Detecting Protein-Protein Interaction Decoys using Fast Free Energy Calculations

Christopher James Langmead[*†], **Hetunandan Kamisetty**[*]

October 2007
CMU-CS-07-156

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

[*]Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA 15213.
[†] Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, PA, USA 15213

E-mail: cjl@cs.cmu.edu

This research is supported by a Young Pioneer Award from the Pittsburgh Lifesciences Greenhouse and a CAREER award from the U.S. Department of Energy.

Keywords: Markov Random Fields, Generalized Belief Propagation, Free Energy, Protein-Protein Interactions

Abstract

We present a physics-based method for identifying native configurations of protein-protein interactions amongst a set of nearly native decoys (< 2.0 Å C_α RMSD to the native structure) using a fast new method for performing free energy calculations. The method uses Markov Random Fields to encode the Boltzmann distribution for a given complex, and Generalized Belief Propagation to perform the free energy calculation. Our method is fast, running in a few minutes on a single-processor workstation, making it an attractive alternative to free-energy calculations based molecular dynamics and Monte Carlo simulations, which can require hours or days on multiprocessor machines. The method is also accurate; in an experiment involving 9 targets with an average of 8 nearly native decoys, our method ranks the native structure number one 67% of the time, and in the top three for the remaining cases.

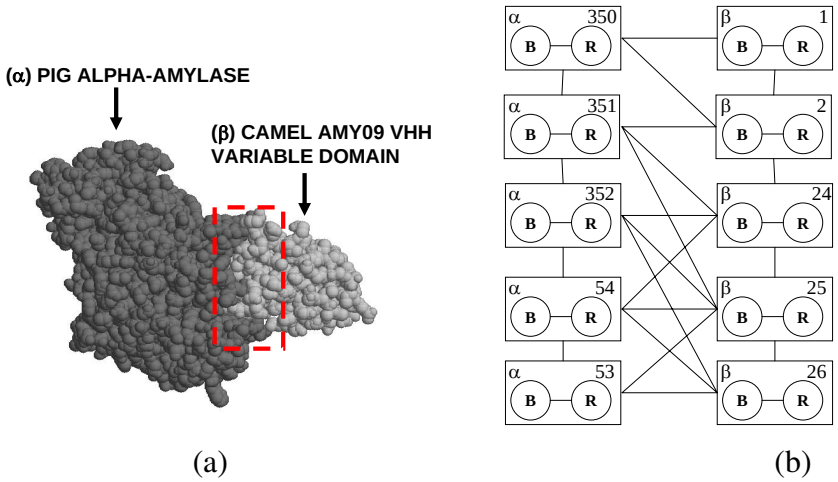
1 Introduction

Protein structure prediction is among the most challenging problems in molecular modeling. There are two primary sub-problems associated with structure prediction. The *first* problem concerns searching the space of possible configurations. The *second* problem involves scoring each configuration for the purpose of selecting the "best" structure. While significant progress has been made over the past decade towards addressing both problems, it is fair to say that the protein structure prediction problem remains an open problem.

The method presented in this paper addresses the problem of scoring putative structures of protein-protein complexes. Specifically, we present a method capable of identifying the native structure of a protein-protein complex among a set of nearly-native *decoys*. In the protein folding literature, decoys are generally defined to be structures that are structurally different than the native configuration, but are nevertheless indistinguishable from the native configuration in terms of their internal energies. We will define the phrase "nearly-native decoy", to mean a decoy that is within < 2.0 Å C_α RMSD of the native structure. Distinguishing native structures from nearly native structures is a particularly challenging task; indeed it has been conjectured that the class of energy functions used in this paper (those involving sums of pairwise interactions) *cannot* distinguish native structures from nearly native decoys [1, 18].

Our approach differs from previous work in that it involves the calculation of a physics-based *free energy*, as opposed to a potential energy or a statical-based free-energy. The free energy of a system is defined as $G = E - TS$, where E is the enthalpy of the system, T is the absolute temperature, and S is the entropy of the system. Existing scoring functions fall into one of two categories. The first class essentially computes the enthalpy (i.e., E), which includes contributions due to hydrogen bonds, electrostatic interactions, etc. Such approaches are effectively modeling the system at 0 Kelvin. The second class includes statistical potentials which are derived from analyses of the contents of the Protein Data Bank (PDB). Our approach, in contrast, involves encoding the Boltzmann distribution over structures using Markov Random Fields (MRFs) and performing free energy calculations (at any temperature) using Generalized Belief Propagation (GBP). Traditional methods for performing free energy calculations invoke either molecular dynamics or Monte Carlo simulations, which can take many hours or days on multi-process machines. GBP-based free energy calculations, on the other hand, can be performed in minutes on a single processor workstation, suggesting it is well-suited for large-scale proteomic studies.

We show that our method is capable of discerning the subtle differences between the native structure and each decoy by examining the differences in the entropic contributions to the free energies of each structure. In particular, we find that the native structure has the highest entropy (and thus the lowest free energy) 67% of the time, and within the top three highest entropies 100% of the time. These results shed new light onto the previously cited conjecture. If the conjecture is true, one likely explanation is that entropy contributions become significant when structures are similar. Our findings are consistent with this hypothesis.



(a) (b)

Figure 1: (a) Complex of pig alpha-amylase and camel amy09 VHH variable domain. The interface region is outlined. (b) Part of the Markov Random Field induced by the outlined residues. Each square (aka plate) contains 2 random variables, B and R, representing the configuration of the backbone atoms and side-chain atoms for that residue, respectively. Upper left corner of each plate indicates the chain. Upper right corner indicates the residue number. Edges between random variables (and plates) indicate the physical interactions being modeling by the MRF. Note, the complete MRF for this complex has 613 plates, one for each residue in the complex.

2 Modeling Protein Complexes with MRFs

Markov Random Fields are a subclass of *probabilistic graphical models* (PGM). Probabilistic graphical models comprise a family of techniques for representing and computing over complex multivariate probability distributions. Markov Random Fields are particularly well suited for physics-based modeling because, unlike some other kinds of PGMs, they are capable of modeling arbitrary interactions among a set of random variables. In the context of molecular modeling, this means that MRFs can model long-range (i.e., non-bonded) interactions.

A MRF for a complex consisting of n atoms has a total of n random variables. However, in order to simplify the explanation of our model and its visualization, we will group atoms together by type (either backbone or side-chain). Thus all random variables are assumed to vector-valued.

In what follows, random variables are represented using upper case variables, sets of random variables appear in bold face, while lower case variables represent specific values that the random variables can assume. For example, if a particular complex consists of chains $\alpha, \beta, \gamma, ...$, we will denote the configuration of all the backbone atoms of chain α as the random variable $\mathbf{B}_\alpha$, and the configuration of all the side chain atoms for that chain as the random variable $\mathbf{R}_\alpha$. B^i_α (resp. R^i_α) is the random variable representing the backbone (resp. side-chain) configuration of the i^{th} residue of chain α, and b^i_α (resp. r^i_α) represents a particular configuration of the backbone (resp. side-chain) of the i^{th} residue in chain α.

Let $\mathbf{X} = \{\mathbf{B}_\alpha, \mathbf{R}_\alpha, \mathbf{B}_\beta, \mathbf{R}_\beta, ...\}$ be the random variable representing the phase space of a com-

plex consisting of chains α, β, etc. Each $\mathbf{B}$ and $\mathbf{R}$ can be encoded in terms of either Cartesian or internal coordinates. A MRF is a compact encoding of $P(\mathbf{X})$ as an undirected graph and a set of potential functions (Fig. 1). More formally, a MRF, M, is a pair $M = (\mathcal{G}, \Phi)$ where $\mathcal{G} = (V, E)$ is an undirected graph, and Φ is a set of potential functions over the maximal cliques in $\mathcal{G}$, $C(\mathcal{G})$. The graph's vertex set $V = \{V_1, V_2, ..., V_n\}$ is isomorphic to the set of variables (i.e., $\mathbf{X}$) and we will make no distinction between the ith vertex and the ith random variable. Each edge $e = \{u, v\} \in E$, represents a dependency between random variables $u \in V$ and $v \in V$. Each potential, ϕ_c, is a mapping from the possible joint assignments of the elements of c to the positive reals.

Figure 1-B depicts a small portion of the MRF encoding the pig alpha-amylase and camel amy09 VHH variable domain complex. Here, we have grouped the random variables for each residue to simplify the visualization. There is one *plate* for each residue, and each plate contains the **B** and **R** variables for that residue. Edges between each **B** and **R** indicate that these random variables are dependent. Edges between plates i and j are a short-hand for the set of edges: $(\mathbf{B}_i, \mathbf{B}_j)$,$(\mathbf{B}_i, \mathbf{R}_j)$,$(\mathbf{R}_i, \mathbf{B}_j)$, and $(\mathbf{R}_i, \mathbf{R}_j)$. The figure only shows ten plates corresponding to ten residues in the interface region of the complex. The full MRF, as used in our experiments, contains one such plate per residue in the complex being studied.

A MRF encodes the following joint probability distribution:

$$P(\mathbf{X} = \mathbf{x}) = \frac{\mathbf{1}}{\mathbf{Z}} \exp\left(\sum_{\mathbf{c} \in \mathbf{C}(\mathcal{G})} \phi_{\mathbf{c}}(\mathbf{x}_{\mathbf{c}}) \right) \tag{1}$$

where $\mathbf{x}_c$ is the state of the variables in clique c, and Z is the partition function:

$$Z = \sum_{\mathbf{x} \in \mathbf{X}} \exp\left(\sum_{c \in C(\mathcal{G})} \phi_c(\mathbf{x}_c) \right). \tag{2}$$

Note that it is *not* necessary for the potential functions to be probability density functions.

Naturally, the probability distribution we are interested in is the one corresponding to *Boltzmann's law*:

$$P_B(\mathbf{x}) = \frac{\mathbf{1}}{\mathbf{Z}(\mathbf{T})} \exp\left(\frac{-\mathbf{E}(\mathbf{x})}{\mathbf{k_B T}} \right) \tag{3}$$

where $Z(T)$ is a temperature-sensitive partition function, $E(\mathbf{x})$ is the potential energy of a particular configuration, $\mathbf{x}$, k_B is Boltzmann's constant, and T is the temperature in degrees Kelvin. We can ensure that the MRF follows this distribution by defining the potential functions in terms of *Boltzmann factors*:

$$\phi_c(\mathbf{x}_c) \equiv \exp\left(\frac{-E(\mathbf{x}_c)}{k_B T} \right) \tag{4}$$

where the potential energy E can be computed using any molecular mechanics or statistical potential.

This completely defines the MRF. In principle, the graph $\mathcal{G}$ should be fully connected, reflecting the fact that each atom interacts with every other atom in the system. Unfortunately, the complexity of the algorithm for performing free energy calculations is dependent on the size of the cliques in the graph which is, of course, a function of the number of edges in the graph. For this reason, we make a common simplifying assumption and restrict the set of edges in the graph to those involving atoms close enough to make a non-negligible contribution to the total energy. Recall that our task is to score a set of *given* structures. We can therefore reduce the number of edges by applying a distance threshold. That is, only those atoms that are less than some pre-defined distance apart in the given structure are connected by an edge. The topology of the graph thus corresponds to the coarse-grained topological features of the complex. An alternative strategy is to define the edges based on the strength of interaction between particles. Using this method one can, for example, include edges between distant strongly charged atoms in an effort to better model long-range electrostatic interactions. Regardless of which criteria are used to select edges, we note that the edges do not prevent the complex from changing configurations; the edges merely dictate which interactions will be modeled.

We note that a MRF-based approach to representing molecules is compatible with traditional representations, like the Protein Data Bank (PDB) format. As outlined above, a PDB file can be used to construct the MRF by using it to define the set of edges. It is also possible to select one (or more) configurations encoded in the MRF and then construct and output a PDB formatted file or any other standard format.

3 Free Energy Calculations with Markov Random Fields

The *Helmholtz free energy*, F, of a system is a thermodynamic quantity defined as:

$$F = -\ln Z. \tag{5}$$

for a closed system at constant temperature. It is minimized when the system is at equilibrium. Thus, we will assume that our complexes are at equilibrium.

There has been a considerable amount of work by physicists at developing various approximations to estimate the value of Eq. 5 (e.g., [2, 9, 13, 14]). While the properties of these approximations have been extensively studied, there have been comparatively few algorithms for actually computing these approximations. Recently, however, it has been shown that a family of inference algorithms are mathematically equivalent to certain approximations of Eq. 5 [24]. In particular, Pearl's *Belief Propagation* algorithm [16] is equivalent to the Bethe approximation [2] of F. This discovery led to the development of the *Generalized Belief Propagation* (GBP) algorithm [23] which is equivalent to the Kikuchi approximation [9] to F. The Kikuchi approximation is a better approximation than the Bethe approximation [24], and GBP has been shown to be much more efficient than previous algorithms for computing the Kikuchi approximation [17].

BP and GBP adopt a variational approach. Consider a computationally tractable probability distribution, $Q(\mathbf{X})$, that approximates the computationally intractable distribution $P(\mathbf{X})$. We define the variational, or *Gibbs free energy*, G, as follows:

$$G = U(Q) - TS(Q) \tag{6}$$

where

$$U(Q) = \sum_{\mathbf{x}} Q(\mathbf{x}) E(\mathbf{x}) \tag{7}$$

is the variational average energy, and

$$S(Q) = -\sum_{\mathbf{x}} Q(\mathbf{x}) \ln Q(\mathbf{x}) \tag{8}$$

is the variational entropy.

It follows from the definitions that

$$G = F + D(Q||P) \tag{9}$$

where

$$D(Q||P) \equiv \sum_{\mathbf{x}} Q(\mathbf{x}) \ln \frac{Q(\mathbf{x})}{P(\mathbf{x})} \tag{10}$$

is the Kullback-Liebler divergence, or relative entropy between $P(\mathbf{X})$ and $Q(\mathbf{X})$. It can be shown that the Kullback-Liebler divergence is always non-negative, and zero only when $P(\mathbf{X}) = Q(\mathbf{X})$. Thus, $G \geq F$. That is, the Gibbs free energy is an upper bound on the Helmholtz free energy.

Notice that any algorithm which minimizes G is also an exact method for computing both $P(\mathbf{X})$ and F. Belief Propagation and GBP are algorithms for minimizing G. The term 'belief' in both BP and GBP refers to the marginal distributions of $Q(\mathbf{X})$ that these algorithms compute. This is the essential idea behind MRF-based free energy calculations. Of course, these same marginals can also be used to identify lowest energy (i.e., highest probability) configurations (e.g., [22]).

3.1 MRF-based Decoy Detection

Our hypothesis is that the native structure can be distinguished from nearly native decoys by considering the free energy of complex (as opposed to the potential energy). Decoys are generally defined as having similar or identical potential energies to the native structure. Thus, if our hypothesis is correct, the native structure is the one with the highest entropy. This suggests a straightforward approach to identifying the native structure — compute the free energy of each complex using GBP and then rank structures by entropy. In Section 5 we describe the results of a set of experiments designed to test our hypothesis.

4 Related Work

Probabilistic graphical models have been used to address a number of problems in structural biology, primarily in the area of secondary structure prediction (e.g., [4]). Applications of graphical models to tertiary structure are generally limited to applications of Hidden Markov Models (HMMs) (e.g., [8]). HMMs make severe independence assumptions to allow for efficient learning and inference, the result of which is that long-range interactions cannot be modeled. Markov Random Field-based molecular modeling was first introduced in 2002 when Yanover and Weiss used it to perform side-chain placement for fixed backbones [22]. Subsequent applications of MRF modeling include: force-field parameterizations [21], constructing protein backbone traces [6] and all-atom models [5] from electron density maps, protein sequence design [20], and fold recognition/threading [10]. Recently [7], we have applied MRFs for the specific task of free energy calculations. In that work we showed that the quantitative predictions concerning mutants and $\Delta\Delta G$ are well-correlated with experimental values. We also showed that free-energy calculations are sufficient for identifying native structures for individual proteins. In this paper, we extend that work to decoy detection for nearly-native protein complexes.

5 Results

Our hypothesis was that free energy calculations can be used to distinguish native from nearly native structures (Sec. 3.1). If this hypothesis is true, then the native structure will be at the top of a list of structures ranked by decreasing entropy. We tested this hypothesis by assembling a number of data sets, performing free energy calculations with GBP, and ranking structures in order of decreasing entropy.

5.1 Data and Assumptions

Data were obtained from the CAPRI (CRITICAL ASSESSMENT OF PREDICTION OF INTERACTIONS) website [15] and a collection from the Vakser lab [19]. The criterion for inclusion were as follows: a) structure data were available for both the native structure and decoys, b) the decoys have 100% sequence identity to the native structure, c) the existence of one or more decoys with $\leq$ 2.0 Å C_α RMSD to the native structure, d) no HETATM records (except waters), and e) no missing backbone atoms. Using these criteria, a total of 9 targets were studied. There were an average of 8 decoys for each of the 9 targets.

The data for each target includes the structure for the native conformation and one or more decoy conformations. For each target, $T_i = \{t_1, t_2, ..., t_n\}$, a separate MRF was constructed, $\{M_1, M_2, ..., M_n\}$, in the manner outlined in Section 2. The set of edges in the graph for each M_i were between those atoms that are within 8.0 Å of each other in the corresponding t_i.

In principle, MRFs are capable of modeling full conformational flexibility. Our implementation, however, makes two simplifying assumptions. *First*, we model the backbones of each chain within each complex, t_i, as being rigid. Using the chain rule of probability, the probability of any

Description	# Residues	# Structures	Rank of Native Str.
CAPRI target 6 [15]	613	8	1
CAPRI target 7 [15]	455	6	1
CAPRI target 12 [15]	193	19	2
CAPRI target 14 [15]	581	17	1
CAPRI target 19 [15]	533	8	1
1a2p-1a19 [19]	197	5	1
1sup-2ci2 [19]	340	3	2
2ptn-4pti [19]	281	4	3
5cha-2ovo [19]	292	4	1

Table 1: **Data and Results** Our experiments were performed on data drawn from two sources (see text for details). Column 1 describes the data set. Column 2 lists the total number of residues in the complex. Column 3 is the size of the data set (number of decoys + the native structure). Column 4 is the rank of the native structure, as determined by our method. The native structure is ranked number one in 6 out of 9 targets.

particular configuration of the complex, $\mathbf{x}$, can be factored as:

$$P(\mathbf{X} = \mathbf{x}|\mathbf{\Theta}) = \mathbf{P}(\mathbf{B} = \mathbf{b})\mathbf{P}(\mathbf{R} = \mathbf{r}|\mathbf{B} = \mathbf{b}, \mathbf{\Theta})$$

, where Θ represents any parameters used to describe this model, including sequence information, temperature etc. Under our assumption of a rigid backbone, $P(\mathbf{B} = \mathbf{b}) = \mathbf{1}$. We do, however, consider the flexibility of the side chains of each chain in the complex. That is, we rank structures based on the conditional distribution $P(\mathbf{R}|\mathbf{B} = \mathbf{b}, \mathbf{\Theta})$. *Second*, we model the flexibility of the side chains using a discrete rotamer library. In particular, we used a backbone-dependent rotamer library [11]. That rotamer library defines prior probabilities on each rotameric configuration. These priors were incorporated into the MRF.

Our potentials were defined using the force-field used in the side-chain placement prediction program SCWRL [3], which consists of a linear approximation to the repulsive van der Waals force. The remaining parameter of our method, temperature, was set to 315 K. Free energy calculations were performed using GBP implemented in the Java programming language and run on a single processor Linux 3.2GHz workstation. Each free energy calculation required 5 minutes or less.

The structures were ranked in terms of decreasing free energy as calculated by GBP. The results are shown in Table 1. The native structure is ranked number one for 6 out of 9 targets (67% accuracy). In the remaining 3 targets, the native structure was ranked number either two or three. The predicted free energy and the RMSD were not significantly correlated (corr. coef. = 0.3). We note that while the present study does not consider the problem of nearly native structures among more distant decoys, as is more natural in the context of structure prediction, we have demonstrated previously [7] that our method is capable of performing that task with high accuracy (84%). We are presently conducting a more comprehensive test of our method for protein complexes.

6 Discussion and Conclusion

Free energy is a thermodynamic quantity that either directly or indirectly governs many of the physical properties of interest to scientists. Consequently, one of the most fundamental and important tasks in molecular modeling is the calculation of free energies. Unfortunately, traditional methods for performing free energy calculations, based on either molecular dynamics (MD) or Monte Carlo (MC) simulations, are expensive, requiring hours or days on multi-processor machines. MRF-based free energy calculations are potentially an attractive alternative to MD and MC-based methods because they are orders of magnitude faster. Our experiments required several minutes on single processor workstation using un-optimized codes.

More importantly, our results suggest that MRF-based free energy calculations may be sufficiently accurate to distinguish native from nearly native structures of protein complexes. This represents a new strategy for decoy detection; existing methods score structures using either a potential energy function or a statistical approximation of the free energy. One of the most interesting aspects of our results is that they are obtained making severe simplifying assumptions (i.e., rigid backbone and discrete side-chain configurations) and an extremely simple energy potential. The use of the simple potential is especially interesting because it has previously been conjectured [1, 18] that such potentials are incapable of distinguishing native from nearly native structures. We believe that highly similar structures may be best evaluated in terms of their free energy and/or entropy. The computational efficiency of GBP makes it well suited to performing these free energy calculations.

We are presently performing decoy detections on a much larger set of benchmark protein complex decoys [12]. Early results are promising and consistent with those presented here. For example, our method ranks the native structure of a multi chain complex (PDB ID: 1AHW) number 1 among a set of 58 nearly native decoys. However, on a different complex (PDB ID: 1AKJ), our method ranks the native structure last among 8 nearly native decoys. Closer inspection revealed that in this case, the native structure has one fewer residues than the decoy structures. The free energies of structures with different sequences are not immediately comparable in the context of decoy detection. This sensitivity to the number of residues is not unexpected, but it does reveal that our approach is brittle when it comes to the precise details of the structures being compared.

There are several additional limitations of our existing implementation of MRF-based free energy calculations. First, our method allows only limited flexibility. In previous work [7], we have demonstrated that MRF-based free energy calculations are well-correlated with experimental values, but the absolute magnitude of the predictions can be very different than experimental values. The limited flexibility of our approach is likely one source of error in our calculations, and we are presently working on the development of new algorithms that allow full flexibility.

Acknowledgments

This research was supported by a U.S. Department of Energy Career Award (DE-FG02-05ER25696), and a Pittsburgh Life-Sciences Greenhouse Young Pioneer Award to C.J.L.

References

[1] M. R. Betancourt and D. Thirumalai. Pair potentials for protein folding: choice of reference states and sensitivity of predicted native states to variations in the interaction schemes. *Protein Science*, 8:361–369, 1999.

[2] H. A. Bethe. Statistical theory of superlattices. *Proc. Roy. Soc. London A*, 150:552–575, 1935.

[3] A. Canutescu, A. A. Shelenkov, and R. L. Dunbrack Jr. A graph theory algorithm for protein side-chain prediction. *Protein Science*, 12:2001–2014, 2003.

[4] W. Chu, Z. Ghahramani, and D. Wild. A graphical model for protein secondary structure prediction. *Proc. 21st Ann. Intl. Conf. on Machine Learning (ICML) Banff, Canada)*, 2004.

[5] F. DiMaio, A. Soni, G. N Jr. Phillips, and J.W. Shavlik. Creating all-atom protein models from electron-density maps using particle-filtering methods. *Bioinformatics*, in press, 2007.

[6] Frank Dimaio, Jude Shavlik, and George N. Phillips. A probabilistic approach to protein backbone tracing in electron density maps. *Bioinformatics*, 22(14):e81–e89, 2006.

[7] H. Kamisetty, E.P. Xing, and C.J. Langmead. Free Energy Estimates of All-atom Protein Structures Using Generalized Belief Propagation. In *Proc. of the 7th Ann. Intl. Conf. on Research in Comput. Biol. (RECOMB)*, pages 366–380, 2007.

[8] K. Karplus, R. Karchin, J. Draper, J. Casper, Y. (Mandel-Gutfreund), M. Diekhans, and R. Hughey. Combining local-structure, fold-recognition, and new-fold methods for protein structure prediction. *Proteins*, 53:491–6, 2003.

[9] R. Kikuchi. A theory of cooperative phenomena. *Phys. Rev*, 81:988–1003, 1951.

[10] Y. Liu, J. Carbonell, P. Weigele, and V. Gopalakrishna. Segmentation conditional random fields (scrfs): A new approach for protein fold recognition. *Proc. of the 9th Ann. Intl. Conf. on Comput. Biol. (RECOMB) Boston, MA, May 14-18)*, pages 408–422, 2005.

[11] S.C. Lovell, J.M. Word, J.S. Richardson, and D.C. Richardson. The Penultimate Rotamer Library. *Structure Function and Genetics*, 40:389–408, 2000.

[12] Mintseris, J. and Wiehe, K. and Pierce, B. and Anderson, R. and Chen, R. and Janin, J. and Weng, Z. Protein-Protein Docking Benchmark 2.0: an Update. *Proteins*, 60(2):214–6, 2005.

[13] T. Morita. Cluster variation method for non-uniform ising and heisenberg models and spin-pair correlation function. *Prog. Theor. Phys.*, 85:243 – 255, 1991.

[14] T. Morita, T. M. Suzuki, K. Wada, and M. Kaburagi. Foundations and applications of cluster variation method and path probability method. *Prog. Theor. Phys. Supplement*, 115, 1994.

[15] National Institute of General Medical Sciences. CAPRI: Critical Assessment of PRediction of Interactions, 2007. URL: `http://capri.ebi.ac.uk/capri.html`.

[16] Judea Pearl. Fusion, propagation, and structuring in belief networks. *Artif. Intell.*, 29(3):241–288, 1986.

[17] A. Pelizzola. Cluster variation method in statistical physics and probabilistic graphical models. *J. phys. A : math. gen*, 38:R309–R339, 2005.

[18] R. Tobi, D. Elber. Distance-dependent, pair potential for protein folding: Results from linear optimization. *Proteins: Structure, Function and Genetics*, 41:40–46, 2000.

[19] Vakser, I.A. Database of protein-protein decoys for the validation of energy functions and refinement procedures. http://www.bioinformatics.ku.edu/files/vakser/decoys/database.html.

[20] C. Yanover, T. Meltzer, and Y. Weiss. Linear programming relaxations and belief propagation – an empirical study. *Journal of Machine Learning Research*, 7:1887–1907, 2006.

[21] C. Yanover, O. Schueler-Furman, and Y. Weiss. Minimizing and Learning Energy Functions for Side-Chain Prediction. In *Proc. of the 7th Ann. Intl. Conf. on Research in Comput. Biol. (RECOMB)*, pages 381–395, 2007.

[22] Yanover, C. and Weiss, Y. Approximate Inference and Protein Folding. *NIPS*, pages 84–86, 2002.

[23] J.S. Yedidia, W.T. Freeman, and Y. Weiss. Generalized belief propagation. *Advances in Neural Information Processing Systems (NIPS)*, 13:689–695, 2000.

[24] 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.