



REVIEW

Sequence co-evolutionary information is a natural partner to minimally-frustrated models of biomolecular dynamics [version 1; referees: 3 approved]

Jeffrey K Noel^{1,2}, Faruck Morcos³, Jose N Onuchic¹

¹Center for Theoretical Biological Physics, Rice University, Houston, TX, USA

²Kristallographie, Max-Delbrück-Centrum für Molekulare Medizin, Berlin, Germany

³Department of Biological Sciences, University of Texas at Dallas, Richardson, TX, USA

v1 First published: 26 Jan 2016, 5(F1000 Faculty Rev):106 (doi: [10.12688/f1000research.7186.1](https://doi.org/10.12688/f1000research.7186.1))

Latest published: 26 Jan 2016, 5(F1000 Faculty Rev):106 (doi: [10.12688/f1000research.7186.1](https://doi.org/10.12688/f1000research.7186.1))

Abstract

Experimentally derived structural constraints have been crucial to the implementation of computational models of biomolecular dynamics. For example, not only does crystallography provide essential starting points for molecular simulations but also high-resolution structures permit for parameterization of simplified models. Since the energy landscapes for proteins and other biomolecules have been shown to be minimally frustrated and therefore funneled, these structure-based models have played a major role in understanding the mechanisms governing folding and many functions of these systems. Structural information, however, may be limited in many interesting cases. Recently, the statistical analysis of residue co-evolution in families of protein sequences has provided a complementary method of discovering residue-residue contact interactions involved in functional configurations. These functional configurations are often transient and difficult to capture experimentally. Thus, co-evolutionary information can be merged with that available for experimentally characterized low free-energy structures, in order to more fully capture the true underlying biomolecular energy landscape.



This article is included in the **F1000 Faculty Reviews** channel.

Open Peer Review

Referee Status:

	Invited Referees		
	1	2	3
version 1 published 26 Jan 2016			

F1000 Faculty Reviews are commissioned from members of the prestigious F1000 Faculty. In order to make these reviews as comprehensive and accessible as possible, peer review takes place before publication; the referees are listed below, but their reports are not formally published.

- 1 **Shoji Takata**, Kyoto University Japan
- 2 **Angel Garcia**, Rensselaer Polytechnic Institute USA
- 3 **Sudip Kundu**, University of Calcutta India

Discuss this article

Comments (0)

Corresponding author: Jose N Onuchic (jonuchic@rice.edu)

How to cite this article: Noel JK, Morcos F and Onuchic JN. **Sequence co-evolutionary information is a natural partner to minimally-frustrated models of biomolecular dynamics [version 1; referees: 3 approved]** *F1000Research* 2016, 5(F1000 Faculty Rev):106 (doi: [10.12688/f1000research.7186.1](https://doi.org/10.12688/f1000research.7186.1))

Copyright: © 2016 Noel JK *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Grant information: This work was supported by the Center for Theoretical Biological Physics sponsored by the National Science Foundation (grants PHY-1427654 and NSF-MCB-1214457). Jeffrey K. Noel is supported in part by the Welch Foundation (grant C-1792). *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

Competing interests: The authors declare that they have no competing interests.

First published: 26 Jan 2016, 5(F1000 Faculty Rev):106 (doi: [10.12688/f1000research.7186.1](https://doi.org/10.12688/f1000research.7186.1))

Introduction

High-resolution structural techniques (e.g., X-ray crystallography and nuclear magnetic resonance) have provided the data necessary to develop and refine a multitude of potential energy functions used in the simulation of biomolecules. In particular, these structures provide the parameterization for simplified models that are based on the energy landscape theory of protein folding. These models construct an energetically unfrustrated (ideal) funneled landscape by including stabilizing interactions between native contacts (i.e., amino acid pairs that are nearby in the three-dimensional native structure of a protein). In cases in which experimental structures are lacking or insufficient, it becomes necessary to supplement these models with other sources of contact information. An emerging technique for contact estimation is via the statistical analysis of residue co-evolution in families of protein sequences. Combinations of high-resolution structural data and predictions from residue co-evolution are proving to be invaluable tools for building models to study protein structure and dynamics.

Understanding the fundamental process of how a heterogeneous polypeptide can reversibly fold into a distinct native three-dimensional structure on biological timescales led to the development of the energy landscape theory of biomolecular folding. This theory is based on the principle of minimal frustration¹ and the folding funnel concept^{2,3}. These physical principles describe an energy landscape that has been molded by evolution such that the native interactions (i.e., the molecular interactions present in low free-energy configurations of folded proteins and RNAs) are, on average, more stabilizing than non-native interactions. The consequence of proteins having sufficiently reduced energetic frustration is that geometry dominates energetic roughness in determining folding mechanisms. Thus, a description of the effective energetics of the folding phenomenon can be attained by including a set of native stabilizing interactions consistent with the native basin of attraction. Potential energy functions of this type, which use experimental information to determine such native interactions, are known as “structure-based models” (SBMs)^{4–6} and, when employed in dynamical models, are powerful tools for understanding the connection between structure, folding, and function. Although these SBMs have been successfully applied to different biomolecules, we will be focusing on proteins for clarity in this review.

Structural information, however, may be limited for many interesting systems. This is particularly true for functional configurations that are transient or partially disordered or both. The recent explosion in genomic information has enabled complementary methods for discovering functionally important amino acid interactions. The minimal frustration principle applies equally to any sequence of amino acids that can robustly fold to a particular native structure. Thus, in a family of sequences where most of them fold to a common structure, residue positions that are in contact will display a correlated mutational record because of the global evolutionary constraint that the native structure imposes for foldability. Of course, additional constraints beyond folding affect sequence evolution, including maintenance of molecular assemblies, enzymatic activity, and allosteric motions. Signals of these functionally relevant contacts are necessarily mixed with those providing robust folding. To identify such relevant interactions involved in folding

and function, a number of methodologies have been developed in recent years that have been successful in uncovering such molecular couplings from sequence data. One of them is direct coupling analysis (DCA)^{7,8}, which is designed to infer a global statistical model from a multiple sequence alignment (MSA) of a single protein family. Using a maximum entropy approach, DCA infers the parameters of an effective energy function consisting of single-site fields and pairwise couplings that is able to approximately reproduce the empirically observed single-site and pairwise amino acid frequencies from the input sequence alignment. The DCA energy function is known as a Potts model, a generalized Ising model that includes non-nearest neighbor interactions and non-constant spin-spin interactions. In practice, couplings of varying strength are computed between all possible pairs of sequence positions. In the past, accurate and tractable approximations of such global models were elusive and detection of direct correlations, as opposed to an aggregate of direct and indirect correlations, was challenging. Other methods are derived from similar theoretical perspectives but have varying computational demands and accuracies^{9–12}. Using an inferred effective energy function, one can estimate pairwise direct probabilities at a particular pair of residue sites. Calculating the Kullback–Leibler divergence between these joint probabilities and single marginal frequencies gives the direct information (DI) score for that residue pair. DI is a proxy of how “directly correlated” two sites are in an MSA. When compared with crystal structures, high DI scores correlate highly with native contacts, and more than 80% overlap, on average, for the top residue pairs in many protein families^{7,13}. The full set of highly scoring contacts amounts to a superset of minimally frustrated and functionally important residue pairs that are spatially localized in the functional configurations of the members of a protein family. Here, we will review the current progress in using residue co-evolution for modeling the structure and dynamics of proteins with a focus on its combination with SBMs.

Residue co-evolutionary constraints are natural input for minimally frustrated protein models

In their simplest form, SBMs idealize minimally frustrated protein energy landscapes by including only native interactions. This model removes any residual non-native energetic roughness and clarifies analysis of the geometrical and topological aspects of protein dynamics and folding. These models faithfully represent the local geometry through bond, angle, dihedral, and excluded volume terms at either single-bead-per-residue or all-atom resolutions. Non-local interactions consist of stabilizing pairwise potentials applied between residue (or atom) pairs that are nearby in the native structure. These pairwise interactions are called native contacts, and the entire set is known as a native contact map. All of the interactions, local and non-local, are set to have an explicit minimum at the native structure, hence the name “structure-based”. The simplified construction of the potential energy function permits for reduced computational requirements, and the explicitly encoded native interactions provide a baseline model that can be used for molecular modeling or studying physical perturbations. For a detailed discussion of the theoretical foundation and construction of SBMs, we refer you to the following reviews,^{14–16} and the references therein.

The quality of contact maps derived from DCA and similar methods have been benchmarked against contact maps calculated from

crystallographic structures, and their accuracy is promising. In general, the larger and more diverse the family of sequences, the higher the quality of contact prediction. The high level of DCA accuracy provided sufficient tertiary constraints to allow folding single domain proteins to within 3 Å from the crystal structure when given knowledge of the secondary structure¹⁷⁻²¹. A rule of thumb is that the number of sequences should be larger than 1000 with less than 80% identity; however, others propose an even lower requirement of a minimum number of sequences close to the length L of the protein polypeptide chain, provided that they are diverse¹⁸. The notoriously difficult problem of predicting membrane protein structures has also been aided by considering evolutionarily coupled pairs^{22,23}.

A native contact map derived from a single native structure is often not sufficient to encode all the functionally relevant, minimally frustrated interactions. This led to the development of a variety of “multi-basin” models, where multiple experimental structures or structural constraints are included in a single SBM²⁴⁻²⁶. As described above, residue pairs with the highest DI scores, the high DI pairs (HDPs), are consistent with the native contact maps. Thus, in an analogous fashion, predicted contact constraints from co-evolution can be merged with contact maps computed from experimental structures in order to more fully capture the true underlying biomolecular energy landscape, including functional transitions and conformations, and therefore to be consistent with multiple structures²⁷⁻²⁹.

Recent advances

Interactions between proteins are fundamental to cellular processes. Where these interactions involve direct contact, multimeric structures, both long-lived and transient, leave correlated mutational patterns between interacting surface residues. A pioneering study used the HDPs between a histidine kinase and its response regulator to make a prediction of the transient protein complex enabling phosphotransfer³⁰. This allowed a prediction for the histidine kinase TM0853 and its response regulator TM0468 that was later confirmed experimentally to be within 3.3 Å³¹. These predictions

are made by minimizing a contact-based energy function consisting of dimeric HDPs. Where dimerization only weakly perturbs the monomer structure, refined rigid-body modeling in combination with co-evolutionary constraints can be employed to estimate protein complexes. When combined with experimental observations, directly coupled amino acids can unveil protein interfaces relevant for the study of disease³². Larger monomer distortions can be readily sampled with SBMs coupled with simulated annealing³³. Current protocols involving HDPs have allowed the large-scale prediction of both homodimers³⁴ and heterodimers^{35,36}. The HDP contact map for a protein family that forms homodimers is a prime example of how ambiguity can arise in co-evolutionary information. The co-evolving dimeric interfacial contacts are mixed with HDPs selected for monomeric folding, but the dimeric contacts can in general be sorted from the monomeric contacts when there is a known monomer structure³⁴. But rarely are there true dichotomies in biology; the existence of domain swapping^{37,38} and structural symmetry^{26,39} highlights some difficulties in assigning particular roles to each HDP. Also, some protein-protein interactions are mediated by disordered regions that order upon binding. The utility of DCA in these cases remains to be tested.

In addition to homo-multimerization, the set of conformations encoded in HDP contact maps can include functional motions. Multi-domain proteins can undergo conformational changes, for example, to accommodate ligands⁴⁰ or in response to phosphorylation⁴¹. In periplasmic ligand binding proteins, there exists an open, ligand-free configuration and a closed, ligand-bound configuration. Molecular dynamics simulations can be performed by using an SBM specific to an open configuration but overlaid with an additional potential term consisting of a set of attractive, short-range interactions for each HDP²⁷. **Figure 1** illustrates an example of this for the leucine-binding protein. The native contact maps for two crystal structures of leucine-binding protein are shown in **Figure 1A**: “open” without ligand and “closed” with ligand. The closed contact map has additional contacts not present in the open structure. The

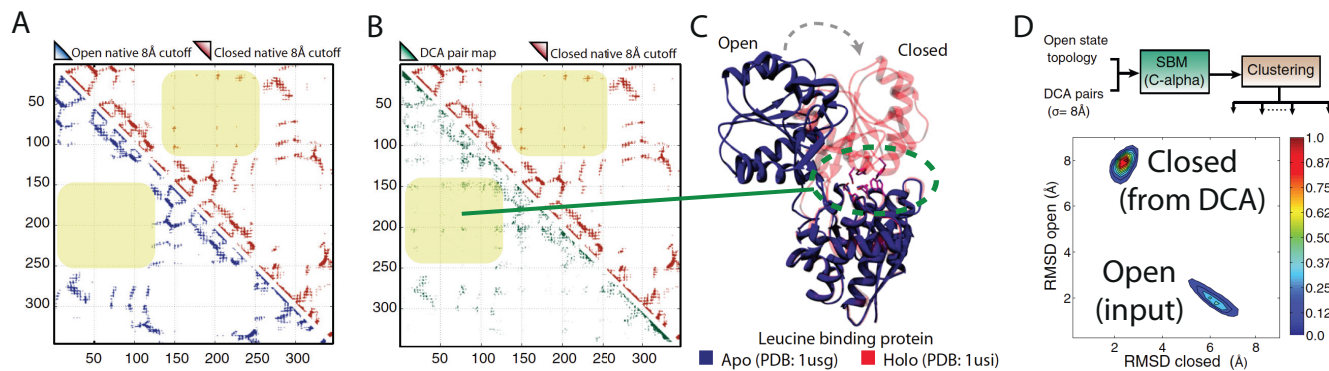


Figure 1. Direct coupling analysis (DCA) contact maps derived from protein family sequence co-evolution are consistent not only with single native structures but also with multiple functional configurations. (A) Leucine-binding protein (LBP) contact maps derived from crystal structures: “open” without ligand and “closed” bound to a ligand. Each triangular region in the map shows a mark if residue pairs are less than 8 Å apart in the experimental structure. The closed contact map (upper triangle) has additional contacts not present in the open structure. (B) The DCA contact map inferred from residue co-evolution (lower triangle) contains a superset of contacts from both open and closed conformations. (C) A cartoon representation of the aligned open (apo) and closed (holo) LBP structures shows a large conformational change upon ligand binding. (D) A structure-based model (SBM) is defined from the apo structure plus contact potentials stabilizing DCA contacts that are not already in the open structure. A two-dimensional root mean square deviation (RMSD) distribution of the states explored by molecular dynamics simulations of this hybrid Hamiltonian shows two peaks within 2 Å of the open and closed states. This shows the ability to uncover functional states via co-evolutionary couplings.

DCA contact map, shown as the lower triangular map in [Figure 1B](#), contains a superset of both the open and closed configuration contacts. An SBM is constructed that is specific to the open structure ([Figure 1D](#)) and additionally contains contact potentials stabilizing all the “non-native” DCA contacts (i.e., any DCA contacts that are not already in the open structure). These additional contacts are each given a stabilizing potential with a minimum at 8 Å. Molecular dynamics simulations of this hybrid SBM+DCA Hamiltonian show two clusters, each within 2 Å of either the open or closed state. Overlaying the DCA contacts does not disrupt the stability of the open structure, and additionally reveals the closed state without including any information from the closed crystal structure. This shows that co-evolutionary information can be used to uncover intermediary, hidden, and functionally relevant conformational states present in many protein families²⁷.

So far, we have discussed how HDP contact maps can be used for structural modeling. However, the fundamental output of the DCA algorithm is not direct information about co-evolving pairs but rather a Potts model Hamiltonian describing the effective energies of interaction for all pairs of residues in a protein family. This Hamiltonian, though not transferable to any sequences outside the family, should, in principle, be able to provide a quantitative window into the stabilities provided by each amino acid in a protein. Strong evidence of the utility of the effective energies comes from their ability to predict the stability changes of single-site mutants^{42,43} and significant correlations to folding rates⁴⁴. Including the so-called single-site fields in addition to the pairwise energies provides even better predictive power⁴⁵. These results suggested that the pairwise energies calculated from co-evolution could be used to inform thermodynamic models of protein folding. Indeed, folding simulations using SBMs with DCA-weighted native contact potentials can better capture transition state ensembles⁴⁶. DCA energies have also been shown to correlate with physical potentials when summed over the entire sequence⁴⁷. Confidence in the ability to estimate energies at both the single-mutant and full-sequence levels is allowing novel methods for investigating the effective energy landscape of evolution, and bridging the gap between biophysics and sequence evolution^{47,48}. These developments are important for integrating the energetics of protein folding and function with protein evolution and selection, which will be crucial to understanding drug resistance and cancer development going forward.

Future directions

The marriage between co-evolutionary information and physical models of biomolecules has been shown to be a fertile research field, where the most important results are yet to come. This field has been focused on rigorously validating the connection and usefulness between evolutionary information with structural modeling and experimental information. However, the true utility of co-evolutionary information is that it allows us to go places that are hard to access by current experimental technologies; important examples are those of membrane protein structure^{22,23} and dynamics, systems with transient conformational states, as well as investigation of large molecular assemblies that resist crystallographic characterization. Although crystal structures exist for FtsH AAA peptidase and the 30S ribosome, recent studies on these two systems^{28,49} show the promise of co-evolutionary information for discovering structural constraints in molecular assemblies. The ability to detect relevant evolutionary interactions has repercussions to our understanding of biomolecular assembly and function. Hopefully, these new tools can be used to alter protein conformation and rewire their interfaces. This has potential applications in the field of protein engineering, as well as systems biology. There is no conceptual hurdle to resisting the application of these ideas to RNA structure and function as well as protein-RNA interactions. Ultimately, we would hope to use all this knowledge to tackle biomedical problems that would help advance human health.

Abbreviations

DCA, direct coupling analysis; DI, direct information; HDP, high direct information pair; MSA, multiple sequence alignment; SBM, structure-based model.

Competing interests

The authors declare that they have no competing interests.

Grant information

This work was supported by the Center for Theoretical Biological Physics sponsored by the National Science Foundation (grants PHY-1427654 and NSF-MCB-1214457). Jeffrey K. Noel is supported in part by the Welch Foundation (grant C-1792).

I confirm that the funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

- Bryngelson JD, Wolynes PG: **Spin glasses and the statistical mechanics of protein folding.** *Proc Natl Acad Sci U S A.* 1987; **84**(21): 7524–8. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Leopold PE, Montal M, Onuchic JN: **Protein folding funnels: a kinetic approach to the sequence-structure relationship.** *Proc Natl Acad Sci USA.* 1992; **89**(18): 8721–5. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Onuchic JN, Wolynes PG: **Theory of protein folding.** *Curr Opin Struct Biol.* 2004; **14**(1): 70–5. [PubMed Abstract](#) | [Publisher Full Text](#)
- Socchi ND, Onuchic JN, Wolynes PG: **Diffusive dynamics of the reaction coordinate for protein folding funnels.** *J Chem Phys.* 1996; **104**(15): 5860–8. [PubMed Abstract](#) | [Publisher Full Text](#)
- Clementi C, Nymeyer H, Onuchic JN: **Topological and energetic factors: what determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? An investigation for small globular proteins.** *J Mol Biol.* 2000; **298**(5): 937–53. [PubMed Abstract](#) | [Publisher Full Text](#)
- Noel JK, Whitford PC, Sanbonmatsu KY, *et al.*: **SMOG@ctbp: simplified**



- deployment of structure-based models in GROMACS. *Nucleic Acids Res.* 2010; **38**(Web Server issue): W657–61.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
7. **F** Morcos F, Pagnani A, Lunt B, *et al.*: **Direct-coupling analysis of residue coevolution captures native contacts across many protein families.** *Proc Natl Acad Sci U S A.* 2011; **108**(49): E1293–301.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
 8. **F** Weigt M, White RA, Szurmant H, *et al.*: **Identification of direct residue contacts in protein-protein interaction by message passing.** *Proc Natl Acad Sci U S A.* 2009; **106**(1): 67–72.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
 9. **F** Taylor WR, Sadowski MI: **Structural constraints on the covariance matrix derived from multiple aligned protein sequences.** *PLoS One.* 2011; **6**(12): e28265.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
 10. **F** Jones DT, Buchan DW, Cozzetto D, *et al.*: **PSICOV: precise structural contact prediction using sparse inverse covariance estimation on large multiple sequence alignments.** *Bioinformatics.* 2012; **28**(2): 184–90.
[PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
 11. **F** Kamisetty H, Ovchinnikov S, Baker D: **Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era.** *Proc Natl Acad Sci U S A.* 2013; **110**(39): 15674–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
 12. **F** Ekeberg M, Lövkvist C, Lan Y, *et al.*: **Improved contact prediction in proteins: using pseudolikelihoods to infer Potts models.** *Phys Rev E Stat Nonlin Soft Matter Phys.* 2013; **87**(1): 012707.
[PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
 13. de Juan D, Pazos F, Valencia A: **Emerging methods in protein co-evolution.** *Nat Rev Genet.* 2013; **14**(4): 249–61.
[PubMed Abstract](#) | [Publisher Full Text](#)
 14. Whitford PC, Sanbonmatsu KY, Onuchic JN: **Biomolecular dynamics: order-disorder transitions and energy landscapes.** *Rep Prog Phys.* 2012; **75**(7): 076601.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 15. Noel JK, Onuchic JN: **The Many Faces of Structure-Based Potentials: From Protein Folding Landscapes to Structural Characterization of Complex Biomolecules.** In: Dokholyan NV, editor. *Computational Modeling of Biological Systems.* Springer US; 2012: 31–54.
[Publisher Full Text](#)
 16. Hills RD Jr, Brooks CL 3rd: **Insights from coarse-grained G_o models for protein folding and dynamics.** *Int J Mol Sci.* 2009; **10**(3): 889–905.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 17. Sulkowska JI, Morcos F, Weigt M, *et al.*: **Genomics-aided structure prediction.** *Proc Natl Acad Sci U S A.* 2012; **109**(26): 10340–5.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 18. **F** Marks DS, Colwell LJ, Sheridan R, *et al.*: **Protein 3D structure computed from evolutionary sequence variation.** *PLoS One.* 2011; **6**(12): e28766.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
 19. Marks DS, Hopf TA, Sander C: **Protein structure prediction from sequence variation.** *Nat Biotechnol.* 2012; **30**(11): 1072–80.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 20. **F** Taylor WR, Jones DT, Sadowski MI: **Protein topology from predicted residue contacts.** *Protein Sci.* 2012; **21**(2): 299–305.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
 21. **F** Nugent T, Jones DT: **Accurate de novo structure prediction of large transmembrane protein domains using fragment-assembly and correlated mutation analysis.** *Proc Natl Acad Sci U S A.* 2012; **109**(24): E1540–7.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
 22. **F** Hopf TA, Colwell LJ, Sheridan R, *et al.*: **Three-dimensional structures of membrane proteins from genomic sequencing.** *Cell.* 2012; **149**(7): 1607–21.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
 23. **F** Wang Y, Barth P: **Evolutionary-guided de novo structure prediction of self-associated transmembrane helical proteins with near-atomic accuracy.** *Nat Commun.* 2015; **6**: 7196.
[PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
 24. **F** Okazaki K, Koga N, Takada S, *et al.*: **Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations.** *Proc Natl Acad Sci U S A.* 2006; **103**(32): 11844–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
 25. Whitford PC, Miyashita O, Levy Y, *et al.*: **Conformational transitions of adenylate kinase: switching by cracking.** *J Mol Biol.* 2007; **366**(5): 1661–71.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 26. Noel JK, Schug A, Verma A, *et al.*: **Mirror images as naturally competing conformations in protein folding.** *J Phys Chem B.* 2012; **116**(23): 6880–8.
[PubMed Abstract](#) | [Publisher Full Text](#)
 27. **F** Morcos F, Jana B, Hwa T, *et al.*: **Coevolutionary signals across protein lineages help capture multiple protein conformations.** *Proc Natl Acad Sci U S A.* 2013; **110**(51): 20533–8.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
 28. Jana B, Morcos F, Onuchic JN: **From structure to function: the convergence of structure based models and co-evolutionary information.** *Phys Chem Chem Phys.* 2014; **16**(14): 6496–507.
[PubMed Abstract](#) | [Publisher Full Text](#)
 29. **F** Dago AE, Schug A, Procaccini A, *et al.*: **Structural basis of histidine kinase autophosphorylation deduced by integrating genomics, molecular dynamics, and mutagenesis.** *Proc Natl Acad Sci U S A.* 2012; **109**(26): E1733–42.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
 30. Schug A, Weigt M, Onuchic JN, *et al.*: **High-resolution protein complexes from integrating genomic information with molecular simulation.** *Proc Natl Acad Sci U S A.* 2009; **106**(52): 22124–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 31. **F** Casino P, Rubio V, Marina A: **Structural insight into partner specificity and phosphoryl transfer in two-component signal transduction.** *Cell.* 2009; **139**(2): 325–36.
[PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
 32. Tamir S, Rotem-Bamberger S, Katz C, *et al.*: **Integrated strategy reveals the protein interface between cancer targets Bcl-2 and NAF-1.** *Proc Natl Acad Sci U S A.* 2014; **111**(14): 5177–82.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 33. **F** Zheng W, Schafer NP, Davtyan A, *et al.*: **Predictive energy landscapes for protein-protein association.** *Proc Natl Acad Sci U S A.* 2012; **109**(47): 19244–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
 34. dos Santos RN, Morcos F, Jana B, *et al.*: **Dimeric interactions and complex formation using direct coevolutionary couplings.** *Sci Rep.* 2015; **5**: 13652.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 35. **F** Ovchinnikov S, Kamisetty H, Baker D: **Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information.** *eLife.* 2014; **3**: e02030.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
 36. **F** Hopf TA, Schärfe CP, Rodrigues JP, *et al.*: **Sequence co-evolution gives 3D contacts and structures of protein complexes.** *eLife.* 2014; **3**: e03430.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
 37. **F** Liu Y, Eisenberg D: **3D domain swapping: as domains continue to swap.** *Protein Sci.* 2002; **11**(6): 1285–99.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
 38. Yang S, Cho SS, Levy Y, *et al.*: **Domain swapping is a consequence of minimal frustration.** *Proc Natl Acad Sci U S A.* 2004; **101**(38): 13786–91.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 39. **F** Brown JH: **Breaking symmetry in protein dimers: designs and functions.** *Protein Sci.* 2006; **15**(1): 1–13.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
 40. Felder CB, Graul RC, Lee AY, *et al.*: **The Venus flytrap of periplasmic binding proteins: an ancient protein module present in multiple drug receptors.** *AAPS PharmSci.* 1999; **1**(2): E2.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 41. Lätzer J, Shen T, Wolynes PG: **Conformational switching upon phosphorylation: a predictive framework based on energy landscape principles.** *Biochemistry.* 2008; **47**(7): 2110–22.
[PubMed Abstract](#) | [Publisher Full Text](#)
 42. **F** Lui S, Tiana G: **The network of stabilizing contacts in proteins studied by coevolutionary data.** *J Chem Phys.* 2013; **139**(15): 155103.
[PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
 43. Cheng RR, Morcos F, Levine H, *et al.*: **Toward rationally redesigning bacterial two-component signaling systems using coevolutionary information.** *Proc Natl Acad Sci U S A.* 2014; **111**(5): E563–71.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 44. **F** Mallik S, Kundu S: **Co-evolutionary constraints of globular proteins correlate with their folding rates.** *FEBS Lett.* 2015; **589**(17): 2179–85.
[PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
 45. **F** Contini A, Tiana G: **A many-body term improves the accuracy of effective potentials based on protein coevolutionary data.** *J Chem Phys.* 2015; **143**(2): 25103.
[PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
 46. Cheng RR, Raghunathan M, Noel JK, *et al.*: **Constructing sequence-dependent protein models using coevolutionary information.** *Protein Sci.* 2016; **25**(1): 111–22.
[PubMed Abstract](#) | [Publisher Full Text](#)
 47. Morcos F, Schafer NP, Cheng RR, *et al.*: **Coevolutionary information, protein folding landscapes, and the thermodynamics of natural selection.** *Proc Natl Acad Sci U S A.* 2014; **111**(34): 12408–13.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 48. **F** Sikosek T, Chan HS: **Biophysics of protein evolution and evolutionary protein biophysics.** *J R Soc Interface.* 2014; **11**(100): 20140419.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
 49. Mallik S, Akashi H, Kundu S: **Assembly constraints drive co-evolution among ribosomal constituents.** *Nucleic Acids Res.* 2015; **43**(11): 5352–63.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Referee Status:



Editorial Note on the Review Process

F1000 Faculty Reviews are commissioned from members of the prestigious F1000 Faculty and are edited as a service to readers. In order to make these reviews as comprehensive and accessible as possible, the referees provide input before publication and only the final, revised version is published. The referees who approved the final version are listed with their names and affiliations but without their reports on earlier versions (any comments will already have been addressed in the published version).

The referees who approved this article are:

Version 1

- 1 **Sudip Kundu**, Department of Biophysics, Molecular Biology and Bioinformatics, University of Calcutta, Calcutta, India
Competing Interests: No competing interests were disclosed.
- 2 **Angel Garcia**, Department of Physics, Rensselaer Polytechnic Institute, Troy, NY, USA
Competing Interests: No competing interests were disclosed.
- 3 **Shoji Takata**, Department of Biophysics, Kyoto University, Kyoto, Japan
Competing Interests: No competing interests were disclosed.