Computational advances in discovering cryptic pockets for drug discovery

Bemelmans, Martijn; Cournia, Zoe; Damm-Ganamet, Kelly L; Gervasio, Francesco Luigi; Pande, Vineet

How to cite

BEMELMANS, Martijn et al. Computational advances in discovering cryptic pockets for drug discovery. In: Current opinion in structural biology, 2025, vol. 90, p. 102975. doi: 10.1016/j.sbi.2024.102975

This publication URL: https://archive-ouverte.unige.ch/unige:184301
Publication DOI: 10.1016/j.sbi.2024.102975

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Computational advances in discovering cryptic pockets for drug discovery

Martijn P. Bemelmans[1,4], Zoe Cournia[2], Kelly L. Damm-Ganamet[3], Francesco L. Gervasio[4] and Vineet Pande[1]

A number of promising therapeutic target proteins have been considered "undruggable" due to the lack of well-defined ligandable pockets. Substantial research in protein dynamics has elucidated the existence of "cryptic" pockets that only exist transiently and become favorable for binding in the presence of a ligand. These pockets provide an avenue to target challenging proteins, inspiring the development of multiple computational methods. This review highlights established cryptic pocket modeling approaches like mixed solvent molecular dynamics and recent applications of enhanced sampling and AI-based methods in therapeutically relevant proteins.

Addresses
[1] Computer-Aided Drug Design, In Silico Discovery, Therapeutics Discovery, Johnson & Johnson Innovative Medicine, Turnhoutseweg 30, 2340 Beerse, Belgium
[2] Biomedical Research Foundation, Academy of Athens, 4 Soranou Ephesiou, Athens 11527, Greece
[3] Computer-Aided Drug Design, In Silico Discovery, Therapeutics Discovery, Johnson & Johnson Innovative Medicine, 3210 Merryfield Row, San Diego, CA 92121, United States
[4] School of Pharmaceutical Sciences, University of Geneva, Rue Michel Servet 1, Geneva, 1206, Switzerland

Corresponding authors: Gervasio, Francesco L. (francesco.gervasio@unige.ch); Pande, Vineet (vpande@its.jnj.com)

Current Opinion in Structural Biology 2025, **90**:102975

This review comes from a themed issue on **New Concepts in Drug Discovery (2025)**

Edited by **Philip Biggin** and **Roderick Hubbard**

For a complete overview see the Issue and the Editorial

Available online xxx

https://doi.org/10.1016/j.sbi.2024.102975

Keywords:
cryptic pockets, allostery, mixed-solvent molecular dynamics, enhanced sampling, artificial intelligence.

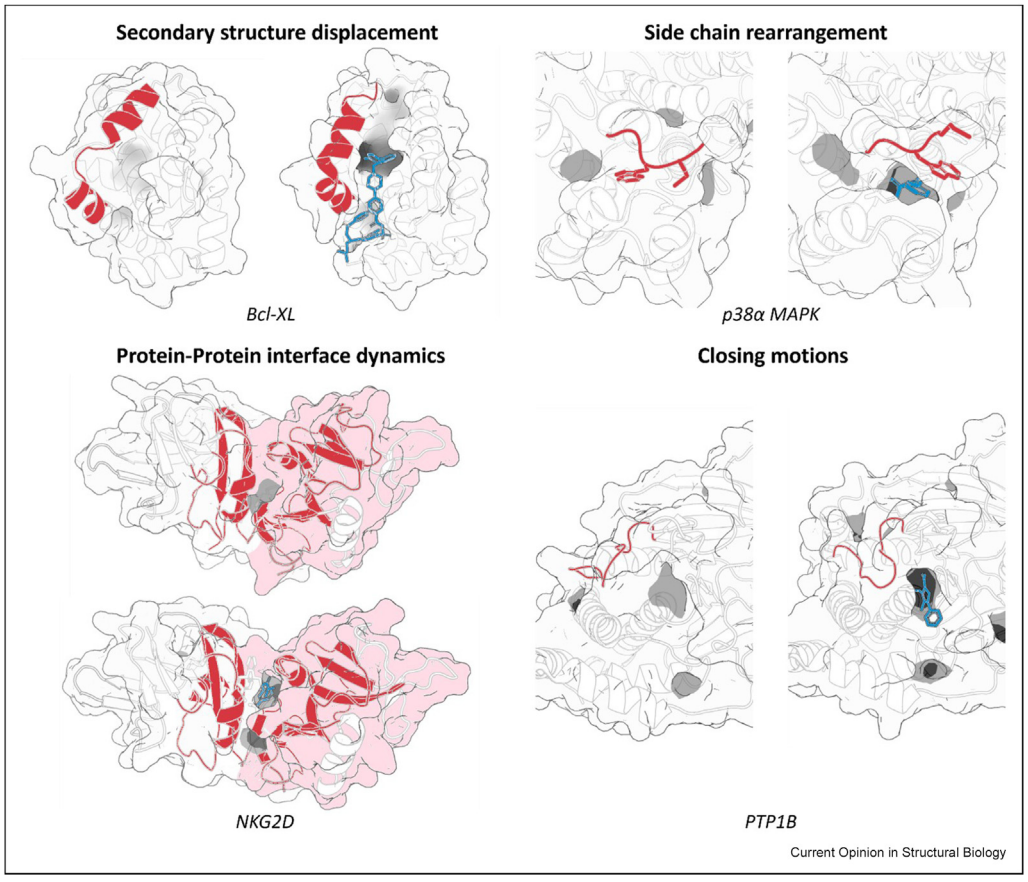
Introduction

Drug discovery efforts traditionally focus on finding molecules aimed at endogenous ligand binding sites of target proteins. This has proven to be difficult for numerous targets, including transcription factors, phosphatases, and GTPases, due to a respective lack of interfacial ligandable pockets, presence of highly conserved and/or poorly ligandable orthosteric sites, and picomolar-affinity endogenous binders [1]. In search of alternative approaches to modulate such difficult targets, the frameworks of allostery and crypticity have recently garnered considerable attention. Allostery can be defined as a modulation of the orthosteric site by a molecule, membrane, or posttranslational modification acting at a distal site that impacts protein function. This has advantages over targeting the orthosteric site, as reviewed elsewhere [2]. But, to realize its potential, it is necessary to first establish the presence of a ligandable secondary site.

As many proteins lack well-defined allosteric sites, cryptic pockets constitute a novel but challenging advance in drug discovery. While their mechanism of formation is still an active area of research and often case-dependent, they appear to emerge in transient excited states by protein structural fluctuations and conformational changes in solution as well as upon ligand binding [3]. After induced-fit, conformational selection, or a combination thereof, the pockets accessible in these minor protein states must be promptly stabilized with a molecule to prevent closing [4]. Consequently, these short-lived pockets are difficult to detect both experimentally and computationally and are aptly labeled "cryptic" [5]. While there is no consensus yet on what constitutes "cryptic" in this regard; we define a cryptic pocket as one that is well-formed with favorable ligand binding properties [6] not observed in the unbound structure(s) available at the start of a drug discovery project. Figure 1 shows examples of cryptic pockets in pharmaceutically relevant targets, highlighting the variety of protein motions that can underlie their formation.

Binding cryptic pockets with allosteric modulators has already displayed pharmaceutical value by enabling tractability of targets that were once considered undruggable, such as K-Ras [7], p53 [8], and Ebola VP35 [9], to name a few. Nevertheless, most cryptic

Figure 1



Cryptic pocket formation in pharmaceutically relevant targets. Key protein elements involved are highlighted in red, and ligands binding the cryptic pockets are colored blue. In Bcl-XL an alpha helix displaces to turn the flat surface present in the *apo* state (PDB ID 1R2D) into a well-formed ligandable pocket in the *holo* state (PDB ID 2YXJ). For p38α, the pocket entrance and exit are already accessible in *apo* form (PDB ID 1WBS), but a tryptophan and methionine residue must move out to make the pocket ligandable (PDB ID 3HVC). NKG2D (PDB ID 1R2D vs 8EA6) exemplifies that cryptic pockets can emerge through dynamics at interdomain or, in this case, protein–protein interface sites (using pink and white to differentiate monomers). In PTP1B, a nearby loop closes in on a preexisting cavity in the *apo* state (PDB ID 2CM2) to form a well-defined ligandable pocket (PDB ID 2H4K). This highlights that cryptic pockets can also emerge by "closing motions" rather than protein elements moving out of the way, as in Bcl-XL and p38α.

pockets so far have been discovered serendipitously while screening compound libraries, where X-ray crystallization revealed a hit compound to bind a pocket that was absent in the unliganded structure, such as recently in IL-1β [10]. Computational approaches may have the ability to investigate cryptic pockets, as they can help prioritize targets for experimental or virtual screening (VS) and provide new information related to cryptic pocket formation [6,11]. Indeed, a recent VS against a cryptic pocket in HSP70 found several micromolar compounds with varying scaffolds [12]. In the following sections we review state-of-the-art *in silico* tools to investigate cryptic pockets (Table 1, Figure 2) through case studies in pharmaceutically relevant proteins, and we conclude on emerging approaches impacting the field.

Mixed-solvent molecular dynamics based methods

Mixed-solvent molecular dynamics (MSMD) is one of the most used computational tools by investigators to assist in uncovering cryptic pockets, with recent applications in integrins [44] and the SARS-CoV2 Spike glycoprotein [45]. Here, cryptic pockets were identified as preferential binding areas of cosolvents. Owing to its many years in use, MSMD has gone through extensive (in)validation. Typical known challenges include i) phase separation between the cosolvent and water, ii) lingering of cosolvents at superficial sites, iii) protein unfolding, and iv) correlation between cosolvent polarity and hydrophilicity of mapped pockets. When these limitations are anticipated, MSMD can help to uncover cryptic pockets, as illustrated by Martinez-Rosell et al. [14]. By using low benzene concentrations ($\leq$0.2M)

Table 1

Overview of computational methods to investigate cryptic pockets and their known challenges, as discussed in the article.

State of the art	Known Challenges	Typical solutions	Additional remarks
Mixed solvent molecular dynamics	Phase separation between probe and solvent	Test and change probe force–field parameters to concur with experimental densities [13]	Requires extensive validation [13]
		Use water-miscible probes [13]	The polarity of the probe is highly correlated with the polarity of the pockets it finds [14,15]
		Use lower probe concentrations [14]	May require longer simulation times, which is resource-intensive and risks protein unfolding [14]
		Use repulsive forces between probes to prevent aggregation [4,14]	Can prevent sampling of larger pockets that can fit multiple probes [16]
		Use coarse-grained MD such that it is sufficient to have only a single probe that can be followed in its trajectory to find the cryptic binding pocket [17]	Flexibility constraints limit the extent of conformational changes that can be sampled [18]
	Spurious site mapping	Only analyze probe occupancies at the end of the simulation to focus on pockets that open later [16]	It is unknown beforehand at what timescale buried pockets become accessible
		Use hydrophobic probes to discourage binding on superficial sites [16]	Risks probe aggregation and protein unfolding may only sample hydrophobic pockets [14,15]
	Protein unfolding	Apply protein positional restraints to keep structural elements close together [4]	Can hamper sampling of buried pockets that require large backbone motions [16]
		Use lower probe concentrations to reduce protein opening motions by probes [16]	May require longer simulation times, which is resource-intensive and risks protein unfolding [14]
		Use many short simulations to subject the protein only shortly to pushing forces [16]	May not allow large conformational shifts necessary for cryptic pocket formation
	Pocket-probe polarity dependence	Use a range of probes covering aromatic, aliphatic, hydrogen bond donating/accepting, and charged moieties to target pockets of varying chemistry [14,15,19]	Each probe risks phase separation, aggregation, and bad force–field parameters [13]
	Pocket-starting structure dependence	Use a range of (AlphaFold generated) starting structures, regardless of quality or binding state [20]	Sometimes only a single structure is available and AlphaFold may fail to generate sufficiently diverse protein conformations to yield a desired impact
	Solvent occluded pockets	Combine mixed-solvent MD with accelerated MD to promote large conformational changes [16,19]	Can still require long simulation times, and the added energy in accelerated MD may force probes to exit cryptic pockets more quickly [19]
		Combine mixed-solvent MD with scaled protein-water interactions (SWISH) to promote large conformational changes [4]	The high probe concentration and scaling factor require flexibility restraints to prevent unfolding [4]
		Combine mixed-solvent MD with Monte Carlo (SILCS) to insert/delete probes at grid points in the protein that may not be easily solvent accessible [21]	The dynamics observed in SILCS typically involve side chain rearrangements [21]
	Application in GPCRs	Only place probes in the water phase [22]	Need to sample longer to allow transmembrane diffusion, making protein unfolding more likely [22]
		Apply a barrier to restrict probes to a certain area [23]	Does not study pocket opening in the full protein
CV-based sampling	Finding instructive CVs	Perform dimensionality reduction methods like principal component analysis on unbiased MD to obtain the motions describing maximal variance [24]	Works best when cryptic pocket formation is seen in unbiased MD, which eliminates the need for a CV(24)
		Markov state modeling to find slow motions from unbiased MD trajectories [25,26]	Highly resource-intensive, generally requiring hundreds of microseconds sampling time [26]
		Normal mode analysis to find slow motions in a protein structure [27]	It is difficult to select modes involved in functional dynamics, and the method generally still struggles with local dynamics and large conformational shifts [27]
		tICA on unbiased MD to reduce many dimensions to a few describing maximal autocorrelations [28]	Does not always evade the need for foreknowledge on the cryptic pocket location [28]
	Using suboptimal CVs	Use a static bias to enhance convergence in the sampled free energy profile [29]	The protein may get trapped in a local minimum, hindering sampling of the state of interest [29]
		Use an adaptive bias to resolve kinetic trapping [29]	May sample so many high-energy states that the state of interest may be difficult to identify
		On-the-fly probability enhanced sampling (adaptive bias that quickly becomes static) [30]	Still requires an instructive CV to evade the issues associated with static or adaptive biases [29]

(*continued on next page*)

Table 1 (*continued*)

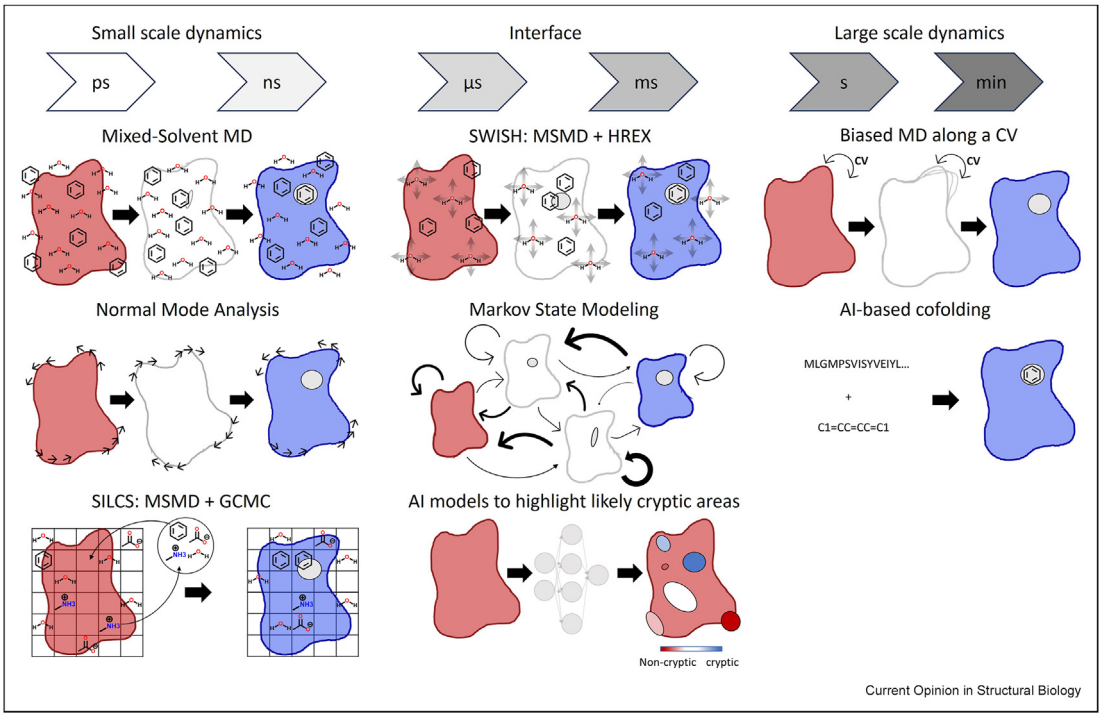
State of the art	Known Challenges	Typical solutions	Additional remarks
		OneOPES (REX + OPES explore) to iterate between adaptive and static biases [31]	Applicability on large systems for cryptic pocket discovery is not widely substantiated
		AF2-RAVE (iteratively build the CV) [32]	Applicability on large systems for cryptic pocket discovery is not established
	Identifying actionable cryptic pockets	Only consider pockets that have a good ligandability score [6,33,34]	Ligandability scoring functions are case sensitive. Also, depending on the MD snapshot used, a pocket may score highly or poorly ligandable [33]
		Establish crosstalk between the putative and orthosteric pockets or use mechanistic knowledge of the target to select pockets that are likely to incite a functional effect upon ligand binding [35,36]	The prior knowledge required for these approaches may not be available
		Focus on pockets that are backed up by supporting data, such as aligning well with known pockets in homologous targets or those composed of residues implicated in coevolution [33,37]	Such data may not be available for the target of interest
AI-based pocket detection	Finding areas likely to be involved in crypticity	Algorithms like PocketMiner can score the likelihood of any residue to be involved in cryptic pocket formation [38]	Still requires sampling with MD on the area identified by the scoring algorithm [38]
	Find instructive data-driven CVs for biased MD	Use cryptic pocket open and closed states to define a progress coordinate underlying cryptic pocket formation [39]	Such data may not be available for the target of interest
	Protein clashes when placing ligands in a rigid protein model	Include flexibility in sidechains after ligand placement [40]	Typically focuses on sidechain rotations and retains an *apo*-like backbone, which can be detrimental for cryptic pocket discovery [41]
		Fold the protein and ligand together in co-folding [42,43]	Has not been widely validated for cryptic pockets

instead of repulsion potentials, prioritizing long-residence-time binding areas, and simulating at short timescales (<50 ns), they could bypass the first three issues and retrieve cryptic pockets for most targets. A limitation that was later addressed by Szabó et al. [15] and Clark et al. [20] was the use of a single cosolvent and starting structure, which may explain why some pockets were missed. Szabó et al. [15] showed in SARS-CoV2 RdRp that different probes can unlock exclusive protein states exhibiting different pockets, and Clark et al. [20] found that the success of retrieving validated pockets depends on the starting structure, irrespective of *apo*/*holo* state or quality (up to 2.5 Å resolution). While the choice and concentration of cosolvent thus dictates the conformational ensemble explored; they only serve as "computational probes" to find novel pocket space and are not intended to be used as ligands themselves.

Advances in MSMD have enabled value for **G-protein-coupled receptors** (GPCRs) and buried pockets. Using MSMD in GPCRs requires extra attention to prevent clashes with membrane lipids and allow transmembrane diffusion. For example, by placing cosolvents in the water phase and simulating longer [22] or using a barrier to confine probes to a select area [23]. Chan et al. [22] recently used the former method to retrieve cryptic pockets in C–C chemokine receptor 2 and cannabinoid receptor type 1, and suggest novel pockets in the M_2 muscarinic and μ-opioid receptors. Notably, the MacKerell Jr. group was able to apply their probe-based method, "Site Identification by Ligand Competitive Saturation" (SILCS), to GPCRs before by iterating MD sampling and probe insertion/deletion with Monte Carlo [21]. While they retrieved (partially) occluded pockets in three GPCRs, the biggest structural change involved sidechain reorientations in the androgen receptor.

When large conformational shifts underlie cosolvent binding, MSMD can be coupled to accelerated MD, which augments protein dynamics with a bias. For instance, Tan's accelerated ligand-mapping MD (16), using 0.2 M benzene for 200 ns, retrieved buried cryptic pockets in all eight test cases, including Bcl-xL and PTP1B. Carlson's aMD + MixMD (19), using 5% v/v organic cosolvents for 100 ns, conversely missed the pocket in four of twelve cases. Namely, cMet, PTP1B, AANAT, and TEM-1 β-lactamase, which all require helix displacements. This illustrates that the success of MSMD is case-dependent when large structural shifts are involved and, as stated by Carlson et al. [19], suggest that probes may not fully drive cryptic pocket opening but rather opportunely bind dynamic opening cavities. Indeed, this aligns with early observations in "Sampling Water Interfaces through Scaled

Figure 2



Overview of different computational approaches to investigate cryptic pockets. Depicting the pocket-closed state in red, potential intermediate states in white, and the pocket-open state in blue. While there is generally no strict limit to the motions accessible to these methods; they are grouped based on the timescale at which the motions typically observed in these approaches naturally occur. That is, smaller-scale dynamics like sidechain rotations and small protein opening motions that occur on the picosecond to nanosecond timescales are generally observed in a standard protocol execution of mixed-solvent MD, SILCS, and NMA. SWISH, Markov state modeling, and AI-based crypticity detection methods generally identify potential binding areas or conformational shifts occurring in the microsecond to millisecond range, and biased MD along a CV or AI-based cofolding approaches can enable the slowest conformational changes such as those occurring in protein folding pathways on the second to minute timescale. SILCS, Site Identification by Ligand Competitive Saturation; NMA, normal mode analysis; SWISH, Sampling Water Interfaces through Scaled Hamiltonians; CV, collective variable.

Hamiltonians" (SWISH) [4], a method that combines MSMD with scaled protein-water interactions. Where benzene probes without SWISH typically failed to bind the cryptic pocket in TEM-1 β-lactamase even after 1 µs, the application of SWISH exposed a cavity that these molecules could enter to fully shape the pocket. In this light, standard MSMD appears suited when small conformational changes are expected, but alternative tools may be desirable to cover large-scale dynamics.

Collective variable based methods

To cover more extensive dynamics, biased sampling methods can be used, which typically conduce rare events by pushing the protein along a collective variable (CV). This is a parameter describing one or more degrees of freedom, like a distance or angle between atoms. Given that CVs force exploration of a conformational subspace, they ideally describe the key motions underlying cryptic pocket formation. But biological systems have many degrees of freedom, most of which do not individually drive state transitions. As such, dimensionality reduction techniques like principal component analysis (PCA), are routinely used to define CVs. Benabderrahmane et al. recently applied PCA on 100 ns unbiased MD of Mcl-1 to obtain CVs that drove cryptic pocket opening in well-tempered metadynamics [24]. While this shows value for PCA, a pocket-opening event notably occurred in the unbiased simulation. This is a golden scenario for CV design, as it captures the dynamics of pocket formation, but it is rare to occur within 100 ns. This turns CV design into a circular problem at short timescales, where defining good CVs requires observing state transitions in MD while observing state transitions in MD requires sampling along good CVs. Alternative tools for CV design include normal mode analysis (NMA), Markov state modeling (MSM), time-lagged independent component analysis (tICA), and Just Exploring Druggability at protein Interfaces (JEDI) [46]. Sampling informed by these tools, respectively, showed value in cryptic pocket detection for exportin 1 (27), methionine aminopeptidase II (25), NPC2 (28),

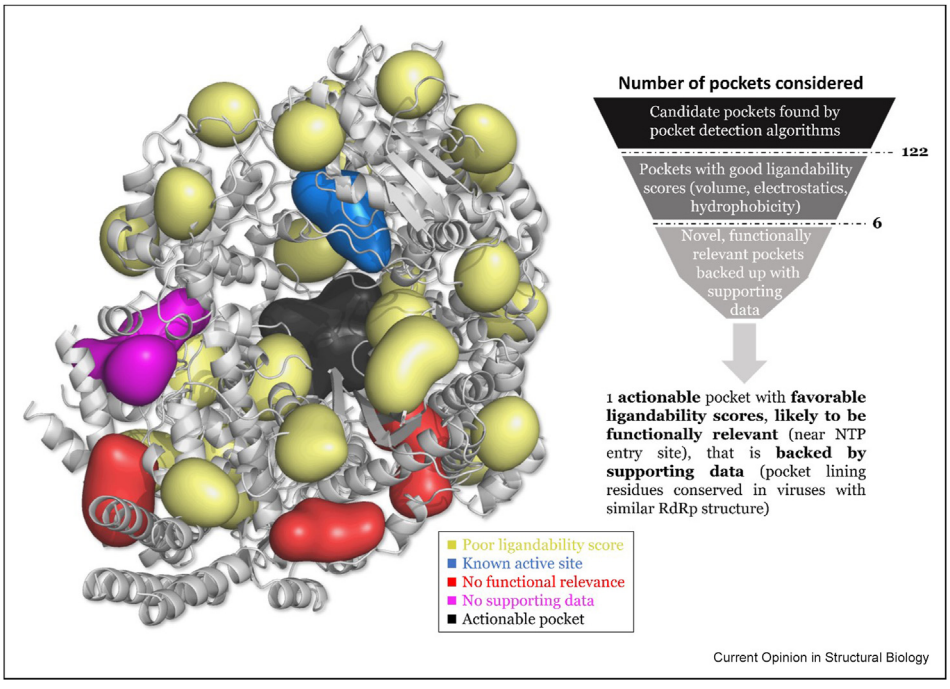
and VHL (46). Still, there are some limitations since NMA struggles with large structural shifts and local dynamics, MSM typically needs hundreds of microseconds of sampling time, tICA at times requires foreknowledge of the cryptic pocket location, and JEDI is an expensive CV [25,27,28,46].

The best achievable CV is still often suboptimal [47] given the challenge in including all relevant motions. Importantly, cryptic pocket formation will remain rare when missing even a single involved motion, which can include water dynamics [48]. While adding more CVs could improve the method; using more than a handful greatly reduces simulation speed. So, one must typically choose between convergence and exploration for complex targets [29]. On one hand, a static bias can be applied on the suboptimal CV to favor convergence but risk kinetic trapping due to the unaccelerated dynamics. Alternatively, a rapidly changing adaptive bias can push the system out of such traps but disallows convergence until becoming (quasi)static. To allow both exploration and fast convergence for suboptimal CVs, Invernizzi and Parinello introduced On-the-fly Probability Enhanced Sampling (OPES) [30]. But CV quality remained critical, leading to the development of an exploration-focused variant [29]. To further reduce dependence on CV quality, the latest OPES variant, termed "oneOPES"(31), combines OPESexplore [29] with replica exchange and multithermal sampling. By raising the temperature along the replica ladder, convergence is favored in bottom replicas while conducing exploration in top replicas even along coordinates ignored by the CV [49*]. These methods are invaluable in efficient sampling of slow motions, such as protein folding pathways, along which cryptic pockets can be uncovered, as illustrated for the dual-specificity tyrosine-phosphorylation-regulated kinase 1A [50].

Finding actionable pockets

Although biased MD can efficiently sample large conformational shifts, it does not implicitly accentuate pockets and thus requires pocket detection algorithms. These often highlight many cavities without appropriate ranking, hindering the selection of actionable pockets from false positives. While MSMD can likewise return many binding areas, its probe data can guide pocket selection by focusing on those with high probe occupancy and residence time [14]. Absent or supporting

Figure 3



Selecting actionable pockets from a pool of candidates. An illustrative example of selecting an actionable pocket from a host of candidates found by pocket detection algorithms, based on a case described by Oraby et al. [34]. Pocket detection suggested 122 candidate pockets in the *apo* RSV L–P complex (PDB 6PZK). Although inspired by the original case, this figure depicts fewer and mostly different pockets (defined by MOE SiteFinder) for illustrative purposes. Selecting only pockets found in consensus by three different pocket detection algorithms and displaying favorable ligandability scores narrowed the pool down to six candidates. Of these, the known active site was removed, leaving five. Finally, the selection was made for the one pocket that was close to the nucleoside triphosphate (NTP) entry site, making it likely that liganding this pocket could lead to a functional effect in the protein, and had its composing residues conserved in a sequence alignment of homologous viruses.

such data, actionable pockets can be identified by three traits: ligandability, functional relevance, and (co)evolution (Figure 3). First, a candidate pocket should have favorable ligand binding properties [6]. This is typically gauged by scoring functions, such as DScore and SiteScore of SiteMap [51], which assess features including volume, shape, and hydrophobicity. Secondly, liganding the pocket should incite a functional effect, though it should be noted that nonfunctional sites can still be useful for molecular glues and PROTACs [52]. This can be inferred from mechanistic knowledge of the target, as illustrated by Balasubramaniam et al. [35], who found a functional ligand by pursuing the only candidate pocket in ApoE4 located near the core of the DNA-binding region. Functional impact can also be verified by establishing a connection between the candidate pocket and orthosteric site. This is exemplified by Ni et al. [36], who found that the interaction free energy of residue pairs in the actionable candidate pocket changed upon NAD + binding in the orthosteric site. The Colombo group similarly showed that monitoring internal protein energetics with and without the orthosteric ligand present can help to define residue groups that are most responsive to ligand binding as (cryptic) allosteric hotspots [53]. Finally, pocket selection can be supported by (co)evolution and similarity data. La Sala et al. [33] showed that functional allosteric pockets typically contain coevolving residues, and Yu et al. [37] discovered a functional Wntless ligand by pursuing the pocket that aligns with known GPCR drug binding sites. Notably, integrating these traits gives optimal results [33,34], and if sufficient computational power is accessible, additional value can be drawn from the generation of quantitative conformational ensembles. For instance, the 0.1s sampling of SARS-CoV-2 by Folding@home enabled a quantitative probability measure of epitope exposure to guide antibody design [54]. While these approaches can aid in pocket selection, experimental validation is a necessary final step. This is typically achieved by binding and/or functional assays with a molecule identified through (virtual) screening against the hypothesized pocket [33,35,37], crystallizing a screening hit [12], or site-directed mutagenesis to validate the effect of perturbing the new pocket [36].

Artificial intelligence based approaches

Cryptic pocket discovery has recently been enriched with AI-based methods [55], either to support the physics-based approaches described above or as standalone solutions. An example of the former is PocketMiner [38], which scores the likelihood of protein residues to be involved in cryptic pocket formation to prioritize targets for further analysis in MD. Another strategy is the use of AlphaFold [56] (AF) to generate an ensemble of target structures that may exhibit partially open cryptic pockets, which can be fully formed in subsequent MD and MSM [57]. AF2-RAVE (32) from the Tiwary lab similarly uses AF-generated protein states but then iterates MD and autoencoder-based analysis to gradually build a CV for enhanced sampling [58]. The use of data-driven CVs in biased MD is emerging [39], though the need for structures in both open and closed states complicates this methodology when only one state is available. In this context, methods capable of generating protein ensembles covering such open and closed conformations can be used, like SPEECH_AF [59]) and AF-cluster [60].

Standalone AI solutions typically focus on the placement of ligands anywhere in a (modeled) protein structure. For instance, CobDock [61], which performs such blind docking, AlphaFill [62], which transplants a ligand from an experimental structure into a protein model, and DynamicBind [40], which places input ligands into a (modeled) protein followed by translation and rotation to include dynamics. Notably, the latter enables large conformational shifts and successfully retrieved a cryptic pocket in the histone methyltransferase SET2D. Still, AI-based blind docking is less generalizable across the proteome than physics-based docking [63], transplantation can lead to protein-ligand clashes, and methods including flexibility often retain an *apo*-like backbone, hampering performance for cryptic pocket discovery [41]. Some generative methods go one step further and take only one-dimensional protein (sequence) and compound (SMILES) information to fold them together, *i.e.* cofolding. These methods, which have the potential advantage of placing ligands in cryptic pockets and directly rank order them, include AlphaFold3 (42) and NeuralPlexer [43]. While the solutions promised by these methods are considerable, their value in cryptic pocket discovery remains to be widely validated.

Emerging approaches and conclusions

While most AI models only recently emerged, their potential in cryptic pocket discovery is evident, such as is the case for coarse-graining (CG) and CG-informed CVs [18]. Reducing protein resolution enables simulation on timescales at which cryptic pockets can emerge, as shown by Diamanti et al. [17], who used CGMD to find a partially hidden site in energy-coupling factor transporter FolT2. Coarse-graining and back-mapping to an all-atom description also enables cryptic pocket modeling in large protein complexes [64]. While parametrization and flexibility constraints are restrictive, developments like the Martini-Gō model and integration of CGMD with MSMD [65] have poised the method for impact in cryptic pocket discovery. Although monomeric soluble or transmembrane proteins have been the main focus, applications in RNA [66] and protein−protein interfaces [49] are establishing cryptic pockets as a universal target to exploit in therapeutics discovery. The use cases described throughout this review, give the

impression that MSMD is most suited to study sidechain rearrangements and local dynamics, while biased MD can help investigate larger dynamics. AI models have the unique position to span the entire field, with clear supportive and standalone roles and applicability in both local and global dynamics. Altogether, the combination of orthogonal AI- and physics-based approaches reviewed here combined with the knowledge of target biochemistry and other available experimental information at hand is set to impact this field in providing novel hypotheses and new chemical starting points for challenging yet high-potential drug targets.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank Simone Marsili, Alberto Borsatto and Herman van Vlijmen for fruitful discussions. M.B. would like to acknowledge funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 956314. F.G. acknowledges funding from the Swiss National Science Foundation under project numbers 200021_204795 and CRSII5_216587, and Bridge project number 40B2-0_203628.

Data availability

No data were used for the research described in the article.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. * Xie X, Yu T, Li X, Zhang N, Foster LJ, Peng C, *et al.*: **Recent advances in targeting the "undruggable" proteins: from drug discovery to clinical trials**. *Signal Transduct Targeted Ther* 2023, **8**:335.
A review discussing "undruggable" targets and how some challenges have been surmounted to bring therapeutics options to the clinic.

2. Chatzigoulas A, Cournia Z: **Rational design of allosteric modulators: challenges and successes**. *WIREs Computational Molecular Science* 2021, **11**, e1529.

3. Kuzmanic A, Bowman GR, Juarez-Jimenez J, Michel J, Gervasio FL: **Investigating cryptic binding sites by molecular dynamics simulations**. *Acc Chem Res* 2020, **53**:654–661.

4. Oleinikovas V, Saladino G, Cossins BP, Gervasio FL: **Understanding cryptic pocket formation in protein targets by enhanced sampling simulations**. *J Am Chem Soc* 2016, **138**:14257–14263.

5. Ni D, Chai Z, Wang Y, Li M, Yu Z, Liu Y, *et al.*: **Along the allostery stream: recent advances in computational methods for allosteric drug discovery**. *WIREs Computational Molecular Science* 2022, **12**, e1585.

6. Brown KK, Hann MM, Lakdawala AS, Santos R, Thomas PJ, Todd K: **Approaches to target tractability assessment – a practical perspective**. *MedChemComm* 2018, **9**:606–613.

7. Lanman BA, Allen JR, Allen JG, Amegadzie AK, Ashton KS, Booker SK, *et al.*: **Discovery of a covalent inhibitor of KRASG12C (AMG 510) for the treatment of solid tumors**. *J Med Chem* 2020, **63**:52–65.

8. ** Fallatah MMJ, Law FV, Chow WA, Kaiser P: **Small-molecule correctors and stabilizers to target p53**. *Trends Pharmacol Sci* 2023, **44**:274–289.
Discussing cryptic pockets as a therapeutic avenue for challenging drug targets. In particular, the ability to target a cryptic pocket within a mutated p53 is discussed.

9. Cruz MA, Frederick TE, Mallimadugula UL, Singh S, Vithani N, Zimmerman MI, *et al.*: **A cryptic pocket in Ebola VP35 allosterically controls RNA binding**. *Nat Commun* 2022, **13**:2269.

10. Hommel U, Hurth K, Rondeau J-M, Vulpetti A, Ostermeier D, Boettcher A, *et al.*: **Discovery of a selective and biologically active low-molecular weight antagonist of human interleukin-1β**. *Nat Commun* 2023, **14**:5497.

11. Shan Y, Mysore VP, Leffler AE, Kim ET, Sagawa S, Shaw DE: **How does a small molecule bind at a cryptic binding site?** *PLoS Comput Biol* 2022, **18**, e1009817.

12. O'Connor S, Le Bihan Y-V, Westwood IM, Liu M, Mak OW, Zazeri G, *et al.*: **Discovery and characterization of a cryptic secondary binding site in the molecular chaperone HSP70**. *Molecules [Internet]* 2022, **27**.

13. Lexa KW, Goh GB, Carlson HA: **Parameter choice matters: validating probe parameters for use in mixed-solvent simulations**. *J Chem Inf Model* 2014, **54**:2190–2199.

14. Martinez-Rosell G, Lovera S, Sands ZA, De Fabritiis G: **PlayMolecule CrypticScout: predicting protein cryptic sites using mixed-solvent molecular simulations**. *J Chem Inf Model* 2020, **60**:2314–2324.

15. Szabó PB, Sabanés Zariquiey F, Nogueira JJ: **Cosolvent and dynamic effects in binding pocket search by docking simulations**. *J Chem Inf Model* 2021, **61**:5508–5523.

16. Tze-Yang Ng J, Tan YS: **Accelerated ligand-mapping molecular dynamics simulations for the detection of recalcitrant cryptic pockets and occluded binding sites**. *J Chem Theor Comput* 2022, **18**:1969–1981.

17. Diamanti E, Souza PCT, Setyawati I, Bousis S, Monjas L, Swier LJYM, *et al.*: **Identification of inhibitors targeting the energy-coupling factor (ECF) transporters**. *Commun Biol* 2023, **6**:1182.

18. Kjølbye LR, Pereira GP, Bartocci A, Pannuzzo M, Albani S, Marchetto A, *et al.*: **Towards design of drugs and delivery systems with the Martini coarse-grained model**. *QRB Discovery* 2022, **3**:e19.

19. Smith RD, Carlson HA: **Identification of cryptic binding sites using MixMD with standard and accelerated molecular dynamics**. *J Chem Inf Model* 2021, **61**:1287–1299.

20. Clark JJ, Orban ZJ, Carlson HA: **Predicting binding sites from unbound versus bound protein structures**. *Sci Rep* 2020, **10**, 15856.

21. Lakkaraju SK, Yu W, Raman EP, Hershfeld AV, Fang L, Deshpande DA, *et al.*: **Mapping functional group free energy patterns at protein occluded sites: nuclear receptors and G-protein coupled receptors**. *J Chem Inf Model* 2015, **55**:700–708.

22. * Chan WKB, Carlson HA, Traynor JR: **Application of mixed-solvent molecular dynamics simulations for prediction of allosteric sites on G protein–coupled receptors**. *Mol Pharmacol* 2023, **103**:274.

An exemplary application of mixed-solvent molecular dynamics on G protein-coupled receptors to find cryptic pockets.

23. Ciancetta A, Gill AK, Ding T, Karlov DS, Chalhoub G, McCormick PJ, *et al.*: **Probe confined dynamic mapping for G protein-coupled receptor allosteric site prediction**. *ACS Cent Sci* 2021, **7**:1847–1862.

24. Benabderrahmane M, Bureau R, Voisin-Chiret AS, Santos JS-dO: **Cryptic pockets repository through pocket dynamics tracking and metadynamics on essential dynamics space: applications to mcl-1**. *J Chem Inf Model* 2021, **61**:5581–5588.

25. Rubina Moin ST, Haider S: **Identification of a cryptic pocket in methionine aminopeptidase-II using adaptive bandit molecular dynamics simulations and Markov state models**. *ACS Omega* 2024, **9**:28534–28545.

26. Ferruz N, Doerr S, Vanase-Frawley MA, Zou Y, Chen X, Marr ES, *et al.*: **Dopamine D3 receptor antagonist reveals a cryptic pocket in aminergic GPCRs**. *Sci Rep* 2018, **8**:897.

27. Zheng W: **Predicting cryptic ligand binding sites based on normal modes guided conformational sampling**. *Proteins: Struct, Funct, Bioinf* 2021, **89**:416–426.

28. Sabanés Zariquiey F, Jacoby E, Vos A, van Vlijmen HWT, Tresadern G, Harvey J: **Divide and conquer. Pocket-opening mixed-solvent simulations in the perspective of docking virtual screening applications for drug discovery**. *J Chem Inf Model* 2022, **62**:533–543.

29. Invernizzi M, Parrinello M: **Exploration vs convergence speed in adaptive-bias enhanced sampling**. *J Chem Theor Comput* 2022, **18**:3988–3996.

30. Invernizzi M, Parrinello M: **Rethinking metadynamics: from bias potentials to probability distributions**. *J Phys Chem Lett* 2020, **11**:2731–2736.

31. * Rizzi V, Aureli S, Ansari N, Gervasio FL: **OneOPES, a combined enhanced sampling method to rule them all**. *J Chem Theor Comput* 2023.

Here a new development in collective variable based simulation is introduced, which allows efficient sampling of large conformational changes with minimal dependence on CV choice.

32. Vani BP, Aranganathan A, Wang D, Tiwary P: **AlphaFold2-RAVE: from sequence to Boltzmann ranking**. *J Chem Theor Comput* 2023.

33. La Sala G, Pfleger C, Käck H, Wissler L, Nevin P, Böhm K, *et al.*: **Combining structural and coevolution information to unveil allosteric sites**. *Chem Sci* 2023, **14**:7057–7067.

34. * Oraby AK, Bilawchuk L, West FG, Marchant DJ: **Structure-based discovery of allosteric inhibitors targeting a new druggable site in the respiratory syncytial virus polymerase**. *ACS Omega* 2024, **9**:22213–22229.

Exemplifying identificaiton of actionable pockets from a large list of candidate cavities. This work aimed to find the binding site of a known binder for the RSV L protein, however describes general principles by which pockets are ranked and selected for compound design.

35. Balasubramaniam M, Narasimhappagari J, Liu L, Ganne A, Ayyadevara S, Atluri R, *et al.*: **Rescue of ApoE4-related lysosomal autophagic failure in Alzheimer's disease by targeted small molecules**. *Commun Biol* 2024, **7**:60.

36. Ni D, Wei J, He X, Rehman AU, Li X, Qiu Y, *et al.*: **Discovery of cryptic allosteric sites using reversed allosteric communication by a combined computational and experimental strategy**. *Chem Sci* 2021, **12**:464–476.

37. Yu J, Liao P-J, Keller TH, Cherian J, Virshup DM, Xu W: **Ultra-large scale virtual screening identifies a small molecule inhibitor of the Wnt transporter Wntless**. *iScience* 2024, **27**.

38. * Meller A, Ward M, Borowsky J, Kshirsagar M, Lotthammer JM, Oviedo F, *et al.*: **Predicting locations of cryptic pockets from single protein structures using the PocketMiner graph neural network**. *Nat Commun* 2023, **14**:1177.

An example of the value of AI to prioritize protein targets.

39. Bonati L, Rizzi V, Parrinello M: **Data-driven collective variables for enhanced sampling**. *J Phys Chem Lett* 2020, **11**:2998–3004.

40. Lu W, Zhang J, Huang W, Zhang Z, Jia X, Wang Z, *et al.*: **DynamicBind: predicting ligand-specific protein-ligand complex structure with a deep equivariant generative model**. *Nat Commun* 2024, **15**:1071.

41. Zhu J, Gu Z, Pei J, Lai L: **DiffBindFR: an SE(3) equivariant network for flexible protein–ligand docking**. *Chem Sci* 2024, **15**:7926–7942.

42. ** Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, *et al.*: **Accurate structure prediction of biomolecular interactions with AlphaFold 3**. *Nature* 2024, **630**:493–500.

The newest AlphaFold model, AF3, is introduced with the capability of cofolding a ligand together with the protein to uncover cryptic pockets.

43. Qiao Z, Nie W, Vahdat A, Miller TF, Anandkumar A: **State-specific protein–ligand complex structure prediction with a multiscale deep generative model**. *Nat Mach Intell* 2024, **6**:195–208.

44. * Ilie IM, Ehrhardt C, Caflisch A, Weitz-Schmidt G: **Decrypting integrins by mixed-solvent molecular dynamics simulations**. *J Chem Inf Model* 2023, **63**:3878–3891.

An example of how mixed-solvent molecular dynamics can be applied to detect novel cryptic pockets, in this case in integrins.

45. Zuzic L, Samsudin F, Shivgan AT, Raghuvamsi PV, Marzinek JK, Boags A, *et al.*: **Uncovering cryptic pockets in the SARS-CoV-2 spike glycoprotein**. *Structure* 2022, **30**. 1062-74.e4.

46. Cuchillo R, Pinto-Gil K, Michel J: **A collective variable for the rapid exploration of protein druggability**. *J Chem Theor Comput* 2015, **11**:1292–1307.

47. Invernizzi M, Parrinello M: **Making the best of a bad situation: a multiscale approach to free energy calculation**. *J Chem Theor Comput* 2019, **15**:2187–2194.

48. Ansari N, Rizzi V, Carloni P, Parrinello M: **Water-triggered, irreversible conformational change of SARS-CoV-2 main protease on passing from the solid state to aqueous solution**. *J Am Chem Soc* 2021, **143**:12930–12934.

49. * Borsatto A, Gianquinto E, Rizzi V, Gervasio FL: **SWISH-X, an expanded approach to detect cryptic pockets in proteins and at protein–protein interfaces**. *J Chem Theor Comput* 2024, **20**:3335–3348.

This work shows how cryptic pockets can be found at protein–protein interfaces.

50. Kii I, Sumida Y, Goto T, Sonamoto R, Okuno Y, Yoshida S, *et al.*: **Selective inhibition of the kinase DYRK1A by targeting its folding process**. *Nat Commun* 2016, **7**, 11391.

51. Halgren TA: **Identifying and characterizing binding sites and assessing druggability**. *J Chem Inf Model* 2009, **49**:377–389.

52. Koldenhof P, Bemelmans MP, Ghosh B, Damm-Ganamet KL, van Vlijmen HWT, Pande V: **Application of AlphaFold models in evaluating ligandable cysteines across E3 ligases**. *Proteins: Struct, Funct, Bioinf* 2024, **92**:819–829.

53. Colombo G: **Computing allostery: from the understanding of biomolecular regulation and the discovery of cryptic sites to molecular design**. *Curr Opin Struct Biol* 2023, **83**, 102702.

54. Zimmerman MI, Porter JR, Ward MD, Singh S, Vithani N, Meller A, *et al.*: **SARS-CoV-2 simulations go exascale to predict dramatic spike opening and cryptic pockets across the proteome**. *Nat Chem* 2021, **13**:651–659.

55. Nerín-Fonz F, Cournia Z: **Machine learning approaches in predicting allosteric sites**. *Curr Opin Struct Biol* 2024, **85**, 102774.

56. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, *et al.*: **Highly accurate protein structure prediction with AlphaFold**. *Nature* 2021, **596**:583–589.

57. Meller A, Bhakat S, Solieva S, Bowman GR: **Accelerating cryptic pocket discovery using AlphaFold**. *J Chem Theor Comput* 2023.

58. Mehdi S, Wang D, Pant S, Tiwary P: **Accelerating all-atom simulations and gaining mechanistic understanding of biophysical systems through state predictive information bottleneck**. *J Chem Theor Comput* 2022, **18**:3231–3238.

59. Stein RA, McHaourab HS: **SPEECH_AF: sampling protein ensembles and conformational heterogeneity with Alphafold2**. *PLoS Comput Biol* 2022, **18**, e1010483.

60. Wayment-Steele HK, Ojoawo A, Otten R, Apitz JM, Pitsawong W, Hömberger M, *et al.*: **Predicting multiple conformations via sequence clustering and AlphaFold2**. *Nature* 2024, **625**: 832–839.

61. Ugurlu SY, McDonald D, Lei H, Jones AM, Li S, Tong HY, *et al.*: **Cobdock: an accurate and practical machine learning-based consensus blind docking method**. *J Cheminf* 2024, **16**:5.

62. Hekkelman ML, de Vries I, Joosten RP, Perrakis A: **AlphaFill: enriching AlphaFold models with ligands and cofactors**. *Nat Methods* 2023, **20**:205–213.

63. Buttenschoen M, Morris GM, Deane CM: **PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences**. *Chem Sci* 2024, **15**: 3130–3139.

64. Alshahrani M, Gupta G, Xiao S, Tao P, Verkhivker G: **Comparative analysis of conformational dynamics and systematic characterization of cryptic pockets in the SARS-CoV-2 omicron BA.2, BA.2.75 and XBB.1 spike complexes with the ACE2 host receptor: confluence of binding and structural plasticity in mediating networks of conserved allosteric sites**. *Viruses* 2023, **15** [Internet].

65. Andreev G, Kovalenko M, Bozdaganyan ME, Orekhov PS: **Colabind: a cloud-based approach for prediction of binding sites using coarse-grained simulations with molecular probes**. *J Phys Chem B* 2024, **128**: 3211–3219.

66. ** Panei FP, Gkeka P, Bonomi M: **Identifying small-molecules binding sites in RNA conformational ensembles with SHAMAN**. *Nat Commun* 2024, **15**:5725.

This work showcases a method that can be used to find cryptic pockets in RNA.