# Investigating Cryptic Binding Sites by Molecular Dynamics Simulations

*Antonija Kuzmanic,[1] Gregory R. Bowman,[2] Jordi Juarez-Jimenez,[3] Julien Michel,[3] and*

*Francesco L. Gervasio[1,4,]\**

[1]Department of Chemistry and Institute of Structural and Molecular Biology, University College London, London, United Kingdom

[2]Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St Louis, Missouri

[3]EaStCHEM School of Chemistry, University of Edinburgh, Edinburgh, United Kingdom

[4]Pharmaceutical Sciences, University of Geneva, Geneva, Switzerland.

CONSPECTUS: This Account highlights recent advances and discusses major challenges in investigations of cryptic (hidden) binding sites by molecular simulations. Cryptic binding sites are not visible in protein targets crystallized without a ligand and only become visible crystallographically upon binding events. These sites have been shown to be druggable and might provide a rare opportunity to target difficult proteins. However, due to their hidden nature, they are difficult to find through experimental screening. Computational methods based on atomistic molecular simulations remain one of the best approaches to identify and characterize cryptic binding sites. However, not all methods are equally efficient. Some are more apt at quickly probing protein dynamics, but do not provide thermodynamic or druggability

1

information, while others that are able to provide such data are demanding in terms of time and resources. Here, we review the recent contributions of mixed-solvent simulations, metadynamics, Markov state models, and other enhanced sampling methods to the field of cryptic site identification and characterization. We discuss how these methods were able to provide precious information on the nature of the site opening mechanisms, to predict previously unknown sites which were used to design new ligands, and to compute the free energy landscapes and kinetics associated with the opening of the sites and the binding of the ligands. We highlight the potential and the importance of such predictions in drug discovery, especially for difficult ("undruggable") targets. We also discuss the major challenges in the field and their possible solutions.
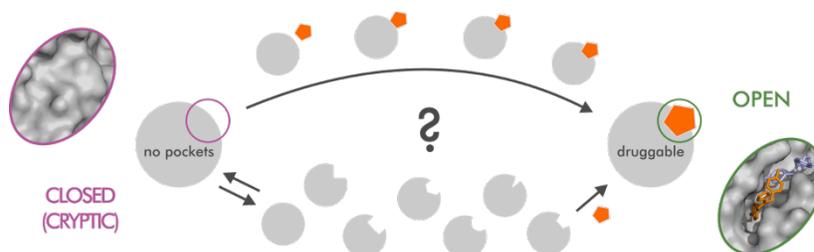


## INTRODUCTION

In recent years, the number of biologically validated drug targets for complex diseases has considerably increased, especially due to large-scale genome sequencing.[1,2] Unfortunately, many of these potential targets are considered "undruggable", meaning they are unamenable to classic substrate-competitive drug discovery strategies.[1] For this reason, significant efforts are put towards alternative strategies, including identifying allosteric ligands and characterizing hidden (cryptic) allosteric pockets. Cryptic binding pockets are absent in unliganded protein structures but open due to protein dynamics. They represent an attractive alternative to substrate-competitive sites that have been exploited in a number of high-profile targets, including K-Ras,[3,4]

an oncogene commonly found in human cancers. Despite their vast potential, the hidden nature of cryptic pockets makes it difficult to use rational drug discovery approaches based on structural experiments or computer modeling. Indeed, most known cryptic sites have been found serendipitously. Even the very mechanism of the cavity opening and whether it occurs through an induced fit, a conformational selection, or perhaps a combination of the two, is still debated in the literature. However, atomistic molecular dynamics (MD) simulations provide ways to both model the opening of unknown cryptic sites and describe possible binding mechanisms.

After many years of development, MD simulations are finally becoming useful in obtaining a detailed molecular description of protein (de)activation and ligand (un)binding. With the technological advancements in specialized computer architectures[5] for MD simulations and distributed computing,[6] as well as the parallelization of the MD simulation packages[7] and their increasing use of GPUs,[8] scientists are now able to reach unprecedented system sizes[9] and timescales,[10] enabling μs-long simulations on a routine basis. At the same time, both protein[11-14] and ligand[15] force fields are sufficiently accurate to capture the important features of target dynamics and ligand binding mechanisms. Prediction of previously unknown cryptic binding pockets that have successively been experimentally validated to design new drugs shows how promising the approach is, as was the case of HIV integrase where MD simulations predicted the opening of a new cleft.[16] Still, despite the substantial progress, conventional approaches in MD simulations cannot adequately sample many biologically and pharmaceutically interesting processes. Among such processes is the opening of buried cryptic binding sites which is a prerequisite for their detection. Recent studies demonstrate how this challenge can be successfully tackled using MD-based approaches, which offer a more dynamic picture of the

system (compared to X-ray crystallography) and provide high-resolution structural data for drug design.

In this Account, we review the recent contributions of MD simulations and modelling combined with experiments in understanding the nature of cryptic pockets and predicting their location.

NATURE OF CRYPTIC POCKETS AND MECHANISMS OF THEIR FORMATION

Similarly to all conformational changes of a target associated with ligand binding, the discussion on whether cryptic pockets emerge through conformational selection[17,18] or induced fit[19] is still ongoing. The former mechanism implies that the ligand's role is to stabilize specific conformations that are also accessible in the unbound state. The latter mechanism proposes that the ligand causes the target to explore regions of the conformational space that are practically inaccessible to the unbound form. As discussed elsewhere,[20] from the experimental point of view, only kinetic experiments might provide a clear distinction of the two mechanisms, as the binding rate in the case of conformational selection is dependent on the concentration of the ligand, while for the induced fit it is not. When the binding mechanism was ascertained experimentally, it emerged that, depending on the target, either induced fit or conformational selection are prevalent and in some cases both play a role.[21,22] In the case of cryptic sites, an extensive analysis of multiple X-ray structures of proteins with validated cryptic sites performed by Beglov *et al.*[23] showed that cryptic sites tend to be quite flexible which would hint towards conformational selection, but they are also almost always close to binding energy hot spots, often exploited by the bound ligands, suggesting that an induced fit component is also important for binding. This finding is in agreement with our simulations of the cryptic pocket found at the interface between α-helices H11 and H12 of TEM1 β-lactamase. Long timescale simulations of

the system have detected this cryptic pocket in the absence of a ligand and suggested roles for both induced fit and conformational selection.[24] The role of conformational selection was supported further by subsequent experiments that detected the cryptic pocket in the absence of ligand.[25] Enhanced sampling simulations of small fragments binding to the pocket are also consistent with a picture in which large fluctuations lead to the opening of the pocket (conformational selection) which are then stabilized by small molecules wedging between the helices (induced fit).[26]

Analysis of ~90 available crystal structures harboring cryptic sites showed that the main conformational changes associated with their opening are linked to: lateral chain rotation, loop movements, secondary structure changes, and interdomain motions (Figure 1).[23,27] Cryptic pockets have been also found to play a role at flexible protein-protein interfaces.[28]

Thus, cryptic pocket opening is associated with conformational changes of the target and the extent of such change plays a role in the *operative definition* of *crypticity* (i.e., a site "being cryptic").   The *operative definition* of *crypticity* is based on whether the pocket is present or visible in the *apo* structure and is typically a continuous measure based on measures of pocket exposure or steric clash with the ligand. Thus, depending on the chosen threshold, a pocket can be considered cryptic or not, and even sites contiguous with the catalytic pocket can end up being labelled as cryptic.

Algorithms for the detection and analysis of pockets *per se* (including their transient formation), are thus instrumental to distinguish and identify the cryptic ones. To this aim, a number of pocket analysis tools have been developed by the community. Tools such as Fpocket,[29] Epock,[30] POVME,[31] TRAPP,[32] and Nanoshaper[33] are able to detect and/or analyse both stable and transient

pockets and exposons focus on transient pockets. When the pockets are not particularly buried and the opening mechanism is mainly conformational selection, these tools can successfully detect the presence of cryptic pockets in the trajectories of the unliganded target.

Finally, the mere detection of a cryptic pocket does not make a target tractable *per se*. The newly found pocket needs to be ligandable. Having a cryptic pocket that is close to a binding hotspot and can accommodate chemically diverse fragments increases the chances that they are ligandable.[23]   Some pocket analysis algorithms and mixed solvent simulations also provide an estimate of the pocket ligandability.[29,32]

If a cryptic pocket does not cdirectly coincide with a functional site, then an allosteric connection of the remote pocket to the catalytic site or other functionally important sites is also needed for its druggability. As allosteric connections are bidirectional, an MD-based analysis of allosteric signal propagation from and to the active site of a target can be used to discover previously unknown binding sites. For instance, in the case of Hsp90, such a method was used to characterize the allosteric "hot spots" involved in interdomain communication pathways. One of these binding hotspots is cryptic and was successively shown to be druggable.[34,35]
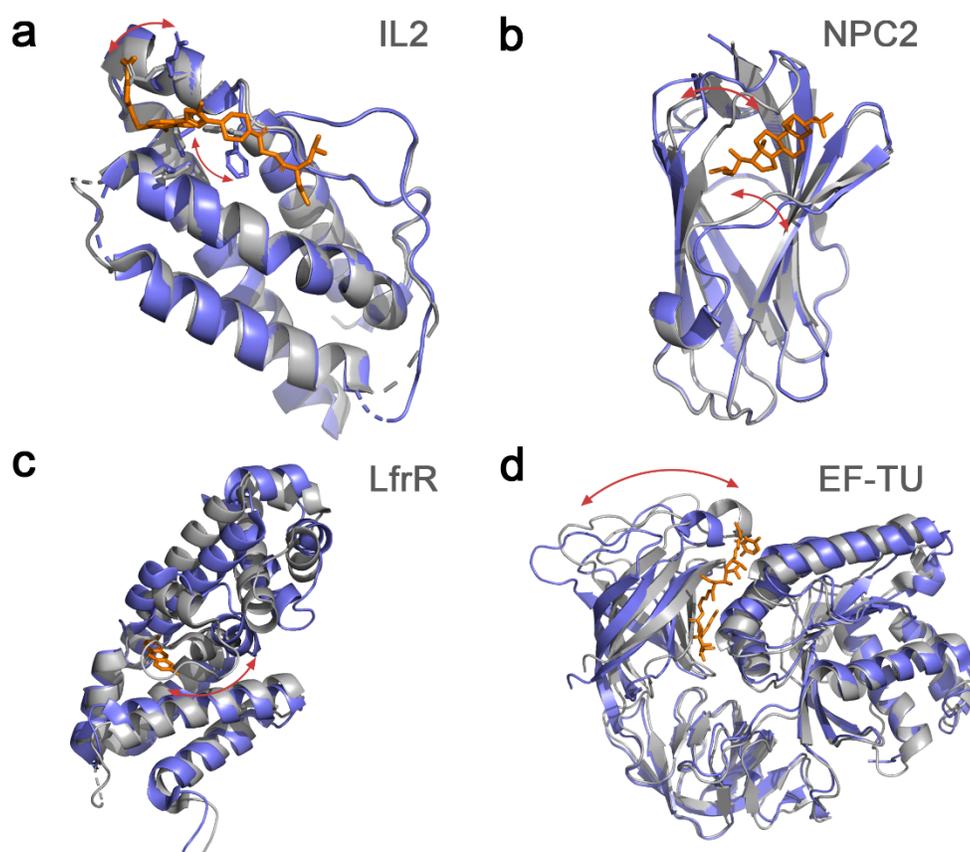
**Figure 1.** Illustration of common structural changes associated with cryptic binding site formation. **a.** Side chain rearrangements as shown in IL2 (*apo*: 1M47_A, *holo*: 1PY2_D), **b.** loop movements in NPC2 (*apo*: 1NEP_A, *holo*: 2HKA_C), **c.** secondary structure changes like the partial helix unfolding in LfrR (*apo*: 2WGB_A, *holo*: 2V57_A), and **d.** interdomain motions in EF-TU (*apo*: 1EXM_A, *holo*: 1HA3_B). *Apo* structures are show in gray and *holo* structures in slate blue, while red arrows highlight the main structural changes.

MIXED-SOLVENT MOLECULAR DYNAMICS SIMULATIONS

To facilitate the exploration of cryptic binding sites, one of the approaches based on MD simulations relies on the use of small probes mixed with water molecules that help open and stabilize cryptic pockets which are often hydrophobic. These simulations are known as mixed-

solvent MD simulations.[36-39] Probes that were successfully tested were either readily miscible with water (e.g., acetic acid, isopropanol, or resorcinol),[36,40] or quite hydrophobic (e.g. benzene)[26,41] and in such cases a repulsive potential is necessary to prevent the probes from clustering.[37,42] Mixed-solvent MD simulations have been successfully used on a number of targets to characterize active and allosteric sites and are relatively simple to set up.[28,36-38,43] Schmidt *et al*. tested the performance of various co-solvent compositions on seven diverse targets with known cryptic sites and reported that a composition of 90% water and 10% phenol was the most effective in opening the cavities without unfolding the proteins.[44] The pocket occupancy and fragment residence time obtained from mixed-solvent MD simulations can also provide druggability estimates, which is important to assess the feasibility of targeting a pocket with small molecules.[16,36] However, their application to cryptic pockets detection still requires substantial sampling of the system. For instance, we showed that, in the case of the cryptic pocket of TEM1 β-lactamase, 6 different hydrophobic ring probes were able to open the cryptic pocket only in a few 100-ns-long simulations out of 32 that we ran for each of the probes. Extending the mixed-solvent simulations with benzene to more than 1 μs resulted in the opening of the pocket in 1/3 of the simulations.[26] Sampling problems also affect the ability of mixed-solvent MD simulations to identify deeply buried cryptic pockets as their exploration relies on the diffusion of probe molecules around the protein surface and/or overcoming high energy barriers. An additional problem with mixed solvent MD is the risk of unfolding the protein with hydrophobic probes[44] which in some cases can be circumvented through a carefully selected set of position/distance restraints.[41]

COLLECTIVE-VARIABLE-DEPENDENT ENHANCED SAMPLING METHODS

When conventional MD simulations prove ineffective in sampling the event of interest, enhanced sampling methods can be used. Such techniques can be practically divided based on their dependence on collective variables (CV). In the case of CV-dependent methods, like metadynamics,[45,46] umbrella sampling,[47] and steered MD,[48] the process of interest is approximated through CVs, i.e. the functions of system's coordinates that can be as simple as a distance between two atoms or as complex as a contact map describing an active state of a protein. Unfortunately, choosing suitable CVs can be far from straightforward which is why these choices are often guided by experimental data. Recently, metadynamics was successfully applied to investigate the binding mechanism of SSR128129E (SSR), a newly developed inhibitor of fibroblast growth factor receptor 1 (FGFR1).[49,50] This receptor tyrosine kinase arose as a potential anticancer drug target due to its involvement in numerous essential cellular processes, such as blood vessel formation. SSR was initially discovered through high-throughput screening and its binding location was narrowed down to the extracellular D3 Ig-like domain, but the exact binding site eluded both X-ray crystallography and NMR due to the disordered nature of the domain. Using a range of CVs that describe not only the binding process, but also the folding of the domain, the authors were able to capture inhibitor's reversible binding to a hidden pocket formed through the elongation of a small α-helix in the D3 domain. The observed binding mechanism was further confirmed with a mutational analysis and subsequently used to design more potent inhibitors. Despite this success story, using CV methods based on geometric criteria to systematically predict cryptic binding sites is challenging as it is difficult to find a set of parameters that adequately captures the complexity of the cryptic binding pocket formation in the absence of previous information on the location and nature of the pocket.[26]

The JEDI methodology is a CV-based approach that overcomes these limitations and allows to bias sampling towards protein conformations of favorable druggability.[51] It uses a 3D grid overlapped on a region of interest on a protein surface to determine the location of putative pockets, and calculates a druggability score, i.e., a quantification of how adequate those pockets are to accommodate a small organic molecule. This druggability scoring was inspired by structural bioinformatics methodologies trained on datasets of X-ray crystallography-derived binding sites,[52,53] with the important difference that the druggability estimator was designed to be smooth and continuously differentiable with respect to protein atomic Cartesian coordinates. This enables the use of the JEDI score as a collective variable for biased MD simulations. Therefore, a typical workflow to seek cryptic pockets with JEDI involves working out the druggability score of a protein region covered by a manually placed 3D grid, and subsequently carrying out biased MD simulations to encourage the protein to adopt new conformations that increase (or decrease) the initial druggability score. Because the estimator scores favorably hydrophobic cavities suitably sized to accommodate a small molecule, the application of JEDI to a protein region that does not contain a pocket will encourage spontaneous sampling of a cavity without requiring *a priori* knowledge of which residues must undergo a conformational change to reveal a cryptic pocket. The process can be very rapid, requiring only a few nanoseconds of MD simulation (Figure 2). Current drawbacks of the first version of the algorithm include the relatively expensive cost of the CV, making long timescale simulations difficult, and support for implicit solvent simulations only. The JEDI CV can also suffer from degeneracy, leading to the biased simulations becoming stuck in an alternative conformational state that may not accurately describe the cryptic pocket owing to the approximate nature of the druggability estimator.
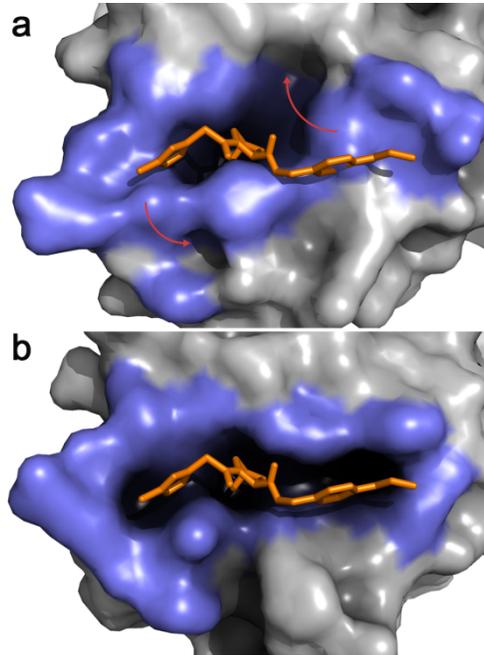
**Figure 2. a.** Representative snapshot from the most populated cluster of an equilibrium MD simulation of *apo* VHL. The binding-site lining residues (highlighted in slate blue) adopt a collapsed conformation with low druggability score **b.** A JEDI-biased MD simulation of *apo* VHL has rearranged the binding-site lining residues in a druggable conformation. In both panels an overlay of the crystallographic structure of a VHL ligand is displayed in orange sticks. Red arrows highlight the main structural changes.

COLLECTIVE-VARIABLE-INDEPENDENT ENHANCED SAMPLING METHODS

One of the most commonly used CV-independent enhanced sampling methods is parallel tempering (i.e., temperature replica exchange MD)[54] where several replicas of the system are run at the same time, but at different temperatures. At regular intervals, an exchange between the neighboring replicas is attempted and either accepted or rejected based on the Metropolis criterion. The method is usually quite effective (albeit computationally expensive as it requires a large number of replicas to span a wide range of temperatures) in studying protein

conformational landscapes. However when it came to studies of cryptic binding sites, it was unable to open the pockets in the targets that were tested, even at temperatures of 400 K.[26] This is probably due to the entropic nature of the energy barriers involved in pocket opening.[26] A different approach, based on non-equilibrium MD simulations (rotamerically induced perturbations or RIP) was proposed by Kokh *et al*. to initiate large-scale protein movements and sample cryptic sites.[55] This approach proved useful for a rapid evaluation of pocket flexibility. However, RIP is not able to predict folding of protein segments, nor the kinetics or thermodynamics associated with the protein motions.[56]

Considering the hydrophobic nature of cryptic binding pockets, Oleinikovas *et al*. devised a method in which they sample water interactions through scaled Hamiltonians (SWISH).[26] The method is based on a Hamiltonian replica exchange approach and, instead of altering the temperature of the replicas as in parallel tempering, it changes the interaction of apolar carbons and sulfurs with water oxygen. This renders the protein less hydrophobic at higher replicas and allows for more effective exploration of cryptic binding sites compared to conventional MD simulations. The method was tested on a range of pharmaceutically relevant targets, such as $\beta$-lactamase, IL2, and PLK1 in the initial study,[26] and NPC2, LfrR, p38$\alpha$, and hPNMT in the follow-up study.[41] SWISH was able to successfully detect known cryptic binding sites in these proteins, typically within 1 $\mu$s of of sampling time with 6-8 replicas. When combined with probes which stabilized the open cryptic binding sites, it became apparent that the formation mechanism of cryptic binding pockets typically involves both conformational selection and induced fit. However, the approach combining SWISH with probes suffers from the same drawbacks as mixed-solvent simulations where the proteins can become denatured when the selected probes are hydrophobic in nature. Such behavior is even more pronounced in SWISH

simulations where the protein is already pushed towards more hydrophilic states. A way to prevent the protein from unfolding is either through a more conservative set of scaling factors or by using a contact-map restraint[41] which is applied on a small set of relevant contacts away from the site of interest that essentially keeps the secondary structure elements folded. Further iterations of the method involve a more targeted approach where only the interactions of hydrophobic residues with water are scaled which would hopefully also reduce the protein unfolding and rendering the contact-map restraint unnecessary. Another advantage of SWISH is that it can be easily combined with other enhanced sampling methods, such as metadynamics, if the cryptic binding pocket of interest forms through a more complex conformational change, as was the case of FGFR1.[50]

MARKOV STATE MODELS (MSMs)

MSMs are a powerful means to integrate many independent simulations into a map of protein's conformational space that captures long-timescale events far beyond the reach of any individual simulation.[57] These methods were originally developed to study protein folding mechanisms.[58] However, they have since proved extremely effective in a wide variety of settings.[59]

Inspired by the conformational selection model, MSMs were first applied to the *apo* TEM1 β-lactamase enzyme introduced above.[24] The resulting model successfully identified the cryptic pocket between helices 11 and 12, supporting a role for conformational selection in the opening of such pockets. A second model built in the presence of a known ligand for this cryptic site also demonstrated a role for induced fit, consistent with previous applications of MSMs to other protein-ligand binding processes[60] that revealed an interplay between conformational selection and induced fit. In addition to recapitulating the known cryptic pocket, subsequent analysis of the

13

β-lactamase MSM revealed at least two additional cryptic pockets with allosteric coupling to the active site.[25] To test this prediction, an experimental thiol labeling assay was applied to detect the exposure of buried residues upon pocket opening. The results confirmed the two newly predicted pockets, a role for conformational selection in the opening of all three sites, and their ability to allosterically impact function.[25] Since then, MSMs have been applied to hunt for cryptic pockets in a number of other proteins.[61-63] New adaptive sampling algorithms have also been developed to reduce the computational cost of building such models.[64] Strikingly, these approaches have revealed yet another cryptic allosteric site that is common to multiple β-lactamase families (Fig. 3).[65] Experimental tests have confirmed the existence of this pocket and demonstrated that it exerts more potent allosteric control over enzymatic activity than the previously identified pockets.[65,66] Applications to multiple related proteins are also beginning to reveal the sequence-dependence of the probabilities of potentially druggable sites.[39,66,67]
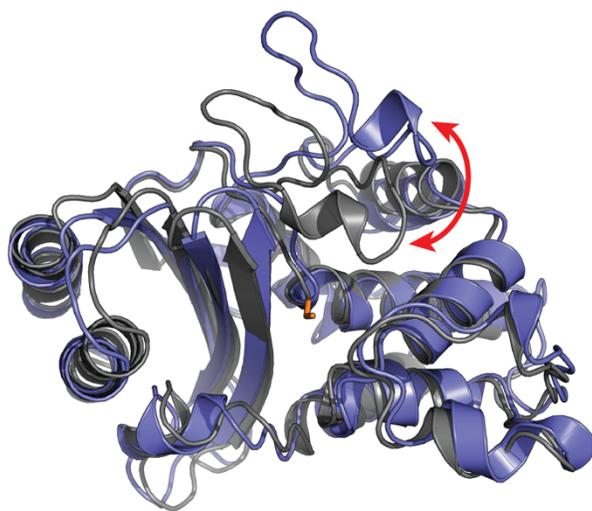


**Figure 3.** A new cryptic pocket in TEM1 β-lactamase that was discovered with MSMs and subsequently confirmed by experimental tests. An open structure (blue) is overlaid on the *apo*

crystal structure (gray) with the key catalytic serine (S70) in sticks (orange). The red arrow highlights a large loop motion that creates a cryptic pocket adjacent to the active site.

Having succeeded in detecting cryptic pockets, MSMs are now actively being used to intentionally target these sites with small molecules. Hypothetically, this could be achieved by taking an individual structure with an open cryptic pocket and applying existing rational drug design tools. However, more effective approaches could systematically take advantage of all the conformational diversity an MSM captures. One proposal for achieving this end, called Boltzmann docking,[67] works by docking a library of compounds against a representative structure from each state in an MSM and then, using the equilibrium probabilities of those states, to calculate a Boltzmann-weighted average docking score. Boltzmann docking is related to ensemble docking.[68] The primary difference is that Boltzmann docking uses the equilibrium probabilities of each state instead of treating all conformations equally. In doing so, Boltzmann docking automatically favors compounds that bind to higher probability states over compounds that bind higher-energy (i.e. lower probability) states where a greater penalty has to be paid to stabilize the conformation. Boltzmann docking has been successfully applied to find new allosteric modulators of TEM1 β-lactamase, including both inhibitors and activators that have been confirmed experimentally.[69] The discovery of activators is intriguing as it suggests the potential for enhancing desirable functions in other settings, e.g., to counteract the deleterious effects of a disease-causing mutation.

OTHER COMPUTATIONAL APPROACHES

Despite the aforementioned advances in computer software and hardware, MD simulations still require a significant amount of computational resources and time to adequately sample protein

conformational landscapes. Thus, the appeal of inexpensive computational methods that rapidly predict cryptic binding sites is evident. Cimermancic *et al*. recently curated a data set of *apo* and *holo* protein pairs from the PDB harboring cryptic binding sites (composed of ~90 proteins) which they used to build a machine learning model (CryptoSite) to predict such sites in proteins considered undruggable.[27] The authors initially used over 50 features they deemed potentially relevant for the cryptic binding site prediction, however, only three of them proved to be statistically significant - the average pocket score calculated for conformations obtained from MD simulations that rely on a simplified energy landscape, the sequence conservation of the site, and the likelihood of binding small-molecule fragments. While the approach is certainly fast and remarkably accurate for the testing set, it drew criticism due to the overall small size of the training and testing sets, as well as the use of a single *apo* structure in each pair which can lead to an overestimation of the site's crypticity as the pocket could be quite apparent in other *apo* structures.[23,70] Beglov *et al*. expanded the original CryptoSite data set by including these structures and analyzed the new set using their own tools (FTMap and FTFlex).[23] The authors showed that cryptic binding sites in *apo* structures typically have a strong binding hot spot in their vicinity and that they exhibit an above-average flexibility. It should be, however, noted that the analysis of X-ray structures has its limitations due to their static nature and the presence of crystal contacts that can introduce a range of artifacts.[71] Efforts are also underway to identify cryptic pockets using Rosetta-based Monte Carlo sampling of the conformational space.[72]

PERSPECTIVES

The past few years have seen the successful execution of a number of proof-of-concept studies that have demonstrated it is possible to computationally identify cryptic pockets and then target them with small molecules. Methods based on MD simulations combined with algorithms for the

detection of pockets, mixed solvents, enhanced sampling, and Markov state models are increasingly effective in sampling and identifying cryptic pockets, providing high-quality structural models for the design of novel drug-like compounds. Some of these methodologies, namely CV-based enhanced sampling methods and MSMs, are also able to provide accurate estimates of the free energy penalty associated with the opening of the pockets, while simulations with mixed solvents and fragments provide precious information on the ligandability of the newly discovered sites. As many cryptic pockets are allosteric, i.e., remote from known functional sites, an important question is whether binding of a small molecule at a cryptic pocket has a functional effect on the protein. Even in this case, methods to characterize allosteric communication networks based on simulations can provide solid predictions. Looking ahead, as many new cryptic pockets are predicted by simulations, experimental validation becomes a crucial bottleneck in the search for cheaper solutions than the ones currently used. Machine learning approaches predict that a vast number of potential drug targets in the human genome harbor cryptic sites. It will be exciting to see how many of those, particularly in validated drug targets that small molecule drugs have yet to be discovered for, can be targeted by leveraging the cryptic sites predicted by simulation-based approaches.

## AUTHOR INFORMATION

**Corresponding Author**

*E-mail: f.l.gervasio@ucl.ac.uk.

**Author Contributions**

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

**Funding Sources**

**Notes**

The authors declare no competing financial interest.

**Biographies**

**Gregory R. Bowman** is an Associate Professor in the Department of Biochemistry & Molecular Biophysics at the Washington University School of Medicine. His research focuses on the mechanisms of allosteric coupling and the opportunities this insight provides for manipulating function with drugs and mutations. A major emphasis is on identifying and targeting cryptic pockets. Half the lab develops and applies novel computational tools, often based on Markov state model (MSM) methods, while the other half experimentally tests small molecule and protein designs.

**Julien Michel** is a Senior Lecturer at the School of Chemistry of the University of Edinburgh. His research interests cover methods and applications of molecular simulations and their integration with biophysical experiments for studies of protein dynamics, protein-ligand

interactions, and for drug discovery. His group has developed the JEDI approach for cryptic site discovery, and characterized cryptic site formation mechanisms in important drug targets

**Jordi Juarez-Jimenez** is a Research Associate at the School of Chemistry of the University of Edinburgh. His research focuses on understanding how protein conformational changes can be exploited in rational drug design. He is currently working on enhancing the functionalities of the JEDI algorithm in combination with different computational tools.

**Antonija Kuzmanic** is a Research Associate at University College London. She splits her time between trying to understand what makes kinases active and cryptic pockets hidden using molecular dynamics simulations and enhanced sampling methods.

**Francesco L. Gervasio** is a full Professor in Chemistry and Structural and Molecular Biology at University College London and a full Professor in Biomolecular and Pharmaceutical Modelling at the University of Geneva. His research interests are in the field of computational chemistry, biophysics, structural biology, and drug discovery. He has co-developed a number of enhanced-sampling algorithms, including the SWISH approach for cryptic site discovery.

REFERENCES

(1)     Hopkins, A. L.; Groom, C. R. The druggable genome. *Nat. Rev. Drug Discov.* **2002**, *1*, 727-730.

(2)     Oprea, T. I.; Bologa, C. G.; Brunak, S.; Campbell, A.; Gan, G. N.; Gaulton, A.; Gomez, S. M.; Guha, R.; Hersey, A.; Holmes, J.; Jadhav, A.; Jensen, L. J.; Johnson, G. L.; Karlson, A.; Leach, A. R.; Ma'ayan, A.; Malovannaya, A.; Mani, S.; Mathias, S. L.; McManus, M. T.; Meehan, T. F.; von Mering, C.; Muthas, D.; Nguyen, D.-T.; Overington, J. P.; Papadatos, G.; Qin, J.; Reich, C.; Roth, B. L.; Schürer, S. C.; Simeonov, A.; Sklar, L. A.; Southall, N.; Tomita, S.; Tudose, I.; Ursu, O.; Vidović, D.; Waller, A.; Westergaard, D.; Yang, J. J.;

Zahoránszky-Köhalmi, G. Unexplored therapeutic opportunities in the human genome. *Nat. Rev. Drug Discov.* **2018**, *17*, 317.

(3)    Ostrem, J. M.; Peters, U.; Sos, M. L.; Wells, J. A.; Shokat, K. M. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature* **2013**, *503*, 548.

(4)    Canon, J.; Rex, K.; Saiki, A. Y.; Mohr, C.; Cooke, K.; Bagal, D.; Gaida, K.; Holt, T.; Knutson, C. G.; Koppada, N.; Lanman, B. A.; Werner, J.; Rapaport, A. S.; Miguel, T. S.; Ortiz, R.; Osgood, T.; Sun, J.-R.; Zhu, X.; McCarter, J. D.; Volak, L. P.; Houk, B. E.; Fakih, M. G.; O'Neil, B. H.; Price, T. J.; Falchook, G. S.; Desai, J.; Kuo, J.; Govindan, R.; Hong, D. S.; Ouyang, W.; Henary, H.; Arvedson, T.; Cee, V. J.; Lipford, J. R. The clinical KRAS(G12C) inhibitor AMG 510 drives anti-tumour immunity. *Nature* **2019**, *575*, 217-223.

(5)    Shaw, D. E.; Grossman, J. P.; Bank, J. A.; Batson, B.; Butts, J. A.; Chao, J. C.; Deneroff, M. M.; Dror, R. O.; Even, A.; Fenton, C. H.; Forte, A.; Gagliardo, J.; Gill, G.; Greskamp, B.; Ho, C. R.; Ierardi, D. J.; Iserovich, L.; Kuskin, J. S.; Larson, R. H.; Layman, T.; Lee, L.-S.; Lerer, A. K.; Li, C.; Killebrew, D.; Mackenzie, K. M.; Mok, S. Y.-H.; Moraes, M. A.; Mueller, R.; Nociolo, L. J.; Peticolas, J. L.; Quan, T.; Ramot, D.; Salmon, J. K.; Scarpazza, D. P.; Schafer, U. B.; Siddique, N.; Snyder, C. W.; Spengler, J.; Tang, P. T. P.; Theobald, M.; Toma, H.; Towles, B.; Vitale, B.; Wang, S. C.; Young, C.: Anton 2: raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*; IEEE Press: New Orleans, Louisana, 2014; pp 41-53.

(6)    Shirts, M.; Pande, V. S. COMPUTING: Screen savers of the world unite! *Science* **2000**, *290*, 1903-1904.

(7)    Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19-25.

(8)    Páll, S.; Abraham, M. J.; Kutzner, C.; Hess, B.; Lindahl, E. Tackling exascale software challenges in molecular dynamics simulations with GROMACS, In *Solving Software Challenges for Exascale*, Cham, 2015; Springer International Publishing.

(9)    von Bulow, S.; Siggel, M.; Linke, M.; Hummer, G. Dynamic cluster formation determines viscosity and diffusion in dense protein solutions. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 9843-9852.

(10)     Voelz, V. A.; Bowman, G. R.; Beauchamp, K.; Pande, V. S. Molecular simulation of ab initio protein folding for a millisecond folder NTL9(1-39). *J. Am. Chem. Soc.* **2010**, *132*, 1526-1528.

(11)     Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696-3713.

(12)     Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmuller, H.; MacKerell, A. D., Jr. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71-73.

(13)     Robustelli, P.; Piana, S.; Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E4758-E4766.

(14)     Kuzmanic, A.; Pritchard, R. B.; Hansen, D. F.; Gervasio, F. L. Importance of the force field choice in capturing functionally relevant dynamics in the von Willebrand factor. *J. Phys. Chem. Lett.* **2019**, *10*, 1928-1934.

(15)     Robertson, M. J.; Tirado-Rives, J.; Jorgensen, W. L. Improved peptide and protein torsional energetics with the OPLS-AA force field. *J. Chem. Theory Comput.* **2015**, *11*, 3499-3509.

(16)     Ghanakota, P.; Carlson, H. A. Moving beyond active-site detection: MixMD applied to allosteric systems. *J. Phys. Chem. B* **2016**, *120*, 8685-8695.

(17)     Monod, J.; Wyman, J.; Changeux, J. P. On the nature of allosteric transitions: A plausible model. *J. Mol. Biol.* **1965**, *12*, 88-118.

(18)     Tsai, C. J.; Ma, B.; Nussinov, R. Folding and binding cascades: shifts in energy landscapes. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 9970-9972.

(19)     Koshland, D. E. Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **1958**, *44*, 98-104.

(20)     Gianni, S.; Dogan, J.; Jemth, P. Distinguishing induced fit from conformational selection. *Biophys. Chem.* **2014**, *189*, 33-39.

(21)     Hammes, G. G.; Chang, Y. C.; Oas, T. G. Conformational selection or induced fit: a flux description of reaction mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 13737-13741.

(22)     Morando, M. A.; Saladino, G.; D'Amelio, N.; Pucheta-Martinez, E.; Lovera, S.; Lelli, M.; Lopez-Mendez, B.; Marenchino, M.; Campos-Olivas, R.; Gervasio, F. L. Conformational selection and induced fit mechanisms in the binding of an anticancer drug to the c-Src kinase. *Sci. Rep.* **2016**, *6*, 24439.

(23)     Beglov, D.; Hall, D. R.; Wakefield, A. E.; Luo, L.; Allen, K. N.; Kozakov, D.; Whitty, A.; Vajda, S. Exploring the structural origins of cryptic sites on proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E3416-E3425.

(24)     Bowman, G. R.; Geissler, P. L. Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 11681-11686.

(25)     Bowman, G. R.; Bolin, E. R.; Hart, K. M.; Maguire, B. C.; Marqusee, S. Discovery of multiple hidden allosteric sites by combining Markov state models and experiments. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 2734-2739.

(26)     Oleinikovas, V.; Saladino, G.; Cossins, B. P.; Gervasio, F. L. Understanding cryptic pocket formation in protein targets by enhanced sampling simulations. *J. Am. Chem. Soc.* **2016**, *138*, 14257-14263.

(27)     Cimermancic, P.; Weinkam, P.; Rettenmaier, T. J.; Bichmann, L.; Keedy, D. A.; Woldeyes, R. A.; Schneidman-Duhovny, D.; Demerdash, O. N.; Mitchell, J. C.; Wells, J. A.; Fraser, J. S.; Sali, A. CryptoSite: Expanding the druggable proteome by characterization and prediction of cryptic binding sites. *J. Mol. Biol.* **2016**, *428*, 709-719.

(28)     Schames, J. R.; Henchman, R. H.; Siegel, J. S.; Sotriffer, C. A.; Ni, H. H.; McCammon, J. A. Discovery of a novel binding trench in HIV integrase. *J. Med. Chem.* **2004**, *47*, 1879-1881.

(29)     Le Guilloux, V.; Schmidtke, P.; Tuffery, P. Fpocket: An open source platform for ligand pocket detection. *Bmc Bioinformatics* **2009**, *10*, 168.

(30)     Laurent, B.; Chavent, M.; Cragnolini, T.; Dahl, A. C.; Pasquali, S.; Derreumaux, P.; Sansom, M. S.; Baaden, M. Epock: rapid analysis of protein pocket dynamics. *Bioinformatics* **2015**, *31*, 1478-1480.

(31)     Wagner, J. R.; Sørensen, J.; Hensley, N.; Wong, C.; Zhu, C.; Perison, T.; Amaro, R. E. POVME 3.0: Software for mapping binding pocket flexibility. *J. Chem. Theory Comput.* **2017**, *13*, 4584-4592.

(32)     Kokh, D. B.; Richter, S.; Henrich, S.; Czodrowski, P.; Rippmann, F.; Wade, R. C. TRAPP: a tool for analysis of transient binding pockets in proteins. *J. Chem. Inf. Model.* **2013**, *53*, 1235-1252.

(33)     La Sala, G.; Decherchi, S.; De Vivo, M.; Rocchia, W. Allosteric communication networks in proteins revealed through pocket crosstalk analysis. *ACS Cent. Sci.* **2017**, *3*, 949-960.

(34)     Morra, G.; Neves, M. A.; Plescia, C. J.; Tsustsumi, S.; Neckers, L.; Verkhivker, G.; Altieri, D. C.; Colombo, G. Dynamics-based discovery of allosteric inhibitors: selection of new ligands for the C-terminal domain of Hsp90. *J. Chem. Theory Comput.* **2010**, *6*, 2978-2989.

(35)     Sattin, S.; Tao, J.; Vettoretti, G.; Moroni, E.; Pennati, M.; Lopergolo, A.; Morelli, L.; Bugatti, A.; Zuehlke, A.; Moses, M.; Prince, T.; Kijima, T.; Beebe, K.; Rusnati, M.; Neckers, L.; Zaffaroni, N.; Agard, D. A.; Bernardi, A.; Colombo, G. Activation of Hsp90 enzymatic activity and conformational dynamics through rationally designed allosteric ligands. *Chemistry* **2015**, *21*, 13598-13608.

(36)     Seco, J.; Luque, F. J.; Barril, X. Binding site detection and druggability index from first principles. *J. Med. Chem.* **2009**, *52*, 2363-2371.

(37)     Guvench, O.; MacKerell, A. D., Jr. Computational fragment-based binding site identification by ligand competitive saturation. *PLoS Comput. Biol.* **2009**, *5*, e1000435.

(38)     Bakan, A.; Nevins, N.; Lakdawala, A. S.; Bahar, I. Druggability assessment of allosteric proteins by dynamics simulations in thepresence of probe molecules. *J. Chem. Theory Comput.* **2012**, *8*, 2435-2447.

(39)     Sultan, M. M.; Kiss, G.; Pande, V. S. Towards simple kinetic models of functional dynamics for a kinase subfamily. *Nat. Chem.* **2018**, *10*, 903-909.

(40)     Kimura, S. R.; Hu, H. P.; Ruvinsky, A. M.; Sherman, W.; Favia, A. D. Deciphering cryptic binding sites on proteins by mixed-solvent molecular dynamics. *J. Chem. Inf. Model.* **2017**, *57*, 1388-1401.

(41)     Comitani, F.; Gervasio, F. L. Exploring cryptic pockets formation in targets of pharmaceutical interest with SWISH. *J. Chem. Theory Comput.* **2018**, *14*, 3321-3331.

(42)     Raman, E. P.; Yu, W.; Guvench, O.; Mackerell, A. D. Reproducing crystal binding modes of ligand functional groups using site-identification by ligand competitive saturation (SILCS) simulations. *J. Chem. Inf. Model.* **2011**, *51*, 877-896.

(43)     Eyrisch, S.; Helms, V. Transient pockets on protein surfaces involved in protein-protein interaction. *J. Med. Chem.* **2007**, *50*, 3457-3464.

(44)     Schmidt, D.; Boehm, M.; McClendon, C. L.; Torella, R.; Gohlke, H. Cosolvent-enhanced sampling and unbiased identification of cryptic pockets suitable for structure-based drug design. *J. Chem. Theory Comput.* **2019**, *15*, 3331-3343.

(45)     Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 826-843.

(46)     Laio, A.; Gervasio, F. L. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep. Prog. Phys.* **2008**, *71*, 126601.

(47)     Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comp. Phys.* **1977**, *23*, 187-199.

(48)     Gullingsrud, J. R.; Braun, R.; Schulten, K. Reconstructing potentials of mean force through time series analysis of steered molecular dynamics simulations. *J. Comp. Phys.* **1999**, *151*, 190-211.

(49)     Bono, F.; De Smet, F.; Herbert, C.; De Bock, K.; Georgiadou, M.; Fons, P.; Tjwa, M.; Alcouffe, C.; Ny, A.; Bianciotto, M.; Jonckx, B.; Murakami, M.; Lanahan, A. A.; Michielsen, C.; Sibrac, D.; Dol-Gleizes, F.; Mazzone, M.; Zacchigna, S.; Herault, J. P.; Fischer, C.; Rigon, P.; Ruiz de Almodovar, C.; Claes, F.; Blanc, I.; Poesen, K.; Zhang, J.; Segura, I.; Gueguen, G.; Bordes, M. F.; Lambrechts, D.; Broussy, R.; van de Wouwer, M.; Michaux, C.; Shimada, T.; Jean, I.; Blacher, S.; Noel, A.; Motte, P.; Rom, E.; Rakic, J. M.; Katsuma, S.; Schaeffer, P.; Yayon, A.; Van Schepdael, A.; Schwalbe, H.; Gervasio, F. L.; Carmeliet, G.; Rozensky, J.; Dewerchin, M.; Simons, M.; Christopoulos, A.; Herbert, J. M.; Carmeliet, P. Inhibition of tumor angiogenesis and growth by a small-molecule multi-FGF receptor blocker with allosteric properties. *Cancer Cell* **2013**, *23*, 477-488.

(50)     Herbert, C.; Schieborr, U.; Saxena, K.; Juraszek, J.; De Smet, F.; Alcouffe, C.; Bianciotto, M.; Saladino, G.; Sibrac, D.; Kudlinzki, D.; Sreeramulu, S.; Brown, A.; Rigon, P.; Herault, J. P.; Lassalle, G.; Blundell, T. L.; Rousseau, F.; Gils, A.; Schymkowitz, J.; Tompa, P.; Herbert, J. M.; Carmeliet, P.; Gervasio, F. L.; Schwalbe, H.; Bono, F. Molecular mechanism of SSR128129E, an extracellularly acting, small-molecule, allosteric inhibitor of FGF receptor signaling. *Cancer Cell* **2013**, *23*, 489-501.

(51)    Cuchillo, R.; Pinto-Gil, K.; Michel, J. A collective variable for the rapid exploration of protein druggability. *J. Chem. Theory Comput.* **2015**, *11*, 1292-1307.

(52)    Ghanakota, P.; DasGupta, D.; Carlson, H. A. Free energies and entropies of binding sites identified by MixMD cosolvent simulations. *J. Chem. Inf. Model.* **2019**, *59*, 2035-2045.

(53)    Halgren, T. New method for fast and accurate binding-site identification and analysis. *Chem. Biol. Drug Des.* **2007**, *69*, 146-148.

(54)    Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141-151.

(55)    Kokh, D. B.; Czodrowski, P.; Rippmann, F.; Wade, R. C. Perturbation approaches for exploring protein binding site flexibility to predict transient binding pockets. *J. Chem. Theory Comput.* **2016**.

(56)    Le Guilloux, V.; Schmidtke, P.; Tuffery, P. Fpocket: An open source platform for ligand pocket detection. *Bmc Bioinformatics* **2009**, *10*.

(57)    Bowman, G.; Pande, V.; Noé, F.: *An introduction to Markov state models and their application to long timescale molecular simulation*, 2014; Vol. 797.

(58)    Bowman, G. R.; Voelz, V. A.; Pande, V. S. Taming the complexity of protein folding. *Curr. Opin. Struct. Biol.* **2011**, *21*, 4-11.

(59)    Chodera, J. D.; Noe, F. Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* **2014**, *25*, 135-144.

(60)    Silva, D.-A.; Bowman, G. R.; Sosa-Peinado, A.; Huang, X. A role for both conformational selection and induced fit in ligand binding by the LAO protein. *PLoS Comput. Biol.* **2011**, *7*, e1002054.

(61)    Wagner, J. R.; Lee, C. T.; Durrant, J. D.; Malmstrom, R. D.; Feher, V. A.; Amaro, R. E. Emerging computational methods for the rational discovery of allosteric drugs. *Chem. Rev.* **2016**, *116*, 6370-6390.

(62)    Knoverek, C. R.; Amarasinghe, G. K.; Bowman, G. R. Advanced methods for accessing protein shape-shifting present new therapeutic opportunities. *Trends Biochem. Sci.* **2019**, *44*, 351-364.

(63)    Ferruz, N.; Doerr, S.; Vanase-Frawley, M. A.; Zou, Y.; Chen, X.; Marr, E. S.; Nelson, R. T.; Kormos, B. L.; Wager, T. T.; Hou, X.; Villalobos, A.; Sciabola, S.; De Fabritiis,

G. Dopamine D3 receptor antagonist reveals a cryptic pocket in aminergic GPCRs. *Sci. Rep.* **2018**, *8*, 897.

(64)     Zimmerman, M. I.; Bowman, G. R. FAST conformational searches by balancing exploration/exploitation trade-offs. *J. Chem. Theory Comput.* **2015**, *11*, 5747-5757.

(65)     Porter, J. R.; Moeder, K. E.; Sibbald, C. A.; Zimmerman, M. I.; Hart, K. M.; Greenberg, M. J.; Bowman, G. R. Cooperative changes in solvent exposure identify cryptic pockets, switches, and allosteric coupling. *Biophys. J.* **2019**, *116*, 818-830.

(66)     Chen, S.; Wiewiora, R. P.; Meng, F.; Babault, N.; Ma, A.; Yu, W.; Qian, K.; Hu, H.; Zou, H.; Wang, J.; Fan, S.; Blum, G.; Pittella-Silva, F.; Beauchamp, K. A.; Tempel, W.; Jiang, H.; Chen, K.; Skene, R. J.; Zheng, Y. G.; Brown, P. J.; Jin, J.; Luo, C.; Chodera, J. D.; Luo, M. The dynamic conformational landscape of the protein methyltransferase SETD8. *eLife* **2019**, *8*, e45403.

(67)     Hart, K. M.; Ho, C. M.; Dutta, S.; Gross, M. L.; Bowman, G. R. Modelling proteins' hidden conformations to predict antibiotic resistance. *Nat. Commun.* **2016**, *7*, 12965.

(68)     Amaro, R. E.; Baudry, J.; Chodera, J.; Demir, O.; McCammon, J. A.; Miao, Y.; Smith, J. C. Ensemble docking in drug discovery. *Biophys. J.* **2018**, *114*, 2271-2278.

(69)     Hart, K. M.; Moeder, K. E.; Ho, C. M. W.; Zimmerman, M. I.; Frederick, T. E.; Bowman, G. R. Designing small molecules to target cryptic pockets yields both positive and negative allosteric modulators. *PLoS One* **2017**, *12*, e0178678.

(70)     Vajda, S.; Beglov, D.; Wakefield, A. E.; Egbert, M.; Whitty, A. Cryptic binding sites on proteins: definition, detection, and druggability. *Curr. Opin. Chem. Biol.* **2018**, *44*, 1-8.

(71)     Sondergaard, C. R.; Garrett, A. E.; Carstensen, T.; Pollastri, G.; Nielsen, J. E. Structural artifacts in protein-ligand X-ray structures: implications for the development of docking scoring functions. *J. Med. Chem.* **2009**, *52*, 5673-5684.

(72)     Johnson, D. K.; Karanicolas, J. Druggable protein interaction sites are more predisposed to surface pocket formation than the rest of the protein surface. *PLoS Comput. Biol.* **2013**, *9*, e1002951.