Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Structural Biology

Hybrid approaches to molecular simulation

Bosco K Ho[1,2], David Perahia[3] and Ashley M Buckle[1]

Molecular dynamics (MD) simulation is an established method for studying the conformational changes that are important for protein function. Recent advances in hardware and software have allowed MD simulations over the same timescales as experiment, improving the agreement between theory and experiment to a large extent. However, running such simulations are costly, in terms of resources, storage, and trajectory analysis. There is still a place for techniques that involve short MD simulations. In order to overcome the sampling paucity of short time-scales, hybrid methods that include some form of MD simulation can exploit certain features of the system of interest, often combining experimental information in surprising ways. Here, we review some recent hybrid approaches to the simulation of proteins.

Addresses
[1] Department of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria 3800, Australia
[2] Monash eResearch Centre, Monash University, Victoria 3800, Australia
[3] LBPA, ENS de Cachan, CNRS, Avenue du Président Wilson, 94235, Cachan, France

Corresponding author: Buckle, Ashley M (ashley.buckle@monash.edu)

Current Opinion in Structural Biology 2012, **22**:1–8

This review comes from a themed issue on
Sequences and topology
Edited by Christine Orengo and James Whisstock

0959-440X/$ – see front matter

http://dx.doi.org/10.1016/j.sbi.2012.05.005

Introduction

The dynamical properties of proteins play a key role in their functions. These can be investigated using molecular dynamics (MD) simulations, where the protein is represented on the atomistic level and also using simplified coarse-grained methods. It is fundamentally important that theoretical methods are benchmarked against experimental observations, but this has been hampered owing to the difference in timescales accessible by the two methods. MD typically accessed timescales of nanoseconds (ns)–microseconds (μs), orders of magnitude shorter than the millisecond (ms) timescales observed by experiment. However, recent advances in hardware and software have begun to close this gap, most notably with the first millisecond simulation [1]. The technological advances allowing long-timescale MD over the past few years have been reviewed extensively, as has the use of MD in protein folding [2–5]. The use of MD simulation is now firmly embedded in the repertoire of biophysical methods. Since the scope of this field is now broad, we focus on recent highlights where MD has contributed to 'hybrid approaches' for elucidating the structure and dynamics of macromolecules, where useful predictions can be extrapolated on even modest computational systems (Table 1).

Predicting allostery through intramolecular pathways

Allostery is one of the underlying principles of signal transduction in proteins where a binding event on one site of a protein induces binding at another site. Although classically, allostery was first found in domain interfaces [6], there has been growing interest in single domain allostery ever since the experimental work on intramolecular pathways of allostery in the PDZ domain (Figure 1A) [7]. In that study, Statistical Coupling Analysis (SCA) was introduced to measure covariation between position pairs across a multiple-sequence alignment of the PDZ family of sequences (Figure 1B). Subsequent mutations at positions that had high scoring SCA values induced significant thermodynamic changes to the PDZ domain. Lockless and Ranganathan [7] made an interesting conjecture that interlocking chains of highly covarying SCA position pairs constituted intramolecular pathways of allostery that travel through the body of the protein.

Hybrid simulations are particularly suited to the study of single domain allostery. The most commonly studied example is the PDZ domain, which functions as a modular scaffold in eukaryotes that bring together receptors and signaling molecules [8]. PDZ domains bind to the C-terminus of other proteins largely through a canonical binding site lined by an α-helix on one side and the last strand of the central β-sheet on the other (Figure 1A) [9]. One example of a PDZ-based scaffold is GRIP [10], which consists of seven PDZ domains, some of which bind AMPA receptors and others bind signaling proteins. By bringing the receptors and signaling proteins together, GRIP enhances the efficiency of synaptic transmission. More importantly, the binding of PDZ domains to the C-terminus of their target proteins is modulated by allosteric interactions to other PDZ domains. Allosteric effectors may thus be identified by the simulation of intramolecular pathways in PDZ domains.

A straightforward way of analyzing such intramolecular pathways of communication simply involves careful analysis of sidechain–sidechain correlations of straight

Please cite this article in press as: Ho BK, et al.: Hybrid approaches to molecular simulation, Curr Opin Struct Biol (2012), http://dx.doi.org/10.1016/j.sbi.2012.05.005

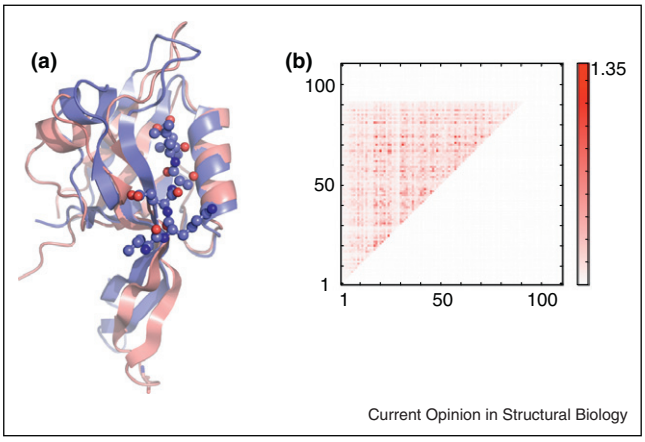
Table 1

Summary of methods

Hybrid method	Pros	Cons
Temperature coupling to residues [14,16]	Simple to implement temperature couple to residue	Requires backbone constraints, and energy leaks into the backbone
Rotamerically induced perturbation [18•,19]	Flexible modeling of local perturbations to identify direct interaction of residues	Only considers local plasticity
Sidechain Monte-Carlo [17]	Exhaustive modeling of sidechain conformations	Since it uses Monte-Carlo of the sidechains over the whole protein, cannot identify direct interactions via energy flow. No backbone motion
Double-minimum allostery [22]	Identifies sidechain correlations during the transition between large motions	Dependent on knowledge of the crystal structure of two states of allostery
Elastic Networks + Replica-Exchange [20,31•]	Effective sampling of the binding site by combining global motions with extensive temperature search	Complicated protocol and replica-exchange is a reasonably expensive calculation
FlexPepDock [29]	Docks arbitrary peptides to ill-defined binding site on a protein	Low resolution docking and expensive computationally
Pepsec [36]	Fast peptide docking with reasonable binding energies for sequence profiles	Requires manual choice of anchor residues and is designed for determining sequence specificity
Peptide binding + protein-design [37•]	Extremely exhaustive analysis of sequence profiles of a peptide–protein system	May not be easily transferable to other systems as it relies on the canonical binding site of PDZ domains
Direct Coupling Analysis [44••,46•,47••]	The most robust definition of covariation in multiple-sequence alignment. Predicts structural contacts, protein–protein interactions, and common-fold structure	Limited to consensus structures of a family of proteins
Normal Mode + Minimization/ Molecular Dynamics/ Metadynamics [61–64,66,67]	Massively enhances the sampling capabilities of molecular dynamics simulations	Restricted to conformational changes that are evident in the starting structure
Coarse-grain domain model of Nuclear Pore complex [93••]	Unprecedented scale in terms of size and time-scales	Will rely on the experimental determination of systems of similar enormous size

MD simulations. Long-range communication is assumed to take place if a statistically significant correlation is found between distal residues. There is considerable freedom in choosing the relevant parameter in which to define the correlation. Over the past few years, different studies have defined correlations over a wide variety of measures: from energy fluctuations in PDZ2 [11], to the entropy of torsional angles in Interleukin-2 [12], and to relative distance fluctuations in ABL and EGFR kinases [13]. In all these studies, interesting distal residues were found to correlate strongly, thus providing a simulation trace of potential allosteric interactions. These studies however, do not propose any concrete mechanism for how the communication is effected between distal sites.

Figure 1



The PDZ domain. **(A)** Schematic of the PDZ domain (6th domain of Partition-Defective-Protein-6). The bound conformation (1RY4) is in blue, and the free conformation (1RZX) in pink. The peptide in the bound conformation is in ball-and-stick and the C-terminal is at the top of the figure. There is significant conformational change upon binding the peptide. **(B)** A heat map of the SCA analysis of covariation in the multiple-sequence alignment of the PDZ family. Red values indicate high values of covariation. This is a noisy graph that has been the basis of many studies of the allostery of the PDZ domain.

Several simulation methods have attempted to directly model intramolecular pathways by isolating certain physical features. The first method to study this was the Anisotropic Thermal Diffusion (ATD) method [14]. In ATD, the PDZ protein is cooled down to 10 K, then, after position constraints are applied to the backbone and surface atoms, a temperature couple is applied to a single residue. Coupling between residues is measured by heat flow, allowing a prediction of residues that physically couple, and thus carry energy from one allosteric site to another. ATD has been used to predict allosteric sites in different systems including a recent analysis of the Liver X Receptor, which predicted several allosteric sites upon binding cholesterol [15]. The Pump-Probe method is a variation of ATD that uses pulsed energies at set frequencies [16].

Please cite this article in press as: Ho BK, et al.: Hybrid approaches to molecular simulation, Curr Opin Struct Biol (2012), http://dx.doi.org/10.1016/j.sbi.2012.05.005

One drawback with the ATD and Pump-Probe methods is that much of the applied energy flows along the backbone. This generates a large background energy flow that swamps direct sidechain interactions with the perturbed residue. One solution to this is a simplified model that focuses exclusively on sidechain fluctuations [17]. The PDZ domain was modeled using Monte-Carlo simulation of only the sidechain degrees of freedom, where correlations between sidechain fluctuations were used to identify allosteric interaction. In this way, the problem of energy leakage in the backbone is avoided, but backbone plasticity is lost. A variation of the ATD method solves the problem of backbone energy leakage in another way [18•]. Rotamerically Induced Perturbation (RIP) applies the perturbation energy to a residue only through rotation of the sidechain dihedral angles. A RIP perturbation does not leak any energy into the backbone and thus, backbone constraints are not needed at all and the backbone can respond to perturbed sidechains in nonlocal contacts. As a result, RIP analysis generates much cleaner signals of physical sidechain–sidechain interactions in the PDZ domain. Furthermore, when driven at high temperature, RIP generates a useful map of surface plasticity [19]. High-temperature RIP applied to residues in a protein equilibrated to room temperature, induces weakly held surface loops to move. This method recapitulated known flexible loops in a number of well-studied proteins [19]. When applied to different PDZ domains [18•], dramatically different surface plasticity was produced that were consistent with the known differences in dynamics between PDZ domains.

Nevertheless these models focus on local interactions of a sidechain with nearby sidechains and nearby backbone. Other models have been introduced to model global responses to perturbation. Perturbation Response Scanning (PRS) focuses on how individual residues affect the global motion of a protein domain [20]. In the PRS method, a matrix is constructed that represents the global motion of the protein using a classical Elastic Network Model approximation [21]. Using a detailed analysis of this matrix, changes in the global motion owing to perturbation of specific residues can be calculated. Another simplified MD-based method investigates collective motions of allostery owing to ligand binding [22]. Artificial double-minimum distance constraints are applied to residue–residue contacts in a protein system, allowing the protein to smoothly oscillate between the bound and free conformation. Another set of artificial potentials force the ligand-binding state into either the bound or free states. Allosteric sites are then identified by correlated sidechain–sidechain motions in the system observed during an MD run.

The main conclusion is that even small rigid domains exhibit subtle dynamics that can be modeled in different ways. Methods have been developed that focus exclusively on sidechains, or sidechains with local backbone plasticity, or global motions. The use of restricted models allows us to isolate specific aspects of motion that induce single domain allostery.

Restricted models of peptide binding

Ligand binding is clearly an important component of allostery. Unfortunately, there is as yet no general solution to the problem of modeling ligand-binding especially if one considers small organic molecules [23]. However, in ligand binding systems where a protein domain exhibits a canonical-binding site for binding peptide ligands, simplified approaches can be taken [24]. Given that the ligands are peptides themselves, the same force-fields as proteins can be used, and this avoids the complication of parameterizing arbitrary organic molecules. Well-known peptide–protein interaction systems include the PDZ, SH2, SH3, WW, PTB, EH, and MHC domains [9,25–27].

Recent developments of peptide–protein systems have focused on how to model protein flexibility. There is a balance between allowing flexibility in the system, and keeping the simulation tractable. Many of these systems are based on the Monte-Carlo Rosetta package [28]. FlexPepDock is a general approach to peptide–protein binding, where the system iterates between a local exploration of the binding site using standard Rosetta local moves [28], and an exploration of the peptide conformation, where the peptide conformations also exploit the Rosetta local fragment library [29]. FlexPepDock was shown to have reasonable predictive power over a wide range of peptide–protein systems. Another approach used a custom Monte-Carlo model of the peptide binding. In this study, the protein was rigidly constrained except for the active site, where the active site residues was subjected to a local 8-residue Monte-Carlo move [30]. This model used custom-fitted solvation parameters, and provided interesting details on the kinetics of binding.

Nevertheless, these methods only apply local methods of modeling the active site. Another approach takes into account global motions of the protein in response to ligand binding [31•]. In that study, Elastic Network Model analysis was used to generate a set of constraints for the global motion of the protein. Using these constraints in a replica-exchange MD simulation, a rich set of protein conformations was sampled, to which peptides were docked using the standard ligand-binding package RosettaLigand [32]. When applied to different PDZ domains, quite subtle interaction differences were found in the peptide binding.

Large-scale proteomic analysis of binding affinities are now available that provide highly detailed peptide-binding profiles for various peptide–protein systems, such as the PDZ domain [33,34], as well as the SH3 and WW

Please cite this article in press as: Ho BK, et al.: Hybrid approaches to molecular simulation, Curr Opin Struct Biol (2012), http://dx.doi.org/10.1016/j.sbi.2012.05.005

domains [35]. In order to connect with experiment, we need to push the peptide-binding models further to generate comprehensive peptide-binding profiles, where the focus is on speed and relative binding energies. Several methods have been developed that generate rich sequence profiles that can be compared to experiment. Pepsec is a Rosetta-based method that uses constraints to anchor certain peptide residues to the binding site in order to limit the search space [36]. Another Rosetta-based study designed a hybrid method that combines Monte-Carlo conformational search and protein design to generate rich sequence profiles [37•]. In this method, the peptide backbone is docked onto the structure, and a large ensemble of structures of the complex is generated. Then, using essentially a protein-design protocol, a search is conducted over both the chemistry and the conformations of the sidechains of the peptide. This method generated rich peptide-sequence profiles of several PDZ domains that were even able to recapitulate changes in the sequence profiles owing to point mutations on the PDZ domain itself.

These methods demonstrate a range of approaches that expand the repertoire of peptide–protein systems. As the coverage of calculated sequence profiles improves, and the allostery of the proteins is understood, it may be possible to predict genome-wide interaction maps through structural analysis.

Simulation models that incorporate sequence alignments

Much of the work on intramolecular pathways of allostery was inspired by the SCA measure [7], which is a formalized measure of covariation analysis [38–40]. Some recent simulations have exploited recent refinements of covariation analysis to produce some useful hybrid simulations of protein–protein interactions, and remarkably, even structure prediction. The highly covarying pairs of positions, identified by the SCA measure, are somewhat ambiguous, and some of them admit no easy interpretation (Figure 1B). There have been various attempts to improve the measure. A correction for the SCA term was found that accounts for phylogenetic correlations [40]. Another study found a weighting term that accounted for uneven levels of conservation at different positions [41]. A clustering analysis of high-scoring SCA position pairs was used to identify groups of coevolving residues [42]. Another approach attempts to disentangle distal from proximate correlations using a Bayesian network model [43].

The most significant improvement of the analysis of covariation was the introduction of a global model of covariation called Direct Coupling Analysis (DCA) [44••]. Instead of considering each pair of positions as independent, the DCA uses a global probability model that treats the correlation between a pair of positions as a marginal distribution over the global model. The DCA analysis was able to filter out indirect correlations from direct correlations, leading to much cleaner predictions. The original application of DCA was applied to a two-protein docking problem, where DCA analysis was carried out over a combined multiple-sequence alignment of a kinase–receptor system. It was found that high-scoring DCA position-pairs between the protein families predicted key residues that formed the protein–protein interaction. A subsequent MD docking protocol was designed around these predicted contacts [45], which was able to recapitulate the known structure of the protein complex.

A faster version of DCA analysis was developed [46•] that allowed a much more extensive analysis over a diverse set of 131 different protein families. It was found that high scoring DCA position-pairs in a multiple scoring alignment correspond to observed contacts in experimentally determined structures of proteins from each family. This confirmed an older hypothesis [39] that the covariation of positions is largely determined by conserved structural contacts in a protein family.

An impressive confirmation and application of this hypothesis was recently developed by Marks and colleagues [47••]. For a diverse range of 15 protein families, they identified high-scoring DCA pairs of positions in the multiple sequence alignment. They interpreted these pairs of positions as structural contacts and fed the pairs as distance constraints into standard NMR refinement software. They generated robust predictions of structure that were typically within 3 or 4 angstroms of known structures of members of the families. Not only was this an impressive confirmation of the interpretation of covariation as structural contacts, but this was also a demonstration of the power of this method to perform ab initio structure prediction.

These impressive results with DCA suggests that covariation analysis of a sequence alignment identifies key structural contacts that we can directly exploit in simulations. However, for purposes of allosteric analysis, it would help to understand exactly how structural contacts contribute to allostery in single domains. There is a theoretical model that postulates that allostery in single-domains is transmitted through the rigidification of flexible-regions upon ligand-binding [48]. Is there such a connection between structural contacts and intrinsic flexibility? One study found such a link in the PDZ domains [18•]. In simulations using high-temperature RIP, it was found that different PDZ domains exhibited different flexibility in a key α-helix. In certain strategically placed structural contacts, in the PDZ domain where the helix was rigid, a strong RIP interaction was found between a sidechain in the body of the protein and the helix, whereas in the PDZ domain where the helix was flexible, there was no such interaction. These key structural contacts scored much higher SCA

Please cite this article in press as: Ho BK, et al.: Hybrid approaches to molecular simulation, Curr Opin Struct Biol (2012), http://dx.doi.org/10.1016/j.sbi.2012.05.005

values in the sequence alignment than other contacts. We can thus interpret highly covarying structural contacts as contacts that are well placed to modulate the flexibility of a protein, and hence, its allosteric response.

Given the impressive results that have already emerged from the covariation analysis, exploiting this rich source of information will surely lead to exciting opportunities for hybrid simulations.

Unbiased predictions of protein conformational changes using normal modes

Many allosteric systems require the dynamic response of a protein to respond correctly to interaction with its partner. As such, there is a place for the prediction of conformational changes of protein structures using hybrid methods. Clearly, methods that use short simulations are important. But it is more important to focus on unbiased methods that do not require the input of predefined transition states, collective variables or pathways. Only then, can we truly make a prediction of conformational change. One such method is the classical Normal Mode (NM) analysis [49–51]. The classical NM method considers the atomic force-field interactions for defining the Hessian matrix of the atomic coordinates. An approximation to this is the Elastic Network Model (EMN) that considers the system as a network of mechanical springs. The eigenvalues and eigenvectors of the Hessian matrix thus describe the collective motion of the protein, where the eigenvalue in particular gives the frequency of the motion. Standard NM analysis has been used to predict the large collective displacements of atoms relevant to enzymatic activity, and binding of ligands or other macromolecules [52–55]. NM analysis has also predicted motions for large-scale displacements corresponding to time scales approaching the millisecond or beyond for large complex assemblies of proteins [56,57], which would otherwise be difficult to study with conventional brute-force MD simulations [58–60].

NM analysis has been successfully combined with other simulation techniques, for example energy minimization where important results have been obtained just by displacing the structures along the lowest NM frequency modes using energy minimization [61,62]. NM analysis has also successfully augmented MD simulation. A new approach, termed 'Consensus Mode' was developed in order to take advantage of the topological characteristics of multiple points on the potential energy surface derived from NM in order to generate more robust modes in free-energy MD calculations [63]. The lowest NM frequency eigenvectors describing these motions have been used as reaction coordinate for computing conformational changes [64] by using the umbrella method [65]. Metadynamics is another useful method for obtaining the free energy profiles in the subspace defined by low frequency NM eignenvectors [66,67].

Several NM studies that were able to corroborate experimental findings deserve highlighting. A number of studies have investigated conformational changes that are linked to enzymatic activity and allostery [68–70]. In the case of the bacterial enzyme Glucosamine-6-phosphate synthase, it was shown that the inner-channel within the protein linking the two active sites is formed owing to the collective motions of the protein [61]. This channel could not be obtained with standard MD simulations. For docking simulations, a greater variety of conformations of the protein can be generated by exploring a low frequency NM subspace. This increased the success of finding new ligands that are likely to bind with high affinity [71–74], and to identify allosteric binding sites [75]. In the prediction of protein–protein complexes where protein partners undergo important global conformational changes, the NM approach was able to generate conformations that were suitable for binding [76–80]. The NM-generated conformations also improved the correct ranking of docked complexes, mimicking the induced fit mechanism by global movements [81–84]. NM analysis has also been applied to the study of macromolecular assemblies from electron microscopy [85–87], as well as the interpretation of SAXS experiments [88–90].

The simulations of proteins discussed here typically operate on an atomic level of detail. In cases where atomic interactions are not so relevant, one can always simplify the representation of the protein to a coarse-grained (CG) model [91,92]. Whilst a general CG discussion is beyond the scope of this review, a recent CG model that has successfully modeled a system of unprecendented size should be highlighted. This impressive example studied the transport of cargo through the nuclear pore complex, which is one of the largest macromolecular assemblies in eukaryotic cells [93••]. New insights were obtained as to how this system can discriminate between the active and inert cargos, the passage time as a function of the size of the cargo, the interactions that are involved during the translocation, and many other properties.

Conclusions and outlook

The ultimate goal of MD simulation is to provide a complete theoretical model of the chemistry of proteins that can accurately and efficiently generate structure, dynamics and interactions. Nevertheless, the kind of proteins that have been selected by evolution only constitute a subset of all possible protein sequences. These proteins belong to the subset of protein sequences that exhibit foldability, marginal stability at physiological temperatures, restricted dynamics, and efficient specificity in interaction. These are all properties that can be exploited in simulation, and many of the hydrid methods explicitly do this. Whilst the progress towards longer and unbiased MD simulations of individual proteins is an exciting development, in order to connect with experiment, especially given the growing amount of proteomic

Please cite this article in press as: Ho BK, et al.: Hybrid approaches to molecular simulation, Curr Opin Struct Biol (2012), http://dx.doi.org/10.1016/j.sbi.2012.05.005

data, there will always be a place for approximate techniques that are efficient enough to generate predictive results for entire groups of proteins.

Acknowledgement

AMB is an NHMRC Senior Research Fellow.

References and recommended reading

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

- • of special interest
- •• of outstanding interest

1. Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, Bank JA, Jumper JM, Salmon JK, Shan Y, Wriggers W: **Atomic-level characterization of the structural dynamics of proteins**. *Science* 2010, **330**:341-346.

2. Best RB: **Atomistic molecular simulations of protein folding**. *Curr Opin Struct Biol* 2012, **22**:52-61.

3. Klepeis JL, Lindorff-Larsen K, Dror RO, Shaw DE: **Long-timescale molecular dynamics simulations of protein structure and function**. *Curr Opin Struct Biol* 2009, **19**:120-127.

4. Schaeffer RD, Fersht A, Daggett V: **Combining experiment and simulation in protein folding: closing the gap for small model systems**. *Curr Opin Struct Biol* 2008, **18**:4-9.

5. Zwier MC, Chong LT: **Reaching biological timescales with all-atom molecular dynamics simulations**. *Curr Opin Pharmacol* 2010, **10**:745-752.

6. Changeux JP, Edelstein SJ: **Allosteric receptors after 30 years**. *Neuron* 1998, **21**:959-980.

7. Lockless SW, Ranganathan R: **Evolutionarily conserved pathways of energetic connectivity in protein families**. *Science* 1999, **286**:295-299.

8. Nourry C, Grant SG, Borg JP: **PDZ domain proteins: plug and play!** *Sci STKE* 2003:RE7.

9. Harris BZ, Lim WA: **Mechanism and role of PDZ domains in signaling complex assembly**. *J Cell Sci* 2001, **114**:3219-3231.

10. Dong H, O'Brien RJ, Fung ET, Lanahan AA, Worley PF, Huganir RL: **GRIP: a synaptic PDZ domain-containing protein that interacts with AMPA receptors**. *Nature* 1997, **386**:279-284.

11. Kong Y, Karplus M: **Signaling pathways of PDZ2 domain: a molecular dynamics interaction correlation analysis**. *Proteins* 2009, **74**:145-154.

12. McClendon CL, Friedland G, Mobley DL, Amirkhani H, Jacobson MP: **Quantifying correlations between allosteric sites in thermodynamic ensembles**. *J Chem Theory Comput* 2009, **5**:2486-2502.

13. Dixit A, Verkhivker GM: **Computational modeling of allosteric communication reveals organizing principles of mutation-induced signaling in ABL and EGFR kinases**. *PLoS Comput Biol* 2011, **7**:e1002179.

14. Ota N, Agard DA: **Intramolecular signaling pathways revealed by modeling anisotropic thermal diffusion**. *J Mol Biol* 2005, **351**:345-354.

15. Burendahl S, Nilsson L: **Computational studies of LXR molecular interactions reveal an allosteric communication pathway**. *Proteins* 2012, **80**:294-306.

16. Sharp K, Skinner JJ: **Pump-probe molecular dynamics as a tool for studying protein motion and long range coupling**. *Proteins* 2006, **65**:347-361.

17. Dubay KH, Bothma JP, Geissler PL: **Long-range intra-protein communication can be transmitted by correlated side-chain fluctuations alone**. *PLoS Comput Biol* 2011, **7**:e1002168.

18. • Ho BK, Agard DA: **Conserved tertiary couplings stabilize elements in the PDZ fold, leading to characteristic patterns of domain conformational flexibility**. *Protein Sci* 2010, **19**:398-411.
This paper provides a method that can effectively study local residue perturbations and heat-flow that overcomes problems with early methods. The Rotamerically Induced Perturbation perturbs a residue by applying rotations directly to the sidechain χ angles, which solves the problems of spurious energy flow along the backbone, and also avoids the need for applying hard positional constraints on the backbone atoms. As the method does not require backbone constraints, this methods allows the modeling of local backbone plasticity as a response to local perturbations. Indeed, Molecular Dynamics must be used to study heat-flow, as Monte-Carlo simulations focus only on conformational space and potential energies, and ignore heat flow in principle.

19. Ho BK, Agard DA: **Probing the flexibility of large conformational changes in protein structures through local perturbations**. *PLoS Comput Biol* 2009, **5**:e1000343.

20. Gerek ZN, Ozkan SB: **Change in allosteric network affects binding affinities of PDZ domains: analysis through perturbation response scanning**. *PLoS Comput Biol* 2011, **7**:e1002154.

21. Gerek ZN, Keskin O, Ozkan SB: **Identification of specificity and promiscuity of PDZ domain interactions through their dynamic behavior**. *Proteins* 2009, **77**:796-811.

22. Weinkam P, Pons J, Sali A: **Structure-based model of allostery predicts coupling between distant sites**. *Proc Natl Acad Sci USA* 2012, **109**:4875-4880.

23. Gilson MK, Zhou HX: **Calculation of protein-ligand binding affinities**. *Annu Rev Biophys Biomol Struct* 2007, **36**:21-42.

24. Rubinstein M, Niv MY: **Peptidic modulators of protein-protein interactions: progress and challenges in computational design**. *Biopolymers* 2009, **91**:505-513.

25. Hu H, Columbus J, Zhang Y, Wu D, Lian L, Yang S, Goodwin J, Luczak C, Carter M, Chen L *et al.*: **A map of WW domain family interactions**. *Proteomics* 2004, **4**:643-655.

26. Kuriyan J, Cowburn D: **Modular peptide recognition domains in eukaryotic signaling**. *Annu Rev Biophys Biomol Struct* 1997, **26**:259-288.

27. Lafuente EM, Reche PA: **Prediction of MHC-peptide binding: a systematic and comprehensive overview**. *Curr Pharm Des* 2009, **15**:3209-3220.

28. Rohl CA, Strauss CE, Misura KM, Baker D: **Protein structure prediction using Rosetta**. *Methods Enzymol* 2004, **383**:66-93.

29. Raveh B, London N, Zimmerman L, Schueler-Furman O: **Rosetta FlexPepDock ab-initio: simultaneous folding, docking and refinement of peptides onto their receptors**. *PLoS ONE* 2011, **6**:e18934.

30. Staneva I, Wallin S: **Binding free energy landscape of domain-peptide interactions**. *PLoS Comput Biol* 2011, **7**:e1002131.

31. • Gerek ZN, Ozkan SB: **A flexible docking scheme to explore the binding selectivity of PDZ domains**. *Protein Sci* 2010, **19**:914-928.
Although Elastic Network Models provide a powerful technique to identify global motions in proteins, local atomic structure is lost in the process. Similarly, Replica-Exchange Molecular-Dynamics allows a thorough exploration of temperature space at the expense of local energy minima. This paper shows that combining these two methods hits a sweet spot of compromise. An initial Elastic Network Model calculation provides meaningful constraints to the Replica Exchange simulation, after which, the Replica-Exchange can thoroughly search the constrained space for local interaction optimizations. This method gives an effective sampling of peptide-binding site distortions that combine both global and local atom-level interactions.

32. Meiler J, Baker D: **ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility**. *Proteins* 2006, **65**:538-548.

33. Stiffler MA, Chen JR, Grantcharova VP, Lei Y, Fuchs D, Allen JE, Zaslavskaia LA, MacBeath G: **PDZ domain binding selectivity is optimized across the mouse proteome**. *Science* 2007, **317**:364-369.

34. Tonikian R, Zhang Y, Sazinsky SL, Currell B, Yeh JH, Reva B, Held HA, Appleton BA, Evangelista M, Wu Y *et al.*: **A specificity map for the PDZ domain family**. *PLoS Biol* 2008, **6**:e239.

Please cite this article in press as: Ho BK, et al.: Hybrid approaches to molecular simulation, Curr Opin Struct Biol (2012), http://dx.doi.org/10.1016/j.sbi.2012.05.005

35. Gfeller D, Butty F, Wierzbicka M, Verschueren E, Vanhee P, Huang H, Ernst A, Dar N, Stagljar I, Serrano L *et al.*: **The multiple-specificity landscape of modular peptide recognition domains**. *Mol Syst Biol* 2011, **7**:484.

36. King CA, Bradley P: **Structure-based prediction of protein-peptide specificity in Rosetta**. *Proteins* 2010, **78**:3437-3449.

37. • Smith CA, Kortemme T: **Structure-based prediction of the peptide sequence space recognized by natural and synthetic PDZ domains**. *J Mol Biol* 2010, **402**:460-474.
This paper used a rich combination of methods to generate an exhaustive mapping of the peptide sequence preferences of PDZ domains. It uses a Monte-Carlo exploration of the structure of the peptide–protein complex, followed by a modified protein-engineering protocol that searches through both sidechain conformation space and sidechain chemistry space of the residues on the peptides. This dedicated system allows an exhaustive search of the binding specificities of an enormous of peptides.

38. Altschuh D, Lesk AM, Bloomer AC, Klug A: **Correlation of co-ordinated amino acid substitutions with function in viruses related to tobacco mosaic virus**. *J Mol Biol* 1987, **193**:693-707.

39. Gobel U, Sander C, Schneider R, Valencia A: **Correlated mutations and residue contacts in proteins**. *Proteins* 1994, **18**:309-317.

40. Dutheil JY: **Detecting coevolving positions in a molecule: why and how to account for phylogeny**. *Brief Bioinform* 2012, **13**:228-243.

41. Gouveia-Oliveira R, Pedersen A: **Finding coevolving amino acid residues using row and column weighting of mutual information and multi-dimensional amino acid representation**. *Algorithms Mol Biol* 2007, **2**:12.

42. Halabi N, Rivoire O, Leibler S, Ranganathan R: **Protein sectors: evolutionary units of three-dimensional structure**. *Cell* 2009, **138**:774-786.

43. Burger L, van Nimwegen E: **Disentangling direct from indirect co-evolution of residues in protein alignments**. *PLoS Comput Biol* 2010, **6**:e1000633.

44. •• Weigt M, White RA, Szurmant H, Hoch JA, Hwa T: **Identification of direct residue contacts in protein-protein interaction by message passing**. *Proc Natl Acad Sci USA* 2009, **106**:67-72.
This paper provides a significant conceptual advance in the study of the covariation of amino-acid preferences in a multiple-sequence alignment. Previous studies have treated each set of pair-wise correlations of positions in an alignment independently. By contrast, the Direct Coupling Analysis treats covariation in the context of a global model where the observed frequencies for a pair of positions is considered as, not just a result of the covariation at those two positions, but includes background contributions from all other positions in the alignment. In this paper, a computationally expensive method, message-passing, is used to deduce the underlying covariation model from the alignment data, and this is applied to a two-protein signaling system.

45. Schug A, Weigt M, Onuchic JN, Hwa T, Szurmant H: **High-resolution protein complexes from integrating genomic information with molecular simulation**. *Proc Natl Acad Sci USA* 2009, **106**:22124-22129.

46. • Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, Sander C, Zecchina R, Onuchic JN, Hwa T, Weigt M: **Direct-coupling analysis of residue coevolution captures native contacts across many protein families**. *Proc Natl Acad Sci USA* 2011, **108**:E1293-E1301.
This paper provides a crucial development of the Direct Coupling Analysis of covariation in a multiple-sequence alignment introduced in Weigt *et al.* (2009). Instead of the expensive message-passing method of deducing the underlying global model, a computationally efficient method is introduced that allows the DCA to be calculated for large-scale bioinformatic analysis. In this paper, an analysis of 131 domain families shows that high-scoring DCA pairs of positions consistently identifies structural contacts, as well as contacts from alternate protein conformations and inter-domain interactions.

47. •• Marks DS, Colwell LJ, Sheridan R, Hopf TA, Pagnani A, Zecchina R, Sander C: **Protein 3D structure computed from evolutionary sequence variation**. *PLoS ONE* 2011, **6**:e28766.
This paper proves a remarkable result: that the co-variation trace of a sufficiently rich multiple-sequence alignment contains sufficient information to generate a medium-resolution 3D structure of the common fold of the family. This work is a confirmation that the DCA measure of covariation of a multiple-sequence alignment identifies structural contacts in the consensus structure. But more importantly, it shows that the high-scoring DCA contacts are sufficient to generate robust predictions of protein structure.

48. Hilser VJ, Thompson EB: **Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins**. *Proc Natl Acad Sci USA* 2007, **104**:8311-8315.

49. Cui Q, Bahar I (Eds): *Normal Mode Analysis. Theory and Applications to Biological and Chemical Systems*. Chapman and Hall/CRC; 2005.

50. Dykeman EC, Sankey OF: **Normal mode analysis and applications in biological physics**. *J Phys Condens Matter* 2010, **22**:423202.

51. Wilson EB, Cross PCDJ: *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*. New York: Dover Publications; 1980.

52. Bahar I, Lezon TR, Bakan A, Shrivastava IH: **Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins**. *Chem Rev* 2010, **110**:1463-1497.

53. Deriu MA, Soncini M, Orsi M, Patel M, Essex JW, Montevecchi FM, Redaelli A: **Anisotropic elastic network modeling of entire microtubules**. *Biophys J* 2010, **99**:2190-2199.

54. Mouawad L, Perahia D: **Motions in hemoglobin studied by normal mode analysis and energy minimization: evidence for the existence of tertiary T-like, quaternary R-like intermediate structures**. *J Mol Biol* 1996, **258**:393-410.

55. Tama F, Sanejouand YH: **Conformational change of proteins arising from normal mode calculations**. *Protein Eng* 2001, **14**:1-6.

56. Li G, Cui Q: **Analysis of functional motions in Brownian molecular machines with an efficient block normal mode approach: myosin-II and Ca2+ -ATPase**. *Biophys J* 2004, **86**:743-763.

57. Thomas A, Field MJ, Mouawad L, Perahia D: **Analysis of the low frequency normal modes of the T-state of aspartate transcarbamylase**. *J Mol Biol* 1996, **257**:1070-1087.

58. Cui Q, Li G, Ma J, Karplus M: **A normal mode analysis of structural plasticity in the biomolecular motor F(1)-ATPase**. *J Mol Biol* 2004, **340**:345-372.

59. Louet M, Perahia D, Martinez J, Floquet N: **A concerted mechanism for opening the GDP binding pocket and release of the nucleotide in hetero-trimeric G-proteins**. *J Mol Biol* 2011, **411**:298-312.

60. Thomas A, Hinsen K, Field MJ, Perahia D: **Tertiary and quaternary conformational changes in aspartate transcarbamylase: a normal mode study**. *Proteins* 1999, **34**:96-112.

61. Floquet N, Durand P, Maigret B, Badet B, Badet-Denisot MA, Perahia D: **Collective motions in glucosamine-6-phosphate synthase: influence of ligand binding and role in ammonia channelling and opening of the fructose-6-phosphate binding site**. *J Mol Biol* 2009, **385**:653-664.

62. Floquet N, M'Kadmi C, Perahia D, Gagne D, Berge G, Marie J, Baneres JL, Galleyrand JC, Fehrentz JA, Martinez J: **Activation of the ghrelin receptor is described by a privileged collective motion: a model for constitutive and agonist-induced activation of a sub-class A G-protein coupled receptor (GPCR)**. *J Mol Biol* 2010, **395**:769-784.

63. Batista PR, Robert CH, Marechal JD, Hamida-Rebai MB, Pascutti PG, Bisch PM, Perahia D: **Consensus modes, a robust description of protein collective motions from multiple-minima normal mode analysis – application to the HIV-1 protease**. *Phys Chem Chem Phys* 2010, **12**:2850-2859.

64. Batista PR, Pandey G, Pascutti PG, Bisch PM, Perahia D, Robert CH: **Free energy profiles along consensus normal modes provide insight into HIV-1 protease flap opening**. *J Chem Theory Comput* 2011, **7**:2348-2352.

65. Torrie GM, Valleau JP: **Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling**. *J Comput Phys* 1977, **23**:187-199.

Please cite this article in press as: Ho BK, et al.: Hybrid approaches to molecular simulation, Curr Opin Struct Biol (2012), http://dx.doi.org/10.1016/j.sbi.2012.05.005

66. Spiwok V, Lipovova P, Kralova B: **Metadynamics in essential coordinates: free energy simulation of conformational changes**. *J Phys Chem B* 2007, **111**:3073-3076.

67. Tribello GA, Ceriotti M, Parrinello M: **A self-learning algorithm for biased molecular dynamics**. *Proc Natl Acad Sci USA* 2010, **107**:17509-17514.

68. Dykeman EC, Twarock R: **All-atom normal-mode analysis reveals an RNA-induced allostery in a bacteriophage coat protein**. *Phys Rev E Stat Nonlin Soft Matter Phys* 2010, **81**:031908.

69. Liu MS, Todd BD, Sadus RJ: **Allosteric conformational transition in adenylate kinase: dynamic correlations and implication for allostery**. *Aust J Chem* 2010, **63**:405-412.

70. Zheng W, Brooks BR, Thirumalai D: **Allosteric transitions in the chaperonin GroEL are captured by a dominant normal mode that is most robust to sequence variations**. *Biophys J* 2007, **93**:2289-2299.

71. Cavasotto CN, Kovacs JA, Abagyan RA: **Representing receptor flexibility in ligand docking through relevant normal modes**. *J Am Chem Soc* 2005, **127**:9632-9640.

72. Chakraborty S, Kumar S, Basu S: **Conformational transition in the substrate binding domain of beta-secretase exploited by NMA and its implication in inhibitor recognition: BACE1-myricetin a case study**. *Neurochem Int* 2011, **58**:914-923.

73. Floquet N, Marechal JD, Badet-Denisot MA, Robert CH, Dauchez M, Perahia D: **Normal mode analysis as a prerequisite for drug design: application to matrix metalloproteinases inhibitors**. *FEBS Lett* 2006, **580**:5130-5136.

74. Sperandio O, Mouawad L, Pinto E, Villoutreix BO, Perahia D, Miteva MA: **How to choose relevant multiple receptor conformations for virtual screening: a test case of Cdk2 and normal mode analysis**. *Eur Biophys J* 2010, **39**:1365-1372.

75. Asses Y, Venkatraman V, Leroux V, Ritchie DW, Maigret B: **Exploring c-Met kinase flexibility by sampling and clustering its conformational space**. *Proteins* 2012, **80**:1227-1238.

76. Bonvin AM: **Flexible protein-protein docking**. *Curr Opin Struct Biol* 2006, **16**:194-200.

77. Floquet N, Dedieu S, Martiny L, Dauchez M, Perahia D: **Human thrombospondin's (TSP-1) C-terminal domain opens to interact with the CD-47 receptor: a molecular modeling study**. *Arch Biochem Biophys* 2008, **478**:103-109.

78. May A, Zacharias M: **Energy minimization in low-frequency normal modes to efficiently allow for global flexibility during systematic protein-protein docking**. *Proteins* 2008, **70**:794-809.

79. Stein A, Rueda M, Panjkovich A, Orozco M, Aloy P: **A systematic study of the energetics involved in structural changes upon association and connectivity in protein interaction networks**. *Structure* 2011, **19**:881-889.

80. Zacharias M: **Accounting for conformational changes during protein-protein docking**. *Curr Opin Struct Biol* 2010, **20**:180-186.

81. Dobbins SE, Lesk VI, Sternberg MJ: **Insights into protein flexibility: the relationship between normal modes and conformational change upon protein-protein docking**. *Proc Natl Acad Sci USA* 2008, **105**:10390-10395.

82. Keseru GM, Kolossvary I: **Fully flexible low-mode docking: application to induced fit in HIV integrase**. *J Am Chem Soc* 2001, **123**:12708-12709.

83. Lindahl E, Delarue M: **Refinement of docked protein-ligand and protein-DNA structures using low frequency normal mode amplitude optimization**. *Nucleic Acids Res* 2005, **33**:4496-4506.

84. Moal IH, Bates PA: **SwarmDock and the use of normal modes in protein-protein docking**. *Int J Mol Sci* 2010, **11**:3623-3648.

85. Falke S, Tama F, Brooks CL 3rd, Gogol EP, Fisher MT: **The 13 angstroms structure of a chaperonin GroEL-protein substrate complex by cryo-electron microscopy**. *J Mol Biol* 2005, **348**:219-230.

86. Kim DN, Nguyen CT, Bathe M: **Conformational dynamics of supramolecular protein assemblies**. *J Struct Biol* 2011, **173**:261-270.

87. Tama F, Wriggers W, Brooks CL 3rd: **Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory**. *J Mol Biol* 2002, **321**:297-305.

88. Norgaard Toft K, Larsen N, Steen Jorgensen F, Hojrup P, Houen G, Vestergaard B: **Small angle X-ray scattering study of calreticulin reveals conformational plasticity**. *Biochim Biophys Acta* 2008, **1784**:1265-1270.

89. Wu Y, Tian X, Lu M, Chen M, Wang Q, Ma J: **Folding of small helical proteins assisted by small-angle X-ray scattering profiles**. *Structure* 2005, **13**:1587-1597.

90. Zheng W, Tekpinar M: **Accurate flexible fitting of high-resolution protein structures to small-angle x-ray scattering data using a coarse-grained model with implicit hydration shell**. *Biophys J* 2011, **101**:2981-2991.

91. Rader AJ: **Coarse-grained models: getting more with less**. *Curr Opin Pharmacol* 2010, **10**:753-759.

92. Saunders MG, Voth GA: **Coarse-graining of multiprotein assemblies**. *Curr Opin Struct Biol* 2012, **22**:144-150.

93. •• Moussavi-Baygi R, Jamali Y, Karimi R, Mofrad MR: **Biophysical coarse-grained modeling provides insights into transport through the nuclear pore complex**. *Biophys J* 2011, **100**:1410-1419.
This ground-breaking simulation has attempted to model a protein complex of unprecedented size. The model leverages the remarkable work of the Sali lab, who determined the domain architecture of the 450 proteins of the nuclear-pore complex, based on an analysis of multiple sources of experimental information. The model itself relies on a relatively small set of estimated biophysical parameters. The resultant simulations provide a synergistic view of cargo transport through the nuclear pore complex, relating concrete mechanistic details to computed kinetic parameters.

Please cite this article in press as: Ho BK, et al.: Hybrid approaches to molecular simulation, Curr Opin Struct Biol (2012), http://dx.doi.org/10.1016/j.sbi.2012.05.005