Investigating drug-target association and dissociation mechanisms using metadynamics-based algorithms

Andrea Cavalli[1,2], Andrea Spitaleri[2], Giorgio Saladino[3], Francesco L. Gervasio*[3]*

[1]Department of Pharmacy and Biotechnology, University of Bologna, via Belmeloro 6, 40126 Bologna, Italy

[2]D3 Computation, Istituto Italiano di Tecnologia, via Morego 30, 16163 Genova, Italy

[3]Department of Chemistry and Institute of Structural and Molecular Biology, University College London, WC1E 6BT London, United Kingdom

Conspectus

This Account highlights recent advances and discusses major challenges in the field of drug-target recognition, binding, and unbinding studied using metadynamics-based approaches, with particular emphasis on their role in structure-based design. Computational chemistry has significantly contributed to drug design and optimization in an extremely broad range of areas, including prediction of target druggability and drug likeness, de novo design, fragment screening, ligand docking, estimation of binding affinity, and modulation of ADMET (absorption, distribution, metabolism, excretion, toxicity) properties. Computationally driven drug discovery must continuously adapt to keep pace with the evolving knowledge of the factors that modulate the pharmacological action of drugs. There is thus an urgent need for novel computational approaches that integrate the vast amount of complex information currently available for small (bio)organic compounds, biologically relevant targets and their complexes, while also accounting accurately for the thermodynamics and kinetics of drug-target association, the intrinsic dynamical behavior of biomolecular systems, and the complexity of protein-protein networks. Understanding the mechanism of drugs binding to and unbinding from biological targets is fundamental for optimizing lead compounds and designing novel biologically active ones. One major challenge is the accurate description of the conformational complexity prior to and upon formation of drug-target complexes. Recently, enhanced sampling methods, including metadynamics and related approaches, have been successfully applied to investigate complex mechanisms of drugs binding to flexible targets. Metadynamics is a family of enhanced sampling techniques aimed at enhancing the rare events and reconstructing the underlying free energy landscape as a function of a set of order parameters, usually referred to as collective variables.

Studies of drug binding mechanisms have predicted the most probable association and dissociation pathways and the related binding free energy profile. In addition, the availability of an efficient open-source implementation, running on cost-effective GPU (i.e. graphical processor unit) architectures, has considerably decreased the learning curve and the computational costs of the methods, and increased their adoption by the community. Here, we review the recent contributions of metadynamics and other enhanced sampling methods to the field of drug-target recognition and binding. We discuss how metadynamics has been used to search for transition states, to predict binding and unbinding paths, to treat conformation flexibility, and compute free energy profiles. We highlight the importance of such predictions in drug discovery. Major challenges in the field and possible solutions will finally be discussed.

1. Introduction

Approximately 12 years and over 1 billion dollars are required to develop one new medicine from the earliest stages of discovery to approval. Only a tiny percentage of the compounds entering the drug discovery pipeline reach the market. High-throughput screening (HTS) is one of the primary pharmaceutical methods for identifying lead compounds, but it has a high false positive rate. Candidates are not routinely translated into new drugs. This leads to the delay or failure of drug discovery projects.[1] The reasons for this include unclear target biology, poor drug-like properties, low potency or selectivity, lack of efficacy, unexpected toxicity, etc. Failures in the later stages are extremely costly. It is therefore of paramount importance to identify early-on lead compounds that interact with the target in a biologically relevant mechanism without inducing adverse modes of action.

In recent years, more dynamic models of molecular recognition have superseded Emil Fischer's rigid lock-and-key binding paradigm. The two major paradigms are now the induced-

fit and the conformational selection models.[2] It is now increasingly clear that both mechanisms contribute to molecular recognition, which requires a dynamic description over timescales spanning several orders of magnitude to be fully understood. This presents a fundamental theoretical and computational challenge.[3]

Molecular dynamics (MD) simulations have greatly contributed to understanding molecular recognition phenomena as a fully dynamical process. But they are limited by the timescales that can be routinely sampled. Very recently, with the introduction of special-purpose machines such as Anton,[4] the porting of MD codes to GPUs,[5] and the evolution of parallel codes,[6,7] the timescales accessible by fully atomistic MD have increased enormously. Simulations lasting up to a few ms (corresponding to 10^{12} time-steps) are now possible. However, drug-like molecules with long residence times (more than an hour) are common, and their unbinding cannot be observed by conventional MD calculations even when specialized hardware is used. This well-recognized limitation of MD is not limited to ligand binding and has led to the development of many innovative algorithms to enhance the sampling of the high-in-free-energy states and the *rare events* that allow the crossing of very high-free-energy barriers.[8]

Enhanced sampling methods speed up conformational sampling by various means. Umbrella sampling,[9] replica exchange,[10] metadynamics,[11] steered MD,[12] accelerated MD,[13] milestoning,[14] transition path sampling,[15] and their many combinations and derivatives are among the most widely used methods to enhance conformational sampling. Of the enhanced sampling methods that fully explore the binding mechanism, metadynamics,[11] especially in the well-tempered formulation,[16] has emerged as a powerful approach for accelerating rare events and computing multidimensional free energy surfaces (FESs). Metadynamics is a family of methods that flatten the FES as a function of a set of *collective variables* (CVs), by introducing a bias that prevents

the system from being trapped in local free energy minima (Figure 1). It shares many similarities with other CV-based algorithms,[17–22] and it has significantly evolved since its introduction. In contrast to other approaches, the reconstruction of the FES can be easily done in more than 2 dimensions and does not require an additional step (such as WHAM[23]). It has been recently proved under general circumstances that well-tempered metadynamics provides an unbiased estimate of the free energy of the system projected onto predefined CVs, albeit the convergence time depends critically on how optimal are the CVs.[24] By "optimal CVs", one generally means a small set (2-4) of variables, function of the atomic positions (e.g. distance), whose combination provide a good description of the reaction coordinate and is able to clearly distinguish between different free energy minima, kinetic basins and binding pathways.

An open source plug-in implementation of metadynamics (PLUMED)[25] working with many widely-used MD codes, as well as native implementations for NAMD, Desmond and other codes are available. In drug discovery, it can provide the location of cryptic pockets, an estimate of the binding free energy, including conformational changes of the target, as well as quantitative information on the metastable minima and transition states, which can be used to optimize the underlying ligand binding and unbinding kinetics.

In this Account, we review recent applications of metadynamics and compare it to other enhanced sampling approaches in drug-discovery-related endeavors, providing some perspective on the advantages, disadvantages, and practical applicability of these methodologies in academic and industrial settings.

2. Protein-ligand binding and unbinding

MD simulations have contributed in major ways to our understanding of molecular recognition and consequently to drug design. HIV-integrase inhibitors are a classic example. MD simulations revealed a previously unknown "trench", which was invisible to X-ray crystallography and traditional docking methods.[26] This suggested a way of designing stronger inhibitors binding to both the catalytic site and the newly discovered trench. A few notable examples in the literature deal with the long time span required for a binding process to occur spontaneously, by running MD simulations lasting a few microseconds.[27–29] However, the timescale problem still hampers an extensive application of fully atomistic MD in drug design. In the context of accelerated sampling methods, metadynamics-based approaches have been successfully applied to drug design, particularly when major conformational motions play a role prior to and upon a drug binding to a target.

The first application of metadynamics to ligand-binding was carried out by Gervasio et al., who studied four different complexes.[30] The correct geometry corresponded to the deepest free energy minimum in all the cases, demonstrating the ability of the method to predict the correct binding mode. In the case of the trypsin-benzamidine complex, the algorithm was extensively compared to 2D umbrella sampling and was found to provide the free energy reconstruction in less time. The experimental binding energy was correctly predicted. In the case of the CDK2/staurosporine complex, albeit the correct docking pose was predicted, the computed binding free energy was significantly overestimated (20 kcal/mol vs. 11 kcal/mol). This was due to: a fast filling time (a total of only 8 ns), an inaccurate ligand force-field, the lack of a correction for the standard volume of the unbound state and sub-optimal CVs. Nowadays, thanks to the significant increase in available computational resources and the introduction of the well-tempered variant of metadynamics, the bias is added at a much slower rate and the FES is usually

filled in hundreds if not thousands of ns. The ligand force-field still remains a major concern. While there has been a significant progress in the accuracy of protein force-fields,[31-32] generally-available ligand force-fields lag behind, and despite recent progress,[33] careful re-parameterization of ligands based on ab-initio calculations is strongly advised. As for the CVs, major progress has been achieved in the design of optimal coordinates. In Gervasio's et al. paper[30], the choice of geometry-based CVs (distance and angle) was guided by three requirements: i) a general applicability in molecular recognition, ii) the need to distinguish all the minima along the reaction coordinate, and iii) the need to keep the number of CVs low. The latter requirement is due to the fact that the time needed to "fill the minima" (see Figure 1) in principle increases exponentially with the number of the CVs.

The need to choose a set of optimal CVs is the main drawback of metadynamics and other CV-based methods. In a subsequent study, Masetti et al. systematically explored different combinations of CVs and faster approaches to converge the free energy estimation.[34] A combination of docking, cluster analysis,[35] and metadynamics limited to exploring the internal cavity up to the transition state (*coarse-metadynamics*) improved scoring in conventional structure-based endeavors. To date *coarse-metadynamics* together with optimal CVs is still one of the best ways to use metadynamics to rank different ligands binding to the same target.

However, to compute the relative free energy of a set of similar compounds (having a similar binding pose), end-point methods, i.e. methods that compute the difference of free energy between the bound and the unbound state, such as thermodynamic integration (TI) are in most cases more efficient than methods that compute the free energy along a physical association pathway.

Another early application was Branduardi et al.'s study of the binding mechanism of the tetramethylammonium (TMA) ion by human acetylcholinesterase (AChE).[36] The AChE peripheral anionic site (PAS), located at the entrance of the gorge, is solvent-exposed, therefore plain MD simulations are able to identify stable docking conformations.[37] However, enhanced sampling approaches are needed to explore binding into the deep internal gorge.[38] Metadynamics was run with two CVs, the distance between ligand and the binding pocket, along with a new CV describing the cation–π interaction, leading to a well-converged free energy reconstruction. Cation–π interactions were shown to play a fundamental role in both the initial recognition between TMA and the PAS, and the gorge binding mechanism. Notably, the cation–π CV must be included to obtain a converged FES.

Beside protein-ligand systems, metadynamics has also been applied to study the association mechanism of molecules binding to DNA and RNA, e.g. the interaction of ligands with DNA G-quadruplexes,[39] and the unbinding of the anticancer drug distamycin from the DNA minor groove.[40]

Other methods, such as steered MD, have also been successfully used to characterize binding poses and to drive the selection of biologically active compounds. Colizzi et al. have exploited docking and steered MD to discern between active and inactive antimalarial ligands.[41] By modeling the force that is required to pulling inhibitors out of their binding pocket complex, it was possible to separate strong binders from weak ones. Strongly bound inhibitors gave profiles with higher peak forces than weakly bound inhibitors, which gave a flat profile (Figure 2). Compared to metadynamics, in steered MD the choice of the CV is less crucial, allowing a quick exploration of many unbinding events. However, it is less efficient in computing fully converged free energy profiles.

3. Reconstructing the free energy landscapes

As discussed above, selecting a limited number of optimal CVs is a crucial step in metadynamics simulations to obtain a converged free energy landscape. The most commonly used CVs in drug-binding simulations are geometry-based and involve a combination of distances, contacts, angles, and dihedrals. In complex cases, several attempts might be needed to find a proper combination of variables (sub-optimal set of variables are easily recognized as the reconstructed FES shows hysteresis as a function of time). The selection of effective CVs in ligand-(un)binding is further complicated by the need to fully explore (fill) the large conformational spaces available to the solvated ligand in the bulk, once unbound from its target.

Metadynamics is not efficient in exploring relatively flat, diffusive energy landscapes and its use to explore unbound and weakly-bound states is not recommended. In such cases, other methods such as swarms of free MD trajectories and Markov-state Models[28] or some metadynamics variants such as Recoinnascence metadynamics or bias-exchange are more effective. When the binding area is known, this issue can be solved by the use of restraints on the explored area or optimal CVs such as the Path Collective Variables (PCVs).[42]

In analogy to similar strategies used with umbrella-sampling,[43] various restraining potentials have been used. An interesting choice is to use a reverse *funnel-shaped* potential that limits the space available to the ligand once it has undocked.[44] *Funnel* metadynamics was used to revisit some of the systems previously studied with geometrical CVs or PCVs, including the trypsin-benzamidine complex and the binding of the drug SC-558 to cyclooxygenase 2 (COX-2, see below). Thanks to the restraining potential, the unbound state can be thoroughly sampled leading to a faster convergence of the energy landscape.

A perhaps more elegant solution is the use of PCVs that describe the position of the system in the configuration space relative to an optimal exit path, in terms of progression along the path and distance from the path itself. PCVs are able to describe complex association pathways including target conformational changes and to restrain the sampling of the unbound state, improving convergence considerably. In the last few years, the PCV approach has been successfully used to study complex conformational changes and drug-target binding.[45,46] Fidelak et al.[47] used PCVs to study the binding of a congeneric series of ligands to CDK2. Identifying the correct order of activity of the compounds was non-trivial, due to the lack of interaction between the protein and the aryl-pyrimidine substituents differentiating the homologous inhibitors. Using well-tempered metadynamics and PCV, the authors converged a full association free energy profile in remarkable agreement with the experimental binding energies. They also fully characterized the association mechanism. In a subsequent work, Saladino et al.[48] studied a well-known series of p38 inhibitors, frequently used to assess the performances of scoring algorithms, whose activity cannot be easily guessed. The authors moved a step further towards making metadynamics suitable for a pharmaceutical workflow and used a semi-automatic setup. With a similar intent, Mason and coworkers[49] at Heptares Therapeutics designed the *MetaScore* routine, a fast metadynamics-based protocol, used to predict the effect of different mutations on the binding of antagonists to the adenosine A_{2A} receptor. The *MetaScore* scoring function is based on a two-step semi-automatic protocol that uses several adiabatic biased MD simulations to obtain an initial association path and construct PCVs, followed by short well-tempered metadynamics runs to calculate the full binding free energy profile. Using this approach, the authors predicted the correct binding energy changes resulting from active site mutations. PCV were also used by

Provasi et al.[50] to access the free energy of binding of several ligands to another G-protein-coupled receptor, the β2-adrenergic receptor.

When the binding is accompanied by large-scale conformational changes, the choice of CVs can be daunting. In such cases, Parallel Tempering Metadynamics (PT-MetaD)[51], an approach that couples metadynamics with a replica exchange algorithms, can enhance the sampling of hidden (slow) degrees of freedom and greatly improve convergence. It has been used to compute the FES associated with the inactive-to-active conformational changes of the cancer target epidermal growth factor receptor.[46] PT-metaD, albeit being a very expensive approach, is the method of choice when large-scale conformational changes are involved and a converged FES is required. Finally, the number of CVs is not an issue in Bias Exchange Metadynamics (BEMD)[52] and Reconnaissance Metadynamics.[53] The latter uses a self-learning algorithm that clusters the visited conformations on-the-fly and bias along a one-dimensional clustering-based CV. These approaches have been effectively used to explore very complex molecular recognition events and are very beneficial when little information is known about the association mechanism and the location of the binding cavity. Both methods, however, require extra efforts to reconstruct the free energy landscapes, and thus are best used as a means to explore different binding mechanisms.

4. Pharmaceutical applications

Metadynamics was recently used to understand the differences in the binding mechanisms of a non-steroidal anti-inflammatory ligand (SC-558) to COX-1 and COX-2 isoforms.[54] Selective inhibition of COX isoforms has been actively sought, leading to new generations of widely used COX-2 selective drugs. Recently, significant effort has been made to understand the selectivity

and the differences in the binding mechanism of inhibitors to COX-1 and COX-2. In this study, metadynamics and the PCV method[42] were used to describe the (un)binding of a ligand and the associated conformational change of three α-helices that form a narrow gate through which the ligand exits from the protein. In the case of the COX-2 isoform, two separate and almost equally favorable minima in the FES are found (Figure 3), while only one deep minimum was found in COX-1. The existence of two binding modes in COX-2 only explain both the selectivity of the ligand and, interestingly, the increased residence time of the ligand in COX-2 relative to COX-1.

A similar approach was used to study the aspartate uptake and internal release mechanism through a membrane protein, namely the amino acid transporter from *Pyrococcus horikoshii*.[55] The transport of the substrate is assisted by the cotransport of Na^+ ions and is mediated by the concerted motion of the transmembrane domain and the two hairpins (HP1 and HP2) that regulate access to either side of the membrane. In addition, it is known that this transporter may assume at least four different conformational states, which further challenges the study of the substrate translocation across the channel. Grazioso et al. used PCV to enhance the highly cooperative transition upon substrate binding and transport across the membrane.[55] The authors thus described the complete opening mechanism involving a large-scale motion of HP2, which allowed substrate uptake, with a similar role played by HP1 in the inward release mechanism (Figure 4).

As a final case study, PT-metaD with PCVs and BEMD have been used to study the mode of action of a novel inhibitor of the fibroblast growth factor receptor (FGFR).[56] FGFRs are tyrosine kinase receptors whose signaling pathway is involved in cell growth, proliferation, and survival. Due to their key role in angiogenesis, FGFR deregulation can lead to the development of malignancies, so this family of receptors is a well-studied target for cancer treatment. The first

extracellular FGFR inhibitor, SSR128129E (hereafter referred to as SSR), was discovered by the same authors through a high-throughput screening campaign. However, several experimental techniques, ranging from NMR to crystallography, failed to identify the binding site of SSR, due to the inherent flexibility of the extracellular domain of FGFR, comprising three immunoglobulin-like domains. Using BEMD, the authors observed the elongation of a small α-helix in the D3 domain in the presence of the drug, leading to the opening of a pocket not present in the apo structure of FGFR. Metadynamics simulations with PCV correctly predicted the dissociation constant of the compound and of several lower potency analogues, validating the in-silico-derived mode of action. As a further proof, the authors used the observed binding mode as a template for rationally designing new inhibitors, identifying 16 new compounds with varying degrees of activity. Once again, the computationally predicted free energy was in remarkable agreement with experimental results.

5. Binding and unbinding kinetics estimation

Obtaining reliable estimates of the kinetics constants associated with ligand binding is an ambitious goal of modern computational drug discovery, with significant potential benefits in the lead discovery and the lead optimization steps of the drug discovery process.[57] The possibility of reconstructing and estimating the binding free energy along physical association pathways provides quantitative information on the metastable minima and transition states. These can then be used to optimize the underlying ligand binding kinetics to achieve therapeutically safe and differentiated responses. Researchers increasingly recognize the importance of ligand target residence time in fine-tuning in vivo efficacy and toxicology.[58] Knowledge of the metastable

states and the residues involved in a ligand's access to the binding cavity could lead to the design of novel drugs that are more potent and selective for therapeutically relevant targets.

Metadynamics in the bias-exchange variant was used by Pietrucci et al. to investigate the key events in the binding of the p2-NC peptide to the Human Immunodeficiency Virus type-1 Protease (HIV1-1 PR).[59] The free energy was computed as a function of seven CVs, in order to characterize the physical interactions that might play a role in stabilizing the complex and the transition states. From the BE-META trajectories, a thermodynamic and kinetic model of the binding process was constructed based on the weighted-histogram approach (Figure 5). Remarkably, mutation of residues along the access region identified by BE caused resistance to some FDA-approved peptidomimetic drugs, providing an experimental validation. Individual water molecules at the interface between ligand and enzyme played a pivotal role throughout the binding process and a CV describing "interfacial water" was needed to discriminate different intermediate states.

In absence of large barriers to binding/unbinding, Markov state models (MSM)[60] might be more efficient than metadynamics-based algorithms in reconstructing the binding kinetics. A collection of 495 100-ns-long trajectories was used to reconstruct the full binding kinetics of the trypsin/benzamidine complex, revealing the existence of long-lived binding intermediates.

Recently, Tiwary et al.[61] adapted to metadynamics the concept of "acceleration factor" introduced by Grubmüller[17] for Conformational Flooding. Provided that no bias is deposited in the transition state region, this approach can predict the rate of the transition, with minimal additional computational cost. Albeit the authors applied their approach to a toy model, preliminary data indicates that it can be applied to predict the binding kinetics in simple (2-state) systems. A more general (and more computationally expensive) approach was developed by

Juraszek and coworkers.[62] By combining PT-metaD-derived free energies with a transition path sampling (TPS) approach they devised a robust method that can be applied to complex ligand-binding events with multiple intermediate states and diffusive barriers. Indeed, the *Transition State-Partial Path Transition Interface Sampling* was explicitly devised to overcome the difficulties of available methods in predicting the kinetics rates of drug binding phenomena. Introducing a new recursive expression for the transmission coefficient, the authors designed a robust approach with very general applicability and high accuracy.

6. Perspectives

In Table 1 and Figure 6, we summarize the suggested metadynamics-based approach to use in drug discovery. In particular we address the cases of cryptic binding sites, scoring different ligands binding to the same target, full free energy reconstruction with or without large conformational changes and binding kinetics. As we have detailed in the preceding sections, metadynamics-based approaches are not always the most effective choice. In Figure 7, we report on the major strengths and weaknesses of these methods. One of major drawbacks of metadynamics is still related to the choice of CVs, which can compromise a proper reconstruction of the FES associated to the process under investigation. In this scenario, some recent and on-going developments, particularly suited to drug discovery, are the use of non-linear dimensionality reduction techniques to automatically find the optimal CVs, new CVs to find and score cryptic binding sites and, going beyond the current methods, a new variational approach to construct an optimal bias potential and reconstruct the FES.[63] All these evolutions will contribute to a further expansion of metadynamics approaches to drug discovery, widening

the plethora of computational tools utilized to estimate binding and unbinding free energy and kinetics, and eventually accelerating the discovery of novel bioactive compounds.

Figure Legends

Figure 1. Pictorial representation of the way the metadynamics algorithm fills the free energy landscapes.

Figure 2. The force applied to the ligand by the spring is calculated as a function of time, generating a force profile that reflects the strength of protein–ligand binding as reported in Colizzi et al.[41]

Figure 3. Free energy surfaces associated with the binding of the SC558 inhibitor to COX-1 and COX-2 as a function of the distance between the ligand and the cavity and the relative orientation of SC558.

Figure 4. Schematic representation of the free energy landscape of substrate uptake and release by the glutamate transporter. Adapted from Grazioso et al.[55]

Figure 5. Kinetic model of the substrate binding to HIV-1 Protease. Adapted from Pietrucci et al.[59]

Figure 6. Schematic representation of the use of metadynamics in drug design, from scoring and binding pose prediction to thermodynamics and kinetics profiles.

Figure 7. SWOT (Strengths, Weakness, Opportunities and Threats) analysis for metadynamics.

BIOGRAPHY INFORMATION

Andrea Cavalli is Professor in Medicinal Chemistry at the University of Bologna (Italy) and Director of D3 Computation at Italian Institute of Technology. His research interests are in the field of computational drug discovery. In particular, he has developed and applied computational approaches to accelerate the discovery of drug-like compounds in several therapeutic areas, including cancer, neurodegenerative diseases, and neglected tropical diseases. Prof. Cavalli is the author of more than 150 publications in high-ranked journals, he is co-inventor in 8 PCT international patents, and he has delivered more than 50 invited lectures and seminars. In 2003, he was awarded the 'Farmindustria Prize for Pharmaceutical Research'.

Andrea Spitaleri is Principal Investigator of AIRC and researcher of Istituto Italiano di Tecnologia (Italy). His main field of research is the study of protein-protein and protein-ligand interactions by computational methods. He is author of many articles in high-impact life science journals.

Giorgio Saladino is a research assistant in the Institute of Structural and Molecular Biology at University College London. His research activity is focused on the study of biomolecules by computational methods. He is author of 11 papers in high-impact journals, including Cancer Cell, Angewandte Chemie, and Nature Chemistry.

Francesco L. Gervasio holds the Chair in Biomolecular Modeling and is full Professor in Chemistry and Structural and Molecular Biology at University College London. His research interests are in the field of computational chemistry, biophysics, and structural biology. He has co-developed several successful Metadynamics derivatives, including PT-metaD and PCV and applied computational approaches to study ligand-target association and target flexibility. Prof.

Gervasio is the author of more than 70 publications in high-ranked journals, including Cancer Cell, PNAS, Angew. Chem., and JACS. He is a member of the management group of the CCP-BioSim consortium and has organized several international conferences.

Corresponding Authors

*To whom correspondence should be addressed: andrea.cavalli@unibo.it; 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.

ACKNOWLEDGMENTS

We acknowledge University of Bologna, Italian Istitute of Tecnology, University College London, and PRIN2012 project for the financial support. We thank Grace Fox for editing and proofreading the manuscript.

ABBREVIATIONS

MD, molecular dynamics, CV, collective variable; FES, free energy surface; PCV, path collective variables; PT-MetaD, parallel tempering metadynamics; BEMD, bias-exchange molecular dynamics; MSM, Markov state model.

REFERENCES

(1) Macarron, R. Critical Review of the Role of HTS in Drug Discovery. *Drug Discov. Today* **2006**, *11*, 277–279.

(2) Csermely, P.; Palotai, R.; Nussinov, R. Induced Fit, Conformational Selection and Independent Dynamic Segments: An Extended View of Binding Events. *Trends Biochem. Sci.* **2010**, *35*, 539–546.

(3) McCammon, J. A. Theory of Biomolecular Recognition. *Curr. Opin. Struct. Biol.* **1998**, *8*, 245–249.

(4) Shaw, D. E.; Chao, J. C.; Eastwood, M. P.; Gagliardo, J.; Grossman, J. P.; Ho, C. R.; Lerardi, D. J.; Kolossváry, I.; Klepeis, J. L.; Layman, T.; McLeavey, C.; Deneroff, M. M.; Moraes, M. A.; Mueller, R.; Priest, E. C.; Shan, Y.; Spengler, J.; Theobald, M.; Towles, B.; Wang, S. C.; Dror, R. O.; Kuskin, J. S.; Larson, R. H.; Salmon, J. K.; Young, C.; Batson, B.; Bowers, K. J. Anton, a Special-Purpose Machine for Molecular Dynamics Simulation. *Commun. ACM* **2008**, *51*, 91–91.

(5) Harvey, M. J.; Giupponi, G.; Fabritiis, G. D. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *J. Chem. Theory Comput.* **2009**, *5*, 1632–1639.

(6) Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845–854.

(7) Stone, J. E.; Phillips, J. C.; Freddolino, P. L.; Hardy, D. J.; Trabuco, L. G.; Schulten, K. Accelerating Molecular Modeling Applications with Graphics Processors. *J. Comput. Chem.* **2007**, *28*, 2618–2640.

(8) Mitsutake, A.; Mori, Y.; Okamoto, Y. Enhanced Sampling Algorithms. *Methods Mol. Biol.* **2013**, *924*, 153–195.

(9) Patey, G. N.; Valleau, J. P. Monte-Carlo Method for Obtaining Interionic Potential of Mean Force in Ionic Solution. *J. Chem. Phys.* **1975**, *63*, 2334–2339.

(10) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.

(11) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562–12566.

(12) Isralewitz, B.; Gao, M.; Schulten, K. Steered Molecular Dynamics and Mechanical Functions of Proteins. *Curr. Opin. Struct. Biol.* **2001**, *11*, 224–230.

(13) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys.* **2004**, *120*, 11919.

(14) Faradjian, A. K.; Elber, R. Computing Time Scales from Reaction Coordinates by Milestoning. *J. Chem. Phys.* **2004**, *120*, 10880.

(15) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. Transition Path Sampling: Throwing Ropes over Rough Mountain Passes, in the Dark. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291–318.

(16) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, 020603.

(17) Grubmller, H. Predicting Slow Structural Transitions in Macromolecular Systems: Conformational Flooding. *Phys. Rev. E* **1995**, *52*, 2893–2906.

(18) Marsili, S.; Barducci, A.; Chelli, R.; Procacci, P.; Schettino, V. Self-Healing Umbrella Sampling: A Non-Equilibrium Approach for Quantitative Free Energy Calculations. *J. Phys. Chem. B* **2006**, *110*, 14011–14013.

(19) Huber, T.; Torda, A. E.; van Gunsteren, W. F. Local Elevation: A Method for Improving the Searching Properties of Molecular Dynamics Simulation. *J. Comput. Aided Mol. Des.* **1994**, *8*, 695–708.

(20) Cvijovicacute, D.; Klinowski, J. Taboo Search: An Approach to the Multiple Minima Problem. *Science* **1995**, *267*, 664–666.

(21) Darve, E.; Rodríguez-Gómez, D.; Pohorille, A. Adaptive Biasing Force Method for Scalar and Vector Free Energy Calculations. *J. Chem. Phys.* **2008**, *128*, 144120.

(22) Wojtas-Niziurski, W.; Meng, Y.; Roux, B.; Bernèche, S. Self-Learning Adaptive Umbrella Sampling Method for the Determination of Free Energy Landscapes in Multiple Dimensions. *J. Chem. Theory Comput.* **2013**, *9*, 1885–1895.

(23) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13*, 1011–1021.

(24) Dama, J. F.; Parrinello, M.; Voth, G. A. Well-Tempered Metadynamics Converges Asymptotically. *Phys. Rev. Lett.* **2014**, *112*, 240602.

(25) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185*, 604–613.

(26) Schames, J. R.; Henchman, R. H.; Siegel, J. S.; Sotriffer, C. A.; Ni, H.; McCammon, J. A. Discovery of a Novel Binding Trench in HIV Integrase. *J. Med. Chem.* **2004**, *47*, 1879–1881.

(27) Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E. How Does a Drug Molecule Find Its Target Binding Site? *J. Am. Chem. Soc.* **2011**, *133*, 9181–9183.

(28) Buch, I.; Giorgino, T.; De Fabritiis, G. Complete Reconstruction of an Enzyme-Inhibitor Binding Process by Molecular Dynamics Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 10184–10189.

(29) Klepeis, J. L.; Lindorff-Larsen, K.; Dror, R. O.; Shaw, D. E. Long-Timescale Molecular Dynamics Simulations of Protein Structure and Function. *Curr. Opin. Struct. Biol.* **2009**, *19*, 120–127.

(30) Gervasio, F. L.; Laio, A.; Parrinello, M. Flexible Docking in Solution Using Metadynamics. *J. Am. Chem. Soc.* **2005**, *127*, 2600–2607.

(31) Lindorff-Larsen K.; Piana S; Palmo K.; Maragakis P.; Klepeis J. L.; Dror R. O.; Shaw D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950-1958.

(32) Cerutti D. S.; Swope W. C.; Rice J. E.; Case D. A. ff14ipq: A self-consistent force field for condensed-phase simulations of proteins. *J. Chem. Theory Comput.,* **2014**, 10, 4515-4534.

(33) Shivakumar D; Harder E.; Damm W.; Friesner R. A.; Sherman W. Improving the Prediction of Absolute Solvation Free Energies Using the Next Generation OPLS Force Field. *J. Chem. Theory Comput.*, **2012**, 8, 2553-2558.

(34) Masetti, M.; Cavalli, A.; Recanatini, M.; Gervasio, F. L. Exploring Complex Protein-Ligand Recognition Mechanisms with Coarse Metadynamics. *J. Phys. Chem. B* **2009**, *113*, 4807–4816.

(35) Bottegoni, G.; Cavalli, A.; Recanatini, M. A Comparative Study on the Application of Hierarchical-Agglomerative Clustering Approaches to Organize Outputs of Reiterated Docking Runs. *J. Chem. Inf. Model.* **2006**, *46*, 852–862.

(36) Branduardi, D.; Gervasio, F. L.; Cavalli, A.; Recanatini, M.; Parrinello, M. The Role of the Peripheral Anionic Site and Cation–π Interactions in the Ligand Penetration of the Human AChE Gorge. *J. Am. Chem. Soc.* **2005**, *127*, 9147–9155.

(37) Cavalli, A.; Bottegoni, G.; Raco, C.; De Vivo, M.; Recanatini, M. A Computational Study of the Binding of Propidium to the Peripheral Anionic Site of Human Acetylcholinesterase. *J. Med. Chem.* **2004**, *47*, 3991–3999.

(38) Bui, J. M.; Henchman, R. H.; McCammon, J. A. The Dynamics of Ligand Barrier Crossing inside the Acetylcholinesterase Gorge. *Biophys. J.* **2003**, *85*, 2267–2272.

(39) Di Leva F. S.; Novellino E.; Cavalli A.; Parrinello M.; Limongelli V. Mechanistic insight into ligand binding to G-quadruplex DNA. *Nucl. Acids Res **2014**,* 42, 5447-5455

(40) Vargiu A. V.; Ruggerone P.; Magistrato A.; Carloni P. Dissociation of minor groove binders from DNA: insights from metadynamics simulations. *Nucl. Acids Res.* **2008,** 36, 5910-5921

(41) Colizzi, F.; Perozzo, R.; Scapozza, L.; Recanatini, M.; Cavalli, A. Single-Molecule Pulling Simulations Can Discern Active from Inactive Enzyme Inhibitors. *J. Am. Chem. Soc.* **2010**, *132*, 7361–7371.

(42) Branduardi, D.; Gervasio, F. L.; Parrinello, M. From A to B in Free Energy Space. *J. Chem. Phys.* **2007**, *126*, 054103.

(43) Deng, Y.; Roux, B. Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. *J. Phys. Chem. B* **2009**, *113*, 2234–2246.

(44) Limongelli, V.; Bonomi, M.; Parrinello, M. Funnel Metadynamics as Accurate Binding Free-Energy Method. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 6358–6363.

(45) Berteotti, A.; Cavalli, A.; Branduardi, D.; Gervasio, F. L.; Recanatini, M.; Parrinello, M. Protein Conformational Transitions: The Closure Mechanism of a Kinase Explored by Atomistic Simulations. *J. Am. Chem. Soc.* **2009**, *131*, 244–250.

(46) Sutto, L.; Gervasio, F. L. Effects of Oncogenic Mutations on the Conformational Free-Energy Landscape of EGFR Kinase. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 10616–10621.

(47) Fidelak, J.; Juraszek, J.; Branduardi, D.; Bianciotto, M.; Gervasio, F. L. Free-Energy-Based Methods for Binding Profile Determination in a Congeneric Series of CDK2 Inhibitors. *J. Phys. Chem. B* **2010**, *114*, 9516–9524.

(48) Saladino, G.; Gauthier, L.; Bianciotto, M.; Gervasio, F. L. Assessing the Performance of Metadynamics and Path Variables in Predicting the Binding Free Energies of p38 Inhibitors. *J. Chem. Theory Comput.* **2012**, *8*, 1165–1170.

(49) Mason, J.; Bortolato, A.; Weiss, D.; Deflorian, F.; Tehan, B.; Marshall, F. High End GPCR Design: Crafted Ligand Design and Druggability Analysis Using Protein Structure, Lipophilic Hotspots and Explicit Water Networks. *In Silico Pharmacol* **2013**, *1*, 23.

(50) Provasi, D.; Artacho, M. C.; Negri, A.; Mobarec, J. C.; Filizola, M. Ligand-Induced Modulation of the Free-Energy Landscape of G Protein-Coupled Receptors Explored by Adaptive Biasing Techniques. *PLoS Comput. Biol.* **2011**, *7*, e1002193.

(51) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. Free-Energy Landscape for B Hairpin Folding from Combined Parallel Tempering and Metadynamics. *J. Am. Chem. Soc.* **2006**, *128*, 13435–13441.

(52) Piana, S.; Laio, A. A Bias-Exchange Approach to Protein Folding. *J. Phys. Chem. B* **2007**, *111*, 4553–4559.

(53) Söderhjelm, P.; Tribello, G. A.; Parrinello, M. Locating Binding Poses in Protein-Ligand Systems Using Reconnaissance Metadynamics. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 5170–5175.

(54) Limongelli, V.; Bonomi, M.; Marinelli, L.; Gervasio, F. L.; Cavalli, A.; Novellino, E.; Parrinello, M. Molecular Basis of Cyclooxygenase Enzymes (COXs) Selective Inhibition. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 5411–5416.

(55) Grazioso, G.; Limongelli, V.; Branduardi, D.; Novellino, E.; De Micheli, C.; Cavalli, A.; Parrinello, M. Investigating the Mechanism of Substrate Uptake and Release in the Glutamate Transporter Homologue Glt $_{Ph}$ through Metadynamics Simulations. *J. Am. Chem. Soc.* **2012**, *134*, 453–463.

(56) 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.

(57) Copeland, R. A.; Pompliano, D. L.; Meek, T. D. Drug-Target Residence Time and Its Implications for Lead Optimization. *Nat. Rev. Drug Discov.* **2006**, *5*, 730–739.

(58) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* **2010**, *330*, 341–346.

(59) Pietrucci, F.; Marinelli, F.; Carloni, P.; Laio, A. Substrate Binding Mechanism of HIV-1 Protease from Explicit-Solvent Atomistic Simulations. *J. Am. Chem. Soc.* **2009**, *131*, 11811–11818.

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

(61) Tiwary, P.; Parrinello, M. From Metadynamics to Dynamics. *Phys. Rev. Lett.* **2013**, *111*, 230602.

(62) Juraszek, J.; Saladino, G.; van Erp, T.; Gervasio, F. Efficient Numerical Reconstruction of Protein Folding Kinetics with Partial Path Sampling and Pathlike Variables. *Phys. Rev. Lett.* **2013**, *110*, 108106.

(63) Valsson O,; Parrinello M. Variational Approach to Enhanced Sampling and Free Energy Calculations. *Phys. Rev. Lett.*, **2014**, 113, 090601.

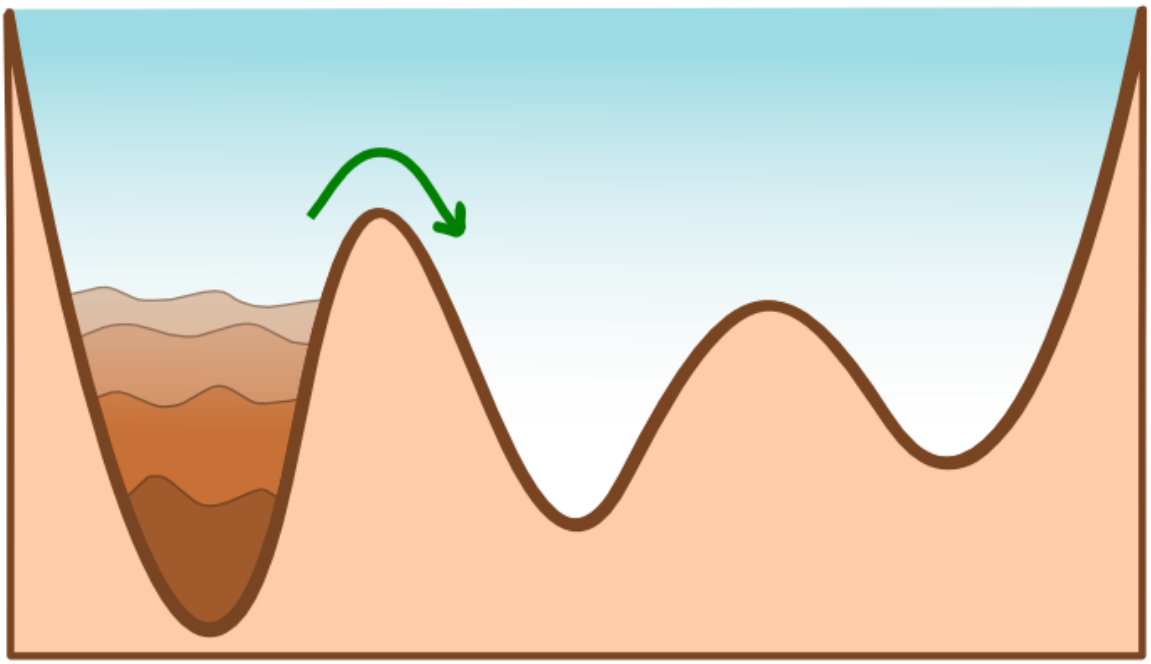

Figure 1

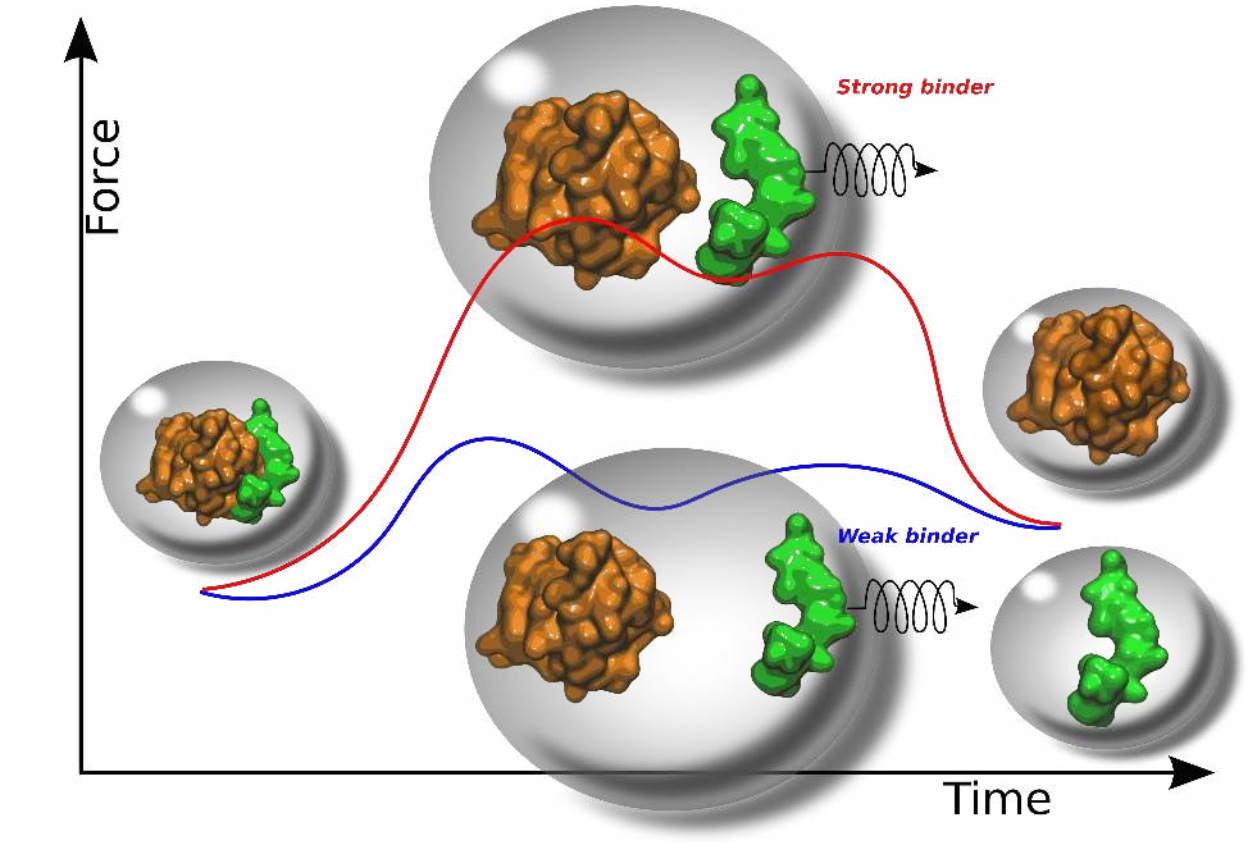

Figure 2

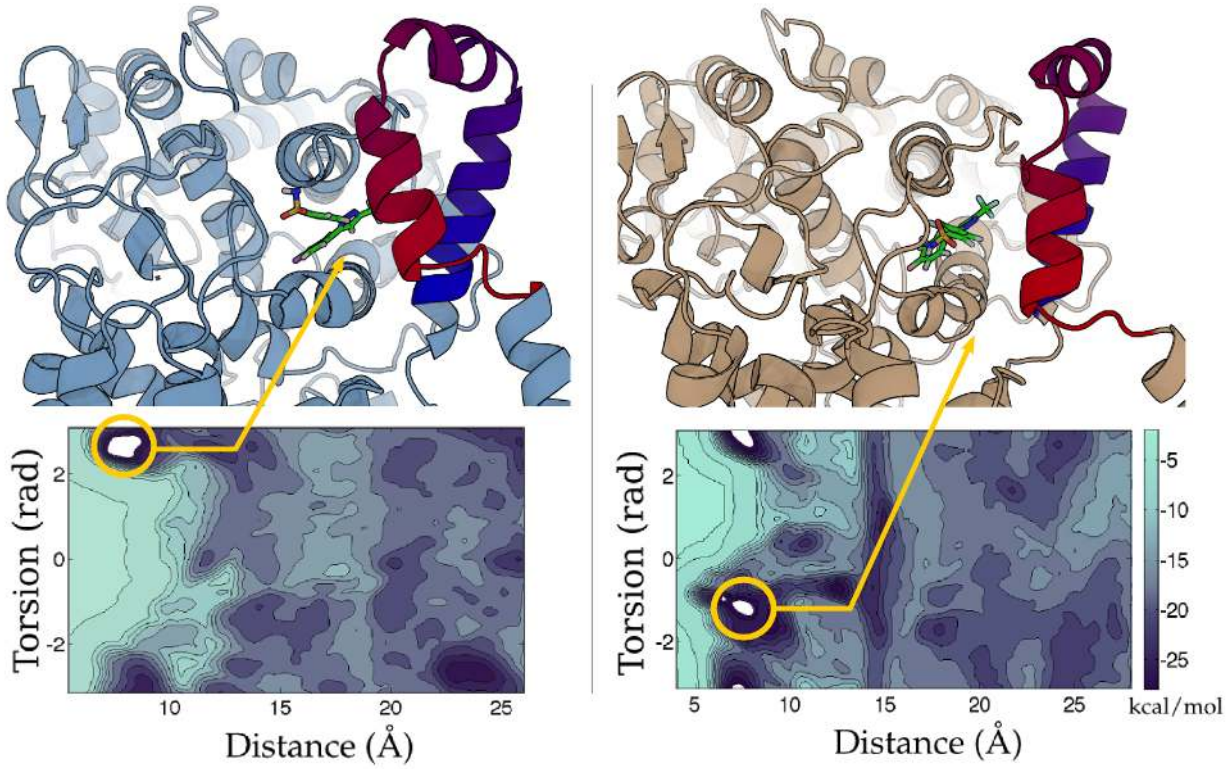
Torsion (rad)
2
0
-2
10
15
20
25
Distance (Å)
Torsion (rad)
2
0
-2
5
10
15
20
25
Distance (Å)
-5
-10
-15
-20
-25
kcal/mol

Figure 3

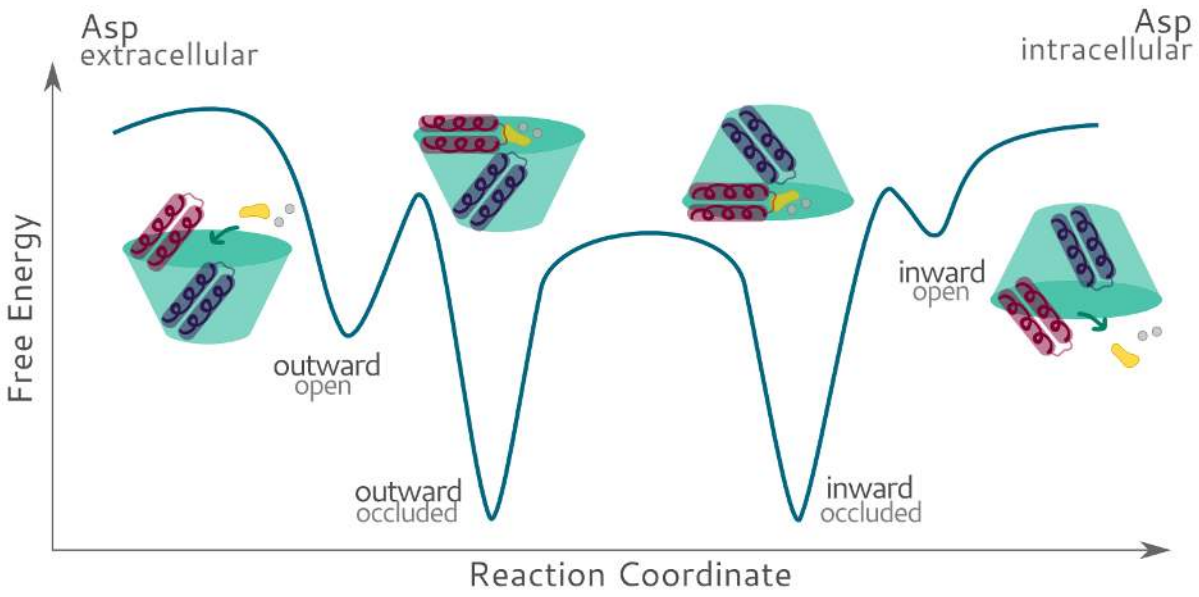
Asp
extracellular
Asp
intracellular
Free Energy
outward
open
outward
occluded
inward
occluded
inward
open
Reaction Coordinate

Figure 4

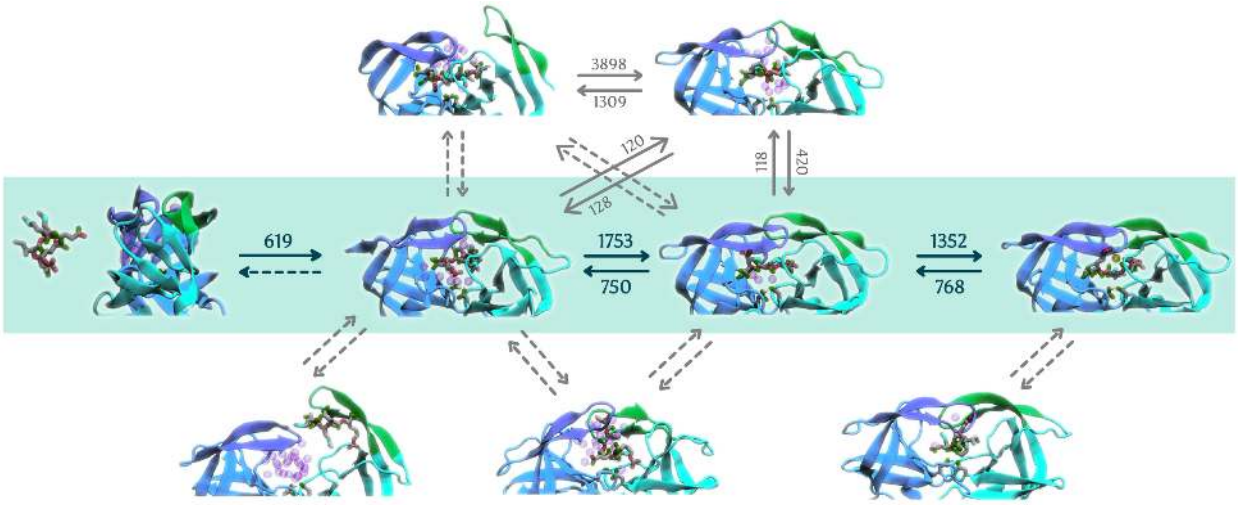

Figure 5

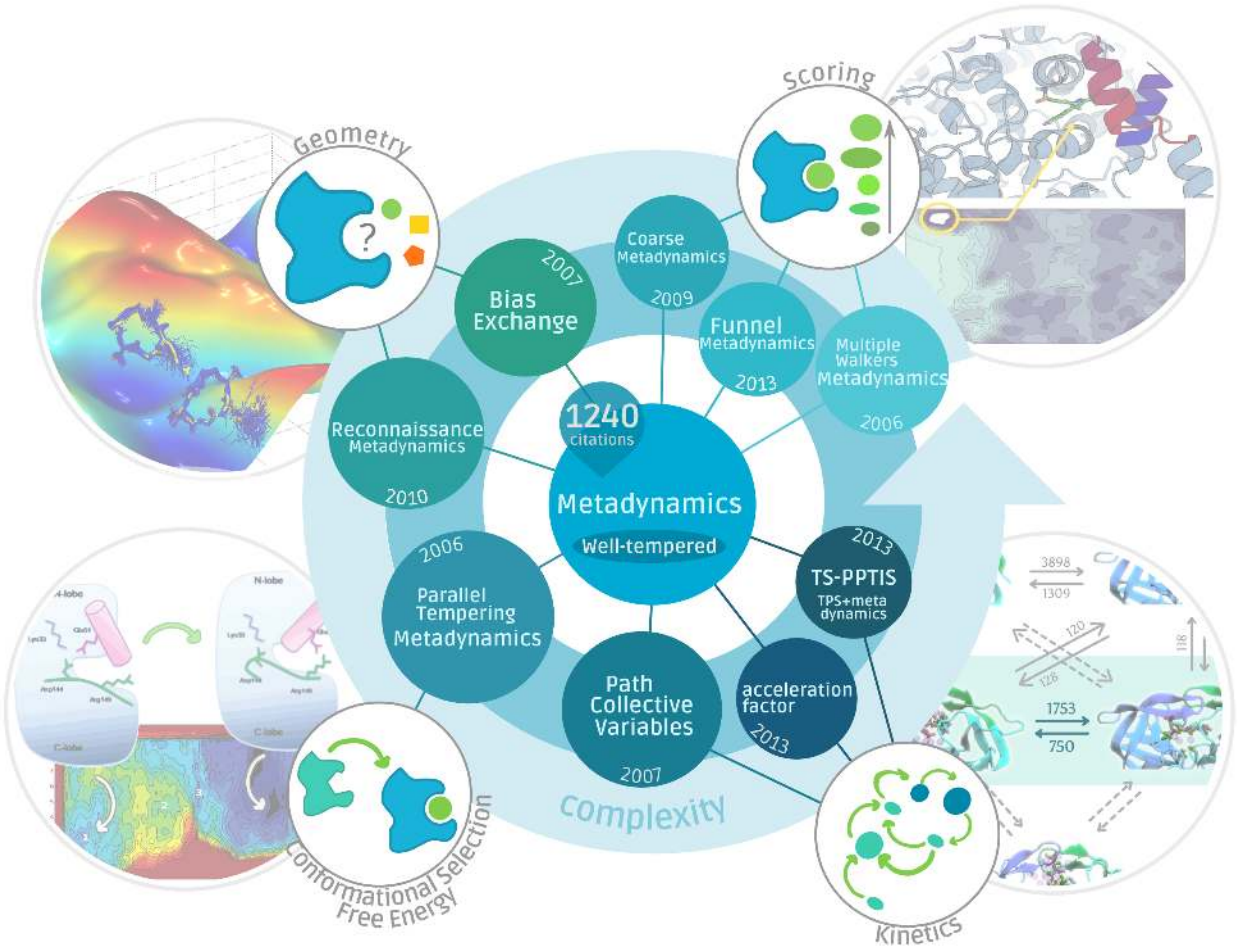
Geometry
scoring
Coarse Metadynamics
2009
2007
Bias Exchange
Funnel Metadynamics
2013
Multiple Walkers Metadynamics
2006
Reconnaissance Metadynamics
2010
1240
citations
Metadynamics
Well-tempered
2006
Parallel Tempering Metadynamics
2013
TS-PPTIS
TPS+meta dynamics
Path Collective Variables
2007
acceleration factor
2013
complexity
Conformational Selection
Free Energy
Kinetics
3898
1309
120
126
118
1753
750

Figure 6


Strenghts
reconstruction of free energy landscapes in more than one dimension
general and easy to implement
available for multiple MD engines through PLUMED
Weaknesses
the efficiency depends on the choice of CVs
reaching convergence can be expensive when using more than 3 CVs
informations on dynamics (and possibly mechanism) might be lost
Opportunities
it can be easily combined with other approaches for better results (e.g. Parallel Tempering or Transition Path Sampling)
new algorithms and approaches are constantly released
several recipes to recover dynamics are being developed
Threats
a converged free energy if projected on suboptimal coordinates can lead to misinterpretation
calculations can become highly expensive (especially with multiple replicas)

Figure 7

Table 1. Summary of the suggested approaches for different drug discovery applications.

Increasing Complexity ↓		Metadynamics Flavour	Other Methods
	Scoring	Coarse Metadynamics Funnel Metadynamics Multiple Walkers Metadynamics	Thermodynamic Integration (TI) for similar ligands with minor receptor changes
	Geometry	Bias Exchange Reconnaissance Metadynamics	
	Conformational Change & Free Energy	Parallel Tempering Metadynamics	Umbrella Sampling (US) re-sample metadynamics-flattened surfaces
	Kinetics	Path Collective Variables Accelleration Factor TS-PPTIS	Markos State Model (MSM) for flat/diffusive free energy surfaces

Conspectus Graphic