100\_cMyc\_TBirA\*\_**GGG**\_SNAP-tag and pLew100\_cMyc\_mTBirA\*\_**GGG**\_SNAP-tag, **HindIII** (20U/μL, NEB, USA) and **XhoI** were employed to digest the pLew100\_cMyc\_BirA\*\_**GGG**\_SNAP-tag vector (to remove the BirA\* moiety) and the cMyc\_TBirA\* or cMyc\_mTBirA\* inserts required for the conception of the two plasmids of interest. NEBuffer™ r3.1 buffer (10X, NEB) was chosen for simultaneous digestion using the two previously mentioned enzymes.

The details of the digestion protocol are summarized in Table 5 below.

**Table 5: Protocol for the digestion of vectors and inserts of interest**

| <i>Components</i>           | <i>Quantities (volume or mass)</i>                           |
|-----------------------------|--------------------------------------------------------------|
| <b>Restriction enzyme 1</b> | 1 μL (of either 10 U/μL or 20 U/ μL depending on the enzyme) |
| <b>Restriction enzyme 2</b> | 1 μL (of either 10 U/μL or 20 U/ μL depending on the enzyme) |
| <b>Plasmid DNA</b>          | 1 μg                                                         |
| <b>NEBuffer 10X</b>         | 2 μL                                                         |
| <b>Total volume</b>         | 20 μL                                                        |

The digestion was performed at 37°C for 1h.

The plasmid vector was then dephosphorylated for 10 minutes at 37°C using a thermosensitive alkaline phosphatase (FastAP™1U/L, ThermoScientific, Cat N°EF0651), and then inactivated for 15 minutes at 65°C or 20 minutes in the event of more heat-resistant restriction enzymes. After the inactivation step, 1% agarose gel electrophoresis was performed for 30 min at 110V to separate the vector from the digested fragment, and finally, the vector and insert of interest were purified using the NucleoSpin® Gel and PCR Clean-up Kit (Macherey-Nagel, Düren, D) according to the following procedure:

The excised gel strip is placed in an Eppendorf with 200 µL of NTI binding buffer per 100 mg of gel, and the mixture is incubated at 50°C for 5–10 minutes to thoroughly dissolve the gel slice.

If the genetic material is not present in the agarose gel, this step can be skipped. Instead, 2 volumes of NTI buffer are added, and the sample is then placed directly onto the NucleoSpin® gel and PCR Clean-up column.

Then, the sample (max 700 µL) is loaded into the column and centrifuged at 11'000 x *g* for 30 seconds. Next, the silica membrane of the column is washed by adding 700 µL of NT3 wash buffer. This last step is repeated once more to obtain a pure product.

The final drops of NT3 buffer are collected in the collecting tube and thrown away after being entirely removed from the silica column membrane by centrifugation at 11000 x *g* for 1 minute. Finally, the column is placed in a new 1.5 mL microcentrifuge tube and 15-30 µL of NE elution buffer is added to elute the genetic material. The addition is followed by a one-minute incubation period at room temperature (RT) and then centrifugation at 11,000 x *g* for 1 minute to collect the DNA, the concentration of which will be measured using the NanoDrop OneC (ThermoScientific).

#### *.2.2.9 Ligation:*

Ligation was performed on the purified vector and insert of interest by adding T4 DNA ligase (400 U/µL) and the corresponding T4 buffer (10X: 500 mM Tris-HCl, 100 mM MgCl<sub>2</sub>, 10 mM ATP, pH 7.5 at 25°C). Differential ratios were prepared between the digested vector and the insert by varying the amount of the latter relative to the plasmid as follows: 1:3, 1:5, 1:7, and 1:9 (details regarding the ligation reactions can be found in Appendix 9). Then, all ratios were

incubated for 1h at RT. Following that, the competent *E. coli* TOP10 bacteria were used to undergo bacterial transformation using different ratios of ligation reactions.

#### *.2.2.10 E. coli Transformation by the heat shock method:*

*E. coli* Top10 competent cell strains were used for the cloning of the majority of plasmids except for those obtained by site-directed mutagenesis where competent *E. coli* NovaBlue strain was chosen. The Heat-Shock approach serves as the foundation for the bacterial transformational process, and the following methodology was applied: Competent cells (Top10 or NovaBlue) were incubated for five minutes with either ligation reactions or a purified plasmid of interest in an ice-filled incubation container. Competent bacteria (20–25 mL) were mixed with either 1–5 mL of purified plasmid at ~1 ng/μl or with a ligation reaction (of various ratios), and the mixture was then placed on ice for 20–30 minutes. After being subjected to a heat shock for 30 seconds at 42°C, the cells were immediately put on ice for 2 min 30 s. After cooling, the cells were supplemented with 200 μl of LB or 2xTY medium, incubated for 1-2 h at 37°C with gentle agitation, placed on an LB or 2xTY plate with ampicillin (100 μg/mL), and then kept at 37°C overnight.

#### *.2.2.11 Plasmid purification*

According to the instructions provided by the NucleoSpin® Plasmid (NoLid) kit (Macherey-Nagel), the generated plasmids of interest were purified as follows: All centrifugation steps were performed at **11'000 x g**.

First, 1 to 5 ml of saturated *E. coli* bacterial culture was centrifuged for 30 seconds. The supernatant was discarded, and the pellet was lysed by first resuspending it in 250 μL of resuspension buffer A1 supplemented with Rnase A, then the mixture was vortexed and 250 μL of lysis buffer A2 was added and mixed gently by inverting the Eppendorf tube containing the lysate, 6 to 8 times. Here, vortexing must be avoided to prevent the shearing of genetic material. The lysate was then incubated for 5 minutes at RT, then 300 μL of neutralization buffer A3 was added and, as before, the lysate was mixed gently 6-8 times by inverting the tube and without vortexing. The lysate was clarified by centrifuging the tube for 5-10 minutes before being loaded into the NucleoSpin® Plasmid (NoLid) column and centrifuged again for 1 minute. The silica membrane on the column was then washed with 500 μL of AW wash buffer

and centrifuged for 1 minute followed by the addition of 600 µL of A4 buffer and centrifugation for 1 minute. The silica membrane was then dried by centrifugation for 2 min and finally, the plasmid was eluted from the column by adding 50 µL of elution buffer AE, in the column previously placed in a 1.5 mL microcentrifuge tube, then an incubation step for 1 min at RT followed by centrifugation for 1 min was performed to retrieve the purified plasmid in the tube.

#### *.2.2.12 Cultivation of *T. brucei* bloodstream form:*

*Tbb* bloodstream-form New York Single Maker (*Tbb* BSF NYSM) cell line (140), a derivative of Lister 427-2 (221 cell line; MiTat 1.2) which co-expresses a tetracycline repressor TetR and a T7 RNA polymerase was used to generate cell lines expressing the different tools. The expression of these fusion proteins is regulated by the addition of tetracycline to the growth medium.

The *Tbb* BSF NYSM strain was chosen as the wild-type strain for the project. This parasite cell line was selected for its lack of human virulence, which makes it safer and more suitable for human handling, as well as its capacity to co-express TetR and the T7 RNA polymerase upon tetracycline addition thereby controlling the expression of genes of interest (140-142).

Transfected parasites were cultivated in Hirumi's modified Iscove's medium-9T (HMI-9) medium supplemented with 10% heat-inactivated fetal calf serum (iFCS) and 2.5 µg /mL hygromycin, at 37°C with 5% CO<sub>2</sub>. To maintain the production of TetR and T7 RNA polymerase, 1 µg/mL Geneticin® (G418) was added to the media needed to grow the different generated strains.

#### *.2.2.13 Parasites transfection and stable cell lines:*

By electroporating the parasite with the desired plasmid using the X-001 program on the AMAXA Nucleofector™II device (Lonza, Switzerland), it was possible to generate stable cell lines with the correct genes inserted into their genomes (Lonza, Switzerland). A homemade transfection buffer (*Tb*-BSF) was used to transfect the parasites (90 mM Sodium phosphate, 5 mM KCl, 0.15 mM CaCl<sub>2</sub>, 50 mM HEPES, pH 7.3). The transfection step was performed as followed:

1-2x10<sup>7</sup> cells in log phase were centrifuged at 1300 x g (2500 rpm), 10 min, 4°C, or RT, and the pellet was resuspended in 100 µL *Tb*-BSF containing 10 µg of DNA plasmid of interest that had previously been linearized with NotI 10 U/µL (NEB, USA), dephosphorylated with FastAP™ 1U/µL and purified using the NucleoSpin® Gel and PCR Clean-up Kit according to the supplier's protocol described above. The mixture was then electroporated using a sterile cuvette (Gene Pulser®, Bio-Rad, Cat N° 165-2086). The details of the digestion and dephosphorylation steps are described in Table 6 hereafter.

Following the transfection procedure, the cells were transferred right away to 20 mL of pre-warmed HMI-9 complete medium, and limited dilutions (1:5, 1:15, 1:50, 1:100, 1:300, and 1:500) were carried out before being distributed across 48-well plates and cultured at 37°C with 5% CO<sub>2</sub>.

After 24h, the hygromycin antibiotic 2.5 µg/mL was added to select clones that had integrated the gene of interest containing the hygromycin resistance. Clonal transformants began to emerge one to three weeks after transfection, and these were transferred to flasks T25 containing the medium supplemented with hygromycin 2.5 g/mL for further experiments.

**Table 6: Protocol used for simultaneous linearization and dephosphorylation of plasmids of interest**

| <i>Components</i>                            | <i>Volume (µL)</i> |
|----------------------------------------------|--------------------|
| Plasmid DNA                                  | <b>*1µg</b>        |
| 10X restriction enzyme buffer                | <b>2 µL</b>        |
| Restriction enzyme                           | <b>1 µL</b>        |
| FastAP™ Thermosensitive Alkaline Phosphatase | <b>1 µL</b>        |
| Water (nuclease-free)                        | <b>To 20 µL</b>    |
| Total volume                                 | <b>**20 µL</b>     |

\*Knowing that one unit is the amount of enzyme needed to dephosphorylate the 5' termini of 1 µg of linearised pUC57 DNA in 10 min at 37 °C in FastAP™ buffer so depending on the amount of plasmid to be linearised, the necessary volumes of FastAP to be added will have to be adapted accordingly.

\*\* Can be adapted

Digestion was carried out for 1 h at 37°C followed by dephosphorylation at 37°C for 10 min. The reaction was then stopped by heating at 65°C for 15 min or at 80°C for 20 min in case the restriction enzymes were not inactivated at 65°C.

#### *.2.2.14 gDNA extraction*

1-5x10<sup>7</sup> cells in log phase were centrifuged at 1300 x g (2500 rpm), 10 min, 4°C, or RT, and after thoroughly removing the supernatant, the pellet was then resuspended in 80 µL of T1 lysis buffer from NucleoSpin® Tissue XS kit (MACHEREY-NAGEL) and all subsequent steps were performed using the manual for extraction of genomic DNA from tissues and cells (NucleoSpin® Tissue XS kit; MACHEREY-NAGEL) and following the dedicated protocol for the extraction of gDNA from cells in culture.

Briefly, after resuspending the pellet in buffer T1, 8 µL of Proteinase K was added and the cells were vortexed twice for 5 seconds before being incubated at 56°C for 10 min. Next, 80 µL of lysis buffer B3 is added and an additional incubation step at 70°C for 5 min was carried out. Following this, 80 L of 96–100% pure ethanol was added, and the lysate was vortexed twice for 5 s. The lysate was then loaded onto the NucleoSpin Tissue XS column and centrifuged at 11,000 x g for 1 min. The flow-through was collected in the collection tube and then, discarded, and 50 µL of wash buffer B5 was put directly into the column before the centrifugation step at 11,000 x g for 1-2 min. This last washing step was repeated once more, afterwards, the collection tube containing the flow-through was discarded and the column was placed in a 1.5 mL microcentrifuge tube (Eppendorf tube) before being eluted with 20 µL of BE elution buffer and centrifuged at 11'000 x g for 1 min. The eluate containing pure gDNA was then collected in the Eppendorf tube and the concentration of the genetic material was measured using the NanoDrop One<sup>c</sup> (ThermoScientific).

#### *.2.2.15 Validation of the correct integration of the gene of interest*

Correct integration of the gene of interest into the parasite genome by homologous recombination was confirmed by PCR using forward and reverse primers to amplify genomic DNA previously extracted from the cells.

The primers that were used to verify the successful integration of the various generated genes of interest are the following: GCTCCAAACGCGTAGATACATG (**oSS0022** reverse primer targeting a part of Chromosome 1 ID X52586.1) and CCAGAGGAATCGACAAGCAGG (**oSS0023** forward primer targeting a part of BirA\*(R118G) gene). A band around **1950 bp** is produced by these primers during PCR to validate the proper integration of the genes that produce the

corresponding proteins: cMyc\_BirA\*\_GGG\_SNAP-tag, cMyc\_BirA\*\_PAPAP\_SNAP-tag, and cMyc\_BirA\*\_PTP\_SNAP-tag.

The subsequent primers GGATGGGCGGAGAAATCAGTC (**oSS0033** forward primer targeting the same part in TBirA\* and mTBirA\* genes) and GGATGGGCGGAGAAATCAGTC (**oSS0034** reverse primer targeting a distinct region of Chromosome 1 ID X52586.1 from that covered by the **oSS0022** primer) were used for further confirmation of the correct integration of the genes encoding for the following proteins:

cMyc\_TBirA\*\_GGG\_SNAP-tag, cMyc\_mTBirA\*\_GGG\_SNAP-tag, cMyc\_BirA\*\_GGGS\_SNAP-tag, and cMyc\_BirA\*\_GGGGG\_SNAP-tag. These primers result in a PCR band of approximately **1550 bp**.

Both primer sets can be used to verify the successful integration of the different genes of interest. The decision will be made in light of the PCR's proper functioning (results in section 3.2 below).

#### *.2.2.16 Microscopic studies*

A total of  $\sim 10^7$  cells (parasites that have been incubated (sample of interest) or not (negative control) in the presence of 20  $\mu$ M BG-PEG<sub>4</sub>-FDA (Fluorescein diacetate)) were centrifuged (900-1300 x g, 4°C, 10-5 min) and then washed three times with 0.5M Tris, pH 8.0. 10  $\mu$ L of the suspension was placed directly onto a positively charged glass microscope slide, on which a well was drawn. Then, 10  $\mu$ L of 4% paraformaldehyde (PFA) was added directly to the drop to fix the cells on the slide (20 min fixation). Next, the glass slide was washed twice with 0.5M Tris, pH 8.0, and 10  $\mu$ L of Hoechst 33342 dye diluted beforehand (1:5000) in 0.5M Tris, pH 8.0 was added to the fixed parasites and a 15-minute rest period in darkness was performed. After staining, the dye was removed and the slide was washed twice with 0.5M Tris, pH 8.0. The last step involved adding 10  $\mu$ L of Mowiol 4-88 (a water-soluble hydrocolloid mucoadhesive) to the slide and waiting 48 hours in the dark to allow the coverslip to set.

If auto-fluorescence is observed, the ReadyProbes™ Tissue Autofluorescence Quenching kit can be used to reduce this phenomenon, which is frequently noticed in cells due to the presence of intrinsic biomolecules that emit fluorescence in the UV-visible spectral range (143).

The slides were visualized using Widefield Fluorescence Microscopy.

#### *.2.2.17 Medium preparation:*

*Tbb* BSF NYSM WT (wild type) were cultivated in Hirumi's modified Iscove's medium-9 (HMI-9)(144) containing 88% of Iscove's Modified Dulbecco's Medium (IMDM) supplemented with GlutaMAX (dipeptide, L-alanyl-L-glutamine), an improved version of L-glutamine that is more stable in aqueous solutions and does not freely breakdown (145, 146)), 10% heat-inactivated serum (iFCS, heat-inactivation for 30 min at 56°C), 1% Hirumi II/ST-9 stock (see Table 7 below for composition), and 1% 2-Mercaptoethanol (2ME) taken from a 20 mM stock solution that had to be freshly prepared as follows: 14 µL 2-ME in 10 mL in ddH<sub>2</sub>O. Before use, the HMI-9 medium was sterilized by filtration ("rapid"-FILTERMAX 250-500, 0.2µm, Techno Plastic Products AG (TPP), Switzerland) under aseptic conditions immediately after preparation, followed by a detoxification step of the 2-ME either at 4°C for 24 hours or 37°C for 2 hours.

After the detoxication period, 5 mL of Penicillin/streptomycin was added to prevent contamination. In order to select only *Tbb* NYSM WT strain, Geneticin® (G418) at a concentration of 1 µg/ml was introduced to the growth medium. Hygromycin 2.5 µg/ml was used only for the selection of transfected cell lines.

Details on the product references are listed in section 2.1 above.

**Table 7: Composition of Hirumi II/ST-9 stock (100X final concentration)**

| <i>Compound</i>              | <i>Molecular Weight<br/>(MW) (g/mol)</i> | <i>Concentration<br/>(mM)</i> | <i>Concentration<br/>(mg/mL)</i> |
|------------------------------|------------------------------------------|-------------------------------|----------------------------------|
| NaPyruvat                    | 110.05                                   | 100                           | 11.0                             |
| Hypoxanthine                 | 136.1                                    | 100                           | 13.6                             |
| Bathocuproindisulfat         | 564.5                                    | 5                             | 2.82                             |
| L-Cysteine or L-Cysteine HCl | 121.16 or 157.6                          | 150                           | 18.2 or 23.6                     |
| Thymidine                    | 242.2                                    | 16                            | 3.9                              |

After preparing the Hirumi stock solution, 5 mL aliquots were stored at -20°C. All compounds from the table above were purchased from Sigma-Aldrich.

#### *.2.2.18 Parasite counting:*

A Neubauer counting chamber (Neubauer hemocytometer 0.1mm depth, 0.0025 mm<sup>2</sup>, Marienfeld, Germany) was used to count the trypanosomes. The procedure entails depositing a volume of 10 µL of the parasite-containing culture medium in each of the two counting chambers.

A counting chamber is divided into four grids, each of which is further divided into sixteen smaller squares (see Figure 17 below). It is necessary to count each parasite in the 16-square grid (the parasites circled in blue in the figure below). The cells in the bottom and the left row of each grid should also be counted (underlined in red in the figure hereafter). However, the cells on the top and right-hand rows should not be considered.

The result of counting corresponds to **Y x 10<sup>4</sup> cells/ml** where **Y** indicates the mean value of the total number of parasites that were counted in the 8 grid squares of the two chambers, then divided by 8. If the preloaded sample has been diluted, the dilution factor f (10 or 100 maximum) must be taken into account and the formula becomes the following: **Y x 10<sup>4</sup> x f**

Example: In Figure 17 hereunder, there are 18 parasites counted in the 4 square grids, considering that 18 parasites were also counted in the second chamber and that the loaded

samples had been diluted 10 times, the concentration of parasites would then be:  $\frac{1}{5} \times 10^4 \times 10$   
 $= 4.5 \times 10^5$  cells/ml.

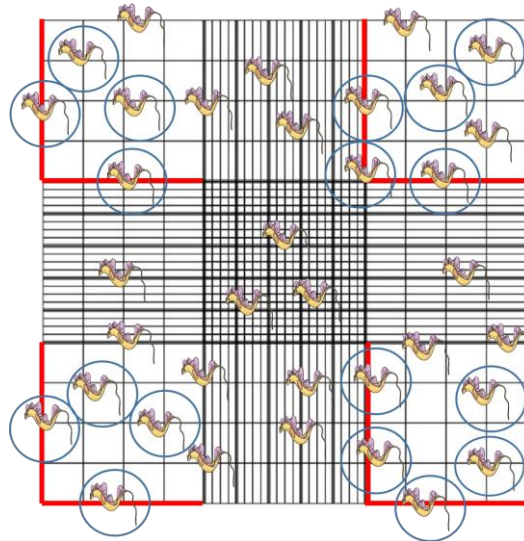


**Figure 17:** Parasites counting using Neubauer hemocytometer (The grid taken from (147) and slightly modified by me): Only parasites circled in blue are taken into account

### .2.2.19 Cultivation of trypanosomes

Trypanosomes have exponential growth and a logarithmic phase ranging from  $1-4 \times 10^6$  cells/mL. They were systematically diluted to  $1 \times 10^4$  cells/mL for a 48 h passage or to  $1 \times 10^3$  cells/mL for a 72 h passage to avoid overgrowth. They were incubated at 37°C, 5% CO<sub>2</sub>. The typical passage entails introducing 100 µl of the well holding the parasites in the log phase (e.g., well C1 as illustrated in Figure 18 below) to a new well containing 900 µL of warm HMI-9 medium and then making 1:10 dilutions from this well throughout the row.

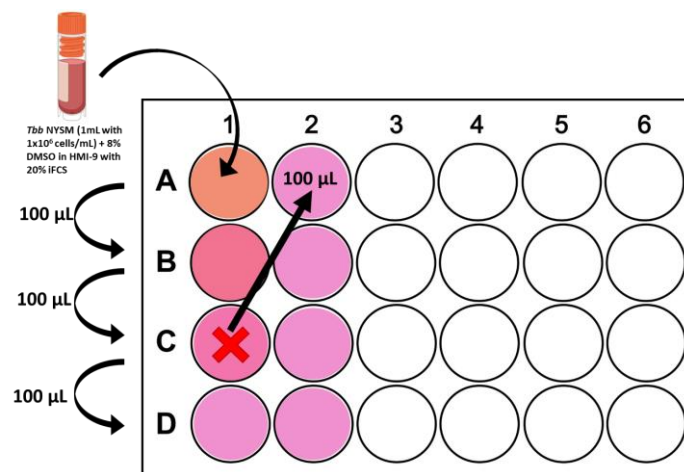


**Figure 18:** Routine Trypanosome passages

#### *.2.2.20 Freezing*

The freezing step for prolonged storage consists of preparing cryotubes containing a dense culture of parasites (towards the end of the log phase  $1-5 \times 10^6$  cells/mL). The protocol used was as follows:  $1-5 \times 10^6$  cells/mL of parasite cultures were centrifuged at  $1300 \times g$  for 10 min either at room temperature (RT) or 4 °C. The supernatant was then removed, and the pellet was resuspended in a new stabilate medium consisting of HMI-9 with an increased proportion of serum (20 % iFCS rather than 10 %) and 8% DMSO (10% glycerine could be used). Next, 1 mL was transferred to a sterile 1.6 mL CryoPure tube (SARSTEDT AG & Co, Germany). The cryotubes were then placed directly into a cell freezing container (Corning® CoolCell™ LX, Germany) which, when placed in a -80°C freezer, will allow freezing at the -1°C/min rate ideal for cell cryopreservation. After 24 h at -80°C, the cryotubes were shifted to liquid nitrogen for long-term storage.

#### *.2.2.21 Thawing*

The frozen trypanosomes were placed on ice and thawed in a water bath at 37 °C as rapidly as feasible.

After transferring the contents of the cryotube to the first well of a 24-well plate (TPP® 1.86 cm<sup>2</sup> tissue culture plates, sterile), which was filled with HMI-9 medium pre-warmed at 37°C, a 1:10 dilution was carried out as previously mentioned in section 2.2.19.

#### *.2.2.22 Growth curves*

The following growth curves were developed to determine the day that corresponded to the log phase of optimal growth: First, on day 1, a well containing  $10^6$  cells/mL was selected and diluted 10X into a new well in the first row (e.g., well A1 Figure 18 above) of a 24-well plate. From this well, the row was subjected to a 10X dilution until  $10^2$  cells/mL were obtained (e.g., in A4). The wells containing parasites in the different dilutions ( $10^2$  cells/mL to  $10^5$  cells/mL) were counted on the following days (day 2 to day 4) using the Neubauer counting chamber and the concentrations (numbers of parasites/well) were recorded to establish the growth curves.

The method for determining the growth rate of each strain was performed as follows: On day 1, a well containing  $10^5$  cells/mL was prepared, incubated for 24 hours, and then counted the following day (day 2). After the Day 2 count, the well was diluted to  $10^5$  cells/mL in a new well

of the same column (e.g., B1), incubated again for 24 hours, and counted the following day. These steps were repeated until day 4 (e.g., well D1). This method allows the determination of the growth factor which is the number of parasites increasing in 24 hours and the calculation of the growth rate.

For each strain, the same methods were applied to obtain different growth curves and rates. The formulas used to calculate the growth factor (GF), growth rate (GR), and population doubling time (PDT) are described below:

$$GF = \frac{N_t}{N_0}$$

$$GR = \frac{\log_2 \left( \frac{N_t}{N_0} \right)}{t}$$

$$PDT = \frac{1}{GR}$$

**N<sub>t</sub>**: Number of cells at t = X hours

**N<sub>0</sub>**: Number of cells at time = 0 hours.

All values obtained from the parasite counting were used to generate growth curves using GraphPad Prism (version 8.4.3).

### **.2.2.23 Western blot**

#### **.2.2.23.1 Preparation of samples and polyacrylamide gel electrophoresis (PAGE):**

10<sup>7</sup> cells in total were harvested by centrifugation (600-1000 X g, 10 min, 4°C) and the pellet was resuspended in PBS supplement with PMSF 1mM and centrifuged again before another last wash step in PBS only. After the centrifugation process, the resulting pellet was re-suspended in 2XSDS sample buffer (with DTT) to achieve a final concentration of 2x10<sup>5</sup> cells/μL. The samples were then heated to 95°C for 5 minutes, and either stored at -20°C or loaded into each well of a polyacrylamide gel (8-15%) using 10 μL per well, equivalent to 2x10<sup>6</sup> cells. Next, the gel was run for 1 hour at 200 V and 400 mA in a running buffer containing tris 20 mM, 191.6 mM glycine, and 0.1% SDS, to separate the proteins based on their size.

The PageRuler™ prestained protein ladder ranging from 10 to 180 kDa was utilized as the marker. However, when gels were stained with Coomassie blue (consisting of Blue G250

0.02%, MeOH 30%, Acetic acid 10%, and Copper sulfate 0.1%), the PageRuler™ unstained protein ladder spanning from 10 to 200 kDa was chosen as the marker.

#### **.2.2.23.2 Proteins blotting and blocking**

After the running step, the gel and four thin Whatman papers were soaked in the transfer buffer (without methanol). The eluted proteins were transferred overnight at 4°C to a nitrocellulose membrane of 0.22 µm pore size in the transfer buffer (Tris Base 25 mM; Glycine 192 mM; MeOH 10%-20%). The western blot sandwich for the transfer was assembled as followed: Cathode core - sponge pad - two thin Whatman papers (filter paper) – gel – blotting membrane - two thin Whatman papers - sponge pad - anode core. After the transfer step, the membrane was blocked overnight at 4°C in the blocking solution (Tris-buffered saline (TBS-T) supplemented with 0.05% tween 20 and 3% BSA (**TBS-T** 10X: 0.5M Tris-HCl; 1.5M NaCl, pH 7.4 ± 0.1 with 0.5% Tween 20)).

The proteins were blotted overnight at 4°C with anti-cMyc mouse monoclonal primary antibodies (1:500) to identify the various tools of interest. Additionally, anti-Bip rabbit polyclonal primary antibodies (1:50'000) were added to detect the trypanosomal binding protein *TbBiP*, which is a 70 kDa housekeeping protein used as a loading control. To detect the *TbBiP* protein and the different tools, Goat anti-Rabbit IgG 800 secondary antibodies (1:10,000) (IRDye® 800CW; LI-COR Biosciences - GmbH) and Goat anti-mouse IgG 680 secondary antibodies (1:10,000) (IRDye® 680RD; LI-COR Biosciences - GmbH) were used, respectively.

All the antibodies were diluted in TBS-T (1X: 0.05M Tris-HCl; 0.15M NaCl, pH 7.4 ± 0.1 with 0.05% Tween 20) supplemented with 3% BSA.

All incubation steps were performed using a rocking platform.

To ensure that proteins were correctly transferred from the gel to the nitrocellulose membrane, a solution of Red Ponceau S staining (Ponceau S 5%; 1% acetic acid) was employed. After the transfer, the gel was Coomassie blue-stained in order to verify the efficacy of the transfer.

Table 8 below summarizes the details of all the antibodies used in the western blot:

**Table 8: Antibodies used for Western Blot**

| <b>Epitope</b>                                                                                                                | <b>Primary antibodies and dilution</b>                                | <b>References</b>                              | <b>Secondary antibodies and dilution</b>                                                   | <b>References</b>                    | <b>Detection device</b>           |
|-------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------|------------------------------------------------|--------------------------------------------------------------------------------------------|--------------------------------------|-----------------------------------|
| cMyc-Tag                                                                                                                      | Mouse anti-cMyc monoclonal antibodies (1:500-1:1000 dilution)         | 9E10; Invitrogen; ThermoFisher Scientific; USA | IRDye®800CW Goat anti-Mouse IgG secondary antibody (green) (1:5'000-1:10'000 dilution)     | 926-32210 Li-Cor®; USA               | Odyssey CLx scanner; Li-Cor®; USA |
| Biotin                                                                                                                        | IRDye®800CW Streptavidin (1:2'500 dilution)                           | 926-32230 Li-Cor®; USA                         | Acting as primary and secondary antibodies                                                 | 926-32230 Li-Cor®; USA               | Odyssey CLx scanner; Li-Cor®; USA |
| <i>TbAK (T. brucei</i> adenosine kinase)                                                                                      | Rabbit polyclonal anti-TbAK2360 EAA/10 antibodies (1:16'000 dilution) | H. S. San Raffaele, Milano Italy               | Goat anti-IgG (H+L) Dylight™680 Conjugated polyclonal antibodies (red) (1:10'000 dilution) | #35568; ThermoFisher Scientific; USA | Odyssey CLx scanner; Li-Cor®; USA |
| <i>TbBip (T. brucei</i> homolog of immunoglobulin heavy chain binding protein (loading control for the housekeeping protein)) | Rabbit polyclonal anti-Bip antibodies (1:50'000 - 1:500'000) dilution | A generous gift from Dr. Jay Bangs             | Goat anti-IgG (H+L) Dylight™680 Conjugated polyclonal antibodies (red) (1:10'000 dilution) | #35568; ThermoFisher Scientific; USA | Odyssey CLx scanner; Li-Cor®; USA |

#### *.2.2.24 Time dependant protein expression*

Once the correct integration of the three different genes of interest was validated, the next step was to verify the overexpression of the tool using western blot analysis. Tetracycline (1 µg/mL) was added to the culture medium to induce the expression of the different fusion proteins. To find out when the overexpression of the tools is highest and reaches its maximum, the different cell lines were incubated for 4 days in the presence of tetracycline (1 µg/mL), and each day a sample was prepared as follows:  $\approx 10^7$  total cells were harvested by centrifugation ( $600\text{--}1300 \times g$ , 10 min, 4°C or RT) and the pellet was resuspended in PBS supplemented with 1mM PMSF and then centrifuged again before another final washing step in PBS only. After centrifugation, the pellet was resuspended in 2XSDS buffer containing DTT in a volume necessary to obtain a final concentration of  $2 \times 10^5$  cells/µl. For example, 50 µL would be required for  $10^7$  cells in total. The samples were then heated at 95°C for 5 min and 10 µL of each sample corresponding to  $2 \times 10^6$  cells were loaded into a 10-15% SDS gel and the latter was run at 200 V, 400 mA for 1-2h. The eluted proteins were then transferred to a nitrocellulose membrane of 0.22 µm pore size and stained with anti-cMyc antibodies to reveal the proteins of interest (i.e. cMyc\_BirA\*\_**GGG**\_SNAP-tag, cMyc\_TBirA\*\_**GGG**\_SNAP-tag, cMyc\_mTBirA\*\_**GGG**\_SNAP-tag, cMyc\_BirA\*\_**GGGS**\_SNAP-tag, cMyc\_BirA\*\_**GGGS**\_SNAP-tag, cMyc\_BirA\*\_**PAPAP**\_SNAP-tag cMyc\_BirA\*\_**PTP**\_SNAP-tag).

Antibodies against Trypanosomal binding protein (*TbBiP*) were added to reveal the housekeeping protein *TbBIP* (70 kDa) used as a loading control. Rabbit polyclonal anti-Bip (1:50,000) and mouse monoclonal anti-cMyc (1:500) antibodies were diluted in 1XTris-buffered saline (TBS: 0.05M Tris-HCl; 0.15M NaCl, pH 7.4 ± 0.1) supplemented with 0.05% Tween 20 and 3% BSA. Goat anti-rabbit IgG 800 secondary antibodies (1:10,000) (IRDye® 800CW; LI-COR Biosciences - GmbH) and goat anti-mouse IgG 680 secondary antibodies (1:10,000) (IRDye® 680RD; LI-COR Biosciences - GmbH), were used to reveal the *TbBIP* protein and the different tools of interest, respectively. All antibodies were diluted in TBS-T supplemented with 0.05% Tween 20 and 3% BSA.

It is worth highlighting that the time-dependent expression studies of the various tools presented in Section .3.6 of the Results and Discussion chapter comprise representative

images derived from conducting multiple experiments. These experiments were performed numerous times, varying from three repetitions to even more in some instances.

#### *.2.2.25 Pull down of biotinylated proteins*

The following protocol is based on that described in the paper by Dr. Brooke Morriswood et al. (106) with slight modifications.

Transfected parasites were grown in a growth medium (HMI-9 with 1 µg/mL G418 and 2.5 µg/mL hygromycin) with the presence (sample of interest) or absence (negative control) of 20-50 µM BG-PEG<sub>4</sub>-Mtx (or 10 µM BG-PEG<sub>4</sub>-C1) until they reached log phase (1-2×10<sup>6</sup> cells/mL). Once a total of 10<sup>8</sup> cells in the log phase were attained (after 3-4 days), tetracycline (1 µg/mL) was added to induce expression of the different tools of interest (according to the incubated strain). For example, for the cMyc\_BirA\*\_**GGG**\_SNAP-tag tool, since the maximum expression of the protein is reached after 24 hours of induction by tetracycline (see results in section 3.6.1 below), the antibiotic should be added one day before the addition of biotin, i.e. 2 days before the pull-down.

All the different cultures were scaled up to a final volume of 300 mL for each condition (sample of interest and negative control) and 50 µM biotin (1 mM stock solution in IMDM without iFCS) was added either overnight (for BioID\_SNP-tag systems) or 10 min before pull-down (for the TurboID\_SNP-tag and miniTurbo\_SNP-tag systems). After 24 h (or 10 minutes) of incubation in the presence of 50 µM biotin (for the BioID\_SNP-tag tools), the cells were centrifuged (1800 × *g*, 5 min, 4°C), harvested, and washed three times with PBS. After the third wash, the cells were centrifuged again with the same parameters, and the harvested parasites were transferred into 1.5 ml Eppendorf tubes, extracted in PEME lysis buffer (2 mM EGTA, 1 mM MgSO<sub>4</sub>, 0.1 mM EDTA, 0.1 M PIPES [piperazine-N,N'-bis (2-ethanesulfonic acid)]-NaOH, pH 6.9, supplemented with 0.5% NP-40 (vol/vol) and 1% protease inhibitor cocktail (Sigma-Aldrich, P8340)), sonicated (3 cycles of 10 seconds, 20 kHz) in order to complete the lysis step and then incubated directly on ice for 30 min with gentle mixing. A 5% sample named **E1** was taken for further analysis. The detergent soluble fraction was separated from the insoluble fraction by centrifugation at 16,000 × *g*, 5-10 min, 4°C. Again, a 5% sample labeled **S1** was taken from the supernatant. The insoluble pellet must be over-extracted with RIPA lysis buffer

(SDS 0.1%, NP-40 1%, Deoxycholate 0.5%) (30 min, 4°C, with moderate mixing) in order to obtain as much protein as possible. A 5% sample designated **P1** was taken from this further extraction for the subsequent experiment. After centrifugation (16,000 x *g* for 5-10 minutes at 4°C) of the resulting lysate allowing the separation of solubilized proteins from cell debris, a 5% sample denoted **S2** was collected from the resulting supernatant.

Approximately 100-500 µL of pre-washed Streptavidin-coated Dynabeads (4-5 times in PBS supplemented with 0.1% BSA for the pre-washing) were added to **S1** and **S2** and an incubation step at 4° for 4 hours (or overnight) with moderate mixing was performed. After the incubation period, the beads added to both fractions were magnetically separated from the unbound material using the Dyna-Mag magnetic device. 5% samples were taken from the unbound fractions designated **F1** and **F2**. The beads were then washed 4-6 times for 2 minutes, twice with 1XPBS or 1XTBS and two to four times with PBS supplemented with 0.1% BSA for a total of 6 washes. The 100% washed beads were finally resuspended in an excess of biotin (5-25 mM), heated at 95°C for 5 minutes, and finally eluted in 4X Laemili buffer (50-100 µL) or 4XSDS sample loading buffer, before being heated a second time at 95°C for 5 minutes to obtain the respective **B1** and **B2** fractions that were used directly for western blot analysis. The collected **E1**, **S1**, **S2**, **P1**, **F1**, **F2**, **B1**, and **B2** samples were equally separated by SDS-PAGE, and the migrated proteins were transferred to a 0.2 µm nitrocellulose membrane and blotted using streptavidin dye (IRDye 800CW Streptavidin, LiCor) at a dilution of 1:2500 in TBS-T. For a more in-depth analysis concerning the nature of the proteins retained on the beads, a mass spectrometry (MS) conducted by the proteomic platform of Dr. Alexandre Hainard and his team (CMU, Geneva, Switzerland) was performed. For this analysis, larger-scale parasite cultures, typically between 300 and 500 mL of culture media (containing ~1-5x10<sup>9</sup> cells in total) were required.

According to the findings of Cheah JS, using PEME lysis buffer supplemented with 1% IGEPAL-CA630 and 0.4% SDS resulted in significantly improved outcomes. The addition of these two detergents increased solubility and facilitated the efficient elution of biotinylated proteins from the beads (148).

Figure 19 below schematically describes the entire biotinylated protein pull-down protocol.

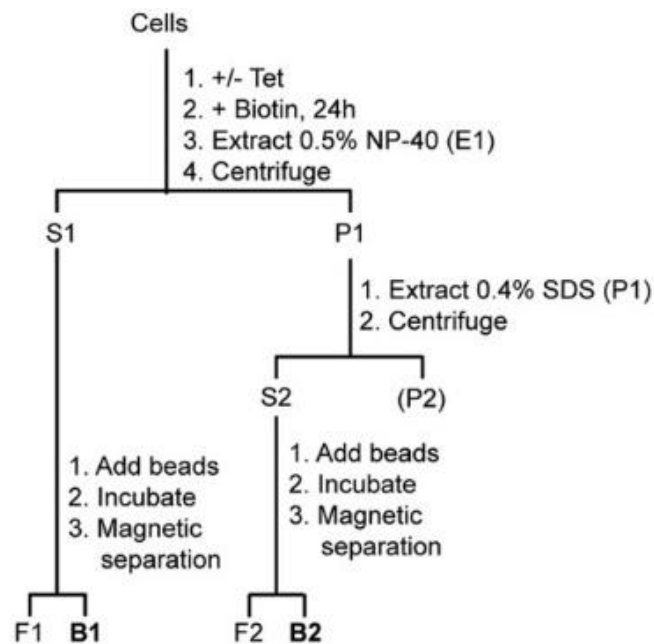


**Figure 19:** The pull-down procedure for capturing biotinylated proteins (taken from Morriswood, Havlicek, et al. 2013, (106))

Please note that the pull-down experiments conducted in Section .3.7 of the Results and Discussion chapter were repeated multiple times. However, only the images depicting the most distinct and conclusive outcomes were selected for inclusion.

#### **.2.2.26 SNAP-tag pure protein extraction**

To confirm the adequate obtention of parasites having integrated only the SNAP-tag gene of interest, a parasite lysate was prepared by centrifuging approximately 20 mL of cell cultures at  $1300 \times g$  for 10 minutes at RT. The resulting pellets were washed three times with 1X PBS before extracting the parasite lysate for 30 minutes in PEME lysis buffer supplemented with 0.5% IGEPAL® CA-630, 0.4% SDS, and 1% protease inhibitor cocktail.

To ensure complete and optimal lysis of the lysate, three cycles of sonication for 10 seconds each were performed. The lysate was then incubated for 20 minutes before being centrifuged at  $16'000 \times g$  for 5 minutes. The concentration of total protein present in the lysate was determined to be 0.25 mg/mL using the Pierce™ BCA Protein Assay Kit.

Finally, Laemmli Sample Buffer 4X was added to the sample, and electrophoretic separation was carried out using a tris-tricine gel 10%. The gel was run at 80V for 20 minutes, 120V for 30 minutes, and 150V for 45 minutes.

The band corresponding to the desired molecular weight (about 20 kDa for the pure SNAP-tag protein) was excised before being sent for MS analysis.

To understand the binding of the pure SNAP-tag protein to its substrate and be able to detect it in MS and thereby compare the results obtained with the *in vivo* experiments (pure SNAP-tag expression in the parasite), an *in-vitro* assay consisting of incubating the pure protein with two different BG-PEG<sub>4</sub> derivatives was performed. The details are mentioned in the table below.

**Table 9: Binding reactions of pure SNAP protein with BG-PEG<sub>4</sub> derivatives performed at 37°C for 1h**

| <i>Sample</i>           | <i>1: SNAP_BG-PEG<sub>4</sub>-Mtx</i> | <i>2: SNAP_BG-PEG<sub>4</sub>-biotin</i> |
|-------------------------|---------------------------------------|------------------------------------------|
| DTT 50 mM               | 10 µL                                 | 10 µL                                    |
| SNAPtag 50 mM (1 mg/mL) | 20 µL                                 | 20 µL                                    |
| BGPEG-Mtx 50 mM         | 5 µL                                  | --                                       |
| BGPEG-biotin 50 mM      | --                                    | 5 µL                                     |
| HEPES 50 mM, pH 7.4     | 65 µL                                 | 65 µL                                    |
| <b>Total</b>            | 100 µL                                | 100 µL                                   |

After one hour of incubation at 37°C, 2XSDS sample buffer was added to the reaction tubes, then after 5 min of incubation at 95°C, an SDS-PAGE electrophoresis was performed (15% gel, 200V, 1-2h) followed by Coomassie blue staining for the revelation of bands of interest.

The obtained bands were cut (bands around 20 kDa) and sent for MS analysis.

### **.3.1 The linker story: the culmination point of BG derivatives and fusion proteins**

#### 3.1.1 The optimal length of PEG in BG-PEG-derivatives

This study aimed to find the optimal number of PEGs to allow both the BG and the compound moieties (Figure 12 chapter 3.1.4) to bind to their specific targets while preserving the permeability of the molecule to cell membranes.

First, let's talk about PEGs, what are they? PEG polymers, the abbreviation for Polyethylene Glycol polymers are flexible amphiphilic molecules that have been shown to increase and improve the solubility of conjugated drugs (150, 151) (Figure 20).

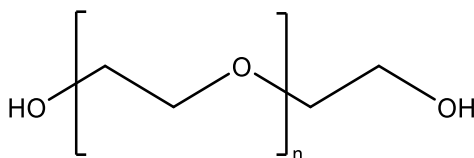


**Figure 20:** General structure of a poly (ethylene glycol) (PEG) chain, (image made with ChemDraw 19.1)

PEGs of linear formula  $\text{H}(\text{OCH}_2\text{CH}_2)_n\text{OH}$  or  $\text{HO}-(\text{CH}_2\text{CH}_2\text{O})_n-\text{CH}_2\text{CH}_2-\text{OH}$  (151) (Figure 20 above) are composed of repeated units of ethylene glycol, each unit of which associates in solution with about two water molecules, making them highly hydrated compounds (151-153). This hydrophilicity limits their diffusion through the lipidic bilayers of the membrane, their good absorption would thus be explained by their passive transport using the paracellular pathway (151-155).

With all this information, the objective here is to predict the optimal number of PEG units that would allow and simplify the formation of a ternary complex with the fusion protein composed of BirA\* and SNAP-tag.

For this purpose, a computational approach was implemented by Margaux Hérítier referring to the article by Dr. Nan Bai et al. (149). In short, the approach described in the paper is based on four steps which are the following: 1) Building diverse binding modes, 2) Generating candidate linker conformations, 3) Building initial models of the ternary complex, and 4)

Analyzing ternary complex models that correspond to the final filtering step in which the fraction of Fully Compatible Complexes (**FCC**) is calculated (149). The calculation of the **FCC** measures and thus determines the extent to which a given linker is geometrically and energetically compatible with the various preferred means of interaction of the fusion protein formed by the two protein domains (149). In other words, the higher the **FCC** value obtained, the better the probability of the formation of the ternary complex between the compound of interest and the different domains forming the fusion protein. In Dr. Nan Bai et al. paper, they calculated the FCC value in order to find out the best linker that can lead to the formation of the PROTAC ternary complex that is composed of a Warhead, linker, and the E3 recruiter (as drawn and shown in Figure 21).

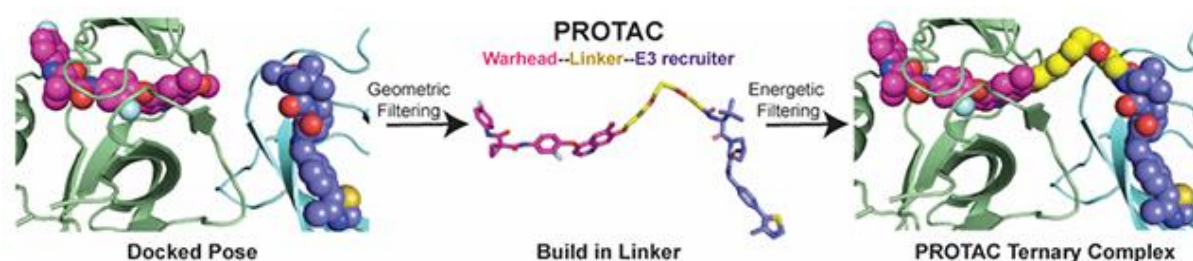


*Figure 21: Method based on the calculation of the FCC (fraction of fully compatible complexes) to model the formation of the ternary complex PROTAC, image taken from Bai, N., et al. (2021) article (149)*

According to the Figure 22, a linker composed of two PEG units (n2) seems to be slightly the best, although it is not statistically significantly different than n4 or n5 or n6. Studies show that increasing the number of PEGs would improve the proper formation of the complex. Indeed, Becker et al. (156) reported in the paper that PEG linkers containing at least three or more units ensured good performance in the Y3H system and were required for the successful formation of the complex of interest (156). Furthermore, linkers composed of four PEG units have been widely used in other studies (157-159). Moreover, Gendreizig, Kindermann, and Johnsson have well described in their paper the synthesis as well as the use of the BG derivative linked to four PEG units and Mtx (159).

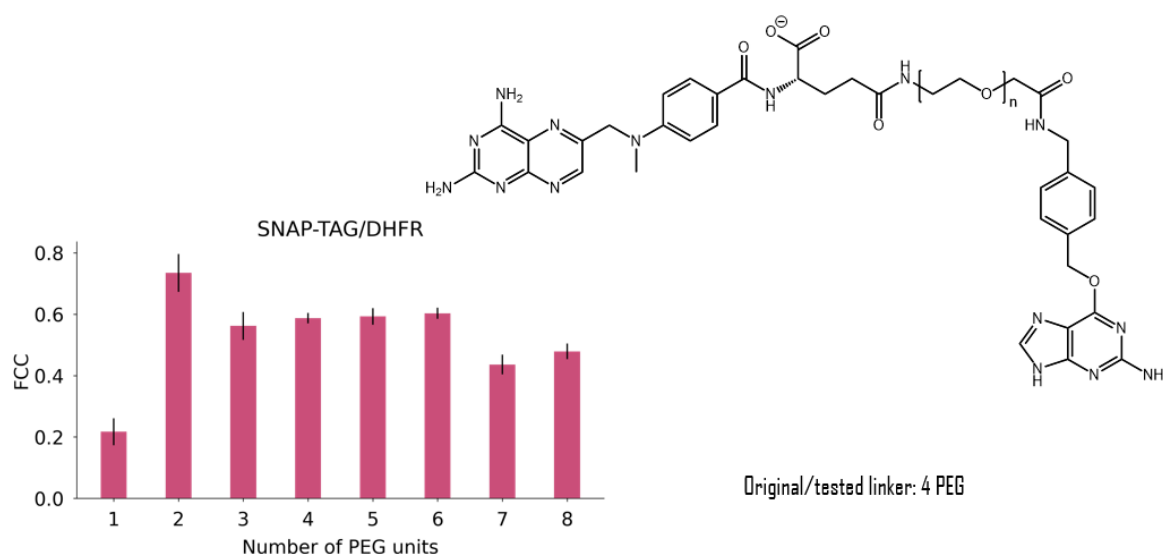


Figure 22: Study of the PEG bond length between the BG moiety and the compound of interest Mtx to form the ternary complex with the DHFR-TS target, using FCC calculation (established by Margaux Héritier & Jiri Bejcek)

As shown in Figure 22 above, a linker composed of five or six PEG units could also have been chosen given the good FCC obtained, but according to some studies, the addition of multiple PEG units could impede the permeability of the compound across biological membranes (due to the reduction of  $\log P$  by adding PEG units) (160-162). Indeed, in a study, 9 PEG units were integrated between the BG and NHS (*N*-Hydroxysuccinimide) fragments to increase the hydrophilicity of the compound and subsequently make the BG-NHS impermeable to the plasma membrane. BG-PEG<sub>9</sub>-NHS was indeed experimentally validated as a membrane-impermeable compound (163). Therefore, a linker composed of 4 PEG units seems to be a good compromise to maintain permeability and ensure the correct formation of the complex of interest. In addition, as mentioned above, a linker consisting of 4 PEG units has been chosen and selected in many studies (163-168). Furthermore, as shown in the graph in Figure 22 above, a plateau starts to appear from 4 units of PEG, showing that there would be no significant difference in the probability of ternary complex formation between linkers composed of 4, 5, or 6 PEG units, thus, PEG<sub>4</sub> has been selected as the spacer linking the BG

moiety to the small molecule of interest (Mtx, C1 or others) to perform all the experiments that will allow the development as well as the validation the tool.

### *.3.1.2 Modeling study to determine the linker length between BirA\* & SNAP-tag*

To avoid steric hindrance and a certain rigidity that would impede each domain forming the fusion protein to play its full action, a linker must be added between the BirA\* and the SNAP-tag proteins.

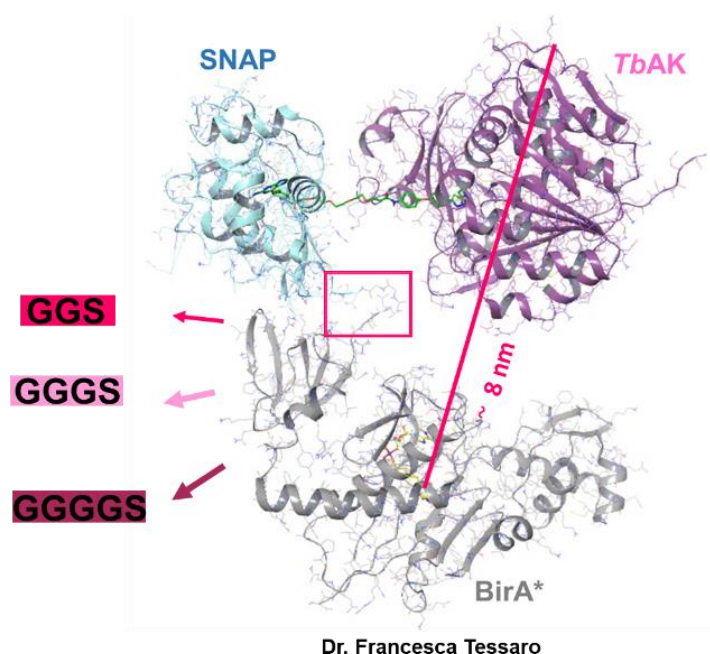


*Figure 23: Modeling study for the determination of the linker length between BirA\* and SNAP-tag parts (made by Dr. Francesca Tessaro)*

A modeling study (performed by my colleague Dr. Francesca Tessaro) was conducted to determine the optimal length of the linker. As indicated in Figure 23 above, with a linker consisting of 3 amino acids, the distance between the active site of BirA\* and the furthest amino acid from the TbAK target which was chosen as an example is about 8 nm, which is within the biotinylation window of 15 nm, and at the same time, BirA\* is sufficiently distant from the target, which allows it to be flexible enough to be fully active.

This selected linker is composed of glycine and serine, these amino acids were chosen as they are known to form flexible spacers (169).

The GGS linker has been the first tested in the development of the tool of interest containing the original system (BioID) or its variants (TurboID and miniTurbo).

The effect of the variation of the length of the flexible GGS linker has also been studied, as it was interesting to know the limit of its length that would allow the maintenance of the proper functioning of the system (see results below). The two other flexible spacers that have been designed were composed of either one or two additional glycines (as shown in Figure 23 above).

Besides the length of the linker, the study of the variation of its amino acids' nature would provide crucial elements concerning the proper functioning of the tool of interest. Keeping in mind the general 3D arrangements needed (Figure 23) such linkers have been constructed and tested and will be discussed hereunder.

#### *.3.1.3 Selection of the most suitable linker using the computational approach*

As indicated above in Section 3.1.1, the FCC calculation predicts how a given linker can be geometrically and energetically compatible with the preferred interaction modes of the different domains of the fusion protein of interest and therefore, the higher the FCC value, the higher the probability of forming a ternary complex between BirA\*, the tested linker and the SNAP-tag.

Unlike the PEG linker between BG and the small molecule of interest (Mtx, C1, or other), the objective here is to maintain a certain distance between the different domains avoiding the agglomeration of the two domains BirA\* and SNAP-tag. Indeed, if agglomeration would occur because of overly flexible linkers such as GGS or GGGGS, BG-PEG<sub>4</sub> derivatives might theoretically have restricted access to the SNAP-tag cavity due to the possible and potential masking of the Cys145 of the SNAP-tag active site by BirA\*, which would prevent the proper functioning of the system. Therefore, a linker that would give the lowest FCC value would be the most adequate because it would avoid the formation of this agglomeration.

The amino acids proline, alanine, and threonine are known to form rigid linkers, especially the (PA)<sub>n</sub> or (PT)<sub>n</sub> repeats (169, 170). According to Xiaoying Chen's paper, the use of rigid linkers could be a good alternative when the biological activity of one of the domains composing the

fusion protein is reduced due to the concealment of its active site by the other part of the protein (169).

It was therefore decided to test at the computational level using the same approach as described in Bai N and *al.* paper (149) and previously used to determine the optimal length of the linker composed of PEG units in BG derivatives, different rigid linkers but this time the result that will give the lowest probability of forming the complex (the lowest FCC) (Figure 24) will be taken into account, and will therefore be selected, and then confirmed experimentally.

The following rigid linkers have been tested computationally by my colleague Margaux H  ritier: PAPAP (chosen according to Xiaoying Chen's paper (169)) and PTP, PTPTP, PAP, and PAPA (chosen by myself) in order to determine which of them will give rise to low FCC values enabling the maintenance of the biological activities of each of the domains forming the fusion protein (Figure 24).

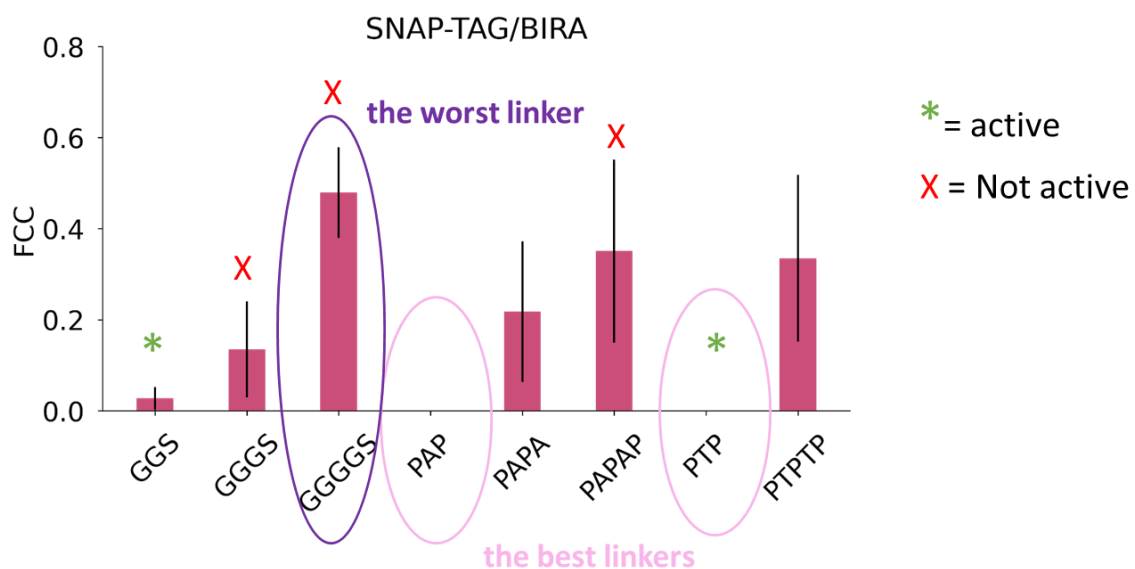


*Figure 24: Determination of FCC as a function of linker nature and length (made by Margaux H  ritier and adapted from the article by Dr. Nan Bai et al. (150))*

According to the analyses (Figure 24 above), the flexible linkers GGGG and GGGGS as well as the rigid PAPAP have high probability of forming the ternary complex and thus agglomerates that can result in masking of the SNAP-tag active site, hence hindering the binding of its

substrate and thus the BG derivative to its active site. On the other hand, GGS, PAP, and PTP gave the lowest FCC values and therefore could potentially be the linkers that will give the best results.

To prove experimentally the predictivity of the method and the hypotheses put forward by this study, the best linkers such as GGS and PTP but also some of the less favourable linkers such as GGGG, GGGGS, and PAPAP have been tested.

## .3.2 Generation of different plasmids for parasite transfection

### .3.2.1 Design & production of plasmids of interest: *BioID* variants

As mentioned in the introduction, two variants of BirA\* have been developed and have been tested to determine which one will lead to the best-performing tool. For these variants, the flexible linker GGS has been selected, being the first one that has been studied in the computational study carried out by Dr. Francesca Tessaro, giving an adequate distance between the different domains of the fusion protein in addition to a correct FCC value close to zero.

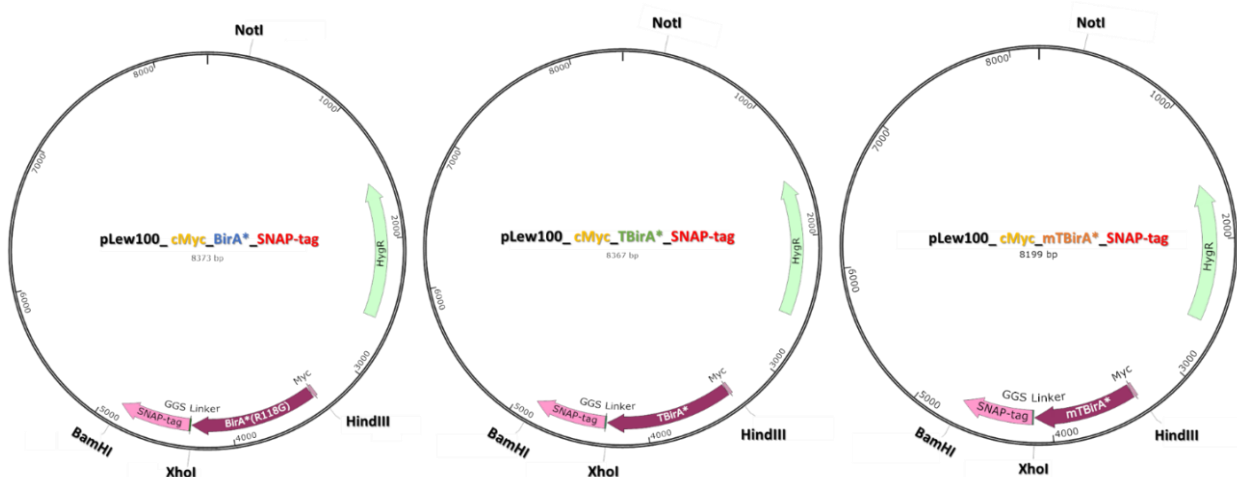


**Figure 25:** The three targeted plasmids containing either the gene encoding the original *BioID* system linked to SNAP-tag or the one coding for the *TurboID* variant or the *miniTurbo*, both linked to SNAP-tag

The three aimed plasmids (Figure 25 above) have been designed, produced, and successfully obtained (as described in sections 2.2.1, 2.2.2, and 2.2.3). They were verified by colony PCR and sequencing (the details of PCRs and sequencing showing the correct insertion of either

TbirA\* or mTbirA\* genes of interest in the pre-digested plasmid from which BirA\* was removed, are present in Appendixes 1, 2, and 3).

For the first plasmid pLew100\_cMyc\_BirA\*\_GGG\_SNAP-tag, the synthesis of the GGS\_SNAP-tag gene was needed before inserting it downstream BirA\* gene (see section 2.2.1). For the two other plasmids, a PCR of either TBirA\* or mTBirA\* genes was needed beforehand in order to amplify these genes and to incorporate the restriction sites of interest (HindIII & XhoI) as well as the cMyc gene (Figure 26, Figure 27). The generated fragments were inserted upstream SNAP-tag gene (see sections 2.2.2 & 2.2.3 for details). PCR and sequencing results showing the correct insertion of the fragments of interest are illustrated in the figures below (details in Appendixes 2 & 3).

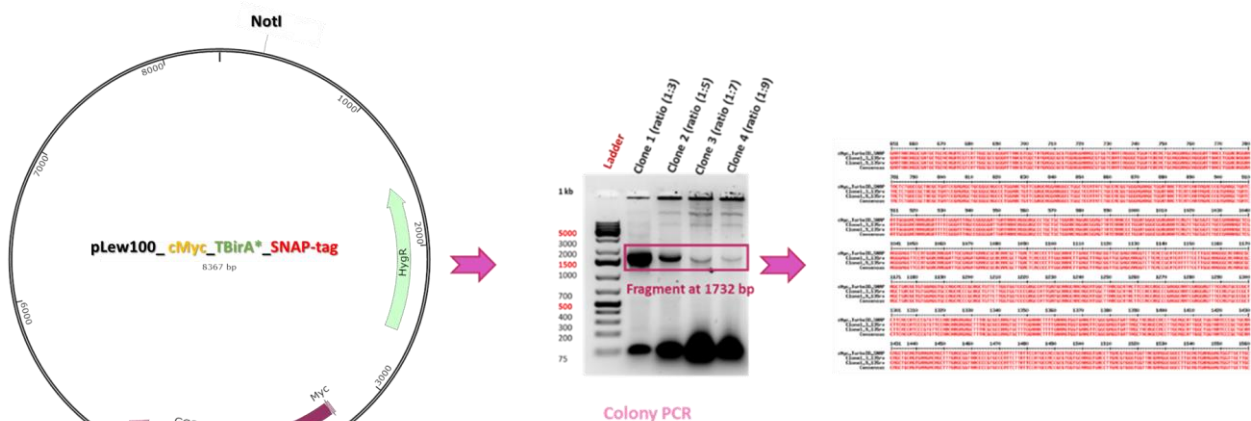


**Figure 26:** Colony PCR and sequencing showing the successful insertion of the HindIII\_cMyc\_TBirA\*\_XhoI fragment upstream of the SNAP-tag gene

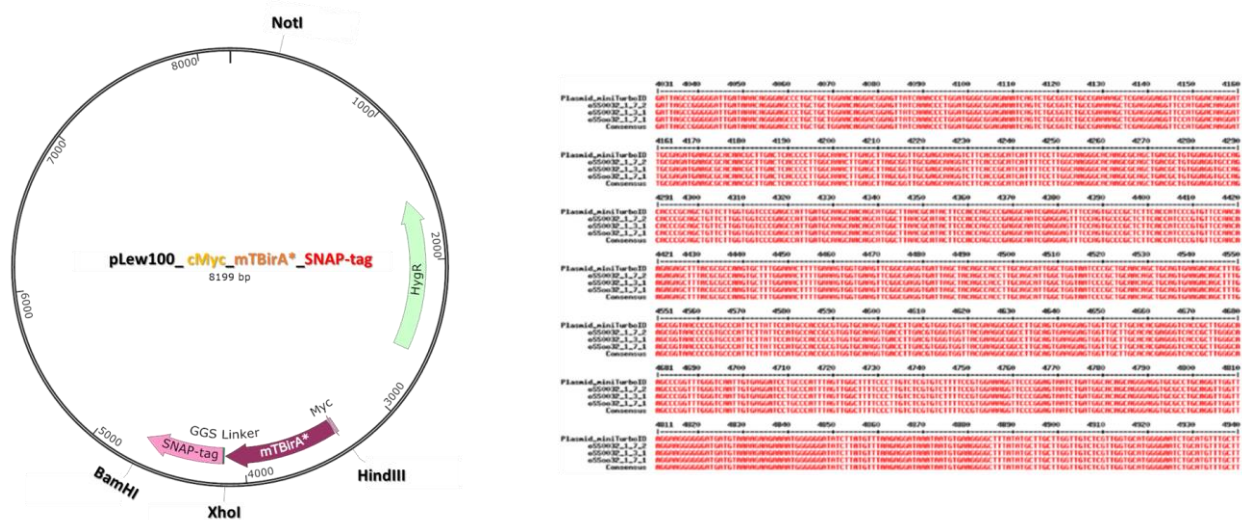


Figure 27: Sequencing validating the successful insertion of the HindIII\_cMyc\_mTBirA\*\_XhoI fragment upstream of the SNAP-tag gene

The results of forward and reverse sequencings (Figure 26: oCA134 TurboID FW 5'-CGCGCCTTCGAGTTTTTTTTTCC-3'; oCA135 TurboID RV 5'-CCTGCTGTGCCATCAGATTAC -3' respectively target the GPEET 5' UTR & amp upstream and ALD 3 & acute long UTR regions contained in the main vector plasmid between which the gene of interest, cMyc\_mTBirA\*\_GGS\_SNAP-tag, is located; Figure 27: oSS0031 miniTurboID RV 5'-CCGCTCGAGcttttcggcagaccgagactg-3' ; oSS0032 miniTurboID FW 5'-CTCTGGCCGCTATGCTGATC-3' respectively target the hygromycin resistance gene region present in the main vector plasmid and the mTBirA\* gene region) perfectly demonstrate the alignments between the theoretical genes of interest and those obtained experimentally (when the nucleotides match, the alignment becomes red as illustrated in Figures 26 and 27).

### 3.2.2 Design & production of plasmids of interest containing: Linkers variants

#### 3.2.2.1 Modification of the length of the original GGS linker

As previously described, glycine and serine are amino acids known to form flexible linkers (169, 171).

The result of the modeling study showed an appropriate distance between BirA\* and SNAP-tag allowing these two domains to play their full role, hence the choice of GGS linker in the first place (see the result above in section 3.1.2).

### ¶ Incorporation of glycine into the GGS linker by site-directed mutagenesis to obtain GGGG:

As shown in the sequencing result of Figure 28, the experimental alignment perfectly matches the theoretical one (part A of Figure 28). In fact, the presence of the linker of interest GGGG is clearly visible. Moreover, when the experimentally obtained sequence is compared with the original theoretical gene containing the GGS linker, the addition of glycine, illustrated by the addition of the **TGG** nucleotides, is well observed in the alignment of the genes in part B of Figure 28, proving once again the successful obtaining of the GGGG linker.

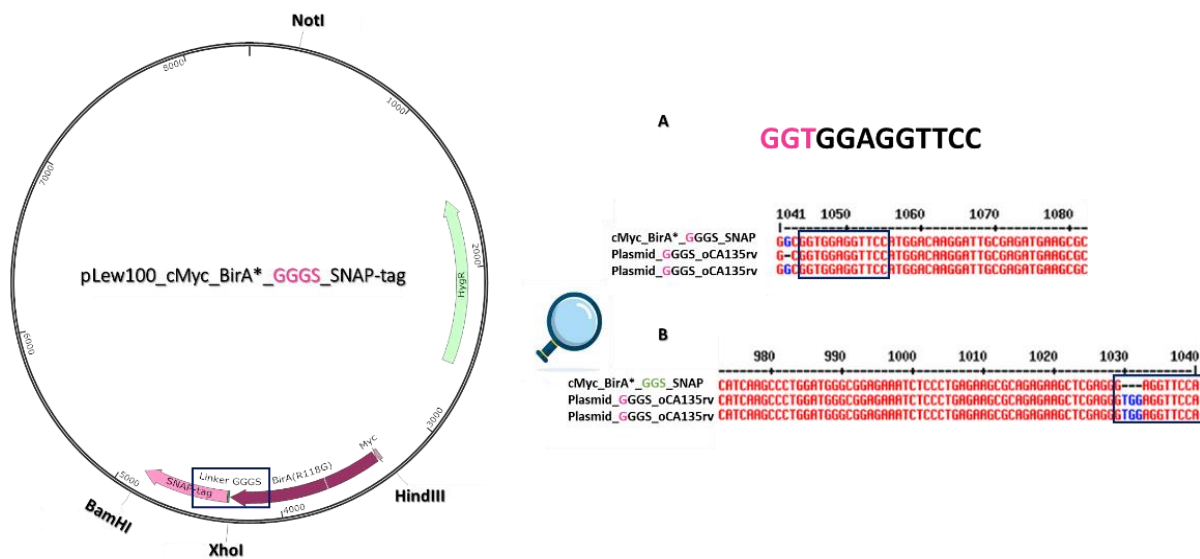


**Figure 28:** **A:** Sequencing comparing the sequence of interest of the newly produced plasmids (Plasmid\_GGGS\_oCA135rv/ oCA134fw) with that of the theoretical gene (cMyc\_BirA\*\_GGGS\_SNAP); **B:** Sequencing comparing the sequence of interest of the newly produced plasmids (Plasmid\_GGGS\_oCA135rv/ oCA134fw) with that of the original BirA\_GGS\_SNAP insert (minus one glycine).

### ¶ Incorporation of glycine into the GGGG linker by site-directed mutagenesis to obtain GGGGS:

The results of the incorporation of glycine into the GGGG linker by site-directed mutagenesis to obtain GGGGS shows that the correct plasmid was obtained. In indeed, the alignment of the linker obtained experimentally with the theoretical one is perfect (as shown in part A of Figure 29) in addition to the evidence of the addition of two glycines as presented in part B of Figure 29.

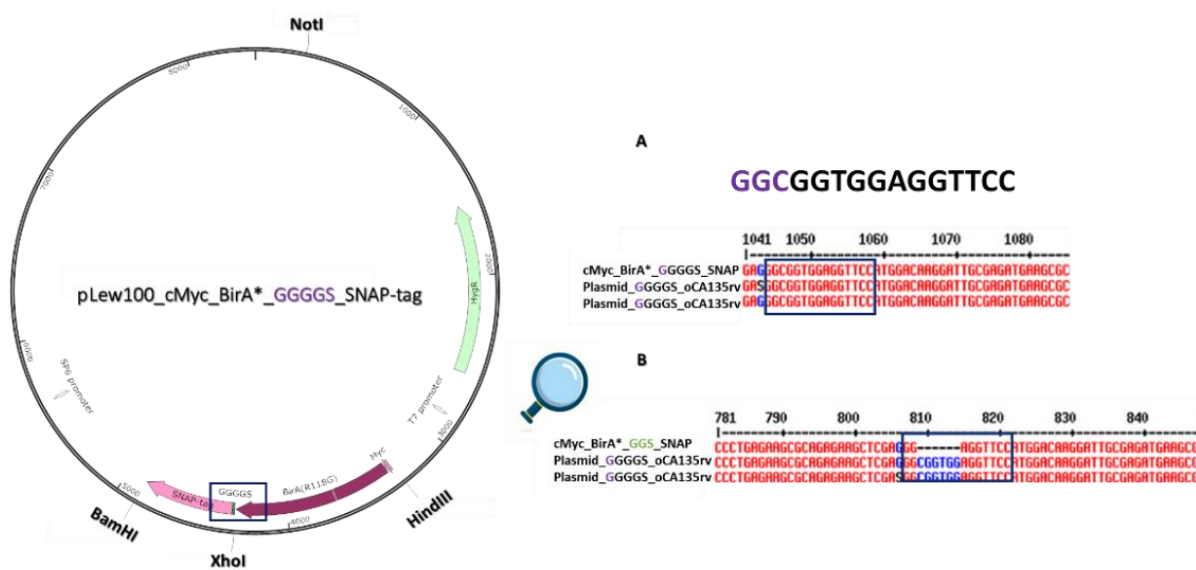


**Figure 29:** **A:** Sequencing comparing the sequence of interest of the newly produced plasmids (Plasmid\_GGGGS\_oCA135rv/ oCA134fw) with that of the theoretical gene (cMyc\_BirA\*\_GGGGS\_SNAPE); **B:** Sequencing comparing the sequence of interest of the newly produced plasmids (Plasmid\_GGGGS\_oCA135rv/ oCA134fw) with the sequence of interest of the original BirA\*\_GGS\_SNAPE insert (minus two glycines)

In conclusion, the site-directed mutagenesis worked well (Figures 28 and 29), and the sequencing results of the obtained plasmids (after their extraction and purification of the transformed nova blue bacteria) reveal the correct insertion of either one or two glycines depending on the parent plasmid used in the reaction.

### 3.2.2.2 Modification of the nature of the amino acids composing the linker

**PAPAP linker:** A rigid linker composed of prolines and alanines which are amino acids that confer rigidity (169).

As detailed in Section 2.2.7, the PAPAP linker was generated by PCR amplification of the SNAP-tag gene using primers allowing the incorporation of the PAPAP-encoding end of DNA upstream of the SNAP-tag-encoding gene. The sequencing result (part A of Figure 30) reveals the perfect match between the theoretical and experimental PAPAP gene.

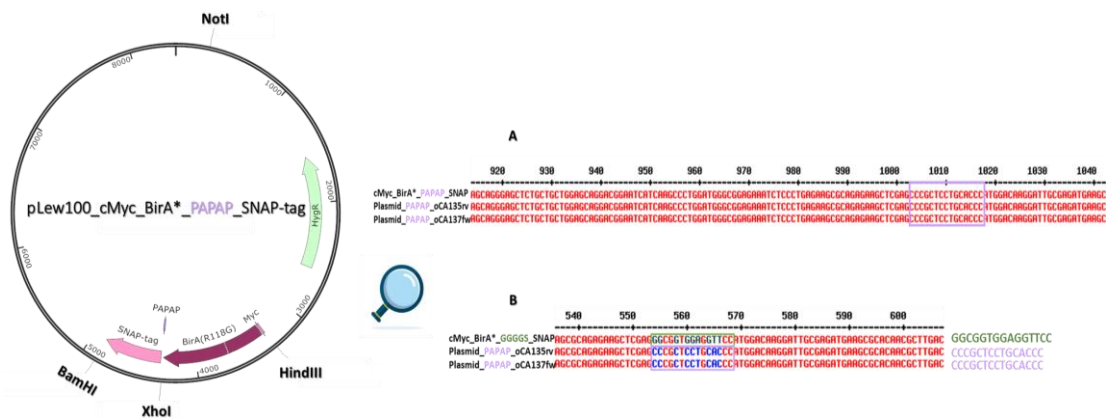


**Figure 30: A** Sequencing comparing the sequence of interest of the newly produced plasmids (Plasmid\_PAPAP\_oCA135rv/ oCA134fw) with that of the theoretical gene (cMyc\_BirA\*\_PAPAP\_SNAP); **B**: Sequencing comparing the sequence of interest of the newly produced plasmids (Plasmid\_PAPAP\_oCA135rv/ oCA134fw) with the sequence of interest of the plasmid containing the same linker length BirA\_GGGGS\_SNAP insert showing the correct insertion of the PAPAP linker of interest

Furthermore, when the gene coding for theoretical GGGGS (chosen because it contains the same number of amino acids as PAPAP) is aligned with the one obtained experimentally, a mismatch appears (part B of Figure 30 above) indicating that this linker no longer exists and that only PAPAP has been generated in the newly produced plasmid.

Details of plasmid construction are given in Appendix 5.

🔧 **PTP linker:** A rigid linker composed of prolines and threonines which are uncharged amino acids and provide a certain rigidity (169, 172)

PCR and cloning proved to be highly successful. The transformation resulted in a multitude of colonies, and a significant majority of them (confirmed through sequencing as shown in the Figure 31) contained the desired newly generated plasmids. Similarly, here, what is observed is the proper alignment between one of the experimentally obtained PTP genes and the theoretical one (part A of Figure 31). Moreover, when the experimental gene is aligned with the theoretical one containing GGS, a mismatch appears (part B of Figure 31) proving that only PTP has been well incorporated upstream of the SNAP-tag gene.

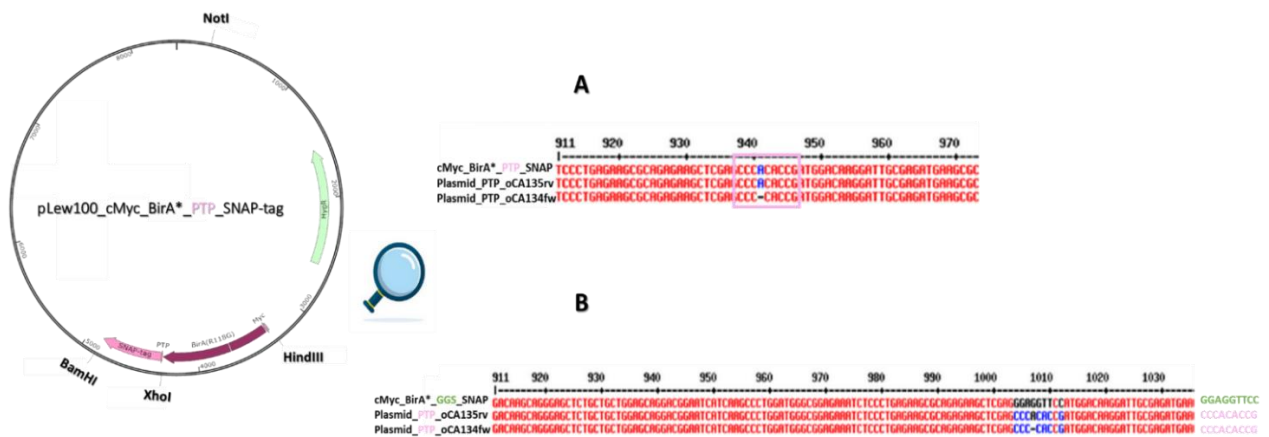


**Figure 31:** **A** Sequencing comparing the sequence of interest of the newly produced plasmids (Plasmid\_PTP\_oCA135rv/ oCA134fw) with that of the theoretical gene (cMyc\_BirA\*\_PTP\_SNAP); **B**: Sequencing comparing the sequence of interest of the newly produced plasmids (Plasmid\_PTP\_oCA135rv/ oCA134fw) with the sequence of interest of the plasmid containing the same linker length BirA\_GGS\_SNAP insert showing the correct insertion of the PTP linker of interest

Details of plasmid construction are given in Appendix 6.

### .3.3 Transfection of plasmids of interests

The aimed plasmids obtained were transfected to parasites using the AMAXA Nucleofector™II by selecting the X-001 program as described in Section 2.2.13 and illustrated in Figure 32.

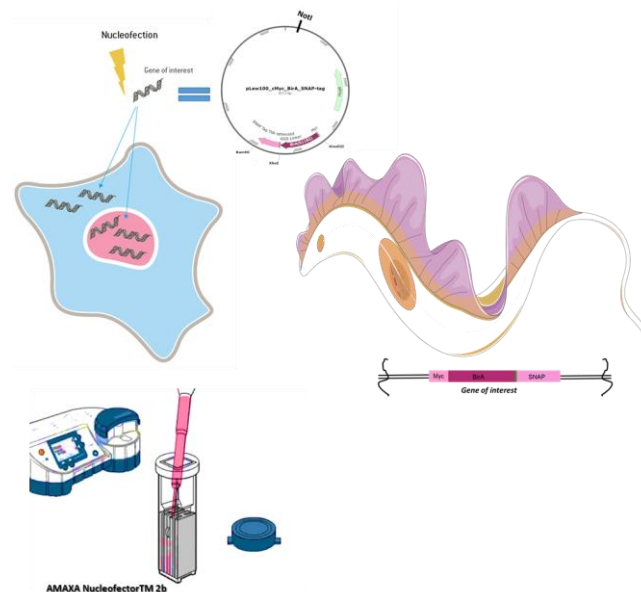


**Figure 32:** Electroporation method used for parasite transfection (adapted from (174))

In order to obtain stable cell lines, limited dilutions 1:5, 1:15, 1:100, 1:150, and 1:300 were performed and after a few days, clones started to appear (Table 10). To verify the correct insertion of the different genes of interest, gDNA extractions followed by PCR were performed. The results are shown in Figure 33.

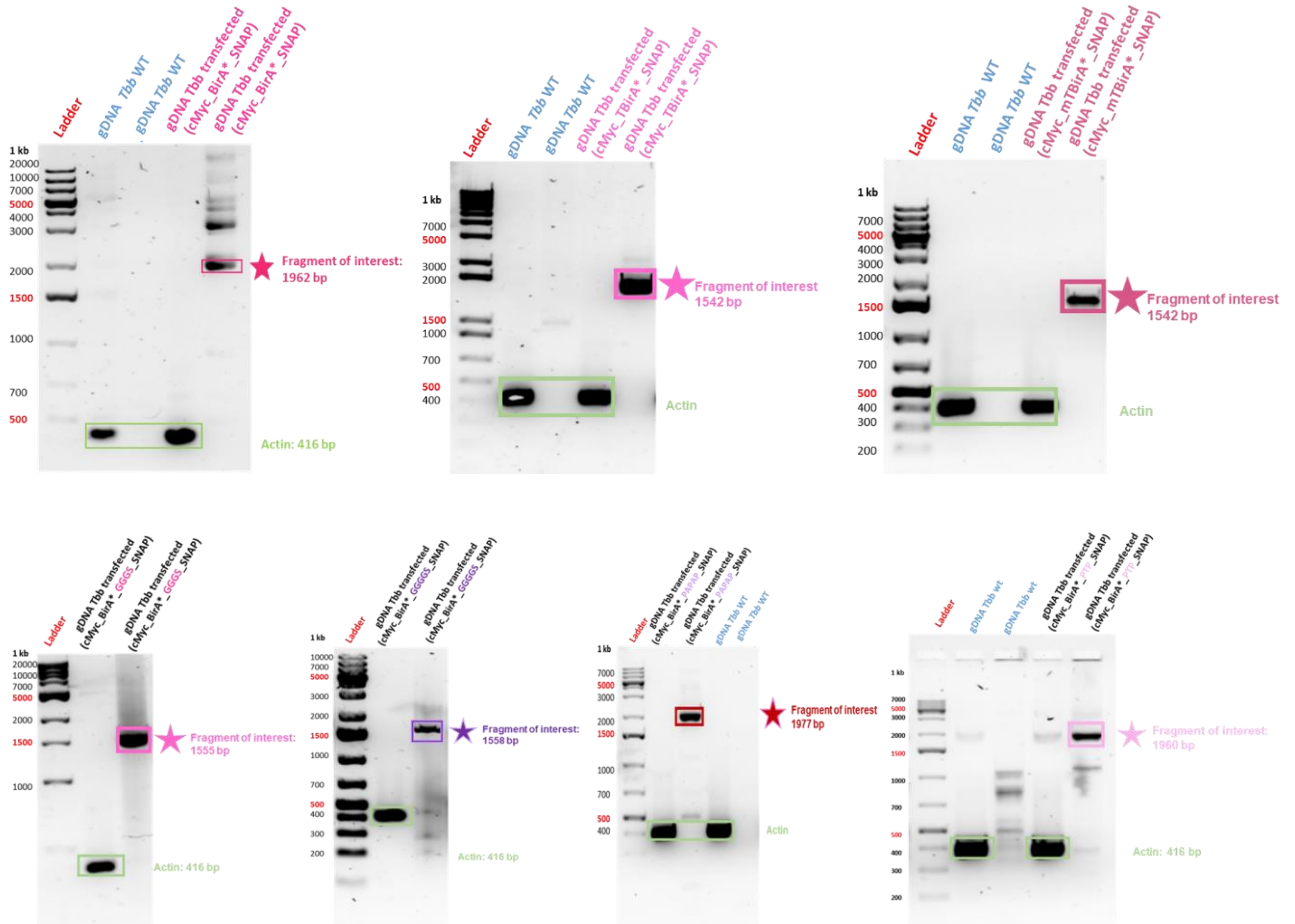


**Figure 33:** PCR results revealing the successful integration of the different genes of interest into the parasite genome

More than 50 primers were tested in order to find and select those that would bind best to the genomic part of the trypanosome (intergenic region) close to the insertion site of the gene of interest encoding the fusion protein where the homologous recombination was produced.

Primers **oSS0022** and **oSS0034** are reverse primers targeting different parts of Chromosome 1 ID X52586.1 and **oSS0023** and **oSS0033** are forward primers that target a part of the integrated gene of interest (details of their genetic sequences in Section 2.2.15) and these were selected for their ability to generate clear and accurate bands in PCR (as shown in Figure 33).

The visible bands around 1950 bp were obtained using primers **oSS0022** and **oSS0023** and those around 1550 bp were generated using primers **oSS0033** and **oSS0034**.

Figure 33 above shows clearly the presence of the different genes of interest validating that the right strains are correctly transfected.

The different clones appeared on average 4 to 7 days (sometimes beyond that, up to max 2 weeks) after transfection and selection by hygromycin.

Table 10 below presents the details of the different dilutions where the various clones were selected:

**Table 10: the appearance of clones as a function of dilution after the addition of hygromycin 2.5 µg/mL 24 h after electroporation**

| <i>Transfected Tbb</i> | <i>Selected clone in dilution (1:X)</i> |
|------------------------|-----------------------------------------|
| <b><i>Tbb_G0</i></b>   | 1:150                                   |
| <b><i>Tbb_TG0</i></b>  | 1:50                                    |
| <b><i>Tbb_mTG0</i></b> | 1:50                                    |
| <b><i>Tbb_G1</i></b>   | 1:100                                   |
| <b><i>Tbb_G2</i></b>   | 1:100                                   |
| <b><i>Tbb_PA</i></b>   | 1:50                                    |
| <b><i>Tbb_PTP</i></b>  | 1:200                                   |

### **.3.4 Fluorescence microscopy**

This experiment was performed to test the permeability of BG-PEG<sub>4</sub>-derived compounds. The test was performed using the BG-PEG<sub>4</sub>-FDA derivative, as the FDA can be excited in a wide range of wavelengths reaching its maximum excitation at 494 nm and emitting fluorescence at 521 nm (in the green spectrum) (174-176). The SpectraViewer website of Thermo Fischer indicates the excitation and emission wavelengths of different fluorescent compounds (176). If the derivative penetrates the cells, the parasites will emit fluorescence in the green field. This is clearly visible in the image of Figure 34. Indeed, the parasites that were incubated with BG-PEG<sub>4</sub>-FDA emit a green fluorescence and the amount emitted is 3.33 times higher than

that produced by the parasites that were not incubated with the compound, showing that this derivative penetrates well into the parasite and is, therefore, permeable to cells.

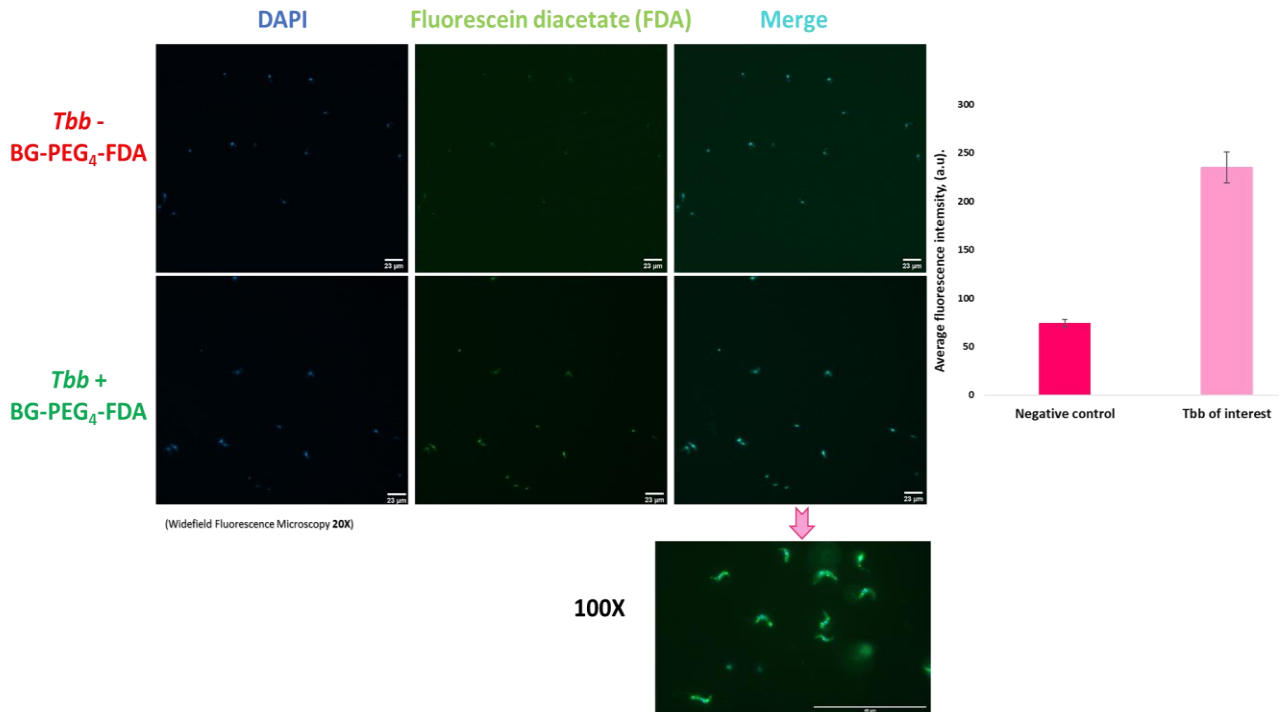


**Figure 34:** Microscopic images obtained by widefield fluorescence microscopy (excitation at 490 nm and emission at 520 nm) showing fluorescent cells revealing the entry of the BG-PEG<sub>4</sub>-FDA derivative into the parasite, indicating that the compound is cell permeable.

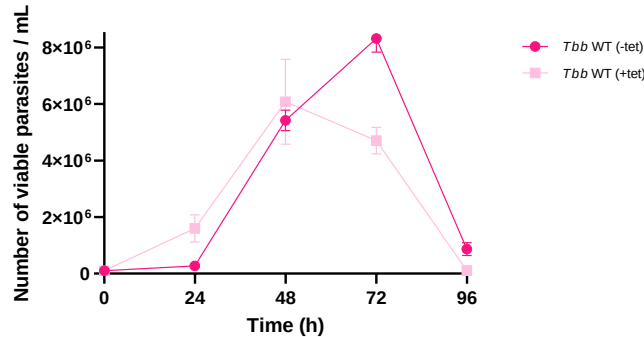
Parasites that have not been incubated with the BG-PEG<sub>4</sub>-FDA derivative emit some fluorescence, which is also visible in the image and histogram above (about 75% of the amount of fluorescence is emitted by the negative control (compared to 250% for the sample of interest)), this is due to the phenomenon of autofluorescence naturally and intrinsically present in eukaryotic and prokaryotic cells, related to the presence of fluorescent cellular components and metabolites (177).

### .3.5 Growth curves of different strains

#### .3.5.1 Growth curves of *Tbb*<sub>WT</sub> with and without tetracycline addition

To determine the experimental windows to be used an exhaustive analysis of *Tbb* growth under different conditions have been performed and the results are presented in this chapter

as well as in 3.3.5.2. As shown in Figure 35, *Tbb\_WT* parasites grow slightly differently depending on whether or not tetracycline is added.



|               | Population doubling time (PDT)<br>(h) (-tet) | Population doubling time<br>(PDT) (h) (+tet) |
|---------------|----------------------------------------------|----------------------------------------------|
| <i>Tbb</i> WT | 7.66                                         | 5.71                                         |

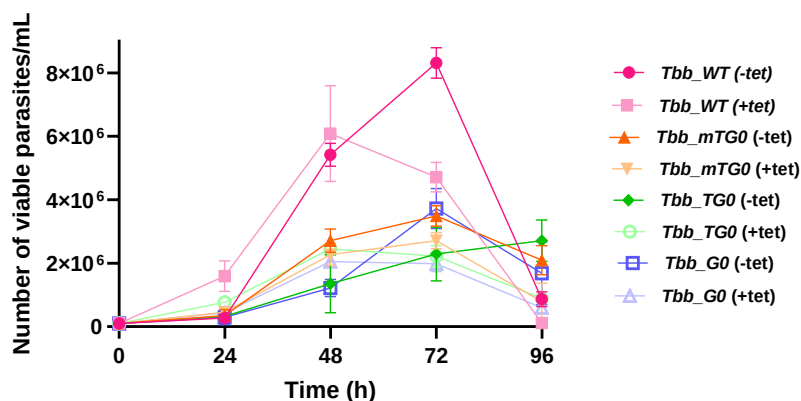
**Figure 35:** Growth curves and PDT of *Tbb\_WT* in the presence or absence of tetracycline (1μg/mL)

Indeed, when the antibiotic is added, they reach the maximum growth faster (PDT of **5.71 h**) after only 48 h compared to parasites that were incubated in the absence of the antibiotic (PDT of **7.66 h** and peak growth achieved after 72 h or 3 days). However, the growth maximum is slightly lower (6x10<sup>6</sup> cells/mL (+tet) vs ~ 8x10<sup>6</sup> cells/mL (-tet)).

NYSM *Tbb\_WT* strains have a tet operator inserted into the robust procyclic acid repeat protein (PARP) promoter, as described in the paper (131). Consequently, the NYSM *Tbb\_WT* strains exhibit tetracycline inducer-dependent expression of reporter genes that are integrated into the chromosome under the regulatory influence of the PARP promoter containing a tet operator (131). Therefore, the difference can be explained by the fact that the addition of tetracycline could induce growth factors that would explain the results obtained. Indeed, it has been shown in some papers that in case of loss of its antibiotic activity in a time-dependent manner, tetracycline can still act as a potent inducer, therefore, if certain growth factors are under the control of the antibiotic, their expression could possibly be induced by its addition in the culture medium which could explain the rapid growth achieved by the parasites (178).

*.3.5.2 Growth curves of transfected Tbb with and without tetracycline addition*

As shown in Figure 36, all the different cell lines reach their growth maximum on day 3 (after 72 hours of incubation), except for the *Tbb\_WT* strain under tetracycline induction, which shows a growth maximum after only 48 hours due to tetracycline-induced growth factors as mentioned above.



|                             | Population doubling time (PDT)<br>(h) (-tet) | Population doubling time<br>(PDT) (h) (+tet) |
|-----------------------------|----------------------------------------------|----------------------------------------------|
| <i>Tbb WT</i>               | 7.66                                         | 5.71                                         |
| <i>Tbb BioID-SNAP</i>       | 10.66                                        | 14.19                                        |
| <i>Tbb TurboID-SNAP</i>     | 10.80                                        | 11.11                                        |
| <i>Tbb miniTurboID-SNAP</i> | 9.61                                         | 12.99                                        |

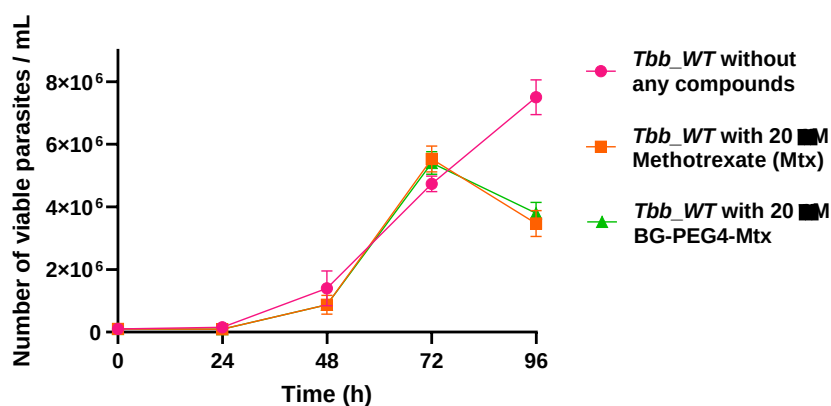
*Figure 36: Growth curves and PDT of different transfected Tbb in the presence or absence of tetracycline (1µg/mL)*

Firstly, it can be noted that tetracycline induction of the expression of the protein of interest does not lead to parasite death, validating the cellular non-toxicity of the different expressed tools. Secondly, for most strains, the maximum growth is reached either on day 2 (with tetracycline) or day 3 (without tetracycline) (except for the *Tbb\_mTGO* strain where the peak is reached on day 3 with and without tetracycline), indicating that subsequent biotinylation studies should be performed within this 3-day window. Finally, the number of transfected viable parasites reaching the growth peak is 2 times lower than that of *Tbb\_WT*, indicating that transfection of the three different genes would influence parasite growth, nevertheless, this number remains sufficient to carry out the necessary target deconvolution experiments.

Also, what can be noticed for parasites expressing the cMyc\_TBirA\*\_GGS\_SNAP-tag gene (*Tbb\_TG0* cell line), is their maximum growth that occurs on day 3 (after 72h) without tetracycline and on day 2 (after 48h) with the presence of the antibiotic just as the *Tbb\_WT*.

To answer the question on how compounds treatment would influence the setup of the experiments, we studied the *Tbb* growth in presence of different compounds. Figure 37 reports the results of *Tbb* growth in presence of Mtx or BG-PEG<sub>4</sub>-Mtx. The Superposition of the growth curves of the parasites that were incubated with either Mtx or BG-PEG<sub>4</sub>-Mtx, is close-to-perfect. From this result two conclusions can be drawn:

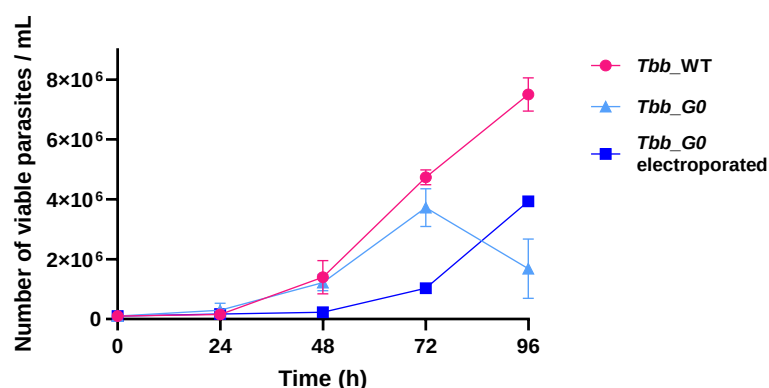
- 1- The efficacy of methotrexate is not impaired by its binding to the BG-PEG<sub>4</sub> moiety indicating that the BG-PEG<sub>4</sub> fragment is unlikely to affect the behaviour of future compounds to be bound to it.
- 2- With concentrations of 20  $\mu$ M (and even higher, indeed, experiments have been performed with 50  $\mu$ M or even 100  $\mu$ M but the corresponding results have not been presented) the parasites remain viable, and their growth starts to decrease only after 72h which is normal and largely sufficient to perform the later experiments of biotinylation.



|                                                     | Population doubling time (PDT) (h) |
|-----------------------------------------------------|------------------------------------|
| <i>Tbb_WT</i>                                       | 7.66                               |
| <i>Tbb_WT</i> + 20 $\mu$ M Mtx                      | 10.28                              |
| <i>Tbb_WT</i> + 20 $\mu$ M BG-PEG <sub>4</sub> -Mtx | 9.13                               |

Figure 37: Growth curves and PDT of *Tbb\_WT* in the presence of either methotrexate (Mtx) or BG-PEG<sub>4</sub>-Mtx or without any compound

In the case of dealing with compounds that are not cell-permeable, what would be the solution? A possible answer would be to electroporate the parasites in the presence of the non-permeable BG-PEG<sub>4</sub> derivative to force the entry of the molecule into the parasite. The question that may arise is whether or not this technique could be fatal to the parasites or if it affects their viability. To answer this question, a growth curve (Figure 38) was performed for the electroporated parasites and compared to the non-electroporated ones to detect any differences and verify cell viability.

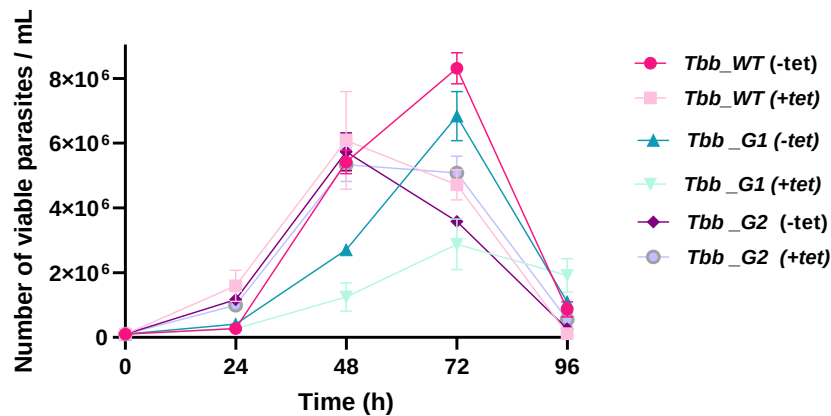


|                              | Population doubling time (PDT) (h) |
|------------------------------|------------------------------------|
| <i>Tbb_WT</i>                | 7.66                               |
| <i>Tbb_G0</i>                | 10.66                              |
| <i>Tbb_G0 electroporated</i> | 15.04                              |

*Figure 38: Growth curves and PDT of either electroporated or non-electroporated Tbb\_G0 (without any compounds)*

According to the results presented in Figure 38, the electroporation technique is not fatal for the parasites, but it shifts the growth peak by 24 hours. This slight alteration will not impact future experiments significantly, as it simply requires extending the waiting period until a sufficient amount of parasites reaches the logarithmic growth phase (refer to section 2.2.18). Electroporation could, therefore, according to the results obtained, be a very good alternative to the situation where non-permeable compounds are involved.

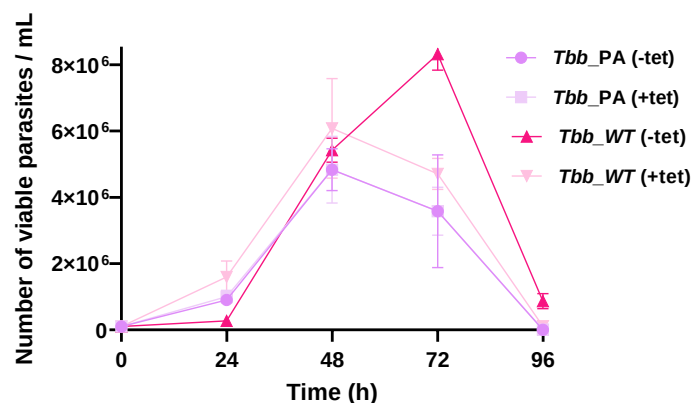
Next, growth curves were generated for the parasites transfected with plasmids containing various linkers. The results are presented in Figure 39 below:



|               | Population doubling time (PDT)<br>(h) (-tet) | Population doubling time<br>(PDT) (h) (+tet) |
|---------------|----------------------------------------------|----------------------------------------------|
| <i>Tbb_WT</i> | 8.56                                         | 5.71                                         |
| <i>Tbb_G1</i> | 8.20                                         | 15.23                                        |
| <i>Tbb_G2</i> | 6.65                                         | 8.00                                         |

**Figure 39:** Growth curves and PDT of the two transfected strains *Tbb\_G1* (*Tbb* transfected with the *cMyc\_BirA\*\_GGGS\_SNAP-tag* gene) & *Tbb\_G2* (*Tbb* transfected with the *cMyc\_BirA\*\_GGGS\_SNAP-tag* gene) in the presence or absence of tetracycline (1 µg/mL) and their comparison with those of *Tbb\_WT* parasites

According to the result of Figure 39, the addition of the antibiotic seems to influence the PDT of strains containing the linker **G1** (GGGS linker), which increases from 8.20 h to 15.23 h. Overexpression of the *cMyc\_BirA\*\_GGGS\_SNAP-tag* fusion protein slows down the growth of the parasite, but they remain viable even after 96 h of expression which is a sufficient time interval to perform the subsequent biotinylation studies. On the other hand, strains transfected with the plasmid containing the linker **G2** (GGGGS linker) do not seem to be much influenced by the addition of tetracycline compared to *Tbb\_G1* parasites. Indeed, their PDT increases slightly (from 6.65 h to 8.00 h) and their growth curves are fairly similar with or without tetracycline. Several hypotheses can be formulated: either the protein of interest was not or only slightly expressed (confirmation of the hypothesis by western blot, results in Section 3.6.5), or the gene of interest was not properly integrated into the genome of the parasite, but this last hypothesis can be refuted given the results of the sequencing of the gDNA showing the correct integration of the fusion protein gene in the parasite genome. The first hypothesis of under-expression of the fusion protein is likely to be accurate, as it will be confirmed by western blot analysis in the time-dependent protein expression studies (see the next results below, Section 3.6.5).



|               | Population doubling time (PDT)<br>(h) (-tet) | Population doubling time (PDT)<br>(h) (+tet) |
|---------------|----------------------------------------------|----------------------------------------------|
| <i>Tbb_WT</i> | 8.56                                         | 5.71                                         |
| <i>Tbb_PA</i> | 6.16                                         | 6.50                                         |

**Figure 40:** Growth curves and PDT of the transfected *Tbb\_PA* (*Tbb\_BirA\**\_PAPAP\_SNAP-tag parasites) in the presence or absence of tetracycline (1 µg/mL) and their comparison with those of *Tbb\_WT* parasites

The newly transfected *Tbb\_BirA\**\_PAPAP\_SNAP-tag cell line, renamed *Tbb\_PA*, surprisingly shows overlapped growth curves between that with and without the addition of tetracycline (Figure 40 above), besides their very similar PDT. Hence, the antibiotic does not seem to affect their growth or PDT. To explain this result, one hypothesis that can be put forward is that the tet operator (whose gene is positioned upstream of the gene of interest) would not be well integrated into the genome, and as a result of that, the gene of interest is constitutively expressed regardless of the addition of the tetracycline.

Another explanation that could be given regarding the similarities in the growth curves, would be that the tool is not overexpressed in large quantities in the cytosol of the parasite. This hypothesis has been tested in the time-dependent protein overexpression studies (see Section .3.6) as well as in the biotinylation studies below in Section .3.7.

Indeed, only 2% of the protein is overexpressed after 24 h (results below Section 3.6.6) against 20% for cMyc\_BirA\*\_GGG\_SNAP-tag, which is 10 times less for the newly designed protein. A study published in 2016 (179) revealed that trypanosomatids do not appear to be highly affected by tetracycline. This can be attributed to the highly divergent nature of their mitochondrial ribosomes. As a result, the impact of tetracycline may not be noticeable when

the protein of interest is expressed in very low amounts. This could explain why similar growth curves were obtained with or without the antibiotic, as shown in Figure 40 above.

Moreover, if we compare the growth curves of *Tbb\_G2* (*Tbb\_BirA\**\_GGGS\_SNAP-tag parasites), with that of *Tbb\_PA* (*Tbb\_BirA\**\_PAPAP\_SNAP-tag parasites), these two strains carry a linker with the same number of amino acids, both strains seem to exhibit similar curves and do not appear to be overly influenced by tetracycline, so this gives rise to an interesting new hypothesis: the more amino acids there are in the linker, the less the fusion protein will be expressed. This hypothesis will be confirmed in the time-dependent overexpression studies below (Section .3.6).

A third hypothesis could be related to constitutive expression without the need for induction by the antibiotic due to a leaky system.

Since the concentration of 1 µg/ml of tetracycline is considered a standard induction dose that does not interfere with *T.brucei* cell division (179), a 10-fold higher concentration has been used in order to verify one of the three hypotheses previously put forward (Figure 41 below).

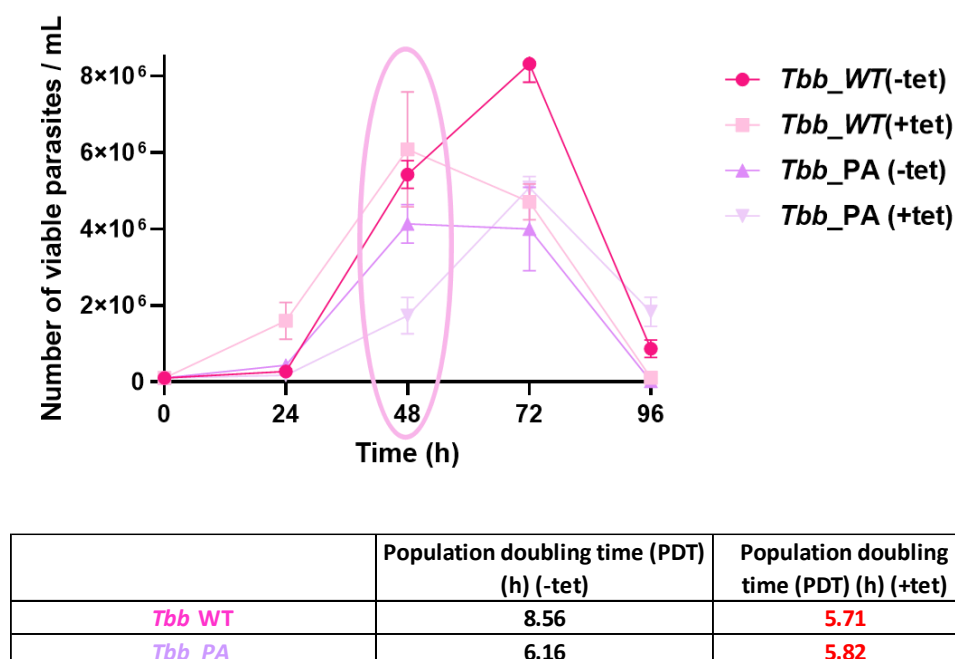
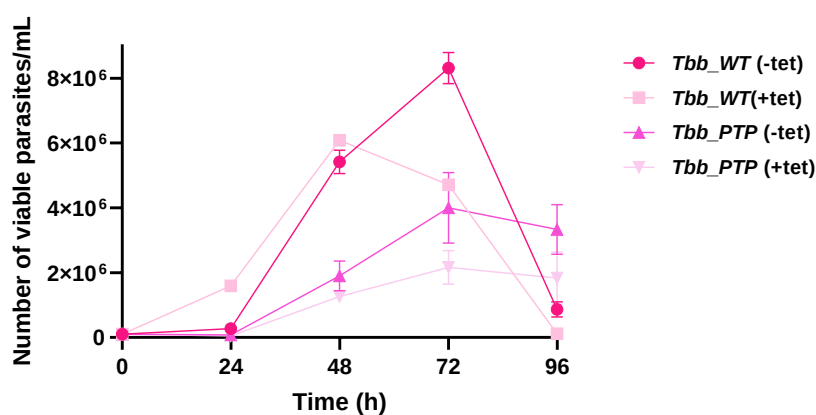


**Figure 41:** Growth curves and PDT of the transfected parasites *Tbb\_PA* in the presence or absence of tetracycline (10 µg/mL) and their comparison with those of the *Tbb\_WT* parasites

Compared to the growth curves of the wild-type strains, the maximum growth rate achieved by the transfected strains (*Tbb\_PA*) with the PAPAP rigid linker remains lower than that of the *Tbb\_WT* strains, showing that the integration of the gene of interest and its subsequent expression, however low, still influences the growth of the parasite.

Moreover, when comparing the growth curve obtained using 1 µg/mL antibiotic, a noticeable distinction is observed. The maximum growth for *Tbb\_PA* (+tet) occurs after 72 hours instead of 48 hours before declining. Nonetheless, similar to *Tbb\_WT*, the parasite doubling time (PDT) is shorter when the antibiotic is introduced into the growth medium. This suggests that tetracycline potentially triggers growth-promoting factors that stimulate accelerated parasite proliferation.

The influence of the stable transfection with the construct with the linker PTP on parasite growth has also been studied. As illustrated in Figure 42, the maximum expression of *Tbb\_PTP* is reached at about  $4 \times 10^6$  cells/mL, half that of *Tbb\_WT* (as with most other transfected strains); transfection of the cMyc\_BirA\*\_PTP\_SNAP-tag gene and its integration into the parasite genome seem to influence the attainment of the growth maximum, but the latter is still reached after 3 days, which is similar to that of *Tbb\_WT* (-tet).



|                | Population doubling time (PDT) (h) (-tet) | Population doubling time (PDT) (h) (+tet) |
|----------------|-------------------------------------------|-------------------------------------------|
| <i>Tbb_WT</i>  | 8.56                                      | 5.71                                      |
| <i>Tbb_PTP</i> | 9.52                                      | 10.08                                     |

**Figure 42:** Growth curves and PDT of the transfected parasites *Tbb\_PTP* (*Tbb\_BirA\*\_PTP\_SNAP-tag* parasites) in the presence or absence of tetracycline (1µg/mL) and their comparison with those of the *Tbb\_WT* parasites

Moreover, the parasites remain viable after 4 days, which represents a large and sufficient time interval to perform all subsequent biotinylation studies.

Regarding the influence of tetracycline on parasite growth, it does have a slight impact, as growth is slowed and peaks at about  $2 \times 10^6$  cells/mL (half that of *Tbb\_PTP* (-tet)), but as before, the parasites remain viable even after 4 days of induction of protein expression by the antibiotic, so the cMyc\_BirA\*\_PTP\_SNAP-tag fusion protein does not seem to be toxic to the parasite.

### **.3.6 Time-dependant protein expression**

In order to assess expression, potential degradation and determine the optimal timing of expression for different fusion proteins, a comprehensive investigation was conducted to study the temporal expression profiles of these tools. This studies presented in this chapter aimed at facilitating protocol adaptation based on the peak expression period for each fusion protein.

#### *.3.6.1 Time-dependant cMyc\_BirA\*\_GGS\_SNAP-tag protein expression*

The first comprehensive study to assess tool expression, potential degradation and identify the optimal timing of expression was carried out with cMyc\_BirA\*\_GGS\_SNAP-tag fusion protein. The membrane stained with Red S Ponceau solution (part A of Figure 43 below) shows the proper transfer of proteins from the gel to the membrane. The scanned membrane (part B of Figure 43) reveals the overexpression of the cMyc\_BirA\*\_GGS\_SNAP-tag protein from day 1 to day 4 (96 h). The results firstly demonstrate an equal expression of the housekeeping protein *TbBIP* (~70 kDa), which can therefore be used to normalize and quantify the results (red bands revealed by anti-*TbBIP* antibodies).

Second, what is clearly visible is the successful expression of the fusion protein as shown by the green bands on the membrane (part B of Figure 43) revealed by the anti-cMyc antibodies. The proper expression was also confirmed and validated by MS (result in Figure 44 below). Part B of Figure 43 above displays a strong band intensity on day 1 (after 24 h of overexpression), indicating that maximum overexpression occurs after 24 h of tetracycline induction. Part C of the above figure, exhibiting the intensity of the signal emitted by cMyc-

tag (revealed by the anti-cMyc antibodies) over that emitted by *TbBIP* (revealed by anti-*TbBIP* antibodies), also supports the same conclusion.

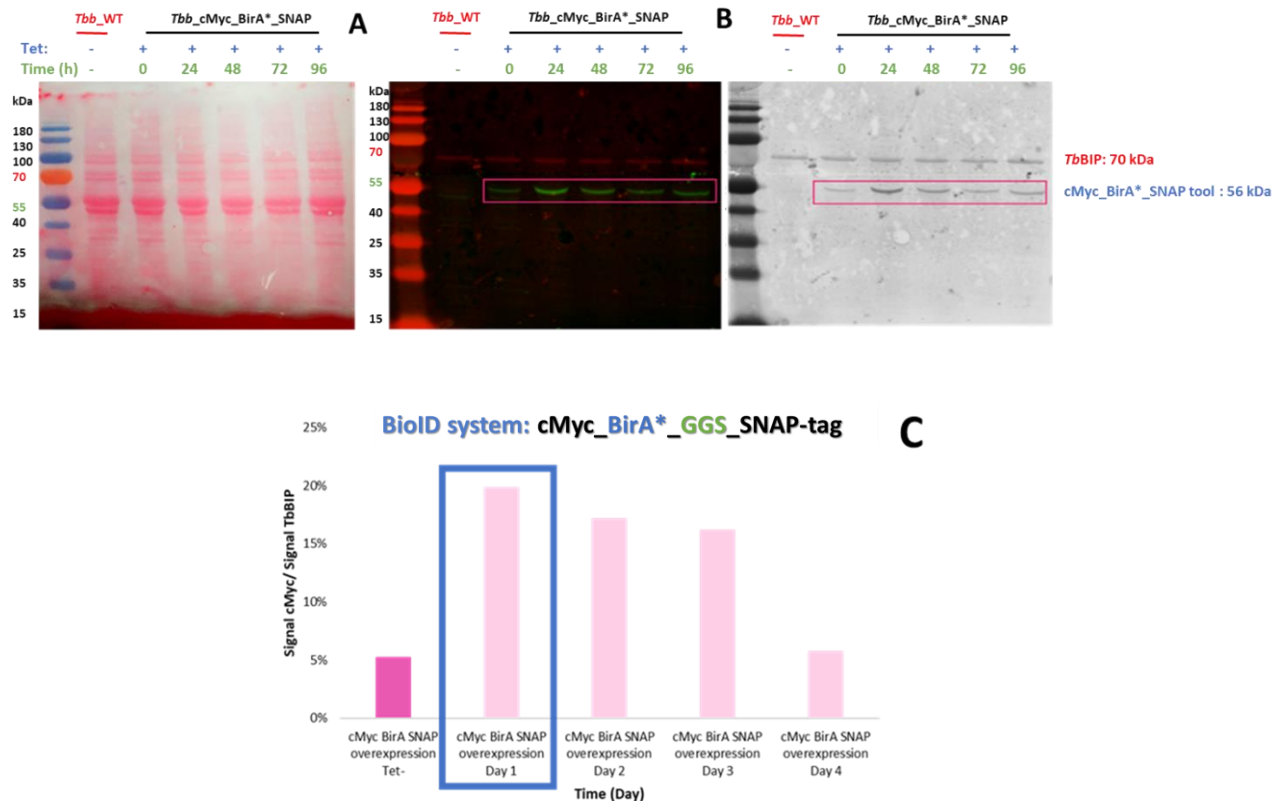


**Figure 43:** Western blot of *cMyc\_BirA\*\_GGG\_SNAP-tag* protein expression as a function of time over 96 h of tetracycline induction. A: membrane (0.22  $\mu$ m) stained with red Ponceau S solution after protein transfer from SDS PAGE gel (12%) for 24 h at 30 V. B: membrane scanned by the LiCor Odyssey imager at two different wavelengths: 700 nm (red) and 800 nm (green) (part B, left image). C: Quantification of the signal emits from the *cMyc-tag* part of the protein over the one from the *TbBIP* housekeeping protein

Once the maximum level of expression is achieved following 24 hours of incubation with tetracycline, the level of expression begins to decrease. Therefore, in order to conduct biotinylation studies, tetracycline should be added 24 hours prior to the pull-down process (simultaneously with biotin addition as detailed in Section .3.7 below). For the BG-PEG<sub>4</sub> derivative, the experiments indicate that it is necessary to add it either at the time of the antibiotic addition or 24 hours later when the expression of the fusion protein is at its peak. Although the results remain similar, it is advisable to add the derivative at the same time as

the tetracycline addition in order to avoid any potential decline in the functionality of the tool due to a reduction in its expression. The viability of parasites was shown to persist even after 96 hours of incubation, albeit with a slight decline, in the presence of 20  $\mu$ M BG-PEG<sub>4</sub>-Mtx, as evidenced by Figure 37 in the section above. Therefore, introducing the compound at the start of the parasite culture could also be a feasible option.

Finally, the protein visible at ~55-60 kDa on the membrane stained with Red Ponceau solution is an HSP60, a heat shock protein belonging to the HSP60/10 chaperonin systems required for parasite development and survival (180).

As mentioned above, the validation of the expression of the tool was verified and confirmed by tryptic digestion followed by MS (Figure 44 below). The LC-MS/MS analysis revealed the presence of cMyc\_BirA\*\_GGG\_SNAP-tag peptides, with a total spectrum count of 28 and 22 unique peptides. This resulted in a protein sequence coverage of 59%, as demonstrated by the analysis in the Figure 44.

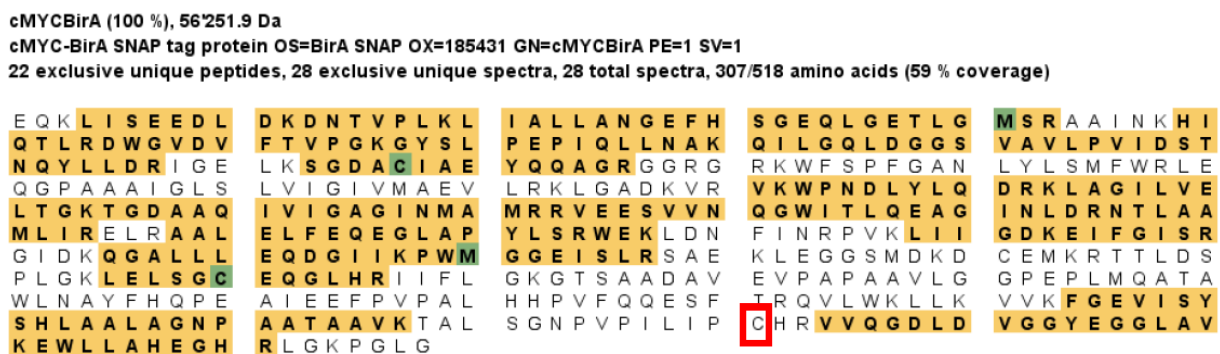


**Figure 44:** Identification in the cell lysate of cMyc\_BirA\*\_GGG\_SNAP-tag peptides by MS (analyses made by the proteomic platform of Dr. Alexandre Hainard and his team)

### .3.6.2 Time-dependant cMyc\_TBirA\*\_GGG\_SNAP-tag protein expression

The second study assessing tool expression, potential degradation and identify the optimal timing of expression was carried out for cMyc\_TBirA\*\_GGG\_SNAP-tag fusion protein. Part A of Figure 45 shows good protein transfer from the gel to the nitrocellulose membrane. Western blot analysis (part B of Figure 45) shows that the house protein *TbBIB* was expressed correctly and similarly, irrespective of time and tetracycline addition.

It is intriguing to note that the expression of cMyc\_TBirA\*\_GGG\_SNAP-tag is initially quite low, with only 8% of maximum expression visible, and it takes two days of incubation with tetracycline before significant expression can be observed. The expression reaches its peak on the third day, as depicted in parts B and C of the figure above, before beginning to decline on subsequent days. Based on the findings, it is recommended to add the BG-PEG<sub>4</sub> derivative either after 72 hours of induction of tool expression, which is three days after the addition of tetracycline (with the pull-down process carried out at this time, as discussed in section .3.6), or simultaneously with the addition of tetracycline or even at the beginning of the parasite culture (this latter option was selected for the majority of the experiments).

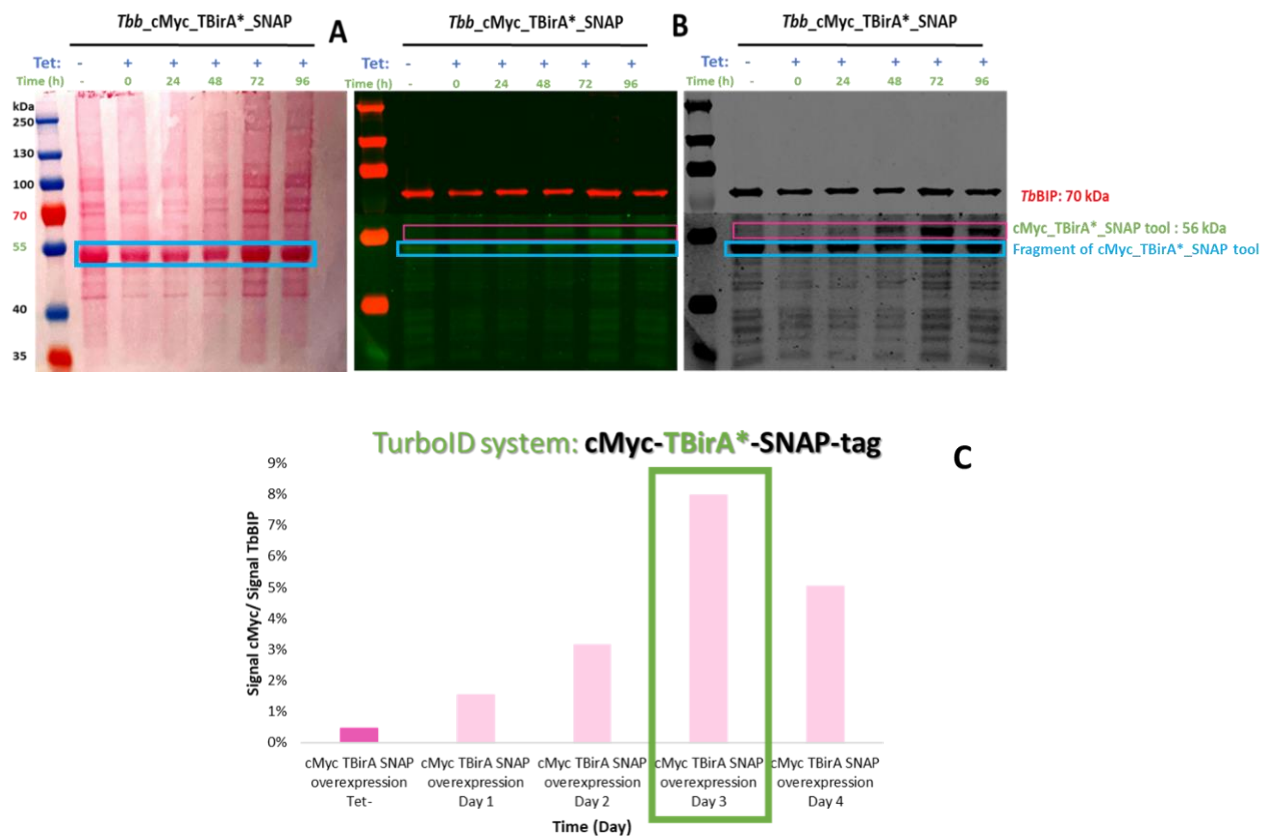


**Figure 45:** Western blot of cMyc\_TBirA\*\_GGG\_SNAP-tag protein expression as a function of time over 96 h of tetracycline induction. **A:** membrane (0.22  $\mu$ m) stained with red Ponceau S stain after protein transfer from SDS PAGE gel (10%) for 24 h at 30 V. **B:** membrane scanned by the LiCor Odyssey imager at two different wavelengths: 700 nm (red) and 800 nm (green) (part B, left image). **C:** Quantification of the signal emits from the cMyc-tag part of the protein over the one from the TbBIP housekeeping protein

What is also evident is a second signal observable just below 55 kDa (Figure 45), and this band is detected regardless of tetracycline addition. This signal could correspond to a fragment of the tool that was co-expressed constitutively, irrespective of antibiotic induction, as a result of the T7 polymerase promoter system, for which leakage is often observed (181, 182).

The intense bands just below 55 kDa visible on the ponceau red stained membrane correspond to chaperone proteins or heat shock proteins (HSPs) and fragments of the tool of interest

Additional validation of the presence of the tool was performed and confirmed by MS (Figure 46 below). Despite the relatively low overall coverage 34% the protein could be clearly identified as the detected peptides are well spread all over the sequence (Figure 46).

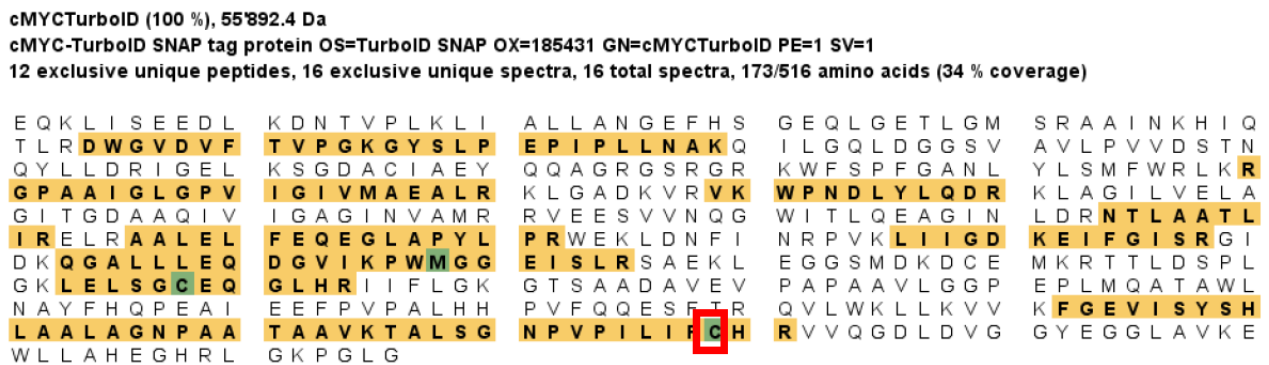


*Figure 46: Identification in the cell lysate of cMyc\_TBirA\*\_GGs\_SNAP-tag peptides by MS (analyses made by the proteomic platform of Dr. Alexandre Hainard and his team)*

### **.3.6.3 Time-dependant cMyc\_mTBirA\*\_GGs\_SNAP-tag protein expression**

As previously, a time-dependant study was conducted to evaluate protein expression, degradation, and determine the optimal expression timing for the cMyc\_mTBirA\*\_GGs\_SNAP-tag fusion protein. The objective was to identify the peak expression period, enabling protocol adaptation as needed. Also in this case like the other constructs, the housekeeping protein *TbBIP* was also here correctly and similarly expressed regardless of time and tetracycline addition (Figure 47).

For the cMyc\_mTBirA\*\_GGs\_SNAP-tag (miniTurboID\_SNAP system), its maximum expression is reached at day 3 (parts B and C of Figure 47) as is that of the cMyc\_TBirA\*\_SNAP-tag fusion protein (TurboID\_SNAP system). These results are similar to those obtained with

cMyc\_TBirA\*\_SNAP-tag. Certainly, the addition of the BG-PEG<sub>4</sub> derivative can be carried out either simultaneously with the addition of tetracycline or at the outset of the culture of the parasites.

After 72h of overexpression, the expression starts to decrease as for the cMyc\_TBirA\*\_SNAP-tag fusion protein, indicating that the pull-down should be performed within this 4-day window.

Low expression of the cMyc\_mTBirA\*\_SNAP-tag protein, as well as that of cMyc\_TBirA\*\_SNAP-tag protein, induced without the addition of tetracycline, can also be observed, although its level is lower than when the antibiotic is added, this is still quite visible. This low constitutive expression can be explained by the fact that as mentioned earlier, expression systems involving T7 polymerase promoters, commonly used to induce recombinant proteins, often manifest an expected leakage of expression (181, 182).

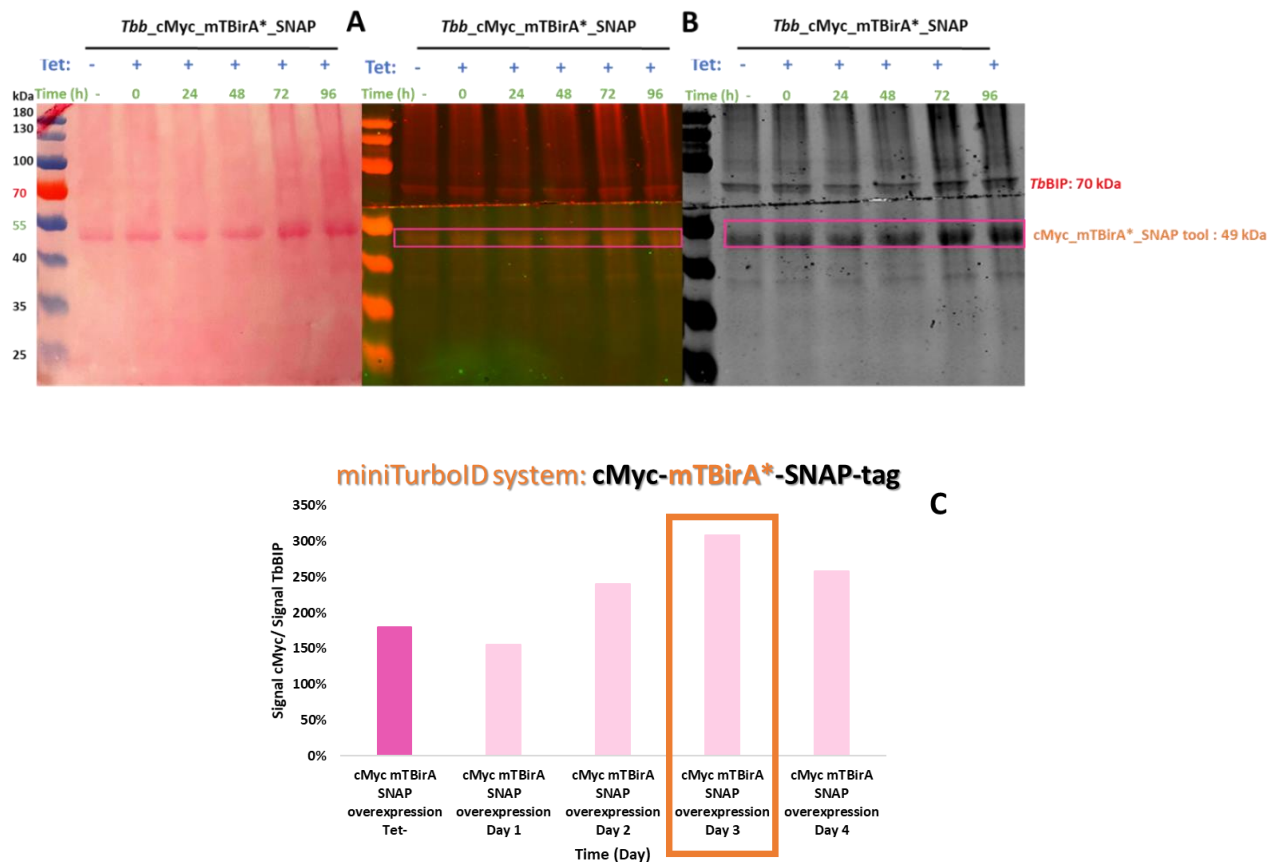


**Figure 47:** Western blot of cMyc\_mTBirA\*\_SNAP-tag protein expression as a function of time over 96 h of tetracycline induction. **A:** membrane (0.22  $\mu$ m) stained with red Ponceau S stain after protein transfer from SDS PAGE gel (12%) for 24 h at 30 V. **B:** membrane scanned by the LiCor Odyssey imager at two different wavelengths: 700 nm (red) and 800 nm (green) (part B, left image). **C:** Quantification of the signal emits from the cMyc-tag part of the protein over the one from the TbbP housekeeping protein

The maximum expression of the cMyc\_TBirA\*\_GGG\_SNP-tag fusion protein is 8% while for the cMyc\_mTBirA\*\_GGG\_SNP-tag fusion protein, its maximum expression reaches 300%. Indeed, in a study conducted on mammalian cells, worms, and flies, a certain toxicity of the TurboID system ligase has been reported on these different hosts (116, 119). Moreover, when the ligase was expressed, a decrease in size and viability of the flies was observed suggesting that the TurboID system would consume all the endogenous and surrounding biotin thus depriving the cells of this compound which leads to the apparent toxicity (119). On the western blot membrane, apparent fragments of the fusion protein cMyc\_TBirA\*\_GGG\_SNP-tag are observed (Figure 45 above, light blue framed fragments). As these were revealed by

anti-cMyc antibodies, these fragments, therefore, contain the cMyc part and probably also the TBirA\* part, given the molecular weight of the protein. The presence of these fragments confirms the instability of the TurboID system observed by some studies (119) and could also explain its low expression (only 8%).

#### .3.6.4 Time-dependant cMyc\_BirA\*\_GGGS\_SNAP tag protein expression

As before, a time-dependant study was conducted to assess protein expression, degradation, and determine the optimal expression timing for the cMyc\_BirA\*\_GGGS\_SNAP-tag fusion protein. The aim was to identify the peak expression period, facilitating protocol adaptation. Based on the results shown in Figure 48, maximum overexpression of cMyc\_BirA\*\_GGGS\_SNAP tag protein was reached after 24 hours, equal to that of the original cMyc-BirA\*-GGGS-SNAP-tag fusion protein. Thus, increasing the linker length to four amino acids does not appear to influence the moment of maximal protein overexpression.

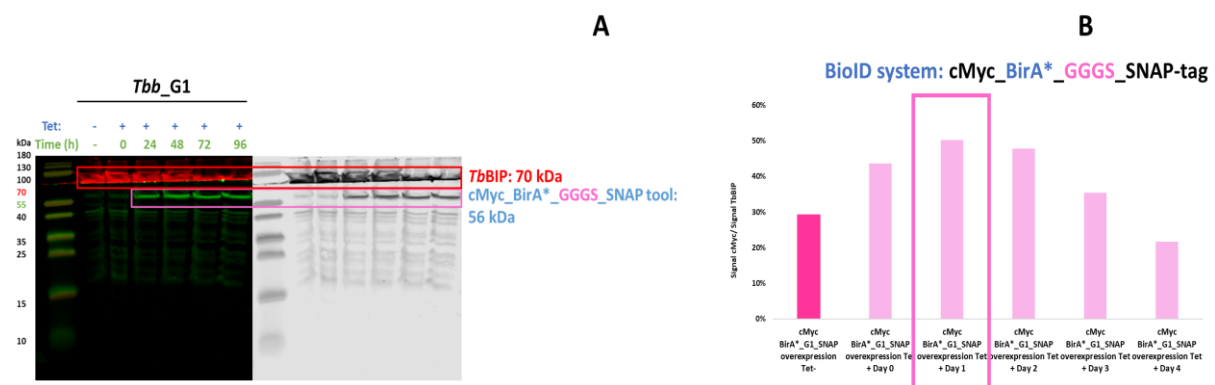


**Figure 48:** Western blot of cMyc\_BirA\*\_GGGS\_SNAP-tag protein expression as a function of time over 96 h of tetracycline induction. **A:** membrane (0.22  $\mu$ m) stained with red Ponceau S stain after protein transfer from SDS PAGE gel (10%) for 24 h at 30V. **B:** membrane scanned by the LiCor Odyssey imager at two different wavelengths: 700 nm (red) and 800 nm (green) (part B, left image). **C:** Quantification of the signal emits from the cMyc-tag part of the protein over the one from the TbbBIP housekeeping protein

These results indicates that the BG-PEG<sub>4</sub> derivative can be added either simultaneously with the tetracycline or at the outset of the culture of the parasites.

#### .3.6.5 Time-dependant cMyc\_BirA\*\_GGGGS\_SNAP-tag protein expression

As for the other tools, we determined the optimal expression time for the cMyc\_BirA\*\_GGGGS\_SNAP-tag fusion protein, enabling protocol adaptation based on the

peak expression period. Based on the findings presented in Figure 49, it appears that increasing the linker length to 5 amino acids causes a delay in maximal overexpression by 24 hours. The optimal expression level is achieved on day 2, which is 42 hours after induction by tetracycline. Additionally, the expression of the protein is comparatively low; only 3% of the tool of interest was expressed compared to 20% for the tool carrying the original GGS linker and 50% for the tool carrying the G1 linker. It is hypothesized that increasing the linker length beyond 4 amino acids impairs protein expression.

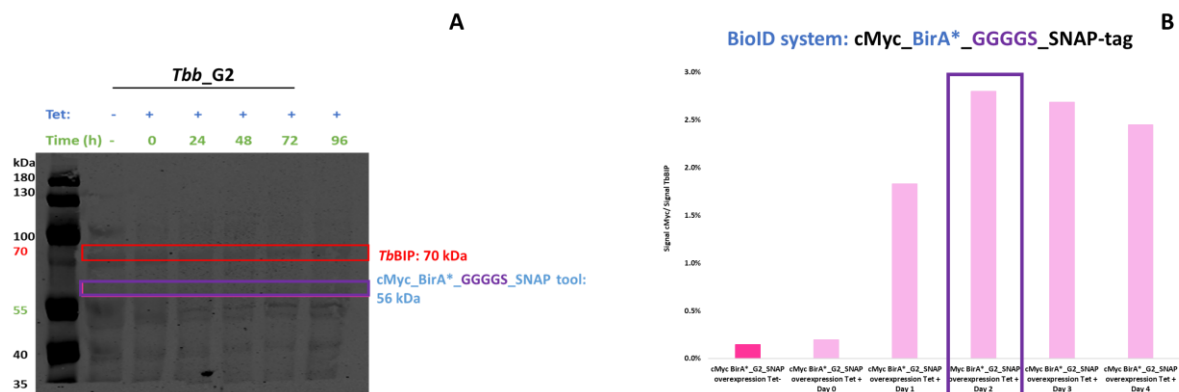


**Figure 49:** Western blot of cMyc\_BirA\*\_GGGGS\_SNAP-tag protein expression as a function of time over 96 h of tetracycline induction. **A:** membrane (0.22  $\mu$ m) stained with red Ponceau S stain after protein transfer from SDS PAGE gel (10%) for 24 h at 30V. **B:** membrane scanned by the LiCor Odyssey imager at two different wavelengths: 700 nm (red) and 800 nm (green) (part B, left image). **C:** Quantification of the signal emits from the cMyc-tag part of the protein over the one from the TbBIP housekeeping protein

The addition of BG-PEG<sub>4</sub> derivative will be added either simultaneously with tetracycline or at the outset of the culture of the parasites.

### .3.6.6 Time-dependant cMyc\_BirA\*\_PAPAP\_SNAP tag protein expression

Similarly, to the other constructs the ability of the transfected parasites to express cMyc\_BirA\_PAPAP\_SNAP-tag tool has been assessed as a function of time. After 24 hours, maximal overexpression is achieved (Figure 50), similar to the original cMyc\_BirA\_GGS\_SNAP-tag fusion protein system. However, the maximum expression is only 2%, which is 10 times less than that of the cMyc\_BirA\_GGS\_SNAP-tag (20%). This indicates that the linker PA, composed of five amino acids, yields similar results to the linker G2, which also has five amino acids (maximum expression only 3%), and confirms the hypothesis presented in Section .3.5.2

during the establishment of growth curves for transfected parasites. In this study, similarities were observed between the growth curves obtained with and without tetracycline for the two strains *Tbb\_G2* and *Tbb\_PA*, both of which contain linkers with the same number of amino acids. Therefore, increasing the number of amino acids beyond four would impede the expression of the tool of interest, regardless of its nature. These findings have important implications for biotinylation studies, as shown in the results below.

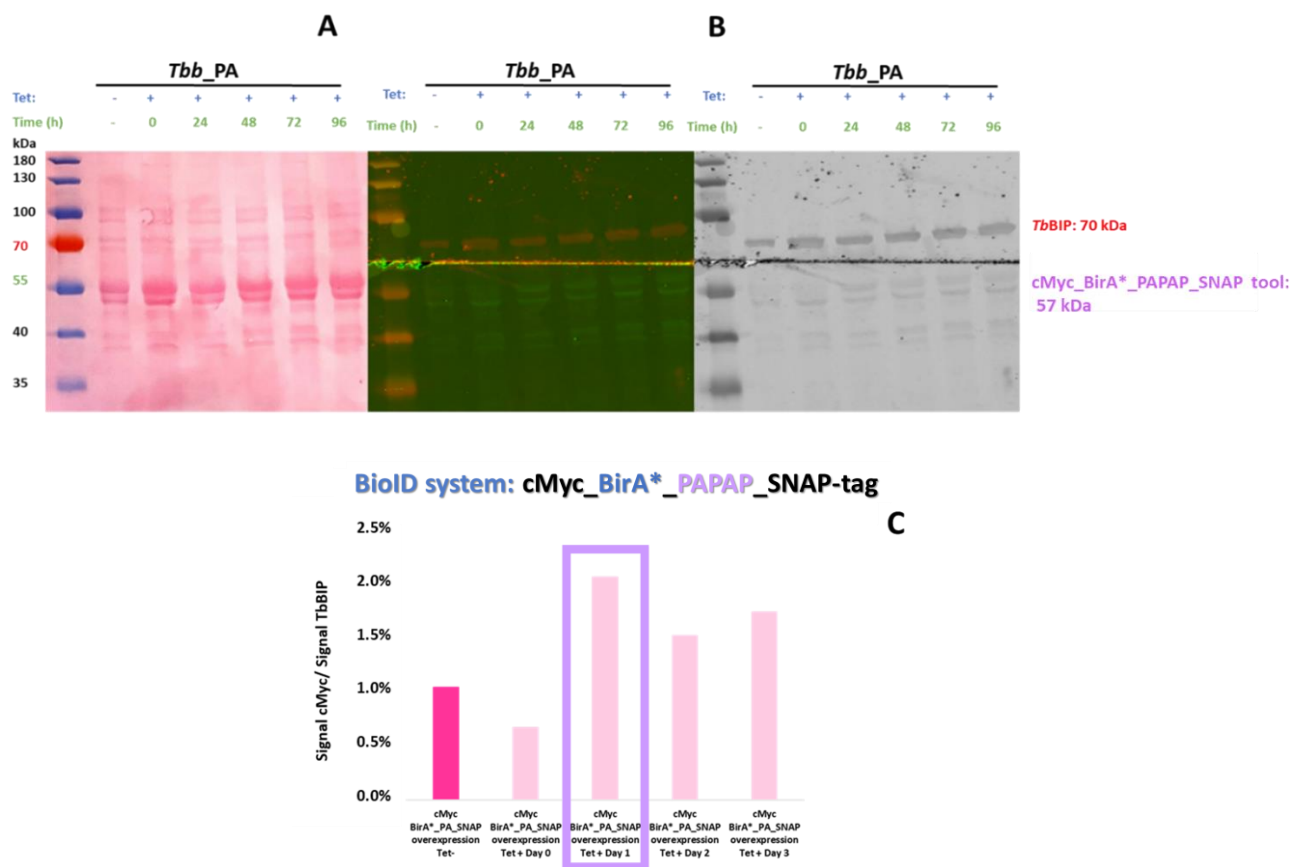


**Figure 50:** Western blot of cMyc\_BirA\*\_PAPAP\_SNAP-tag protein expression as a function of time over 96 h of tetracycline induction. **A:** membrane (0.22  $\mu$ m) stained with red Ponceau S stain after protein transfer from SDS PAGE gel (10%) for 24 h at 30 V. **B:** membrane scanned by the LiCor Odyssey imager at two different wavelengths: 700 nm (red) and 800 nm (green) (part B, left image). **C:** Quantification of the signal emits from the cMyc-tag part of the protein over the one from the TbBIP housekeeping protein

The expression of the tool was also verified and confirmed by the MS (results below in Section .3.7.1). Regarding the original tool containing the linker G0, the addition of the BG-PEG<sub>4</sub> derivative will be performed either simultaneously with the tetracycline or at the beginning of the parasite culture.

*.3.6.7 Time-dependant cMyc\_BirA\*\_PTP\_SNAP tag protein expression*

As described above, the ability of transfected parasites to express the cMyc\_BirA\_PTP\_SNAP-tag tool has been evaluated as a function of time. The tool reaches its highest expression level (>20%) after 24 hours of induction under tetracycline, followed by a gradual decrease in expression (Figure 51). After 96 hours of induction, the expression slightly increases again. It is worth noting that the bands observed on the membrane for the 24 and 48-hour expression conditions are very weak. This is likely due to sample loss during gel loading. However, this issue was addressed by normalizing the obtained signals. As for the original tool, the BG-PEG<sub>4</sub> derivative can be added either simultaneously with tetracycline or at the beginning of the parasite culture.

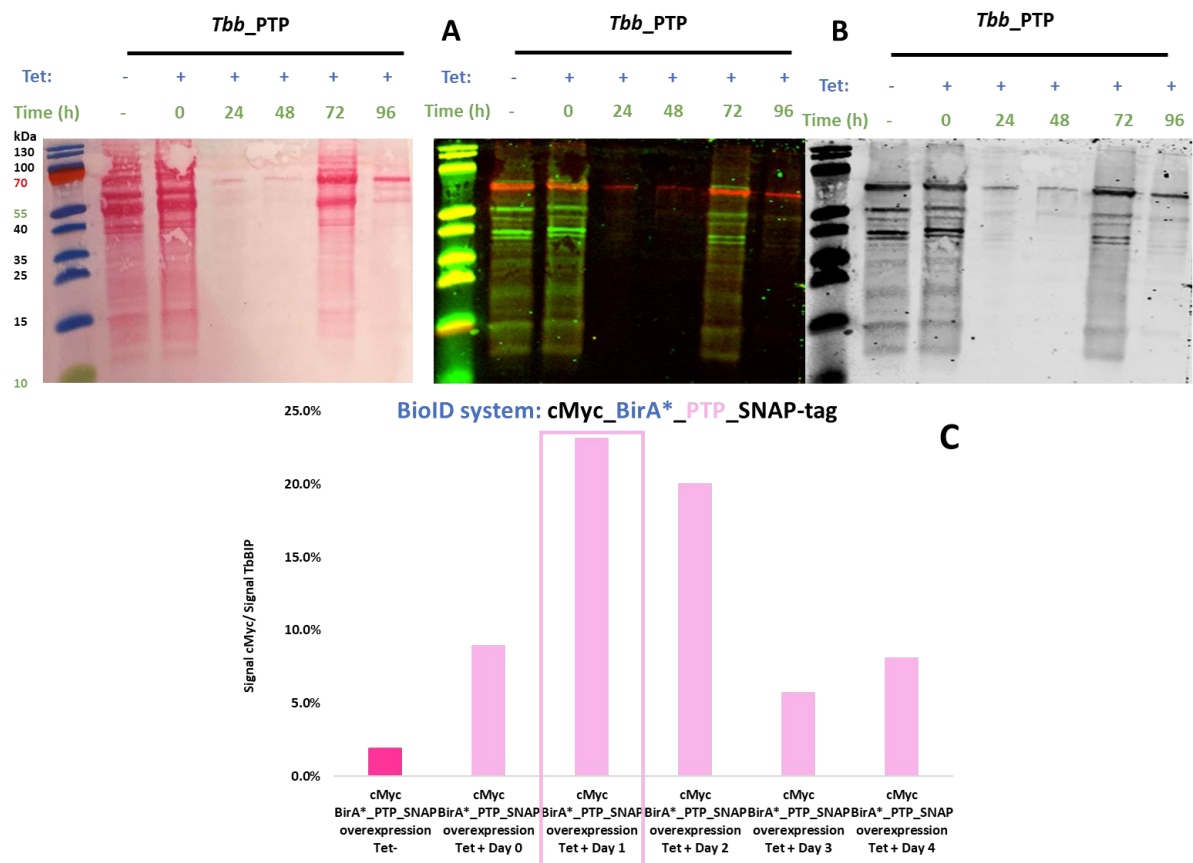


**Figure 51:** Western blot of cMyc\_BirA\*\_PTP\_SNAP-tag protein expression as a function of time over 96 h of tetracycline induction. **A:** membrane (0.22  $\mu$ m) stained with red Ponceau S stain after protein transfer from SDS PAGE gel (10%) for 24 h at 30 V. **B:** membrane scanned by the LiCor Odyssey imager at two different wavelengths: 700 nm (red) and 800 nm (green) (part B, left image). **C:** Quantification of the signal emits from the cMyc-tag part of the protein over the one from the TbbIP housekeeping protein

### **.3.7 Biotinylation studies**

The biotinylation studies have been performed aiming at identifying DHFR-TS as the first target to reach the proof of concept of the new deconvolution tool.

The various fractions collected following the protocol outlined in Section 2.2.25 and depicted in Figure 19 were subjected to western blot analysis. The findings regarding the capacity of the different tools to capture the DHFR-TS target of interest are presented below.

#### *.3.7.1 Biotinylation of DHFR-TS by cMyc\_BirA\*\_GGS\_SNAP-tag tool*

##### *.3.7.1.1 Non-electroporated parasites conditions*

A biotinylation study was conducted to evaluate the capability of the cMyc\_BirA\*\_GGS\_SNAP-tag tool to capture the target of interest DHFR-TS in presence of BG-PEG4-Mtx using non-electroporated parasite strains. The results are depicted in Figure 52.

Firstly, the results presented in Figure 52 clearly shows a well-defined band in the B1 fraction around 55 kDa (in both parts A and B), indicating that the pull-down was successful. This suggests that the streptavidin beads effectively captured the biotinylated proteins since the tool undergoes biotinylation.

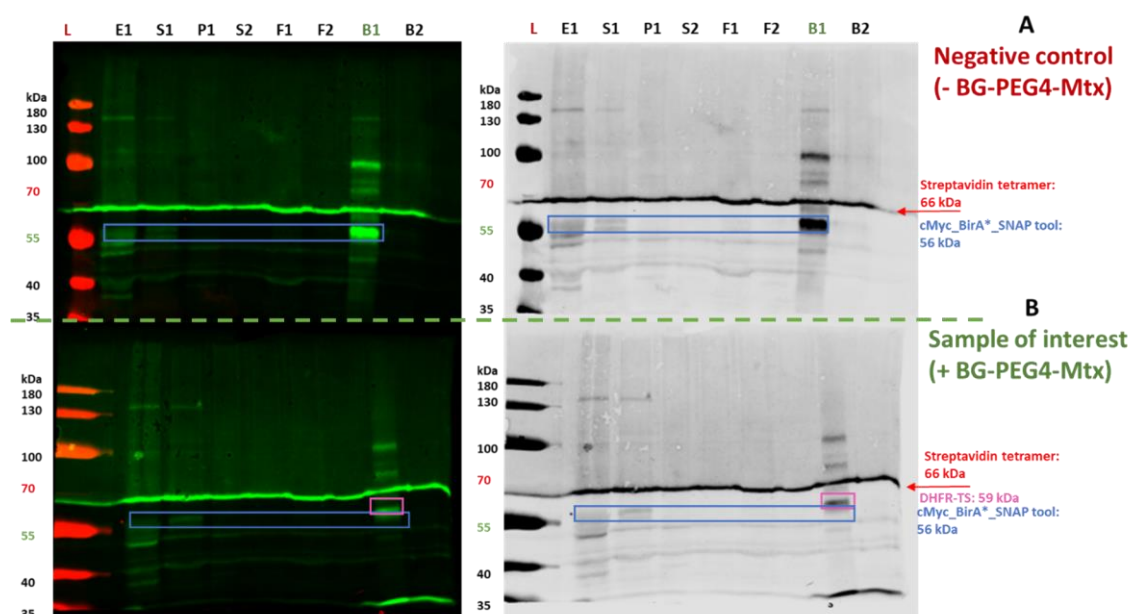


**Figure 52: Purification of candidate DHFR and cMyc\_BirA\_GGS\_SNAP-tag (Gel SDS PAGE 12%).** **Part A:** *Tbb\_G0* parasites (transfected with the cMyc\_BirA\*\_GGS\_SNAP-tag gene) incubated with  $1\mu\text{g/mL}$  tetracycline, and an excess of biotin ( $50\mu\text{M}$ ) added 24 h prior to extraction. **Part B:** *Tbb\_G0* parasites (transfected with the cMyc\_BirA\*\_GGS\_SNAP-tag gene) incubated with  $20\mu\text{M}$  BG-PEG4-Mtx,  $1\mu\text{g/mL}$  tetracycline, and an excess of biotin ( $50\mu\text{M}$ ) added 24 h prior to extraction. After overnight biotinylation, the cells were lysed with lysis buffer (0.5% NP-40, 0.5% deoxycholate, and 0.1% SDS) (E1), separated by centrifugation into detergent-soluble (S1) and detergent-insoluble (P1) fractions. The detergent-insoluble fraction (P1) was then further extracted with RIPA buffer (0.1% SDS). Fractions S1 and S2 were incubated with streptavidin conjugated Dynabeads, and the free unbound fractions (F1 and F2, respectively) were separated magnetically. Fractions B1 and B2 are, respectively, the bound cytoplasmic and cytoskeletal eluates and B1 should contain DHFR, cMyc\_BirA\*\_GGS\_SNAP-tag, and eventually near neighbors.

Moreover, the negative control (Figure 52 part A) exhibits a distinct band corresponding to the biotinylated tool, which is weaker in the sample of interest (Figure 52 part B). This can be explained by the higher fluorescence intensity emitted by the band corresponding to the biotinylated DHFR-TS (around 60 kDa, in B1, part B above), which may obscure that emitted by the biotinylated tool. Lastly, the absence of any observable bands in the S2 and B2 fractions confirms that the DHFR-TS target of interest is mainly present in the cytosolic part of the parasite rather than the insoluble portion. Therefore, future pull-downs aiming at detecting DHFR-TS can be conducted solely in the cytosol, without considering other parasite sub-compartments.

The pink box (part B in Figures 52) indicates the successful capture of the DHFR-TS target protein by the tool and its biotinylation. While the identification by fluorescence and western blot clearly indicates the ability of the deconvolution tool to depict the right target, due to the low amount recovered, mass spectrometry was unable to identify the biotinylated DHFR-TS protein or the biotinylated tool. To retrieve the maximum amount of proteins and obtain a bright and visible band of biotinylated DHFR-TS (in fraction B1), an additional lysis step in RIPA buffer (0.1% SDS, 1% NP-40, 0.5% deoxycholate) and sonication, were required in addition to the one in PEME buffer.

To enhance protein visualization and achieve clear differentiation between the bands representing the tool and the protein of interest, as well as to evaluate the effects of an increased concentration of BG-PEG<sub>4</sub>-Mtx, the experiment was repeated utilizing a 10% gel instead of the previous 12% gel, and the concentration of BG-PEG<sub>4</sub>-Mtx was adjusted to 50  $\mu$ M from the previous 20  $\mu$ M. The findings from this experiment are illustrated in Figure 53.

The blue framed part of the western blot image (part A in Figure 53) revealed by anti-cMyc antibodies shows, on the one hand, the correct expression of the cMyc\_BirA\*\_GGs-SNAP-tag tool (on fractions E1 & S1) and, on the other hand, similarly to the previous experiment, the successful capture of the tool by streptavidin beads (fraction B1) (the tool was biotinylated through its BirA\* domain).

The presence of a clear and intense band within the pink box (part B in Figures 53) indicates the successful biotinylation and capture of the DHFR-TS target protein by the tool.

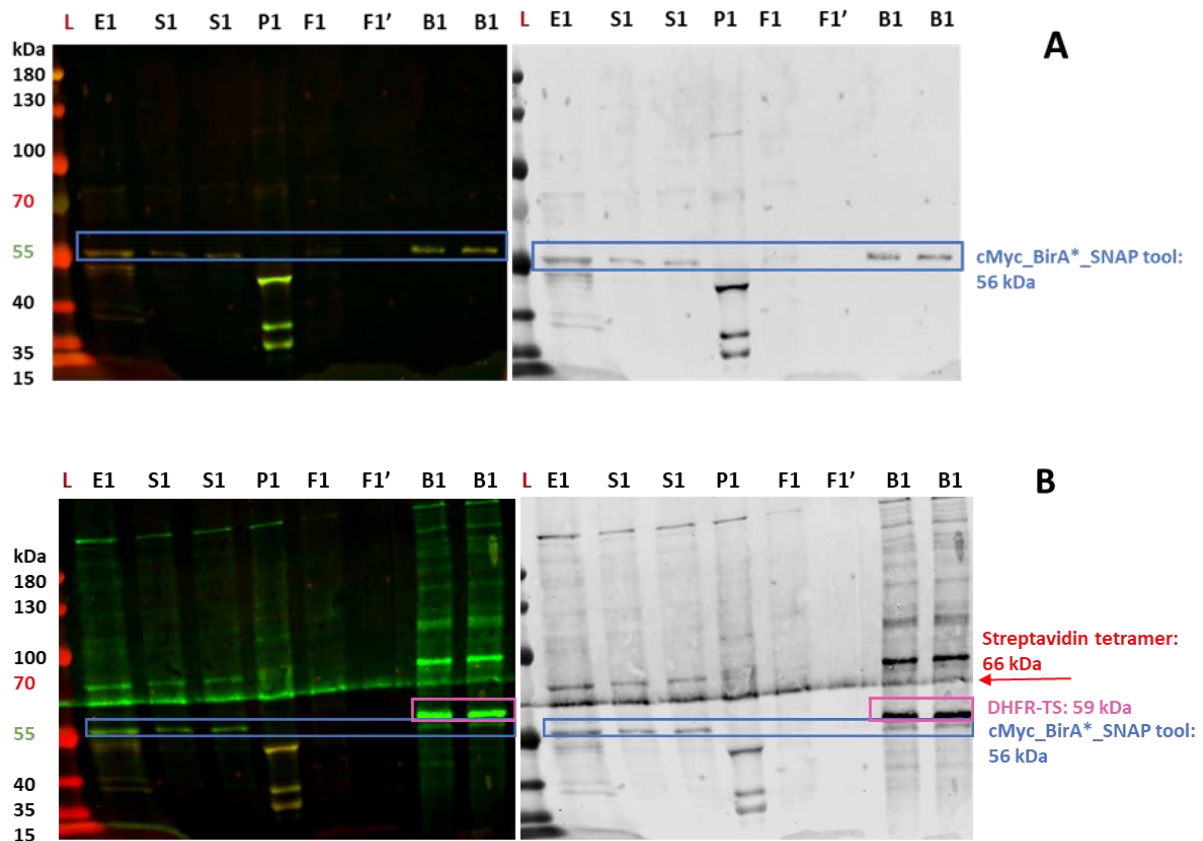


**Figure 53: Purification of candidate DHFR-TS and cMyc\_BirA\_GGS\_SNAP-tag (Gel SDS PAGE 10%).** *Part A:* the membrane was incubated with anti-cMyc antibodies, and *Part B:* the membrane was incubated with streptavidin dye ((IRDye 800CW Streptavidin). Parasites were incubated with BG-PEG<sub>4</sub>-Mtx (50  $\mu$ M), tetracycline (1 $\mu$ g/mL), and an excess of biotin (50  $\mu$ M) added 24 h before the pull-down. After overnight biotinylation, the cells were lysed with lysis buffer (0.5% NP-40, 0.5% deoxycholate, and 0.1% SDS), sonicated for 10 seconds 3 cycles (E1), and separated by centrifugation into detergent-soluble (S1) and detergent-insoluble (P1) fractions. The detergent-insoluble fraction (P1) was then further extracted with RIPA buffer (0.1% SDS). Fractions S1 were incubated with streptavidin-conjugated Dynabeads, and the free unbound fractions (F1 and F1' which are respectively the first and second flow-throughs) were separated magnetically. Fraction B1 is the bound cytoplasmic eluate and should contain DHFR-TS, cMyc\_BirA\*\_GGS\_SNAP-tag, and eventually near neighbors.

Different concentrations of BG-PEG<sub>4</sub>-Mtx were tested but the clearest results were obtained by incubating the parasites with 50  $\mu$ M of the compound, although lower concentrations have worked as well (Figure 52).

To validate the protocol including identification of the band by mass spectrometry, either band cut directly from the gel (10% SDS PAGE gels) (Figure 54 below) or streptavidin beads in 1X PBS, containing the captured biotinylated proteins (Figure 55 below), were sent for LC-

MS/MS analysis (analyses were performed on beads (fraction B1) that were pre-incubated in PBS 1X and subjected to direct trypsin digestion. This experiment was conducted by the proteomic platform of Dr. Alexandre Hainard and his team). While the fusion protein was detected, neither the cysteine-containing peptide of interest (Cys145) carrying the modification related to BG-PEG<sub>4</sub>-Mtx binding nor the biotinylated DHFR-TS could be detected. The lack of detection can be attributed to insufficient protein quantities in the samples despite their apparent visibility on the western blot (Figure 53 above).

us|cMYC-BIRA-SNAP|cMYC-BIRA-SNAP\_REC (100%), 56,624.4 Da

cMYC-BirA-SNAP-tag protein

18 exclusive unique peptides, 22 exclusive unique spectra, 22 total spectra, 247/521 amino acids (47% coverage)

|                     |                     |                     |                     |                     |                     |
|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| MEQK                | L I S E E D         | L D K D N T V P L K | L I A L L A N G E F | H S G E Q L G E T L | G M S R A A I N K H |
| I Q T L R           | D W G V D           | V F T V P G K G Y S | L P E P I Q L L N A | K Q I L G Q L D G G | S V A V L P V I D S |
| T N Q Y L L D R I G | E L K S G D A C I A | E Y Q Q A G R G G R | G R K W F S P F G A | N L Y L S M F W R L | Q D R K L A G I L V |
| E Q G P A A A I G L | S L V I G I V M A E | V L R K L G A D K V | R V K W P N D L Y L | N Q G W I T L Q E A | G I N L D R N T L A |
| E L T G K T G D A A | Q I V I G A G I N M | A M R R V E E S V V | N F I N R P V K L I | E K L E G G S M D K | I G D K E I F G I S |
| A M L I R E L R A A | L E L F E Q E G L A | P Y L S R W E K L D | F C H R V V Q G D L | D C E M K R T T L D | G G P E P L M Q A T |
| R G I D K Q G A L L | LEQDGI I K P W      | M G G E I S L R S A | V E V P A P A A V L | K V V K F G E V I S | D V G G Y E G G L A |
| S P L G K L E L S G | C E Q Q L H R I I F | L G K G T S A A D A | F T R Q V L W K L L |                     |                     |
| AWL N A Y F H Q P   | E A I E E F P V P A | L H H P V F Q Q E S |                     |                     |                     |
| Y S H L A A L A G N | P A A T A A V K T A | L S G N P V P I L I |                     |                     |                     |
| V K E W L L A H E G | H R L G K P G L G Q | L                   |                     |                     |                     |

**Figure 54: On-beads MS identification in fraction B1 of peptides (after on-beads digestion with trypsin) corresponding to the cMyc\_BirA\*\_GGs\_SNAP-tag. Neither the modification of Cys145 by BG-PEG<sub>4</sub>-Mtx nor the biotinylated DHFR-TS was detected. The cysteine of interest (Cys 145) is boxed in red (analyses made by the proteomic platform of Dr. Alexandre Hainard and his team)**

us|cMYC-BIRA-SNAP|cMYC-BIRA-SNAP\_REC (100%), 56,624.4 Da

cMYC-BirA-SNAP-tag protein

17 exclusive unique peptides, 20 exclusive unique spectra, 20 total spectra, 223/521 amino acids (43% coverage)

|                     |                     |                     |                     |                     |                     |
|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| MEQK                | L I S E E D         | L D K D N T V P L K | L I A L L A N G E F | H S G E Q L G E T L | G M S R A A I N K H |
| I Q T L R           | D W G V D           | V F T V P G K G Y S | L P E P I Q L L N A | K Q I L G Q L D G G | S V A V L P V I D S |
| T N Q Y L L D R I G | E L K S G D A C I A | E Y Q Q A G R G G R | G R K W F S P F G A | N L Y L S M F W R L | Q D R K L A G I L V |
| E Q G P A A A I G L | S L V I G I V M A E | V L R K L G A D K V | R V K W P N D L Y L | N Q G W I T L Q E A | G I N L D R N T L A |
| E L T G K T G D A A | Q I V I G A G I N M | A M R R V E E S V V | N F I N R P V K L I | E K L E G G S M D K | I G D K E I F G I S |
| A M L I R E L R A A | L E L F E Q E G L A | P Y L S R W E K L D | F C H R V V Q G D L | D C E M K R T T L D | G G P E P L M Q A T |
| R G I D K Q G A L L | LEQDGI I K P W      | M G G E I S L R S A | V E V P A P A A V L | K V V K F G E V I S | D V G G Y E G G L A |
| S P L G K L E L S G | C E Q Q L H R I I F | L G K G T S A A D A | F T R Q V L W K L L |                     |                     |
| AWL N A Y F H Q P   | E A I E E F P V P A | L H H P V F Q Q E S |                     |                     |                     |
| Y S H L A A L A G N | P A A T A A V K T A | L S G N P V P I L I |                     |                     |                     |
| V K E W L L A H E G | H R L G K P G L G Q | L                   |                     |                     |                     |

**Figure 55: In-gel MS identification in fraction B1 of peptides (after in-gel digestion with trypsin) corresponding to the cMyc\_BirA\*\_GGs\_SNAP-tag. Neither the modification of Cys145 by BG-PEG<sub>4</sub>-Mtx nor the biotinylated DHFR-TS was detected. The cysteine of interest (Cys 145) (analyses made by the proteomic platform of Dr. Alexandre Hainard and his team)**

To enhance the protein quantity for MS detection, larger cell lysates (~300 mL of growth medium) were used instead of the 50 mL from previous pull-down assays. The lysates, with or without BG-PEG<sub>4</sub>-Mtx, were electrophoretically separated, and the bands corresponding to

the target of interest (DHFR-TS) and cMyc\_BirA\*\_GGs\_SNAP-tag were excised from the gel. Following trypsin digestion, MS analysis was performed, and the findings are presented in Figure 56 below.

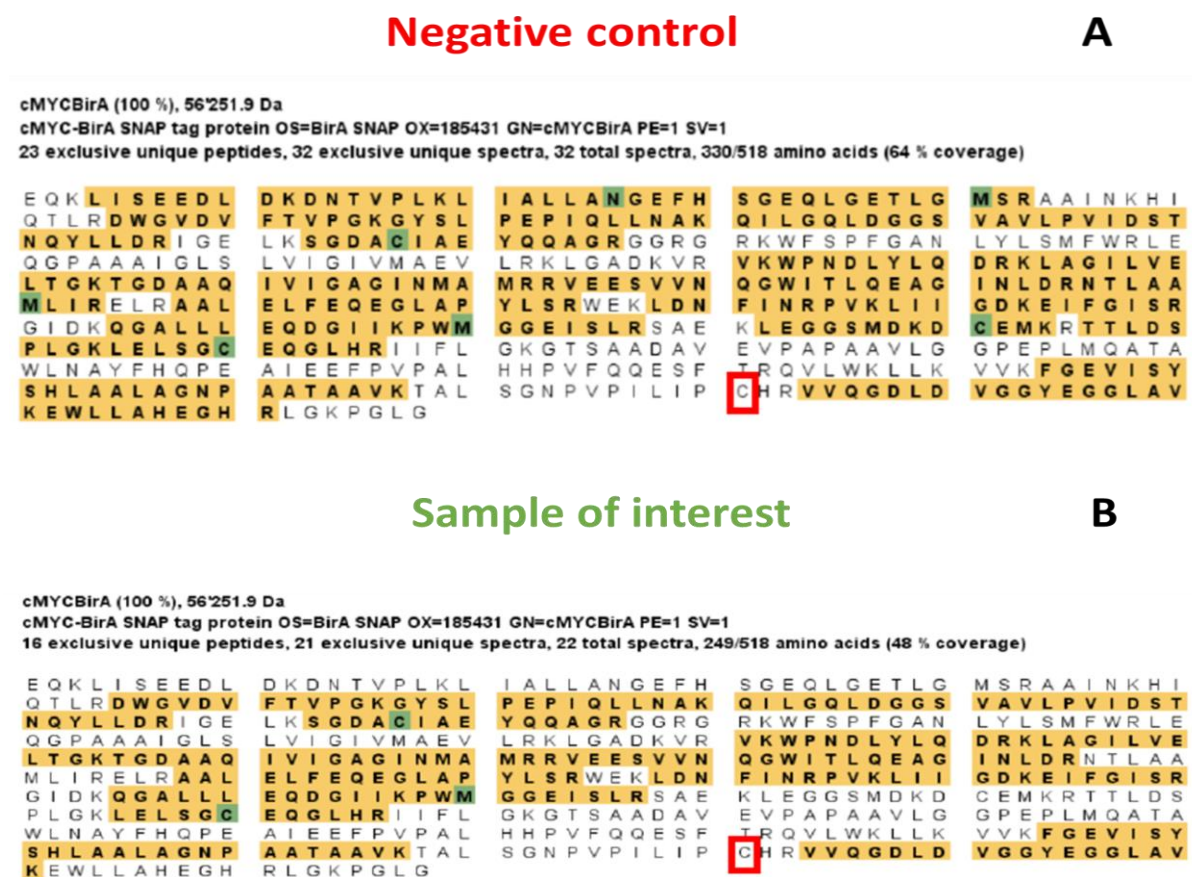


**Figure 56:** LC-MS/MS analysis of cell lysates incubated with BG-PEG<sub>4</sub>-Mtx (B) or without (A). The two distinct cell lysates (A and B) were subjected to MS detection for the identification of peptides matching the cMyc\_BirA\*\_GGs\_SNAP-tag fusion protein (analyses performed by the proteomic platform of Dr. Alexandre Hainard and his team)

Based on the data presented in Figure 56, there was no detectable difference concerning the binding of the compound of interest to the cysteine (Cys145) of the SNAP-tag active site between the negative control (part A: parasite sample incubated without BG-PEG<sub>4</sub>-Mtx) and the sample of interest (part B: parasite sample incubated without BG-PEG<sub>4</sub>-Mtx). To overcome this limitation, an electroporation-based approach was used to force the entry of BG-PEG<sub>4</sub>-Mtx into the parasites. The rationale behind this strategy was to increase the amount of compound that penetrates the cell, potentially allowing for the capture of more DHFR-TS and

facilitating the detection of the BG-PEG<sub>4</sub> moiety's binding to the SNAP-tag domain. The results of this electroporation experiment are presented in .3.7.1.2 below.

#### **.3.7.1.2 Electroporated parasites conditions**

As previously stated, the band that could potentially correspond to the biotinylated DHFR-TS target is faint, despite being visible on the membrane, and therefore could not be identified through MS analysis. To address this issue, the idea arose to enhance the quantity of captured target and thereby confirm its presence via MS. This approach involves electroporation of the parasites using the Nucleofactor device, which was previously used to transfect the parasites. The underlying concept was to force the entry of a greater quantity of BG-PEG<sub>4</sub>-Mtx compounds. Successful implementation of this method would be highly advantageous when non-cell membrane permeable BG-PEG<sub>4</sub> derivatives are utilized. Additionally, as demonstrated in Section 3.5.2, electroporation has a minimal impact on parasite viability, making it a suitable method for this experiment.

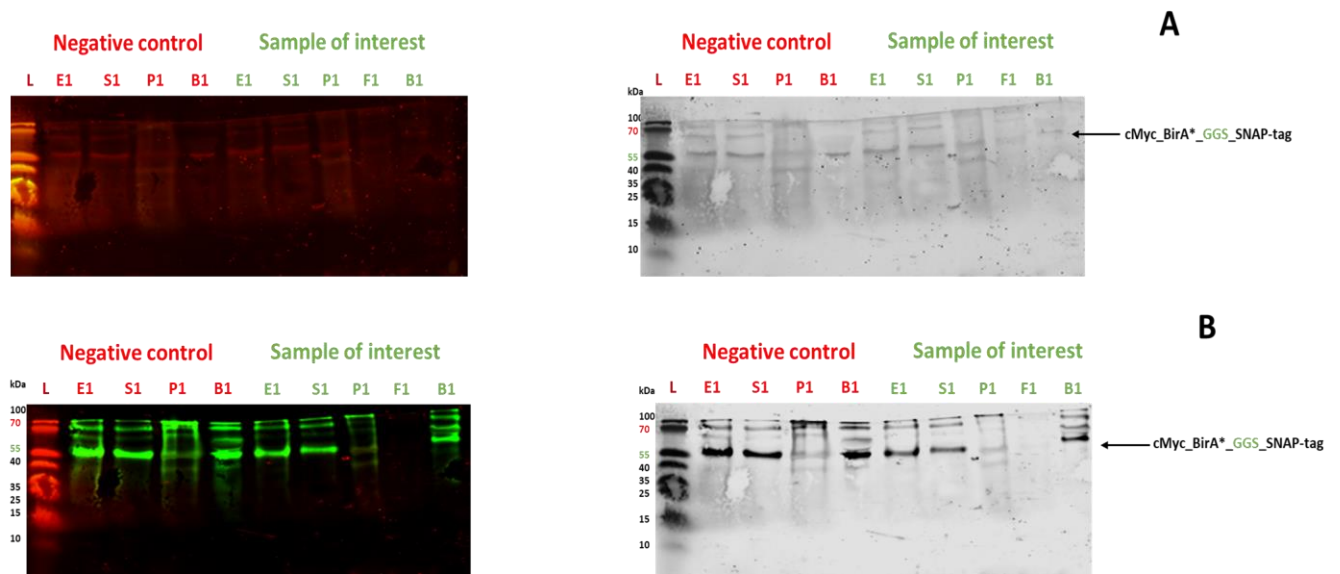


**Figure 57: Purification of candidate DHFR and cMyc\_BirA\_GGS\_SNAP-tag (Gel Tris-tricine SDS-PAGE 10%).** **Part A:** the membrane was incubated with anti-cMyc antibodies, and **Part B:** the membrane was incubated with streptavidin dye ((IRDye 800CW Streptavidin). Parasites were incubated with **50  $\mu$ M** of BG-PEG<sub>4</sub>-Mtx (only for the sample of interest), tetracycline (**1 $\mu$ g/mL**), and an excess of biotin (**50  $\mu$ M**) added 24 h before the pull-down. After overnight biotinylation, the cells were lysed with lysis buffer (0.5% NP-40, 0.5% deoxycholate, and 0.1% SDS), sonicated for 10 seconds 3 cycles (**E1**), separated by centrifugation into detergent-soluble (**S1**) and detergent-insoluble (**P1**) fractions. The detergent-insoluble fraction (**P1**) was then further extracted with RIPA buffer (0.1% SDS). Fractions **S1** was incubated with streptavidin conjugated Dynabeads, and the free unbound fraction (**F1**) was separated magnetically from the beads. Fraction **B1** is the bound cytoplasmic eluate and should contain the biotinylated DHFR-TS, the biotinylated tool, and eventually biotinylated near neighbors.

First, what is apparent on the membrane of part A of Figure 57 above, is a slightly visible band on all fractions (negative control or sample of interest), the membrane being revealed by anti-cMyc antibodies, this band, therefore, corresponds to the tool (the right molecular weight of the band (55 kDa) also validates its presence). Then, this band becomes very intense in most of the fractions revealed by the streptavidin dye except in P1 and F1 demonstrating respectively the absence of the expression of the tool in the insoluble compartment of the parasite, and in the flow through. This last result indicates the efficacy of the beads in capturing biotinylated proteins.

Afterward, what is noteworthy is the absence of difference between the negative control (parasites electroporated without the presence of the compound) and the sample of interest (parasites electroporated and incubated with 50  $\mu$ M of BG-PEG<sub>4</sub>-Mtx). The B1 fraction of both

conditions was therefore subjected to MS analysis to confirm the presence or absence of biotinylated DHFR-TS, as well as the binding of the tool to the derivative.

Two additional biotinylation experiments were conducted in the presence of 20  $\mu$ M BG-PEG<sub>4</sub>-Mtx, and the results are presented in Figure 58. In part A of the figure, a strong band is visible, most likely corresponding to the biotinylated DHFR-TS. Additionally, a clear difference is observed between **B1-** (negative control) and **B1 el+** (sample of interest). Part B of the figure reveals a band at approximately 59 kDa, exclusively present in **B1 el+**. This band is probably corresponding to the biotinylated DHFR-TS target, validating the efficiency of the tool.

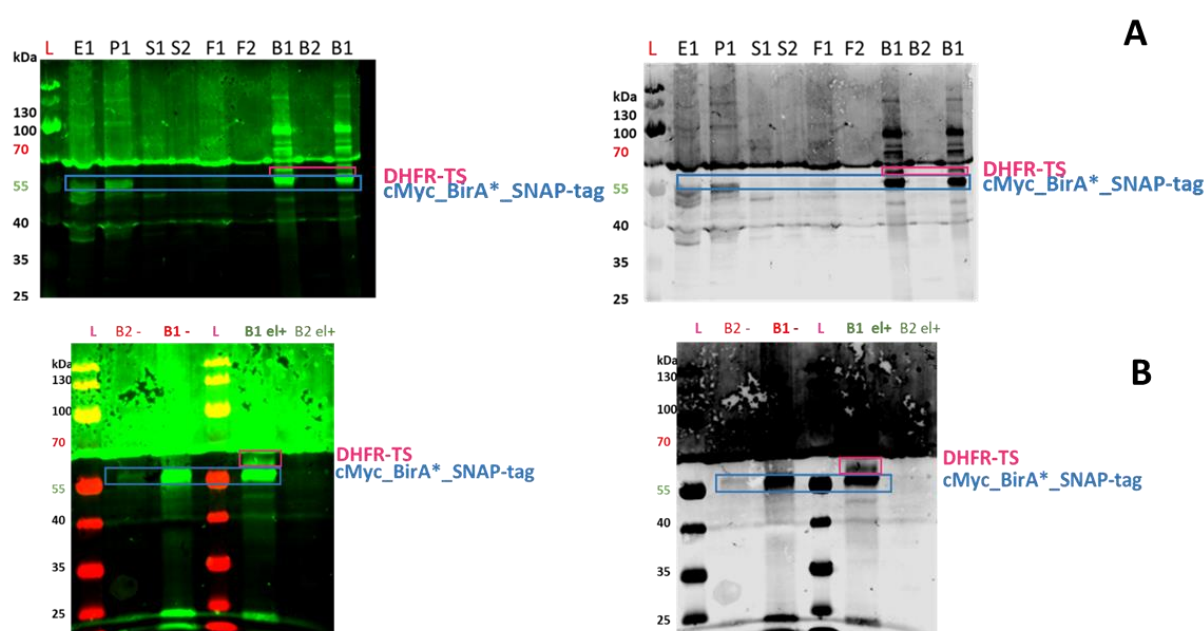


**Figure 58: Purification of candidate DHFR and cMyc\_BirA\_GGS\_SNAP-tag (Gel SDS PAGE 12%).** **Part A:** The parasites were subjected to electroporation and immediately incubated with tetracycline (1 $\mu$ g/mL) and BG-PEG<sub>4</sub>-Mtx (20  $\mu$ M). An excess of biotin (50  $\mu$ M) was introduced 24 hours prior to conducting the pull-down procedure. After overnight biotinylation, the cells were lysed using lysis buffer (0.5% NP-40, 0.5% deoxycholate, and 0.1% SDS) designated as **E1**, and then centrifuged to separate them into detergent-soluble (**S1**) and insoluble (**P1**) fractions. The detergent-insoluble fraction (**P1**) was further extracted with RIPA buffer (0.1% SDS). Fractions **S1** and **S2** were incubated with streptavidin conjugated Dynabeads, and the resulting free unbound fractions (**F1** and **F2**, respectively) were separated magnetically. The cytoplasmic and cytoskeletal eluates that are bound to the beads are designated as **B1** and **B2**, respectively, with **B1** expected to contain DHFR, cMyc\_BirA\*\_GGS\_SNAP-tag, and possibly nearby molecules. **Part B:** The parasites were subjected to electroporation and immediately incubated with tetracycline (1 $\mu$ g/mL) and either with BG-PEG<sub>4</sub>-Mtx (20  $\mu$ M) (**B1 el+**, **B2 el+**) or without (**B1-**, **B2**), all bands correspond to fractions **B** (beads attached to the biotinylated proteins).

Unfortunately, the MS could not confirm the presence of biotinylated peptides for either the tool or the DHFR-TS target. Furthermore, the modification of Cys145 by the BG-PEG<sub>4</sub>-Mtx compound could not be identified either.

In order to verify whether this is a permeability problem of the BG-PEG<sub>4</sub>-Mtx compound or not, a pull-down using rabbit anti-cMyc antibodies covalently immobilized on agarose resin (high-affinity immunoprecipitation resin, reference in section 2.1) was performed after lysis of the parasites previously incubated with 50  $\mu$ M of BG-PEG<sub>4</sub>-Mtx. MS analysis was then conducted after trypsin digestion of the beads. The fusion protein cMyc\_BirA\*\_GGG\_SNAP-tag was clearly detected, but no peptide containing the Cys145 modified by the covalent binding of the BG-PEG<sub>4</sub>-Mtx derivative could be observed (results in Figure 59 below) despite the very good coverage (50%).

**cMYC-BirA-SNAP-tag protein**

18 exclusive unique peptides, 24 exclusive unique spectra, 34 total spectra, 259/521 amino acids (50 % coverage)

|                     |                     |                     |                     |                     |                     |
|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| MEQK                | L I S E E D         | L D K D N T V P L K | L I A L L A N G E F | H S G E Q L G E T L | G M S R A A I N K H |
| I Q T L R           | D W G V D           | V F T V P G K G Y S | L P E P I Q L L N A | K Q I L G Q L D G G | S V A V L P V I D S |
| T N Q Y L L D R I G | E L K S G D A C I A | E Y Q Q A G R G G R | V L R K L G A D K V | G R K W F S P F G A | N L Y L S M F W R L |
| E Q G P A A A I G L | S L V I G I V M A E | A M R R V E E S V V | N Q G W I T L Q E A | R V K W P N D L Y L | Q D R K L A G I L V |
| E L T G K T G D A A | Q I V I G A G I N M | P Y L S R W E K L D | N F I N R P V K L I | E K L E G G S M D K | G I N L D R N T L A |
| A M L I R E L R A A | L E L F E Q E G L A | M G G E I S L R S A | V E V P A P A A V L | I G D K E I F G I S | D C E M K R T T L D |
| R G I D K Q G A L L | L E Q D G I I K P W | L G K G T S A A D A | F T R Q V L W K L L | K V V K F G E V I S | G G P E P L M Q A T |
| S P L G K L E L S G | C E Q G L H R I I F | L H H P V F Q Q E S | P C H R V V Q G D L | D V G G Y E G G L A |                     |
| A W L N A Y F H Q P | E A I E E F P V P A | L S G N P V P I L I |                     |                     |                     |
| Y S H L A A L A G N | P A A T A A V K T A |                     |                     |                     |                     |
| V K E W L L A H E G | H R L G K P G L G Q |                     |                     |                     |                     |

**Figure 59:** LC-MS/MS analysis of the resulting fraction from immunoprecipitation with anti-cMyc antibodies from lysed cells (after **on-beads** digestion with trypsin). Only peptides corresponding to cMyc\_BirA\*\_GGG\_SNAP-tag were identified. However, the modification of Cys145 (red box) by BG-PEG<sub>4</sub>-Mtx could not be detected (analyses carried out by the proteomic platform of Dr. Alexandre Hainard and his team).

Two hypotheses can be proposed here. The first one is linked to the non-accessibility of the compound, particularly the BG moiety, to the active site of the SNAP-tag that could be concealed by the BirA\* domain, which was predicted (albeit with low probability as the FCC value is very low) in computational studies as being a possible 3D arrangement of the fused SNAP and BirA\* domains (refer to Section .3.1.3). The value obtained from the FCC with a GGS linker is almost zero but not entirely null, which might explain why the tool works occasionally but not consistently. The biotinylation assays were not reproducible, mainly with non-electroporated parasites (refer to the above results). To confirm this hypothesis, as outlined in Section .3.1.3, various linkers that join the SNAP-tag and BirA\* domains have been tested

and assessed. These linkers include the ones with good FCC values such as the PTP linker and the ones with poorer values including the GGGs, GGGG, and PAPAP linkers.

The second hypothesis pertains to the permeability of BG-PEG<sub>4</sub>-Mtx through the parasite membrane. When BG-PEG<sub>4</sub> is attached to Mtx, it might result in a non-permeable compound (unlike the BG-PEG<sub>4</sub>-FDA derivative, which displayed good permeability based on microscopic studies). To verify this assertion, parasites that express only the pure SNAP-tag protein (without the BirA\* domain) have been generated and then incubated with either BG-PEG<sub>4</sub>-Mtx or BG-PEG<sub>4</sub>-biotin (since biotin is a smaller molecule and has different physicochemical properties than Mtx). Moreover, to compare the binding of these two derivatives to the SNAP-tag active site, purified pure SNAP-tag protein will be incubated with these same compounds at 37°C for 1-3 hours, and will then be subjected to MS analysis (see **Figure 76** below). The results are presented in sections 3.7.2 and 3.8 below.

Note that BG-PEG<sub>4</sub>-biotin was selected since it produced excellent outcomes with a project that employed a three-hybrid system containing the SNAP-tag protein.

### *.3.7.2 Verification of BG moiety binding to SNAP-Tag*

As reported in .3.7.2, the results obtained from the MS analysis were not able to identify the modification of the Cys145 active site of SNAP-tag by BG-PEG<sub>4</sub>-Mtx. To test one of the two hypotheses mentioned earlier, which suggests that the derivative is either inaccessible or impermeable, the decision was made to generate parasites expressing only the pure SNAP-tag protein (as mentioned above). This was accomplished by transfecting the gene of interest, obtained by simple digestion of the plasmid pLew100\_BirA\_GGS\_SNAP-tag using XhoI and BamHI to eliminate the BirA part. Electroporation was used to carry out the transfection. The lysate was prepared for MS analysis following the protocol outlined in Section 2.2.26.

Unfortunately, despite multiple attempts, it has never been possible to obtain strains expressing only the pure SNAP-tag protein (MS analysis of the parasite lysate confirmed the lack of expression). As a result, it was not possible to conduct the incubation test of these parasites in the presence of BG-PEG<sub>4</sub>-Mtx or other BG-PEG<sub>4</sub>-derivatives (such as BG-PEG<sub>4</sub>-Biotin). Consequently, it was decided to test only the compound's binding directly on the pure protein.

To this end, pure SNAP-tag protein at a concentration of 10  $\mu$ M was incubated with 50  $\mu$ M of BG-PEG<sub>4</sub>-Mtx at 37°C for an hour. To prevent disulfide bonding, 50 mM DTT was added to the reaction. The resulting product was loaded onto a gel at a concentration of 0.226 mg/mL (11.3 mg for 50  $\mu$ L loaded), followed by electrophoresis separation. The band of interest was then excised from the gel and sent to the MS proteomics platform for comparison with *the in vivo* results (parasites exclusively producing SNAP-tag protein in the cytosol were exposed to BG-PEG<sub>4</sub>-Mtx). The analysis revealed that the compound (+788 Da) had bound to the Cys145 (1687 Da) of the SNAP-tag active site, as detailed in the results shown in Figure 60.

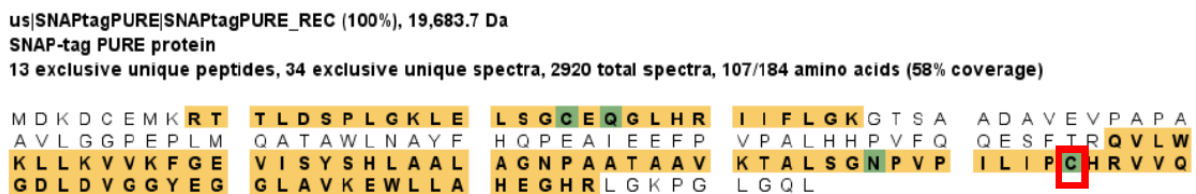


**Figure 60:** LC-MS/MS analysis of peptides corresponding to the pure SNAP-tag protein with the modification of Cys145 (red box) obtained by its binding with the BG-PEG<sub>4</sub>-Mtx compound: Identified peptides satisfying the minimum peptide threshold are emphasized in yellow, and modifications are marked in green (analyses carried out by the proteomic platform of Dr. Alexandre Hainard and his team).

However, this above result could only be observed by lowering the thresholds of the identification scores. Only a few MS spectra were able to identify the peptide with the BG-PEG<sub>4</sub>-Mtx modification (+788 Da) on Cys145 with a mass of 2475 Da (1687 Da+ 788 Da), and these peptides had low probability scores and unsatisfactory quality spectra. Most of the identified peptides containing the Cys145 of interest had a carbamidomethylated cysteine, indicating a free cysteine without modification by the compound of interest. This result may explain the challenge of using MS to visualize the modification of the Cys145 of the SNAP-tag active when it is expressed *in vivo* in the parasite cytosol as a fusion protein. Nonetheless, the lack of confirmation by MS does not exclude the proper functioning of the tools of interest and in particular of the following system: cMyc\_BirA\*\_GGG\_SNAP-tag (Figures 52, 53, and 58 above clearly demonstrate the effective detection of DHFR-TS on western blot membranes).

Additional *in-vitro* tests have been carried out using pure SNAP-tag protein along with the two derivatives BG-PEG<sub>4</sub>-Mtx and BG-PEG<sub>4</sub>-Biotin. A detailed description of these assays can be found in Section 2.2.26 and the results in Section 3.8 below.

### 3.7.3 Biotinylation of DHFR-TS by the cMyc\_TBirA\*\_GGG\_SNAP-tag tool

To determine the effectiveness of the cMyc\_TBirA\*\_GGG\_SNAP-tag tool in capturing the DHFR-TS target of interest, a biotinylation study was conducted on parasites transfected with the cMyc\_TBirA\*\_GGG\_SNAP-tag gene. The results are presented in Figure 61 below:

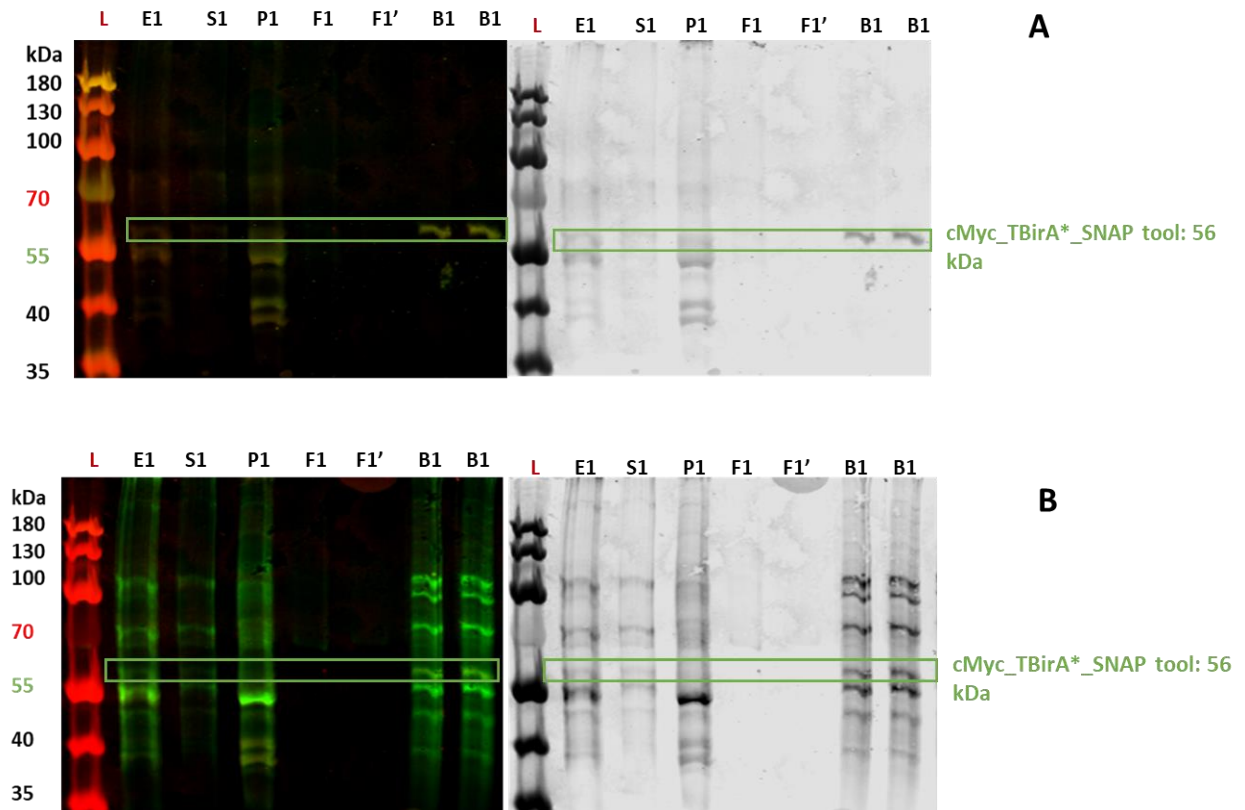


**Figure 61: Purification of candidate DHFR-TS and cMyc\_TBirA\*\_GGG\_SNAP-tag (Gel SDS PAGE 10%).** *Part A: the membrane was incubated with anti-cMyc antibodies, and Part B: the membrane was incubated with streptavidin dye ((IRDye 800CW Streptavidin). Parasites were incubated with BG-PEG<sub>4</sub>-Mtx (50  $\mu$ M) (only for the sample of interest), tetracycline (1  $\mu$ g/mL), and an excess of biotin (50  $\mu$ M) added 10 minutes before the pull-down. Following the biotinylation stage, the subsequent steps were carried out identically to the previously performed ones, using the same designations E1, S1, P1, F1, F1' (which are respectively the first and the second flow through) and B1 (which refers to the beads that have bound to the cytoplasmic eluate, that should contain the biotinylated target, the tool, and potentially nearby biotinylated neighbors).*

The streptavidin beads captured the cMyc\_TBirA\*\_GGG\_SNAP-tag, as demonstrated by the image of the membrane after incubation with anti-cMyc antibodies or streptavidin dye (Figure 61). This confirms the successful biotinylation of the tool via its TBirA moiety (TurboID

technology). However, DHFR-TS was not detected on the western blot membrane despite multiple attempts using varied processing conditions.

Furthermore, there is a distinct indication of a band positioned beneath the one that corresponds to the tool, with an approximate weight of 55 kDa, on the membrane. This specific band was not discernible by anti-cMyc antibodies, however, it was detected using streptavidin dye. It is inferred that this band is a result of a fragment of the fusion protein that lacks the cMyc-tag since its identification was absent when utilizing specific antibodies. This discovery validates the instability of the system, which has previously been reported in various studies (119).

This instability also confirms the results obtained during the study of the time-dependent expression of the tool (section 3.6.2), where this fragment was already clearly visible (figure 45) and where the maximum quantity of protein of interest expressed was only 8%.

Additional experiments were conducted where the biotin was added at different intervals before the pull-down. Instead of adding it 10 minutes prior to the pull-down, which is a unique feature of the TurboID and miniTurbo variants that can biotinylate in 10 minutes compared to 24 hours, the biotin was added at 1 hour, 2 hours, and 24 hours before the pull-down. However, the results were comparable to the ones previously obtained, which are not shown here. Therefore, the cMyc\_TBirA\_GGS\_SNAP-tag fusion protein based deconvolution tool was excluded and not chosen for subsequent experiments (119).

#### *.3.7.4 Biotinylation of DHFR-TS by the cMyc\_mTBirA\*\_GGS\_SNAP-tag tool*

A biotinylation study was performed on parasites transfected with the cMyc\_mTBirA\_GGS\_SNAP-tag gene to assess the efficacy of the cMyc\_mTBirA\_GGS\_SNAP-tag tool in capturing the DHFR-TS target of interest. The results are illustrated in Figure 62 below:

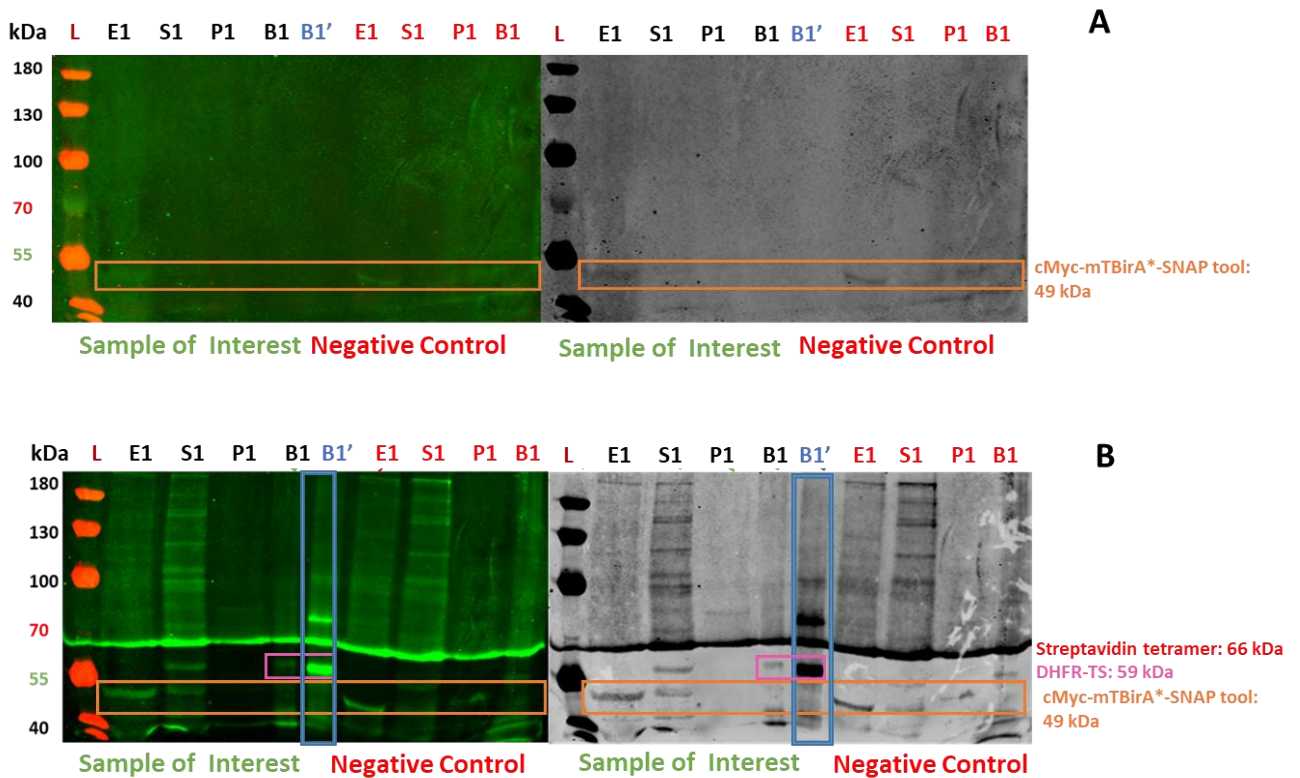


**Figure 62: Purification of candidate DHFR and cMyc\_mTBirA\_GGS\_SNAP-tag (Gel SDS PAGE 8%).** Part A: the membrane was incubated with anti-cMyc antibodies, and Part B: the membrane was incubated with streptavidin dye (IRDye 800CW Streptavidin). Parasites were incubated with BG-PEG<sub>4</sub>-Mtx (**50  $\mu$ M**) (only for the sample of interest), tetracycline (**1  $\mu$ g/mL**), and an excess of biotin (**50  $\mu$ M**) added **10 minutes** before the pull-down. Following the biotinylation stage, the subsequent steps were carried out identically to the previously performed ones, using the same designations E1, S1, P1, F1, B1 (which refers to the beads that have bound to the cytoplasmic eluate, which should contain the biotinylated target, the tool, and potentially nearby biotinylated neighbors), and **B1'** (the bound cytoplasmic eluate that should contain DHFR-TS and the biotinylated original tool with BirA\*).

The orange boxed area in the western blot image revealed by anti-cMyc antibodies (part A in Figure 62 above) and streptavidin dye (part B in the same figure) indicates the correct expression of the cMyc\_mTBirA\*\_GGS\_SNAP-tag (bands around 50 kDa of low intensity in the E1 fractions of part A and higher intensity in those of part B as revealed with streptavidin dye).

The highlighted area (fraction B1 in part B of the figure above) confirms the successful biotinylation of the DHFR-TS target protein using the tool, as well as its proper capture by the streptavidin-coated beads. It is noteworthy that the intensity of the band of interest is greatly attenuated compared to the B1' fraction captured by the original cMyc\_BirA\*\_GGS\_SNAP-tag

tool. Despite scaling up the culture media to 300 mL, the quantities were still very low and neither the biotinylated DHFR-TS nor the biotinylated tool could be identified by MS.

Additional experiments were conducted to modify the timing of biotin addition (either 1 hour, 2 hours, or 24 hours before pull-down). Nonetheless, the outcomes were akin to the earlier results obtained using the TurboID\_SNAP-tag tool (these are not presented here).

In conclusion, given the evident results showing that the BioID\_SNAP-tag system performs better than the other two variants (TurboID\_SNAP-tag and miniTurbo\_SNAP-tag), the original system was selected for further refinement.

After obtaining inconclusive results with the non-electroporated strains, the biotinylation studies involving the electroporation of parasites *Tbb\_TG0* (*Tbb* carrying the cMyc\_ **TBirA**\_GGG\_SNAP-tag gene) and *Tbb\_mTG0* (*Tbb* carrying the cMyc\_ **mTBirA**\_GGG\_SNAP-tag gene) in the presence of BG-PEG<sub>4</sub>-Mtx were not pursued. Consequently, these two variants were excluded from further experimentation, and the focus shifted exclusively to the original *Tbb\_GO* variant (*Tbb* carrying the cMyc\_ **BirA\***\_GGG\_SNAP-tag gene) given its positive outcomes (refer to Section .3.7).

#### *.3.7.5 Biotinylation of DHFR-TS by the cMyc\_BirA\*\_GGG\_SNAP-tag tool*

To evaluate the effectiveness of the cMyc\_BirA\_GGG\_SNAP-tag tool and the role of the linker in capturing the DHFR-TS target of interest, a biotinylation study was conducted on parasites transfected with the cMyc\_BirA\_GGG\_SNAP-tag gene, following the established methodology used for the other tools.

The results from the negative control, which involved parasites that were not exposed to BG-PEG<sub>4</sub>-Mtx, and the results from parasites that were electroporated with or without the compound of interest (Figure 63) were comparable to the findings observed in Figure 56 in section 3.7 (MS results). The MS analysis (data not shown) did not detect biotinylated DHFR-TS or the modification of Cys145 by BG-PEG<sub>4</sub>-Mtx at the SNAP-tag active site.

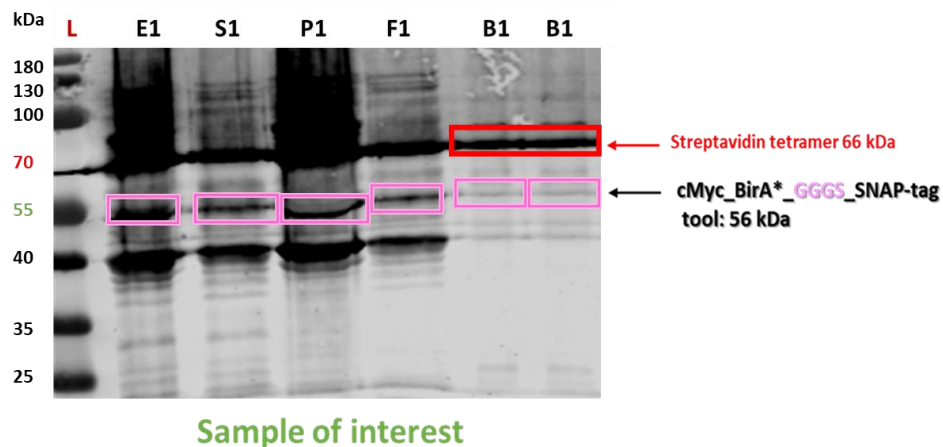


**Figure 63: Purification of candidate DHFR and cMyc\_BirA\*\_GGGS\_SNAP-tag (Gel SDS PAGE 10%).** The membrane was incubated with streptavidin dye ((IRDye 800CW Streptavidin). Parasites were incubated with BG-PEG<sub>4</sub>-Mtx (50  $\mu$ M), tetracycline (1 $\mu$ g/mL), and an excess of biotin (50  $\mu$ M) 24 h before the pull-down. Following the overnight biotinylation, the subsequent steps were carried out identically to the previously performed ones, using the same designations E1, S1, P1, F1, and B1 (which refers to the beads that have bound to the cytoplasmic eluate, which should contain the biotinylated target, the tool, and potentially nearby biotinylated neighbors).

This result confirms the computational analysis, which indicated that SNAP and BirA\* tend to form an agglomerate complex with a GGGS linker, thereby covering the active site of the SNAP-tag domain and preventing BG-PEG<sub>4</sub>-Mtx from accessing Cys145. Consequently, the tool with this linker was rejected and was not pursued for further experiments. It is worth noting that the high-intensity band seen around 65 kDa in all fractions corresponds to the streptavidin tetramer, most likely released by the beads, which may be due to improper bead pre-washing.

### .3.7.6 Biotinylation of DHFR-TS by the cMyc\_BirA\*\_GGGS\_SNAP-tag tool

A biotinylation study was carried out on parasites transfected with the cMyc\_BirA\_GGGS\_SNAP-tag gene to assess how the highly flexible linker (GGGS) affects the capability of the tool in capturing the DHFR-TS target of interest. The study adhered to the same methodology utilized for the other tools.

#### .3.7.6.1 Non-electroporated parasites conditions

The study was first carried out with non-electroporated strains as illustrated below:

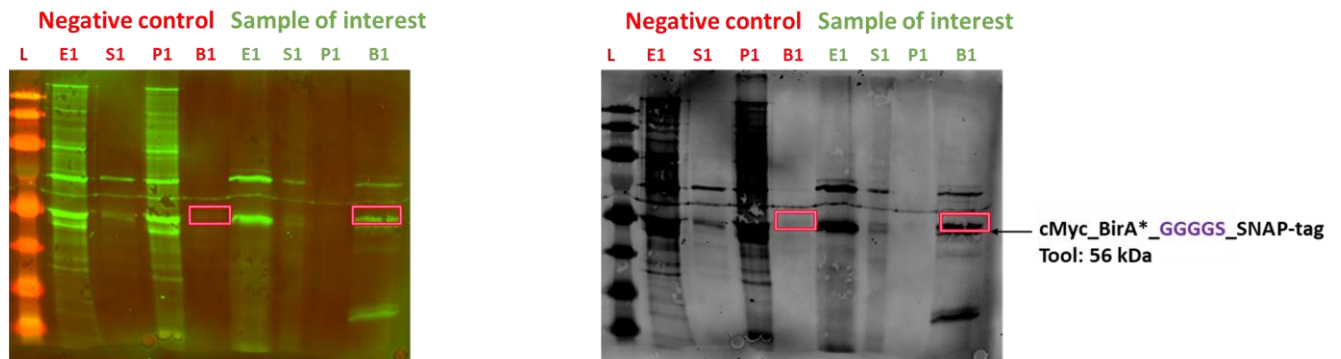


**Figure 64: First attempt: Purification of candidate DHFR and cMyc\_BirA\*\_GGGGS\_SNAP-tag (Gel SDS PAGE 12%).** The membrane was incubated with streptavidin dye ((IRDye 800CW Streptavidin). Parasites were incubated with BG-PEG<sub>4</sub>-Mtx (50  $\mu$ M) (only for the sample of interest), tetracycline (1 $\mu$ g/mL), and an excess of biotin (50  $\mu$ M) added 24 h before the pull-down. Following the overnight biotinylation, the subsequent steps were carried out identically to the previously performed ones, using the same designations E1, S1, P1, F1, and B1 (which refers to the beads that have bound to the cytoplasmic eluate, which should contain the biotinylated target, the tool, and potentially nearby biotinylated neighbors).

As illustrated in Figure 64 above, DHFR-TS is not detected in any of the fractions except for the tool of interest. In addition, the only notable discrepancy between the negative control and the sample of interest is the greater visibility of the biotinylated tool at around 55 kDa in B1. Further attempts were made to determine if these results were caused by handling errors (see below). Despite the presence of the bubble (Figure 65), the tool captured by the beads in the B1 fractions of both the negative control and the sample of interest was clearly visible. Specifically, a faint band around 55 kDa was visible in both conditions in part A of Figure 65, which was further intensified after the revelation with streptavidin dye in part B. Since this band was highlighted by the anti-cMyc antibodies in part A and considering its molecular weight, this band certainly corresponds to the biotinylated fusion protein.

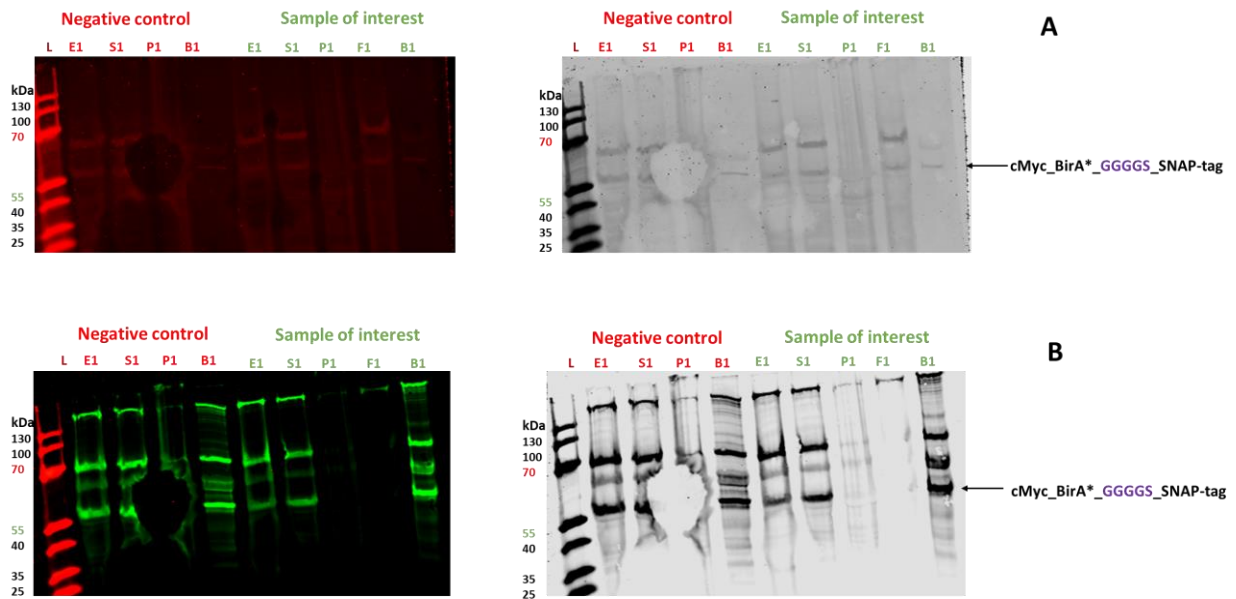


**Figure 65: Second attempt: Purification of candidate DHFR and cMyc\_BirA\*\_GGGGS\_SNAP-tag (Tris-Tricine SDS PAGE 10%).** Part A: the membrane was incubated with anti-cMyc antibodies, and Part B: the membrane was incubated with streptavidin dye ((IRDye 800CW Streptavidin). Parasites were incubated with BG-PEG<sub>4</sub>-Mtx (50  $\mu$ M) (only for the sample of interest), tetracycline (1 $\mu$ g/mL), and an excess of biotin (50  $\mu$ M) added 24 h before the pull-down. Following the overnight biotinylation, the subsequent steps were carried out identically to the previously performed ones, using the same designations E1, S1, P1, F1, and B1.

Upon observing an air bubble on the S1 and B1 fractions of the negative control membrane (Figure 65), it was decided to perform an additional test to improve the visualization of the bands and ensure the reproducibility of results.

This additional test increased the number of repeats and improved the accuracy of the results. The pull-down operation is deemed successful based on the evident clarity and distinction of the bands observed in fraction B1 of both the negative control and the sample of interest displayed in part B of Figure 66. The particularly noteworthy bright band at approximately 55 kDa is indicative of the biotinylated fusion protein, confirmed by its accurate molecular weight and detection by the anti-cMyc antibodies in part A of the same figure. The presence of this band further supports the successful capture of the tool of interest.

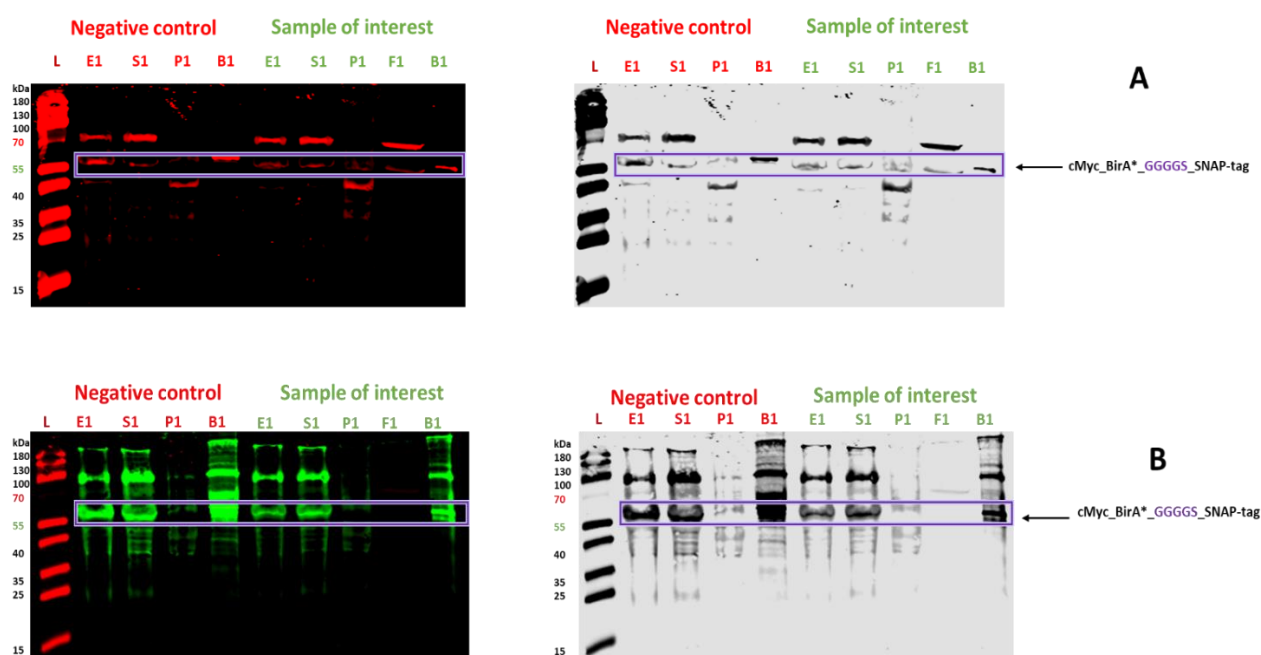


**Figure 66: Third attempt: Purification of candidate DHFR and cMyc\_BirA\*\_GGGGS\_SNAP-tag (Tris-Tricine SDS PAGE 10%).** Part A: the membrane was incubated with anti-cMyc antibodies, and Part B: the membrane was incubated with streptavidin dye ((IRDye 800CW Steptavidin). Parasites were incubated with BG-PEG<sub>4</sub>-Mtx (50  $\mu$ M) (only for the sample of interest), tetracycline (1  $\mu$ g/mL), and an excess of biotin (50  $\mu$ M) added 24 h before the pull-down. Following the overnight biotinylation, the subsequent steps were carried out identically to the previously performed ones, using the same designations E1, S1, P1, F1, and B1.

A third test was conducted using 10% Tris-Tricine gels to enable differentiation between proteins with similar molecular weights, specifically chaperone proteins and DHFR-TS, both of which weigh approximately 60 kDa (180). Indeed, this type of gel allows the production of thinner and more distinguishable bands. Thus, represent an improvement in the protocol.

The absence of the tool in the P1 fraction (Figure 66) clearly indicates that it was expressed exclusively in the parasite's cytosol and not in the subcellular compartment. furthermore, examination of the membranes from the three attempts demonstrates no apparent dissimilarities between the negative control and the sample of interest. These findings reinforce the idea that a highly flexible linker, such as the GGGGS linker, is unsuitable for this investigation. This conclusion aligns with the predictions made by the computational analysis, which proposed that the formation of an agglomerate complex and subsequent masking of the SNAP-tag active site was probable. For the seek of further testing the other elements of

the protocol, additional experiments with electroporated strains were nonetheless performed and the results can be found below.

#### **.3.7.6.2 Electroporated parasites conditions**

Subsequently, the study was conducted using electroporated strains to compare the efficiency of capture with non-electroporated strains. As with the tool with the GGGG linker (*Tbb\_G1*, *Tbb* carrying the cMyc\_BirA\*\_GGGS\_SNAP-tag gene), the results obtained with electroporated (Figures 64, 65 and 66) or non-electroporated (Figure 67) *Tbb\_G2* (*Tbb* carrying the cMyc\_BirA\*\_GGGS\_SNAP-tag gene) in the presence or absence of the compound of interest are similar. Indeed, no significative differences were observed between non-electroporated and electroporated strains in the presence of the compound (Figures 64, 65, 66 and 67). This result confirms the conclusion drawn already using the non-electroporation conditions suggesting that with the GGGG linker the Cys145 within active site of SNAP is not accessible to BG-PEG<sub>4</sub>-Mtx because of the formation of a possible SNAP-BirA\*agglomerate as suggested by the computational analysis (Section .3.1.3).

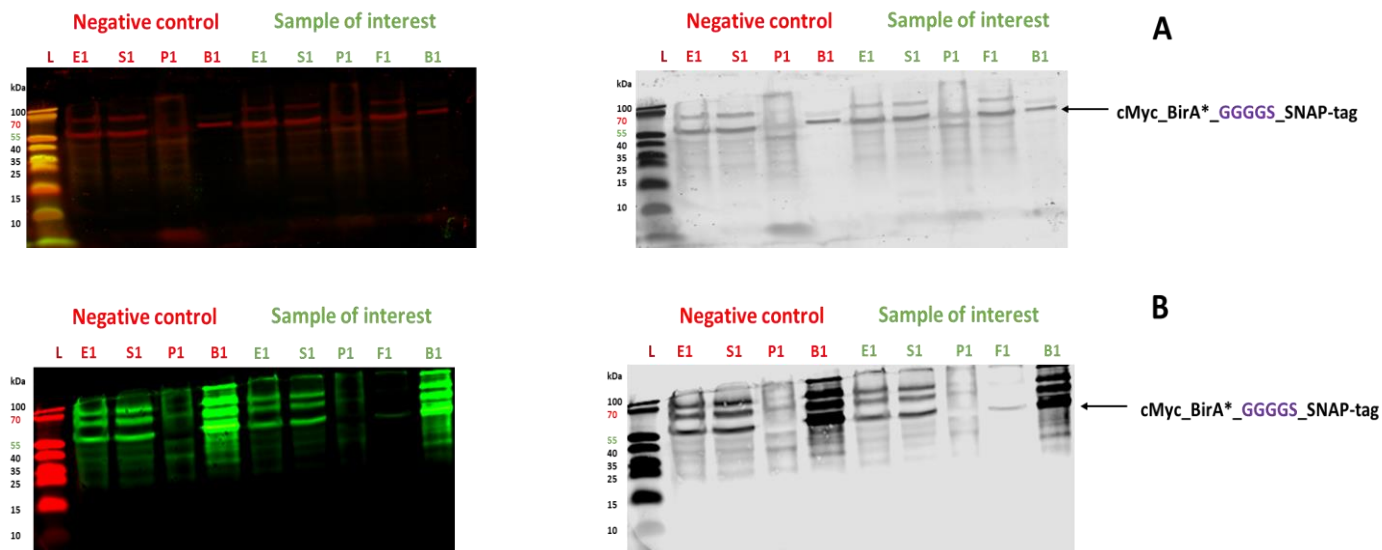


**Figure 67: Purification of candidate DHFR and cMyc\_BirA\*\_GGGGS\_SNAP-tag (Tris-tricine gel SDS PAGE 10%).**

**Part A:** the membrane was incubated with anti-cMyc antibodies, and **Part B:** the membrane was incubated with streptavidin dye ((IRDye 800CW Streptavidin). Parasites were incubated with BG-PEG<sub>4</sub>-Mtx (**50  $\mu$ M**) (only for the sample of interest), tetracycline (**1 $\mu$ g/mL**), and an excess of biotin (**50  $\mu$ M**) added 24 h before the pull-down. Following the overnight biotinylation, the subsequent steps were carried out identically to the previously performed ones, using the same designations E1, S1, P1, F1, and B1.

Furthermore, MS analysis did not detect the presence of biotinylated DHFR-TS or any modification of Cys145 by the BG of the BG-PEG<sub>4</sub>-Mtx compound. This was expected as there was no corresponding band in the western blot results for the target of interest.

Several experiments were conducted, including increasing the volume of culture media, modifying the pull-down protocol, and using the electroporation method. None of these experiments with either the G1 (GGGS) or G2 (GGGGS) linker were able to yield conclusive and convincing results.

In conclusion, by increasing the number of amino acids, the flexibility of the linker increases, which seems to favor the formation of a complex between the BirA\* domain and the SNAP-tag domain, which prevents the BG-PEG<sub>4</sub>-Mtx derivative from binding to the active site of the SNAP-tag domain. The experimental findings are in agreement with the results obtained with the computational approach.

In conclusion, initially, more flexible linkers than the properly working GGS linker and chosen according to Xiaoying Chen's paper (169) were tested, but, neither western blot analysis of

the biotinylated proteins nor the MS demonstrated the presence of biotinylated DHFR-TS target selected as a proof-of-concept. Furthermore, the covalent binding of the BG moiety of the BG-PEG<sub>4</sub>-Mtx derivative to the Cys145 of the SNAP-tag active site could never be identified by MS, confirming the results of the computational approach assuming that the BirA\* module would hide the SNAP active site due to the flexibility of the linker used (linkers G1 & G2) which would increase the mobility of BirA\* and therefore prevent the accessibility of BG-PEG<sub>4</sub>-Mtx to its active site, resulting to a loss of biological activity of the fusion protein of interest.

Despite the many advantages of flexible linkers including the guarantee of a certain degree of mobility of the fusion protein, the lack of rigidity can be a limitation (169). A good example of loss of biological activity linked to too high flexibility is the one of the fusion protein composed of the G protein and *Vargula* luciferase (183). The introduction of a flexible linker between these two domains, consisting of 5 amino acids GGGGS, did not recover the binding capacity of the G protein to immunoglobulins (169, 183).

Moreover, as we also observed with our study, low expression yields and even expression failures have been observed when a flexible linker composed of (GSSSS)<sub>3</sub> was inserted between granulocyte colony-stimulating factor (G-CSF) and transferrin (Tf) (169, 184, 185).

In both examples, the flexible linkers did not allow adequate and efficient separation of the different domains composing the fusion proteins. It is in such situations that rigid linkers might be more appropriate to maintain some fixed separation between the different domains in order to preserve each of their biological activities (169) (As predicted in Section 3.1.3, especially with PTP and PAP linkers).

However, the lack of conclusive MS results for the G0 linker (GGS) (as a reminder, the western blot results and especially those obtained with electroporated strains were convincing and showed successful capture of DHFR-TS) may be related to a limited amount of the available target that could be detected.

In line with the idea of exploring the effect of a more rigid linker, one potential linker to be evaluated is based a combination of amino acids known leading to restricted flexibility, specifically either prolines and alanine or proline and threonine as mentioned by Xiaoying Chen and coworkers (169) . Based on computational analysis, it was determined that the

PAPAP linker, despite its rigidity, would not impede the generation of the SNAP-BirA\* agglomerate. However, the PTP linker would prevent any complex formation, leaving the SNAP-tag active site exposed to BG-PEG<sub>4</sub>-Mtx. Despite its high FCC value, it was nevertheless decided to test the PAPAP rigid linker in order to verify the hypothesis raised in the mentioned study (in Section 3.1.3) and also to be able to compare the results with those obtained with the GGGGS flexible linker containing the same number of amino acids. Thus, to validate the computational study's hypotheses and seeking to further ameliorate the target deconvolution tool, both the PTP and PAPAP linkers were examined and evaluated through pull-down tests and MS analyses. The outcomes of these evaluations are presented in .3.7.7 below.

### **.3.7.7 Biotinylation of DHFR-TS by the cMyc\_BirA\*\_PAPAP\_SNAP-tag tool**

A biotinylation study was conducted to evaluate the efficiency of DHFR-TS capture by the cMyc\_BirA\*\_PAPAP\_SNAP-tag tool. This study included both non-electroporated and electroporated strains, following the same methodology as previously performed.

#### **.3.7.7.1 Non-electroporated parasites conditions**

Figure 68 displays the results of the initial biotinylation study conducted on non-electroporated strains containing the cMyc\_BirA\*\_PAPAP\_SNAP-tag gene to assess DHFR-TS capture:

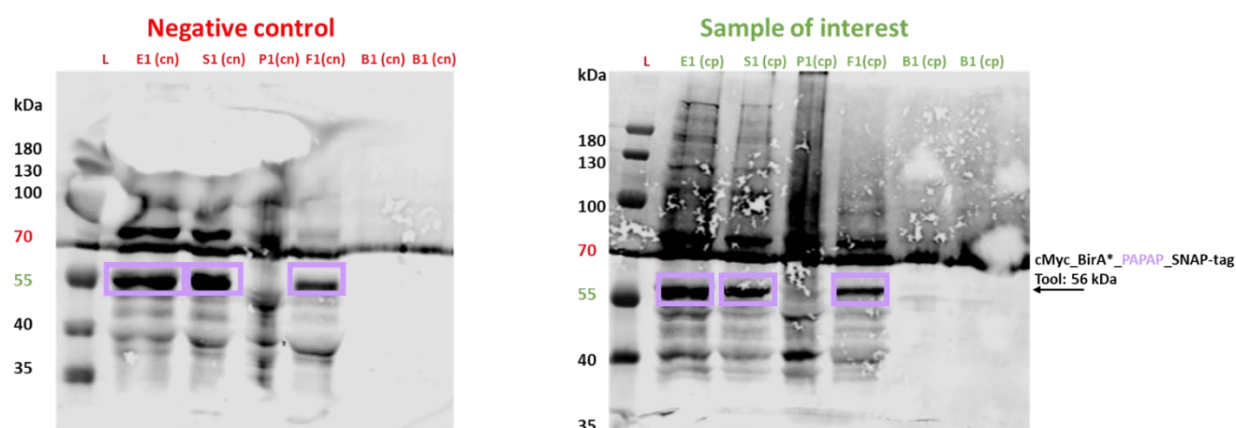


**Figure 68: Purification of candidate DHFR and cMyc\_BirA\*\_PAPAP\_SNAP-tag (Gel Tris-tricine SDS-PAGE 8%).** The two membranes were incubated with streptavidin dye ((IRDye 800CW Streptavidin). Parasites were incubated with 50  $\mu$ M of BG-PEG<sub>4</sub>-Mtx (only for the sample of interest), tetracycline (1  $\mu$ g/mL), and an excess of biotin (50  $\mu$ M) added 24 h before the pull-down. Following the overnight biotinylation, the subsequent steps were carried out identically to the previously performed ones, using the same designations E1, S1, P1, F1, and B1.

As illustrated in Figure 68, only the biotinylated tool is visible in fractions E1, S1, and F1 on the membranes of the two different conditions. The tool's presence in the flow-through fraction (F1) could indicate either saturation of the beads from capturing biotinylated proteins, resulting in an excess appearing in the eluate, or poor capture of biotinylated proteins, possibly due to a handling error. However, this outcome is unexpected as previous results did not show such a strong band in the eluate, suggesting a greater likelihood of a handling error, such as insufficient incubation time with the presence of the beads.

It is evident that there is minimal visibility in the two B1 fractions that contain the proteins extracted from the beads. The only visible protein is the streptavidin tetramer around 70 kDa, which is related to its detachment from the beads. This protein's presence is likely due to errors during handling, particularly inadequate bead washing. Although 0.1% SDS was added to the wash buffer to dissolve the proteins, it caused the streptavidin to detach from the beads, as illustrated in Figure 68. Further examination revealed that SDS could cause the proteins to detach from the beads (according to the general guidelines Dynabeads M-280 Streptavidin; Invitrogen; Cat N°11205D) (148). To address this issue, 0.1% BSA was added to the wash buffer (PBS or 1X TBS), which effectively resolved the observed outcomes.

Several other attempts were made (results not presented as they were comparable to those shown in Figure 68), but unfortunately, no other proteins, especially DHFR-TS, were detected except for the target tool. However, an experiment was conducted on electroporated parasites, and the outcomes are presented hereafter.

#### **.3.7.7.2 Electroporated parasites conditions**

Figure 69 presents the results of the biotinylation study conducted on electroporated strains carrying the cMyc\_BirA\*\_PAPAP\_SNAP-tag gene in order to evaluate DHFR-TS capture. The goal was to confirm any potential increase in DHFR-TS detection:

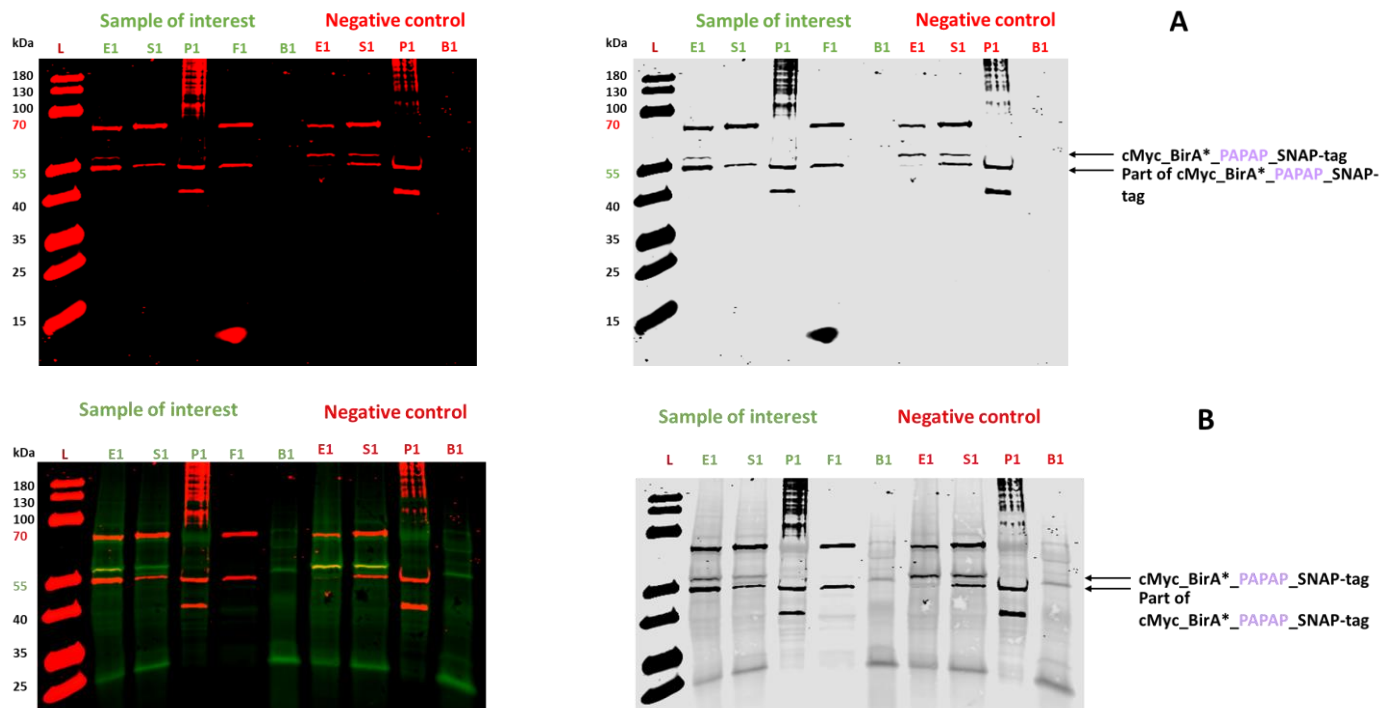


**Figure 69: Purification of candidate DHFR and cMyc\_BirA\*\_PAPAP\_SNAP-tag (Tris-tricine gel SDS PAGE 10%).**

**Part A:** the membrane was incubated with anti-cMyc antibodies (**Red** fluorescence), and **Part B:** the membrane was incubated with streptavidin dye ((IRDye 800CW Streptavidin). Parasites were incubated with BG-PEG<sub>4</sub>-Mtx (**50  $\mu$ M**) (only for the sample of interest), tetracycline (**1 $\mu$ g/mL**), and an excess of biotin (**50  $\mu$ M**) added 24 h before. Following the overnight biotinylation, the subsequent steps were carried out identically to the previously performed ones, using the same designations E1, S1, P1, F1, and B1.

To begin with, it is evident that several bands are visible on the gel (part A of Figure 69), and these are the outcome of anti-cMyc antibodies' unspecific binding. However, despite this, the band around 55 kDa, which corresponds to the target tool, is clearly visible in both parts A and B. After revelation with streptavidin dye, the green and red signals overlap to produce an orange band around 55 kDa, indicating the presence of the biotinylated system. Additionally, the absence of the target band (which should be around 60 kDa) is notable, along with a weak band of the system in both B1 fractions, corresponding to the negative control and the sample of interest. These findings suggest that the beads failed to capture the biotinylated proteins appropriately.

In addition, there is no observable difference between the negative control and the sample of interest. Despite its rigidity, the PAPAP linker had being predicted by the FCC analysis to have a high likelihood of forming complexes between BirA\* and SNAP-tag domains. The results

confirm this prediction. This occurrence prevented the compound of interest, BG-PEG<sub>4</sub>-Mtx, from accessing the active site. Moreover, the time-dependent protein expression study (section 3.6.6) showed weak overexpression of the tool (only 2%), which explains the faint bands on the western blot membranes.

In summary, the rigid PAPAP linker did not enhance the fusion protein's performance since DHFR-TS could not be captured, confirming the computational study's predictions (refer to Section 3.1.3).

As previously mentioned and illustrated in Figure 69 above, the bands observed in the two fractions B1 were considerably less intense in comparison to those observed in fraction S1 for both the negative control and the sample of interest. As a result, it was decided that only the S1 fractions from both conditions, after electrophoretic migration and excision of bands between 55 and 60 kDa, would be sent for MS analysis. The MS results (Figure 70) confirmed the presence of the tool. Indeed, LC-MS/MS analysis successfully identified cMyc\_BirA\*\_PAPAP\_SNAP-tag fusion protein, revealing 21 total spectra and 15 unique peptides, providing a protein sequence coverage of 40% (Figure 70). Unfortunately, neither peptides containing the Cys145 of interest nor biotinylated or BG-PEG<sub>4</sub>-Mtx modified peptides were detected.

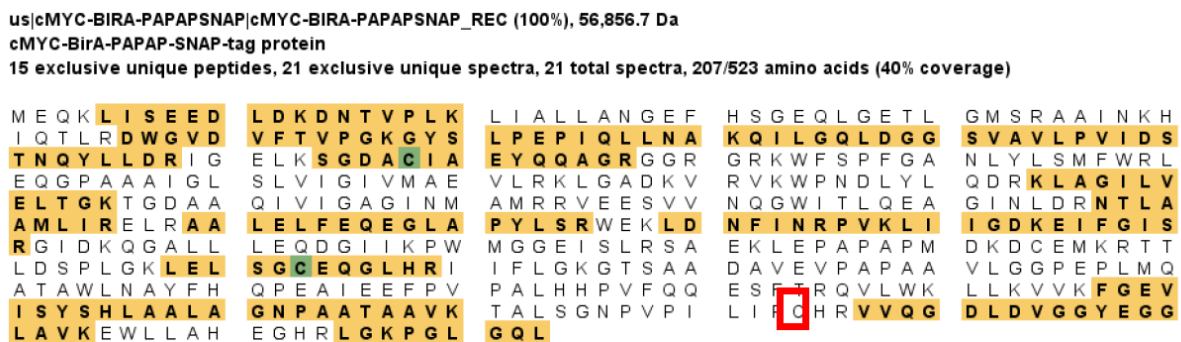


**Figure 70:** Identification of cMyc\_BirA\*\_PAPAP\_SNAP-tag peptides by LC-MS/MS analysis: The tool was detected with 21 total spectra, 15 unique peptides and the protein coverage was 40% (analyses made by the proteomic platform of Dr. Alexandre Hainard and his team)

The culture media were scaled up to 500 mL for each strain expressing the different tools. However, in this particular case, DHFR-TS was only detected by MS in *Tbb* strain expressing the cMyc\_BirA\*\_PAPAP\_SNAP-tag tool (Figure 71). Although DHFR-TS was observed in sample

S1, which corresponds to the pull-down supernatant (Figure 71), unfortunately, no biotinylated peptide associated with DHFR-TS could be identified. This further validates the previous findings where DHFR-TS was not detectable by western blot analysis, regardless of whether the parasites underwent electroporation in the presence of BG-PEG<sub>4</sub>-Mtx (Figures 68 and 69). Consequently, the MS-based detection of DHFR-TS is likely a result of the solubilization of cytosolic proteins present in the supernatant, wherein DHFR-TS was detected alongside other proteins within the fraction.

**Q582G3 (100%), 58,852.0 Da**

**Bifunctional dihydrofolate reductase-thymidylate synthase OS=Trypanosoma brucei brucei (strain 927/4 GUTat10.1) OX=185431 GN=Tb9  
17 exclusive unique peptides, 19 exclusive unique spectra, 19 total spectra, 246/527 amino acids (47% coverage)**

|                     |                     |                     |                     |                     |
|---------------------|---------------------|---------------------|---------------------|---------------------|
| M L S L T R I L R K | K I P V H E L A G K | I S R P P L R P F S | V V V A S D E K G G | I G D G G T I P W E |
| I P E D M Q Y F R R | V T T N L R G K N V | K P S P S K R N A V | V M G R K T W D S L | P P K F R P L S N R |
| L N V V L S R S A T | K E Q L L A G I P D | P I K R A E A A N D | V V A V N G G L E D | A L R M L V S K E H |
| T S S I E T V F C I | G G G T I Y K Q A L | C A P C V N V L Q A | I H R T V V R P A S | N S C S V F F D I P |
| A A G T K T L E G L | E L V R E S I T D E | R V S T G A G G K K | Y Q F E K L V P R N | S E E E Q Y L N L V |
| G R I I D E G C T K | C D R T G V G T R S | L F G A Q M R F S L | R N N R L P L L T T | K R V F W R G V C E |
| E L L W F L R G E T | N A K L L S D K G I | H I W D G N G S R A | F L D S R G L T D Y | D E M D L G P V Y G |
| F Q W R H F G A D Y | I S C K E D S E G K | G V D Q I A N I V K | S L I E N P D D R R | M I C T A W N P A A |
| L P R M A L P P C H | M M A Q F Y V S N G | E L S C M L Y Q R S | C D M G L G V P F N | I A S Y A L L T F L |
| M A K A S G L R P G | E L V H T L G D A H | V Y S N H V E P C R | K Q L K R V P R P F | P F I V F K Q D K E |
| F L E D F Q E S D I | E V I D Y S P Y P V | I S M E M A V       |                     |                     |

**Figure 71:** Identification of DHFR-TS peptides in supernatant S1 by LC-MS/MS analysis: The target was detected with 19 total spectra, 17 unique peptides and the protein coverage was 47% (analyses made by the proteomic platform of Dr. Alexandre Hainard and his team)

Based on the results, it can be concluded that the PAPAP linker enhances the likelihood of forming a ternary complex among the fusion protein's domains. This complex masks the SNAP-tag active site, obstructing access to BG-PEG<sub>4</sub>-Mtx and ultimately leading to the non-capture of the DHFR-TS target.

Testing a single linker implies a substantial amount of work, as the molecular biology steps including cloning, bacterial transformation, validation of the plasmid of interest as well as the parasite transfection steps including clone selection followed by verification of correct gene insertion in the selected strain are all time-consuming and labor-intensive. Furthermore, the pull-down assays of biotinylated proteins and their identification via western blot and MS must be repeated for each linker. However, since the modeling approach was employed beforehand (as described in Section 3.1.3) to predict the most suitable linker, it was possible to avoid the significant workload of repeating the experiments. After experimentally confirming the suitability of the GGS, GGGs, and PAPAP linkers based on theoretical

predictions, only the PTP linker remains to be tested among the potentially effective ones. The following section presents the results of this testing.

#### *.3.7.8 Biotinylation of DHFR-TS by the cMyc\_BirA\*\_PTP\_SNAP-tag tool*

As previously mentioned, computational studies have shown that the formation of an agglomerate between BirA\*, SNAP-tag, and the linker may result in steric interference, leading to a loss of biological functions of the fusion protein components. Additionally, the active site of the SNAP-tag may be masked, hindering substrate binding. This occurrence becomes even more likely as the length of the linker increases, regardless of its amino acid composition. Experimental evidence confirms that linkers GGGGS and PAPAP, both composed of the same number of amino acids but with different compositions, yielded inconclusive results and low expression levels of the tool. However, with a rigid linker composed of Proline (P) and Threonine (T) (PTP), the probability of forming the complex is null and therefore would maintain the proper functioning of the tool. This linker has been experimentally tested in both non-electroporated and electroporated cell lines expressing cMyc\_BirA\*\_PTP\_SNAP-tag and the results are presented below.

##### *.3.7.8.1 Non-electroporated parasites conditions*

The crucial observation is the overexpression of the protein of interest and its efficient capture by the beads, as evidenced by the intense band at approximately 55 kDa in fraction B1 under both experimental conditions (Figure 72).

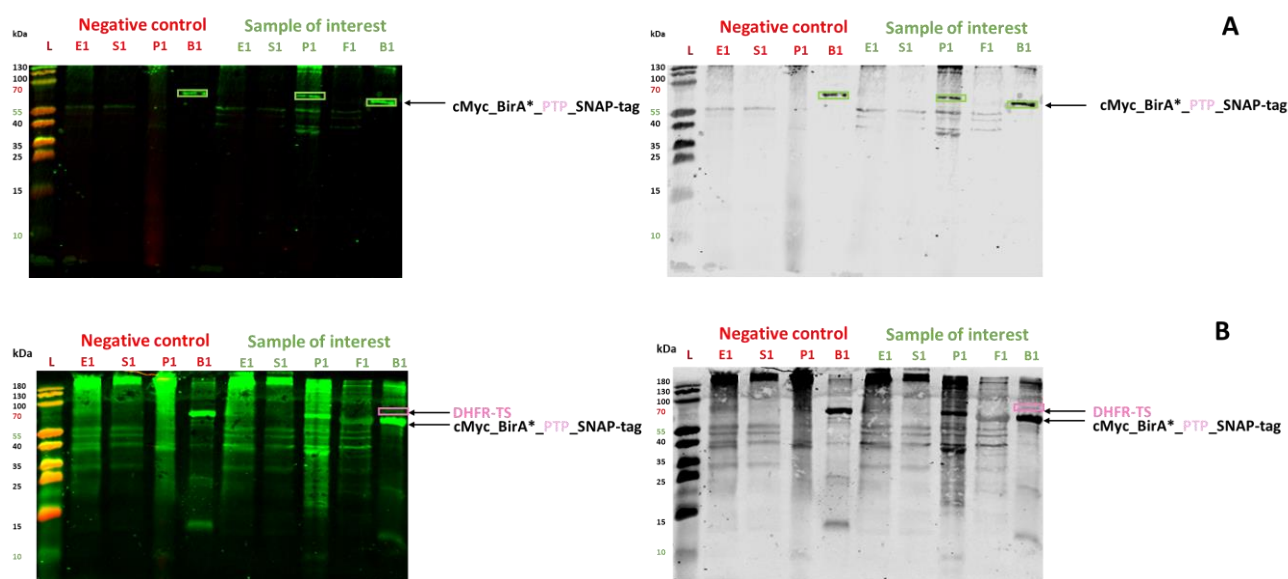


**Figure 72: Purification of candidate DHFR and cMyc\_BirA\*\_PTP\_SNAP-tag (Tris-tricine gel SDS PAGE 15%).** Part A: the membrane was incubated with anti-cMyc antibodies (green fluorescence), and Part B: the membrane was incubated with streptavidin dye ((IRDye 800CW Streptavidin). Parasites were incubated with BG-PEG<sub>4</sub>-Mtx (50  $\mu$ M) (only for the sample of interest), tetracycline (1 $\mu$ g/mL), and an excess of biotin (50  $\mu$ M) added 24 h before. Following the overnight biotinylation, the subsequent steps were carried out identically to the previously performed ones, using the same designations E1, S1, P1, F1, and B1.

However, a slightly less prominent band at around 59 kDa in fraction B1 of the sample of interest, absent in the negative control, is likely to correspond to the desired target protein DHFR-TS (Figure 72). The low intensity of this band could be due to either a low amount of target protein or the fluorescence emitted by the tool masking the target protein's fluorescence, as can happen when one protein emits significantly more fluorescence than others (186).

In comparison to the band observed under identical conditions with the G0 linker (GGs) (refer to Figures 52 and 53 in Section .3.7.1), the band potentially corresponding to DHFR-TS is notably less intense. Additionally, considering the molecular weight of the heat shock protein HSP60 (60 kDa), the band prominently visible in the B1 fraction of the negative control is likely associated with that protein.

To further investigate, a subsequent experiment was conducted solely on the B1 fraction of both the negative control and the sample of interest. Figure 73 below depicts the outcome of this experiment.

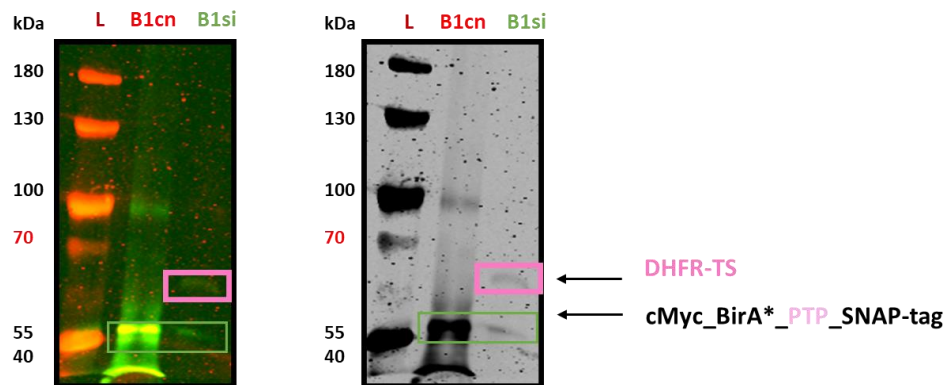


*Figure 73: Purification of candidate DHFR and cMyc\_BirA\*\_PTP\_SNAP-tag (Tris-tricine gel SDS PAGE 8%). Membrane incubated with anti-cMyc (red) and IRDye 800CW Streptavidin (green). The fraction B1 is the bound cytoplasmic eluate and should contain the biotinylated tool, and eventually biotinylated near neighbors.*

The above figure clearly shows that both the sample of interest (**B1si**) and the negative control (**B1cn**) demonstrate the presence of the tool through proper capture by the beads. However, the observed differences in intensity are most likely due to the varying concentrations of proteins in the two samples. A low-intensity band, which is absent in the negative control, is detected in the B1 fraction of the sample of interest. Based on its molecular weight (58-59 kDa), this band expectedly corresponds to the target of interest DHFR-TS. These results demonstrate that the tool is functioning efficiently at a level comparable to the initial tool with the G0 linker (GGs linker).

#### .3.7.8.2 Electroporated parasites conditions

To evaluate the effectiveness of DHFR-TS capture, transfected strains carrying the cMyc\_BirA\*\_PTP\_SNAP-tag gene underwent electroporation, followed by a subsequent biotinylation study. The aim was to verify the potential increase in DHFR-TS capture. The results are presented in Figure 74 below.

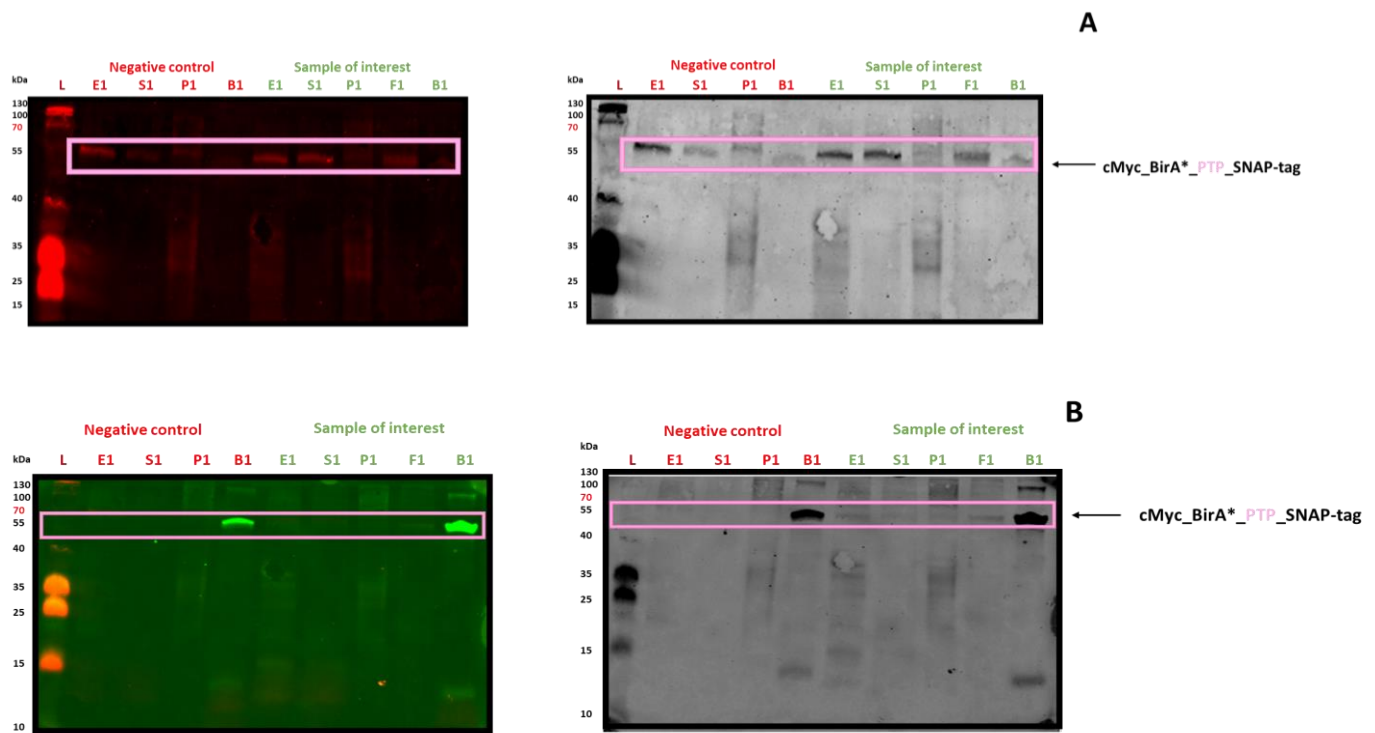


**Figure 74: Purification of candidate DHFR and cMyc\_BirA\*\_PTP\_SNAP-tag (Tris-tricine gel SDS PAGE 15%).**

**Parts A & B: not the same membranes:** Part A: Membrane incubated with anti-cMyc antibodies (**Red** fluorescence); Part B: Another membrane incubated with IRDye 800CW Streptavidin (**green**). Parasites were treated with BG-PEG<sub>4</sub>-Mtx (100  $\mu$ M) (added only for the sample of interest), tetracycline (1 $\mu$ g/mL), and excess biotin (50  $\mu$ M) (added 24 h prior). Following the overnight biotinylation, the subsequent steps were carried out identically to the previously performed ones, using the same designations E1, S1, P1, F1, and B1.

What can be noticed is the tool's appropriate expression in all of the fractions (part A of the figure above) and the streptavidin beads' successful capture of it (B1, part B of Figure 74).

Nevertheless, because the separation of the various proteins above 55 kDa is not particularly apparent, the DHFR is not discernible; for this reason, a second experiment using 8% acrylamide gel was conducted, and the results are shown in Figure 75.

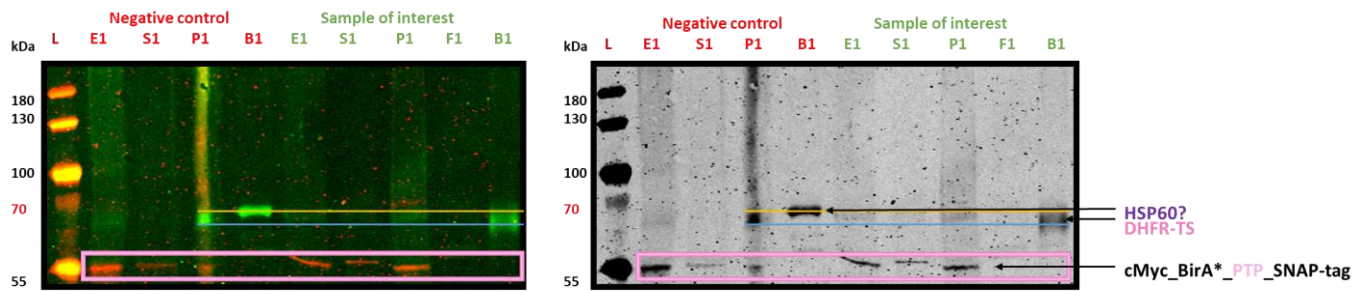


**Figure 75: Purification of candidate DHFR and cMyc\_BirA\*\_PTP\_SNAP-tag (Tris-tricine gel SDS PAGE 8%).** Membrane incubated with anti-cMyc (red) and IRDye 800CW Streptavidin (green). Parasites were treated with BG-PEG<sub>4</sub>-Mtx (100  $\mu$ M) (added only for the sample of interest), tetracycline (1 $\mu$ g/mL), and excess biotin (50  $\mu$ M) (added 24 h prior). Following the overnight biotinylation, the subsequent steps were carried out identically to the previously performed ones, using the same designations E1, S1, P1, F1, and B1.

The red bands evident in all fractions of both conditions at approximately 55 kDa in the above image (framed in pink) amply demonstrate the proper expression of the tool. A distinct separation of proteins is observable in the membrane above. This was achieved by reducing the acrylamide percentage in the gel and applying a slower migration rate of 180V over 3:30 hours.

It is noteworthy that a band at approximately 60 kDa is not present in the sample of interest, whereas it is evident in B1 of the negative control. Given its molecular weight, this band may correspond to HSP60. The existence of a second, extremely bright band between 58 and 59 kDa, which generates a greater fluorescence and can cover that of other bands, is thought to be the cause of HSP60's absence in the B1 fraction of the interest sample (Figure 75). Given its molecular weight of 58–59 kDa, the latter, detected in the B1 fraction of the sample of interest would most likely be the biotinylated DHFR-TS.

In conclusion, the first finding is that biotinylated DHFR-TS was successfully captured by both electroporated and non-electroporated parasites. Second, the most conclusive and promising results came from the two tools, cMyc\_BirA\*\_**GGG**\_SNAP-tag and cMyc\_BirA\*\_**PTP**\_SNAP-tag. Thereby, these ought to be utilized in future studies with other BG-PEG<sub>4</sub> derivatives.

### .3.8 MS detection of the Cys145 modification of the SNAP-tag active site by BG-PEG<sub>4</sub>-Mtx

In order to comprehend how BG-PEG<sub>4</sub>-Mtx interacts with SNAP, the following experiment was carried out: Pure SNAP-tag protein was incubated with the BG-PEG<sub>4</sub>-Mtx derivative under two different conditions: The first condition in which 10  $\mu$ M of pure protein were incubated with 10  $\mu$ M of the compound (1:1 equivalence) for 1 h at 37°C and the second, where 10  $\mu$ M of protein were incubated with 50  $\mu$ M of the compound (1:5 equivalence) for 3 h at 37°C (unlike the experiment described in 3.7.2 above where the incubation was only for 1 h at 37°C). Then, a run in SDS-PAGE followed by gel staining with Coomassie blue, a cut of the appropriate band, and in-gel protein digestion with trypsin (this last step, as well as the MS analysis, were performed by the proteomic platform (Dr. Alexandre Hainard and his team, proteomic platform, CMU, Geneva)) were performed. The results of the LC-MS/MS analysis (Figure 76 below) showing a coverage of 60 % demonstrate the presence of the peptide of interest containing the Cys145 of the SNAP-tag active bound to BG-PEG<sub>4</sub>-Mtx. However, as already seen in Section 3.7.2, this detection was only visible after lowering the identification score thresholds and thus, the peptides of interest carry low probability scores not allowing the validation of the binding. Two theories could account for these findings: the first is that BG-PEG<sub>4</sub>-Mtx is difficult to identify by MS, and the second hypothesis proposes that the incorporation of Mtx into the BG-PEG<sub>4</sub> moiety might result in a slight impediment to the molecule's accessibility to the active site of the SNAP tag.

us|SNAPtagPURE|SNAPtagPURE\_REC (100%), 19,683.7 Da  
 SNAP-tag PURE protein  
 12 exclusive unique peptides, 28 exclusive unique spectra, 2183 total spectra, 110/184 amino acids (60% coverage)

|       |       |    |       |        |       |        |       |       |       |        |
|-------|-------|----|-------|--------|-------|--------|-------|-------|-------|--------|
| MDKDC | EMK   | RT | TLDSP | LGLKE  | LSGCE | QGLHR  | IIFLG | KGTSA | ADAVE | VPAPA  |
| AVLGG | PEPLM |    | QATAW | LNAYF  | HQPEA | IEEFP  | VPALH | HPVFQ | QESFT | RQVLW  |
| KLLKV | VKFGE |    | VISYS | HLLAAL | AGNPA | AATAAV | KTALS | GNPVP | ILIP  | CHRVVQ |
| GDLDV | GGYEG |    | GLAVK | EWLLA  | HEGHR | LKGPG  | LGQL  |       |       |        |

*Figure 76: Identification by LC-MS/MS analysis of peptides corresponding to the pure SNAP-tag protein carrying the modification of Cys145 (red box) obtained by its binding with the BG-PEG<sub>4</sub>-Mtx compound: Identified peptides satisfying the minimum peptide threshold are emphasized in yellow, and modifications are marked in green. (analyses carried out by the proteomic platform of Dr. Alexandre Hainard and his team).*

To verify these two hypotheses, both a competition experiment (Figure 77) and a follow up an MS analysis (Figure 78) of the pure protein incubated either with BG-PEG<sub>4</sub>-Mtx or BG-PEG<sub>4</sub>-Biotin, as well as a competition test with BG-PEG<sub>4</sub>-FDA were conducted. The results are shown below:

What is noticeable in the Figure 77 is that when the fluorescent compound BG-PEG<sub>4</sub>-FDA is added after the following compounds have been incubated: BG-PEG<sub>4</sub>-N<sub>3</sub> and BG-PEG<sub>4</sub>-biotin (C & D), the fluorescence decreases slightly in comparison to the pure protein incubated only with BG-PEG<sub>4</sub>-FDA. The result is even more evident when the protein is dimerized or trimerized.

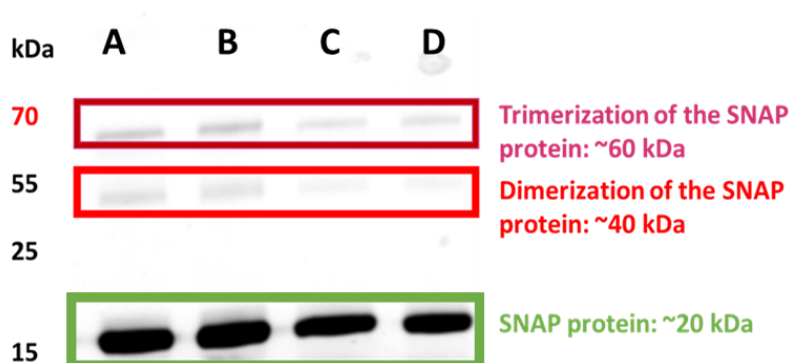


**Figure 77: Analysis of the emitted fluorescence (366 nm):** L: Ladder; A: SNAP incubated with 10  $\mu$ M BG-PEG<sub>4</sub>-FDA for 1 h at 37°C; B: SNAP incubated with 10  $\mu$ M BG-PEG<sub>4</sub>-Mtx for 1 h at 37°C then addition of 10  $\mu$ M BG-PEG<sub>4</sub>-FDA followed by 1 h incubation at 37°C; C: SNAP incubated with 10  $\mu$ M BG-PEG<sub>4</sub>-N<sub>3</sub> for 1 h at 37°C then the addition of 10  $\mu$ M BG-PEG<sub>4</sub>-FDA followed by 1 h incubation at 37°C; D: SNAP incubated with 10  $\mu$ M BG-PEG<sub>4</sub>-biotin for 1 h at 37°C then the addition of 10  $\mu$ M BG-PEG<sub>4</sub>-FDA followed by 1 h incubation at 37°C

However, when SNAP was first incubated with BG-PEG<sub>4</sub>-Mtx and then with the fluorescent compound, the fluorescence intensity of the band was identical to that of SNAP that had been incubated only with BG-PEG<sub>4</sub>-FDA (B). This indicates that the BG-PEG<sub>4</sub>-Mtx derivative accesses the protein's active site with a little more difficulty than BG-PEG<sub>4</sub>-FDA, as BG-PEG<sub>4</sub>-FDA out competed BG-PEG<sub>4</sub>-Mtx.

This experiment shows that BG-PEG<sub>4</sub>-Mtx binds slightly less well to accessible active sites than the other compounds which might explain why the modification of Cys145 by BG-PEG<sub>4</sub>-Mtx could never be identified by MS analysis (Figure 76 above) despite the convincing results

**Part 3: Development of a new target deconvolution tool**

obtained with the cMyc\_BirA\*\_GGs\_SNAP-tag and cMyc\_BirA\*\_PTP\_SNAP-tag tools carrying the “GGs” (GO) or PTP linkers leading to a functional tool.

An additional experiment was conducted to evaluate the hypothesis that MS would have difficulty detecting BG-PEG<sub>4</sub>-Mtx. This experiment involved incubating pure SNAP-tag protein with either BG-PEG<sub>4</sub>-Mtx or BG-PEG<sub>4</sub>-biotin under identical conditions. The description of this experiment can be found in Section 2.2.26 above.

Once the reactions were completed, a gel was run and then stained and the bands of interest were excised. Subsequently, the samples underwent MS analysis. The results are shown below in Figure 78.

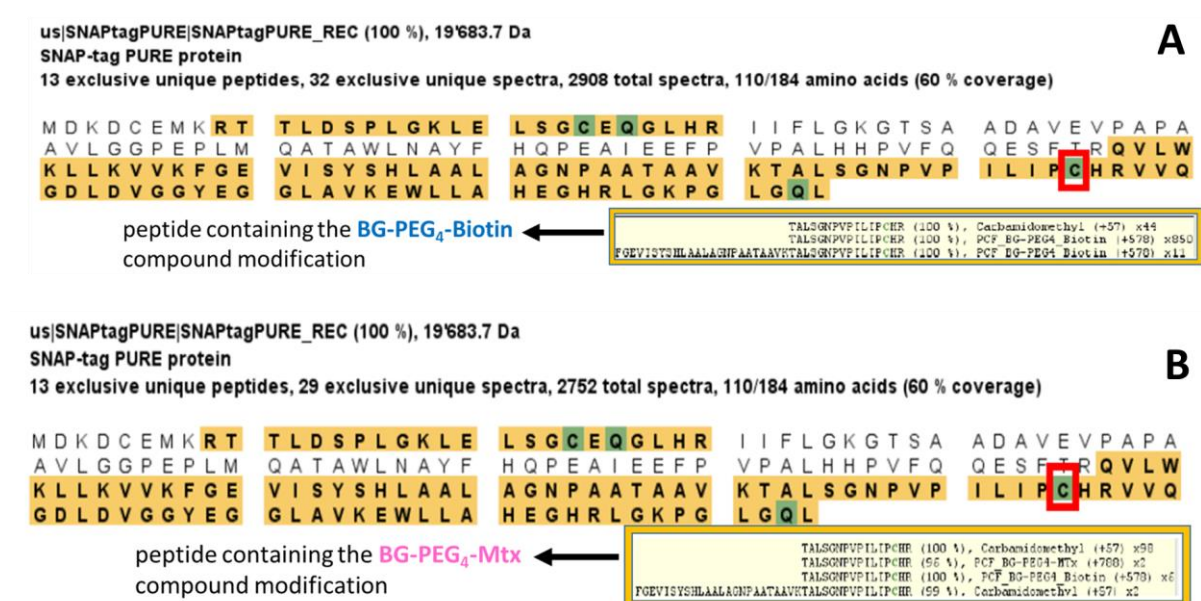


**Figure 78: Part A:** LC-MS/MS analysis of the protein incubated with BG-PEG<sub>4</sub>-Mtx: Identification of peptides corresponding to the pure SNAP-tag protein carrying the modification of the cysteine of interest Cys145 (red box) obtained by its binding with BG-PEG<sub>4</sub>-Mtx, **Part B:** LC-MS/MS analysis of the protein incubated with BG-PEG<sub>4</sub>-Biotin: Identification of peptides corresponding to the pure SNAP-tag protein carrying the modification of the cysteine of interest Cys145 (red box) obtained by its binding with BG-PEG<sub>4</sub>-Biotin. The identified peptides satisfying the minimum peptide threshold are emphasized in yellow, and modifications are marked in green (analyses carried out by the proteomic platform of Dr. Alexandre Hainard and his team). The yellow box emphasizes the number of peptides that include the modified cysteine

The sequence coverage of both samples of pure SNAP protein was 60% after incubation with either BG-PEG<sub>4</sub>-Biotin or BG-PEG<sub>4</sub>-Mtx (Figure 78). However, there was a marked difference

in the number of identified peptides that contained the modification at Cys145. The number of peptides with the target cysteine modified by BG-PEG<sub>4</sub>-Biotin was 861, while there were only 2 peptides with the modification from BG-PEG<sub>4</sub>-Mtx. The issue appears to have arisen during the fragmentation step of the target peptide TALSGNPVPILIP<sup>CHR</sup> [TALSGNPVPILIPC(BG-PEG<sub>4</sub>-Mtx)HR] bearing the BG-PEG<sub>4</sub>-Mtx modification. This fragmentation step is critical for confirming the presence of the compound's binding to the active site. However, for unknown reasons, the fragmentation of this specific peptide failed to result in its identification (further discussion provided below).

It is also worth mentioning that while 44 peptides with a carbamidomethylated variant of Cys145 were found for BG-PEG<sub>4</sub>-Biotin, 100 such peptides were found for BG-PEG<sub>4</sub>-Mtx. It can be inferred from a large number of generated spectra (about 2000) that only a small fraction (100) carried a free (carbamidomethylated) cysteine. This suggests that the remaining unidentified spectra likely carried the modification from BG-PEG<sub>4</sub>-Mtx, but were not assigned due to difficulties encountered during the fragmentation process (see Appendix 10 for more details)

Several factors could explain the issue with BG-PEG<sub>4</sub>-Mtx. The primary one is that Mtx is highly charged, resulting in the accumulation of too many electrons, making efficient fragmentation difficult.

The peak with the highest intensity at  $z=4$  is quadruply charged, but its fragmentation did not result in clear identification. The two detected peptides originate from the peak at  $z=2$ , which has a low intensity and, therefore, a low probability of identification (Figure 79). Another possible factor is the greater molecular weight of BG-PEG<sub>4</sub>-Mtx (939.99 g/mol) compared to BG-PEG<sub>4</sub>-Biotin (729.33 g/mol). However, the absence of visible MS evidence for the binding of BG-PEG<sub>4</sub>-Mtx to Cys145 after the fragmentation step does not necessarily mean that the modification is absent in the SNAP-tag's active site.

Suzanne Sherihan SAHRAOUI  
**Part 3: Development of a new target deconvolution tool**

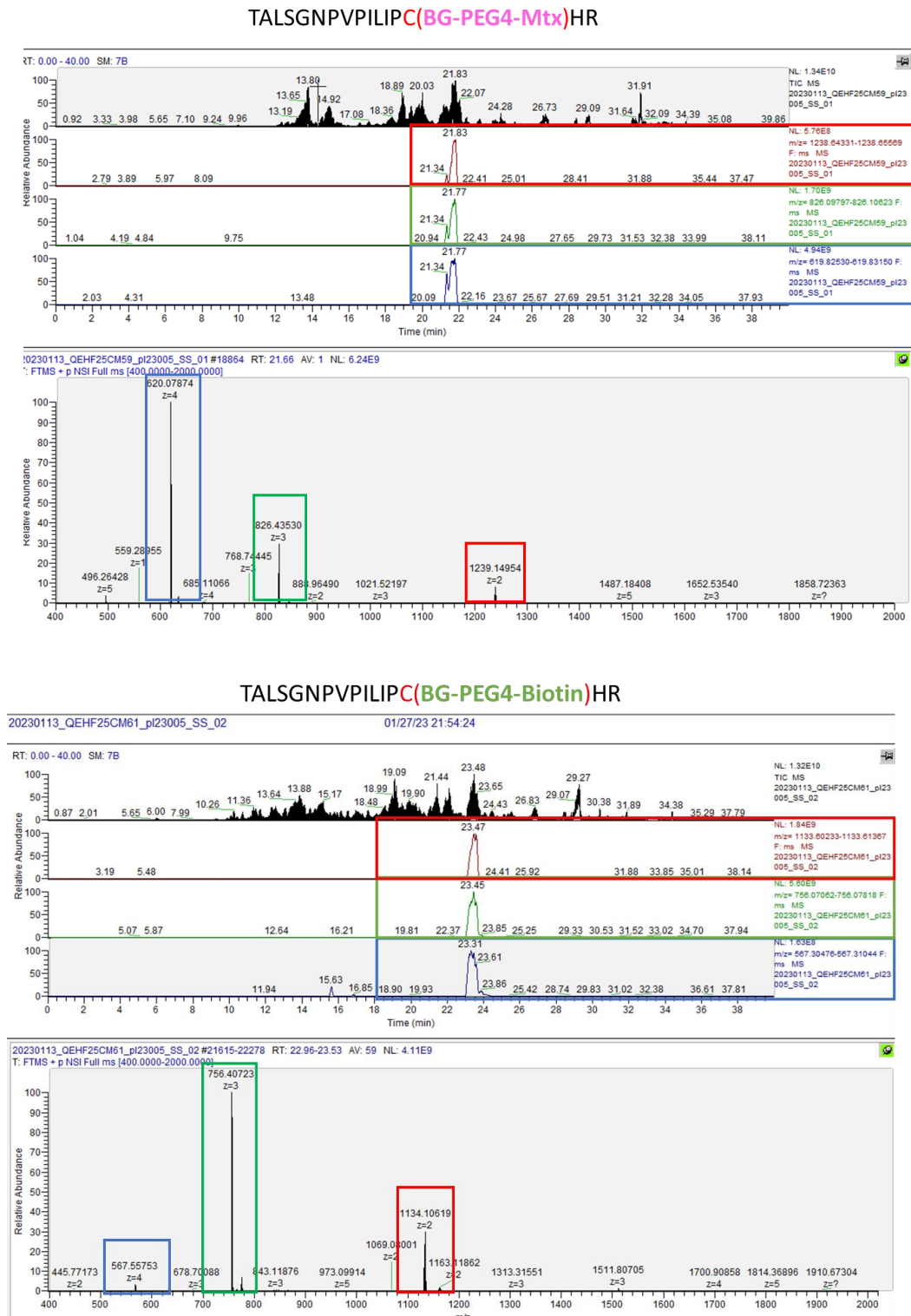


**Figure 79:** MS Raw data and XIC analysis of the peptide of interest [TALSGNPVPILIP<sup>C</sup>HR] modified by either BG-PEG<sub>4</sub>-Mtx (A) or BG-PEG<sub>4</sub>-Biotin (B): Analysis undertaken prior to the fragmentation stage

Indeed, as presented in Figure 79 above, the interest's peptides containing either BG-PEG<sub>4</sub>-Mtx or BG-PEG<sub>4</sub>-Biotin exhibit extremely comparable patterns in the MS raw data prior to the fragmentation phase (Figure 79). In fact, the peaks appear virtually simultaneously (21 and 23 min respectively for BG-PEG<sub>4</sub>-Mtx and BG-PEG<sub>4</sub>-Biotin). This demonstrates that the binding did occur despite the assignment being challenging following the fragmentation stage for an unforeseen reason.

In conclusion, the competition experiment with the other derivatives followed by MS analysis (Figure 78 and Figure 79) showed that BG-PEG<sub>4</sub>-Mtx binds to Cys145 albeit less effectively than BG-PEG<sub>4</sub>-Biotin.

It's also crucial to remember that Mtx cannot be easily detected by MS due to an unclear issue that restricts the compound's identification. The possibility of similar occurrences with other Mtx-like molecules highlights the necessity for an alternative MS-based analytical technique, orthogonal to western blotting. Such techniques may involve the use of other ionization methods like MALDI (MALDI-TOF), or alternative fragmentation techniques to HCD (higher energy collisional dissociation), such as collision-induced dissociation (CID) or electron transfer dissociation (ETD) (187). Additionally, it may be beneficial to optimize the collision energy used for the fragmentation of the modified peptide. All of these potential approaches should be taken into consideration for future experiments.

### **.3.9 Key findings from DHFR-TS biotinylation studies across various tools**

In summary, the computational approach predicted more favorable outcomes with the G0 (GGG) and PTP linkers, which were subsequently verified through experimental confirmation. These linkers displayed distinct and discernible bands of biotinylated DHFR-TS that were effectively captured by the beads, especially when electroporated parasites were examined.

Conversely, the flexible linkers G1 (GGGG) and G2 (GGGGG) as well as the rigid PA (PAPAP linker) gave the worst results, supporting the theoretical predictions. These linkers have high FCC values, which increases the likelihood of forming a ternary complex, thereby preventing access of BG-PEG<sub>4</sub>-Mtx to its active site. This finding was confirmed experimentally. No

significant differences were detected between the negative control and the sample of interest when utilizing the tool with the different linkers (G1, G2, and PA), regardless of whether electroporated parasites were used or not.

Although the MS part of the work needs more works to come to affirm validation of the whole protocol, the rest of the results strongly suggest that the novel target deconvolution tool is efficacious, particularly in light of the numerous positive outcomes attained through western blot analysis. Moreover, this technique employs highly specific antibodies, rendering it a more sensitive method than MS.

### **.3.10 Early findings of *Tb*AK biotinylation and capture using the cMyc\_BirA\*\_GGs\_SNAP-tag tool**

BG-PEG4-C1 synthesis was performed internally, and the specifics can be found in Appendix 8. In contrast, BG-PEG4-Mtx was outsourced to WuXi due to its intricate synthesis process involving multiple steps and the cytotoxic nature of Mtx, requiring specialized conditions. The assay was carried out with non-electroporated parasites. To perform the experiments, 1 µg/mL of tetracycline and 10 µM of BG-PEG<sub>4</sub>-C1 were added to the growth medium just before the log phase was reached, at around 10<sup>5</sup> parasites/mL. The parasites were then incubated for 5 days, and biotin was added on the 4th day at a concentration of 50 µM, i.e., 24 hours before the pull-down.

Subsequently, the pull-down procedure was performed according to the established protocol (details in Section .2.2.25), and the resulting findings are displayed in Figure 80.

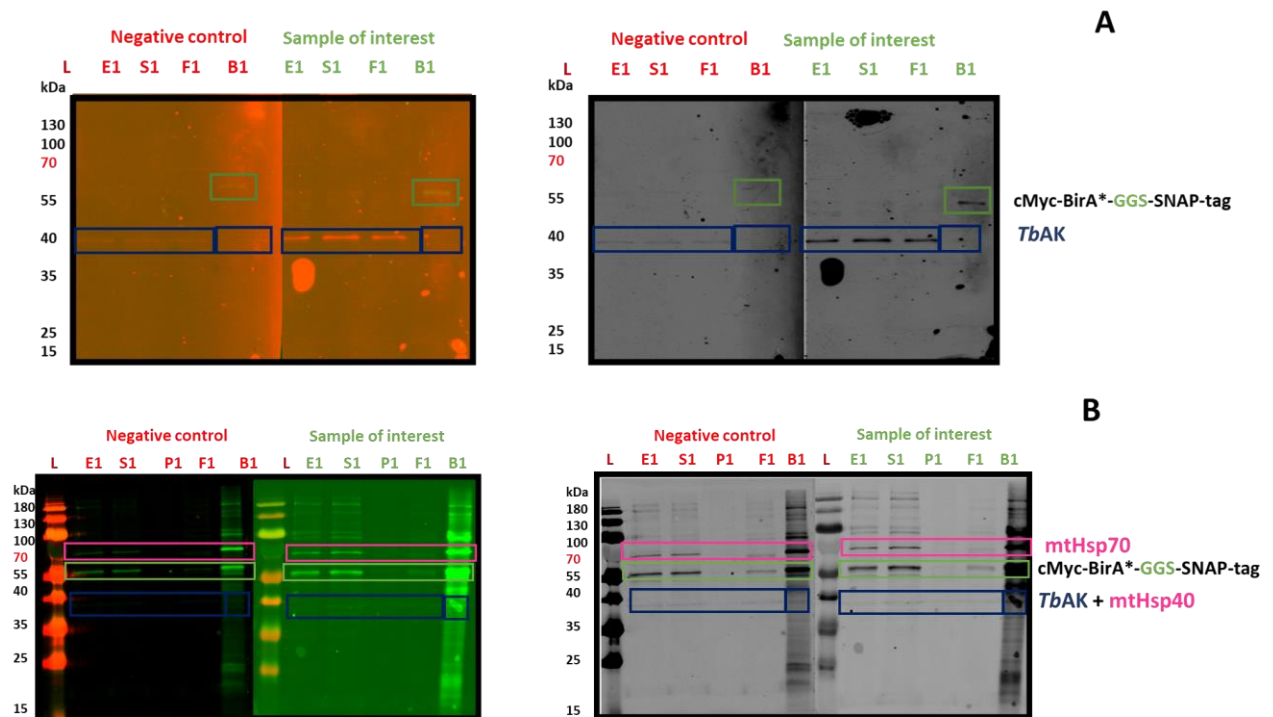


**Figure 80: Purification of candidate *TbAK* and *cMyc\_BirA\*\_GGG\_SNAP-tag* (Gel SDS PAGE 12%).** Part A: The membranes were incubated with anti-cMyc antibodies to reveal the tool and anti-TbAK to reveal TbAK and Part B: The membranes were incubated with streptavidin dye (IRDye 800CW Streptavidin) for biotinylated proteins revelation. Parasites were incubated with BG-PEG<sub>4</sub>-C1 (10  $\mu$ M) (only for the sample of interest), tetracycline (1 $\mu$ g/mL), and an excess of biotin (50  $\mu$ M) added 24 h before the pull-down. Following the overnight biotinylation, the subsequent steps were carried out identically to the previously performed ones, using the same designations E1, S1, P1, F1, and B1.

Compared to the previous experiments, the revelations carried out with the different antibodies or by streptavidin dye were performed on different membranes.

What is well visible in the Figure 80 above and particularly in A (membranes revealed with anti-TbAK antibodies) are the bands corresponding to the *TbAK* target of molecular weight of 38 kDa (188). Their intensity is even more important in the sample of interest showing that the tool was able to capture the target.

It is also noticeable that the band in the B1 fraction has disappeared. This phenomenon may be due to the biotinylation of the target and its subsequent capture by the beads, which could cause it to become unrecognizable by anti-TbAK. As a result, the corresponding band is no longer visible through these antibodies.

As previously mentioned, some chaperones (mtHsp70/mtHsp40), with molecular weights of 70 kDa and 40 kDa are abundant and therefore, can be biotinylated by the tool (139). Their bands and in particular those corresponding to mtHsp40 can overlap with that of *TbAK* (38 kDa). This overlapping does not allow discrimination of the band corresponding to the target of interest from that of the mtHSP40 chaperone. This is apparent in the above membranes (Figure 80). Indeed, their bands are slightly visible in the different fractions and their intensity is stronger on the membrane corresponding to the sample of interest.

Noteworthy is that experiments with the various other tools and in particular with the cMyc\_BirA\*\_PTP\_SNAP-tag system could not be carried out yet but would be worth trying in the future.

### **.3.11 Proof of concept of the Successful Functioning of the Tool**

To vividly visualize and illustrate the tool's sensitivity and selectivity, along with its capacity to exclusively capture the desired target based on the specific BG-PEG<sub>4</sub> derivative added in the culture medium, a comprehensive experiment was conducted. This latter involved analyzing all B fractions obtained from samples containing either *TbAK* or DHFR-TS, followed by an additional western blot analysis. The primary objective was to obtain conclusive results that showcase the various target proteins captured by the specific tools employed, namely cMyc\_BirA\_GGS\_SNAP-tag and cMyc\_BirA\_PTP\_SNAP-tag. The outcomes of this analysis are presented in Figure 81.

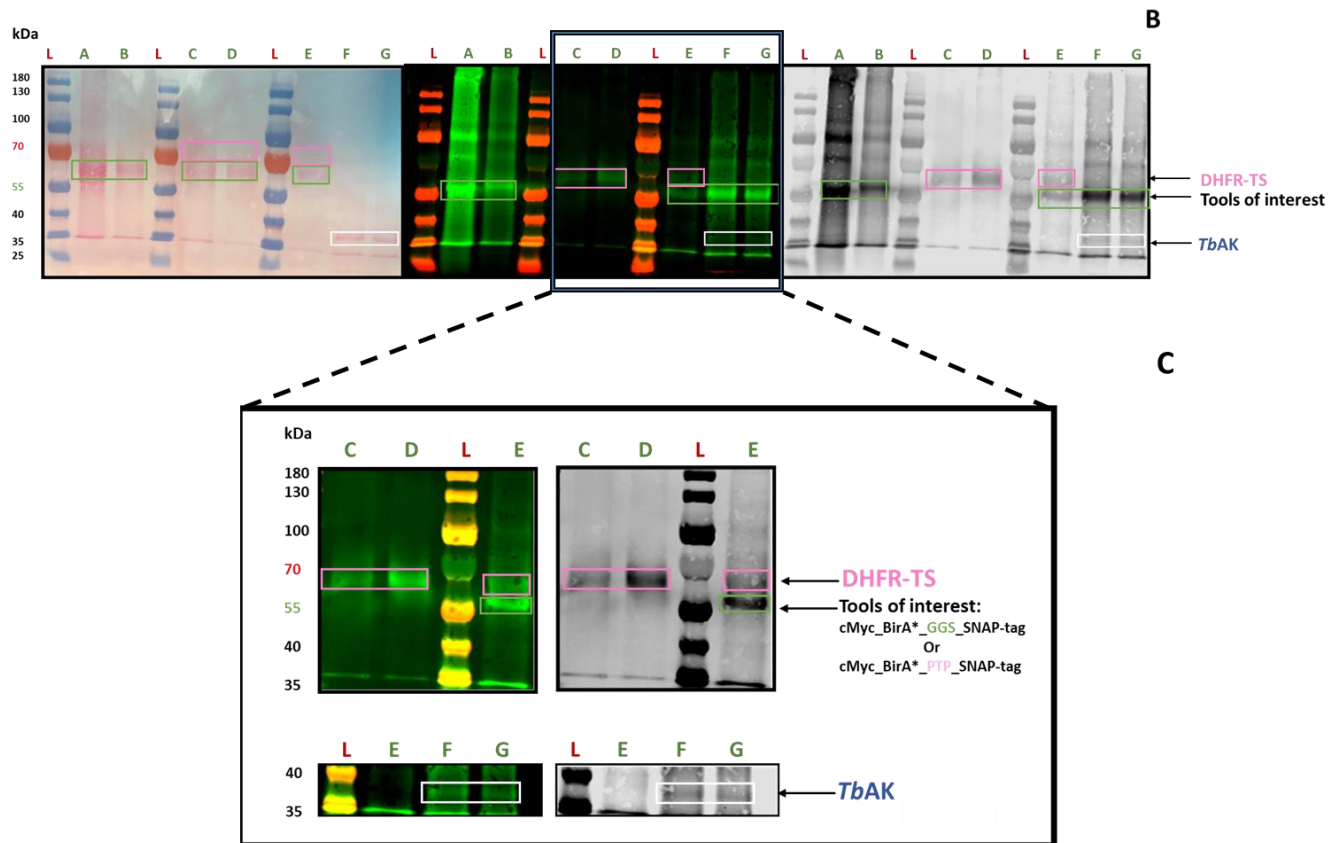


**Figure 81: Purification of candidate *TbAK*, *DHFR-TS* using, *cMyc\_BirA\*\_GG5\_SNAP-tag* (G0), and *cMyc\_BirA\*\_PTP\_SNAP-tag* (PTP) (Gel SDS PAGE 10%).** **A:** the membrane was incubated with anti-*TbAK* to reveal *TbAK* and **B:** the membrane was incubated with streptavidin dye (IRDye 800CW Streptavidin) to reveal biotinylated proteins. All samples correspond to fraction B1, which comprises the bound cytoplasmic eluate containing the biotinylated tools and targets, and potentially biotinylated neighbors. Specifically, *Tbb-G0* parasites were treated with BG-PEG<sub>4</sub>-C1 at a concentration of 10  $\mu$ M, while *Tbb-PTP* was treated with BG-PEG<sub>4</sub>-Mtx at a concentration of 50  $\mu$ M. The negative control (NC) lacked the presence of BG-PEG<sub>4</sub> derivatives, while the sample of interest (SI) was incubated in the presence of the compounds. A: non-electroporated *Tbb-PTP* (NC); B: non-electroporated *Tbb-PTP* (SI) with BG-PEG<sub>4</sub>-Mtx; C & D: electroporated *Tbb-PTP* (SI) with BG-PEG<sub>4</sub>-Mtx; E: non-electroporated *Tbb-G0* (SI) with BG-PEG<sub>4</sub>-Mtx; F & G: non-electroporated *Tbb-G0* (SI) with BG-PEG<sub>4</sub>-C1

First and foremost, it is evident that the biotinylated tools carrying the selected linkers (G0 linker or PTP linker) are effectively captured by the beads.

Secondly, there is no noticeable distinction between band A (the non-electroporated *Tbb\_G0* negative control) and band B (the non-electroporated *Tbb\_G0* sample of interest incubated with BG-PEG<sub>4</sub>-Mtx). However, following electroporation (bands C and D), a band of approximately 60 kDa, likely corresponding to *DHFR-TS*, is observed. The intense fluorescence

of this band obscures the fluorescence of the tool, leading to the absence of the corresponding band. Once again, this outcome serves as a strong testament to the remarkable efficiency of electroporation method in capturing the intended target.

Bands E, F, and G depict *Tbb\_GO* parasites that were not electroporated and were treated with either BG-PEG<sub>4</sub>-Mtx (E) or BG-PEG<sub>4</sub>-C1 (F and G). The presence of the band at 60 kDa (DHFR-TS) in E, which is absent in F and G, and the presence of a second band in F and G at around 40 kDa, likely corresponding to *TbAk* but missing in E, are both evident. This unequivocally demonstrates that the tool was successful in capturing the intended targets in light of the substance that was put to the growth medium (BG-PEG<sub>4</sub>-Mtx or BG-PEG<sub>4</sub>-C1 having as target DHFR-TS or *TbAK*, respectively).

To improve the repeatability of experiments, a second western blot analysis was conducted. This time, the B1 fraction was compared between two different conditions where the parasites were incubated with either BG-PEG<sub>4</sub>-C1 or BG-PEG<sub>4</sub>-Mtx. The results are presented by Figure 82.

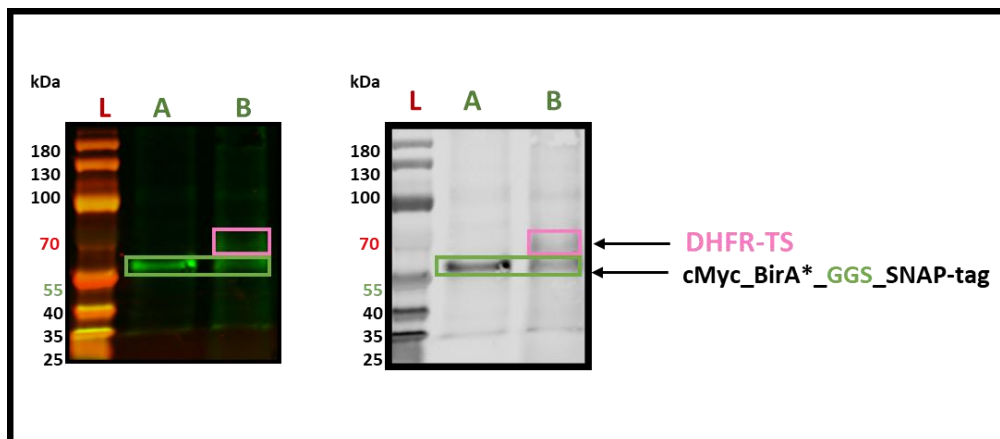


**Figure 82: Purification of candidate DHFR-TS and cMyc\_BirA\*\_GGG\_SNAP-tag (Gel SDS PAGE 10%).** The membrane was incubated with streptavidin dye (IRDye 800CW Streptavidin) to reveal biotinylated proteins. The two samples correspond to fraction B1 which comprises the bound cytoplasmic eluate containing the biotinylated tools and targets, and potentially biotinylated neighbors. **A:** *Tbb-GO* parasites that have been treated with BG-PEG<sub>4</sub>-C1 at a concentration of 10  $\mu$ M; **B:** *Tbb-PTP* parasites that have been treated with BG-PEG<sub>4</sub>-Mtx at a concentration of 50  $\mu$ M.

The obtained results demonstrate that the targets of interest were captured depending on the BG-PEG<sub>4</sub> derivative incubated with the transfected parasites. Notably, the DHFR-TS

protein was easily discernible in B, indicating a high level of selectivity and specificity of the tool. However, due to the low amount of *TbAK* protein captured, its fluorescence was overshadowed by the intense fluorescence emitted by the biotinylated tool, leading to the absence of its band in A.

Overall, the cMyc\_BirA\_GGS\_SNAP-tag and cMyc\_BirA\_PTP\_SNAP-tag target deconvolution tools show great promise for future optimization and experimentation.

## **.4 Conclusions & Outlooks**

### **.4.1 General conclusion**

This project aimed to create a novel tool that is both selective and specific, designed for detecting therapeutic targets of molecules, including newly discovered and promising compounds, those that are currently available on the market, as well as hits and leads. The development and optimization of this tool were a great challenge. Indeed, many parameters came into play and had to be studied in a very precise and meticulous way. Several parameters were thoroughly studied to optimize the system and generate the most efficient tool. These included determining the optimal concentration of the BG-PEG<sub>4</sub> derivative, identifying the appropriate time interval for tetracycline addition, ascertaining the ideal moment for tool expression, finding the appropriate concentration of the tetracycline inducer, identifying the best linker between the BirA\* and SNAP-tag domains, and exploring different BirA\* variants.

Identifying the captured targets and validating their presence by MS has been challenging due to several obstacles. The parasite environment is complex, and therapeutic targets are often rare and present in low amounts in parasites. Additionally, a wide range of detergents and lysis buffers must be tested and optimized to solubilize the proteins of interest without degrading them. Furthermore, different manipulations can lead to the loss of proteins at each step, which further complicates the process.

Moreover, the challenges encountered in obtaining conclusive results with the tool's various variants carrying the different linkers can partly be attributed to the limitations of MS in detecting the binding of the BG-PEG<sub>4</sub>-Mtx derivative to the SNAP-tag active site.

Therefore, it would be of interest to evaluate other BG-PEG<sub>4</sub> derivatives and examine the MS response, such as the BG-PEG<sub>4</sub>-C1 compound, which binds to the *TbAK* target and has undergone preliminary testing.

The study examined the BirA\* domain of the BioID system as well as its variants, TBirA\* and mTBirA\* corresponding respectively to the TurboID and miniTurbo systems. The findings indicated that the original BirA\* yielded the most favorable outcomes. Specifically, the DHFR-TS band on the nitrocellulose membrane revealed the strongest intensity with the BioID system, and no degradation of the fusion protein was observed in contrast to the TurboID technology. Additionally, the TurboID variant showed some degree of toxicity. Even though the growth of the parasites' cells that were transfected with the BioID\_SNAP-tag gene (*Tbb\_G0*), as well as those transfected with the two BioID mutants (TurboID and miniTurbo) linked to the SNAP-tag domain (*Tbb\_TG0* and *Tbb\_mTG0*), was slowed down when induced by tetracycline to express the different protein tools, the TurboID\_SNAP-tag system was only weakly expressed (maximum of 8% expression) and several degradation fragments were observed in western blot analysis over time, demonstrating the instability of the tool in the parasitic environment.

To summarize, the original BirA\* produced the best results and was therefore chosen for further investigations.

The second line of study was to investigate the influence of the nature and length of the linker connecting the two domains BirA\* and SNAP-tag. According to computational studies, the initial linker selection not only supported the functionality of BirA\* and SNAP but also limited biotinylation to a range of 10-15 nm, preventing any potential background interference resulting from indiscriminate biotinylation of adjacent proteins. This optimal linker is comprised of a sequence of three amino acids, namely Glycine-Glycine-Serine (GGG). Indeed, the GGS amino acid sequence is known to form a flexible linker, which allows for unimpeded action of each domain comprising the fusion protein by avoiding steric hindrance. Western blot analysis indicated that this linker exhibited good activity. To enhance the performance, several other linkers were investigated. The length of the flexible linker was initially increased to five amino acids, but this modification led to a loss of system functionality. Longer flexible

linkers may increase the likelihood of a ternary complex formation between the BirA\* and SNAP-tag domains and the linker. This could lead to the masking of the SNAP-tag active site and hinder the accessibility of the BG-PEG<sub>4</sub>-Mtx compound, causing the non-binding of the BG moiety to the Cys145 of the SNAP-tag domain.

The preceding statement is supported by findings from a computational analysis that calculated the FCC and the likelihood of creating a ternary complex. The study showed that utilizing lengthy and flexible linkers could impede the tool's functionality. Conversely, a linker made up of three rigid amino acids, specifically, Proline-Theonine-Proline (PTP) or Proline-Alanine-Proline (PAP), resulted in a more favorable outcome than the original GGS linker. These linkers were thus designed and produced, and the newly introduced parasites (*Tbb*\_PTP) were evaluated. Biotinylation studies validated the results obtained from the computational methods, demonstrating that tools with either G0 or PTP linkers produced the best results and should be utilized in future experiments with various BG-PEG<sub>4</sub> derivatives in order to identify their targets.

Finally, the comparison of the binding of BG-PEG<sub>4</sub>-Biotin and BG-PEG<sub>4</sub>-Mtx to the SNAP active site revealed suboptimal fragmentation of the Cys145-containing peptide bound to BG-PEG<sub>4</sub>-Mtx. This made it difficult to confirm the presence of the compound in the active site, despite the obvious raw data indicating good binding. BG-PEG<sub>4</sub>-Biotin did not have this issue. Hence, it is essential to consider this factor when binding compounds similar to Mtx to the BG-PEG<sub>4</sub> moiety for target identification. In such situations, it is necessary to explore an alternative orthogonal approach to western blot or a different MS fragmentation method.

## **.4.2 Outlooks**

A promising technology worth exploring is the CLIP-tag protein, which is a modified version of the SNAP-tag that also relies on O<sup>6</sup>-alkylguanine-DNA-alkyltransferase (hAGT). While CLIP-tag and SNAP-tag function similarly, the key distinction between them is the substrate they use. Specifically, CLIP-tag employs O<sup>2</sup>-benzylcytosine (BC) derivatives as its group of substrates, while SNAP-tag uses O<sup>6</sup>-benzylguanine derivatives (189, 190). Both proteins consist of 182 amino acids and have a molecular weight of 19.4 kDa. The ligands bound to their substrates

can be fluorescent dyes, biotin, beads, cell-permeable or cell-free dyes, etc. Both domains can be used simultaneously with different substrates in the same cell (125, 191).

There are two ways to test the CLIP protein: either as a fusion protein with BirA\* alone, bound with the selected linker (GGG or PTP), or fused with a SNAP-BirA\* to compare its effects and selectivity to the original cMyc\_BirA\_SNAP-tag tool. Theoretically, the addition of CLIP to the BirA\*-SNAP-tag is expected to enhance selectivity, but experimental validation is necessary. Furthermore, the choice of the linker is critical, and different ones will need to be studied computationally and experimentally to guarantee proper functionality and preserve the biological activities of the fusion protein's different domains. While linkers G0 and PTP have produced the best outcomes, their effectiveness with a new fusion protein cannot be ensured.

The introduction of CLIP requires the study of various domain insertions in the fusion protein. Two possible options are cMyc\_BirA\_Linkers\_SNAP\_Linkers\_CLIP-tag or cMyc\_BirA\_Linkers\_CLIP\_Linkers\_SNAP, with the linker corresponding to the ones selected, such as the PTP and G0 linkers, or new binders that can be evaluated, such as PAP, which gave the best theoretical and computational results as the PTP.

Introducing CLIP into the SNAP-BirA\* fusion protein could enhance selectivity, but this raises other scientific questions since the protein would become larger, at approximately 75 kDa, which could potentially impact its expression within the parasite.

Another approach that would represent an alternative to biotinylation is the photocrosslinking technique (Figure 83 below). This method has been well described and used in some papers (192-198). Moreover, it has been employed to identify the molecular target of a class of potent antimalarial drugs (199).

The proposed method involves the formation of an irreversible covalent bond between the target protein and the BG-PEG<sub>4</sub> derivative that had been bound to SNAP. To achieve this goal, a photoreactive cross-linking agent would be attached to the BG-PEG<sub>4</sub>-SM (SM referring to the small molecule) block, and upon UV irradiation, it would conjugate to any available functional group in nearby proteins. Once the complex between SNAP and the BG-PEG<sub>4</sub>-SM containing

cross-linker has formed with the unknown target of the SM, this latter can be isolated through pull-down using an affinity tag included in the SNAP-tag sequence.

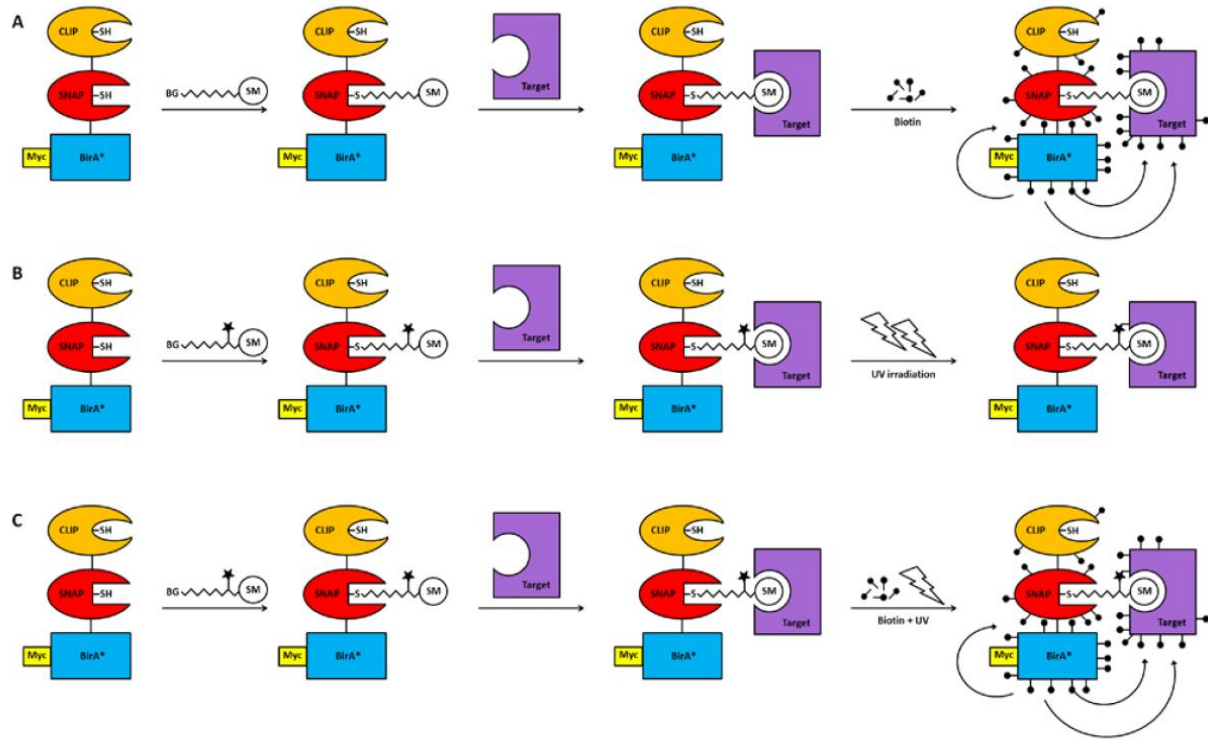


**Figure 83: (A) Biotinylation strategy** (used during the thesis project): The overexpressed cMyc-BirA\*-SNAP-CLIP-tag fusion protein interacts *in vivo* with the BG-PEG(n)-SM derivative. The interaction of the small molecule with its target will bring it close to the BirA\* domain which, upon the addition of biotin, will be activated in order to biotinylate the target of interest. Subsequently, the biotinylated proteins will be isolated using biotin-mediated affinity purification with the aid of streptavidin-coated Dynabeads. **(B) Photo-reactive affinity capture technique strategy:** The BG derivative bearing a photocrosslinking (black star) will form covalent bonds with the target protein upon UV irradiation. Isolation of the target of interest will be performed by cMyc-mediated affinity purification (e.g., anti-Myc agarose). **(C) Combination of both strategies:** This combination will provide the advantage of a dual purification procedure by sorting out biotinylated proteins that were covalently captured and eliminating those that were close but were not the target of the small molecule (SM)(by Dr. Remo Perozzo).

In addition to testing the CLIP-tag protein, various BG derivatives containing different numbers of PEG units could also be tested. As per the computational approach discussed in Section 3.1, the interaction between BG-(PEG)*n*-Mtx, SNAP-tag, and DHFR-TS resulted in a higher FCC value (probability of ternary complex formation) when the number of PEG units was equal to 2. However, for *n*>4, the FCCs remained relatively constant, and a plateau was observed on

the generated graph (Figure 21). Therefore, it can be expected that similar results would be obtained with a number of PEG units higher than 4. However, for  $n=2$  PEG units, significant differences could potentially be observed, and hence, it would be worth exploring further.

However, regarding the BG-(PEG) $_n$ -C1 interaction (Figure 84 below), the computational study shows that the FCC value reaches its maximum when the number of PEG units is equal to 1 and decreases as the number of units increases, suggesting that the likelihood of forming a ternary complex between the SNAP-tag domain binding to the BG moiety and the C1 moiety engaging to its *TbAK* target is high when these two molecules are linked by a single PEG unit, therefore, further investigation of this phenomenon is warranted

Moreover, the MS analysis reveals that the binding of BG-PEG $_4$ -Mtx derivative to SNAP-tag is not accurately identified in comparison to BG-PEG $_4$ -biotin compound, thereby, it would be interesting to test other BG-PEG $_4$  derivatives that could potentially give better MS results (e.g., continue the experiments undertaken with the BG-PEG $_4$ -C1 derivative).

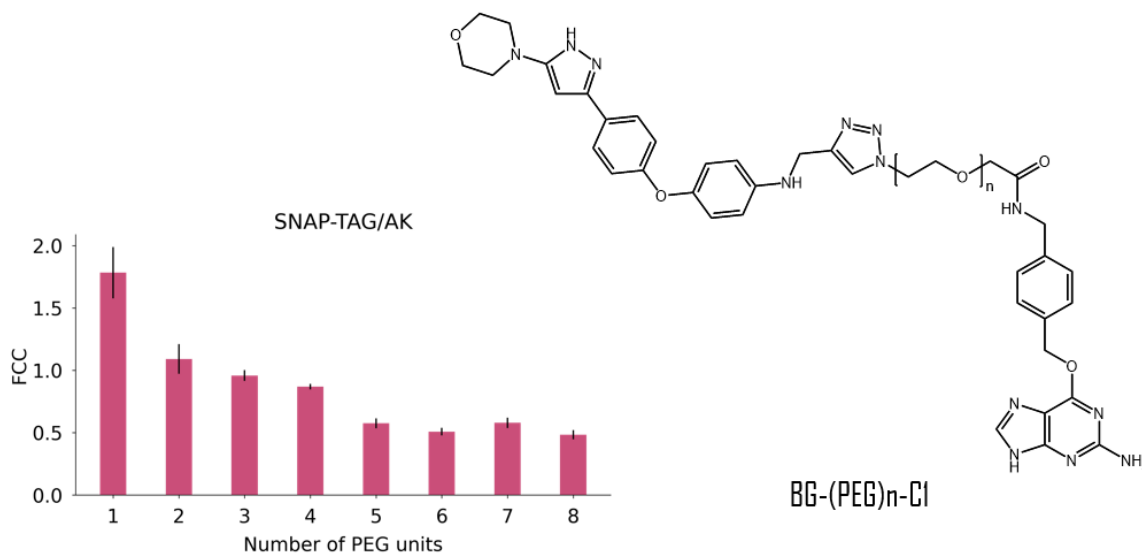


*Figure 84: Study of the PEG bond length between the BG moiety and the compound of interest C1 to form the ternary complex with the *TbAK* target, using FCC calculation (established by Margaux Héritier & Jiri Bejcek)*

Finally, it is worth noting that this project has the potential to be expanded and applied to other species, such as human embryonic kidney cells (HEK293), human cervix-carcinoma cells (heLa), or various parasite species such as *Tbr*, *Tbg*, *Trypanosoma cruzi*, *Leishmania spp* or

also *Leishmania tarentolae* (*L. tarentolae*) which is a protozoan parasite belonging to the *Leishmania* genus, that exhibits the unique characteristic of infecting only lizard species and not mammals. Additionally, it does not cause any harm to humans (as *Tbb*) or to mice suffering from severe combined immunodeficiency (SCID) that lack B and T cells. These distinctive features make this species an excellent choice for various scientific studies and experiments. (200, 201). The lack of pathogenicity of *L. tarentolae* could be explained by a study revealing, after sequencing the parasite's genome, the loss of genes presents in the intracellular amastigote form of *Leishmania* species pathogenic to humans (200, 202).

Upon validating the tool and identifying the best linker, an appropriate signal peptide can be selected for translocating the fusion protein into a subcellular compartment, such as the glycosome, nucleus, or mitochondria. The DNA encoding the desired signal peptide sequence can be introduced upstream of the cMyc-tag gene, or downstream of the SNAP-CLIP or CLIP-SNAP gene, before the stop codon. There are various well-defined signal peptide sequences available, such as the C-terminal SKL (serine-lysine-leucine) that facilitates translocation into the glycosome. Adding this sequence to the fusion protein of interest can direct it to this specific organelle (203). The N-terminal RGHRSRE guides the fusion protein to the nucleus, while the N-terminal MFRRCFPIF steers it toward the mitochondria (204, 205).

The successful translocation of the fusion protein into the target organelle and the optimization of lysis and isolation protocols for various subcellular targets are critical scientific questions that need to be addressed. Overcoming these challenges will be essential to complete this project, as my thesis results have clearly demonstrated the intricacy and difficulty of the subject matter.

## .5 References

1. Achcar F, Kerkhoven EJ, Barrett MP. Trypanosoma brucei: meet the system. Current opinion in microbiology. 2014;20:162-9.
2. Prevention CCFDCa. Life cycle African Trypanosomiasis [updated 2019; cited 2019. Available from: <https://www.cdc.gov/parasites/sleepingsickness/biology.html>.
3. Kallinikova VD. The kinetoplast as a cell organelle. Int Rev Cytol. 1981;69:105-56.
4. Clayton CE. Genetic manipulation of kinetoplastida. Parasitology today (Personal ed). 1999;15(9):372-8.
5. Shapiro TA, Englund PT. The structure and replication of kinetoplast DNA. Annual review of microbiology. 1995;49:117-43.
6. Wheeler RJ, Gluenz E, Gull K. The limits on trypanosomatid morphological diversity. PLoS One. 2013;8(11):e79581.
7. Kaufer A, Ellis J, Stark D, Barratt J. The evolution of trypanosomatid taxonomy. Parasites & Vectors. 2017;10(1):287.
8. Kennedy PGE. Update on human African trypanosomiasis (sleeping sickness). Journal of neurology. 2019;266(9):2334-7.
9. Vanhamme L, Pays E. The trypanosome lytic factor of human serum and the molecular basis of sleeping sickness. Int J Parasitol. 2004;34(8):887-98.
10. Raper J, Fung R, Ghiso J, Nussenzweig V, Tomlinson S. Characterization of a novel trypanosome lytic factor from human serum. Infection and immunity. 1999;67(4):1910-6.
11. Raper J, Portela MPM, Lugli E, Frevert U, Tomlinson S. Trypanosome lytic factors: novel mediators of human innate immunity. Current opinion in microbiology. 2001;4(4):402-8.
12. Thomson R, Samanovic M, Raper J. Activity of trypanosome lytic factor: a novel component of innate immunity. Future microbiology. 2009;4(7):789-96.
13. Capewell P, Cooper A, Clucas C, Weir W, Macleod A. A co-evolutionary arms race: trypanosomes shaping the human genome, humans shaping the trypanosome genome. Parasitology. 2015;142 Suppl 1(Suppl 1):S108-19.
14. Vanhamme L, Paturiaux-Hanocq F, Poelvoorde P, Nolan DP, Lins L, Van Den Abbeele J, et al. Apolipoprotein L-I is the trypanosome lytic factor of human serum. Nature. 2003;422(6927):83-7.
15. De Greef C, Chimfwembe E, Kihang'a Wabacha J, Bajyana Songa E, Hamers R. Only the serum-resistant bloodstream forms of Trypanosoma brucei rhodesiense express the serum resistance associated (SRA) protein. Ann Soc Belg Med Trop. 1992;72 Suppl 1:13-21.
16. Bart JM, Cordon-Obras C, Vidal I, Reed J, Perez-Pastrana E, Cuevas L, et al. Localization of serum resistance-associated protein in Trypanosoma brucei rhodesiense and transgenic Trypanosoma brucei brucei. Cellular microbiology. 2015;17(10):1523-35.
17. Kieft R, Capewell P, Turner CM, Veitch NJ, MacLeod A, Hajduk S. Mechanism of Trypanosoma brucei gambiense (group 1) resistance to human trypanosome lytic factor. Proc Natl Acad Sci U S A. 2010;107(37):16137-41.
18. Smith AB, Esko JD, Hajduk SL. Killing of trypanosomes by the human haptoglobin-related protein. Science (New York, NY). 1995;268(5208):284-6.
19. Shiflett AM, Bishop JR, Pahwa A, Hajduk SL. Human high density lipoproteins are platforms for the assembly of multi-component innate immune complexes. J Biol Chem. 2005;280(38):32578-85.
20. Molina-Portela Mdel P, Lugli EB, Recio-Pinto E, Raper J. Trypanosome lytic factor, a subclass of high-density lipoprotein, forms cation-selective pores in membranes. Mol Biochem Parasitol. 2005;144(2):218-26.
21. Molina-Portela MP, Samanovic M, Raper J. Distinct roles of apolipoprotein components within the trypanosome lytic factor complex revealed in a novel transgenic mouse model. The Journal of experimental medicine. 2008;205(8):1721-8.
22. Widener J, Nielsen MJ, Shiflett A, Moestrup SK, Hajduk S. Hemoglobin is a co-factor of human trypanosome lytic factor. PLoS Pathog. 2007;3(9):1250-61.
23. Vanhollebeke B, De Muylder G, Nielsen MJ, Pays A, Tebabi P, Dieu M, et al. A haptoglobin-hemoglobin receptor conveys innate immunity to Trypanosoma brucei in humans. Science (New York, NY). 2008;320(5876):677-81.
24. Drain J, Bishop JR, Hajduk SL. Haptoglobin-related protein mediates trypanosome lytic factor binding to trypanosomes. J Biol Chem. 2001;276(32):30254-60.

25. Uzureau P, Uzureau S, Lecordier L, Fontaine F, Tebabi P, Homblé F, et al. Mechanism of *Trypanosoma brucei* gambiense resistance to human serum. *Nature*. 2013;501(7467):430-4.
26. Berberof M, Pérez-Morga D, Pays E. A receptor-like flagellar pocket glycoprotein specific to *Trypanosoma brucei* gambiense. *Mol Biochem Parasitol*. 2001;113(1):127-38.
27. Capewell P, Clucas C, DeJesus E, Kieft R, Hajduk S, Veitch N, et al. The TgsGP gene is essential for resistance to human serum in *Trypanosoma brucei* gambiense. *PLoS Pathog*. 2013;9(10):e1003686.
28. Gould M. Putative phosphodiesterase inhibitors as potential new chemotherapies against African Trypanosomiasis 2009.
29. K V, Myler P, Kd S. African trypanosomiasis. 1993. p. 170-212.
30. Mugnier MR, Stebbins CE, Papavasiliou FN. Masters of Disguise: Antigenic Variation and the VSG Coat in *Trypanosoma brucei*. *PLoS Pathog*. 2016;12(9):e1005784.
31. Silva Pereira S, Jackson AP, Figueiredo LM. Evolution of the variant surface glycoprotein family in African trypanosomes. *Trends in Parasitology*. 2022;38(1):23-36.
32. Hovel-Miner G, Mugnier M, Papavasiliou FN, Pinger J, Schulz D. A Host-Pathogen Interaction Reduced to First Principles: Antigenic Variation in *T. brucei*. Results and problems in cell differentiation. 2015;57:23-46.
33. Vassella E, Oberle M, Urwyler S, Renggli CK, Studer E, Hemphill A, et al. Major surface glycoproteins of insect forms of *Trypanosoma brucei* are not essential for cyclical transmission by tsetse. *PLoS One*. 2009;4(2):e4493.
34. Pays E, Coquelet H, Tebabi P, Pays A, Jefferies D, Steinert M, et al. *Trypanosoma brucei*: constitutive activity of the VSG and procyclin gene promoters. *The EMBO journal*. 1990;9(10):3145-51.
35. Magez S, Stijlemans B, Baral T, De Baetselier P. VSG-GPI anchors of African trypanosomes: their role in macrophage activation and induction of infection-associated immunopathology. *Microbes and infection*. 2002;4(9):999-1006.
36. Vassella E, Bütikofer P, Engstler M, Jelk J, Roditi I. Procyclin Null Mutants of *Trypanosoma brucei* Express Free Glycosylphosphatidylinositols on Their Surface. *Mol Biol Cell*. 2002;14(4):1308-18.
37. Hutchinson S, Foulon S, Crouzols A, Menafrá R, Rotureau B, Griffiths AD, et al. The establishment of variant surface glycoprotein monoallelic expression revealed by single-cell RNA-seq of *Trypanosoma brucei* in the tsetse fly salivary glands. *PLoS Pathog*. 2021;17(9):e1009904.
38. Radwanska M, Chamekh M, Vanhamme L, Claes F, Magez S, Magnus E, et al. The serum resistance-associated gene as a diagnostic tool for the detection of *Trypanosoma brucei* rhodesiense. *The American journal of tropical medicine and hygiene*. 2002;67(6):684-90.
39. Langousis G, Hill KL. Motility and more: the flagellum of *Trypanosoma brucei*. *Nat Rev Microbiol*. 2014;12(7):505-18.
40. Satir P. Landmarks in cilia research from Leeuwenhoek to us. *Cell motility and the cytoskeleton*. 1995;32(2):90-4.
41. Ooi CP, Bastin P. More than meets the eye: understanding *Trypanosoma brucei* morphology in the tsetse. *Frontiers in cellular and infection microbiology*. 2013;3:71.
42. Geiger A, Ravel S, Mateille T, Janelle J, Patrel D, Cuny G, et al. Vector competence of *Glossina palpalis* gambiensis for *Trypanosoma brucei* s.l. and genetic diversity of the symbiont *Sodalis glossinidius*. *Molecular biology and evolution*. 2007;24(1):102-9.
43. Wamwiri FN, Changasi RE. Tsetse Flies (*Glossina*) as Vectors of Human African Trypanosomiasis: A Review. *BioMed research international*. 2016;2016:6201350.
44. Bouteille B, Dumas M. Human African Trypanosomiasis. In: Aminoff MJ, Daroff RB, editors. *Encyclopedia of the Neurological Sciences*. New York: Academic Press; 2003. p. 587-94.
45. Krinsky WL. Tsetse Flies (*Glossinidae*)\_ Medical and Veterinary Entomology. 3rd Edition ed. Durden GRMaLA, editor 2019 October 2, 2018.
46. Marcello L, Barry JD. Analysis of the VSG gene silent archive in *Trypanosoma brucei* reveals that mosaic gene expression is prominent in antigenic variation and is favored by archive substructure. *Genome Res*. 2007;17(9):1344-52.
47. Weiden M, Osheim YN, Beyer AL, Ploeg LHvd. Chromosome structure: DNA nucleotide sequence elements of a subset of the minichromosomes of the protozoan *Trypanosoma brucei*. *Molecular and Cellular Biology*. 1991;11(8):3823-34.
48. Melville SE, Leech V, Gerrard CS, Tait A, Blackwell JM. The molecular karyotype of the megabase chromosomes of *Trypanosoma brucei* and the assignment of chromosome markers. *Molecular and Biochemical Parasitology*. 1998;94(2):155-73.

49. Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, et al. The genome of the African trypanosome *Trypanosoma brucei*. *Science (New York, NY)*. 2005;309(5733):416-22.
50. Van der Ploeg LHT, Cornelissen AWCA, Michels PAM, Borst P. Chromosome rearrangements in *trypanosoma brucei*. *Cell*. 1984;39(1):213-21.
51. Johnson PJ, Borst P. Mapping of VSG genes on large expression-site chromosomes of *Trypanosoma brucei* separated by pulsed-field gradient electrophoresis. *Gene*. 1986;43(3):213-20.
52. Williams RO, Young JR, Majiwa PAO. Genomic environment of *T. brucei* VSG genes: presence of a minichromosome. *Nature*. 1982;299(5882):417-21.
53. Sloof P, Menke HH, Caspers MPM, Borst P. Size fractionation of *Trypanosoma brucei* DNA: localization of the 177-bp repeat satellite DNA and a variant surface glycoprotein gene in a mini-chromosomal DNA fraction. *Nucleic Acids Research*. 1983;11(12):3889-901.
54. Amodeo S, Jakob M, Ochsenreiter T. Characterization of the novel mitochondrial genome replication factor MiRF172 in *Trypanosoma brucei*. *J Cell Sci*. 2018;131(8):jcs211730.
55. Liu B, Wang J, Yaffe N, Lindsay ME, Zhao Z, Zick A, et al. Trypanosomes have six mitochondrial DNA helicases with one controlling kinetoplast maxicircle replication. *Molecular cell*. 2009;35(4):490-501.
56. Klingbeil MM, Motyka SA, Englund PT. Multiple Mitochondrial DNA Polymerases in *Trypanosoma brucei*. *Molecular cell*. 2002;10(1):175-86.
57. Liu B, Liu Y, Motyka SA, Agbo EE, Englund PT. Fellowship of the rings: the replication of kinetoplast DNA. *Trends Parasitol*. 2005;21(8):363-9.
58. Shlomai J. The structure and replication of kinetoplast DNA. *Current molecular medicine*. 2004;4(6):623-47.
59. Simpson L, Sbicego S, Aphasizhev R. Uridine insertion/deletion RNA editing in trypanosome mitochondria: a complex business. *RNA (New York, NY)*. 2003;9(3):265-76.
60. Stuart KD, Schnauffer A, Ernst NL, Panigrahi AK. Complex management: RNA editing in trypanosomes. *Trends in biochemical sciences*. 2005;30(2):97-105.
61. Clayton C. Regulation of gene expression in trypanosomatids: living with polycistronic transcription. *Open Biol*. 2019;9(6):190072.
62. Waithaka A, Maiakovska O, Grimm D, do Nascimento LM, Clayton C. Sequences and proteins that influence mRNA processing in *Trypanosoma brucei*: Evolutionary conservation of SR-domain and PTB protein functions. *PLoS neglected tropical diseases*. 2022;16(10):e0010876.
63. Begolo D, Vincent IM, Giordani F, Pöhner I, Witty MJ, Rowan TG, et al. The trypanocidal benzoxaborole AN7973 inhibits trypanosome mRNA processing. *PLoS Pathog*. 2018;14(9):e1007315.
64. Michaeli S. Trans-splicing in trypanosomes: machinery and its impact on the parasite transcriptome. *Future microbiology*. 2011;6(4):459-74.
65. Clayton C, Michaeli S. 3' processing in protists. *Wiley interdisciplinary reviews RNA*. 2011;2(2):247-55.
66. Matthews KR. The developmental cell biology of *Trypanosoma brucei*. *Journal of cell science*. 2005;118(Pt 2):283-90.
67. Sharma R, Gluenz E, Peacock L, Gibson W, Gull K, Carrington M. The heart of darkness: growth and form of *Trypanosoma brucei* in the tsetse fly. *Trends in Parasitology*. 2009;25(11):517-24.
68. Matthews KR, Ellis JR, Paterou A. Molecular regulation of the life cycle of African trypanosomes. *Trends Parasitol*. 2004;20(1):40-7.
69. Barry JD, McCulloch R. Antigenic variation in trypanosomes: enhanced phenotypic variation in a eukaryotic parasite. *Advances in parasitology*. 2001;49:1-70.
70. Vassella E, Den Abbeele JV, Bütikofer P, Renggli CK, Furger A, Brun R, et al. A major surface glycoprotein of *trypanosoma brucei* is expressed transiently during development and can be regulated post-transcriptionally by glycerol or hypoxia. *Genes & development*. 2000;14(5):615-26.
71. Vickerman K. Developmental cycles and biology of pathogenic trypanosomes. *British Medical Bulletin*. 1985;41(2):105-14.
72. Fernandes MC, Flannery AR, Andrews N, Mortara RA. Extracellular amastigotes of *Trypanosoma cruzi* are potent inducers of phagocytosis in mammalian cells. *Cellular microbiology*. 2013;15(6):977-91.
73. Kima PE. The amastigote forms of *Leishmania* are experts at exploiting host cell processes to establish infection and persist. *Int J Parasitol*. 2007;37(10):1087-96.
74. Teixeira DE, Benchimol M, Rodrigues JC, Crepaldi PH, Pimenta PF, de Souza W. The cell biology of *Leishmania*: how to teach using animations. *PLoS Pathog*. 2013;9(10):e1003594.
75. Brun R, Don R, Jacobs RT, Wang MZ, Barrett MP. Development of novel drugs for human African trypanosomiasis. *Future microbiology*. 2011;6(6):677-91.

76. Kennedy PGE. The continuing problem of human African trypanosomiasis (sleeping sickness). *Annals of Neurology*. 2008;64(2):116-26.
77. Kennedy PGE. Update on human African trypanosomiasis (sleeping sickness). *Journal of neurology*. 2019;266(9):2334-7.
78. Brun R, Don R, Jacobs RT, Wang MZ, Barrett MP. Development of novel drugs for human African trypanosomiasis. *Future microbiology*. 2011;6(6):677-91.
79. Checchi F, Filipe JAN, Barrett MP, Chandramohan D. The Natural Progression of Gambiense Sleeping Sickness: What Is the Evidence? *PLoS neglected tropical diseases*. 2008;2(12):e303.
80. WHO. Human African trypanosomiasis (sleeping sickness) 2021 [updated 2021. Available from: [https://www.who.int/health-topics/human-african-trypanosomiasis#tab=tab\\_1](https://www.who.int/health-topics/human-african-trypanosomiasis#tab=tab_1).
81. Kennedy PGE. Sleeping sickness – human African trypanosomiasis. *Practical Neurology*. 2005;5(5):260.
82. Liu Z, Ulrich vonBargen R, McCall L-I. Central role of metabolism in *Trypanosoma cruzi* tropism and Chagas disease pathogenesis. *Current opinion in microbiology*. 2021;63:204-9.
83. Nunes MCP, Beaton A, Acquatella H, Bern C, Bolger AF, Echeverría LE, et al. Chagas Cardiomyopathy: An Update of Current Clinical Knowledge and Management: A Scientific Statement From the American Heart Association. *Circulation*. 2018;138(12):e169-e209.
84. Reisenman CE, Lawrence G, Guenstein PG, Gregory T, Dotson E, Hildebrand JG. Infection of kissing bugs with *Trypanosoma cruzi*, Tucson, Arizona, USA. *Emerging infectious diseases*. 2010;16(3):400-5.
85. da Costa KM, Marques da Fonseca L, Dos Reis JS, Santos M, Previato JO, Mendonça-Previato L, et al. *Trypanosoma cruzi* trans-Sialidase as a Potential Vaccine Target Against Chagas Disease. *Frontiers in cellular and infection microbiology*. 2021;11:768450.
86. Organization WWH. Chagas disease (also known as American trypanosomiasis) 2022 [updated 13.04.2022. Available from: [https://www.who.int/news-room/fact-sheets/detail/chagas-disease-\(american-trypanosomiasis\)](https://www.who.int/news-room/fact-sheets/detail/chagas-disease-(american-trypanosomiasis))).
87. Gramiccia M, Gradoni L. The current status of zoonotic leishmaniasis and approaches to disease control. *International Journal for Parasitology*. 2005;35(11):1169-80.
88. Organization WWH. Leishmaniasis. 2022 [updated 08.01.2022. Available from: <https://www.who.int/news-room/fact-sheets/detail/leishmaniasis>].
89. Prevention CCoDca. CDC Centers of Disease Control and Prevention- Parasites- Leishmaniasis -Biology 2020 [updated 18.02.2020. Available from: <https://www.cdc.gov/parasites/leishmaniasis/biology.html>].
90. Sunter J, Gull K. Shape, form, function and *Leishmania* pathogenicity: from textbook descriptions to biological understanding. *Open Biology*. 2017;7(9):170165.
91. Ahuir-Baraja AE, Ruiz MP, Garijo MM, Llobat L. Feline Leishmaniosis: An Emerging Public Health Problem. *Veterinary sciences*. 2021;8(9):173.
92. Goto H, Lauletta Lindoso JA. Cutaneous and Mucocutaneous Leishmaniasis. *Infectious Disease Clinics of North America*. 2012;26(2):293-307.
93. Kushwaha AK, Scorza BM, Singh OP, Rowton E, Lawyer P, Sundar S, et al. Domestic mammals as reservoirs for *Leishmania donovani* on the Indian subcontinent: Possibility and consequences on elimination. *Transboundary and emerging diseases*. 2022;69(2):268-77.
94. Paz C, Samake S, Anderson JM, Faye O, Traore P, Tall K, et al. *Leishmania major*, the predominant *Leishmania* species responsible for cutaneous leishmaniasis in Mali. *The American journal of tropical medicine and hygiene*. 2013;88(3):583-5.
95. Talas J, Mielcarek K, Wu J, Brunner M, Steinhoff M, Zouboulis CC. Kutane Leishmaniasis in Deutschland – Weiterhin eine Reisedermatose. *Der Hautarzt*. 2022;73(2):146-51.
96. Alcover MM, Riera MC, Fisa R. Leishmaniosis in Rodents Caused by *Leishmania infantum*: A Review of Studies in the Mediterranean Area. *Frontiers in veterinary science*. 2021;8:702687.
97. Rodríguez-Serrato MA, Salinas-Carmona MC, Limón-Flores AY. Immune response to *Leishmania mexicana*: the host–parasite relationship. *Pathogens and Disease*. 2020;78(8):ftaa060.
98. Volpedo G, Huston RH, Holcomb EA, Pacheco-Fernandez T, Gannavaram S, Bhattacharya P, et al. From infection to vaccination: reviewing the global burden, history of vaccine development, and recurring challenges in global leishmaniasis protection. *Expert review of vaccines*. 2021;20(11):1431-46.
99. Taghipour A, Abdoli A, Ramezani A, Abolghazi A, Mofazzal Jahromi MA, Maani S, et al. Leishmaniasis and Trace Element Alterations: a Systematic Review. *Biological trace element research*. 2021;199(10):3918-38.
100. Kumari D, Singh K. Exploring the paradox of defense between host and *Leishmania* parasite. *International immunopharmacology*. 2022;102:108400.

101. Müller J, Müller J, Hemphill A. New approaches for the identification of drug targets in protozoan parasites. *Int Rev Cell Mol Biol.* 2013;301:359-401.
102. Lee WC, Lee KH. Applications of affinity chromatography in proteomics. *Anal Biochem.* 2004;324(1):1-10.
103. Müller J, Hemphill A. Chapter Seven - New Approaches for the Identification of Drug Targets in Protozoan Parasites. In: Jeon KW, editor. *International review of cell and molecular biology.* 301: Academic Press; 2013. p. 359-401.
104. Müller J, Hemphill A. Drug target identification in intracellular and extracellular protozoan parasites. *Current topics in medicinal chemistry.* 2011;11(16):2029-38.
105. Müller J, Hemphill A. Drug target identification in protozoan parasites. *Expert Opin Drug Discov.* 2016;11(8):815-24.
106. Morriswood B, Havlicek K, Demmel L, Yavuz S, Sealey-Cardona M, Vidilaseris K, et al. Novel bilobe components in *Trypanosoma brucei* identified using proximity-dependent biotinylation. *Eukaryotic cell.* 2013;12(2):356-67.
107. Roux KJ, Kim DI, Burke B. BioID: a screen for protein-protein interactions. *Curr Protoc Protein Sci.* 2013;74:19.23.1-19.23.14
108. Varnaite R, MacNeill SA. Meet the neighbors: Mapping local protein interactomes by proximity-dependent labeling with BioID. *Proteomics.* 2016;16(19):2503-18.
109. Kim DI, Jensen SC, Noble KA, Kc B, Roux KH, Motamedchaboki K, et al. An improved smaller biotin ligase for BioID proximity labeling. *Mol Biol Cell.* 2016;27(8):1188-96.
110. Chojnowski A, Sobota RM, Ong PF, Xie W, Wong X, Dreesen O, et al. 2C-BioID: An Advanced Two Component BioID System for Precision Mapping of Protein Interactomes. *iScience.* 2018;10:40-52.
111. Ashley A, Adile AV. Use of BioID to detect protein-protein interactions. *Canadian Journal of Undergraduate Research.* 2018.
112. Sears RM, May DG, Roux KJ. BioID as a Tool for Protein-Proximity Labeling in Living Cells. *Methods Mol Biol.* 2019;2012:299-313.
113. Martell JD, Deerinck TJ, Sancak Y, Poulos TL, Mootha VK, Sosinsky GE, et al. Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. *Nat Biotechnol.* 2012;30(11):1143-8.
114. Lam SS, Martell JD, Kamer KJ, Deerinck TJ, Ellisman MH, Mootha VK, et al. Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nature Methods.* 2015;12(1):51-4.
115. Kalocsay M. APEX Peroxidase-Catalyzed Proximity Labeling and Multiplexed Quantitative Proteomics. *Methods Mol Biol.* 2019;2008:41-55.
116. Branon TC, Bosch JA, Sanchez AD, Udeshi ND, Svinkina T, Carr SA, et al. Efficient proximity labeling in living cells and organisms with TurboID. *Nat Biotechnol.* 2018;36(9):880-7.
117. Kaewsapsak P, Shechner DM, Mallard W, Rinn JL, Ting AY. Live-cell mapping of organelle-associated RNAs via proximity biotinylation combined with protein-RNA crosslinking. *Elife.* 2017;6:e29224.
118. Roux KJ, Kim DI, Raida M, Burke B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol.* 2012;196(6):801-10.
119. May DG, Scott KL, Campos AR, Roux KJ. Comparative Application of BioID and TurboID for Protein-Proximity Biotinylation. *Cells.* 2020;9(5):1070.
120. Gronemeyer T, Chidley C, Juillerat A, Heinis C, Johnsson K. Directed evolution of O6-alkylguanine-DNA alkyltransferase for applications in protein labeling. *Protein engineering, design & selection: PEDS.* 2006;19(7):309-16.
121. Pegg AE. Repair of O6-alkylguanine by alkyltransferases. *Mutation Research/Reviews in Mutation Research.* 2000;462(2):83-100.
122. Gronemeyer T, Chidley C, Juillerat A, Heinis C, Johnsson K. Directed evolution of O6-alkylguanine-DNA alkyltransferase for applications in protein labeling. *Protein Engineering, Design and Selection.* 2006;19(7):309-16.
123. Keppler A, Gendreizig S, Gronemeyer T, Pick H, Vogel H, Johnsson K. A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat Biotechnol.* 2003;21(1):86-9.
124. Hinner MJ, Johnsson K. How to obtain labeled proteins and what to do with them. *Current Opinion in Biotechnology.* 2010;21(6):766-76.
125. Gautier A, Juillerat A, Heinis C, Corrêa IR, Kindermann M, Beaufils F, et al. An Engineered Protein Tag for Multiprotein Labeling in Living Cells. *Chemistry & Biology.* 2008;15(2):128-36.

126. Keppler A, Kindermann M, Gendreizig S, Pick H, Vogel H, Johnsson K. Labeling of fusion proteins of O6-alkylguanine-DNA alkyltransferase with small molecules in vivo and in vitro. *Methods*. 2004;32(4):437-44.
127. Bodor DL, Rodríguez MG, Moreno N, Jansen LE. Analysis of protein turnover by quantitative SNAP-based pulse-chase imaging. *Curr Protoc Cell Biol*. 2012;Chapter 8:Unit8.
128. Cole NB. Marquage spécifique des protéines avec les SNAP-tags. *Curr Protoc Protein Sci*. 2013;73:30.1.1-30.1.16.
129. Chidley C. A Novel Yeast-Based Platform for Small Molecule Target Deconvolution Reveals Previously Unknown Binding Partners for the Approved Drugs Erlotinib, Atorvastatin, and Sulfasalazine: EPFL; 2010.
130. Juillerat A, Gronemeyer T, Keppler A, Gendreizig S, Pick H, Vogel H, et al. Directed Evolution of O6-Alkylguanine-DNA Alkyltransferase for Efficient Labeling of Fusion Proteins with Small Molecules In Vivo. *Chemistry & Biology*. 2003;10(4):313-7.
131. Wirtz E, Leal S, Ochatt C, Cross GM. A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*. *Molecular and Biochemical Parasitology*. 1999;99(1):89-101.
132. Rao VS, Srinivas K, Sujini GN, Kumar GNS. Protein-Protein Interaction Detection: Methods and Analysis. *International Journal of Proteomics*. 2014;2014:147648.
133. Patil A. Protein-Protein Interaction Databases. In: Ranganathan S, Gribskov M, Nakai K, Schönbach C, editors. *Encyclopedia of Bioinformatics and Computational Biology*. Oxford: Academic Press; 2019. p. 849-55.
134. Kuettel S, Greenwald J, Kostrewa D, Ahmed S, Scapozza L, Perozzo R. Crystal Structures of T. b. rhodesiense Adenosine Kinase Complexed with Inhibitor and Activator: Implications for Catalysis and Hyperactivation. *PLoS neglected tropical diseases*. 2011;5(5):e1164.
135. Kuettel S, Mosimann M, Maser P, Kaiser M, Brun R, Scapozza L, et al. Adenosine Kinase of T. b. Rhodesiense identified as the putative target of 4-[5-(4-phenoxyphenyl)-2H-pyrazol-3-yl]morpholine using chemical proteomics. *PLoS neglected tropical diseases*. 2009;3(8):e506.
136. Kuettel S, Zambon A, Kaiser M, Brun R, Scapozza L, Perozzo R. Synthesis and evaluation of antiparasitic activities of new 4-[5-(4-phenoxyphenyl)-2H-pyrazol-3-yl]morpholine derivatives. *Journal of medicinal chemistry*. 2007;50(23):5833-9.
137. Gibson MW, Dewar S, Ong HB, Sienkiewicz N, Fairlamb AH. *Trypanosoma brucei* DHFR-TS Revisited: Characterisation of a Bifunctional and Highly Unstable Recombinant Dihydrofolate Reductase-Thymidylate Synthase. *PLoS neglected tropical diseases*. 2016;10(5):e0004714.
138. Radwanska M, Magez S, Michel A, Stijlemans B, Geuskens M, Pays E. Comparative analysis of antibody responses against HSP60, invariant surface glycoprotein 70, and variant surface glycoprotein reveals a complex antigen-specific pattern of immunoglobulin isotype switching during infection by *Trypanosoma brucei*. *Infection and immunity*. 2000;68(2):848-60.
139. Týč J, Klingbeil MM, Lukeš J. Mitochondrial heat shock protein machinery hsp70/hsp40 is indispensable for proper mitochondrial DNA maintenance and replication. *mBio*. 2015;6(1):e02425-14.
140. Wirtz E, Leal S, Ochatt C, Cross GA. A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*. *Mol Biochem Parasitol*. 1999;99(1):89-101.
141. Wirtz E, Clayton C. Inducible gene expression in trypanosomes mediated by a prokaryotic repressor. *Science (New York, NY)*. 1995;268(5214):1179-83.
142. Tomlinson S, Raper J. The lysis of *Trypanosoma brucei brucei* by human serum. *Nat Biotechnol*. 1996;14(6):717-21.
143. Croce AC, Bottiroli G. Autofluorescence spectroscopy and imaging: a tool for biomedical research and diagnosis. *European journal of histochemistry : EJH*. 2014;58(4):2461.
144. Hirumi H, Hirumi K. Continuous cultivation of *Trypanosoma brucei* blood stream forms in a medium containing a low concentration of serum protein without feeder cell layers. *J Parasitol*. 1989;75(6):985-9.
145. Christie A, Butler M. Glutamine-based dipeptides are utilized in mammalian cell culture by extracellular hydrolysis catalyzed by a specific peptidase. *Journal of biotechnology*. 1994;37(3):277-90.
146. Tritsch GL, Moore GE. Spontaneous decomposition of glutamine in cell culture media. *Experimental cell research*. 1962;28(2):360-4.
147. Barbedo J. Automatic Object Counting In Neubauer Chambers. 2013: p.1.
148. Cheah JS, Yamada S. A simple elution strategy for biotinylated proteins bound to streptavidin conjugated beads using excess biotin and heat. *Biochem Biophys Res Commun*. 2017;493(4):1522-7.

- 
149. Bai N, Miller SA, Andrianov GV, Yates M, Kirubakaran P, Karanicolas J. Rationalizing PROTAC-Mediated Ternary Complex Formation Using Rosetta. *Journal of Chemical Information and Modeling*. 2021;61(3):1368-82.
  150. Roberts MJ, Bentley MD, Harris JM. Chemistry for peptide and protein PEGylation. *Adv Drug Deliv Rev*. 2002;54(4):459-76.
  151. Gursahani H, Riggs-Sauthier J, Pfeiffer J, Lechuga-Ballesteros D, Fishburn CS. Absorption of Polyethylene Glycol (PEG) Polymers: The Effect of PEG Size on Permeability. *Journal of Pharmaceutical Sciences*. 2009;98(8):2847-56.
  152. He Y-L, Murby S, Warhurst G, Gifford L, Walker D, Ayrton J, et al. Species Differences in Size Discrimination in the Paracellular Pathway Reflected by Oral Bioavailability of Poly(ethylene glycol) and d-Peptides. *Journal of Pharmaceutical Sciences*. 1998;87(5):626-33.
  153. Turecek PL, Bossard MJ, Schoetens F, Ivens IA. PEGylation of Biopharmaceuticals: A Review of Chemistry and Nonclinical Safety Information of Approved Drugs. *J Pharm Sci*. 2016;105(2):460-75.
  154. Watson CJ, Rowland M, Warhurst G. Functional modeling of tight junctions in intestinal cell monolayers using polyethylene glycol oligomers. *American Journal of Physiology-Cell Physiology*. 2001;281(2):C388-C97.
  155. Monaco A, Ovryn B, Axis J, Amsler K. The Epithelial Cell Leak Pathway. *Int J Mol Sci*. 2021;22(14):7677.
  156. Becker F, Murthi K, Smith C, Come J, Costa-Roldán N, Kaufmann C, et al. A three-hybrid approach to scanning the proteome for targets of small molecule kinase inhibitors. *Chem Biol*. 2004;11(2):211-23.
  157. Chidley C, Haruki H, Pedersen MG, Muller E, Johnsson K. A yeast-based screen reveals that sulfasalazine inhibits tetrahydrobiopterin biosynthesis. *Nature chemical biology*. 2011;7(6):375-83.
  158. Merlo R, Caprioglio D, Cillo M, Valenti A, Mattosovich R, Morrone C, et al. The SNAP-tag technology revised: an effective chemo-enzymatic approach by using a universal azide-based substrate. *Journal of enzyme inhibition and medicinal chemistry*. 2021;36(1):85-97.
  159. Gendreizig S, Kindermann M, Johnsson K. Induced Protein Dimerization in Vivo through Covalent Labeling. *Journal of the American Chemical Society*. 2003;125(49):14970-1.
  160. Klein VG, Townsend CE, Testa A, Zengerle M, Maniaci C, Hughes SJ, et al. Understanding and Improving the Membrane Permeability of VH032-Based PROTACs. *ACS Medicinal Chemistry Letters*. 2020;11(9):1732-8.
  161. Klein VG, Bond AG, Craigon C, Lokey RS, Ciulli A. Amide-to-Ester Substitution as a Strategy for Optimizing PROTAC Permeability and Cellular Activity. *Journal of medicinal chemistry*. 2021;64(24):18082-101.
  162. Foley CA, Potjewyd F, Lamb KN, James LI, Frye SV. Assessing the Cell Permeability of Bivalent Chemical Degraders Using the Chloroalkane Penetration Assay. *ACS chemical biology*. 2020;15(1):290-5.
  163. Shi G, Azoulay M, Dingli F, Lamaze C, Loew D, Florent J-C, et al. SNAP-tag Based Proteomics Approach for the Study of the Retrograde Route. *Traffic*. 2012;13(7):914-25.
  164. Bannwarth M, Corrêa IR, Sztretye M, Pouvreau S, Fellay C, Aebischer A, et al. Indo-1 Derivatives for Local Calcium Sensing. *ACS chemical biology*. 2009;4(3):179-90.
  165. Kovtun YV, Audette CA, Mayo MF, Jones GE, Doherty H, Maloney EK, et al. Antibody-Maytansinoid Conjugates Designed to Bypass Multidrug Resistance. *Cancer Research*. 2010;70(6):2528-37.
  166. Li X, Yang X, Li Z, Zheng X, Peng Y-j, Lin W, et al. Development of a Radiotracer for PET Imaging of the SNAP Tag. *ACS Omega*. 2022;7(9):7550-5.
  167. Zhang C, Sheng W, Al-Rawe M, Mohiuddin TM, Niebert M, Zeppernick F, et al. EpCAM- and EGFR-Specific Antibody Drug Conjugates for Triple-Negative Breast Cancer Treatment. *Int J Mol Sci*. 2022;23(11):6122.
  168. Fang Y, Zhang A, Li S, Sproviero M, Xu M-Q. Enzyme Immobilization for Solid-Phase Catalysis. *Catalysts*. 2019;9(9):732.
  169. Chen X, Zaro JL, Shen W-C. Fusion protein linkers: property, design and functionality. *Adv Drug Deliv Rev*. 2013;65(10):1357-69.
  170. Skaf MS, Polikarpov I, Stanković IM. A linker of the proline-threonine repeating motif sequence is bimodal. *Journal of Molecular Modeling*. 2020;26(7):178.
  171. van Rosmalen M, Krom M, Merkx M. Tuning the Flexibility of Glycine-Serine Linkers To Allow Rational Design of Multidomain Proteins. *Biochemistry*. 2017;56(50):6565-74.
  172. Green SM, Meeker AK, Shortle D. Contributions of the polar, uncharged amino acids to the stability of staphylococcal nuclease: evidence for mutational effects on the free energy of the denatured state. *Biochemistry*. 1992;31(25):5717-28.
  173. bioscience L. Electroporation and Nucleofector® Technology [Available from: [https://bioscience.lonza.com/lonza\\_bs/CH/en/nucleofector-technology](https://bioscience.lonza.com/lonza_bs/CH/en/nucleofector-technology).

- 
174. Powell HM, Armour AD, Boyce ST. Fluorescein diacetate for determination of cell viability in 3D fibroblast-collagen-GAG constructs. *Methods Mol Biol.* 2011;740:115-26.
  175. Armour AD, Powell HM, Boyce ST. Fluorescein diacetate for determination of cell viability in tissue-engineered skin. *Tissue engineering Part C, Methods.* 2008;14(1):89-96.
  176. Scientific T. Fluorescence SpectraViewer ThermoFisher Scientific [Available from: <https://www.thermofisher.com/order/fluorescence-spectraviewer#!/>].
  177. Surre J, Saint-Ruf C, Collin V, Orenge S, Ramjeet M, Matic I. Strong increase in the autofluorescence of cells signals struggle for survival. *Scientific Reports.* 2018;8(1):12088.
  178. Carroll P, Muttucumaru DG, Parish T. Use of a tetracycline-inducible system for conditional expression in *Mycobacterium tuberculosis* and *Mycobacterium smegmatis*. *Applied and environmental microbiology.* 2005;71(6):3077-84.
  179. Hashimi H, Kaltenbrunner S, Zíková A, Lukeš J. Trypanosome Mitochondrial Translation and Tetracycline: No Sweat about Tet. *PLoS Pathog.* 2016;12(4):e1005492.
  180. Stevens M, Abdeen S, Salim N, Ray AM, Washburn A, Chitre S, et al. HSP60/10 chaperonin systems are inhibited by a variety of approved drugs, natural products, and known bioactive molecules. *Bioorganic & medicinal chemistry letters.* 2019;29(9):1106-12.
  181. Du F, Liu Y-Q, Xu Y-S, Li Z-J, Wang Y-Z, Zhang Z-X, et al. Regulating the T7 RNA polymerase expression in *E. coli* BL21 (DE3) to provide more host options for recombinant protein production. *Microbial Cell Factories.* 2021;20(1):189.
  182. Kato Y. Extremely Low Leakage Expression Systems Using Dual Transcriptional-Translational Control for Toxic Protein Production. *Int J Mol Sci.* 2020;21(3):705.
  183. Maeda Y, Ueda H, Kazami J, Kawano G, Suzuki E, Nagamune T. Engineering of Functional Chimeric Protein G–VargulaLuciferase. *Analytical Biochemistry.* 1997;249(2):147-52.
  184. Bai Y, Shen WC. Improving the oral efficacy of recombinant granulocyte colony-stimulating factor and transferrin fusion protein by spacer optimization. *Pharmaceutical research.* 2006;23(9):2116-21.
  185. Amet N, Lee HF, Shen WC. Insertion of the designed helical linker led to increased expression of tf-based fusion proteins. *Pharmaceutical research.* 2009;26(3):523-8.
  186. Ito T, Ando H, Suzuki T, Ogura T, Hotta K, Imamura Y, et al. Identification of a primary target of thalidomide teratogenicity. *Science (New York, NY).* 2010;327(5971):1345-50.
  187. Shen Y, Tolić N, Xie F, Zhao R, Purvine SO, Schepmoes AA, et al. Effectiveness of CID, HCD, and ETD with FT MS/MS for degradomic-peptidomic analysis: comparison of peptide identification methods. *Journal of proteome research.* 2011;10(9):3929-43.
  188. Lüscher A, Onal P, Schweingruber AM, Mäser P. Adenosine kinase of *Trypanosoma brucei* and its role in susceptibility to adenosine antimetabolites. *Antimicrob Agents Chemother.* 2007;51(11):3895-901.
  189. Hoehnel S, Lutolf MP. Capturing Cell–Cell Interactions via SNAP-tag and CLIP-tag Technology. *Bioconjugate Chemistry.* 2015;26(8):1678-86.
  190. Gautier A, Juillerat A, Heinis C, Corrêa IR, Jr., Kindermann M, Beaufils F, et al. An engineered protein tag for multiprotein labeling in living cells. *Chem Biol.* 2008;15(2):128-36.
  191. Cole NB. Site-specific protein labeling with SNAP-tags. *Current protocols in protein science.* 2013;73:30.1.1-30.1.16.
  192. Geurink PP, Prely LM, van der Marel GA, Bischoff R, Overkleeft HS. Photoaffinity labeling in activity-based protein profiling. *Topics in current chemistry.* 2012;324:85-113.
  193. Jessani N, Cravatt BF. The development and application of methods for activity-based protein profiling. *Curr Opin Chem Biol.* 2004;8(1):54-9.
  194. Köster H, Little DP, Luan P, Muller R, Siddiqi SM, Marappan S, et al. Capture compound mass spectrometry: a technology for the investigation of small molecule protein interactions. *Assay and drug development technologies.* 2007;5(3):381-90.
  195. Lenz T, Fischer JJ, Dreger M. Probing small molecule–protein interactions: A new perspective for functional proteomics. *Journal of Proteomics.* 2011;75(1):100-15.
  196. Li N, Overkleeft HS, Florea BI. Activity-based protein profiling: an enabling technology in chemical biology research. *Curr Opin Chem Biol.* 2012;16(1-2):227-33.
  197. Mishra PK, Yoo C-M, Hong E, Rhee HW. Photo-crosslinking: An Emerging Chemical Tool for Investigating Molecular Networks in Live Cells. *ChemBioChem.* 2020;21(7):924-32.
  198. Brüninghoff K, Wulff S, Dörner W, Geiss-Friedlander R, Mootz HD. A Photo-Crosslinking Approach to Identify Class II SUMO-1 Binders. *Frontiers in chemistry.* 2022;10:900989.

199. Brunner R, Ng CL, Aissaoui H, Akabas MH, Boss C, Brun R, et al. UV-triggered affinity capture identifies interactions between the Plasmodium falciparum multidrug resistance protein 1 (PfMDR1) and antimalarial agents in live parasitized cells. *J Biol Chem.* 2013;288(31):22576-83.
200. Niimi T. Leishmania tarentolae for the Production of Multi-subunit Complexes. *Advances in experimental medicine and biology.* 2016;896:155-65.
201. Breton M, Tremblay MJ, Ouellette M, Papadopolou B. Live nonpathogenic parasitic vector as a candidate vaccine against visceral leishmaniasis. *Infection and immunity.* 2005;73(10):6372-82.
202. Raymond F, Boisvert S, Roy G, Ritt JF, Légaré D, Isnard A, et al. Genome sequencing of the lizard parasite Leishmania tarentolae reveals loss of genes associated to the intracellular stage of human pathogenic species. *Nucleic Acids Res.* 2012;40(3):1131-47.
203. Sommer JM, Cheng QL, Keller GA, Wang CC. In vivo import of firefly luciferase into the glycosomes of Trypanosoma brucei and mutational analysis of the C-terminal targeting signal. *Mol Biol Cell.* 1992;3(7):749-59.
204. Marchetti MA, Tschudi C, Kwon H, Wolin SL, Ullu E. Import of proteins into the trypanosome nucleus and their distribution at karyokinesis. *Journal of cell science.* 2000;113 ( Pt 5):899-906.
205. Häusler T, Stierhof YD, Blattner J, Clayton C. Conservation of mitochondrial targeting sequence function in mitochondrial and hydrogenosomal proteins from the early-branching eukaryotes Crithidia, Trypanosoma and Trichomonas. *European journal of cell biology.* 1997;73(3):240-51.

## .6 Appendix

### APPENDIX 1

Maps of all synthesized plasmids with details

#### Plasmid pLew100\_cMyc\_GGS\_BirA\*\_SNAP-tag

Detailed plasmid map hereafter:

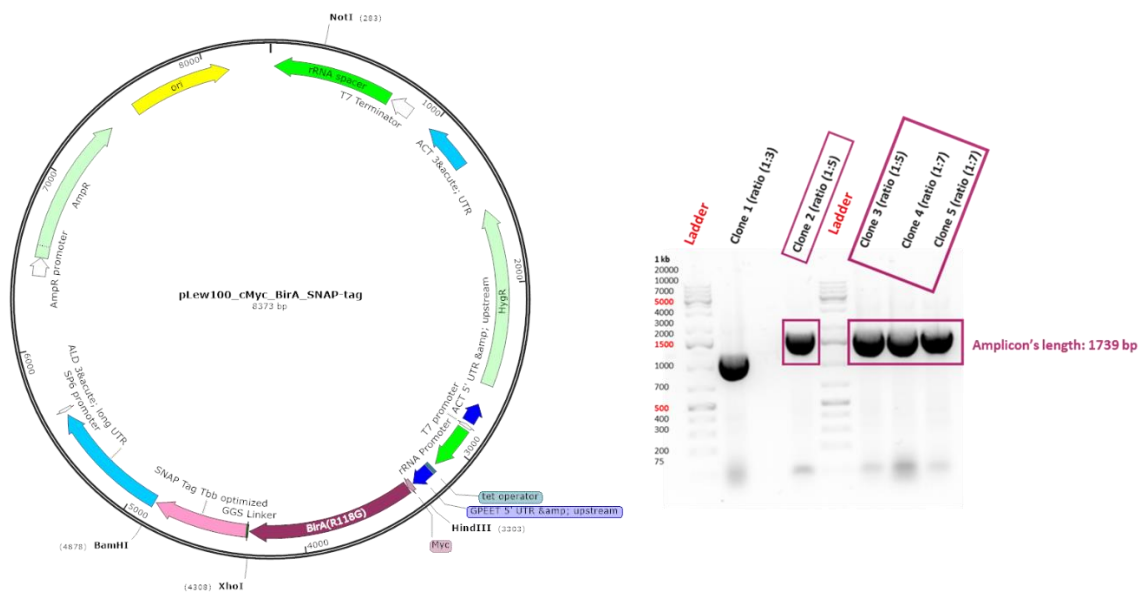


**Figure S1:** Plasmid Map and Successful *E. coli* Transformation with cMyc\_BirA\*\_GGS\_SNAP-tag

The pLew100\_cMyc\_BirA\*\_GGS\_SNAP-tag was generated as described in Section .2.2.1, after digestion of the original plasmid pLew100\_cMyc\_BirA\* as well as the synthesized GGS-SNAP-tag gene, followed by ligation of the insert (GGG-SNAP-tag) into the digested plasmid.

Part 3: Development of a new target deconvolution tool

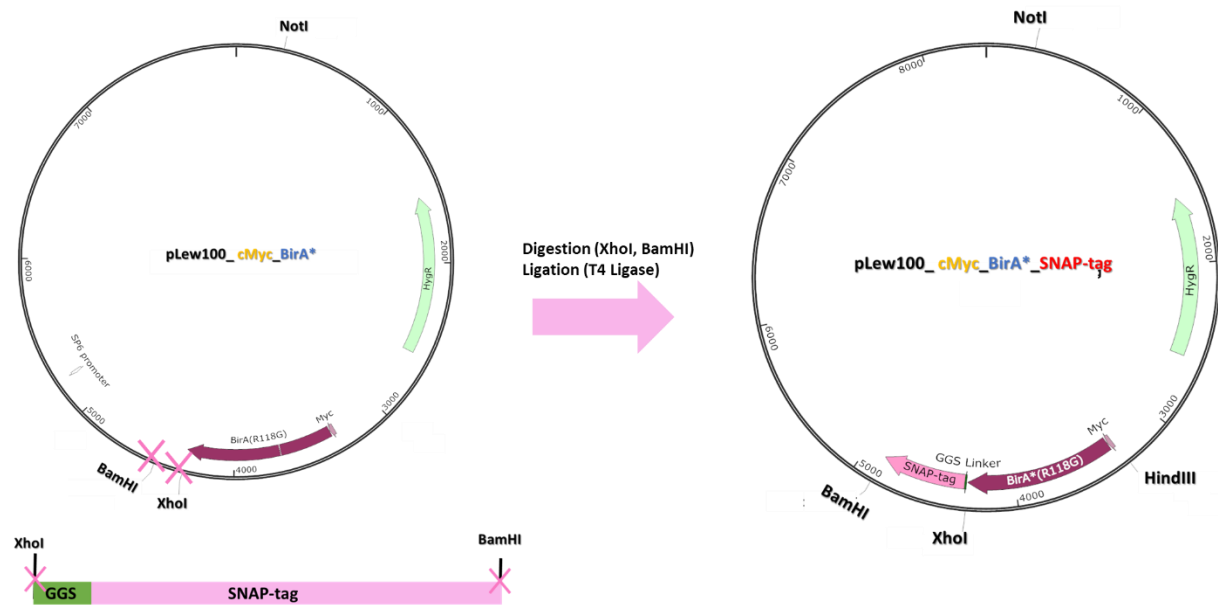


**Figure S2:** Plasmid pLew100\_cMyc\_BirA\* Digestion and GGS\_SNAP-tag Gene Ligation: Generating cMyc\_BirA\*\_GGS\_SNAP-tag Plasmid

After bacterial transformation, the correct integration of the SNAP-tag into the plasmid was verified by colony PCR followed by sequencing of purified plasmids from the isolated clones (figure beneath).

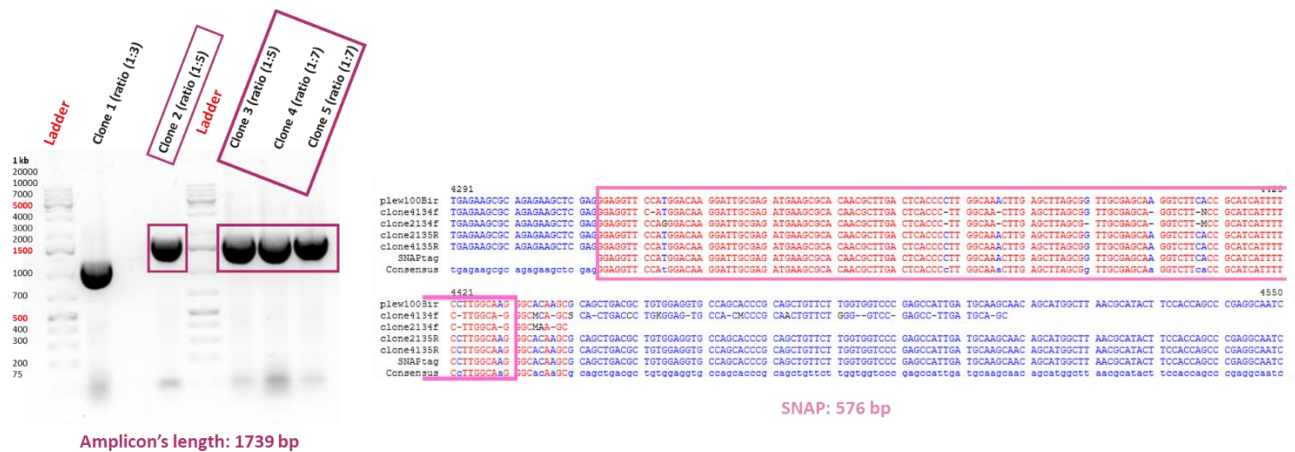


**Figure S3:** Sequencing verification confirming accurate integration of GGS\_SNAP-tag into the Plasmid

## APPENDIX 2

## Plasmid pLew100\_cMyc\_TBirA\*\_GGG\_SNP-tag

**A: Design and production of the insert of interest HindIII\_cMyc\_TBirA\*\_XhoI**

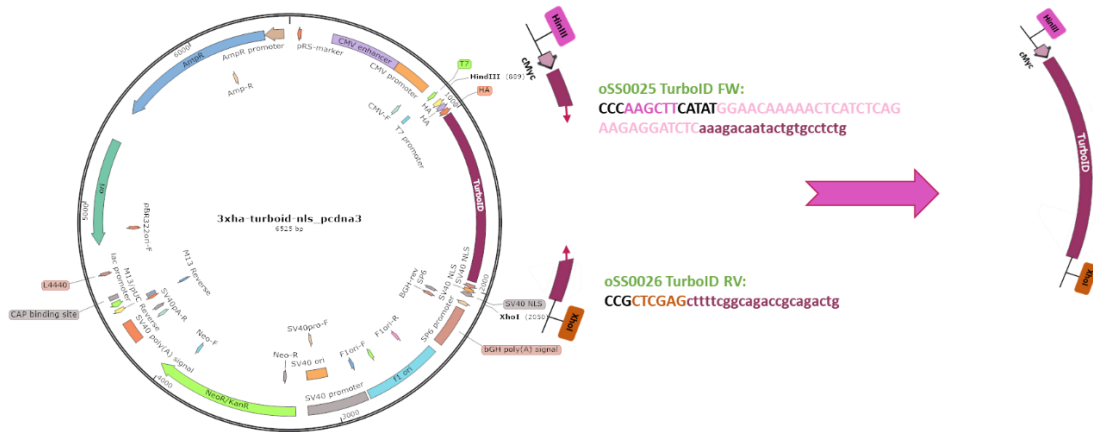


**Figure S4:** Generation of the *HindIII\_cMyc\_TBirA\*\_XhoI* insert of interest by PCR

### ***B: Verification by PCR and sequencing of the correct acquisition of the fragment of interest***

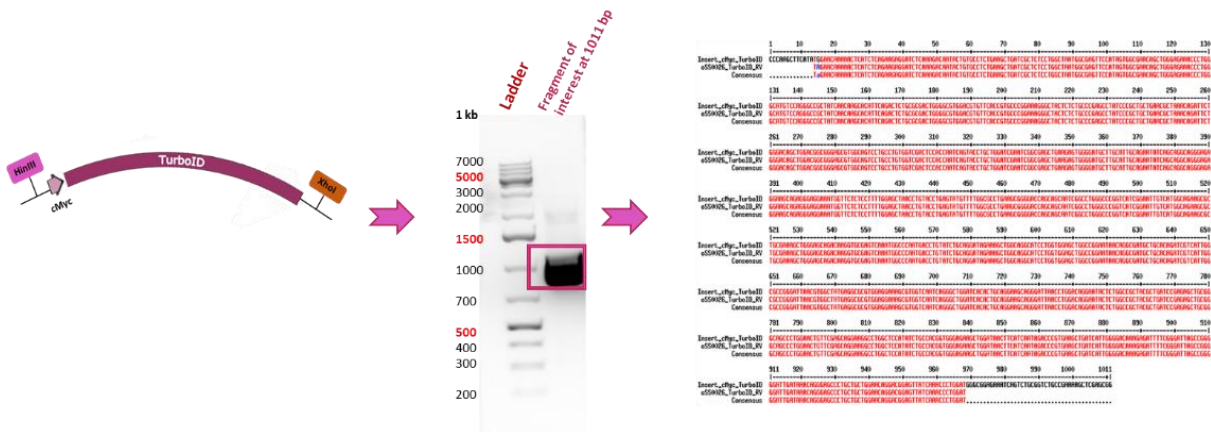


**Figure S5:** Verification of the acquisition of the *HindIII\_cMyc\_TBirA\*\_XhoI* insert of interest by sequencing

***C: Verification by PCR and sequencing of the correct insertion of the fragment of interest into the pre-digested plasmid pLew100 cMyc BirA\* SNAP***

- **oCA134 FW:** 5'-CGCGCCTTCGAGTTTTTTTCC-3'==> targeting GPEET 5' UTR & amp upstream region contained in the main vector plasmid

- **oCA135 RV:** 5'-CCTGCTGTGCCATCAGATTAC -3'==> targeting ALD 3 & acute long UTR region  
contained in the main vector plasmid

Suzanne Sherihan SAHRAOUI  
**Part 3: Development of a new target deconvolution tool**

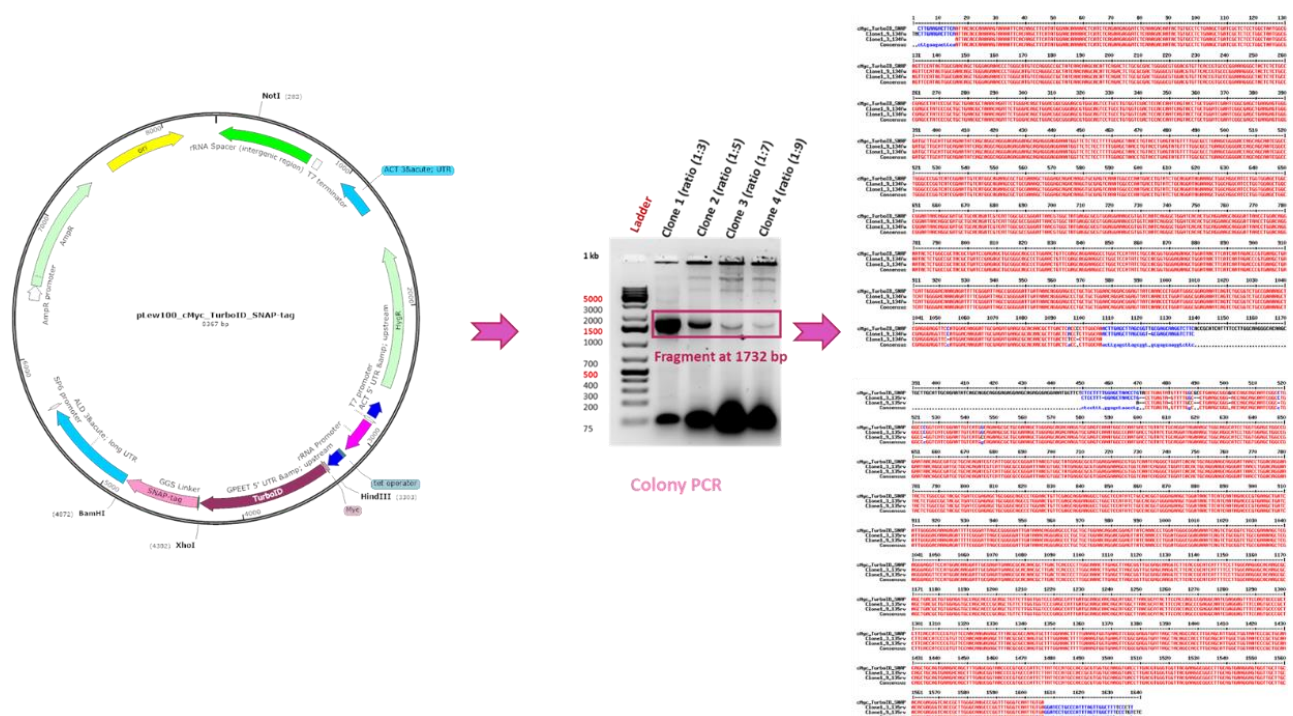


**Figure S6:** PCR and Sequencing Verification of Successful Generation of pLew100\_cMyc\_TBirA\*\_GGG\_SNAP-tag Plasmid

## APPENDIX 3

### Plasmid pLew100\_cMyc\_mTBirA\*\_GGG\_SNAP-tag

**A: Design and production of the insert of interest HindIII\_cMyc\_mTBirA\*\_XhoI**

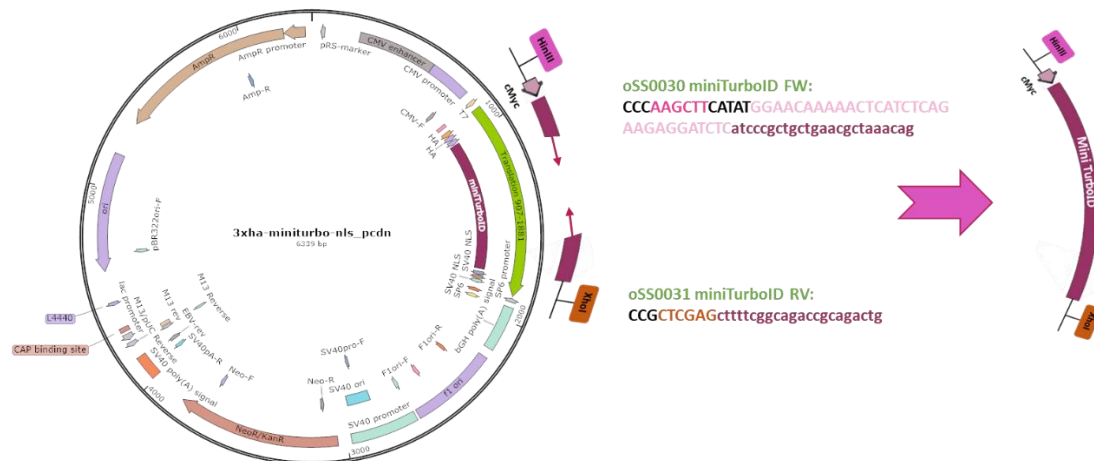


**Figure S7:** Generation of the *HindIII\_cMyc\_mTBirA\*\_XhoI* insert of interest by PCR

### ***B: Verification by PCR and sequencing of the correct acquisition of the fragment of interest***

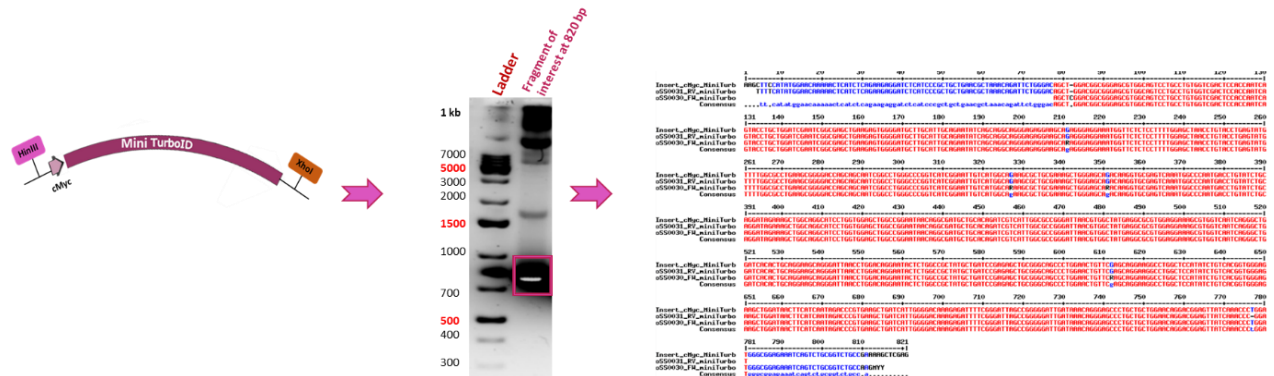


**Figure S8:** Verification of the acquisition of the *HindIII* cMyc mTBirA\* *XhoI* insert of interest by sequencing

***C: Verification by PCR and sequencing of the correct insertion of the fragment of interest into the pre-digested plasmid pLew100 cMyc BirA\* SNAP***

- **oSS0031 miniTurboID RV** 5'-CCGCTCGAGcttttcggcagaccgcagactg-3' targeting the hygromycin resistance gene region present in the main vector plasmid
- **oSS0032 miniTurboID FW** 5'-CTCTGGCCGCTATGCTGATC-3' targeting the **mTBirA\*** gene region



## APPENDIX 4

### 1- Plasmid pLew100\_cMyc\_BirA\*\_GGGS\_SNAP-tag

The parental plasmid pLew100\_cMyc\_BirA\*\_GGG\_SNAP-tag (8373 bp) was used as a template to perform site-directed mutagenesis to insert an additional glycine into the GGG linker.

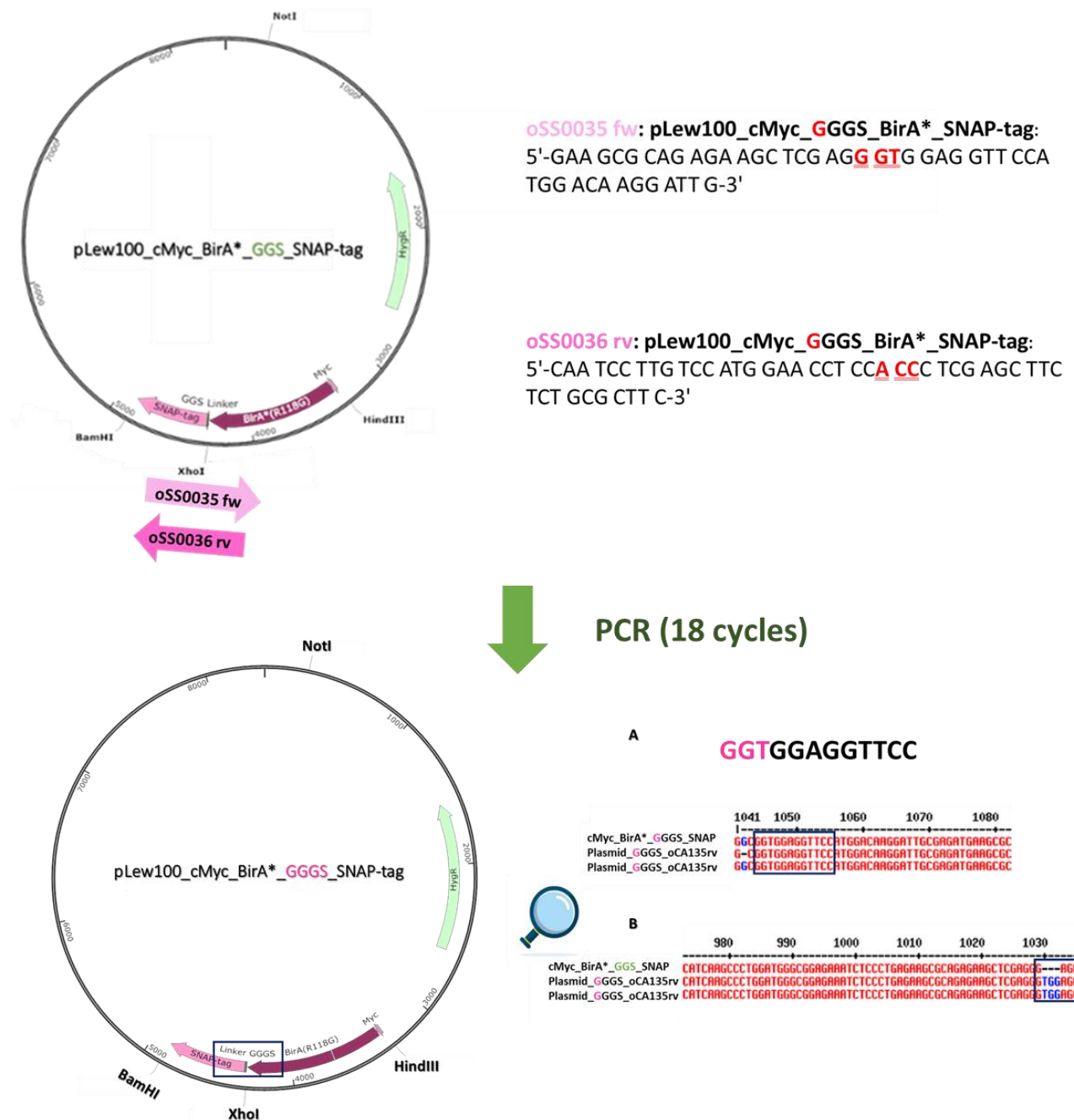


**Figure S10:** Steps for Directed Mutagenesis to Insert GGT Codon and Generate G1 Linker (GGGS)

## 2- PLew100\_cMyc\_BirA\*\_GGGS\_SNP-tag

The parental plasmid pLew100\_cMyc\_BirA\*\_GGGS\_SNP-tag (8376 bp) was used as a template to perform site-directed mutagenesis to insert an additional glycine into the GGGS linker.

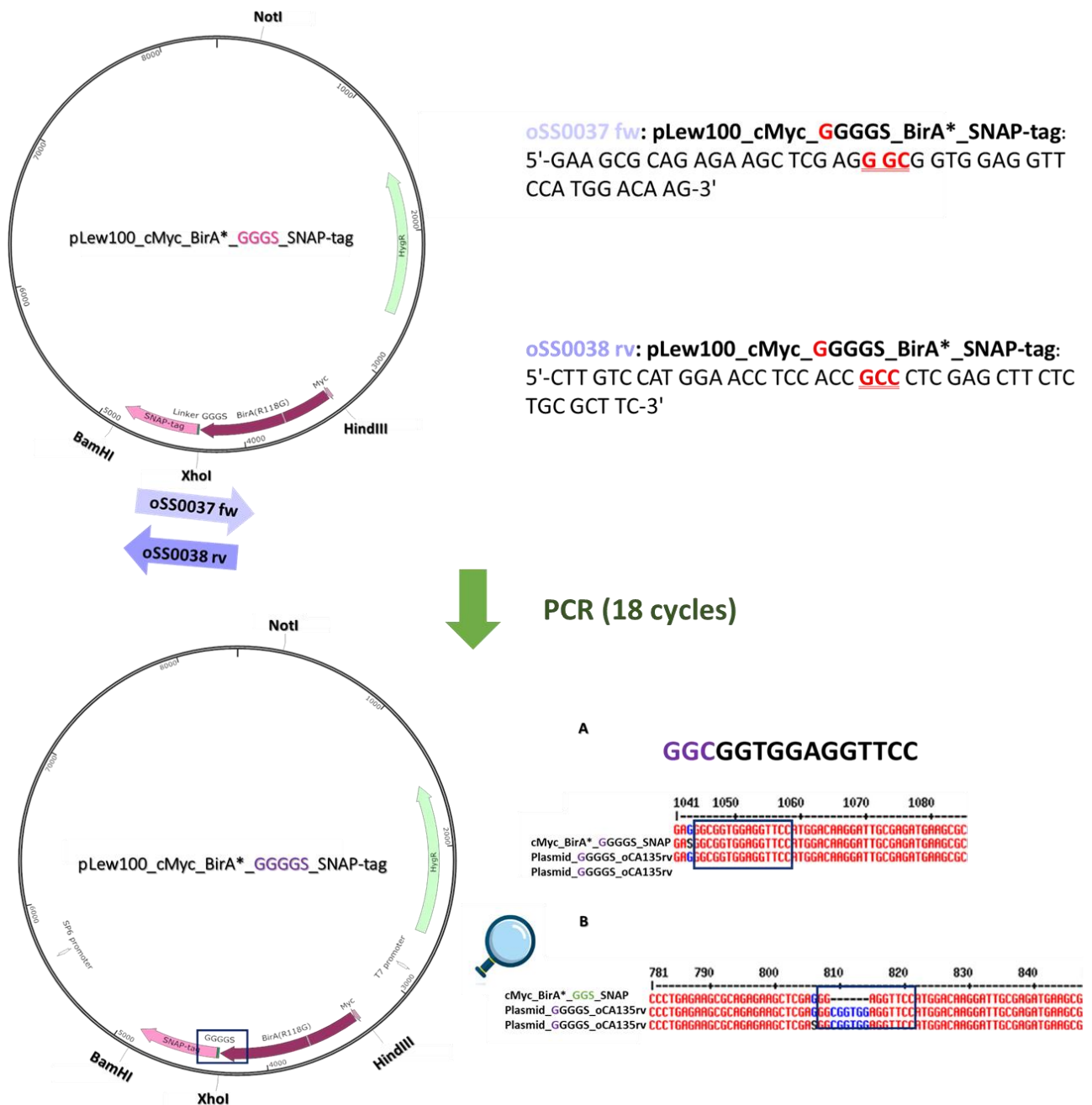


**Figure S11:** Steps for Directed Mutagenesis to Insert GGC Codon and Generate G2 Linker (GGGS)

## APPENDIX 5

### Plasmid pLew100\_cMyc\_BirA\*\_**PAPAP**\_SNAP-tag

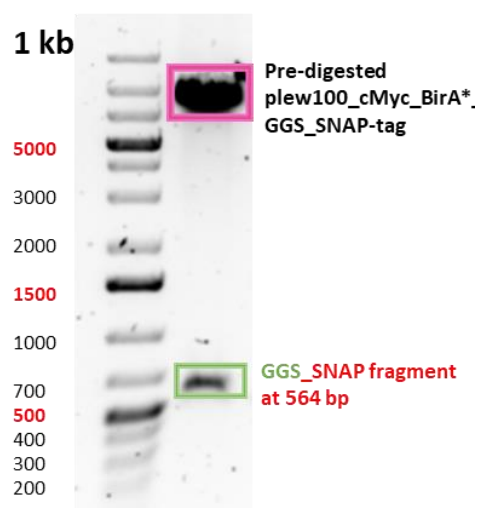
#### *A: Design and production of the insert of interest XhoI\_PAPAP\_SNAP-tag\_BamHI*



**Figure S12:** Generation of the XhoI\_PAPAP\_SNAP-tag\_BamHI insert of interest by PCR

#### *B: Verification by PCR and sequencing of the correct insertion of the fragment of interest into the pre-digested plasmid pLew100\_cMyc\_BirA\*\_SNAP*

The pre-digested plasmid was obtained after digestion with XhoI and BamHI, in which the new **PAPAP**\_SNAP-tag fragment of interest will be integrated after digestion with the same restriction enzymes.



Suzanne Sherihan SAHRAOUI  
**Part 3: Development of a new target deconvolution tool**

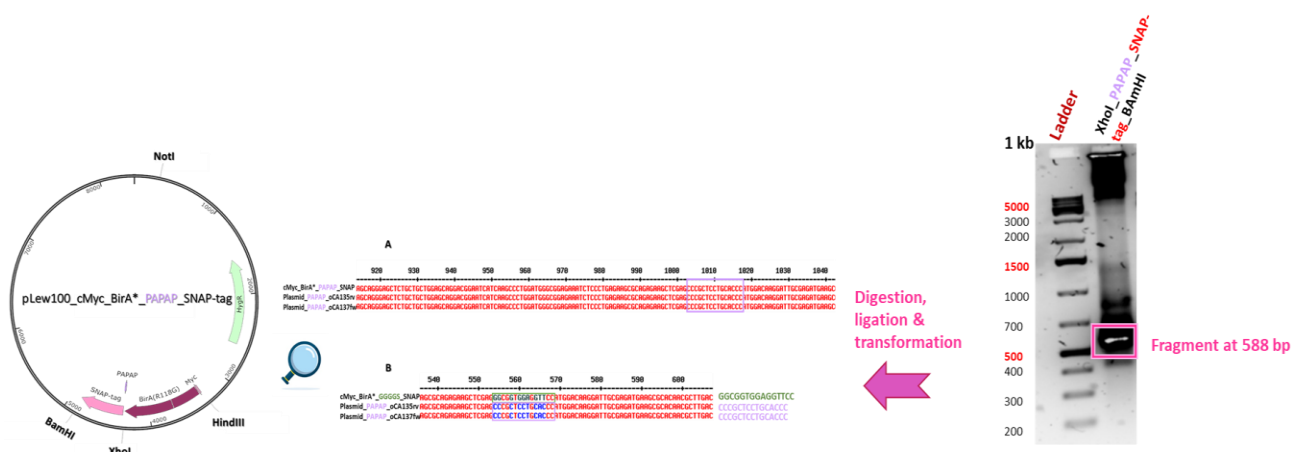


**Figure S13: PCR and Sequencing Verification of Successful Generation of pLew100\_cMyc\_BirA\*\_PAPAP\_SNAP-tag Plasmid**

**C: Verification by PCR and sequencing of the correct insertion of the fragment of interest into the genome of *Tbb\_WT***

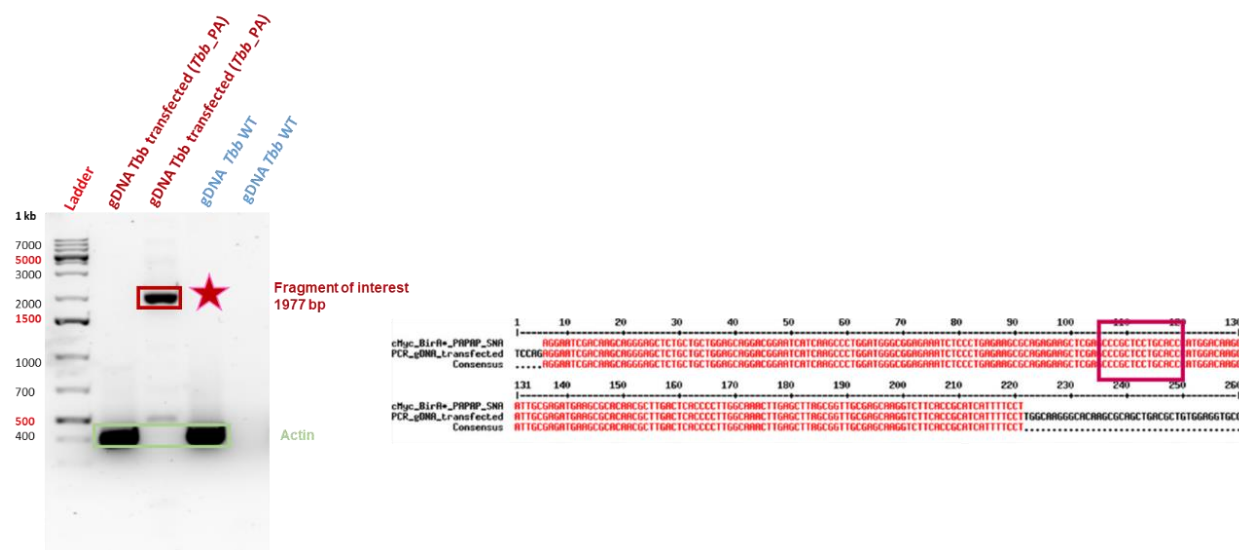


**Figure S14: PCR and Sequencing Verification of Accurate cMyc\_BirA\*\_PAPAP\_SNAP-tag Gene Insertion into *Tbb\_WT* Genome**

Genomic sequencing results (below) with the oSS00134 fw primer show the proper integration of the insert of interest cMyc\_BirA\*\_PAPAP\_SNAP-tag

# Part 3: Development of a new target deconvolution tool



**Figure S15:** Sequencing Verification of Accurate cMyc\_BirA\*\_PAPAP\_SNP-tag Gene Insertion into the genome of selected transfected *Tbb* clones (using oCA134 FW and oCA135 RV primers)

- **oCA134 FW:** 5'-CGCGCCTTCGAGTTTTTTTTTCC-3'==> targeting GPEET 5' UTR & amp upstream region contained in the main vector plasmid

- **oCA135 RV:** 5'-CCTGCTGTGCCATCAGATTAC-3'==> targeting ALD 3 & acute long UTR region contained in the main vector plasmid

## APPENDIX 6

### Plasmid pLew100\_cMyc\_BirA\*\_PTP\_SNAP-tag

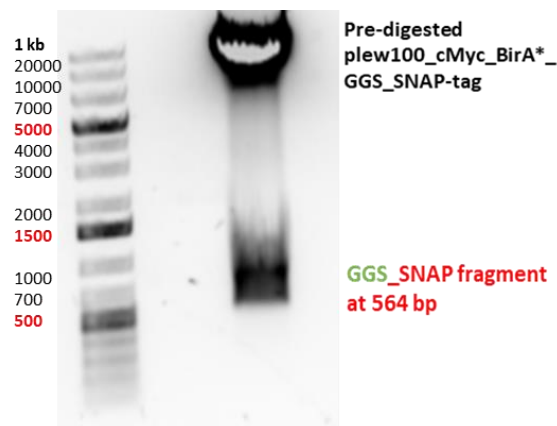
#### A: Design and production of the insert of interest *XhoI\_PTP\_SNAP-tag\_BamHI*



**Figure S16:** Generation of the *XhoI\_PTP\_SNAP-tag\_BamHI* insert of interest by PCR

#### B: Verification by PCR and sequencing of the correct insertion of the fragment of interest into the pre-digested plasmid pLew100\_cMyc\_BirA\*\_SNAP

The pre-digested plasmid was obtained after digestion with *XhoI* and *BamHI*, in which the new **PTP**\_SNAP-tag fragment of interest will be integrated after digestion with the same restriction enzymes.



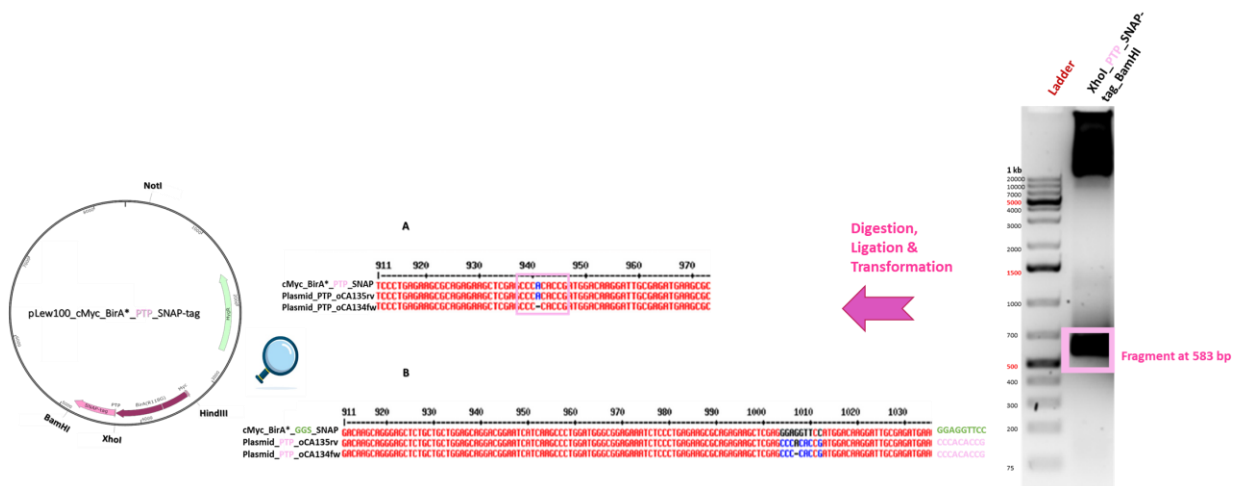
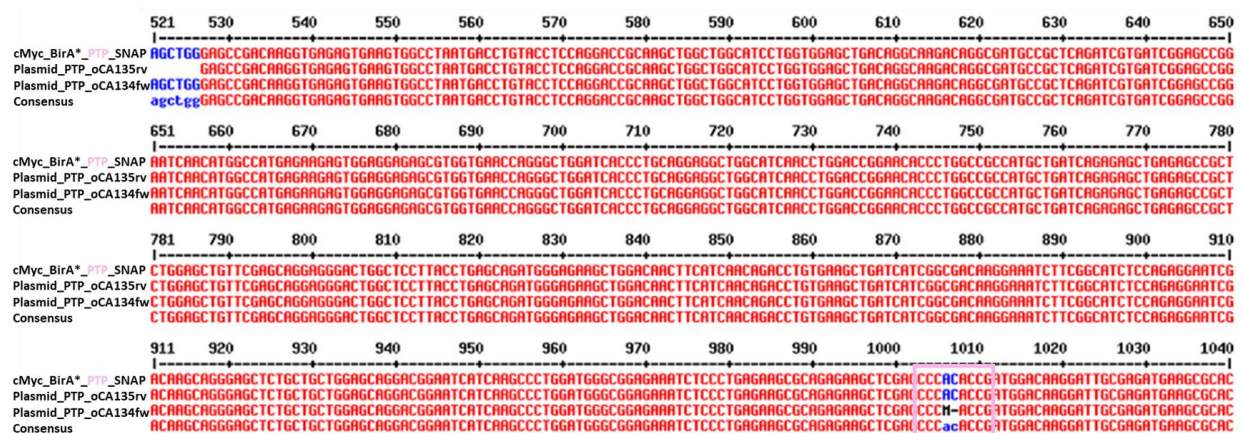
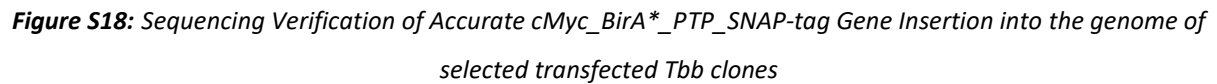


**Figure S17:** PCR and Sequencing Verification of Successful Generation of pLew100\_cMyc\_BirA\*\_PTP\_SNAP-tag Plasmid

The details of sequencing results: showing the good alignment between the insert of interest cMyc\_BirA\*\_PTP\_SNAP-tag and the resulting plasmid after cloning. The pink box shows the successful replacement of the GGS linker with the PTP linker



**C: Verification by PCR and sequencing of the correct insertion of the fragment of interest into the genome of Tbb\_WT**



Suzanne Sherihan SAHRAOUI  
*Part 3: Development of a new target deconvolution tool*

## APPENDIX 7

### 1- Growth curves *Tbb\_WT*

#### A: *Tbb\_WT* without tetracycline

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 1.00E+03                 | 2.50E+04                 | 2.25E+05                 |  | Tue | 24       | 2.00                     | 3.00                     | 4.40                     | 5.35                     |
| Wed | 48       | 1.75E+04                 | 1.10E+05                 | 1.00E+06                 | 5.63E+06                 |  | Wed | 48       | 4.24                     | 5.04                     | 6.00                     | 6.75                     |
| Thu | 72       | 3.25E+05                 | 2.70E+06                 | 6.50E+06                 | 8.75E+06                 |  | Thu | 72       | 5.51                     | 6.43                     | 6.81                     | 6.94                     |
| Fri | 96       | 2.60E+06                 | 4.00E+06                 | 7.75E+06                 | 1.00E+06                 |  | Fri | 96       | 6.41                     | 6.60                     | 6.89                     | 6.00                     |

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 1.00E+03                 | 2.25E+04                 | 2.40E+05                 |  | Tue | 24       | 2.00                     | 3.00                     | 4.35                     | 5.38                     |
| Wed | 48       | 1.75E+04                 | 9.90E+04                 | 7.25E+05                 | 5.00E+06                 |  | Wed | 48       | 4.24                     | 5.00                     | 5.86                     | 6.70                     |
| Thu | 72       | 2.50E+05                 | 2.80E+06                 | 6.13E+06                 | 8.40E+06                 |  | Thu | 72       | 5.40                     | 6.45                     | 6.79                     | 6.92                     |
| Fri | 96       | 1.40E+06                 | 4.80E+06                 | 4.63E+06                 | 1.00E+06                 |  | Fri | 96       | 6.15                     | 6.68                     | 6.67                     | 6.00                     |

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 3.75E+03                 | 6.25E+03                 | 4.25E+04                 | 3.60E+05                 |  | Tue | 24       | 3.57                     | 3.80                     | 4.63                     | 5.56                     |
| Wed | 48       | 1.60E+04                 | 9.75E+04                 | 7.25E+05                 | 5.63E+06                 |  | Wed | 48       | 4.20                     | 4.99                     | 5.86                     | 6.75                     |
| Thu | 72       | 4.75E+05                 | 2.70E+06                 | 6.00E+06                 | 7.80E+06                 |  | Thu | 72       | 5.68                     | 6.43                     | 6.78                     | 6.89                     |
| Fri | 96       | 5.75E+06                 | 4.50E+06                 | 1.10E+07                 | 6.00E+05                 |  | Fri | 96       | 6.76                     | 6.65                     | 7.04                     | 5.78                     |

#### B: *Tbb\_WT* with tetracycline (1 µg/mL)

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 2.50E+03                 | 1.38E+04                 | 1.20E+05                 | 2.05E+06                 |  | Tue | 24       | 3.40                     | 4.14                     | 5.08                     | 6.31                     |
| Wed | 48       | 1.63E+04                 | 1.25E+05                 | 1.61E+06                 | 7.50E+06                 |  | Wed | 48       | 4.21                     | 5.10                     | 6.21                     | 6.88                     |
| Thu | 72       | 1.75E+05                 | 2.38E+06                 | 7.50E+06                 | 4.50E+06                 |  | Thu | 72       | 5.24                     | 6.38                     | 6.88                     | 6.65                     |
| Fri | 96       | 2.41E+06                 | 1.30E+07                 | 6.63E+06                 | 1.50E+05                 |  | Fri | 96       | 6.38                     | 7.11                     | 6.82                     | 5.18                     |

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 1.13E+04                 | 1.24E+05                 | 1.09E+06                 |  | Tue | 24       | 2.00                     | 4.05                     | 5.09                     | 6.04                     |
| Wed | 48       | 1.38E+04                 | 1.38E+05                 | 1.13E+06                 | 4.50E+06                 |  | Wed | 48       | 4.14                     | 5.14                     | 6.05                     | 6.65                     |
| Thu | 72       | 6.25E+04                 | 1.75E+06                 | 8.63E+06                 | 4.38E+06                 |  | Thu | 72       | 4.80                     | 6.24                     | 6.94                     | 6.64                     |
| Fri | 96       | 4.50E+06                 | 6.63E+06                 | 1.03E+07                 | 8.75E+04                 |  | Fri | 96       | 6.65                     | 6.82                     | 7.01                     | 4.94                     |

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 1.25E+04                 | 1.44E+05                 | 1.65E+06                 |  | Tue | 24       | 2.00                     | 4.10                     | 5.16                     | 6.22                     |
| Wed | 48       | 1.63E+04                 | 1.13E+05                 | 1.33E+06                 | 6.25E+06                 |  | Wed | 48       | 4.21                     | 5.05                     | 6.12                     | 6.80                     |
| Thu | 72       | 4.63E+05                 | 4.25E+06                 | 7.38E+06                 | 5.25E+06                 |  | Thu | 72       | 5.67                     | 6.63                     | 6.87                     | 6.72                     |
| Fri | 96       | 6.13E+06                 | 1.10E+07                 | 6.38E+06                 | 1.00E+05                 |  | Fri | 96       | 6.79                     | 7.04                     | 6.80                     | 5.00                     |

**Part 3: Development of a new target deconvolution tool****2- Growth curves *Tbb\_G0*****A: *Tbb\_G0* electroporated *without tetracycline***

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 1.25E+03                 | 1.13E+04                 | 1.60E+05                 |  | Tue | 24       | 2.00                     | 3.10                     | 4.05                     | 5.20                     |
| Wed | 48       | 1.00E+02                 | 1.25E+03                 | 1.50E+04                 | 2.25E+05                 |  | Wed | 48       | 2.00                     | 3.10                     | 4.18                     | 5.35                     |
| Thu | 72       | 1.00E+02                 | 1.13E+04                 | 5.62E+04                 | 1.00E+06                 |  | Thu | 72       | 2.00                     | 4.05                     | 4.75                     | 6.00                     |
| Fri | 96       | 1.00E+02                 | 5.75E+04                 | 5.00E+05                 | 4.00E+06                 |  | Fri | 96       | 2.00                     | 4.76                     | 5.70                     | 6.60                     |

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 1.25E+03                 | 1.00E+04                 | 1.90E+05                 |  | Tue | 24       | 2.00                     | 3.10                     | 4.00                     | 5.28                     |
| Wed | 48       | 1.00E+02                 | 1.25E+03                 | 2.10E+04                 | 2.60E+05                 |  | Wed | 48       | 2.00                     | 3.10                     | 4.32                     | 5.41                     |
| Thu | 72       | 1.00E+02                 | 3.75E+03                 | 3.75E+04                 | 1.00E+06                 |  | Thu | 72       | 2.00                     | 3.57                     | 4.57                     | 6.00                     |
| Fri | 96       | 1.00E+02                 | 3.75E+03                 | 5.00E+04                 | 3.80E+06                 |  | Fri | 96       | 2.00                     | 3.57                     | 4.70                     | 6.58                     |

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 1.00E+03                 | 1.50E+04                 | 1.50E+05                 |  | Tue | 24       | 2.00                     | 3.00                     | 4.18                     | 5.18                     |
| Wed | 48       | 1.00E+02                 | 1.00E+03                 | 1.50E+04                 | 2.00E+05                 |  | Wed | 48       | 2.00                     | 3.00                     | 4.18                     | 5.30                     |
| Thu | 72       | 1.00E+02                 | 5.00E+03                 | 2.10E+04                 | 1.10E+06                 |  | Thu | 72       | 2.00                     | 3.70                     | 4.32                     | 6.04                     |
| Fri | 96       | 1.00E+02                 | 2.25E+04                 | 1.75E+05                 | 4.00E+06                 |  | Fri | 96       | 2.00                     | 4.35                     | 5.24                     | 6.60                     |

The electroporation does not affect the viability of parasites

**A: *Tbb\_G0* *without tetracycline***

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.25E+03                 | 3.75E+03                 | 3.88E+04                 | 2.25E+05                 |  | Tue | 24       | 3.10                     | 3.57                     | 4.59                     | 5.35                     |
| Wed | 48       | 2.50E+03                 | 7.50E+03                 | 7.50E+04                 | 1.53E+06                 |  | Wed | 48       | 3.40                     | 3.88                     | 4.88                     | 6.18                     |
| Thu | 72       | 2.50E+03                 | 1.00E+04                 | 2.63E+05                 | 4.11E+06                 |  | Thu | 72       | 3.40                     | 4.00                     | 5.42                     | 6.61                     |
| Fri | 96       | 2.50E+03                 | 3.25E+04                 | 1.19E+06                 | 9.00E+05                 |  | Fri | 96       | 3.40                     | 4.51                     | 6.07                     | 5.95                     |

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 1.00E+03                 | 1.75E+04                 | 5.63E+05                 |  | Tue | 24       | 2.00                     | 3.00                     | 4.24                     | 5.75                     |
| Wed | 48       | 1.25E+03                 | 3.75E+03                 | 5.25E+04                 | 1.03E+06                 |  | Wed | 48       | 3.10                     | 3.57                     | 4.72                     | 6.01                     |
| Thu | 72       | 2.50E+03                 | 2.75E+04                 | 6.38E+05                 | 4.06E+06                 |  | Thu | 72       | 3.40                     | 4.44                     | 5.80                     | 6.61                     |
| Fri | 96       | 2.50E+03                 | 1.75E+04                 | 1.23E+06                 | 1.34E+06                 |  | Fri | 96       | 3.40                     | 4.24                     | 6.09                     | 6.13                     |

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 1.25E+03                 | 2.25E+04                 | 1.10E+05                 |  | Tue | 24       | 3.10                     | 3.10                     | 4.35                     | 5.04                     |
| Wed | 48       | 1.25E+03                 | 3.75E+03                 | 7.75E+04                 | 1.10E+06                 |  | Wed | 48       | 3.40                     | 3.57                     | 4.89                     | 6.04                     |
| Thu | 72       | 2.50E+03                 | 3.63E+04                 | 4.13E+05                 | 3.00E+06                 |  | Thu | 72       | #REF!                    | 4.56                     | 5.62                     | 6.48                     |
| Fri | 96       | 7.50E+03                 | 2.20E+05                 | 2.40E+06                 | 2.80E+06                 |  | Fri | 96       | 3.88                     | 5.34                     | 6.38                     | 6.45                     |

**B: *Tbb\_G0* *with tetracycline* (1 µg/mL)**

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 6.25E+03                 | 5.38E+04                 | 5.75E+05                 |  | Tue | 24       | 2.00                     | 3.80                     | 4.73                     | 5.76                     |
| Wed | 48       | 2.50E+03                 | 3.50E+04                 | 3.38E+05                 | 2.00E+06                 |  | Wed | 48       | 3.40                     | 4.54                     | 5.53                     | 6.30                     |
| Thu | 72       | 2.50E+03                 | 6.50E+04                 | 6.25E+05                 | 1.75E+06                 |  | Thu | 72       | 3.40                     | 4.81                     | 5.80                     | 6.24                     |
| Fri | 96       | 5.00E+03                 | 2.13E+05                 | 1.39E+06                 | 5.63E+05                 |  | Fri | 96       | 3.70                     | 5.33                     | 6.14                     | 5.75                     |

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 1.25E+03                 | 3.63E+04                 | 3.25E+05                 |  | Tue | 24       | 2.00                     | 3.10                     | 4.56                     | 5.51                     |
| Wed | 48       | 2.50E+03                 | 3.50E+04                 | 2.75E+05                 | 1.96E+06                 |  | Wed | 48       | 3.40                     | 4.54                     | 5.44                     | 6.29                     |
| Thu | 72       | 2.50E+03                 | 2.88E+04                 | 4.75E+05                 | 1.95E+06                 |  | Thu | 72       | 3.40                     | 4.46                     | 5.68                     | 6.29                     |
| Fri | 96       | 3.75E+03                 | 1.13E+05                 | 1.38E+06                 | 5.50E+05                 |  | Fri | 96       | 3.57                     | 5.05                     | 6.14                     | 5.74                     |

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 1.25E+03                 | 4.13E+04                 | 4.13E+05                 |  | Tue | 24       | 2.00                     | 3.10                     | 4.62                     | 5.62                     |
| Wed | 48       | 1.25E+03                 | 2.88E+04                 | 2.88E+05                 | 2.20E+06                 |  | Wed | 48       | 3.10                     | 4.46                     | 5.46                     | 6.34                     |
| Thu | 72       | 2.50E+03                 | 5.25E+04                 | 4.13E+05                 | 2.25E+06                 |  | Thu | 72       | 3.40                     | 4.72                     | 5.62                     | 6.35                     |
| Fri | 96       | 3.75E+03                 | 1.00E+05                 | 1.10E+06                 | 6.75E+05                 |  | Fri | 96       | 3.57                     | 5.00                     | 6.04                     | 5.83                     |

## Part 3: Development of a new target deconvolution tool

3- Growth curves *Tbb*\_TG0A: *Tbb*\_TG0 without tetracycline

| Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon 0    | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| 24       | 1.00E+02                 | 2.50E+03                 | 3.38E+04                 | 2.00E+05                 |  | Tue 24   | 2.00                     | 3.40                     | 4.53                     | 5.30                     |
| 48       | 1.00E+02                 | 1.00E+03                 | 1.38E+04                 | 4.75E+05                 |  | Wed 48   | 2.00                     | 3.00                     | 4.14                     | 5.68                     |
| 72       | 1.00E+02                 | 1.00E+03                 | 6.50E+04                 | 2.00E+06                 |  | Thu 72   | 2.00                     | 3.00                     | 4.81                     | 6.30                     |
| 96       | 1.00E+02                 | 6.25E+03                 | 7.50E+04                 | 3.05E+06                 |  | Fri 96   | 2.00                     | 3.80                     | 4.88                     | 6.48                     |

| Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon 0    | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| 24       | 1.00E+02                 | 3.75E+03                 | 2.25E+04                 | 3.13E+05                 |  | Tue 24   | 2.00                     | 3.57                     | 4.35                     | 5.49                     |
| 48       | 3.75E+03                 | 1.38E+04                 | 1.41E+05                 | 1.29E+06                 |  | Wed 48   | 3.57                     | 4.14                     | 5.15                     | 6.11                     |
| 72       | 1.00E+02                 | 3.00E+04                 | 7.50E+05                 | 3.25E+06                 |  | Thu 72   | 2.00                     | 4.48                     | 5.88                     | 6.51                     |
| 96       | 1.00E+02                 | 9.63E+04                 | 2.11E+06                 | 3.13E+06                 |  | Fri 96   | 2.00                     | 4.98                     | 6.32                     | 6.49                     |

| Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon 0    | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| 24       | 3.80E+03                 | 1.10E+04                 | 4.60E+04                 | 4.63E+05                 |  | Tue 24   | 3.58                     | 4.04                     | 4.66                     | 5.67                     |
| 48       | 1.00E+02                 | 6.25E+03                 | 6.25E+04                 | 2.30E+06                 |  | Wed 48   | 2.00                     | 3.80                     | 4.80                     | 6.36                     |
| 72       | 2.50E+03                 | 1.13E+04                 | 1.20E+05                 | 1.63E+06                 |  | Thu 72   | 3.40                     | 4.05                     | 5.08                     | 6.21                     |
| 96       | 3.75E+03                 | 6.25E+04                 | 8.13E+05                 | 1.96E+06                 |  | Fri 96   | 3.57                     | 4.80                     | 5.91                     | 6.29                     |

B: *Tbb*\_TG0 with tetracycline (1 µg/mL)

| Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon 0    | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon 0    | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue 24   | 2.50E+03                 | 1.25E+04                 | 1.00E+05                 | 9.50E+05                 |  | Tue 24   | 3.40                     | 4.10                     | 5.00                     | 5.98                     |
| Wed 48   | 1.00E+04                 | 1.88E+04                 | 3.50E+05                 | 2.64E+06                 |  | Wed 48   | 4.00                     | 4.27                     | 5.54                     | 6.42                     |
| Thu 72   | 1.88E+04                 | 7.50E+04                 | 1.16E+06                 | 2.24E+06                 |  | Thu 72   | 4.27                     | 4.88                     | 6.07                     | 6.35                     |
| Fri 96   | 2.00E+04                 | 3.75E+05                 | 2.31E+06                 | 8.88E+05                 |  | Fri 96   | 4.30                     | 5.57                     | 6.36                     | 5.95                     |

| Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon 0    | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon 0    | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue 24   | 1.00E+02                 | 5.25E+03                 | 7.50E+04                 | 6.38E+05                 |  | Tue 24   | 2.00                     | 3.72                     | 4.88                     | 5.80                     |
| Wed 48   | 1.00E+02                 | 2.75E+04                 | 1.38E+05                 | 2.25E+06                 |  | Wed 48   | 2.00                     | 4.44                     | 5.14                     | 6.35                     |
| Thu 72   | 7.50E+03                 | 1.03E+05                 | 1.24E+06                 | 1.88E+06                 |  | Thu 72   | 3.88                     | 5.01                     | 6.09                     | 6.27                     |
| Fri 96   | 1.25E+04                 | 5.50E+05                 | 2.25E+06                 | 8.13E+05                 |  | Fri 96   | 4.10                     | 5.74                     | 6.35                     | 5.91                     |

| Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon 0    | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon 0    | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue 24   | 1.00E+02                 | 6.25E+03                 | 5.75E+04                 | 7.13E+05                 |  | Tue 24   | 2.00                     | 3.80                     | 4.76                     | 5.85                     |
| Wed 48   | 1.25E+03                 | 1.63E+04                 | 2.00E+05                 | 2.46E+06                 |  | Wed 48   | 3.10                     | 4.21                     | 5.30                     | 6.39                     |
| Thu 72   | 2.50E+03                 | 4.00E+04                 | 8.88E+05                 | 2.54E+06                 |  | Thu 72   | 3.40                     | 4.60                     | 5.95                     | 6.40                     |
| Fri 96   | 2.25E+04                 | 2.00E+05                 | 2.15E+06                 | 9.63E+05                 |  | Fri 96   | 4.35                     | 5.30                     | 6.33                     | 5.98                     |

4- Growth curves *Tbb*\_mTG0A: *Tbb*\_mTG0 without tetracycline

| Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon 0    | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon 0    | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue 24   | 1.00E+02                 | 1.25E+03                 | 3.13E+04                 | 1.50E+05                 |  | Tue 24   | 2.00                     | 3.10                     | 4.49                     | 5.18                     |
| Wed 48   | 5.00E+03                 | 2.00E+04                 | 2.21E+05                 | 2.84E+06                 |  | Wed 48   | 3.70                     | 4.30                     | 5.34                     | 6.45                     |
| Thu 72   | 7.50E+03                 | 7.63E+04                 | 1.03E+06                 | 3.75E+06                 |  | Thu 72   | 3.88                     | 4.88                     | 6.01                     | 6.57                     |
| Fri 96   | 2.88E+04                 | 3.00E+05                 | 5.38E+06                 | 1.99E+06                 |  | Fri 96   | 4.46                     | 5.48                     | 6.73                     | 6.30                     |

| Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon 0    | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon 0    | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue 24   | 1.25E+03                 | 3.75E+03                 | 4.25E+04                 | 3.68E+05                 |  | Tue 24   | 3.10                     | 3.57                     | 4.63                     | 5.57                     |
| Wed 48   | 2.50E+03                 | 3.75E+03                 | 1.60E+05                 | 3.00E+06                 |  | Wed 48   | 3.40                     | 3.57                     | 5.20                     | 6.48                     |
| Thu 72   | 3.75E+03                 | 4.38E+04                 | 1.00E+06                 | 3.60E+06                 |  | Thu 72   | 3.57                     | 4.64                     | 6.00                     | 6.56                     |
| Fri 96   | 3.25E+04                 | 4.20E+05                 | 3.70E+06                 | 2.60E+06                 |  | Fri 96   | 4.51                     | 5.62                     | 6.57                     | 6.41                     |

| Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon 0    | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon 0    | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue 24   | 1.00E+02                 | 1.00E+03                 | 6.00E+04                 | 5.25E+05                 |  | Tue 24   | 2.00                     | 3.00                     | 4.78                     | 5.72                     |
| Wed 48   | 1.25E+03                 | 5.00E+03                 | 6.63E+04                 | 2.30E+06                 |  | Wed 48   | 3.10                     | 3.70                     | 4.82                     | 6.36                     |
| Thu 72   | 1.25E+03                 | 5.00E+03                 | 1.64E+05                 | 3.13E+06                 |  | Thu 72   | 3.10                     | 3.70                     | 5.21                     | 6.49                     |
| Fri 96   | 1.25E+03                 | 1.13E+04                 | 6.85E+05                 | 1.70E+06                 |  | Fri 96   | 3.10                     | 4.05                     | 5.84                     | 6.23                     |

Suzanne Sherihan SAHRAOUI  
**Part 3: Development of a new target deconvolution tool**

**B: Tbb\_mTGO with tetracycline (1 µg/mL)**

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 3.75E+03                 | 3.75E+04                 | 5.63E+05                 |  | Tue | 24       | 2.00                     | 3.57                     | 4.57                     | 5.75                     |
| Wed | 48       | 6.25E+03                 | 2.88E+04                 | 2.53E+05                 | 2.60E+06                 |  | Wed | 48       | 3.80                     | 4.46                     | 5.40                     | 6.41                     |
| Thu | 72       | 6.25E+03                 | 8.00E+04                 | 7.50E+05                 | 2.50E+06                 |  | Thu | 72       | 3.80                     | 4.90                     | 5.88                     | 6.40                     |
| Fri | 96       | 1.13E+04                 | 1.13E+05                 | 1.18E+06                 | 9.13E+05                 |  | Fri | 96       | 4.05                     | 5.05                     | 6.07                     | 5.96                     |

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 2.50E+03                 | 6.50E+04                 | 2.88E+05                 |  | Tue | 24       | 2.00                     | 3.40                     | 4.81                     | 5.46                     |
| Wed | 48       | 3.75E+03                 | 1.88E+04                 | 2.09E+05                 | 2.28E+06                 |  | Wed | 48       | 3.57                     | 4.27                     | 5.32                     | 6.36                     |
| Thu | 72       | 1.00E+04                 | 2.63E+04                 | 5.50E+05                 | 2.63E+06                 |  | Thu | 72       | 4.00                     | 4.42                     | 5.74                     | 6.42                     |
| Fri | 96       | 7.50E+03                 | 6.25E+04                 | 9.25E+05                 | 1.58E+05                 |  | Fri | 96       | 3.88                     | 4.80                     | 5.97                     | 5.20                     |

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 6.25E+03                 | 3.88E+04                 | 5.00E+05                 |  | Tue | 24       | 2.00                     | 3.80                     | 4.59                     | 5.70                     |
| Wed | 48       | 2.50E+03                 | 1.25E+04                 | 1.08E+05                 | 1.93E+06                 |  | Wed | 48       | 3.40                     | 4.10                     | 5.03                     | 6.28                     |
| Thu | 72       | 3.75E+03                 | 3.00E+04                 | 3.50E+05                 | 3.00E+06                 |  | Thu | 72       | 3.57                     | 4.48                     | 5.54                     | 6.48                     |
| Fri | 96       | 5.00E+03                 | 7.50E+04                 | 8.75E+05                 | 1.30E+06                 |  | Fri | 96       | 3.70                     | 4.88                     | 5.94                     | 6.11                     |

**5- Growth curves Tbb\_G1**

**A: Tbb\_G1 (Clone 1 limited dilution 1:100) without tetracycline**

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 7.50E+03                 | 4.40E+04                 | 4.70E+05                 |  | Tue | 24       | 2.00                     | 3.88                     | 4.64                     | 5.67                     |
| Wed | 48       | 7.50E+03                 | 4.80E+04                 | 4.50E+05                 | 2.80E+06                 |  | Wed | 48       | 3.88                     | 4.68                     | 5.65                     | 6.45                     |
| Thu | 72       | 2.75E+04                 | 4.50E+05                 | 3.50E+06                 | 7.50E+06                 |  | Thu | 72       | 4.44                     | 5.65                     | 6.54                     | 6.88                     |
| Fri | 96       | 8.00E+05                 | 3.50E+06                 | 3.00E+06                 | 1.00E+06                 |  | Fri | 96       | 5.90                     | 6.54                     | 6.48                     | 6.00                     |

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.25E+03                 | 6.30E+03                 | 4.00E+04                 | 3.60E+05                 |  | Tue | 24       | 3.10                     | 3.80                     | 4.60                     | 5.56                     |
| Wed | 48       | 7.50E+03                 | 3.25E+04                 | 2.75E+05                 | 2.80E+06                 |  | Wed | 48       | 3.88                     | 4.51                     | 5.44                     | 6.45                     |
| Thu | 72       | 2.00E+04                 | 3.00E+05                 | 2.25E+06                 | 6.00E+06                 |  | Thu | 72       | 4.30                     | 5.48                     | 6.35                     | 6.78                     |
| Fri | 96       | 8.75E+05                 | 2.50E+06                 | 3.00E+06                 | 1.10E+06                 |  | Fri | 96       | 5.94                     | 6.40                     | 6.48                     | 6.04                     |

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.25E+03                 | 6.50E+03                 | 4.20E+04                 | 4.00E+05                 |  | Tue | 24       | 3.10                     | 3.81                     | 4.62                     | 5.60                     |
| Wed | 48       | 6.80E+03                 | 4.00E+04                 | 3.00E+05                 | 2.50E+06                 |  | Wed | 48       | 3.83                     | 4.60                     | 5.48                     | 6.40                     |
| Thu | 72       | 2.25E+04                 | 3.25E+05                 | 3.20E+06                 | 7.00E+06                 |  | Thu | 72       | 4.35                     | 5.51                     | 6.51                     | 6.85                     |
| Fri | 96       | 7.80E+05                 | 3.80E+06                 | 2.70E+06                 | 1.20E+06                 |  | Fri | 96       | 5.89                     | 6.58                     | 6.43                     | 6.08                     |

**B: Tbb\_G1 (Clone 1 limited dilution 1:100) with tetracycline (1 µg/mL)**

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 3.00E+03                 | 1.20E+04                 | 2.75E+05                 |  | Tue | 24       | 2.00                     | 3.48                     | 4.08                     | 5.44                     |
| Wed | 48       | 2.50E+03                 | 1.00E+04                 | 1.25E+05                 | 1.00E+06                 |  | Wed | 48       | 3.40                     | 4.00                     | 5.10                     | 6.00                     |
| Thu | 72       | 2.00E+04                 | 1.25E+05                 | 7.25E+05                 | 2.63E+06                 |  | Thu | 72       | 4.30                     | 5.10                     | 5.86                     | 6.42                     |
| Fri | 96       | 8.75E+04                 | 3.75E+05                 | 4.25E+06                 | 1.75E+06                 |  | Fri | 96       | 4.94                     | 5.57                     | 6.63                     | 6.24                     |

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 2.00E+03                 | 1.00E+04                 | 3.00E+05                 |  | Tue | 24       | 2.00                     | 3.30                     | 4.00                     | 5.48                     |
| Wed | 48       | 5.00E+03                 | 3.75E+04                 | 3.00E+05                 | 1.75E+06                 |  | Wed | 48       | 3.70                     | 4.57                     | 5.48                     | 6.24                     |
| Thu | 72       | 5.00E+03                 | 1.75E+05                 | 9.75E+05                 | 3.75E+06                 |  | Thu | 72       | 3.70                     | 5.24                     | 5.99                     | 6.57                     |
| Fri | 96       | 7.00E+04                 | 6.25E+05                 | 1.50E+06                 | 1.50E+06                 |  | Fri | 96       | 4.85                     | 5.80                     | 6.18                     | 6.18                     |

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 2.50E+03                 | 1.50E+04                 | 2.25E+05                 |  | Tue | 24       | 2.00                     | 3.40                     | 4.18                     | 5.35                     |
| Wed | 48       | 2.50E+03                 | 1.00E+04                 | 3.00E+05                 | 1.00E+06                 |  | Wed | 48       | 3.40                     | 4.00                     | 5.48                     | 6.00                     |
| Thu | 72       | 7.50E+03                 | 2.50E+05                 | 9.25E+05                 | 2.25E+06                 |  | Thu | 72       | 3.88                     | 5.40                     | 5.97                     | 6.35                     |
| Fri | 96       | 7.50E+04                 | 4.75E+05                 | 2.50E+06                 | 2.50E+06                 |  | Fri | 96       | 4.88                     | 5.68                     | 6.40                     | 6.40                     |

**Part 3: Development of a new target deconvolution tool****6- Growth curves *Tbb\_G2*****A: *Tbb\_G2* (Clone 1 limited dilution 1:100) *without tetracycline***

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.25E+03                 | 1.80E+04                 | 9.80E+04                 | 1.30E+06                 |  | Tue | 24       | 3.10                     | 4.26                     | 4.99                     | 6.11                     |
| Wed | 48       | 4.50E+04                 | 3.00E+05                 | 2.70E+06                 | 6.40E+06                 |  | Wed | 48       | 4.65                     | 5.48                     | 6.43                     | 6.81                     |
| Thu | 72       | 3.50E+05                 | 2.00E+05                 | 1.50E+06                 | 3.50E+06                 |  | Thu | 72       | 5.54                     | 5.30                     | 6.18                     | 6.54                     |
| Fri | 96       | 3.50E+06                 | 4.00E+06                 | 4.75E+06                 | 3.75E+05                 |  | Fri | 96       | 6.54                     | 6.60                     | 6.68                     | 5.57                     |

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.25E+03                 | 1.00E+04                 | 9.50E+04                 | 1.20E+06                 |  | Tue | 24       | 3.10                     | 4.00                     | 4.98                     | 6.08                     |
| Wed | 48       | 2.90E+04                 | 2.40E+05                 | 2.20E+06                 | 5.30E+06                 |  | Wed | 48       | 4.46                     | 5.38                     | 6.34                     | 6.72                     |
| Thu | 72       | 4.00E+05                 | 2.20E+05                 | 1.75E+06                 | 3.75E+06                 |  | Thu | 72       | 5.60                     | 5.34                     | 6.24                     | 6.57                     |
| Fri | 96       | 4.90E+06                 | 3.00E+06                 | 5.00E+06                 | 1.50E+05                 |  | Fri | 96       | 6.69                     | 6.48                     | 6.70                     | 5.18                     |

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+03                 | 1.20E+04                 | 9.20E+04                 | 1.00E+06                 |  | Tue | 24       | 3.00                     | 4.08                     | 4.96                     | 6.00                     |
| Wed | 48       | 3.00E+04                 | 2.50E+05                 | 2.30E+06                 | 5.50E+06                 |  | Wed | 48       | 4.48                     | 5.40                     | 6.36                     | 6.74                     |
| Thu | 72       | 3.75E+05                 | 2.25E+05                 | 2.00E+06                 | 3.50E+06                 |  | Thu | 72       | 5.57                     | 5.35                     | 6.30                     | 6.54                     |
| Fri | 96       | 4.20E+06                 | 3.00E+06                 | 4.50E+06                 | 2.75E+05                 |  | Fri | 96       | 6.62                     | 6.48                     | 6.65                     | 5.44                     |

**B: *Tbb\_G2* (Clone 1 limited dilution 1:100) *with tetracycline (1 µg/mL)***

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 5.00E+03                 | 2.00E+05                 | 1.10E+06                 |  | Tue | 24       | 2.00                     | 3.70                     | 5.30                     | 6.04                     |
| Wed | 48       | 1.25E+04                 | 1.20E+05                 | 9.75E+05                 | 4.75E+06                 |  | Wed | 48       | 4.10                     | 5.08                     | 5.99                     | 6.68                     |
| Thu | 72       | 2.50E+05                 | 2.20E+06                 | 7.00E+06                 | 4.50E+06                 |  | Thu | 72       | 5.40                     | 6.34                     | 6.85                     | 6.65                     |
| Fri | 96       | 2.80E+06                 | 4.75E+06                 | 3.00E+06                 | 7.75E+05                 |  | Fri | 96       | 6.45                     | 6.68                     | 6.48                     | 5.89                     |

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 7.50E+03                 | 1.40E+05                 | 9.50E+05                 |  | Tue | 24       | 2.00                     | 3.88                     | 5.15                     | 5.98                     |
| Wed | 48       | 1.75E+04                 | 1.30E+05                 | 1.10E+06                 | 5.75E+06                 |  | Wed | 48       | 4.24                     | 5.11                     | 6.04                     | 6.76                     |
| Thu | 72       | 4.70E+05                 | 2.00E+06                 | 7.25E+06                 | 5.50E+06                 |  | Thu | 72       | 5.67                     | 6.30                     | 6.86                     | 6.74                     |
| Fri | 96       | 4.30E+06                 | 5.60E+06                 | 3.25E+06                 | 4.75E+05                 |  | Fri | 96       | 6.63                     | 6.75                     | 6.51                     | 5.68                     |

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.25E+03                 | 5.00E+03                 | 3.40E+05                 | 9.25E+05                 |  | Tue | 24       | 3.10                     | 3.70                     | 5.53                     | 5.97                     |
| Wed | 48       | 2.00E+04                 | 1.10E+05                 | 1.20E+06                 | 5.50E+06                 |  | Wed | 48       | 4.30                     | 5.04                     | 6.08                     | 6.74                     |
| Thu | 72       | 4.40E+05                 | 2.20E+06                 | 5.75E+06                 | 5.25E+06                 |  | Thu | 72       | 5.64                     | 6.34                     | 6.76                     | 6.72                     |
| Fri | 96       | 3.80E+06                 | 5.10E+06                 | 3.90E+06                 | 4.00E+05                 |  | Fri | 96       | 6.58                     | 6.71                     | 6.59                     | 5.60                     |

**7- Growth curves *Tbb\_PA*****A: *Tbb\_PA* (Clone 2 limited dilution 1:50) *without tetracycline***

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 1.00E+04                 | 1.20E+05                 | 8.30E+05                 |  | Tue | 24       | 2.00                     | 4.00                     | 5.08                     | 5.92                     |
| Wed | 48       | 7.50E+03                 | 2.50E+05                 | 1.90E+06                 | 4.75E+06                 |  | Wed | 48       | 3.88                     | 5.40                     | 6.28                     | 6.68                     |
| Thu | 72       | 1.00E+06                 | 4.50E+06                 | 5.50E+06                 | 5.50E+06                 |  | Thu | 72       | 6.00                     | 6.65                     | 6.74                     | 6.74                     |
| Fri | 96       | 6.50E+06                 | 8.75E+06                 | 3.00E+05                 | 1.25E+03                 |  | Fri | 96       | 6.81                     | 6.94                     | 5.48                     | 3.10                     |

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 2.00E+04                 | 1.00E+05                 | 9.30E+05                 |  | Tue | 24       | 2.00                     | 4.30                     | 5.00                     | 5.97                     |
| Wed | 48       | 3.25E+04                 | 4.25E+05                 | 2.10E+06                 | 5.50E+06                 |  | Wed | 48       | 4.51                     | 5.63                     | 6.32                     | 6.74                     |
| Thu | 72       | 6.25E+05                 | 3.50E+06                 | 6.75E+06                 | 3.00E+06                 |  | Thu | 72       | 5.80                     | 6.54                     | 6.83                     | 6.48                     |
| Fri | 96       | 7.75E+06                 | 9.50E+06                 | 4.50E+05                 | 1.25E+03                 |  | Fri | 96       | 6.89                     | 6.98                     | 5.65                     | 3.10                     |

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 2.25E+04                 | 1.20E+05                 | 9.50E+05                 |  | Tue | 24       | 2.00                     | 4.35                     | 5.08                     | 5.98                     |
| Wed | 48       | 4.00E+04                 | 2.50E+05                 | 2.60E+06                 | 4.25E+06                 |  | Wed | 48       | 4.60                     | 5.40                     | 6.41                     | 6.63                     |
| Thu | 72       | 9.25E+05                 | 5.10E+06                 | 9.75E+06                 | 2.25E+06                 |  | Thu | 72       | 5.97                     | 6.71                     | 6.99                     | 6.35                     |
| Fri | 96       | 5.75E+06                 | 7.00E+06                 | 3.00E+05                 | 1.25E+03                 |  | Fri | 96       | 6.76                     | 6.85                     | 5.48                     | 3.10                     |

**Part 3: Development of a new target deconvolution tool**

**B: Tbb\_PA (Clone 2 limited dilution 1:50) with tetracycline (1 µg/mL)**

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.25E+03                 | 1.25E+04                 | 6.00E+04                 | 8.25E+05                 |  | Tue | 24       | 3.10                     | 4.10                     | 4.78                     | 5.92                     |
| Wed | 48       | 3.25E+04                 | 3.00E+05                 | 1.50E+06                 | 3.75E+06                 |  | Wed | 48       | 4.51                     | 5.48                     | 6.18                     | 6.57                     |
| Thu | 72       | 3.00E+05                 | 3.00E+06                 | 5.00E+06                 | 2.75E+06                 |  | Thu | 72       | 5.48                     | 6.48                     | 6.70                     | 6.44                     |
| Fri | 96       | 6.00E+06                 | 8.00E+06                 | 4.00E+06                 | 1.25E+04                 |  | Fri | 96       | 6.78                     | 6.90                     | 6.60                     | 4.10                     |

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.25E+03                 | 1.00E+04                 | 8.25E+04                 | 9.80E+05                 |  | Tue | 24       | 3.10                     | 4.00                     | 4.92                     | 5.99                     |
| Wed | 48       | 4.00E+04                 | 1.75E+05                 | 1.00E+06                 | 5.00E+06                 |  | Wed | 48       | 4.60                     | 5.24                     | 6.00                     | 6.70                     |
| Thu | 72       | 5.25E+05                 | 3.75E+06                 | 6.25E+06                 | 4.00E+06                 |  | Thu | 72       | 5.72                     | 6.57                     | 6.80                     | 6.60                     |
| Fri | 96       | 3.50E+06                 | 5.50E+06                 | 2.00E+06                 | 1.25E+04                 |  | Fri | 96       | 6.54                     | 6.74                     | 6.30                     | 4.10                     |

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.25E+03                 | 2.00E+04                 | 8.00E+04                 | 1.20E+06                 |  | Tue | 24       | 3.10                     | 4.30                     | 4.90                     | 6.08                     |
| Wed | 48       | 4.75E+04                 | 4.75E+05                 | 1.00E+06                 | 5.75E+06                 |  | Wed | 48       | 4.68                     | 5.68                     | 6.00                     | 6.76                     |
| Thu | 72       | 3.50E+05                 | 2.50E+06                 | 3.50E+06                 | 4.00E+06                 |  | Thu | 72       | 5.54                     | 6.40                     | 6.54                     | 6.60                     |
| Fri | 96       | 4.50E+06                 | 7.75E+06                 | 4.00E+06                 | 1.25E+04                 |  | Fri | 96       | 6.65                     | 6.89                     | 6.60                     | 4.10                     |

**C: Tbb\_PA (Clone 2 limited dilution 1:50) with tetracycline (10 µg/mL)**

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL | 10 µg/ml Tet |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--------------|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |              | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 5.00E+03                 | 1.75E+04                 | 1.50E+05                 |              | Tue | 24       | 2.00                     | 3.70                     | 4.24                     | 5.18                     |
| Wed | 48       | 1.00E+02                 | 1.50E+04                 | 1.75E+05                 | 1.90E+06                 |              | Wed | 48       | 2.00                     | 4.18                     | 5.24                     | 6.28                     |
| Thu | 72       | 3.75E+04                 | 4.90E+05                 | 2.50E+06                 | 4.75E+06                 |              | Thu | 72       | 4.57                     | 5.69                     | 6.40                     | 6.68                     |
| Fri | 96       | 9.75E+05                 | 4.10E+06                 | 5.00E+06                 | 1.50E+06                 |              | Fri | 96       | 5.99                     | 6.61                     | 6.70                     | 6.18                     |

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL | 10 µg/ml Tet |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--------------|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |              | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 7.50E+03                 | 2.00E+04                 | 2.25E+05                 |              | Tue | 24       | 2.00                     | 3.88                     | 4.30                     | 5.35                     |
| Wed | 48       | 1.00E+02                 | 1.00E+04                 | 1.25E+05                 | 1.20E+06                 |              | Wed | 48       | 2.00                     | 4.00                     | 5.10                     | 6.08                     |
| Thu | 72       | 4.25E+04                 | 3.75E+05                 | 2.20E+06                 | 5.25E+06                 |              | Thu | 72       | 4.63                     | 5.57                     | 6.34                     | 6.72                     |
| Fri | 96       | 8.50E+05                 | 3.30E+06                 | 5.75E+06                 | 1.75E+06                 |              | Fri | 96       | 5.93                     | 6.52                     | 6.76                     | 6.24                     |

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL | 10 µg/ml Tet |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--------------|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |              | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 7.50E+03                 | 1.50E+04                 | 1.50E+05                 |              | Tue | 24       | 2.00                     | 3.88                     | 4.18                     | 5.18                     |
| Wed | 48       | 1.00E+02                 | 3.75E+04                 | 3.25E+05                 | 2.10E+06                 |              | Wed | 48       | 2.00                     | 4.57                     | 5.51                     | 6.32                     |
| Thu | 72       | 1.75E+04                 | 4.00E+05                 | 2.60E+06                 | 5.25E+06                 |              | Thu | 72       | 4.24                     | 5.60                     | 6.41                     | 6.72                     |
| Fri | 96       | 6.00E+05                 | 3.10E+06                 | 5.00E+06                 | 2.25E+06                 |              | Fri | 96       | 5.78                     | 6.49                     | 6.70                     | 6.35                     |

**8- Growth curves Tbb\_PTP**

**A: Tbb\_PTP (Clone 1 limited dilution 1:200) without tetracycline**

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 1.00E+03                 | 2.00E+04                 | 8.25E+04                 |  | Tue | 24       | 2.00                     | 3.00                     | 4.30                     | 4.92                     |
| Wed | 48       | 7.50E+03                 | 1.00E+04                 | 2.20E+05                 | 1.80E+06                 |  | Wed | 48       | 3.88                     | 4.00                     | 5.34                     | 6.26                     |
| Thu | 72       | 6.50E+04                 | 1.00E+06                 | 1.60E+06                 | 3.50E+06                 |  | Thu | 72       | 4.81                     | 6.00                     | 6.20                     | 6.54                     |
| Fri | 96       | 1.10E+06                 | 3.00E+06                 | 3.25E+06                 | 4.00E+06                 |  | Fri | 96       | 6.04                     | 6.48                     | 6.51                     | 6.60                     |

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 1.00E+03                 | 1.50E+04                 | 7.50E+04                 |  | Tue | 24       | 2.00                     | 3.00                     | 4.18                     | 4.88                     |
| Wed | 48       | 5.00E+03                 | 2.25E+04                 | 1.20E+05                 | 2.40E+06                 |  | Wed | 48       | 3.70                     | 4.35                     | 5.08                     | 6.38                     |
| Thu | 72       | 8.75E+04                 | 6.75E+05                 | 1.90E+06                 | 5.25E+06                 |  | Thu | 72       | 4.94                     | 5.83                     | 6.28                     | 6.72                     |
| Fri | 96       | 2.10E+06                 | 3.25E+06                 | 5.25E+06                 | 2.50E+06                 |  | Fri | 96       | 6.32                     | 6.51                     | 6.72                     | 6.40                     |

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 8.00E+04                 |  | Tue | 24       | 2.00                     | 3.00                     | 4.00                     | 4.90                     |
| Wed | 48       | 5.00E+03                 | 2.50E+04                 | 1.90E+05                 | 1.50E+06                 |  | Wed | 48       | 3.70                     | 4.40                     | 5.28                     | 6.18                     |
| Thu | 72       | 6.00E+04                 | 1.10E+06                 | 2.50E+06                 | 3.25E+06                 |  | Thu | 72       | 4.78                     | 6.04                     | 6.40                     | 6.51                     |
| Fri | 96       | 1.30E+06                 | 3.25E+06                 | 7.50E+06                 | 3.50E+06                 |  | Fri | 96       | 6.11                     | 6.51                     | 6.88                     | 6.54                     |

Suzanne Sherihan SAHRAOUI  
**Part 3: Development of a new target deconvolution tool**

**B: Tbb\_PTP (Clone 1 limited dilution 1:200) with tetracycline (1 µg/mL)**

|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|--|-----|----------|--------------------------|--------------------------|--------------------------|--------------------------|
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 5.00E+04                 |  | Tue | 24       | 2.00                     | 3.00                     | 4.00                     | 4.70                     |
| Wed | 48       | 1.25E+04                 | 2.25E+04                 | 2.20E+05                 | 1.20E+06                 |  | Wed | 48       | 4.10                     | 4.35                     | 5.34                     | 6.08                     |
| Thu | 72       | 1.70E+05                 | 5.25E+05                 | 2.75E+06                 | 1.75E+06                 |  | Thu | 72       | 5.23                     | 5.72                     | 6.44                     | 6.24                     |
| Fri | 96       | 9.50E+05                 | 2.80E+06                 | 4.25E+06                 | 1.50E+06                 |  | Fri | 96       | 5.98                     | 6.45                     | 6.63                     | 6.18                     |
|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 5.50E+04                 |  | Tue | 24       | 2.00                     | 3.00                     | 4.00                     | 4.74                     |
| Wed | 48       | 7.50E+03                 | 2.75E+04                 | 2.30E+05                 | 1.40E+06                 |  | Wed | 48       | 3.88                     | 4.44                     | 5.36                     | 6.15                     |
| Thu | 72       | 1.10E+05                 | 8.00E+05                 | 2.60E+06                 | 2.00E+06                 |  | Thu | 72       | 5.04                     | 5.90                     | 6.41                     | 6.30                     |
| Fri | 96       | 7.00E+05                 | 3.90E+06                 | 3.00E+06                 | 1.25E+06                 |  | Fri | 96       | 5.85                     | 6.59                     | 6.48                     | 6.10                     |
|     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |  |     | Time (h) | 10 <sup>2</sup> cells/mL | 10 <sup>3</sup> cells/mL | 10 <sup>4</sup> cells/mL | 10 <sup>5</sup> cells/mL |
| Mon | 0        | 1.00E+02                 | 1.00E+03                 | 1.00E+04                 | 1.00E+05                 |  | Mon | 0        | 2.00                     | 3.00                     | 4.00                     | 5.00                     |
| Tue | 24       | 1.00E+02                 | 1.00E+03                 | 7.50E+03                 | 3.50E+04                 |  | Tue | 24       | 2.00                     | 3.00                     | 3.88                     | 4.54                     |
| Wed | 48       | 2.10E+04                 | 3.50E+04                 | 1.80E+05                 | 1.20E+06                 |  | Wed | 48       | 4.32                     | 4.54                     | 5.26                     | 6.08                     |
| Thu | 72       | 6.25E+04                 | 9.50E+05                 | 1.30E+06                 | 2.75E+06                 |  | Thu | 72       | 4.80                     | 5.98                     | 6.11                     | 6.44                     |
| Fri | 96       | 6.50E+05                 | 1.70E+06                 | 3.50E+06                 | 2.75E+06                 |  | Fri | 96       | 5.81                     | 6.23                     | 6.54                     | 6.44                     |

## APPENDIX 8

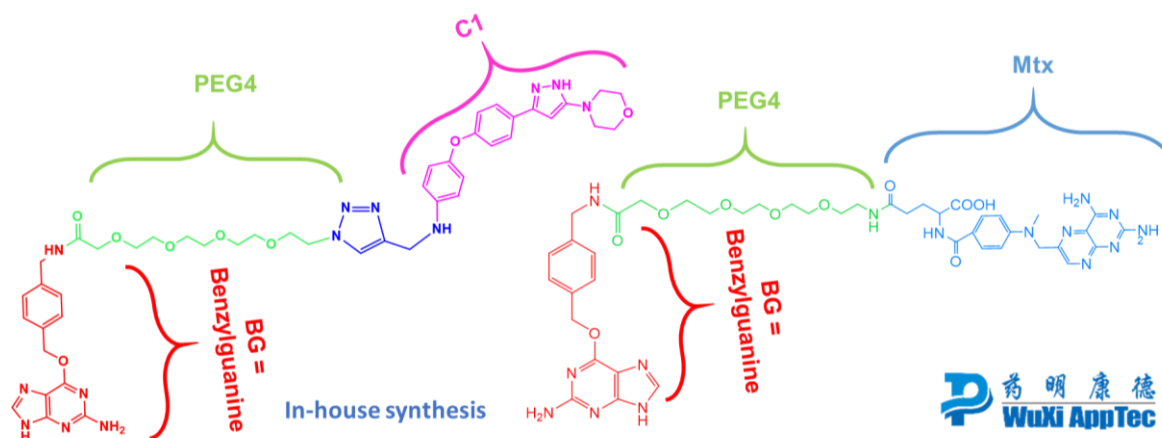


Figure S19: structures of the two BG-PEG4 derivatives: BG-PEG4-C1(left) and BG-PEG4-Mtx (right)

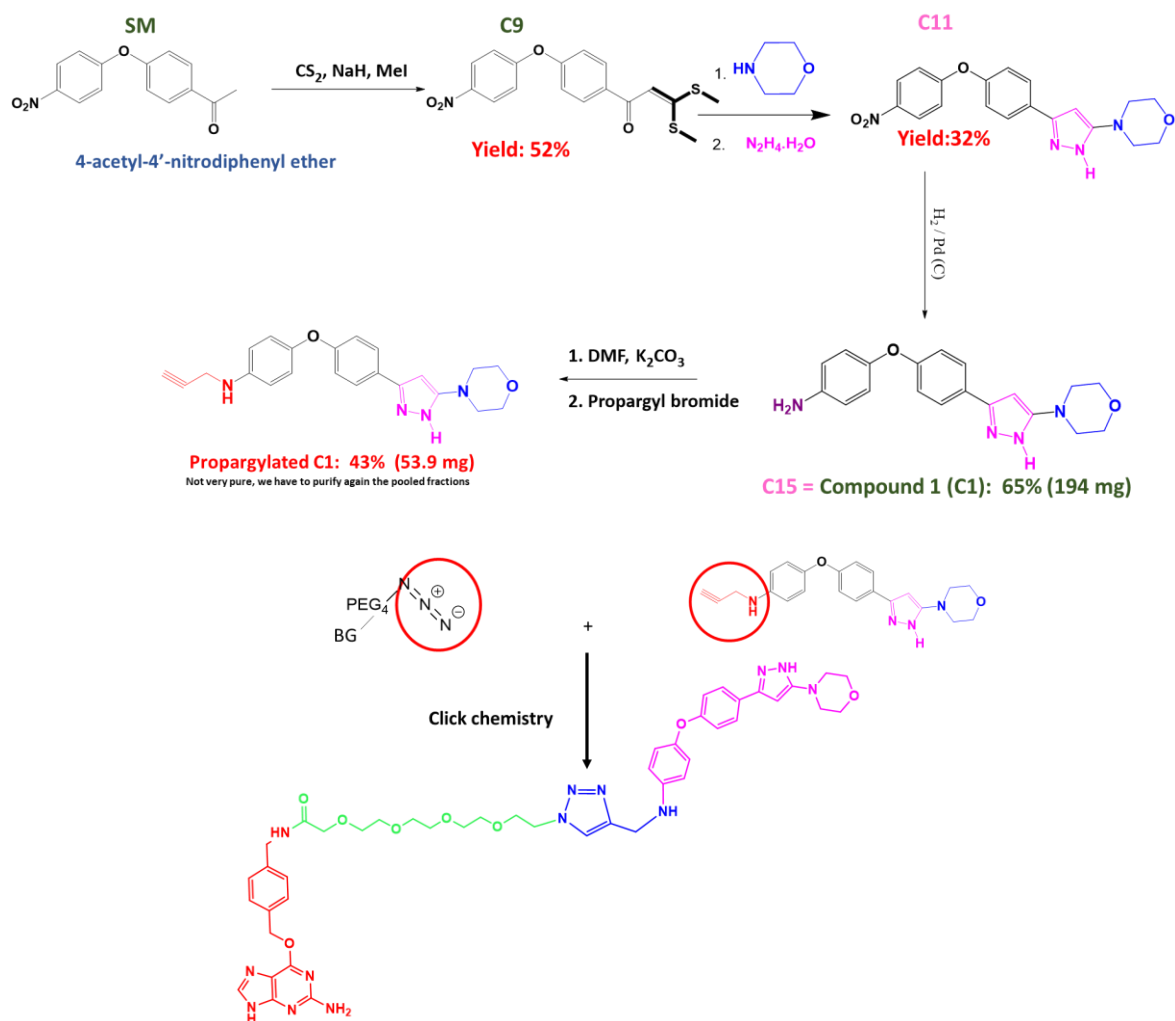


Figure S20: BG-PEG4-C1 synthesis steps

\*SM: starting material

The synthesis of the compound is completed by click chemistry by reacting the azide (the N3 part of BG-PEG<sub>4</sub>-N3) with the propargylated C1 to give the final molecule BG-PEG<sub>4</sub>-C1.

## APPENDIX 9

Different ratios have been carried out and then tested by bacterial transformation to determine which ones will lead to the formation of colonies.

¶ Ligation reactions to ligate digested **GGG\_SNAP-tag** into digested pLew100\_cMyc\_BirA\* to generate **pLew100\_cMyc\_BirA\*\_GGG\_SNAP-tag**

| plasmid                      | size (kb) | conc. (ng/ul) |
|------------------------------|-----------|---------------|
| Digested pLew 100_cMyc_BirA* | 7.824     | 35.4          |

| insert                         | size (kb) | conc. (ng/ul) |
|--------------------------------|-----------|---------------|
| Digested Insert (GGG_SNAP-tag) | 0.576     | 12.7          |

|                     |    |
|---------------------|----|
| Volume total (ul)   | 10 |
| Volume plasmid (ul) | 1  |

|              | control 1 | control 2 | ratio 1:<br>1 | ratio 1:<br>3 | ratio 1:<br>5 | ratio 1:<br>7 | ratio 1:<br>9 |
|--------------|-----------|-----------|---------------|---------------|---------------|---------------|---------------|
| contents     |           |           |               |               |               |               |               |
| buffer 10X   | 1         | 1         | 1             | 1             | 1             | 1             | 1             |
| plasmid      | 1         | 1         | 1             | 1             | 1             | 1             | 1             |
| insert       | 0         | 0         | 0.2           | 0.6           | 1.0           | 1.4           | 1.8           |
| ligase       | 1         | 0         | 1             | 1             | 1             | 1             | 1             |
| water        | 7         | 8         | 6.8           | 6.4           | 6.0           | 5.6           | 5.2           |
| total volume | 10        | 10        | 10            | 10            | 10            | 10            | 10            |

¶ Ligation reactions to ligate cMyc\_TBirA\* into digested pLew100\_GGS\_SNAP-tag to generate **pLew100\_cMyc\_TBirA\*\_GGG\_SNAP-tag**

| plasmid                        | size (kb) | conc. (ng/ul) |
|--------------------------------|-----------|---------------|
| Digested pLew 100_GGS_SNAP-tag | 8.367     | 32.5          |

| insert                        | size (kb) | conc. (ng/ul) |
|-------------------------------|-----------|---------------|
| Digested Insert (cMyc_TBirA*) | 1.011     | 27            |

|                     |    |
|---------------------|----|
| Volume total (ul)   | 10 |
| Volume plasmid (ul) | 1  |

|              | control 1 | control 2 | ratio 1:<br>1 | ratio 1:<br>3 | ratio 1:<br>5 | ratio 1:<br>7 | ratio 1:<br>9 |
|--------------|-----------|-----------|---------------|---------------|---------------|---------------|---------------|
| contents     |           |           |               |               |               |               |               |
| buffer 10X   | 1         | 1         | 1             | 1             | 1             | 1             | 1             |
| plasmid      | 1         | 1         | 1             | 1             | 1             | 1             | 1             |
| insert       | 0         | 0         | 0.1           | 0.4           | 0.7           | 1.0           | 1.3           |
| ligase       | 1         | 0         | 1             | 1             | 1             | 1             | 1             |
| water        | 7         | 8         | 6.9           | 6.6           | 6.3           | 6.0           | 5.7           |
| total volume | 10        | 10        | 10            | 10            | 10            | 10            | 10            |

¶ Ligation reactions to ligate cMyc\_mTBirA\* into digested pLew100\_GGS\_SNAP-tag to generate **pLew100\_cMyc\_mTBirA\*\_GGG\_SNAP-tag**

**Part 3: Development of a new target deconvolution tool**

| plasmid                      | size (kb) | conc. (ng/ul) |
|------------------------------|-----------|---------------|
| Digested pLew 100_cMyc_BirA* | 8.373     | 52            |

| insert                         | size (kb) | conc. (ng/ul) |
|--------------------------------|-----------|---------------|
| Digested Insert (cMyc_mTBirA*) | 0.825     | 12            |

|                     |    |
|---------------------|----|
| Volume total (ul)   | 10 |
| Volume plasmid (ul) | 1  |

|              | control 1 | control 2 | ratio 1:<br>1 | ratio 1:<br>3 | ratio 1:<br>5 | ratio 1:<br>7 | ratio 1:<br>9 |
|--------------|-----------|-----------|---------------|---------------|---------------|---------------|---------------|
| contents     |           |           |               |               |               |               |               |
| buffer 10X   | 1         | 1         | 1             | 1             | 1             | 1             | 1             |
| plasmid      | 1         | 1         | 1             | 1             | 1             | 1             | 1             |
| insert       | 0         | 0         | 0.4           | 1.3           | 2.1           | 3.0           | 3.8           |
| ligase       | 1         | 0         | 1             | 1             | 1             | 1             | 1             |
| water        | 7         | 8         | 6.6           | 5.7           | 4.9           | 4.0           | 3.2           |
| total volume | 10        | 10        | 10            | 10            | 10            | 10            | 10            |

¶ Ligation reactions to ligate digested **PAPAP**\_SNAP-tag into digested pLew100\_cMyc\_BirA\* to generate **pLew100\_cMyc\_BirA\*\_PAPAP\_SNA**P-tag

| plasmid                     | size (kb) | conc. (ng/ul) |
|-----------------------------|-----------|---------------|
| Digested pLew100_cMyc_BirA* | 8.367     | 18.5          |

| insert                    | size (kb) | conc. (ng/ul) |
|---------------------------|-----------|---------------|
| Digested Insert PAPAP_SNA | 0.588     | 13.1          |

|                     |    |
|---------------------|----|
| Volume total (ul)   | 10 |
| Volume plasmid (ul) | 1  |

|              | control 1 | control 2 | ratio 1:<br>1 | ratio 1:<br>3 | ratio 1:<br>5 | ratio 1:<br>7 | ratio 1:<br>9 |
|--------------|-----------|-----------|---------------|---------------|---------------|---------------|---------------|
| contents     |           |           |               |               |               |               |               |
| buffer 10X   | 1         | 1         | 1             | 1             | 1             | 1             | 1             |
| plasmid      | 1         | 1         | 1             | 1             | 1             | 1             | 1             |
| insert       | 0         | 0         | 0.1           | 0.3           | 0.5           | 0.7           | 0.9           |
| ligase       | 1         | 0         | 1             | 1             | 1             | 1             | 1             |
| water        | 7         | 8         | 6.9           | 6.7           | 6.5           | 6.3           | 6.1           |
| total volume | 10        | 10        | 10            | 10            | 10            | 10            | 10            |

¶ Ligation reactions to ligate digested **PTP**\_SNAP-tag into digested pLew100\_cMyc\_BirA\* to generate **pLew100\_cMyc\_BirA\*\_PTP\_SNA**P-tag

*Part 3: Development of a new target deconvolution tool*

| plasmid                     | size (kb) | conc. (ng/ul) |
|-----------------------------|-----------|---------------|
| Digested pLew100_cMyc_BirA* | 8.367     | 99.2          |

| insert                       | size (kb) | conc. (ng/ul) |
|------------------------------|-----------|---------------|
| Digested Insert PTP_SNAP-tag | 0.583     | 39.9          |

|                     |    |
|---------------------|----|
| Volume total (ul)   | 10 |
| Volume plasmid (ul) | 1  |

|                 | control 1 | control 2 | ratio 1:<br>1 | ratio 1:<br>3 | ratio 1:<br>5 | ratio 1:<br>7 | ratio 1:<br>9 |
|-----------------|-----------|-----------|---------------|---------------|---------------|---------------|---------------|
| contents        |           |           |               |               |               |               |               |
| buffer 10X      | 1         | 1         | 1             | 1             | 1             | 1             | 1             |
| plasmid         | 1         | 1         | 1             | 1             | 1             | 1             | 1             |
| insert          | 0         | 0         | 0.2           | 0.5           | 0.9           | 1.2           | 1.6           |
| ligase          | 1         | 0         | 1             | 1             | 1             | 1             | 1             |
| water           | 7         | 8         | 6.8           | 6.5           | 6.1           | 5.8           | 5.4           |
| total<br>volume | 10        | 10        | 10            | 10            | 10            | 10            | 10            |

## APPENDIX 10

MS results of BG-PEG<sub>4</sub>-Mtx VS BG-PEG<sub>4</sub>-Biotin linked to the peptide of interest containing Cys145:

Peptide of interest containing Cys145: TALSGNPVPILIP**C**HR

¶ BG-PEG<sub>4</sub>-Mtx (+788.360573)

User AA Formula 1: C<sub>2</sub> H<sub>3</sub> N<sub>1</sub> O<sub>1</sub>

| MH <sup>+</sup> 1(av) | MH <sup>+</sup> 1(mono) | MH <sup>+</sup> 2(av) | MH <sup>+</sup> 2(mono) | MH <sup>+</sup> 3(av) | MH <sup>+</sup> 3(mono) | MH <sup>+</sup> 4(av) | MH <sup>+</sup> 4(mono) |
|-----------------------|-------------------------|-----------------------|-------------------------|-----------------------|-------------------------|-----------------------|-------------------------|
| 2477.4058             | 2476.2918               | 1239.2066             | 1238.6495               | 826.4735              | 826.1021                | 620.1070              | 619.8284                |

¶ BG-PEG<sub>4</sub>-Biotin (+578.277421)

User AA Formula 1: C<sub>2</sub> H<sub>3</sub> N<sub>1</sub> O<sub>1</sub>

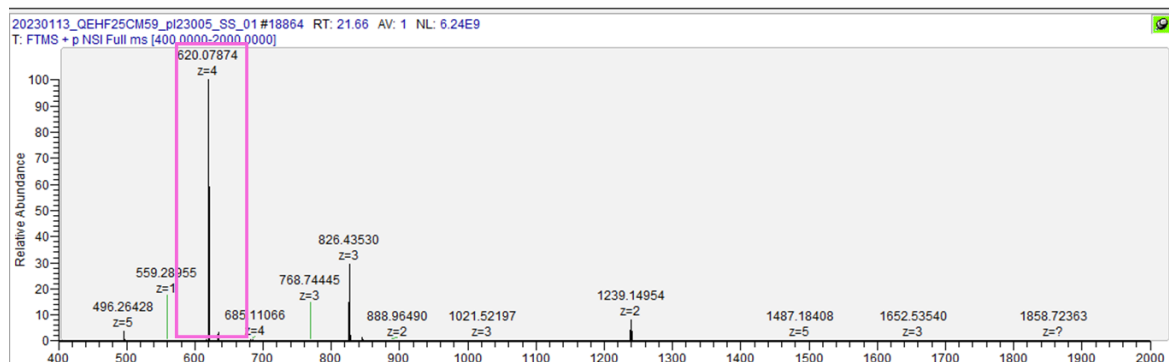
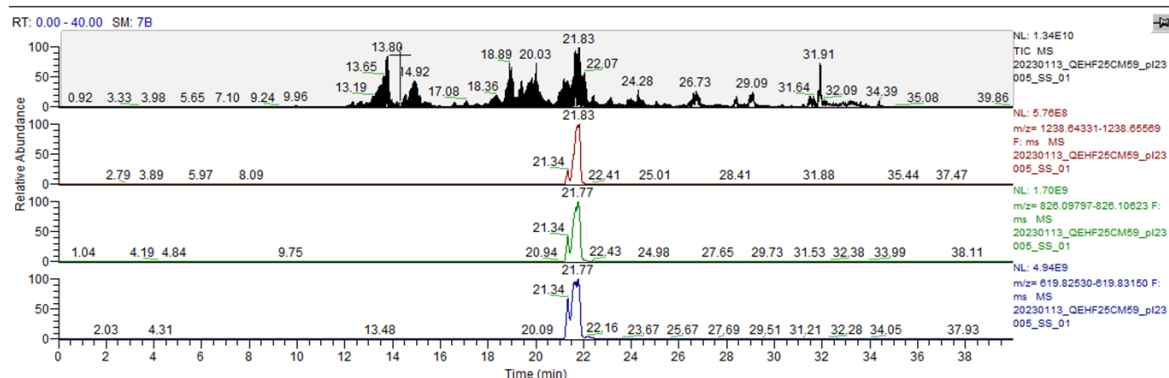
| MH <sup>+</sup> 1(av) | MH <sup>+</sup> 1(mono) | MH <sup>+</sup> 2(av) | MH <sup>+</sup> 2(mono) | MH <sup>+</sup> 3(av) | MH <sup>+</sup> 3(mono) | MH <sup>+</sup> 4(av) | MH <sup>+</sup> 4(mono) |
|-----------------------|-------------------------|-----------------------|-------------------------|-----------------------|-------------------------|-----------------------|-------------------------|
| 2267.3226             | 2266.2086               | 1134.1650             | 1133.6080               | 756.4458              | 756.0744                | 567.5862              | 567.3076                |

¶ Raw data of BG-PEG<sub>4</sub>-Mtx linked to Cys145 of the peptide of interest

TALSGNPVPILIP**C**HR : TALSGNPVPILIP**C**(BG-PEG<sub>4</sub>-Mtx)HR

20230113\_QEHF25CM59\_pl23005\_SS\_01

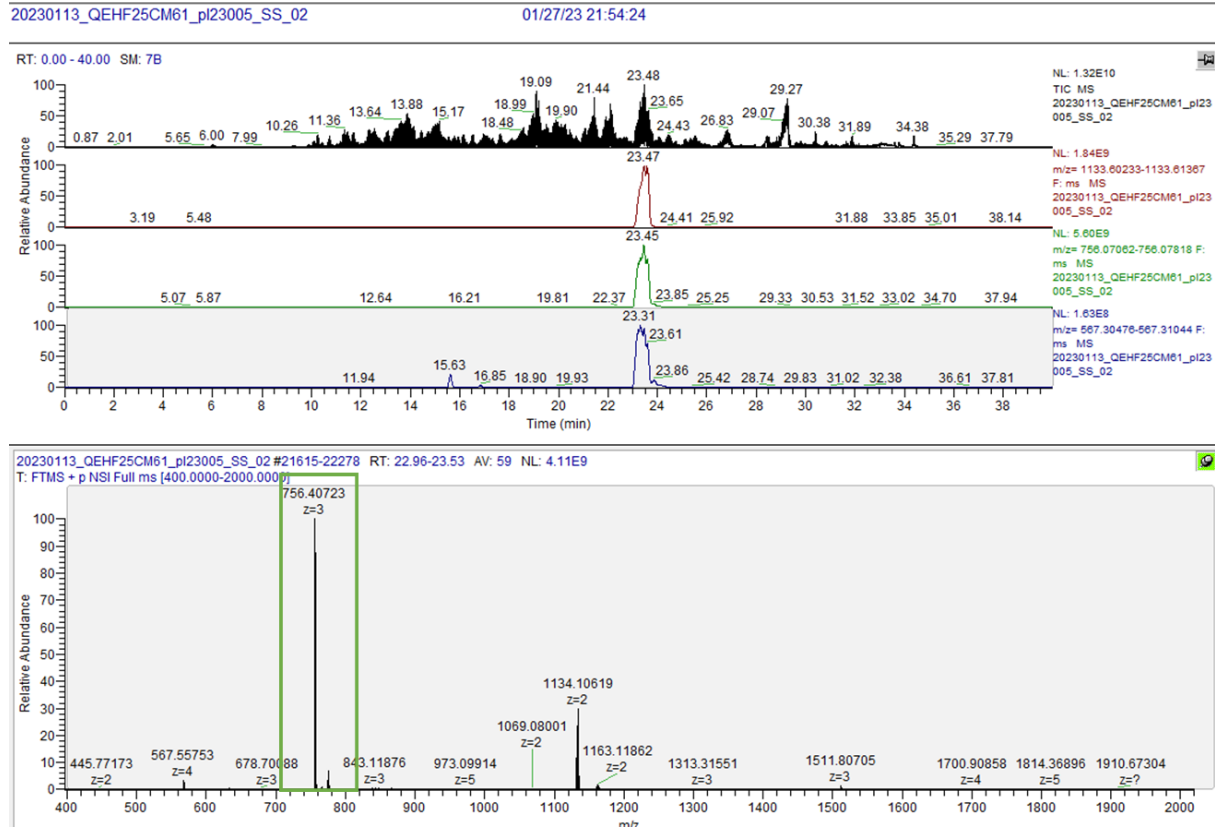
01/27/23 19:33:05



Suzanne Sherihan SAHRAOUI  
*Part 3: Development of a new target deconvolution tool*

¶ Raw data of BG-PEG<sub>4</sub>-Biotin linked to Cys145 of the peptide of interest

TALSGNPVPILIPCHR: TALSGNPVPILIPC(BG-PEG<sub>4</sub>-Biotin)HR



The two peptides of interest carrying the two Cys145-linked derivatives emerge about at the same time, which is clearly apparent (23 min for BG-PEG<sub>4</sub>-Biotin and 21 min for BG-PEG<sub>4</sub>-Mtx). The peak's evident presence indicates that Cys145 has indeed been modified by the BG-PEG<sub>4</sub>-Mtx derivative.

Suzanne Sherihan SAHRAOUI  
**Part 3: Development of a new target deconvolution tool**

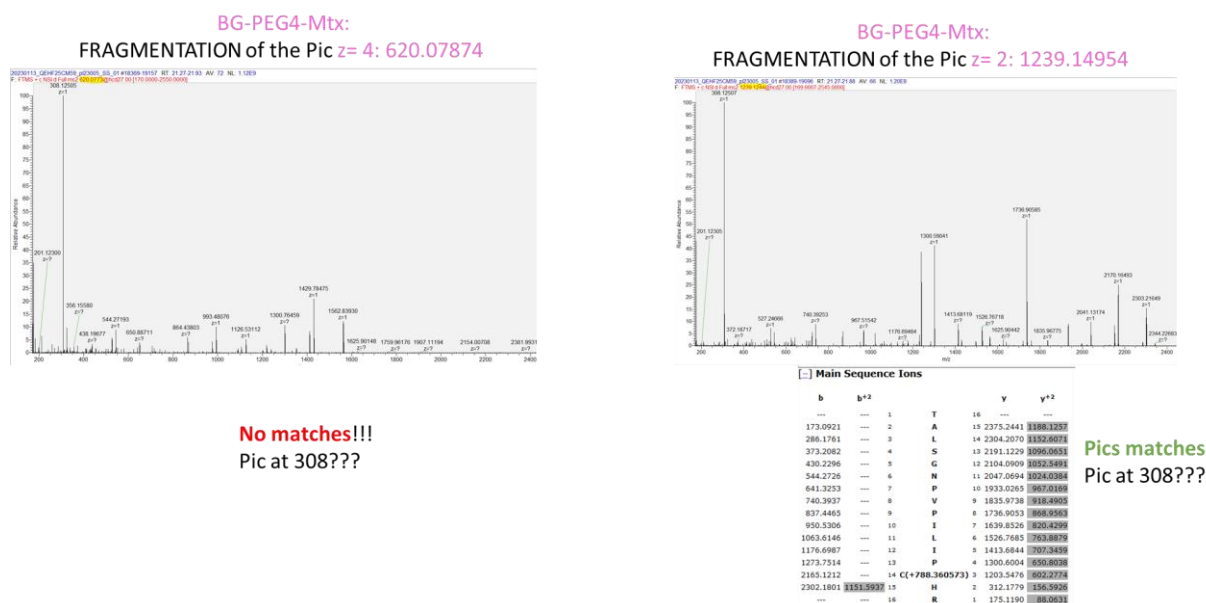


**Figure S21:** LC-MS/MS Fragmentation: No Match Found for Strongly Charged Major Peak ( $z=4$ ) of the Target Peptide [TALSGNPVILIPCHR] Modified by BG-PEG4-Mtx

As can be observed from the graphs above, no match could be found in the fragmentation of the strongly charged major peak  $z=4$  ( $z=4$ : 620.07874), but a correlation was found in the fragmentation of the weak peak  $z=2$  ( $z=2$ : 1239.14954). However, the match's score is significantly low to allow for the conclusion that Cys145 was modified by the BG-PEG<sub>4</sub>-Mtx derivative. Moreover, a distinct and powerful peak at 308 that is unknown arises.

However, in the instance of BG-PEG<sub>4</sub>-Biotin (graph below), the fragmentation of the major peak at  $z=3$  (756.0744) and the minor signal at  $z=2$  (1134.10619) clearly correlate, and this allows one to infer the presence of the derivative's binding to the relevant cysteine in the SNAP active site.

**Suzanne Sherihan SAHRAOUI**  
**Part 3: Development of a new target deconvolution tool**

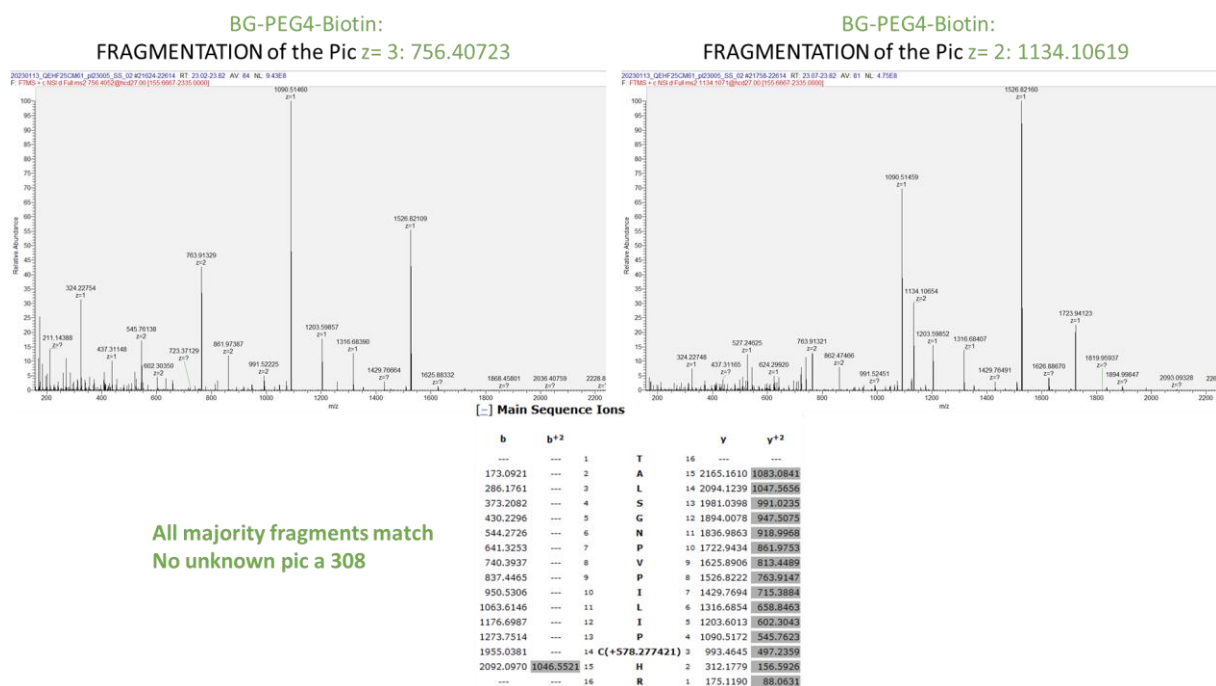


**Figure S22: LC-MS/MS Fragmentation: Clear Match Found for Strongly Charged Major Peak (z=3) and Minor Signal at z=2 of the Target Peptide [TALSGNPVILIPCHR] Modified by BG-PEG4-Biotin**

## **.7 Posters, talks, and publications**

### **.7.1 Talks**

- **EFMC-ISMIC 2021 XXVI EFMC International Symposium on Medicinal Chemistry** (August 29th, 2021)
- **P.h.D day** (November 18th, 2022): Suzanne Sherihan Sahraoui: A novel tool for small molecule target identification in cytosolic and organellar sub-proteomes of Trypanosoma under in vivo conditions
- **LS2 Annual Meeting 2022\_Talk & Poster\_ 2022:** LS2 – Life Sciences Switzerland University of Zurich, Campus Irchel, 20-22 April 2022
- **Swiss Society of Pharmacology and Toxicology 2021: Talk:** A novel tool for small molecule target identification in Trypanosoma brucei parasites: Suzanne Sherihan Sahraoui, Sébastien Tardy, Oscar Vadas, Aurélie Goullier, Olivier Petermann, Leonardo Scapozza
- "Ma thèse en 180" (MT180) édition 2023: Geneva finals reached: link to the video of the presentation: <https://www.unige.ch/cite/evenements/ma-these-en-180-secondes/ma-these-en-180-secondes-2023/les-candidat-es/suzanne-sherihan-sahraoui>

### **.7.2 Posters**

- **LS2 Annual Meeting 2021\_Poster\_ 2021 (February 17th, 2021):** Drug Discovery Chemical Biology Section: Validation of a novel tool for small molecule target identification in cytosolic and organellar sub-proteomes of Trypanosoma brucei under in vivo conditions
- **Swiss Society of Pharmacology and Toxicology 2021 (SSPT meeting 2021)** (Flash talk presentation), April 15th, 2021



# A novel tool for small molecule target identification in cytosolic and organellar sub-proteomes of *Trypanosoma brucei* under *in-vivo* conditions

Suzanne Sherihan Sahraoui<sup>1,2</sup>, Sébastien Tardy<sup>1,2</sup>, Oscar Vadas<sup>3</sup>, Aurélie Gouiller<sup>1,2</sup>, Olivier Petermann<sup>1,2</sup>, Leonardo Scapozza<sup>1,2</sup>

IPSP0  
INSTITUT DES SCIENCES  
PHARMACEUTIQUES  
DE GENÈVE

<sup>1</sup> Institute of Pharmaceutical Sciences of Western Switzerland, University of Geneva, Geneva, Switzerland  
<sup>2</sup> Department of Microbiology and Molecular Medicine, University of Geneva, Geneva, Switzerland  
<sup>3</sup> Institute of Pharmaceutical Sciences of Western Switzerland, University of Geneva, Geneva, Switzerland

## INTRODUCTION

- ❖ Neglected diseases as the Human African Trypanosomiasis (HAT) caused by *Trypanosoma brucei* (Tb) (Tb rhodesiense (acute form), Tb gambiense (chronic form)) [Figures 1,2] rely on drugs associated with many side effects which cause a real problem for the effective treatment of populations [1]
- ❖ The phenotype screening approach is widely used for drug discovery but targeted proteins or the mechanism of action remain unknown
- ❖ Some methods have been developed to identify targets and elucidate mechanisms of action but these methods present some limitations [Figure 3]
- ❖ This project has been designed to overcome the limitations of existing target isolation techniques [Figure 4,5]

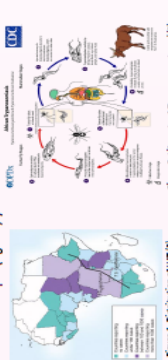


Figure 1. Distribution of HAT (2).

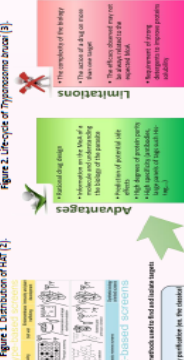


Figure 2. Life-cycle of Trypanosoma brucei (3).

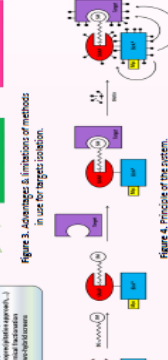


Figure 3. Advantages & Limitations of methods in use for target location.

**Aim of the project**  
The main objective of the project is to develop a new system that will allow the identification of small molecules targets identification in cytosolic and organellar sub-proteomes of *Trypanosoma brucei* under *in-vivo* conditions. The project will therefore provide a sensitive identification tool suitable for small molecules that interact with moderate to the high affinity with their target proteins.

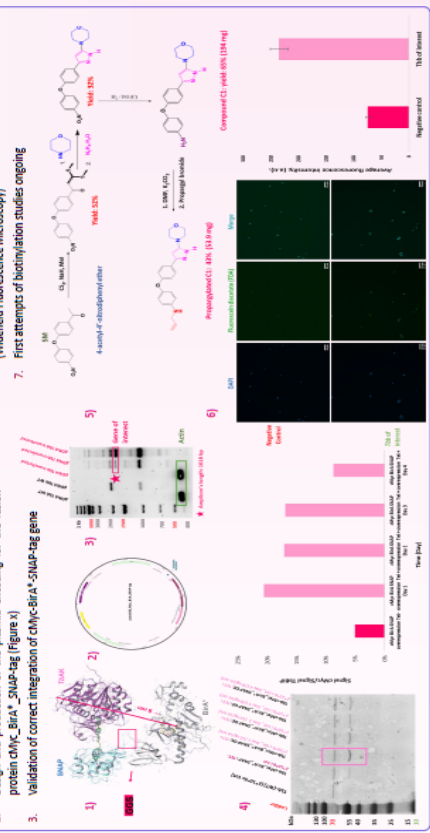
## METHODS

1. Cell culture: *Trypanosoma brucei* brucei (Tb) bloodstream form (BSF)
2. Molecular biology part: plex100 carrying hygromycin resistance and BioL technology is chosen as a template to generate plex100-BirA\* SNUAP-tag plasmid encoding for the fusion protein of interest.
3. Genetics: The electroporation method is used to transfect Tb with the plasmid of interest containing the *rimM*-*spacer* sequence for the homologous recombination necessary to obtain stable cell lines.
4. Chemical part: BG-PEG4-C1 compound is used as a test compound for the proof-of-concept as it targets the adenosine kinase of Tb (TbAK). The synthesis of C1 moiety is completed and will be linked to BG-PEG4 part by click chemistry (step 5 ongoing).
5. Targets biotinylation: Biotinylation occurs after adding in the medium the tetracycline for the induction of the fusion protein expression, and the biotin for BirA\* protein part activation.
6. Targets isolation and identification: This step is performed by western blot and LC/MS/MS Mass Spectrometry.

Figure 4. Biotinylation technique for target location.

## RESULTS

1. Some steps have been successfully completed:
2. Modeling study for the evaluation of a flexible linker length necessary between the two main domains (BirA\* and SNUAP tag) composing the fusion protein.
3. Design and production of the plasmid encoding for the fusion protein cMyC-BirA\* SNUAP-tag (Figure 1).
4. Validation of correct integration of cMyC-BirA\* SNUAP-tag gene.
5. Validation of cMyC-BirA\* SNUAP-tag expression in Tb (LS-MS/MS and Western blot (Figure 1)).
6. Synthesis of C1 moiety of the compound test BG-PEG4-C1.
7. Validation of permeation of BG-PEG4-FDA (fluorescein diacetate) (Widefield Fluorescence Microscopy).
8. First attempts of biotinylation studies ongoing.



## CONCLUSIONS

- ❖ The Biotin approach: a recently developed method that has proven its advantage in terms of detection of weak and transient interactions that may be missed or lost during harsh purification procedures (4).
- ❖ Its combination with SNAP-tag technology brings the very specific and selective labeling, which will lead to the development of an innovative tool and will allow the discovery of therapeutic targets.
- ❖ The specific labeling of this tool will make it potentially transferable to other more pathogenic parasites with a medical interest such as *Plasmodium falciparum*, *Toxoplasma*, *Leishmania*...

## OUTLOOKS

- ❖ The next step for the progress of the project:
- ❖ Continuity of biotinylation studies for the proof of concept, through the use of other compounds having different affinities for their targets such as BG-PEG4-memotretate (Mn) that binds to the dihydrodipicolinate reductase (DHFR) (Figure 6).
- ❖ Expression of the tool in the organelles of the parasite (translocation of the system to subcellular organelles).
- ❖ Use of the developed system for target identification of existing drugs such as suramin, pentamidine, nifurtimox, of drug candidates, and even of hits and leads under development.

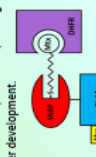


Figure 6. Capture of the DHFR target via the interaction between cMyC-BirA\* SNUAP-tag and BG-PEG4-Mn.

## REFERENCES

1. deSilva P, Sanyal A, Olyos M, Ishiguro-Saito A, Homan K. A current analysis of chemotherapy strategies for the treatment of human African trypanosomiasis. *Parasitology* (2015) 145:52-62.
2. Biss R, van E, Jacobs E, Wang M, Barrett M. Development of novel drugs for human African trypanosomiasis. *Parasitology* (2015) 145:52-62.
3. CDC African trypanosomiasis. <https://www.cdc.gov/parasites/trypanosomiasis/index.html> (2015) Jan 145. Available from: <https://www.cdc.gov/parasites/trypanosomiasis/index.html>
4. Montanari B, Hedrick K, Chinnai L, Yess S, Sanyal A, Cardenas A, Williams K, et al. Novel biotinylation components in Trypanosoma brucei identified using gravity-dependent biotinylation. *Eukaryotic cell* (2015) 15:1258-67.

## CONTACT

Suzanne Sherihan Sahraoui  
E-mail: [suzanne.sahraoui@unige.ch](mailto:suzanne.sahraoui@unige.ch)  
University of Geneva-Pharmaceutical Biochemistry unit  
Address: Rue Michel Servet 1 / CHU-8007 1521A / CH-1211 GENEVA 4  
LinkedIn: <http://fr.linkedin.com/in/suzanne-sherihan-sahraoui-23615106>



