BELIEF PROPAGATION REVEALS ALLOSTERIC MECHANISMS IN PROTEINS

Hetunandan Kamisetty
Computer Science Department, Carnegie Mellon University,
Pittsburgh, PA 15213, USA
Email: hetu@cs.cmu.edu

Arvind Ramanathan
Lane Center for Computational Biology, Carnegie Mellon University,
Pittsburgh, PA 15213, USA
Email: aramanat@andrew.cmu.edu

Christopher James Langmead*
Computer Science Department, Carnegie Mellon University, and
Lane Center for Computational Biology, Carnegie Mellon University,
Pittsburgh, PA 15213, USA
**Email: cjl@cs.cmu.edu*

The physical mechanisms involved in allosteric regulation remain unclear. We present a novel and efficient method for investigating the propagation of regulatory signals in protein structures. Our approach utilizes undirected graphical models to efficiently encode the Boltzmann distribution over geometric configurations. Belief Propagation is then invoked to efficiently compute: (a) free energies and (b) allosteric couplings between distal residues. We present results from two kinds of experiments. First, we show that our method accurately predicts changes in free energy upon activation and/or mutation. Specifically, our method achieves a high correlation with experimentally determined $\Delta\Delta G$s ($R^2 = 0.90$ for core residues). Significantly, our method is capable of identifying those residues experiencing the largest relative changes in enthalpy and/or entropy. Second, we use our method to study the allosteric behavior of cyclophilin A in enzyme catalysis. Our analysis reveals the allosteric coupling between residues separated by as much as 20 angstroms from the active site. These results correspond well with experimental measurements. Our method requires a few minutes per protein, making it suitable for large-scale studies. Taken together, these results suggest that our method provides an effective means for investigating allosteric regulation at the proteome scale.

1. INTRODUCTION

Proteins are inherently dynamic molecules[6]. Increasing evidence from experiments as well as computational work suggests that this inherent flexibility is largely responsible for a protein's function. Several studies indicate the presence of distal couplings between residues, whereby an event such as binding is transmitted across the entire protein to influence catalysis or signaling (Fig. 1). The presence of such *allosteric* couplings between distal sites naturally leads to the question if there are preferred pathways through which conformational changes may propagate[5]. At the present time, there are very few methods for elucidating these pathways within protein structures.

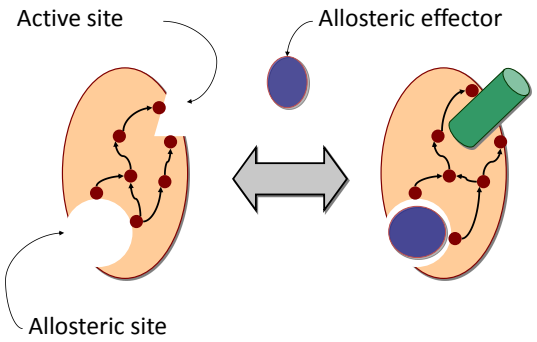

Fig. 1. Allosteric Couplings The binding of an allosteric effector induces structural changes at the active site, thus regulating the behavior of the protein.

*Corresponding author.

2. METHODS

We describe here a novel approach to reveal mechanisms of how conformational changes at one site can propagate to distal locations along the protein's structure. Our approach is based on a probabilistic modeling of the side-chain conformational space of a protein. As is common practice, we discretize the conformational space of each residue type using a rotamer library. If we use $\mathbf{R} = \langle R_1, R_2, \ldots, R_n \rangle$ to represent the n variate random variable representing the side-chain conformations of a protein backbone $\mathbf{b}$, our approach encodes $P(\mathbf{R}|\mathbf{b})$ accurately and efficiently using a *Markov Random Field* (MRF).

The MRF encoding exploits the fact that energetic interactions fall off rapidly with distance. This generic property of physical systems at this scale leads to conditional independencies between atoms with negligible interaction energies. Figure 2 illustrates a toy MRF for a 4-residue protein fragment. The MRF is overlayed on the protein. Nodes correspond to random variables over rotameric states.

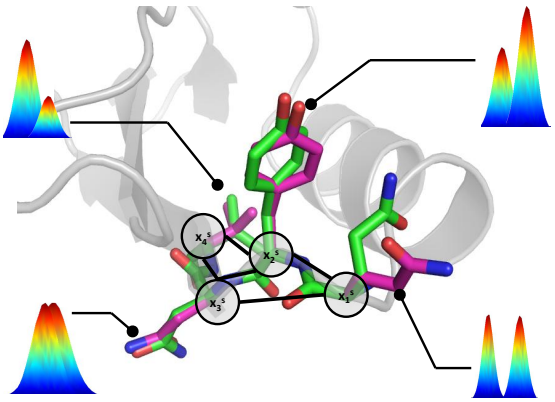

Fig. 2. A toy 4-residue model.

Given a MRF of a protein structure, we can compute the marginal conditional probabilities $P(R_i = r_i|\mathbf{b}, e)$ of each rotameric state r_i for every residue i of the protein using Belief Propagation[12]. Figure 2 illustrates these marginals as histograms. Our algorithm examines changes in these marginals in response to conformational perturbations, like binding. In previous work, we have shown that this probabilistic modeling accurately computes the global properties of the conformational distribution in proteins[9–11] and protein complexes[7, 8]. Moreover, running BP only takes a few minutes per structure, making it ideally suited to performing systematic perturbation studies.

3. RESULTS

We first demonstrate that our method accurately predicts changes in free energy upon mutation. As shown in Figure 3, our method achieves a high correlation between experimentally determined[4] and predicted $\Delta\Delta$Gs ($R^2 = 0.9$) for five point mutations in the core of eglin C (V13A, V14A, V34A, V54A, V62A).

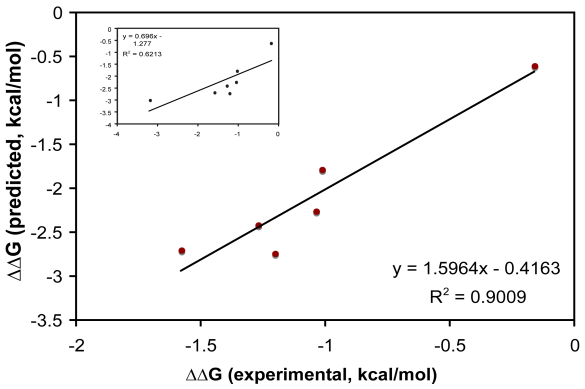

Fig. 3. Predicted $\Delta\Delta$G values show good agreement with experimental values. Inset shows predicted $\Delta\Delta$G values for all mutants showing outliers which are solvent exposed.

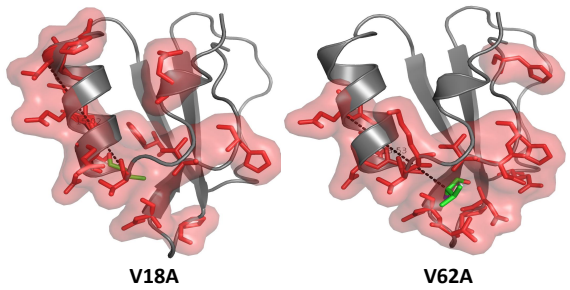

Fig. 4. Residues in eglin c showing significant changes in free-energy upon mutation.

An important feature of our method is that we are able to compute the residue-specific changes in free energy upon mutation. We use this information to identify

communication pathways. Figure 4 shows the results of two such mutations. In each case, the residue highlighted in green is the residue mutated, while red surface is used to highlight side-chain that are significantly perturbed by the mutation. Notice that a single mutation can induce changes throughout the protein.

This phenomenon is not unique to eglin C, nor does it require mutations. Figure 5 illustrates that significant free energy changes are detected upon ligation in regions distal to the active site in four unrelated proteins. The inactive protein shown as gray cartoon. The activated protein is shown as a backbone trace, colored by the change in free energy.

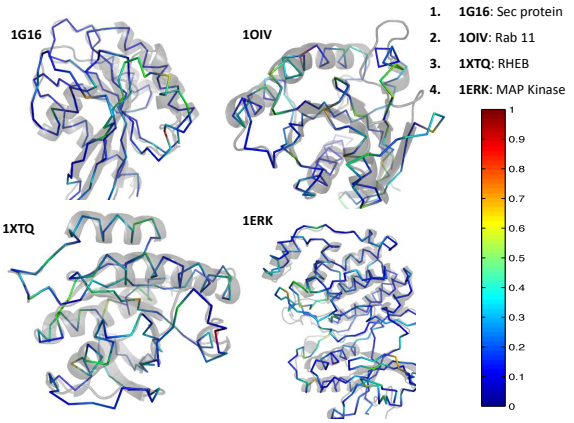

Fig. 5. Four allosteric proteins.

3.1. Identifying Allosteric Couplings

In this section, we examine in detail the allosteric coupling networks revealed by belief propagation in *cyclophilin A (cypA)* from 3 different species (H. sapiens, B. taurus, P. yeolii). CypA is an important receptor for several immunosuppressive drugs and HIV infection. Experimental work[3] and molecular dynamics (MD) simulations[1, 2] have identified a network of coupled motions that influences the substrate isomerization process. These networks consist of 27 residues exhibiting correlated motions, extending from the flexible surface regions all the way to the active site of the enzyme. The study provides us an ideal testing ground since the networks have been extensively characterized across multiple species and multiple substrates.

For every residue i in each structure of cypA, we systematically fixed its sidechain conformation to every possible rotamer r_i to assess the nature of conformational coupling that is intrinsic to the structure. We then compared the change in the marginal probabilities before and after the conformational change for every residue using the symmetric KL divergence ($KL_{sym}(P(R_j|\mathbf{b}, r_i), P(R_j|\mathbf{b}, r_j))$). For every residue across all conformations, we counted the number of residues that showed significant deviations in their marginals (N_a) by defining a suitable threshold. Using this metric and the distance of separation (D) between the two residues, we assessed if our method can distinguish between the behavior of network residues versus non-network residues. As shown in Table 1, network residues (**N**) affect a larger proportion of non-neighbor residues at longer distances than non-network residues (non-**N**). The ability of network residues to influence the conformational state of distal residues is statistically significant across structures, as well in each individual structure.

Table 1. Summary of network (N) and non-network (non-N) residue behavior across all species of CypA. μ and σ represent the mean and standard deviations of the respective classes of residues in each of the six categories (a) Hydrophobic/ Polar (H/P), (b) Number of affected residues (N_a), (c) average distance to affected residues (Avg(D)), (d) maximum distance to affected residue (Max(D)) (e) Number of neighbors at 10 angstrom cut-off (N_n) and (f) ratio of (c) and (e).

	μ(N)	σ(N)	μ(non-N)	σ(non-N)	p-value
H/P	0.637	0.4826	0.4085	0.492	$\ll 0.001$
N_a	14.3926	5.7732	9.6285	5.1538	$\ll 0.001$
Avg(D)	8.4333	1.1245	7.8289	1.4344	$\ll 0.001$
Max(D)	13.8646	3.181	12.2153	3.5086	$\ll 0.001$
N_n	21.8815	4.6857	18.0352	4.9721	$\ll 0.001$
N_a/N_n	0.656	0.2357	0.5367	0.2543	$\ll 0.001$

Next we examined whether conformational changes within the network residues affected similar sets of residues across different species. For this, we compared the intersection of residues affected by the changes in conformation of every network residue pair separated by at least 10 angstroms. We observed that irrespective of the species or substrate bound, the network residues affect the conformation of specific residues located across the entire protein. These residues are located on dynamically coupled regions in CypA affecting the catalytic process as observed in previous studies.

For suitable control, we considered those residues that were not part of the network yet conserved across different species. These residues are located proximal to

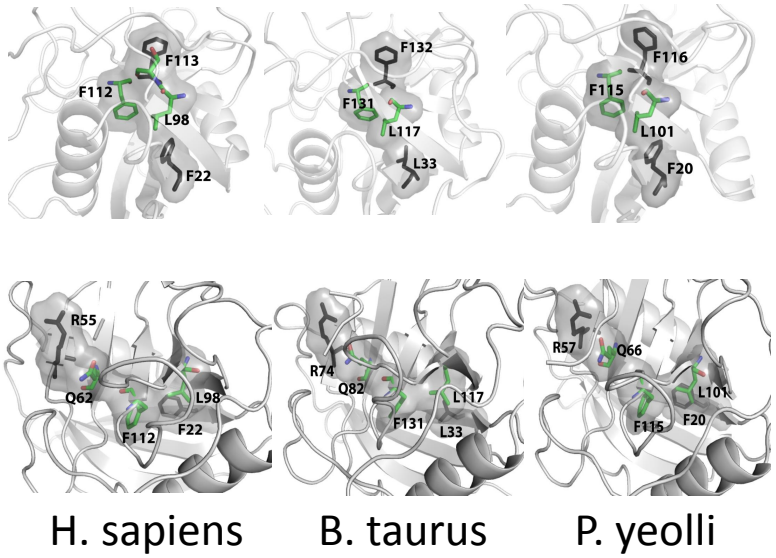

Fig. 6. Conservation of Allosteric pathways in cyclophilin A (CypA). Top row: Pathways detected via belief propagation analysis from F22-F113. Bottom row: Pathways detected via belief propagation analysis from F22-R55.

the network residues both structurally and sequentially. We observed that although the average distances between the network and non-network pairs of residues were similar, the non-network residues did not affect distally separated residues, in agreement with previous studies regarding coupled motions in CypA. This indicates a clear bias in the nature of conformational coupling exhibited by network residues to those of the non-network residues.

Finally, a careful inspection of the residues affected by changes in the conformations of the network also reveal similar pathways of conformational connectivity across multiple species. Example pathways are shown in Fig. 6. The top row left figure shows the pathway between F22 and F113 (human). The pathways between the corresponding residue in cow and *P. yoelii* are shown in the top middle and right figures. F22 is in the core of the protein, and F113 is on the surface and is where substrate binding happens[3]. These two residues are separated by over 12 angstroms. The belief propagation analysis reveals that these residues are connected via a similar set of hydrophobic interactions, irrespective of sequence homology. Additionally, this pathway exhibits a high correlation in terms of coupled motions as observed from MD simulations [2]. The bottom row shows conserved pathways from F22 to a different active residue (R55). This pathway is mediated by hydrogen bonds and hydrophobic interactions. Once again, the belief propagation analysis reveals a conservation in the pathways.

4. CONCLUSIONS

We have presented a novel approach to reveal mechanisms of distal conformational coupling within protein structures. The approach is physically based as it incorporates standard molecular mechanical force fields for computing internal energies and computes a rigorous approximation to the entropic contributions to the free energy. Our results suggest that belief propagation can be used to identify networks of residues that respond to various perturbations (mutations, binding, etc) in an efficient manner. These networks appear in a variety of proteins and our experiments on cypA suggests that they may be conserved to some degree across species. We believe that by using our approach, one may predict mechanisms of energy transfer between different parts of the protein and analyze allosteric regulation in protein structures. The availability of such accurate and efficient methods to understand protein function could be of significant use to biologists wanting to understand protein-protein interaction networks at a structural level.

5. ONGOING WORK

We are presently pursuing a number of extensions to this preliminary study. First, we have recently developed a MRF capable of modeling both backbone and side-chain flexibility[7, 8]. We are presently re-running these experiments to examine the role that backbone flexibility plays in signal transduction. Second, we are interested in the physical interpretation of message passing algorithms. It is known that the fixpoints of the belief propagation algorithm correspond to those of well-studied free energy-approximations[13]. However, the physical interpretation of the actual messages used in belief propagation is not well understood. Finally, we are examining a collection of well-characterized mutants to characterize our method's ability to predict which mutations affect the couplings between residues.

References

1. P. K. Agarwal. Cis/trans isomerization in hiv-1 capsid protein catalyzed by cyclophilin a: Insights from computational and theoretical studies. *Proteins: Struct., Funct., Bioinformatics*, 56:449–463, 2004.
2. P. K. Agarwal, A. Geist, and A. Gorin. Protein dynamics and enzymatic catalysis: Investigating the peptidyl-prolyl cis-trans isomerization activity of cyclophilin a. *Biochemistry*, 43(33):10605–10618, 2004.
3. Daryl A. Bosco, Elan Z. Eisenmesser, Susan Pochapsky, Wesley I. Sundquist, and Dorothee Kern. Catalysis of cis/trans isomerization in native hiv-1 capsid by human cyclophilin a. *Proc. Natl. Acad. Sci. USA*, 99(8):5247–5252, 2002.
4. Michael W. Clarkson, Steven A. Gilmore, Marshall H. Edgell, and Andrew L. Lee. Dynamic coupling and allosteric behavior in a nonallosteric proteinx2020;. *Biochemistry*, 45(25):7693–7699, 2006.
5. Nina M. Goody and Stephen J. Benkovic. Allosteric regulation and catalysis emerge via a common route. *Nat. Chem. Biol.*, 4(8):474–482, 2008.
6. K. Henzler-Wildman and D. Kern. Dynamic personalities of proteins. *Nature*, 450:964–972, 2007.
7. H. Kamisetty, C. Bailey-Kellogg, and C.J. Langmead. A graphical model approach for predicting free energies of association for protein-protein interactions under backbone and side-chain flexibility. Technical Report CMU-CS-08-162, Carnegie Mellon University, 2008.
8. H. Kamisetty, C. Bailey-Kellogg, and C.J. Langmead. A Graphical Model Approach for Predicting Free Energies of Association for Protein-Protein Interactions Under Backbone and Side-Chain Flexibility. In *Proc. Structural Bioinformatics and Computational Biophysics (3DSIG)*, pages 67–68, 2009.
9. H. Kamisetty, B. Ghosh, C. Bailey-Kellogg, and C.J. Langmead. Modeling and Inference of Sequence-Structure Specificity. In *Proc. of the 8th International Conference on Computational Systems Bioinformatics (CSB)*, pages 91–101, 2009.
10. H. Kamisetty, E. P. Xing, and C. J. Langmead. Free energy estimates of all-atom protein structures using generalized belief propagation. *J. Comp. Biol.*, 15(7):755–766, September 2008.
11. H. Kamisetty, E.P. Xing, and C.J. Langmead. Free Energy Estimates of All-atom Protein Structures Using Generalized Belief Propagation. In *Proc. 7th Ann. Intl. Conf. on Research in Comput. Biol. (RECOMB)*, pages 366–380, 2007.
12. J. Pearl. Fusion, propagation, and structuring in belief networks. *Artif. Intell.*, 29(3):241–288, 1986.
13. J.S. Yedidia, W.T. Freeman, and Weiss Y. Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Transactions on Information Theory*, 51:2282–2312, 2005.