Emerging Computational Methods for the Rational Discovery of Allosteric Drugs
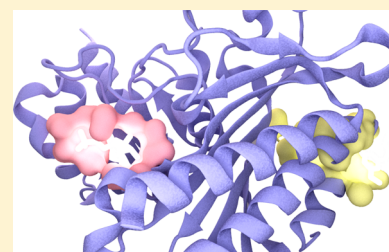
Jeffrey R. Wagner,[†] Christopher T. Lee,[†] Jacob D. Durrant,[†,‡] Robert D. Malmstrom,[†,‡] Victoria A. Feher,[†] and Rommie E. Amaro*[,†,‡]

[†]Department of Chemistry & Biochemistry and [‡]National Biomedical Computation Resource, University of California, San Diego, La Jolla, California 92093, United States

ABSTRACT: Allosteric drug development holds promise for delivering medicines that are more selective and less toxic than those that target orthosteric sites. To date, the discovery of allosteric binding sites and lead compounds has been mostly serendipitous, achieved through high-throughput screening. Over the past decade, structural data has become more readily available for larger protein systems and more membrane protein classes (e.g., GPCRs and ion channels), which are common allosteric drug targets. In parallel, improved simulation methods now provide better atomistic understanding of the protein dynamics and cooperative motions that are critical to allosteric mechanisms. As a result of these advances, the field of predictive allosteric drug development is now on the cusp of a new era of rational structure-based computational methods. Here, we review algorithms that predict allosteric sites based on sequence data and molecular dynamics simulations, describe tools that assess the druggability of these pockets, and discuss how Markov state models and topology analyses provide insight into the relationship between protein dynamics and allosteric drug binding. In each section, we first provide an overview of the various method classes before describing relevant algorithms and software packages.

CONTENTS

1. REVIEW MOTIVATION AND ORGANIZATION

To date, most allosteric drugs have been discovered through high-throughput screening. But growing databases of biomolecular structure and sequence data, in conjunction with increases in computing power and improvements in predictive algorithms, are enabling the rational de novo design of allosteric drugs. Given the large number of published algorithms for predicting allosteric mechanisms, it can be difficult to select the most appropriate method for a given target. This review serves as an introduction for those who wish to use computational techniques to develop allosteric drugs.

After a broad overview of allosteric drug discovery, this review is divided into three sections. First, we discuss bioinformatics and molecular-dynamics methods to identify allosterically important sequence positions. Second, we summarize the computational

methods to predict druggable pockets at these functionally relevant sites. Finally, we describe how Markov state models and topological analyses can tie these single sequence sites to global protein function and dynamics.

2. INTRODUCTION

Allosteric drugs offer a number of advantages that make them desirable as drug candidates. Allosteric effectors, by definition, alter protein activity by binding to a site distinct from the orthosteric pocket. One example of an allosteric system is fructose 1,6-bisphosphatase (shown in Figure 2). Because allosteric sites are typically less evolutionarily conserved, allosteric drugs can be highly selective, even among other members of the same protein family.[1−8] In some cases, allosteric sites are so unique among proteins that an effector is said to have "absolute subtype specificity."[2,3,9,10]

Allosteric modulators may have spatiotemporal specificity. For example, they can be active only in the presence of the endogenous ligand, thus restricting their effect to certain tissues at certain times,[2−4,9,11] which may slow desensitization.[10,12]

Allosteric effectors are generally saturable, meaning that they have a maximal effect that does not necessarily correspond to complete inhibition or activation.[2,4−6,8,10,11] This saturability enables safer dosing. For example, if the maximal effect is an 80% reduction in signaling, overdosing will not fully eliminate an essential signal.[1,2,4]

Other advantages can include noncompetitive inhibition (i.e., drug activity cannot be "overwhelmed" by high concentrations of the endogenous ligand) and pathway- or substrate-specific modulation, which reduces unwanted activity by specifically targeting a single protein function.[2,4,10] For example, if a protein is involved in multiple pathways, an allosteric effector may impact the activity of each pathway differently depending on the systems-biology context. If a protein acts on multiple substrates, the impact on activity may depend on the biological context.

Despite many potential advantages of allosteric therapeutics, it has been challenging to identify predictive approaches to discovering allosteric drugs. In recent decades, the pharmaceutical industry has favored more traditional targets for three primary reasons: the relative ease of assay development around orthosteric sites; access to high-throughput, high-resolution X-ray crystallography; and advances in ligand- and receptor-based computational methods to optimize ligand-binding affinity at a substrate-competitive site. This structure-based approach is thought to significantly reduce the time and cost of hit-to-lead and lead-to-drug development by reducing the number of compounds that need be synthesized.[13,14] Work by Doman et al. comparing computer-aided drug discovery (CADD) and high-throughput screening (HTS) reported that the two methods had hit rates of 35 and 0.021%, respectively.[15]

In contrast, allosteric drugs are uniquely challenging from a rational drug-design perspective. Because experimental assays typically measure orthosteric function rather than ligand binding at the allosteric site, efficient development of allosteric drugs requires that the complex structure−activity relationships (SARs) governing both binding affinity and allosteric activity be considered simultaneously.[5,10,16] Further, allosteric sites are less likely to be evolutionarily conserved. While this enables increased subtype specificity, it also increases the chances of evolved resistance[5,9,10] and can complicate testing in evolutionarily distant animal models.[10]

Additionally, allosteric effectors are particularly susceptible to "mode switching," where relatively minor chemical changes can drastically affect ligand efficacy.[10,16] Structurally similar drug metabolites, therefore, may have varying and unpredictable distributions and allosteric effects.[10,16] Optimizing allosteric modes of action requires methods that are very different from those used in orthosteric drug discovery.[5]

Multifunctional allosteric proteins are particularly challenging. While drug designers may desire to target a single protein function, an allosteric effector may also alter other functions, hindering a full mechanistic understanding of the pharmacology.[10] Also, the benefits of spatiotemporal specificity are lost if the distribution of the endogenous ligand changes with progression of the disease state.[10] Finally, assessment of the limited number of known allosteric pockets indicates that they are generally shallow[5] and present flat SARs.[10] These structural features similarly challenge existing rational drug-discovery paradigms and the general practice of developing selective compounds by optimizing affinity.

Despite these challenges, allosteric drug discovery has gained momentum recently due to a number of developments.[10] First, several allosteric drugs across a broad range of pharmacological target classes have been rationally designed,[17−23] encouraging pursuit of others, as evidenced by the number of allosteric drugs currently in clinical trials.[24] The recent elucidation of new membrane protein crystal structures for GPCRs[25,26] and ion channels[27,28] have assisted in the structure-based design approach to these successes. Finally, advances in our understanding of allosteric mechanisms have supported development of additional rational design strategies (see below).

Our understanding of allosteric mechanisms has advanced considerably since the initial conception of Monod, Wyman, and Changeux.[29] Modern models of allostery consider conformational ensembles.[5,8,9,30−43] This revised view supports newly established and emerging computational advances that comprehensively map conformational landscapes and predict communication between allosteric and orthosteric sites. For example, the physical mechanisms of allostery generally alter the entropic and enthalpic factors that define the conformational landscape and, therefore, govern protein function.[5,8,9] The observed correlation between allosteric modulation and protein structural dynamics is varied: Major conformational rearrangements occur in some cases, as compared with subtle shifts in conformational populations in others.[4−6,9,10,35,37,44] An excellent metaphor for these phenomena is Kornev and Taylor's classification of "domino" versus "violin" models of allosteric signal transduction[45] (Figure 1). Further, allosteric signals are transmitted through a range of structural motifs, from rigid core regions to flexible linkers.[46] Allostery may occur through essential residues along a single allosteric path[47] or through many weak pathways connecting one site to another, acting in concert.[48] As such, it is not surprising that many of the allostery-prediction methods discussed in this review (some of which do not use structure/geometry information at all) in practice may identify noncontiguous groups of residues as being allosterically linked. Such predictions should not immediately be assumed to be wrong but rather may indicate a non-"domino" model of allostery.

Recent work has also revealed that protein allostery is not merely a transition between two discrete protein conformations, as initially thought, but rather a shift in the equilibrium populations of many conformations, induced by effector binding.[31,35,38,41,42,51,52] It is becoming increasingly clear that the kinetics of these transitions define the mechanisms of allostery.[38,51] Empowered with these new understandings and advances in molecular simulation (in terms of speed of
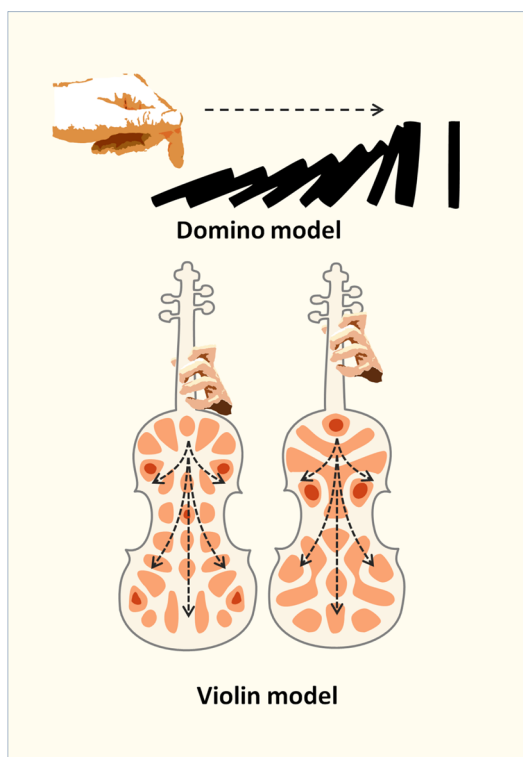
Figure 1. When a small molecule binds to the allosteric site of a protein, information is transferred through the protein molecule to its active site. Two different methods of transmission can be defined. The first mechanism, here defined as the "domino model", is a sequential set of events propagating linearly from the allosteric site to the active site. Binding of the effector triggers local structural changes that sequentially propagate via a single pathway to the active site. It was suggested that this mechanism is applicable for the PDZ domain family,[49] G protein-coupled receptors, the chymotrypsin class of serine proteases, and hemoglobin.[50] The second mechanism, defined here conceptually as a "violin model", is based on vibration pattern changes inside the protein. In a violin its pitch can be changed by a slight movement of the violin player's finger on the fingerboard. Information about the finger movement is, thus, transferred throughout the whole body of the violin with no specific pathway for the signal transduction. By analogy, protein allosteric site is a fingerboard of the protein and a small signaling molecule is the player's finger. If a protein is in a particular vibration mode, it is possible to suggest that binding a small effector molecule to a specific site can change this mode. The signal, thus, will be spread throughout the whole protein including its active site. The "domino model" is a reliable way to transfer information in a macro world, but on a molecular level, with significant thermal motions of the protein, this mechanism will be prone to random triggering of the domino chain reaction, creating noise in the signaling system. Thermal motions in the case of the "violin model" do not hinder the transduction. In fact, the permanent motion of the molecule is a prerequisite for this mechanism. Reproduced with permission from ref 45. Copyright 2015 Cell Trends in Biological Sciences.

calculation and improved methodologies), the era of allosteric drug discovery is now on the cusp of radical advancement. I'

2.1. Emerging Rational Design Principles

So-called tried-and-true "design principles" are still being developed. However, a few general principles have begun to emerge. For example, many argue that it is insufficient to design a ligand that merely binds to an allosteric site; rather, the effector must make contact with certain key binding-pocket atoms to have the desired effect.[5,10] These key atoms can often be identified through mutagenesis experiments and crystallographic studies of other allosteric ligands.

A promising set of design principles is encapsulated in the "allo-network" strategy, a rational approach that adopts two simultaneous but orthogonal approaches to ligand design.[6] On the protein-structure level, the primary focus is to target a single protein function or an interaction with a single partner. On the signaling-pathway level, the "allo-network" strategy suggests targeting less-connected upstream proteins instead of the more direct, though potentially highly connected, signaling proteins themselves. When applied to early stage design, the allo-network method is predicted to increase the likelihood that a given allosteric effector will proceed through the drug-approval process.[6,12,55]

Several published examples for recently approved allosteric drugs serve to illustrate the current state of the art for emerging allosteric drug design principles. They also represent the significant advances that have been made to utilize structure-based methods for challenging druggable sites such as protein–protein interfaces and for membrane proteins such as ion channels. The available details of the discovery and optimization of these compounds do not include the methods discussed in this review; however, they highlight where these predictive techniques could contribute to the allosteric design process.

In 2011, Gilmartin et al. of GlaxoSmithKline reported the discovery of a pharmacokinetically optimized allosteric MEK inhibitor, GSK1120212.[56] The first generation inhibitor was discovered by high throughput screening,[57] and the subsequent ternary crystal structure showed the allosteric pocket to be adjacent to the orthosteric ATP-bound site.[58] By 2012 there were 14 allosteric MEK1/2 inhibitors in clinical trials,[59] because it was recognized that an inhibitor developed for this allosteric pocket afforded two very unique opportunities to avoid adverse clinical effects: the high doses required to compete against 1 mM cellular ATP concentrations and inhibition of closely related ATP-binding sites in other kinases. The unique efficacy properties of GSK1120212 highlight both the opportunity and challenge of allosteric drug design. Gilmartin et al. report that, although some other MEK allosteric drugs demonstrate inhibition of the ERK1/2 pathway in vitro, this has not translated into efficacy in patients.[56] GSK1120212 has since been approved in the United States under the name Trametinib for treatment of metastatic melanoma caused by the V600E mutation.

In 2012, Saalau-Bethell et al. of Astex reported the discovery of allosteric inhibitors for the HCV NS3 protein.[17] These inhibitors produce an allosteric effect by binding at an interdomain interface and stabilizing a preexisting autoinhibited state of the protein. The original discovery of the allosteric site was accomplished using a fragment-based HTS technique, followed by optimization using X-ray crystallography and structure-based SAR. The authors discuss their experience with a few confounding factors in the allosteric design process, namely the need to use the full-length protein in their screening construct to observe the exerted allosteric effect, and the subsequent directed evolution study of resistance mutations that could occur at the allosteric site.

Hackos et al. of Genentech published on the discovery of positive allosteric modulators (PAMs) for GluN2A-containing NMDA receptors in 2016.[60] PAMs are allosteric ligands which increase the effect of the endogenous signaling molecule and do not cause a change in its absence. The allosteric site in this case was at a protein–protein interface, and was discovered using HTS. Subsequent medicinal chemistry efforts then optimized the
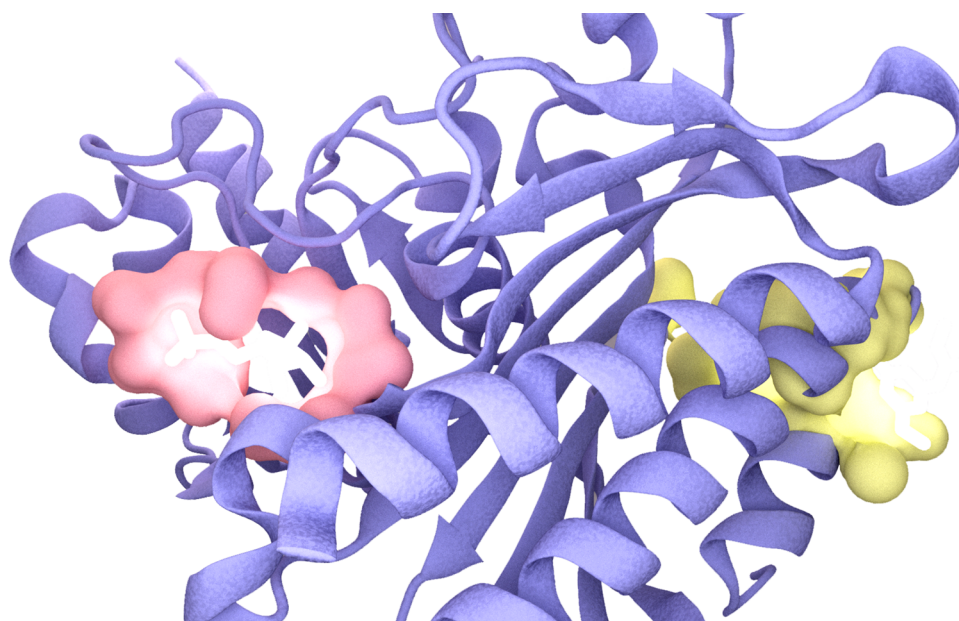
Figure 2. Allosteric protein fructose 1,6-bisphosphatase, shown for illustration. Orthosteric and allosteric pockets (yellow and red, respectively) are bound to an endogenous ligand and an allosteric effector, respectively. Note that the allosteric site is distant from the orthosteric site such that there is no overlap between the bound poses of the allosteric and orthosteric ligands. Despite the distance between them, the allosteric effector measurably modifies the enzymatic activity at the orthosteric site. Illustration derived from PDB IDs 2Y5K[53] and 3IFC.[54]

early hit molecule. The authors note that the validation of this allosteric site was reinforced by its similarity to an analogous allosteric site in AMPA receptors, but that the NMDA receptor site has elements of asymmetry that the AMPA receptor site did not. In comparing two similar compounds, GNE-6901 and GNE-8324, the authors make comments that indicate evidence of mode switching or a shallow SAR landscape, and they further characterize the details of the allosteric mechanism using mutagenesis experiments.

In summary these examples demonstrate allosteric drug discovery can be successful at protein sites often considered to be undruggable. It is apparent that these successes can be further built upon through computational methods that allow for rational rather than serendipitous HTS discovery of new allosteric binding sites and a deeper understanding of allosteric mechanisms that overcome design challenges such as mode switching.

3. COMPUTATIONAL METHODS FOR STUDYING ALLOSTERY

3.1. Protein-Sequence Analysis Methods

3.1.1. Introduction. Protein-sequence analysis is a useful tool to detect and characterize allosteric pathways and pockets. Here, we classify sequence-based methods into two groups: (1) "single site" methods, which produce a list of individual functional sequence positions; (2) "coupled site" methods, which produce a list of groups comprised of two or more sequence positions that appear to be functionally linked based on their coevolution.

All sequence-based analysis methods share some challenges. These challenges include how to select and aggregate clean, relevant sequences as input; interpret the output; and integrate sequence-analysis results with other forms of data. Determining the biological meaning of a strong signal is also problematic. While many analysis methods identify evolutionarily important residues, the specific biological role of these residues cannot be

inferred without additional knowledge. For example, it is difficult to determine, based on sequence alone, whether an evolutionarily significant residue plays an allosteric role, or whether its role is related to another essential process (e.g., substrate binding, maintaining protein structure, etc.).[61,62] Indeed, it is likely that a given residue serves multiple purposes simultaneously.

Input sequence selection and alignment also present challenges. Most techniques require many sequences to establish statistical significance. To obtain the required number of sequences, researchers often lower the stringency of their search parameters, resulting in alignments that contain sequences with lower similarity or incomplete coverage of the original query. While some analysis methods manage to detect meaningful coevolution over a wide range of multiple sequence alignment (MSA) conservation and noise levels, others are more susceptible to messy data.[63,64] For a more complete discussion of these topics and how they affect coevolution analysis methods, readers are directed to an excellent recent review by de Juan et al.[65]

3.1.2. Single-Site Evolutionary Analysis Methods. By our definition, single-site evolutionary analysis methods return a list of predicted functional sequence positions but do not suggest specific linkages between sites. Once a researcher has constructed an MSA, the conservation or phylogenetic relevance of each column can be used to infer the evolutionary importance of each sequence position. This importance is sometimes a hallmark of thermodynamically critical residues that participate in allostery. Though single-site methods only return a list of single high-scoring sequence positions, the inner workings of some single-site methods are based on the aggregate or correlated behaviors of multiple sequence positions (e.g., to determine baseline residue probabilities within a multiple-sequence alignment or construct a phylogenetic tree).

Single-site methods for detecting allostery are advantageous because they lack much of the noise often associated with

correlation analysis. These analyses are also appealing because of their simplicity: There are usually fewer parameters to set, and the results can be visualized directly by highlighting key residues on a three-dimensional (3D) protein structure.

3.1.2.1. Single-Position Entropy. Shannon entropy, one of the simplest nontrivial sequence-analysis metrics,[66] was used widely in early works to identify conserved sequence positions for drug-design or mutagenesis experiments.[67] Similar in form to thermodynamic entropy from statistical mechanics, Shannon entropy measures the population diversity of residues at a single MSA position. It is also central to mutual information (MI), a popular coupled-pair metric. The MI of two sequence positions is defined as the sum of the individual position entropies, minus the entropy of the positions considered jointly. While we do not cover the mathematical details of these methods here, interested readers are directed to previous articles on these topics.[64,68]

Shannon entropy does not consider amino acid similarity (e.g., in the Shannon entropy framework, a leucine-to-isoleucine mutation is considered mathematically equivalent to a leucine-to-arginine mutation). Other entropy measures, such as the relative Shannon entropy (also called the Kullback–Leibler divergence (KLD))[69] and the von Neumann entropy,[70,71] attempt to overcome this limitation and, as a result, may be more useful in the search for allosteric sites. Relative Shannon entropy/KLD accounts for some measure of the protein's chemical environment by considering each mutation with respect to the background amino acid frequencies calculated from the MSA. This analysis may be particularly useful when searching for allosteric sites in proteins that reside in membranes or other noncytosolic compartments, where background residue probabilities or mutational preferences may be biased due to different biochemical contexts. In contrast, von Neumann entropy, a concept borrowed from quantum statistical mechanics, is calculated using amino acid similarity matrices. Identifying an optimal amino acid similarity metric is nontrivial and may well depend on the nature of the system (e.g., in a well-packed protein, residue size may be a sensitive metric, whereas surface-site comparison may require the user to prioritize charge). In a recent publication describing these types of entropy, Johansson and Toh explored how the two metrics can be mixed to detect enzyme active sites with maximum sensitivity.[64]

Zhang et al. constructed a variety of new analysis methods in 2008 by combining Shannon or von Neumann entropy, phylogenetic tree structure, and a novel gap-treatment approach.[70] In benchmarking their method, they compared their results to Evolutionary Trace and ConSurf (discussed in greater detail below). Two of their hybrid approaches outperformed all other techniques in detecting significant residues across a variety of proteins: the Improved Zoom method, which incorporates a tree breakdown of subalignments, and the Physiochemical Similarity Zoom method, which extends the Improved Zoom method with von Neumann entropy and tree-branch-size normalization.

3.1.2.2. Evolutionary Trace. Lichtarge, Bourne, and Cohen pioneered the evolutionary trace (ET) method. The approach has become quite popular, largely because the algorithm is intuitive and its results are readily visualizable.[72] ET aligns a number of sequences and constructs a phylogenetic tree, and then monitors the conservation of sequence positions at major tree branching points. By slicing the tree at different similarity cutoffs, the algorithm extracts the cluster-defining sequence positions. The evolutionary significance of these sequence positions is implied by their conservation in the sequences

beyond the next branch. In their first paper,[72] the authors demonstrated that ET can detect functionally important sites in SH2, SH3, and DNA-binding domains. Work has since been published on ET validation, parameter optimization, and best-use practices.[73]

In a method often referred to as "difference-ET," the user runs ET on two related proteins and considers differences in the high-ranking residues and their scores. The sequence positions with strongly varying scores may suggest specificity determinants or differences in allosteric and/or orthosteric mechanisms. Notably, difference-ET has been used in the study of GPCR specificity.[73−75]

To better account for varying rates of evolution in different subtrees and correlated mutations, in 2004 Mihalek et al. developed real-valued ET.[76] This method incorporates entropy information into the standard ET framework. This work also introduces the zoom ET method (not related to improved zoom, above), which adds higher weight to sequences that are more similar to the protein of interest. In the introductory work, they used real-valued and zoom ET to detect the functional residues in a kinase domain, and then compared the performance of both methods to regular (integer-valued) ET and entropy. Given unpruned sequence data sets, the real-valued ET and zoom ET methods outperformed the others by a significant margin. In contrast, integer-valued ET prevailed in most respects when pruned data was available. An automated web server is available to perform real-valued ET calculations, generate reports, and visualize results (http://mammoth.bcm.tmc.edu/ETserver.html).[77]

3.1.2.3. H2r(s). In 2008, Merkl et al. introduced a method called H2r that serves as a segue between single-site and coupled-site approaches.[78] H2r generates a mutual-information matrix for an MSA, and then discards all but the strongest detected coupled pairs. For each sequence position k, the method returns conn(k), the number of top-ranked pairs that include k. Initial work proved that H2r can successfully detect functionally significant residues across a range of proteins. More recently, H2rs, an improved version of H2r, has been released.[79] This method modifies the original by using von Neumann instead of Shannon entropy and performing more detailed checks for statistical significance. H2rs is available as a web server and a stand-alone application at http://www-bioinf.uni-regensburg.de/.

3.1.3. Coupled-Site Evolutionary Analysis Methods.
Second-order sequence analysis detects residue pairs that mutate in concert more frequently than would be expected given random genetic events. Coevolving residue pairs are assumed to be functionally linked, often because they serve essential roles in allostery or structural integrity.

The immediate output of second-order allostery analysis is a list of residue pairs with associated correlation strengths. Combining these individual pairwise correlations into a single picture of the entire protein is a separate task. On the most basic level, the strongest correlations that include a residue or site of interest can suggest thermodynamic coupling to other sites, possibly related to allostery. More complex analyses use hierarchical clustering or principal component analysis to analyze these linkages and uncover strongly linked networks of coevolving residues.

3.1.3.1. Basic Coupled-Site Analyses. Several simple yet reliable residue-coupling analyses have maintained a presence in the literature over the past decades. These basic approaches are advantageous because they are easier to understand and have been shown to score consistently well in a wide range of tests.

However, they may fail to detect correlations in more complex cases.[80,81] Though more complex methods exist, many of these basic methods still appear as analysis options in coevolution-detection software packages and web servers. In this review, we focus on a few that are still widely used.

3.1.3.2. Mutual Information. Mutual information (MI) is one of the most straightforward and long-lived coupling metrics. The MI between two sequence positions is defined as the sum of the Shannon entropies of both positions, minus their joint entropy. Due to its simplicity and favorable mathematical properties, MI analysis is the basis for a number of more complex coevolution methods. However, MI does present certain shortcomings. For example, uncorrelated pairs of high-entropy sequence positions are likely to have a higher MI than uncorrelated pairs of low-entropy positions.[82,83] To compensate for this and other shortcomings, various software packages have implemented a number of mathematical corrections to MI.[84−88] Further, methods to estimate baseline values for correlation (e.g., resampling or sequence shuffling) can improve MI analysis.[83,89,90]

Another relatively direct coevolution metric, the McLachlan-based substitution correlation (McBASC),[91] looks for similar patterns of variation in the columns of an MSA, weighting for residue similarity using the McLachlan scoring matrix.[92] Analogous methods can be constructed using different substitution matrices, but McBASC continues to be a popular choice in the literature.[63,83,93]

In 2002, Kass and Horovitz[94] analyzed the GroEL complex using a chi-squared test to detect significant residue coevolution in an MSA. The analysis suggested intra- and interchain contact pairs and has continued to appear in the literature under the name "observed minus expected squared" (OMES).[63,80,81,83,95]

3.1.3.3. Statistical Coupling Analysis. The statistical coupling analysis (SCA) method developed by Lockless and Ranganathan is perhaps the most widely used sequence-based method for allostery prediction.[49] SCA draws an analogy to statistical physics by calculating a "coupling energy" between each sequence-position pair. The original SCA method computes a conservation value for each sequence position i in an MSA, applies one of several types of perturbation to another position j (depending on the SCA version[96]), and finally recalculates the conservation at position i for the sequences that satisfy the perturbation. By calculating the change in individual and joint conservation over a variety of perturbations, SCA establishes a "coupling energy" that indicates the evolutionary coupling of positions i and j.

The output of the SCA method is an $N \times N$ matrix of coupling energies, where N is the number of sequence positions in the alignment. In early work, the researchers manually identified strongly coupled residue pairs that included one functional member (per experiment). More recent versions of SCA have grouped this matrix into meaningful clusters of coevolving residues using hierarchical or spectral clustering.[50]

Refinements of SCA have achieved improved statistical properties by resampling the original distribution.[97] In a 2011 paper, SCA was effectively used to engineer a light-sensitive LOV2 domain onto the surface of DHFR at a location that SCA had identified as energetically linked to the enzyme active site. Some variants of the resulting protein chimera were found to have acquired light-dependent activity.[98]

Further work has used SCA to design artificial WW domains.[99,100] In 2011, an SCA analysis of antigen 85C from *Mycobacterium tuberculosis* suggested new sites that potentially could be exploited in drug design.[101] A number of projects have

also demonstrated how SCA can be used to target mutations that affect protein function.[102−108]

Inspired by earlier work on the sequence correlation entropy (SCE) method,[109] Dima and Thirumalai published an SCA variant in 2006.[110] This variant controls for specific protein composition by calculating the background probability that a given amino acid will be present at a random sequence position. This probability is determined by considering only the sequences being analyzed, as opposed to all sequences in the SWISS-PROT database.[111,112] Further, they borrowed a coupled two-way clustering procedure from gene-sequence analysis to define the sectors.[113] In validating this method, the authors analyzed the PDZ, GPCR, and lectin families of proteins and were able to quantitatively predict functional residues, which were in agreement with experimental findings.

3.1.3.4. Explicit Likelihood of Subset Covariation. As mentioned above, SCA is a "perturbation-based" method in which correlation is established by excluding certain sequences from an MSA and monitoring how entropies change. Another popular perturbation-based method was published in 2004 by Dekker et al.[114] This method, explicit likelihood of subset covariation (ELSC), relies on similar principles but returns correlation scores in the form of probabilities rather than statistical energies. ELSC was shown to be superior to SCA in contact prediction when tested on a range of protein families. It has since been implemented on web servers[115,116] and has been a popular benchmark method in the literature.[80,81,93,95]

3.1.3.5. Direct Coupling Analysis. In 2009, Weigt et al. proposed a mutual-information-based method called direct coupling analysis (DCA) that disentangles directly interacting residues from large networks of indirectly coupled sequence positions.[117] While this method is typically used in structure prediction to identify spatially adjacent sequence positions, it may find application in the study of short-range allosteric interactions. A more efficient implementation of the DCA method, known as "mean field" (as opposed to the original "message-passing" implementation) was published in 2011.[118] Both introductory papers show that DCA is a robust predictor of both intra- and interprotein contacts and that it can hint at the existence of unobserved protein conformations. Related work has shown that DCA can be used in conjunction with structural models to generate predictive models of protein complexes,[119−121] determine the sequence positions that contribute to protein-interaction specificity,[122] and describe the conformational ensembles of proteins in crystallographic or near-crystallographic states.[123] A web server and software package are available to perform DCA analysis at http://dca.rice.edu/portal/dca/home.

3.1.3.6. PSICOV. PSICOV is another popular contact-detection method that may find productive use in the study of allostery.[124] Mathematically, PSICOV relies on an estimated inverse of the MSA covariance matrix, which acts as a matrix of correlations between all sequence-position pairs that inherently controls for the variations in all other positions. PSICOV was successful at predicting contacting protein residues based on MSA data. The code has been published online at http://bioinfadmin.cs.ucl.ac.uk/downloads/PSICOV/.

3.1.3.7. Recurrence Quantification Analysis. Recurrence quantification analysis (RQA), another second-order sequence-analysis technique, is best used when much is already known about the mechanism under investigation (e.g., physiochemical amino acid properties such as charge or hydrophobicity are known to drive the allostery). RQA itself is a general method in

nonlinear dynamics:[125] In the context of protein sequences, it considers a scalar-value vector that represents some property of a given sequence. In introductory work by Zbilut et al.,[126] the method was used to properly classify 56 TEM-1/β-lactamase mutants with impaired function based on their hydrophobicity profiles. Further RQA work used hydrophobicity scores to classify proteins as allosteric or nonallosteric,[127] study p53 mutants,[128] and reveal interaction partners in viral-envelope proteins.[129]

In 2005, Colafranceschi et al. investigated the effect of choosing different physicochemical amino acid descriptors and changing the numerical parameters of the RQA algorithm.[130] More recently, a comparison method based on RQA measurements, known as cross-RQA, effectively detected protein allostery.[131] Interested readers are directed to a review by Zbilut-Webber, which provides examples of RQA applied to a range of computational biology problems.[132]

3.1.3.8. Comparative Analyses. Some work has been done to competitively benchmark the performance of these methods. In 2004, Fodor and Aldrich compared OMES, MI, SCA, and McBASC in a variety of tests. In short, the study found that performance is largely dependent on the way that different methods determine background residue probabilities and handle positional conservation.[63] A follow-up study investigated how effectively coevolution analysis finds thermodynamically linked residue pairs.[62] In general, spatially contiguous linked pairs were detected, but long-range couplings did not agree with experiments.

In 2010, Brown and Brown introduced a new pair-scoring method, called Z-scored-product Normalized Mutual Information (ZNMI), and compared it to the accuracy and reproducibility of MI, two versions of SCA, OMES, and ELSC.[81] The authors presented a thorough meta analysis of method performance and the impact of input-parameter selection. Though none of the tools tested was particularly powerful, ZNMI was the most robust prediction tool. Brown and Brown also found that the use of multiple subalignments produced more accurate and reproducible results.

A comparative analysis of SCA and DCA revealed that the top 35 "sectons" found via spectral clustering of the DCA matrix corresponded to pairs, triplets, and quadruplets of spatially contiguous residues.[133] In contrast, a similar analysis of the SCA matrix produced spatially adjacent clusters of many residues each. These different results validate the stated goals of each method: DCA aims to find contacting pairs, whereas SCA aims to find potentially distant groups that are thermodynamically linked in a certain function.

In 2014, Pelé et al. investigated seven coevolution analysis methods to find the hallmark covarying pairs in GPCR alignments.[80] They considered three variants of MI, McLachlan based substitution correlation, SCA, ELSC, and OMES. OMES and ELSC were the most robust methods for finding the residues responsible for subfamily divergence. Their article also included an insightful discussion of the methods.

Mao et al. published a comparative analysis in 2015.[134] Their study tests OMES, two variants of MI, SCA, PSICOV, and DI, and finds that PSICOV and DI are best at identifying contacting residues. OMES and MIp excel at removing false positives from the lists of predicted contacts. While the authors focused on detecting inter- and intramolecular contacts, their analysis also provided useful insights to guide the productive use of each method. For example, all methods benefit from repeatedly shuffling the MSA and rerunning the analyses in order to provide

a baseline and remove false positives. Finally, the authors found that the consensus of DI and PSICOV provides a more robust prediction of contacting residues than any single prediction method alone. The software used to perform this analysis is available through the ProDy Evol program (http://prody.csb.pitt.edu/evol/).[135]

In the course of introducing new types of MI analysis (dbZPX2, dgbZPX2, and nbZPX2) and evaluating the effectiveness of MSA simulation (a topic beyond the scope of this paper), Ackerman et al. in 2012 compared many different coevolution analyses in their ability to predict contacting residue pairs.[95] These comparisons found that the "new" methods (the ZPX2 family, DCA, and $\log(R)$ (not discussed here)) were significantly superior to the "old" methods (OMES, McBASC, ELSC, and SCA).

3.1.3.9. Web Servers. Several web servers perform and visualize sequence analyses. Given a PDB code, Contact Map WebViewer (CMWeb)[116] automatically constructs an MSA and visualizes a variety of coevolution analyses: mutual information, SCA, ELSC, OMES, and an early method presented by Göbel et al.[136] The same server can also compare the results of these methods to user-uploaded data (e.g., results the user obtained using some other type of analysis). The CMWeb server can be accessed at http://cmweb.enzim.hu/.

The Coevolution Analysis of Protein Residues server hosted by the Gerstein Lab[115] (http://coevolution.gersteinlab.org/coevolution/) can perform a large number of the coupled-site analyses presented in this review, including SCA, ELSC, MI, and McBASC-type methods employing different scoring matrices. The server can also validate the results of these methods in predicting residue distances in a crystal structure.

MISTIC (Mutual Information Server to Infer Coevolution) is an automated web server that accepts user-submitted MSAs or collects them from PFAM.[82] MISTIC uses a corrected form of MI to infer coevolving pairs and offers several analysis methods that combine structure and coevolution.[87] It can be accessed at http://mistic.leloir.org.ar/.

CAPS (Coevolution Analysis using Protein Sequences) is a unique algorithm that combines phylogenetic, 3D, and MSA data to predict coupled sequence positions.[137,138] Versions 1 and 2 are hosted on web servers at http://bioinf.gen.tcd.ie/caps/ and http://caps.tcd.ie/, respectively.

The Interprotein-COrrelated Mutations Server (I-COMS, http://i-coms.leloir.org.ar/index.php) focuses on detecting contacts at protein–protein interfaces, though it can also return intrachain correlations.[139] The server automatically builds alignments; performs MI, DCA, or PSICOV analysis; generates visualizations of the results; and allows users to download data taken from various points in the data-collection and analysis workflow.

In 2012, Jeong and Kim published a study describing a close MI variant.[140] They employed an automated workflow to control for various types of noise in sequence alignments, using the MSA sequence profile to establish prior knowledge about the protein. While the authors only studied a few MI variants, they stressed that their profile-based method could be extended to more complex analysis techniques. Their approach, Correlated Mutation Analysis Tool (CMAT), is available on a web server at http://binfolab12.kaist.ac.kr/cmat/.

Access to ConSurf, a single-site detection method similar to ET, is available at http://consurf.tau.ac.il/.[141−146]

3.1.3.10. Software. The ProDy Python package, referred to above, can compute a variety of coevolution metrics.[134,147,148] In

Table 1. Table of Selected Coevolution Web Servers/Software Packages and Their Capabilities

	web server	downloadable	generates MSA	Shannon entropy	relative Shannon/KLD	MI	SCA	DCA	PSICOV	OMES	ELSC	notes	URL
CMWeb[136]	X	X	X			X	X			X	X	also computes an early method from Gobel et al.	http://cmweb.enzim.hu/
MISTIC[82]	X	X			X	X						calculates a corrected version of MI	http://mistic.leloir.org.ar/
CMAT[140]	X	X	X		X	X						many noise-filtering parameters to customize; returns MI, MIp, and MIc	http://binfolab12.kaist.ac.kr/cmat/
I-COMS[139]	X	X	X			X		X				offers two variants of DCA	http://i-coms.leloir.org.ar/
ET[77]	X	X										runs evolutionary trace	http://mammoth.bcm.tmc.edu/ETserver.html
coevolution analysis of protein residues[115]	X	X				X	X			X	X	also computes similarity matrix-based methods (including McBASC), chi square, and quartets coevolution metrics	http://coevolution.gersteinlab.org/
ConSurf[146]	X		X									performs a single-site type of analysis similar to ET	http://consurf.tau.ac.il/
DCA[117,118]	X	X						X				returns pairwise DI	http://dca.rice.edu/portal/dca/home
CAPS[137,138]	X	X										runs a nonstandard coevolution analysis technique	http://caps.tcd.ie/
H2r[78]	X	X				X						runs a nonstandard single-site coevolution analysis technique	http://www-bioinf.uni-regensburg.de/
H2rs[79]		X				X						runs a nonstandard single-site coevolution analysis technique	http://www-bioinf.uni-regensburg.de/
Bio3D[149]		X		X								R package; useful for creating and modifying sequence alignments	http://thegrantlab.org/bio3d/
CorMut[151]		X				X						R package; also computes K_a/K_s and Jaccard index	https://www.bioconductor.org/packages/release/bioc/html/CorMut.html
ProDy[147,148]		X		X		X	X			X		Python package; can compute DI and mutual information correction/normalization	http://prody.csb.pitt.edu/evol/

2014, Skjærven et al.[149] released the latest version of the powerful Bio3D R package for protein structure and sequence analysis.[150] While it focuses on structural analysis, the package can compute Shannon entropy and offers useful functions to create and manipulate sequence alignments. Also in 2014, Li et al.[151] published the CorMut software package for R, which computes MI, a metric called the "Jaccard index,"[152] and the conditional selection pressure metric K_a/K_s.[153] Table 1 summarizes the selected coevolution web servers/software packages and their capabilities discussed here.

3.2. Simulation Methods/Correlated Motions

3.2.1. Molecular Dynamics and Monte Carlo Methods.
Protein allostery is intimately tied to protein dynamics. Allosteric effectors communicate with orthosteric pockets by altering protein dynamics, either through large-scale structural changes or through more subtle changes in correlated residue motions. A proper understanding of the allosteric mechanism is impossible unless one fully appreciates the underlying dynamic conformational flexibility and energetic landscape.

Computational methods to characterize receptor flexibility include molecular dynamics (MD) and Monte Carlo simulations. Briefly, MD simulations represent molecular systems as beads (e.g., atoms) connected by springs (e.g., bonds). Newton's equations of motion are numerically integrated to propagate the dynamics of this classically formulated system.[154−157] Simulation time is typically discretized into steps of 1 or 2 fs, as required to accurately simulate the fastest atomic motions (i.e., hydrogen bond stretching). At each step, the potential energy of the system is calculated by considering the energies associated with bonded atoms (e.g., due to bond stiffness, angle bending, torsion rotation, etc.) and nonbonded atom pairs (e.g., due to electrostatic and van der Waals forces). The forces on each atom, calculated from the negative gradient of the potential energy, are then used to update the atomic positions at each step. In a related method, called "accelerated MD" (aMD),[158−163] the underlying potential energy landscape is modified by raising the energy wells. These modifications facilitate transitions between states that do not normally occur on the time scales of traditional MD simulations. Interested readers are directed to recent reviews that describe MD simulations in the context of drug discovery.[164−167]

In contrast, Monte Carlo based methods take a stochastic approach to conformational sampling.[168,169] At each Monte Carlo step, a Markov-chain procedure is used to generate a slightly modified system conformation. This new conformation is then accepted or rejected at random, biased by the relative potential energy of the new conformation as compared to that of the previous step (the so-called "Metropolis criterion"[168]). As these simulations are stochastic, they do not typically explore conformational space via the same time-dependent path traced in an MD simulation. Indeed, a key limitation of the Monte Carlo method is that the time information connecting the states is lost. Nevertheless, Monte Carlo simulations are widely used because they tend to more efficiently sample system conformations that are statistically representative of the equilibrium state. A recently published book by Landau and Binder describes the technique in detail.[170]

3.2.2. Network Representations and Protein Allostery.
As mentioned above, allosteric signals can be transmitted through large-scale conformational changes or subtle shifts in the correlated motions/interactions of individual residues.[31,35] Accessing the time scales required to observe many large-scale

conformational changes is presently computationally intractable using unbiased simulation-based methods,[171,172] but these simulations are well-suited to the study of allosteric signals that are transmitted via fast, local fluctuations (i.e., nanosecond-scale changes in the coordinated motions of adjacent residues).

Distilling the dynamics sampled by these simulation-based methods into a simple network representation of protein motions often facilitates allosteric analysis.[173−179] Rather than tracking the position of every atom, each protein constituent (e.g., amino acid) is represented as a single node (located, for example, at the residue center of mass). Each pair of nodes is connected by an edge whose length is inversely proportional to the degree of interdependence between their motions, such that nodes connected by short edges are highly interdependent. Different metrics of "interdependence" have been used, including metrics based on correlated motions (e.g., mutual information,[180,181] fluctuations in atomic positions,[182] etc.) or the number of specific noncovalent interactions.[182]

Once a network representation of protein dynamics has been constructed, various network-analysis techniques (described below) can identify important pathways of communication between distant residues that may contribute to allostery. Interested readers are directed to a recent review by Atilgan et al. for more detailed information.[183]

3.2.3. Dynamical Network Analysis.
A number of simulation-analysis techniques consider the interdependence of individual residues along an allosteric path or paths. In 2008, Bradley et al. analyzed molecular dynamics simulations of *Escherichia coli* NikR, a transcriptional repressor of the NikABCDE nickel permease, using a novel network-space clustering approach.[182] Two types of residue−residue interdependence were considered, based on fluctuations in residue motions and the number of polar contacts between residue-pair members. Bradley et al. first identified sets of residues whose motions and/or contact counts were correlated with residues in both allosteric and orthosteric pockets. These sets were then clustered using the "unweighted pair group method with arithmetic mean" (UPGMA), a hierarchical agglomerative approach that groups residue sets according to the degree of similarity in their correlation patterns. The authors ultimately identified residue sets likely to mediate communication between the NikR allosteric Ni^{2+}-binding pocket and the orthosteric DNA-binding pocket.

Others have pursued orthogonal methods to represent and analyze protein dynamics in silico. In 2009, McClendon et al. introduced MutInf, a novel MD-analysis technique for identifying statistically significant correlated motions.[180] Unlike other techniques, MutInf uses an internal coordinate scheme. Rather than considering the positional coordinates of residue atoms directly, it quantifies correlated motions using a mutual-information metric derived from residue torsion angles. This approach makes MutInf particularly well-suited to study allosteric signals transmitted through the gearlike correlated motions of side-chain rotamers. Using multiple short simulations, the authors were able to identify statistically significant correlations. Applying the technique to interleukin-2 provided a viable theory of allosteric transmission involving a population-shift-mediated signal transmitted between the allosteric and orthosteric pockets via a hydrophobic core. More recent work has incorporated the Kullback−Leibler divergence metric to improve the sensitivity of MutInf analysis.[181]

A paper by Van Wart et al., published in 2012, studied imidazole glycerol phosphate synthase (IGPS), an important

enzyme in the histidine biosynthetic pathway of plants, fungi, and microbes.[177] Following a molecular dynamics simulation of the enzyme, they built a representative network of residue interactions as inferred by correlated motions. The edges connecting physically distant protein residues were ignored to emphasize correlations resulting from immediate physical interactions. The single shortest path (in network space) between the allosteric and orthosteric sites was then identified. Two of the residues along this path had been shown previously by experiment to be involved in allostery.

Van Wart et al. later expanded on their work by creating the freely available Weighted Implementation of Suboptimal Paths (WISP) program.[184] WISP can identify not only the single optimal path connecting allosteric and orthosteric sites, but also multiple other near-optimal paths. While allosteric signaling may occur through a single optimal path in some cases, for many proteins it is likely the summed (or perhaps synergistic) effect of signaling through multiple near-optimal pathways.

While several efficient algorithms exist to calculate the single optimal path between two nodes in a network space (e.g., the Floyd–Warshall algorithm, Dijkstra's algorithm, etc.), calculating near-optimal pathways is far more computation intensive. WISP uses several methods to first simplify the network representation of system dynamics. First, the edges between physically distant nodes are removed regardless of the interdependence of their motions, as in Van Wart's original paper. Second, all nodes that cannot possibly participate in optimal or near-optimal pathways of allosteric communication are deleted. For each node, Dijkstra's algorithm is used to calculate the shortest path connecting the source node, the node being considered (n_i), and the sink node. If this shortest path is still too long (in network space) to be an optimal or near-optimal pathway, no other path passing through n_i can be optimal or near-optimal. Therefore, n_i is deleted from the network. After thus greatly simplifying the network, it becomes computationally tractable to calculate notable pathways of allosteric communication between source and sink residues using a recursive, bidirectional approach.

In 2014, LeVine et al. introduced a novel analysis framework, called N-body Information Theory (NbIT), that uses multivariate mutual information to measure residue interdependence. This technique is unique in that it considers not only the interdependence (mutual information) between two residues, but also the mutual information between *groups* of up to N residues. Specifically, a metric called "coordination information" is used to measure the degree to which the correlated motions of a set of residues are shared with another residue not in the set (i.e., the "contribution of a site to all possible *n*-body correlations with another site"). The authors use this technique to analyze MD simulations of the bacterial transporter LeuT and identify many of the same residues previously found through experiment.[185]

3.2.4. Community Analysis.
A number of simulation-analysis techniques expand on calculations of residue–residue interdependence to consider how highly interconnected (and often rigid) *communities* of residues interact with one another. In this view, allostery is achieved when largely isolated communities of highly interconnected residue nodes ("small-world networks") engage in long-range interactions. For example, in 2012, Rivalta et al. studied the allosteric mechanism of PRFAR binding to the IGPS heterodimer.[186] After simulating the system in both the absence and presence of the allosteric modulator, they calculated the mutual information between the positional vectors of all residue–residue pairs. Ignoring the correlations between nodes that were not frequently physically adjacent (≤ 5.0 Å for at least 75% of the simulation), the authors next used the Girvan–Newman algorithm to identify communities of highly interconnected nodes. This analysis provides a powerful illustration of how an allosteric modulator might manipulate entire protein "modules" to transmit a signal. Supported by NMR experiments, the authors ultimately concluded that PRFAR binding decreases the correlated motions at sideL of the heterodimer and induces conformational changes in sideR that ultimately impact the hydrogen-bond network near the active site.[186] Sethi et al. used a similar technique in 2008 to study allostery in tRNA:protein complexes.[173]

The use of accelerated MD (aMD) simulations in this same mutual-information/community-analysis workflow has also proved fruitful. aMD simulations are particularly helpful when the allosteric mechanism being studied occurs on time scales that are longer than those accessible through traditional MD. For example, in 2012, Gasper et al. used a mutual-information/community-analysis technique to study α-thrombin.[174] Ultimately, their results suggested that α-thrombin may have two allosteric pathways. One occurs on slower time scales that might not have been accessible were it not for the aMD. Others have used aMD to similarly study allosteric effects in the M2 muscarinic receptor.[187]

In 2014, Blacklock and Verkhivker used a similar technique to analyze MD trajectories of the Hsp90 chaperone in various functional states.[188] To verify the existence of communities in this system, they first calculated force constants for each residue, corresponding to the energy cost of residue displacement over the course of the equilibrium simulation. As residues with high force constants are typically rigid, community boundaries (e.g., the boundary between a rigid module and a flexible interdomain hinge site) often correspond to locations where these constants change abruptly. In fact, the "hot spot" regions did correspond to residues known experimentally to be important hinge sites or sites of dimerization.

To further characterize this community-based organization, the authors next used protein-structure network analysis to build a network representation of Hsp90 dynamics. They calculated cross-correlation matrices from the MD simulations and identified residues with high degrees of "centrality" (the number of interacting residues) and "betweenness" (the frequency with which a given node lies along the shortest path in network space between two other nodes). Based on this analysis, the authors concluded that Hsp allosteric communication is mediated by highly stable central nodes, surrounded by residues with high degrees of betweenness that may "shield" the more critical residues from the random Brownian/thermal motions common to all proteins.

3.3. Pocket Detection

3.3.1. Introduction.
Simple pocket-finding techniques can be useful to identify potential allosteric sites. One straightforward approach is to consider any identified pocket other than the orthosteric to be allosteric. Pocket-detecting algorithms can be used to analyze static crystal structures. However, coupling these algorithms with techniques that account for pocket dynamics is often useful. Some allosteric pocket conformations are extremely rare in the absence of the allosteric effector itself. These "cryptic pockets" are not readily evident in static apo- or ligand-bound crystal structures, but they may be predictively sampled with molecular dynamics simulations.[189−193] Computational methods

for pocket detection can be classified as geometry, knowledge, and energy based. We describe these three types of methods in sections 3.3.2, 3.3.3, and 3.3.4.

3.3.2. Geometry-Based Pocket Detection.

Geometry-based methods identify pockets by considering only the positions of the receptor atoms themselves. These methods are well-suited to situations where speed of computation is critical (e.g., when pocket detection is applied to conformations extracted from entire MD trajectories) or where pockets are particularly well-defined. On the other hand, accuracy in pocket detection suffers when pockets are shallow, partially collapsed, or highly flexible. And pocket ranking by simple geometric metrics (e.g., volume) does not always correlate strongly with ligand binding affinities or druggability because geometry-based methods do not consider the chemical properties of a given pocket.[194−204]

Many geometry-based algorithms have been described in the literature. They can be classified into three subcategories: sphere based, α-shape based,[205] and grid based. Sphere-based methods, such as PASS,[197] PHECOM,[206] POCASA,[207] and SURFNET,[208] first cover protein surfaces with spheres. Each sphere is evaluated and eliminated if judged unlikely to occupy a surface cavity. The locations of the remaining spheres identify likely binding pockets.

In contrast, methods such as APROPOS,[209] CAST,[210] CASTp,[211] and Fpocket[198] use α-shape schemes to identify protein pockets. In brief, an α-shape is like a convex hull, except that it potentially follows the surface of the source points more closely than a convex hull. The degree to which it deviates from the convex hull is controlled by the α value (for more details, see Edelsbrunner et al.[212]). APROPOS compares multiple α-shape representations of a protein at different values of α.[209] Pockets are identified by subtracting "higher-resolution" α-shapes from "lower-resolution" shapes. In contrast, CAST generates a Voronoi diagram based on protein-atom locations, removes any Voronoi edges and vertices that fall completely outside the receptor, and identifies pockets as collections of empty triangles (for details, see ref 210). Fpocket uses a Voronoi tessellation to calculate receptor α-spheres (i.e., spheres that include four protein atoms on their surface but no protein atoms in their interior).[198] When clustered, these spheres tend to congregate at potential pocket sites.

Grid-based geometric methods for pocket detection include GHECOM,[213] LIGSITE,[214] LIGSITE^CSC,[215] POCKET,[216] Pocket-Picker,[217] PocketFinder,[218] Q-SiteFinder,[219] and POVME/POVME Pocket ID.[194,220] These techniques consider points spaced along a grid that encompasses the protein receptor. Each point is evaluated for the likelihood of pocket occupancy, and high-scoring points are clustered to identify binding pockets, whether orthosteric or allosteric. Various metrics are used to evaluate the points, depending on the algorithm. They include the presence of steric clashes with the protein, detected protein−solvent−protein events,[214,216] buriedness indices,[217] predicted methyl−probe interaction energies,[218,219] etc.

As a representative example of the geometry-based class, consider POVME/POVME Pocket ID, a grid-based pocket-finding algorithm.[194,220] POVME Pocket ID identifies surface and internal cavities by first creating a low-resolution 3D grid of equidistant points that encompasses the entire protein. Points are removed if they are physically near any protein heavy atom or fall outside the convex hull defined by the protein α-carbons. The volumes corresponding to the remaining low-resolution points, which tend to encompass protein pockets, are then replaced with smaller, higher-resolution 3D grids, and the same point-

eliminating process is repeated. Finally, any remaining high-resolution points that have fewer than a user-defined number of neighbors are eliminated iteratively until no such points remain, and stretches of contiguous points are grouped together into distinct pockets.

The locations of these identified pockets can be fed into the POVME program, which calculates pocket metrics such as volume and shape. The latest version of POVME can efficiently analyze entire MD trajectories, allowing the user to monitor pocket changes over time and identify sampled conformations that are geometrically distinct. This approach is useful for identifying cryptic pockets that only occasionally manifest themselves over the course of an MD simulation.

3.3.3. Knowledge-Based Pocket Detection.

Knowledge-based algorithms draw on large structural and genomic databases to determine binding-pocket locations based on sequence and overall structure rather than pocket shape. Algorithms such as 3DLigandSite,[221] FINDSITE,[222] and Pocketome[223] infer these locations by querying existing databases for ligand-bound proteins that are structurally similar to the target protein. These algorithms work particularly well when the target protein has a highly conserved binding site. Other methods, such as FRpred, use sequence conservation, in conjunction with supporting methods like surface-residue-prediction algorithms, to identify surface residues that are likely biologically important.[224] Finally, some hybrid methods, such as the evolutionary-trace method,[225] ConSeq,[141] and ConCavity,[226] combine sequence and structural homology to identify likely binding pockets. The section Single-Site Evolutionary Analysis Methods describes methods like these in further detail.

3.3.4. Energy-Based Pocket Detection.

Energy-based methods identify binding sites by evaluating whether or not a given protein region interacts favorably with docked small organic probes. These methods are particularly useful to detect a given pocket *and* assess its druggability. This additional information does come at a cost; energy-based methods are more computationally demanding and so do not scale well, limiting their applicability when analyzing large-scale structural data sets or MD trajectories.

Methods such as SiteMap[201] and SITEHOUND[227] use single probes (typically methane or water molecules) to assess the likelihood of small-molecule/protein binding. In contrast, methods such as FTMap,[202,228,229] GRID,[230] and MCSS[231] consider multiple chemically diverse probes and so identify not only pockets, but also druggable hot spots that are more likely to function as orthosteric or allosteric binding sites.

FTMap[202,228,229,234] in particular warrants further discussion. FTMap is the computational equivalent of the multiple solvent crystal structures (MSCS) technique.[232,233] Rather than superimposing multiple crystal structures obtained by soaking a protein in six to eight different organic solvents, FTMap computationally docks 16 distinct organic probes into a user-provided protein model. The docked positions of these probes are then clustered, and the clusters are ranked by their Boltzmann-averaged energies. In an alternate implementation called FTSite,[287] adjacent, mutually accessible clusters are combined into "sites" that are then ranked by the number of nonbonded contacts between the protein and the corresponding probes. Both these techniques can be applied to multiple structures, whether derived from X-ray crystallography or simulation, to potentially identify cryptic sites that are not always evident when single structures are considered. FTMap can be used to identify possible allosteric sites. For example,

Ivetac et al. used it to detect five potentially druggable sites on β_1AR and β_2AR.[235,236] It has also been used to evaluate the druggability of p53 cryptic pockets, which were subsequently used to prospectively discover novel reactivation compounds.[191]

3.4. Markov State Models

3.4.1. Introduction. Over the past 10 years, Markov state models (MSMs) have been used increasingly to analyze molecular dynamics simulations, suggesting that they are already promising tools to study allostery in a drug-discovery context.[193,237−241] In brief, a Markov state model is a stochastic kinetic model that describes the probability of transitioning between discrete states at a fixed time interval.[237,242,243] These models are required to have the Markovian property (i.e., that the probability of transitioning between discrete states is independent of previous transitions).

By clustering protein structures extracted from an MD trajectory, it is possible to identify discrete conformational states for use in MSMs.[237,242,243] Transitions between conformational clusters observed over the course of an MD trajectory are tallied, and the MSM is then built from the transition probabilities between these distinct clusters. As they are built solely on these transition probabilities, MSMs can draw upon multiple trajectories, enabling efficient sampling of the entire conformational space through many independent simulations carried out in parallel. MSMs can be built with the software package MSMBuilder, available at http://msmbuilder.org/latest/.[242] EMMA,[244] available at http://www.emma-project.org/latest/, is another useful building tool. Because independent trajectories are aggregated as a postprocessing step, the simulations can be carried out on a variety of computational resources, including independent desktop machines, local clusters, or high-performance computing resources. A recent example carried out by Kohlhoff et al. used tens of thousands of trajectories carried out on Google's exascale cloud computing resources. By subsequently "knitting together" the simulation data within a MSM framework, the researchers elucidated the activation pathway of GPCR β_2AR.[245]

MSM/MD analysis provides access to the thermodynamic, kinetic, and structural characteristics of the protein conformational ensemble (i.e., a robust description of the free-energy landscape of the protein).[237,242−244,246,247] The thermodynamics of the various conformational states can be calculated from the equilibrium distribution. It is also possible to resolve the transition kinetics between individual states, the concerted or principal protein motions, metastable states, and the transition pathways between discrete states.[242,244,246] Lastly, the MD trajectories also provide representative receptor snapshots for use in structure-based drug design.[247]

3.4.2. MSMs and Drug Discovery. MD-based MSMs were originally developed to study protein folding,[238,248−253] but they can also be used to study molecular phenomena that are directly applicable to allostery and drug discovery, including cryptic site identification,[190,193,245] protein−ligand interactions,[254,255] and protein function.[51,245,248,256]

MSMs have been used to study binding-site conformational heterogeneity, revealing conformations that are not readily evident in any crystal structure. In 2012, Bowman and co-workers used MSMs to identify and validate cryptic allosteric sites in TEM-1 β-lactamase, interleukin-2, and RNase H.[190,193] Similarly, in a recent study of GPCR conformational transitions, Kohlhoff et al. identified intermediate conformations that they used successfully in a subsequent virtual-screening campaign.[245]

Shukla and co-workers also identified a key conformational intermediate in the Src kinase activation pathway that could be used as a target for drug discovery.[257] While traditional molecular dynamics simulations have been used to identify cryptic allosteric sites and elucidate binding-site conformational diversity, Markov state models can characterize the probability and dynamics of these conformations, permitting rational design against dynamic pockets.

Researchers have also employed MSMs to gain insights into ligand−protein binding that are critical to both allosteric and orthosteric structure-based drug discovery. Buch et al. used MSMs derived from multiple independent MD simulations to accurately determine the k_{on} and binding free energy of benzamidine to trypsin.[254] Building on that work, Tiwary et al. used MSMs in conjunction with metadynamics, a variant of molecular dynamics, to determine the k_{off} for the same system.[255] While this approach is currently too computationally demanding for virtual screening, it has identified transitory sites important for binding and shed new light on the interactions that govern ligand-binding kinetics.

MSMs also provide useful insights into how a protein conformational ensemble begets function and transmits allosteric signals. Most studies have focused on the transition between two functional states. In this context, MSMs have been used to study Src kinase,[257] the β_2 adrenergic receptor (β_2AR),[245] and bacterial response regulator NtrC.[52,256] Malmstrom et al. recently used this same technique to study allostery in the cyclic-nucleotide binding domain (CBD) of protein kinase A.[51] By building and simulating models of the CBD, with and without bound cAMP, the study provided insight into how ligand binding modulates the protein conformational ensemble.

As it is challenging to determine the impact of an effector molecule on the conformational ensemble, most MSM studies to date have not examined allostery directly. Nevertheless, two relevant themes emerge from existing studies. First, kinetics play an important role in the transition between functional states.[51,52,256] This is best shown in our study of the PKA CBD, where the effector molecule cAMP significantly modulates the kinetics of the transition from the active to the inactive states of the protein, while simultaneously leaving the reverse transition between the inactive and active states unmodified.[51] This modulation seems to be regulated by transient interactions between the effector molecule and the protein. Second, recent MSM studies suggest that there is not one but many pathways between functional states. Studies of β_2AR,[245] NtrC,[52,256] and the PKA CBD[51] each revealed multiple transition pathways that arose from transient interactions within the system, suggesting a probabilistic model of allostery in addition to one that describes allostery in terms of concerted pathways.[256] Studies like these can provide useful insights applicable to structure-based drug discovery targeting transitional conformational states.

MSMs can provide significant insights into protein−ligand interactions and allostery, but building these models is computationally demanding, limiting their use as screening tools. This weakness aside, MSMs do provide a strong foundation for rational drug discovery and enhanced understanding of allosteric mechanisms.

4. ENERGY LANDSCAPE AND TOPOLOGICAL ANALYSES

4.1. Introduction

Like MSMs, energy-landscape theories were originally developed to study protein folding. But they too can be applied to the study of allosteric mechanisms and protein dynamical modes. Per statistical mechanics, the underlying free-energy landscape of a system determines the probability of observing a particular configuration. Any topological analysis aimed at understanding allostery must be based on a thermodynamic understanding of the impact that effector binding has on this energy landscape.

4.2. Protein Frustration

It is helpful to view allosteric binding as an event that perturbs the conformational exchange, or folding pathways, of a protein. The principles of statistical mechanics applied to polymers in solution have provided much insight into the protein-folding process. As proteins fold, the number of native contacts increases, and the entropy of the polymer decreases. Folding is impossible when the required conformational transitions are occluded by large free-energy barriers or wells. For example, a large local minimum along the folding pathway could force a protein to undergo a glass transition, leading to a kinetically trapped state. Typical folding must be governed by the "principle of minimum frustration" (i.e., the folding pathway must be funnel shaped and relatively smooth).[258−263]

In terms of allostery and energy-landscape theory, a positive allosteric modulator can be seen as a compound that increases the likelihood of the binding state by deepening the binding-state well or decreasing the depth of competing wells. On the other hand, a negative allosteric modulator may shift the location of the binding-site well, shallow out the well, or increase the depth of competing wells. Using these guiding principles, heuristics can be designed to identify each of the above scenarios.

Ferreiro et al. have developed a heuristic to pick out locally frustrated residues.[264,265] While the overall folding process obeys the principle of minimum frustration, this does not preclude the possibility of having small local barriers that must be overcome to fold. To detect these local frustrations, the authors systematically perturbed the amino acid sequence and compared the perturbed total energy with that of the native state. This comparison provides the effective stabilization energy associated with each amino acid pair. If a given native-state interaction is highly stabilizing (i.e., the interaction energy is favorable compared to mutants), it is said to be "minimally frustrated." On the other hand, if a pair of residues is sufficiently destabilizing (i.e., the interaction energy is weak compared with that of the mutants), it is "frustrated." Frustrated residues can be thought of as weak points in the global structure, contributing to barriers that may introduce slow degrees of freedom. As such, they may be involved in ligand binding or allostery. Jenik et al. have implemented this heuristic as a web server http://www.frustratometer.tk/.[266]

Weinkam et al. used a dual-topology Gō model[267] to survey the free-energy landscape of the calmodulin-GFP Ca²⁺ sensor protein, the maltose binding protein, and the CSL transcription factor. They introduced a truncated Gaussian distance term to the soft-sphere atom overlap term implemented in MODELLER,[268−271] resulting in either one basin corresponding to the bound/unbound distances or two corresponding to both, depending on the pairwise distance and cutoff. Subsequent MD simulations of the resulting model were analyzed using their pseudocorrelation map algorithm.[272] They found that their model reproduces allosteric motions and sheds insight into the macroscopic mechanism.

4.3. Normal-Mode Analysis

Protein dynamics and subsequent perturbations from binding events can also be studied using normal-mode analysis (NMA).[273−275] NMA assumes that the potential energy landscape in the vicinity of a minimized atomic structure is approximately harmonic. This simplifying assumption allows for diagonalization of the Hessian matrix. Solving the eigenvalue problem produces the eigenvectors (movement direction) and eigenvalues (vibrational frequencies). NMA is known to reproduce the slow degrees of freedom of protein motion well.[273,276] Applying additional constraints to simulate the impact of an effector on binding-site residues can yield new fundamental modes that may provide insight into the allosteric mechanism.

Ming and Wall applied normal-mode analysis to investigate the effect of tri-N-acetyl-D-glucosamine upon lysozyme.[277] They assumed that greater perturbation to the conformational distribution corresponds to increased likelihood that a binding event will occur at a regulatory or active site. The degree of perturbation is monitored by comparing the distributions predicted by NMA of the protein with and without the ligand bound, via the Kullback−Leibler divergence. In this work, Ming and Wall generated many arbitrary poses and investigated the allosteric potential, defined by the KLD. They showed that sites with large KLD correspond with identified functional lysozyme residues.

The use of normal-mode analysis removes the need to integrate the equations of motion. However, for the harmonic approximations to be accurate, the starting structure needs to be in a local minima. An initial molecular mechanics minimization can ensure that proteins of interest are near what might be seen in their native environments (e.g., a GPCR embedded in a lipid bilayer with heterogeneous composition). However, alternate options include simplified interaction Hamiltonians such as those offered by elastic network models (ENMs).

4.3.1. Elastic Network Models.

The complexity and computational cost of simulating detailed atomic potentials inspired development of simpler ENMs in which uniform harmonic potentials are used to model all interactions.[278−280] A popular ENM is the Gaussian network model (GNM), wherein neighboring residues are connected by virtual springs to create a network/graph of interactions. Here the interatomic potential energy, U, of the system can be expressed as

$$ U = \frac{\gamma}{2}\left[\sum_{i,j}^{N} \delta R_i\, \Gamma\, \delta R_j\right] \tag{1} $$

where γ is the uniform force constant of the harmonic springs, δR is the N-dimensional column vector of the instantaneous fluctuations of specific atoms, and Γ is the $N \times N$ Kirchoff (connectivity) matrix of residue interactions given by

$$ \Gamma_{ij} = \begin{cases} -1 & \text{if } i \neq j \text{ and } R_{ij} \leq R_c \\ 0 & \text{if } i \neq j \text{ and } R_{ij} > R_c \\ -\sum_{i,i\neq j}^{N} \Gamma_{ij} & \text{if } i = j \end{cases} \tag{2} $$

where R_c is a residue cutoff distance.

Ming and Wall have developed a backbone-enhanced elastic network model (BENM) in which the interactions between connected residues are scaled by an additional factor. Using an objective function defined by the KLD of the marginal distribution of all-atom and C_α atomic coordinates, the authors showed that BENM can reproduce the atomistic mean-squared displacement for bovine pancreatic trypsin inhibitor (BPTI).[281] To consider the effects of allostery, the authors again used the KLD of local conformational distributions before and after ligand binding.[281]

To locate allosteric binding sites, Su et al. recently identified allosterically coupled regions in a GNM using a thermodynamic method.[282] All neighboring residues (represented by their corresponding C_α atoms) were connected by virtual harmonic springs of equivalent strength. A spherical probe then sampled the protein surface to locate the potential effector interaction points. Additional springs were attached between the effector point and the local residues in the protein–effector model. The free-energy difference ($\Delta\Delta G$) between the protein–ligand and protein–effector–ligand systems was calculated. Regions with large $\Delta\Delta G$ values were assumed to be important for the allosteric mechanism. Using this method, Su et al. studied the Hsp70 and the GluA2 AMPA receptors. These allosteric predictions corresponded to experimental results.[282]

Berezovsky et al. used a similar coarse-grained network-interaction model to predict the location of allosteric sites.[283] To identify potential pockets for effector/substrate binding, the authors generated a "residue interaction graph" (RIG), with residues as nodes and interatomic contacts as edges. They then evaluated the "local closeness," defined as the number of total neighboring residues weighted by $1/r$, where r is the inter-residue distance.[284] Regions with high local closeness were likely to be in pockets and, thus, were putative binding sites. The extent to which each binding site was coupled to the intrinsic protein motions was then explored by estimating the strain (so-called "binding leverage") on ligand–protein contacts under deformations described by NMA. A ligand that bound to a site with high binding leverage could restrict protein dynamics.[285]

The binding leverage describes the extent to which changes in binding sites and protein conformational degrees of freedom are coupled; it does not specifically predict allosterically linked sites. To predict allostery, Mitternacht and Berezovsky introduced "leverage coupling," which selects for sites that are strongly coupled to the same degree of freedom.[285] The mode associated with the most binding leverage is considered most relevant, based on the assumption that there is one dominant motion promoting allostery. These methods have been implemented in the online SPACER server http://allostery.bii.a-star.edu.sg/.[286]

5. CONCLUSIONS

Over the past decade, advances in computing power and predictive algorithms coupled with the increased availability of structural and biochemical data have revealed new opportunities for rational design of allosteric drugs. The emergence of novel computational approaches to describe and predict allosteric phenomena across a range of scales, from the coordinated atomic movements in a single receptor molecule to complex allosteric signaling networks, is ushering in a new era wherein computational methods can be used to prospectively predict, discover, and characterize allosteric sites and effector molecules. Within the context of a drug-discovery program, such approaches hold the potential for developing drugs with increased specificity and selectivity, as well as the ability to gain new and more comprehensive understanding of old targets. The convergence of advances in (i) theoretical MSM-based frameworks and MSM building software, (ii) community MD codes that can achieve >100 ns/day sampling for realistic sized systems on single gaming/commodity GPU processors, and (iii) pocket and druggable site-detection algorithms now make it possible for researchers even in industrial settings, with fast-paced timelines and stringent quality standards, to apply these approaches to drug targets already in their arsenals. The application of these methods to kinases and GPCRs seems particularly worthwhile, given the existence of assays and structural data, and the challenges faced by existing drug candidates in the clinic.

AUTHOR INFORMATION

Corresponding Author

*E-mail: ramaro@ucsd.edu.

Author Contributions

J.R.W. and C.T.L. contributed equally to this work.

Notes

The authors declare the following competing financial interest(s): R.E.A. is a co-founder of Actavalon, Inc., a company developing anticancer therapeutics.

Biographies

Jeffrey R. Wagner was born in Glendale, California, and earned a B.A. in chemistry and physics from Claremont McKenna College in 2011. He then worked with Prof. Nagarajan Vaidehi at City of Hope and Dr. Abhi Jain at the Jet Propulsion Laboratory, and is currently a chemistry Ph.D. student at UC San Diego in the lab of Prof. Rommie E. Amaro. His research activities include modeling work in support of drug-design projects, studying binding pocket shapes, and development of data-fusion tools for general use in drug discovery.

Christopher T. Lee received a B.Sc. in Chemistry and a B.A. in Computer Science from the University of Virginia (UVA) in 2011. He also received a M.Sc. degree in chemistry from UVA in 2013, under the supervision of professors Linda Columbus and Cameron Mura, studying enzyme kinetics. He is currently a Ph.D. student in the laboratories of Prof. Rommie E. Amaro and Prof. J. Andrew McCammon, both at UC San Diego. His research focuses on diffusional processes, membrane permeability, and building models of subcellular electron tomographic data.

Jacob D. Durrant received his Ph.D. in biomedical sciences from UC San Diego, under the supervision of Prof. J. Andrew McCammon (2010). He currently works with Prof. Rommie E. Amaro, also at UC San Diego, as a postdoctoral fellow. His research interests include protein dynamics, protein–ligand interactions, mesoscale simulation, and computer-aided drug discovery.

Robert D. Malmstrom was born in Santa Clara, California, but was raised in Texas. He received his B.S. in biochemistry from Brigham Young University in 2002. He earned his Ph.D. in biochemistry from the University of Texas Medical Branch in 2011. He did his postdoctoral studies at UC San Diego with Prof. Rommie E. Amaro. He now serves as the Executive Director of the National Biomedical Computation Resource and is a consultant in computer-aided drug design for the Center for Translational Drug Discovery. He also maintains a research program in computational biophysical chemistry studying the role of protein dynamics and conformational ensembles in protein–ligand interactions and allostery.

Victoria Feher received her Ph.D. in 1996 in chemistry from the University of Oregon's Department of Chemistry and Institute of

Molecular Biology with a research emphasis in structural biophysics. As a graduate student and postdoctoral fellow, she studied protein dynamics related to protein−ligand exchange and allosteric mechanisms of bacterial signaling response regulators, as well as fragment-based drug discovery. She spent her early career in the pharmaceutical industry and is co-inventor of a number of clinical candidates including the MEK1/2 allosteric drug TAK-733. Dr. Feher now is pursuing her research interests in protein dynamics and drug discovery at UC San Diego. She is Co-director of the Drug Design Data Resource and Assistant Director of the Center for Drug Discovery Innovation.

Rommie E. Amaro is an Associate Professor in the Department of Chemistry and Biochemistry at UC San Diego. She received her B.S. in chemical engineering (1999) and Ph.D. in Chemistry (2005) from the University of Illinois at Urbana−Champaign. She was a NIH postdoctoral fellow with Prof. J. Andrew McCammon at UC San Diego. Dr. Amaro is the recipient of an NIH New Innovator Award, the Presidential Early Career Award for Scientists and Engineers, the ACS COMP OpenEye Outstanding Junior Faculty Award, and the ACS Kavli Foundation Emerging Leader in Chemistry National Lecturer (spring 2016). She is Director of the NIH P41 National Biomedical Computation Resource and Co-director of the NIH U01 Drug Design Data Resource. Her research focuses on development and application of state-of-the-art computational methods to address outstanding questions in drug discovery and multiscale molecular and cellular biophysics.

REFERENCES

(1) Christopoulos, A.; May, L.; Avlani, V.; Sexton, P. G-protein-coupled receptor allosterism: the promise and the problem(s). *Biochem. Soc. Trans.* **2004**, *32*, 873−877.

(2) Groebe, D. R. Screening for positive allosteric modulators of biological targets. *Drug Discovery Today* **2006**, *11*, 632−639.

(3) May, L. T.; Leach, K.; Sexton, P. M.; Christopoulos, A. Allosteric Modulation of G Protein-Coupled Receptors. *Annu. Rev. Pharmacol. Toxicol.* **2007**, *47*, 1−51.

(4) Kenakin, T. P. Ligand Detection in the Allosteric World. *J. Biomol. Screening* **2010**, *15*, 119−130.

(5) Nussinov, R.; Tsai, C.-J. The Different Ways through Which Specificity Works in Orthosteric and Allosteric Drugs. *Curr. Pharm. Des.* **2012**, *18*, 1311−1316.

(6) Szilagyi, A.; Nussinov, R.; Csermely, P. Allo-Network Drugs: Extension of the Allosteric Drug Concept to Protein- Protein Interaction and Signaling Networks. *Curr. Top. Med. Chem.* **2013**, *13*, 64−77.

(7) Grover, A. K. Use of Allosteric Targets in the Discovery of Safer Drugs. *Med. Princ. Pract.* **2013**, *22*, 418−426.

(8) Ma, B.; Nussinov, R. Druggable Orthosteric and Allosteric Hot Spots to Target Protein-protein Interactions. *Curr. Pharm. Des.* **2014**, *20*, 1293−1301.

(9) Gunasekaran, K.; Ma, B.; Nussinov, R. Is allostery an intrinsic property of all dynamic proteins? *Proteins: Struct., Funct., Genet.* **2004**, *57*, 433−43.

(10) Wenthur, C. J.; Gentry, P. R.; Mathews, T. P.; Lindsley, C. W. Drugs for Allosteric Sites on Receptors. *Annu. Rev. Pharmacol. Toxicol.* **2014**, *54*, 165−184.

(11) Groebe, D. R. In search of negative allosteric modulators of biological targets. *Drug Discovery Today* **2009**, *14*, 41−49.

(12) Csermely, P.; Nussinov, R.; Szilagyi, A. Editorial (Hot Topic: From Allosteric Drugs to Allo-Network Drugs: State of the Art and Trends of Design, Synthesis and Computational Methods). *Curr. Top. Med. Chem.* **2013**, *13*, 2−4.

(13) Muchmore, S. W.; Hajduk, P. J. Crystallography, NMR and virtual screening: integrated tools for drug discovery. *Curr. Opin. Drug Discovery Dev.* **2003**, *6*, 544−549.

(14) Jorgensen, W. L. The Many Roles of Computation in Drug Discovery. *Science (Washington, DC, U. S.)* **2004**, *303*, 1813−1818.

(15) Doman, T. N.; McGovern, S. L.; Witherbee, B. J.; Kasten, T. P.; Kurumbail, R.; Stallings, W. C.; Connolly, D. T.; Shoichet, B. K. Molecular Docking and High-Throughput Screening for Novel Inhibitors of Protein Tyrosine Phosphatase-1B. *J. Med. Chem.* **2002**, *45*, 2213−2221.

(16) Wood, M. R.; Hopkins, C. R.; Brogan, J. T.; Conn, P. J.; Lindsley, C. W. Molecular Switches on mGluR Allosteric Ligands That Modulate Modes of Pharmacology. *Biochemistry* **2011**, *50*, 2403−2410.

(17) Saalau-Bethell, S. M.; Woodhead, A. J.; Chessari, G.; Carr, M. G.; Coyle, J.; Graham, B.; Hiscock, S. D.; Murray, C. W.; Pathuri, P.; Rich, S. J.; Richardson, C. J.; Williams, P. A.; Jhoti, H. Discovery of an allosteric mechanism for the regulation of HCV NS3 protein function. *Nat. Chem. Biol.* **2012**, *8*, 920−925.

(18) Rawal, R.; Murugesan, V.; Katti, S. Structure-Activity Relationship Studies on Clinically Relevant HIV-1 NNRTIs. *Curr. Med. Chem.* **2012**, *19*, 5364−5380.

(19) Reynolds, C.; de Koning, C. B.; Pelly, S. C.; van Otterlo, W. a. L.; Bode, M. L. In search of a treatment for HIV - current therapies and the role of non-nucleoside reverse transcriptase inhibitors (NNRTIs). *Chem. Soc. Rev.* **2012**, *41*, 4657.

(20) Gentry, P. R.; Sexton, P. M.; Christopoulos, A. Novel Allosteric Modulators of G Protein-coupled Receptors. *J. Biol. Chem.* **2015**, *290*, 19478−19488.

(21) Regan, M. C.; Romero-Hernandez, A.; Furukawa, H. A structural biology perspective on NMDA receptor pharmacology and function. *Curr. Opin. Struct. Biol.* **2015**, *33*, 68−75.

(22) Rutkowski, P.; Lugowska, I.; Kosela-Paterczyk, H.; Kozak, K. Trametinib: a MEK inhibitor for management of metastatic melanoma. *OncoTargets Ther.* **2015**, *8*, 2251.

(23) Meng, H.; McClendon, C. L.; Dai, Z.; Li, K.; Zhang, X.; He, S.; Shang, E.; Liu, Y.; Lai, L. Discovery of Novel 15-Lipoxygenase Activators To Shift the Human Arachidonic Acid Metabolic Network toward Inflammation Resolution. *J. Med. Chem.* **2015**, DOI: 10.1021/acs.jmedchem.5b01011.

(24) Wu, P.; Nielsen, T. E.; Clausen, M. H. FDA-approved small-molecule kinase inhibitors. *Trends Pharmacol. Sci.* **2015**, *36*, 422−439.

(25) Jazayeri, A.; Dias, J. M.; Marshall, F. H. From G Protein-coupled Receptor Structure Resolution to Rational Drug Design. *J. Biol. Chem.* **2015**, *290*, 19489−19495.

(26) Tautermann, C. S. GPCR structures in drug design, emerging opportunities with new structures. *Bioorg. Med. Chem. Lett.* **2014**, *24*, 4073−4079.

(27) Karakas, E.; Furukawa, H. Crystal structure of a heterotetrameric NMDA receptor ion channel. *Science (Washington, DC, U. S.)* **2014**, *344*, 992−997.

(28) Sciara, G.; Mancia, F. Highlights from recently determined structures of membrane proteins: a focus on channels and transporters. *Curr. Opin. Struct. Biol.* **2012**, *22*, 476−481.

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

(30) Weber, G. Ligand binding and internal equilibriums in proteins. *Biochemistry* **1972**, *11*, 864−878.

(31) Cooper, A.; Dryden, D. T. Allostery without conformational change. A plausible model. *Eur. Biophys. J.* **1984**, *11*, 103−9.

(32) Jardetzky, O. Protein dynamics and conformational transitions in allosteric proteins. *Prog. Biophys. Mol. Biol.* **1996**, *65*, 171−219.

(33) Kumar, S.; Ma, B.; Tsai, C.-J.; Sinha, N.; Nussinov, R. Folding and binding cascades: Dynamic landscapes and population shifts. *Protein Sci.* **2000**, *9*, 10−19.

(34) Kern, D.; Zuiderweg, E. R. The role of dynamics in allosteric regulation. *Curr. Opin. Struct. Biol.* **2003**, *13*, 748−757.

(35) Tsai, C.-J.; del Sol, A.; Nussinov, R. Allostery: Absence of a Change in Shape Does Not Imply that Allostery Is Not at Play. *J. Mol. Biol.* **2008**, *378*, 1−11.

(36) Boehr, D. D.; Nussinov, R.; Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **2009**, *5*, 789−96.

(37) Kar, G.; Keskin, O.; Gursoy, A.; Nussinov, R. Allostery and population shift in drug discovery. *Curr. Opin. Pharmacol.* **2010**, *10*, 715−722.

(38) Zhou, H.-X. From Induced Fit to Conformational Selection: A Continuum of Binding Mechanism Controlled by the Timescale of Conformational Transitions. *Biophys. J.* **2010**, *98*, L15−L17.

(39) Nussinov, R.; Ma, B.; Tsai, C. J. Multiple conformational selection and induced fit events take place in allosteric propagation. *Biophys. Chem.* **2014**, *186*, 22−30.

(40) Changeux, J.-P. Allostery and the Monod-Wyman-Changeux Model After 50 Years. *Annu. Rev. Biophys.* **2012**, *41*, 103−133.

(41) Greives, N.; Zhou, H.-X. Both protein dynamics and ligand concentration can shift the binding mechanism between conformational selection and induced fit. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 10197−202.

(42) Guo, J.; Pang, X.; Zhou, H.-X. Two Pathways Mediate Interdomain Allosteric Regulation in Pin1. *Structure* **2015**, *23*, 237−247.

(43) Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J. The ensemble nature of allostery. *Nature* **2014**, *508*, 331−339.

(44) Tsai, C.-J.; del Sol, A.; Nussinov, R. Protein allostery, signal transmission and dynamics: a classification scheme of allosteric mechanisms. *Mol. BioSyst.* **2009**, *5*, 207.

(45) Kornev, A. P.; Taylor, S. S. Dynamics-Driven Allostery in Protein Kinases. *Trends Biochem. Sci.* **2015**, *40*, 628−647.

(46) Ma, B.; Tsai, C.-J.; Haliloğlu, T.; Nussinov, R. Dynamic Allostery: Linkers Are Not Merely Flexible. *Structure* **2011**, *19*, 907−917.

(47) del Sol, A.; Fujihashi, H.; Amoros, D.; Nussinov, R. Residues crucial for maintaining short paths in network communication mediate signaling in proteins. *Mol. Syst. Biol.* **2006**, *2*, 2006.0019.

(48) del Sol, A.; Tsai, C.-J.; Ma, B.; Nussinov, R. The Origin of Allosteric Functional Modulation: Multiple Pre-existing Pathways. *Structure* **2009**, *17*, 1042−1050.

(49) Lockless, S. W. Evolutionarily Conserved Pathways of Energetic Connectivity in Protein Families. *Science (Washington, DC, U. S.)* **1999**, *286*, 295−299.

(50) Süel, G. M.; Lockless, S. W.; Wall, M. A.; Ranganathan, R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat. Struct. Biol.* **2003**, *10*, 59−69.

(51) Malmstrom, R. D.; Kornev, A. P.; Taylor, S. S.; Amaro, R. E. Allostery through the computational microscope: cAMP activation of a canonical signalling domain. *Nat. Commun.* **2015**, *6*, 7588.

(52) Pontiggia, F.; Pachov, D.; Clarkson, M.; Villali, J.; Hagan, M.; Pande, V.; Kern, D. Free energy landscape of activation in a signalling protein at atomic resolution. *Nat. Commun.* **2015**, *6*, 7284.

(53) Hebeisen, P.; et al. Orally active aminopyridines as inhibitors of tetrameric fructose-1,6-bisphosphatase. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 3237−3242.

(54) Zarzycki, M.; Kołodziejczyk, R.; Maciaszczyk-Dziubinska, E.; Wysocki, R.; Jaskolski, M.; Dzugaj, A. Structure of E69Q mutant of human muscle fructose-1,6-bisphosphatase. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 1028−1034.

(55) Nussinov, R.; Tsai, C.-J.; Csermely, P. Allo-network drugs: harnessing allostery in cellular networks. *Trends Pharmacol. Sci.* **2011**, *32*, 686−693.

(56) Gilmartin, A. G.; Bleam, M. R.; Groy, A.; Moss, K. G.; Minthorn, E. A.; Kulkarni, S. G.; Rominger, C. M.; Erskine, S.; Fisher, K. E.; Yang, J.; Zappacosta, F.; Annan, R.; Sutton, D.; Laquerre, S. G. GSK1120212 (JTP-74057) is an inhibitor of MEK activity and activation with favorable pharmacokinetic properties for sustained in vivo pathway inhibition. *Clin. Cancer Res.* **2011**, *17*, 989−1000.

(57) Sebolt-Leopold, J. S.; Dudley, D. T.; Herrera, R.; Van Becelaere, K.; Wiland, a.; Gowan, R. C.; Tecle, H.; Barrett, S. D.; Bridges, a.; Przybranowski, S.; Leopold, W. R.; Saltiel, a. R. Blockade of the MAP kinase pathway suppresses growth of colon tumors in vivo. *Nat. Med.* **1999**, *5*, 810−816.

(58) Ohren, J. F.; et al. Structures of human MAP kinase kinase 1 (MEK1) and MEK2 describe novel noncompetitive kinase inhibition. *Nat. Struct. Mol. Biol.* **2004**, *11*, 1192−1197.

(59) Zhao, Y.; Adjei, A. A. The clinical development of MEK inhibitors. *Nat. Rev. Clin. Oncol.* **2014**, *11*, 385−400.

(60) Hackos, D. H.; Lupardus, P. J.; Grand, T.; Chen, Y.; Wang, T.-m.; Reynen, P.; Gustafson, A.; Wallweber, H. J. A.; Volgraf, M.; Sellers, B. D.; Schwarz, J. B.; Paoletti, P.; Sheng, M.; Zhou, Q.; Hanson, J. E. Positive Allosteric Modulators of GluN2A-Containing NMDARs with Distinct Modes of Action and Impacts on Circuit Function. *Neuron* **2016**, *89*, 983−999.

(61) Talavera, D.; Lovell, S. C.; Whelan, S. Covariation Is a Poor Measure of Molecular Coevolution. *Mol. Biol. Evol.* **2015**, *32*, 2456−2468.

(62) Fodor, A. A.; Aldrich, R. W. On Evolutionary Conservation of Thermodynamic Coupling in Proteins. *J. Biol. Chem.* **2004**, *279*, 19046−19050.

(63) Fodor, A. A.; Aldrich, R. W. Influence of conservation on calculations of amino acid covariance in multiple sequence alignments. *Proteins: Struct., Funct., Genet.* **2004**, *56*, 211−221.

(64) Johansson, F.; Toh, H. Relative von Neumann entropy for evaluating amino acid conservation. *J. Bioinf. Comput. Biol.* **2010**, *08*, 809−823.

(65) de Juan, D.; Pazos, F.; Valencia, A. Emerging methods in protein co-evolution. *Nat. Rev. Genet.* **2013**, *14*, 249−261.

(66) Shannon, C. E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379−423.

(67) Shenkin, P. S.; Erman, B.; Mastrandrea, L. D. Information-theoretical entropy as a measure of sequence variability. *Proteins: Struct., Funct., Genet.* **1991**, *11*, 297−313.

(68) Vinga, S. Information theory applications for biological sequence analysis. *Briefings Bioinf.* **2014**, *15*, 376−389.

(69) Kullback, S.; Leibler, R. A. On Information and Sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79−86.

(70) Zhang, S.-W.; Zhang, Y.-L.; Pan, Q.; Cheng, Y.-M.; Chou, K.-C. Estimating residue evolutionary conservation by introducing von Neumann entropy and a novel gap-treating approach. *Amino Acids* **2008**, *35*, 495−501.

(71) Nussinov, R.; Tsai, C.-J. Allostery in Disease and in Drug Discovery. *Cell* **2013**, *153*, 293−305.

(72) Lichtarge, O.; Bourne, H. R.; Cohen, F. E. An evolutionary trace method defines binding surfaces common to protein families. *J. Mol. Biol.* **1996**, *257*, 342−58.

(73) Madabushi, S.; Gross, A. K.; Philippi, A.; Meng, E. C.; Wensel, T. G.; Lichtarge, O. Evolutionary Trace of G Protein-coupled Receptors Reveals Clusters of Residues That Determine Global and Class-specific Functions. *J. Biol. Chem.* **2004**, *279*, 8126−8132.

(74) Raviscioni, M.; He, Q.; Salicru, E. M.; Smith, C. L.; Lichtarge, O. Evolutionary identification of a subtype specific functional site in the ligand binding domain of steroid receptors. *Proteins: Struct., Funct., Genet.* **2006**, *64*, 1046−1057.

(75) Rodriguez, G. J.; Yao, R.; Lichtarge, O.; Wensel, T. G. Evolution-guided discovery and recoding of allosteric pathway specificity determinants in psychoactive bioamine receptors. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 7787−7792.

(76) Mihalek, I.; Reš, I.; Lichtarge, O. A Family of Evolution-Entropy Hybrid Methods for Ranking Protein Residues by Importance. *J. Mol. Biol.* **2004**, *336*, 1265−1282.

(77) Mihalek, I.; Res, I.; Lichtarge, O. Evolutionary trace report_-maker: a new type of service for comparative analysis of proteins. *Bioinformatics* **2006**, *22*, 1656−1657.

(78) Merkl, R.; Zwick, M. H2r: Identification of evolutionary important residues by means of an entropy based analysis of multiple sequence alignments. *BMC Bioinf.* **2008**, *9*, 151.

(79) Janda, J.-O.; Popal, A.; Bauer, J.; Busch, M.; Klocke, M.; Spitzer, W.; Keller, J.; Merkl, R. H2rs: Deducing evolutionary and functionally important residue positions by means of an entropy and similarity based analysis of multiple sequence alignments. *BMC Bioinf.* **2014**, *15*, 118.

(80) Pelé, J.; Moreau, M.; Abdi, H.; Rodien, P.; Castel, H.; Chabbert, M. Comparative analysis of sequence covariation methods to mine evolutionary hubs: Examples from selected GPCR families. *Proteins: Struct., Funct., Genet.* **2014**, *82*, 2141−2156.

(81) Brown, C. A.; Brown, K. S. Validation of Coevolving Residue Algorithms via Pipeline Sensitivity Analysis: ELSC and OMES and ZNMI, Oh My! *PLoS One* **2010**, *5*, e10779.

(82) Simonetti, F. L.; Teppa, E.; Chernomoretz, A.; Nielsen, M.; Marino Buslje, C. MISTIC: mutual information server to infer coevolution. *Nucleic Acids Res.* **2013**, *41*, W8−W14.

(83) Horner, D. S.; Pirovano, W.; Pesole, G. Correlated substitution analysis and the prediction of amino acid structural contacts. *Briefings Bioinf.* **2007**, *9*, 46−56.

(84) Gao, H.; Dou, Y.; Yang, J.; Wang, J. New methods to measure residues coevolution in proteins. *BMC Bioinf.* **2011**, *12*, 206.

(85) Dunn, S.; Wahl, L.; Gloor, G. Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction. *Bioinformatics* **2008**, *24*, 333−340.

(86) Wollenberg, K. R.; Atchley, W. R. Separation of phylogenetic and functional associations in biological sequences by using the parametric bootstrap. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 3288−3291.

(87) Buslje, C. M.; Santos, J.; Delfino, J. M.; Nielsen, M. Correction for phylogeny, small number of observations and data redundancy improves the identification of coevolving amino acid pairs using mutual information. *Bioinformatics* **2009**, *25*, 1125−1131.

(88) Clark, G. W.; Ackerman, S. H.; Tillier, E. R.; Gatti, D. L. Multidimensional mutual information methods for the analysis of covariation in multiple sequence alignments. *BMC Bioinf.* **2014**, *15*, 157.

(89) Martin, L. C.; Gloor, G. B.; Dunn, S. D.; Wahl, L. M. Using information theory to search for co-evolving residues in proteins. *Bioinformatics* **2005**, *21*, 4116−4124.

(90) Crooks, G. E.; Wolfe, J.; Brenner, S. E. Measurements of protein sequence-structure correlations. *Proteins: Struct., Funct., Genet.* **2004**, *57*, 804−810.

(91) Olmea, O.; Valencia, A. Improving contact predictions by the combination of correlated mutations and other sources of sequence information. *Folding Des.* **1997**, *2*, S25−S32.

(92) McLachlan, A. Tests for comparing related amino-acid sequences. Cytochrome c and cytochrome c551. *J. Mol. Biol.* **1971**, *61*, 409−424.

(93) Halperin, I.; Wolfson, H.; Nussinov, R. Correlated mutations: Advances and limitations. A study on fusion proteins and on the Cohesin-Dockerin families. *Proteins: Struct., Funct., Genet.* **2006**, *63*, 832−845.

(94) Kass, I.; Horovitz, A. Mapping pathways of allosteric communication in GroEL by analysis of correlated mutations. *Proteins: Struct., Funct., Genet.* **2002**, *48*, 611−617.

(95) Ackerman, S. H.; Tillier, E. R.; Gatti, D. L. Accurate Simulation and Detection of Coevolution Signals in Multiple Sequence Alignments. *PLoS One* **2012**, *7*, e47108.

(96) Ranganathan, R.; Rivoire, O. *Note 109: A Summary of SCA Calculations*; 2012; pp 1−11.

(97) Halabi, N.; Rivoire, O.; Leibler, S.; Ranganathan, R. Protein Sectors: Evolutionary Units of Three-Dimensional Structure. *Cell* **2009**, *138*, 774−786.

(98) Reynolds, K. A.; McLaughlin, R. N.; Ranganathan, R. Hot Spots for Allosteric Regulation on Protein Surfaces. *Cell* **2011**, *147*, 1564−1575.

(99) Russ, W. P.; Lowery, D. M.; Mishra, P.; Yaffe, M. B.; Ranganathan, R. Natural-like function in artificial WW domains. *Nature* **2005**, *437*, 579−583.

(100) Socolich, M.; Lockless, S. W.; Russ, W. P.; Lee, H.; Gardner, K. H.; Ranganathan, R. Evolutionary information for specifying a protein fold. *Nature* **2005**, *437*, 512−518.

(101) Baths, V.; Roy, U. Identification of distant co-evolving residues in antigen 85C from Mycobacterium tuberculosis using statistical coupling analysis of the esterase family proteins. *J. Biomed. Res.* **2011**, *25*, 165−169.

(102) McLaughlin, R. N., Jr.; Poelwijk, F. J.; Raman, A.; Gosal, W. S.; Ranganathan, R. The spatial architecture of protein function and adaptation. *Nature* **2012**, *491*, 138−142.

(103) Peterson, F. C.; Penkert, R. R.; Volkman, B. F.; Prehoda, K. E. Cdc42 Regulates the Par-6 PDZ Domain through an Allosteric CRIB-PDZ Transition. *Mol. Cell* **2004**, *13*, 665−676.

(104) Hatley, M. E.; Lockless, S. W.; Gibson, S. K.; Gilman, A. G.; Ranganathan, R. Allosteric determinants in guanine nucleotide-binding proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 14445−14450.

(105) Shulman, A. I.; Larson, C.; Mangelsdorf, D. J.; Ranganathan, R. Structural Determinants of Allosteric Ligand Activation in RXR Heterodimers. *Cell* **2004**, *116*, 417−429.

(106) Smock, R. G.; Rivoire, O.; Russ, W. P.; Swain, J. F.; Leibler, S.; Ranganathan, R.; Gierasch, L. M. An interdomain sector mediating allostery in Hsp70 molecular chaperones. *Mol. Syst. Biol.* **2010**, *6*, 414.

(107) Chi, C. N.; Elfstrom, L.; Shi, Y.; Snall, T.; Engstrom, A.; Jemth, P. Reassessing a sparse energetic network within a single protein domain. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 4679−4684.

(108) Fuentes, E. J.; Der, C. J.; Lee, A. L. Ligand-dependent dynamics and intramolecular signaling in a PDZ domain. *J. Mol. Biol.* **2004**, *335*, 1105−15.

(109) Dima, R. I.; Thirumalai, D. Proteins associated with diseases show enhanced sequence correlation between charged residues. *Bioinformatics* **2004**, *20*, 2345−2354.

(110) Dima, R. I. Determination of network of residues that regulate allostery in protein families using sequence analysis. *Protein Sci.* **2006**, *15*, 258−268.

(111) O'Donovan, C. High-quality protein knowledge resource: SWISS-PROT and TrEMBL. *Briefings Bioinf.* **2002**, *3*, 275−284.

(112) Boeckmann, B. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **2003**, *31*, 365−370.

(113) Getz, G.; Levine, E.; Domany, E. Coupled two-way clustering analysis of gene microarray data. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 12079−12084.

(114) Dekker, J. P.; Fodor, A.; Aldrich, R. W.; Yellen, G. A perturbation-based method for calculating explicit likelihood of evolutionary co-variance in multiple sequence alignments. *Bioinformatics* **2004**, *20*, 1565−1572.

(115) Yip, K. Y.; Patel, P.; Kim, P. M.; Engelman, D. M.; McDermott, D.; Gerstein, M. An integrated system for studying residue coevolution in proteins. *Bioinformatics* **2008**, *24*, 290−292.

(116) Kozma, D.; Simon, I.; Tusnady, G. E. CMWeb: an interactive on-line tool for analysing residue-residue contacts and contact prediction methods. *Nucleic Acids Res.* **2012**, *40*, W329−W333.

(117) Weigt, M.; White, R. A.; Szurmant, H.; Hoch, J. A.; Hwa, T. Identification of direct residue contacts in protein-protein interaction by message passing. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 67−72.

(118) Morcos, F.; Pagnani, A.; Lunt, B.; Bertolino, A.; Marks, D. S.; Sander, C.; Zecchina, R.; Onuchic, J. N.; Hwa, T.; Weigt, M. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, E1293−E1301.

(119) Schug, A.; Weigt, M.; Onuchic, J. N.; Hwa, T.; Szurmant, H. High-resolution protein complexes from integrating genomic information with molecular simulation. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 22124−22129.

(120) Sulkowska, J. I.; Morcos, F.; Weigt, M.; Hwa, T.; Onuchic, J. N. Genomics-aided structure prediction. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 10340−10345.

(121) Szurmant, H.; Hoch, J. A. Statistical analyses of protein sequence alignments identify structures and mechanisms in signal activation of sensor histidine kinases. *Mol. Microbiol.* **2013**, *87*, 707−712.

(122) Procaccini, A.; Lunt, B.; Szurmant, H.; Hwa, T.; Weigt, M. Dissecting the Specificity of Protein-Protein Interaction in Bacterial Two-Component Signaling: Orphans and Crosstalks. *PLoS One* **2011**, *6*, e19729.

(123) Morcos, F.; Jana, B.; Hwa, T.; Onuchic, J. N. Coevolutionary signals across protein lineages help capture multiple protein conformations. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 20533−20538.

(124) Jones, D. T.; Buchan, D. W. A.; Cozzetto, D.; Pontil, M. PSICOV: precise structural contact prediction using sparse inverse covariance estimation on large multiple sequence alignments. *Bioinformatics* **2012**, *28*, 184−190.

(125) Eckmann, J.-P.; Kamphorst, S. O.; Ruelle, D. Recurrence Plots of Dynamical Systems. *Europhys. Lett.* **1987**, *4*, 973−977.

(126) Zbilut, J. P.; Giuliani, A.; Webber, C. L.; Colosimo, A. Recurrence quantification analysis in structure-function relationships of proteins: an overview of a general methodology applied to the case of TEM-1 beta-lactamase. *Protein Eng., Des. Sel.* **1998**, *11*, 87−93.

(127) Namboodiri, S.; Verma, C.; Dhar, P. K.; Giuliani, A.; Nair, A. S. Sequence signatures of allosteric proteins towards rational design. *Syst. Synth. Biol.* **2010**, *4*, 271−280.

(128) Porrello, A.; Soddu, S.; Zbilut, J. P.; Crescenzi, M.; Giuliani, A. Discrimination of single amino acid mutations of the p53 protein by means of deterministic singularities of recurrence quantification analysis. *Proteins: Struct., Funct., Genet.* **2004**, *55*, 743−755.

(129) Giuliani, A.; Tomasi, M. Recurrence quantification analysis reveals interaction partners in paramyxoviridae envelope glycoproteins. *Proteins: Struct., Funct., Genet.* **2002**, *46*, 171−176.

(130) Colafranceschi, M.; Colosimo, A.; Zbilut, J. P.; Uversky, V. N.; Giuliani, A. Structure-related statistical singularities along protein sequences: A correlation study. *J. Chem. Inf. Model.* **2005**, *45*, 183−189.

(131) Namboodiri, S.; Giuliani, A.; Nair, A. S.; Dhar, P. K. Looking for a sequence based allostery definition: a statistical journey at different resolution scales. *J. Theor. Biol.* **2012**, *304*, 211−8.

(132) Zbilut, J. P.; Sirabella, P.; Giuliani, A.; Manetti, C.; Colosimo, A.; Webber, C. L. Review of Nonlinear Analysis of Proteins Through Recurrence Quantification. *Cell Biochem. Biophys.* **2002**, *36*, 67−88.

(133) Rivoire, O. Elements of Coevolution in Biological Sequences. *Phys. Rev. Lett.* **2013**, *110*, 178102.

(134) Mao, W.; Kaya, C.; Dutta, A.; Horovitz, A.; Bahar, I. Comparative study of the effectiveness and limitations of current methods for detecting sequence coevolution. *Bioinformatics* **2015**, *31*, 1929−1937.

(135) Bakan, A.; Meireles, L. M.; Bahar, I. ProDy: Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics* **2011**, *27*, 1575−1577.

(136) Göbel, U.; Sander, C.; Schneider, R.; Valencia, A. Correlated mutations and residue contacts in proteins. *Proteins: Struct., Funct., Genet.* **1994**, *18*, 309−317.

(137) Fares, M. A. A Novel Method for Detecting Intramolecular Coevolution: Adding a Further Dimension to Selective Constraints Analyses. *Genetics* **2006**, *173*, 9−23.

(138) Fares, M. A.; McNally, D. CAPS: coevolution analysis using protein sequences. *Bioinformatics* **2006**, *22*, 2821−2822.

(139) Iserte, J.; Simonetti, F. L.; Zea, D. J.; Teppa, E.; Marino-Buslje, C. I-COMS: Interprotein-COrrelated Mutations Server. *Nucleic Acids Res.* **2015**, *43*, W320−W325.

(140) Jeong, C.-S.; Kim, D. Reliable and robust detection of coevolving protein residues. *Protein Eng., Des. Sel.* **2012**, *25*, 705−713.

(141) Berezin, C.; Glaser, F.; Rosenberg, J.; Paz, I.; Pupko, T.; Fariselli, P.; Casadio, R.; Ben-Tal, N. ConSeq: the identification of functionally and structurally important residues in protein sequences. *Bioinformatics* **2004**, *20*, 1322−1324.

(142) Glaser, F.; Pupko, T.; Paz, I.; Bell, R. E.; Bechor-Shental, D.; Martz, E.; Ben-Tal, N. ConSurf: Identification of Functional Regions in Proteins by Surface-Mapping of Phylogenetic Information. *Bioinformatics* **2003**, *19*, 163−164.

(143) Landau, M.; Mayrose, I.; Rosenberg, Y.; Glaser, F.; Martz, E.; Pupko, T.; Ben-Tal, N. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **2005**, *33*, W299−W302.

(144) Armon, A.; Graur, D.; Ben-Tal, N. ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. *J. Mol. Biol.* **2001**, *307*, 447−63.

(145) Ashkenazy, H.; Erez, E.; Martz, E.; Pupko, T.; Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **2010**, *38*, W529−W533.

(146) Celniker, G.; Nimrod, G.; Ashkenazy, H.; Glaser, F.; Martz, E.; Mayrose, I.; Pupko, T.; Ben-Tal, N. ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function. *Isr. J. Chem.* **2013**, *53*, 199−206.

(147) Liu, Y.; Gierasch, L. M.; Bahar, I. Role of Hsp70 ATPase Domain Intrinsic Dynamics and Sequence Evolution in Enabling its Functional Interactions with NEFs. *PLoS Comput. Biol.* **2010**, *6*, e1000931.

(148) Liu, Y.; Bahar, I. Sequence Evolution Correlates with Structural Dynamics. *Mol. Biol. Evol.* **2012**, *29*, 2253−2263.

(149) Skjærven, L.; Yao, X.-Q.; Scarabelli, G.; Grant, B. J. Integrating protein structural dynamics and evolutionary analysis with Bio3D. *BMC Bioinf.* **2014**, *15*, 399.

(150) Grant, B. J.; Rodrigues, A. P. C.; ElSawy, K. M.; McCammon, J. A.; Caves, L. S. D. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **2006**, *22*, 2695−2696.

(151) Li, Z.; Huang, Y.; Ouyang, Y.; Jiao, Y.; Xing, H.; Liao, L.; Jiang, S.; Shao, Y.; Ma, L. CorMut: an R/Bioconductor package for computing correlated mutations based on selection pressure. *Bioinformatics* **2014**, *30*, 2073−2075.

(152) Rhee, S.-Y.; Liu, T. F.; Holmes, S. P.; Shafer, R. W. HIV-1 Subtype B Protease and Reverse Transcriptase Amino Acid Covariation. *PLoS Comput. Biol.* **2007**, *3*, e87.

(153) Hurst, L. D. The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet.* **2002**, *18*, 486−487.

(154) Alder, B. J.; Wainwright, T. E. Studies in Molecular Dynamics. I. General Method. *J. Chem. Phys.* **1959**, *31*, 459.

(155) Rahman, A. Correlations in the Motion of Atoms in Liquid Argon. *Phys. Rev.* **1964**, *136*, A405−A411.

(156) Karplus, M.; McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **2002**, *9*, 646−652.

(157) Adcock, S. A.; McCammon, J. A. Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem. Rev.* **2006**, *106*, 1589−615.

(158) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J. Chem. Phys.* **2004**, *120*, 11919.

(159) Wang, Y.; Harrison, C. B.; Schulten, K.; McCammon, J. A. Implementation of accelerated molecular dynamics in NAMD. *Comput. Sci. Discovery* **2011**, *4*, 015002.

(160) Miao, Y.; Feher, V. A.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theory Comput.* **2015**, *11*, 3584−3595.

(161) Miao, Y.; Feixas, F.; Eun, C.; McCammon, J. A. Accelerated molecular dynamics simulations of protein folding. *J. Comput. Chem.* **2015**, *36*, 1536−1549.

(162) Pierce, L. C.; Salomon-Ferrer, R.; Augusto F. de Oliveira, C.; McCammon, J. A.; Walker, R. C. Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 2997−3002.

(163) Lindert, S.; Bucher, D.; Eastman, P.; Pande, V.; McCammon, J. A. Accelerated Molecular Dynamics Simulations with the AMOEBA Polarizable Force Field on Graphics Processing Units. *J. Chem. Theory Comput.* **2013**, *9*, 4684−4691.

(164) Amaro, R. E.; Baron, R.; McCammon, J. A. An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 693−705.

(165) Amaro, R.; Li, W. Emerging Methods for Ensemble-Based Virtual Screening. *Curr. Top. Med. Chem.* **2010**, *10*, 3−13.

(166) Durrant, J. D.; McCammon, J. A. Molecular dynamics simulations and drug discovery. *BMC Biol.* **2011**, *9*, 71.

(167) Nichols, S. E.; Swift, R. V.; Amaro, R. E. Rational Prediction with Molecular Dynamics for Hit Identification. *Curr. Top. Med. Chem.* **2012**, *12*, 2002−2012.

(168) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **1953**, *21*, 1087.

(169) Hastings, W. K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **1970**, *57*, 97−109.

(170) Landau, D. P.; Binder, K. *A Guide to Monte Carlo Simulations in Statistical Physics*; Cambridge University Press: Cambridge, U.K., 2009.

(171) Zwier, M. C.; Chong, L. T. Reaching biological timescales with all-atom molecular dynamics simulations. *Curr. Opin. Pharmacol.* **2010**, *10*, 745−752.

(172) Noé, F. Beating the Millisecond Barrier in Molecular Dynamics Simulations. *Biophys. J.* **2015**, *108*, 228−229.

(173) Sethi, A.; Eargle, J.; Black, A. A.; Luthey-Schulten, Z. Dynamical networks in tRNA:protein complexes. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 6620−6625.

(174) Gasper, P. M.; Fuglestad, B.; Komives, E. A.; Markwick, P. R. L.; McCammon, J. A. Allosteric networks in thrombin distinguish procoagulant vs. anticoagulant activities. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 21216−22.

(175) Freddolino, P. L.; Gardner, K. H.; Schulten, K. Signaling mechanisms of LOV domains: new insights from molecular dynamics studies. *Photochem. Photobiol. Sci.* **2013**, *12*, 1158.

(176) Yan, W.; Zhou, J.; Sun, M.; Chen, J.; Hu, G.; Shen, B. The construction of an amino acid network for understanding protein structure and function. *Amino Acids* **2014**, *46*, 1419−1439.

(177) Van Wart, A. T.; Eargle, J.; Luthey-Schulten, Z.; Amaro, R. E. Exploring Residue Component Contributions to Dynamical Network Models of Allostery. *J. Chem. Theory Comput.* **2012**, *8*, 2949−2961.

(178) Ghosh, A.; Vishveshwara, S. A study of communication pathways in methionyl- tRNA synthetase by molecular dynamics simulations and structure network analysis. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 15711−15716.

(179) Collier, G.; Ortiz, V. Emerging computational approaches for the study of protein allostery. *Arch. Biochem. Biophys.* **2013**, *538*, 6−15.

(180) McClendon, C. L.; Friedland, G.; Mobley, D. L.; Amirkhani, H.; Jacobson, M. P. Quantifying Correlations Between Allosteric Sites in Thermodynamic Ensembles. *J. Chem. Theory Comput.* **2009**, *5*, 2486−2502.

(181) McClendon, C. L.; Hua, L.; Barreiro, G.; Jacobson, M. P. Comparing conformational ensembles Using the Kullback-Leibler divergence expansion. *J. Chem. Theory Comput.* **2012**, *8*, 2115−2126.

(182) Bradley, M. J.; Chivers, P. T.; Baker, N. A. Molecular Dynamics Simulation of the Escherichia coli NikR Protein: Equilibrium Conformational Fluctuations Reveal Interdomain Allosteric Communication Pathways. *J. Mol. Biol.* **2008**, *378*, 1155−1173.

(183) Atilgan, C.; Okan, O. B.; Atilgan, A. R. Network-Based Models as Tools Hinting at Nonevident Protein Functionality. *Annu. Rev. Biophys.* **2012**, *41*, 205−225.

(184) Van Wart, A. T.; Durrant, J.; Votapka, L.; Amaro, R. E. Weighted Implementation of Suboptimal Paths aWISPa: An Optimized Algorithm and Tool for Dynamical Network Analysis. *J. Chem. Theory Comput.* **2014**, *10*, 511−517.

(185) LeVine, M. V.; Weinstein, H. NbIT - A New Information Theory-Based Analysis of Allosteric Mechanisms Reveals Residues that Underlie Function in the Leucine Transporter LeuT. *PLoS Comput. Biol.* **2014**, *10*, e1003603.

(186) Rivalta, I.; Sultan, M. M.; Lee, N.-S.; Manley, G. A.; Loria, J. P.; Batista, V. S. Allosteric pathways in imidazole glycerol phosphate synthase. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, E1428.

(187) Miao, Y.; Nichols, S. E.; Gasper, P. M.; Metzger, V. T.; McCammon, J. A. Activation and dynamic network of the M2 muscarinic receptor. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 10982−7.

(188) Blacklock, K.; Verkhivker, G. M. Computational modeling of allosteric regulation in the hsp90 chaperones: a statistical ensemble

(189) Frembgen-Kesner, T.; Elcock, A. H. Computational Sampling of a Cryptic Drug Binding Site in a Protein Receptor: Explicit Solvent Molecular Dynamics and Inhibitor Docking to p38 MAP Kinase. *J. Mol. Biol.* **2006**, *359*, 202−214.

(190) 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.

(191) Wassman, C. D.; Baronio, R.; Demir, Ö; Wallentine, B. D.; Chen, C.-K.; Hall, L. V.; Salehi, F.; Lin, D.-W.; Chung, B. P.; Hatfield, G. W.; Chamberlin, A. C.; Luecke, H.; Lathrop, R. H.; Kaiser, P.; Amaro, R. E. Computational identification of a transiently open L1/S3 pocket for reactivation of mutant p53. *Nat. Commun.* **2013**, *4*, 1407.

(192) Nair, P. C.; Miners, J. O. Molecular dynamics simulations: from structure function relationships to drug discovery. *Silico Pharmacol.* **2014**, *2*, 4.

(193) 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.

(194) Durrant, J. D.; De Oliveira, C. A. F.; McCammon, J. A. POVME: An algorithm for measuring binding-pocket volumes. *J. Mol. Graphics Modell.* **2011**, *29*, 773−776.

(195) Chovancova, E.; Pavelka, A.; Benes, P.; Strnad, O.; Brezovsky, J.; Kozlikova, B.; Gora, A.; Sustr, V.; Klvana, M.; Medek, P.; Biedermannova, L.; Sochor, J.; Damborsky, J. CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures. *PLoS Comput. Biol.* **2012**, *8*, e1002708.

(196) Eyrisch, S.; Helms, V. Transient Pockets on Protein Surfaces Involved in Protein-Protein Interaction. *J. Med. Chem.* **2007**, *50*, 3457−3464.

(197) Brady, G. P., Jr.; Stouten, P. F. Fast prediction and visualization of protein binding pockets with PASS. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 383−401.

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

(199) Schmidtke, P.; Bidon-Chanal, A.; Luque, F. J.; Barril, X. MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics* **2011**, *27*, 3276−3285.

(200) Halgren, T. New Method for Fast and Accurate Binding-site Identification and Analysis. *Chem. Biol. Drug Des.* **2007**, *69*, 146−148.

(201) Halgren, T. A. Identifying and Characterizing Binding Sites and Assessing Druggability. *J. Chem. Inf. Model.* **2009**, *49*, 377−389.

(202) Brenke, R.; Kozakov, D.; Chuang, G.-Y.; Beglov, D.; Hall, D.; Landon, M. R.; Mattos, C.; Vajda, S. Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics* **2009**, *25*, 621−627.

(203) Votapka, L.; Amaro, R. E. Multistructural hot spot characterization with FTProd. *Bioinformatics* **2013**, *29*, 393−394.

(204) Zheng, X.; Gan, L.; Wang, E.; Wang, J. Pocket-Based Drug Design: Exploring Pocket Space. *AAPS J.* **2013**, *15*, 228−241.

(205) Edelsbrunner, H.; Mücke, E. P. Three-dimensional alpha shapes. *ACM Trans. Graph.* **1994**, *13*, 43−72.

(206) Kawabata, T.; Go, N. Detection of pockets on protein surfaces using small and large probe spheres to find putative ligand binding sites. *Proteins: Struct., Funct., Genet.* **2007**, *68*, 516−529.

(207) Yu, J.; Zhou, Y.; Tanaka, I.; Yao, M. Roll: a new algorithm for the detection of protein pockets and cavities with a rolling probe sphere. *Bioinformatics* **2010**, *26*, 46−52.

(208) Laskowski, R. A. SURFNET: A program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J. Mol. Graphics* **1995**, *13*, 323−330.

(209) Peters, K. P.; Fauck, J.; Frömmel, C. The Automatic Search for Ligand Binding Sites in Proteins of Known Three-dimensional Structure Using only Geometric Criteria. *J. Mol. Biol.* **1996**, *256*, 201−213.

(210) Liang, J.; Woodward, C.; Edelsbrunner, H. Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design. *Protein Sci.* **1998**, *7*, 1884−1897.

(211) Binkowski, T. A. CASTp: Computed Atlas of Surface Topography of proteins. *Nucleic Acids Res.* **2003**, *31*, 3352−3355.

(212) Edelsbrunner, H.; Kirkpatrick, D.; Seidel, R. On the shape of a set of points in the plane. *IEEE Trans. Inf. Theory* **1983**, *29*, 551−559.

(213) Kawabata, T. Detection of multiscale pockets on protein surfaces using mathematical morphology. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 1195−1211.

(214) Hendlich, M.; Rippmann, F.; Barnickel, G. LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. *J. Mol. Graphics Modell.* **1997**, *15*, 359−363.

(215) Huang, B.; Schroeder, M. LIGSITEcsc: predicting ligand binding sites using the Connolly surface and degree of conservation. *BMC Struct. Biol.* **2006**, *6*, 19.

(216) Levitt, D. G.; Banaszak, L. J. POCKET: a computer graphics method for identifying and displaying protein cavities and their surrounding amino acids. *J. Mol. Graphics* **1992**, *10*, 229−34.

(217) Weisel, M.; Proschak, E.; Schneider, G. PocketPicker: analysis of ligand binding-sites with shape descriptors. *Chem. Cent. J.* **2007**, *1*, 7.

(218) An, J. Pocketome via Comprehensive Identification and Classification of Ligand Binding Envelopes. *Mol. Cell. Proteomics* **2005**, *4*, 752−761.

(219) Laurie, A. T. R.; Jackson, R. M. Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics* **2005**, *21*, 1908−1916.

(220) Durrant, J. D.; Votapka, L.; Sørensen, J.; Amaro, R. E. POVME 2.0: An Enhanced Tool for Determining Pocket Shape and Volume Characteristics. *J. Chem. Theory Comput.* **2014**, *10*, 5047−5056.

(221) Wass, M. N.; Kelley, L. A.; Sternberg, M. J. E. 3DLigandSite: predicting ligand-binding sites using similar structures. *Nucleic Acids Res.* **2010**, *38*, W469−W473.

(222) Brylinski, M.; Skolnick, J. A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 129−134.

(223) Kufareva, I.; Ilatovskiy, A. V.; Abagyan, R. Pocketome: an encyclopedia of small-molecule binding sites in 4D. *Nucleic Acids Res.* **2012**, *40*, D535−D540.

(224) Fischer, J. D.; Mayer, C. E.; Soding, J. Prediction of protein functional residues from sequence by probability density estimation. *Bioinformatics* **2008**, *24*, 613−620.

(225) Lichtarge, O.; Sowa, M. E. Evolutionary predictions of binding surfaces and interactions. *Curr. Opin. Struct. Biol.* **2002**, *12*, 21−27.

(226) Capra, J. A.; Laskowski, R. A.; Thornton, J. M.; Singh, M.; Funkhouser, T. A. Predicting Protein Ligand Binding Sites by Combining Evolutionary Sequence Conservation and 3D Structure. *PLoS Comput. Biol.* **2009**, *5*, e1000585.

(227) Hernandez, M.; Ghersi, D.; Sanchez, R. SITEHOUND-web: a server for ligand binding site identification in protein structures. *Nucleic Acids Res.* **2009**, *37*, W413−W416.

(228) Kozakov, D.; Hall, D. R.; Chuang, G.-Y.; Cencic, R.; Brenke, R.; Grove, L. E.; Beglov, D.; Pelletier, J.; Whitty, A.; Vajda, S. Structural conservation of druggable hot spots in protein-protein interfaces. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 13528−13533.

(229) Kozakov, D.; Grove, L. E.; Hall, D. R.; Bohnuud, T.; Mottarella, S. E.; Luo, L.; Xia, B.; Beglov, D.; Vajda, S. The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat. Protoc.* **2015**, *10*, 733−755.

(230) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.

(231) Miranker, A.; Karplus, M. Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins: Struct., Funct., Genet.* **1991**, *11*, 29−34.

(232) Allen, K. N.; Bellamacina, C. R.; Ding, X.; Jeffery, C. J.; Mattos, C.; Petsko, G. A.; Ringe, D. An Experimental Approach to Mapping the Binding Surfaces of Crystalline Proteins. *J. Phys. Chem.* **1996**, *100*, 2605−2611.

(233) Mattos, C.; Ringe, D. Locating and characterizing binding sites on proteins. *Nat. Biotechnol.* **1996**, *14*, 595−9.

(234) Ngan, C. H.; Bohnuud, T.; Mottarella, S. E.; Beglov, D.; Villar, E. A.; Hall, D. R.; Kozakov, D.; Vajda, S. FTMAP: extended protein mapping with user-selected probe molecules. *Nucleic Acids Res.* **2012**, *40*, W271−W275.

(235) Ivetac, A.; Andrew McCammon, J. Mapping the Druggable Allosteric Space of G-Protein Coupled Receptors: a Fragment-Based Molecular Dynamics Approach. *Chem. Biol. Drug Des.* **2010**, *76*, no−no.

(236) Ivetac, A.; McCammon, J. A. A Molecular Dynamics Ensemble-Based Approach for the Mapping of Druggable Binding Sites. *Methods Mol. Biol.* **2012**, *819*, 3−12.

(237) Pande, V. S.; Beauchamp, K.; Bowman, G. R. Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* **2010**, *52*, 99−105.

(238) Schwantes, C. R.; McGibbon, R. T.; Pande, V. S. Perspective: Markov models for long-timescale biomolecular dynamics. *J. Chem. Phys.* **2014**, *141*, 090901.

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

(240) Malmstrom, R. D.; Lee, C. T.; Van Wart, A. T.; Amaro, R. E. Application of Molecular-Dynamics Based Markov State Models to Functional Proteins. *J. Chem. Theory Comput.* **2014**, *10*, 2648−2657.

(241) Shukla, D.; Hernández, C. X.; Weber, J. K.; Pande, V. S. Markov State Models Provide Insights into Dynamic Modulation of Protein Function. *Acc. Chem. Res.* **2015**, *48*, 414−422.

(242) Beauchamp, K. A.; Bowman, G. R.; Lane, T. J.; Maibaum, L.; Haque, I. S.; Pande, V. S. MSMBuilder2: Modeling Conformational Dynamics on the Picosecond to Millisecond Scale. *J. Chem. Theory Comput.* **2011**, *7*, 3412−3419.

(243) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.* **2011**, *134*, 174105.

(244) Senne, M.; Trendelkamp-Schroer, B.; Mey, A. S.; Schütte, C.; Noé, F. EMMA: A Software Package for Markov Model Building and Analysis. *J. Chem. Theory Comput.* **2012**, *8*, 2223−2238.

(245) Kohlhoff, K. J.; Shukla, D.; Lawrenz, M.; Bowman, G. R.; Konerding, D. E.; Belov, D.; Altman, R. B.; Pande, V. S. Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways. *Nat. Chem.* **2013**, *6*, 15−21.

(246) Prinz, J.-H.; Keller, B.; Noé, F. Probing molecular kinetics with Markov models: metastable states, transition pathways and spectroscopic observables. *Phys. Chem. Chem. Phys.* **2011**, *13*, 16912.

(247) Cronkite-Ratcliff, B.; Pande, V. MSMExplorer: visualizing Markov state models for biomolecule folding simulations. *Bioinformatics* **2013**, *29*, 950−952.

(248) Lane, T. J.; Bowman, G. R.; Beauchamp, K.; Voelz, V. a.; Pande, V. S. Markov state model reveals folding and functional dynamics in ultra-long MD trajectories. *J. Am. Chem. Soc.* **2011**, *133*, 18413−18419.

(249) Bowman, G. R.; Ensign, D. L.; Pande, V. S. Enhanced Modeling via Network Theory: Adaptive Sampling of Markov State Models. *J. Chem. Theory Comput.* **2010**, *6*, 787−794.

(250) Lapidus, L. J.; Acharya, S.; Schwantes, C. R.; Wu, L.; Shukla, D.; King, M.; DeCamp, S. J.; Pande, V. S. Complex Pathways in Folding of Protein G Explored by Simulation and Experiment. *Biophys. J.* **2014**, *107*, 947−955.

(251) Li, S.; Xiong, B.; Xu, Y.; Lu, T.; Luo, X.; Luo, C.; Shen, J.; Chen, K.; Zheng, M.; Jiang, H. Mechanism of the All-α to All-β Conformational Transition of RfaH-CTD: Molecular Dynamics Simulation and Markov State Model. *J. Chem. Theory Comput.* **2014**, *10*, 2255−2264.

(252) Schor, M.; Mey, A. S. J. S.; Noé, F.; MacPhee, C. E. Shedding Light on the Dock-Lock Mechanism in Amyloid Fibril Growth Using Markov State Models. *J. Phys. Chem. Lett.* **2015**, *6*, 1076−1081.

(253) 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.

(254) 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.

(255) Tiwary, P.; Limongelli, V.; Salvalaglio, M.; Parrinello, M. Kinetics of protein-ligand unbinding: Predicting pathways, rates, and rate-limiting steps. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E386−E391.

(256) Vanatta, D. K.; Shukla, D.; Lawrenz, M.; Pande, V. S. A network of molecular switches controls the activation of the two-component response regulator NtrC. *Nat. Commun.* **2015**, *6*, 7283.

(257) Shukla, D.; Meng, Y.; Roux, B.; Pande, V. S. Activation pathway of Src kinase reveals intermediate states as targets for drug design. *Nat. Commun.* **2014**, *5*, 3397.

(258) Bryngelson, J. D.; Wolynes, P. G. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84*, 7524−7528.

(259) Wolynes, P.; Onuchic, J.; Thirumalai, D. Navigating the folding routes. *Science (Washington, DC, U. S.)* **1995**, *267*, 1619−1620.

(260) Onuchic, J. N.; Socci, N. D.; Luthey-Schulten, Z.; Wolynes, P. G. Protein folding funnels: the nature of the transition state ensemble. *Folding Des.* **1996**, *1*, 441−450.

(261) Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. Theory of Protein Folding: The Energy Landscape Perspective. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545−600.

(262) Wolynes, P. G. Energy landscapes and solved protein-folding problems. *Philos. Trans. R. Soc., A* **2005**, *363*, 453−467.

(263) Karplus, M. Behind the folding funnel diagram. *Nat. Chem. Biol.* **2011**, *7*, 401−4.

(264) Ferreiro, D. U.; Hegler, J. A.; Komives, E. A.; Wolynes, P. G. Localizing frustration in native proteins and protein assemblies. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 19819−24.

(265) Ferreiro, D. U.; Hegler, J. A.; Komives, E. A.; Wolynes, P. G. On the role of frustration in the energy landscapes of allosteric proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 3499−3503.

(266) Jenik, M.; Parra, R. G.; Radusky, L. G.; Turjanski, A.; Wolynes, P. G.; Ferreiro, D. U. Protein frustratometer: a tool to localize energetic frustration in protein molecules. *Nucleic Acids Res.* **2012**, *40*, W348−51.

(267) Gō, N. The consistency principle in protein structure and pathways of folding. *Adv. Biophys.* **1984**, *18*, 149−164.

(268) Sali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779−815.

(269) Martí-Renom, M. A.; Stuart, A. C.; Fiser, A.; Sánchez, R.; Melo, F.; Šali, A. Comparative Protein Structure Modeling of Genes and Genomes. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 291−325.

(270) Fiser, A.; Do, R. K. G.; Šali, A. Modeling of loops in protein structures. *Protein Sci.* **2000**, *9*, 1753−1773.

(271) Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M.; Eramian, D.; Shen, M.-Y.; Pieper, U.; Sali, A. *Current Protocols in Protein Science*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2007; Chapter 2, Unit 2.9.

(272) Weinkam, P.; Pons, J.; Sali, A. Structure-based model of allostery predicts coupling between distant sites. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 4875−80.

(273) Bahar, I.; Rader, A. Coarse-grained normal mode analysis in structural biology. *Curr. Opin. Struct. Biol.* **2005**, *15*, 586−592.

(274) Zheng, W.; Brooks, B. R.; Thirumalai, D. Low-frequency normal modes that describe allosteric transitions in biological nanomachines are robust to sequence variations. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 7664−7669.

(275) Zheng, W.; Brooks, B.; Thirumalai, D. Allosteric Transitions in Biological Nanomachines are Described by Robust Normal Modes of Elastic Networks. *Curr. Protein Pept. Sci.* **2009**, *10*, 128−132.

(276) Ma, J. Usefulness and Limitations of Normal Mode Analysis in Modeling Dynamics of Biomolecular Complexes. *Structure* **2005**, *13*, 373−380.

(277) Ming, D.; Wall, M. E. Quantifying allosteric effects in proteins. *Proteins: Struct., Funct., Genet.* **2005**, *59*, 697−707.

(278) Tirion, M. M. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Phys. Rev. Lett.* **1996**, *77*, 1905−1908.

(279) Bahar, I.; Atilgan, A. R.; Erman, B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding Des.* **1997**, *2*, 173−181.

(280) Haliloglu, T.; Bahar, I.; Erman, B. Gaussian Dynamics of Folded Proteins. *Phys. Rev. Lett.* **1997**, *79*, 3090−3093.

(281) Ming, D.; Wall, M. E. Allostery in a Coarse-Grained Model of Protein Dynamics. *Phys. Rev. Lett.* **2005**, *95*, 198103.

(282) Su, J. G.; Qi, L. S.; Li, C. H.; Zhu, Y. Y.; Du, H. J.; Hou, Y. X.; Hao, R.; Wang, J. H. Prediction of allosteric sites on protein surfaces with an elastic-network-model-based thermodynamic method. *Phys. Rev. E* **2014**, *90*, 022719.

(283) Berezovsky, I. N. Thermodynamics of allostery paves a way to allosteric drugs. *Biochim. Biophys. Acta, Proteins Proteomics* **2013**, *1834*, 830−835.

(284) Mitternacht, S.; Berezovsky, I. N. A geometry-based generic predictor for catalytic and allosteric sites. *Protein Eng., Des. Sel.* **2011**, *24*, 405−409.

(285) Mitternacht, S.; Berezovsky, I. N. Binding Leverage as a Molecular Basis for Allosteric Regulation. *PLoS Comput. Biol.* **2011**, *7*, e1002148.

(286) Goncearenco, A.; Mitternacht, S.; Yong, T.; Eisenhaber, B.; Eisenhaber, F.; Berezovsky, I. N. SPACER: server for predicting allosteric communication and effects of regulation. *Nucleic Acids Res.* **2013**, *41*, W266−W272.

(287) Ngan, C. H.; Hall, D. R.; Zerbe, B. S.; Grove, L. E.; Kozakov, D.; Vajda, S. FTSite: high accuracy detection of ligand binding sites on unbound protein structures. *Bioinformatics* **2012**, *28*, 286−287.