F1000Research

Check for updates

REVIEW

Co-evolution techniques are reshaping the way we do structural bioinformatics [version 1; referees: 2 approved]

Saulo de Oliveira, Charlotte Deane 📄 iD

Department of Statistics, University of Oxford, Oxford, UK

Abstract

Co-evolution techniques were originally conceived to assist in protein structure prediction by inferring pairs of residues that share spatial proximity. However, the functional relationships that can be extrapolated from co-evolution have also proven to be useful in a wide array of structural bioinformatics applications. These techniques are a powerful way to extract structural and functional information in a sequence-rich world.

Open Peer Review

Referee Status: ✔ ✔

	Invited Referees	
	1	**2**
version 1 published 25 Jul 2017	✔	✔

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 **David Baker**, Howard Hughes Medical Institute, University of Washington, USA

2 **Johannes Söding**, Max Planck Institute for Biophysical Chemistry, Germany

Discuss this article

Comments (0)

Introduction

A large number of structural bioinformatics applications rely on extracting structural features from a protein's sequence. This is traditionally done by performing multiple sequence alignments (MSAs) of homologues. MSAs have been used as input to predict features such as secondary structure, torsion and bond angles, solvent accessibility, disorder regions, and domain boundaries. The main limitation of most of these descriptors such as predicted secondary structure is that, although often highly accurate, they provide information only about a protein's local conformation. For instance, they may tell us how a set of residues comprise an alpha-helix, but they do not provide any information as to how different alpha-helices are oriented with respect to one another. Techniques based on co-evolution go a step further by extracting non-local structural information from MSAs. These techniques are based on the notion that two residues which mutate in a correlated fashion, so that a mutation in one is often compensated by a mutation in the other, can be considered to be co-evolving. Co-evolution is interpreted as functional dependence, i.e. if two residues are co-evolving, there is a cost in fitness for mutating only one of these residues. Although these techniques were originally conceived and applied to protein structure prediction, they are now established tools with a diverse set of applications in structural bioinformatics.

Initial attempts at identifying co-evolving residues were implemented by calculating the correlation between columns in an MSA[1–6]. To quantify the precision of different methods, protein contacts (residues that share spatial proximity; usually C-βs less than 8 Å apart) were considered as true positives. These early attempts presented low precision and therefore limited usefulness. Methods based on calculating the Mutual Information (MI) between MSA columns were able to extend the applicability of these approaches[7–9], but predictions were still not precise enough to be useful for most cases[10].

The exponential growth in the number of protein sequences combined with the application of existing statistical techniques that solved the inverse statistical problem to infer evolutionary couplings have allowed the development of methods with a precision range that has proved useful for many applications[11–14]. Direct Coupling Analysis (DCA) techniques are based on a generalised Ising model and, unlike MI and previous approaches, addressed the problem of transitivity by considering the correlation amongst all columns in the MSA as background to establish if two residues are co-evolving. Subsequent implementations based on similar ideas attempted to relax some of the assumptions of the original model and yielded progressively better results[15–18]. Although close in conception, these methods managed to produce a significant number of non-overlapping predicted correlations[19]. Meta-predictors were then developed to combine the non-overlapping set of predictions to produce a consensus[19–21], further improving the precision of co-evolution inference. A large-scale comparative study (~3,500 cases) has shown that the most precise of these methods, metaPSICOV[22], achieved a precision greater than 50% for its top L predictions, where L is the protein length, for over 68% of test cases[23]. Other methods were developed with specific applicability, such as inferring co-evolving residues in membrane proteins[24–26] or between β-sheets[27]. The precision of predicted correlated mutations has continued to increase with improved methods using physico-chemical information[28] and ultra-deep learning[29].

Co-evolution and protein structure prediction

The implementation of DCA led to consistent and accurate *de novo* structure prediction for both soluble[14,15,30] and transmembrane proteins[31,32] when sufficient sequence information is available. Recent results from the critical assessment of methods of protein structure prediction[33] have shown that in the presence of a sufficiently accurate number of predictions, topology prediction can be performed consistently and accurately. However, a few challenges remain regarding the identification and assignment of domain boundaries, longer proteins, and, most importantly, for cases where the number of available sequences is insufficient for accurate co-evolution inference. This latter problem is the main limitation; without enough diverse sequence information, accurate evolutionary coupling inference is currently impossible. When considering the results of the Critical Assessment of methods for protein Structure Prediction (CASP), a blind community-wide experiment that evaluates different prediction methods, protein structure prediction has been applied to a large number of cases where the target structure was unknown, providing reliable large-scale information about unknown folds[34].

Understanding protein–protein interactions in light of co-evolution

Co-evolution analysis of paired sequences from interacting proteins has been shown to be effective in identifying pairs of residues involved in complex formation[35]. A subsequent study has shown that when the number of paired sequences exceeds the average length of the proteins in the complex, most of the co-evolving residues are in contact at the protein–protein interface[36].

Co-evolution has also been shown to assist in protein–protein docking during the rounds 28–35 of the critical assessment of prediction of interactions. A potential based on co-evolution inference called InterEVScore was used in conjunction with ZDock, SOAP-PP, and Rosetta refinement to produce correct predictions for 10 out of 18 targets[37,38]. Co-evolution has also been used to identify protein–protein interactions and was shown to predict the only two experimentally known interactions of the trp operon[39]. The main limitation that co-evolution techniques encounter when used to infer protein–protein interactions is that these methods require a large number of pairs of protein sequences of the same organism, which currently restricts its applicability to a small number of cases. Furthermore, pairing of same-organism sequences is particularly difficult in the presence of paralogues, and methods have been proposed to address this problem[40,41].

Co-evolving residues may be suggestive of multimerisation

The concept of using co-evolution to predict protein-protein interactions can be further extended to include protein multimerisation. This offsets the limitation of dependence on paired sequences. However, multimerisation prediction is more challenging than the identification of co-evolving residues in a protein–protein interaction interface, since it is necessary to discriminate between multimeric contacts and intra-monomer contacts.

Co-evolution techniques were used to correctly identify multimeric contacts for 18 dimeric complexes[42] and to validate a suggested dimeric interface between two Hsp70 molecules in the DnaK crystal[43]. This success in multimeric prediction suggests that the existing quality assessment of co-evolution techniques may be underestimating their precision. This is because of the fact that pairs of residues interacting in the multimeric conformation would not necessarily share spatial proximity when considering the monomeric protein chain and thus would be incorrectly considered as false positives.

Predicting domain boundaries by means of correlated mutations

Domain boundary identification is particularly useful for, but not restricted to, protein structure prediction, and it has been reported as one of the main challenges encountered in the free-modelling category of CASP[44]. Protein contacts have been used for automatic domain boundary assignment and prediction by means of minimising the inter-domain contacts whilst maximising the number of intra-domain contacts[45,46]. However, these contact-based methods depend on an existing structure for the target sequence and are therefore not applicable when predicting new structures. This limitation, however, can potentially be overcome by using co-evolution inference to predict protein contacts. Correlated mutations output by MI led to the successful prediction of domain boundaries[47]. A more precise co-evolution inference method has also been used for domain prediction. It was shown to produce better results for 368 targets compared to sequence-based methods and comparable results to homology-based methods[48].

Identifying alternative conformations, allostery, and flexibility by means of co-evolution

Co-evolution provides a way of assessing the biological relevance of different conformations observed in coarse-grained structural-based models or molecular dynamics simulations. Co-evolving residues have been used to guide coarse-grained simulations either towards the native conformation or to explore conformational ensembles that are supported by evolution[49]. They have also been used to identify distinct functional conformational states suggested to be observed between apo and holo conformations[50–53]. In another study, co-evolution was used to identify a framework for allostery for the MutS DNA mismatch repair protein[54] by means of Statistical Coupling Analysis (SCA). This approach differs from the traditional DCA, as it aims to construct a network of co-evolving residues as opposed to performing the correlation assessment on a pairwise level.

Identification of alternative conformations and allostery using experimental techniques is challenging, suggesting co-evolution techniques may be a powerful tool for exploring and targeting conformational dynamics. The success of co-evolution approaches suggests that co-evolving residues can be in contact only in a subset of a protein's conformations. Once again, this highlights that the precision of co-evolution methods may be underestimated if they are tested against a single protein structure.

Co-evolution can assist in experimental determination

Structural models produced *ab initio* can be used to assist in crystallographic protein structure determination, particularly when no other structural information is available. In these scenarios, *ab initio* models are used in molecular replacement protocols to solve the phasing problem. However, this is limited by the quality and reliability of the input models. Up until the advent of more precise co-evolution methods, *ab initio* protein structure prediction led to poor modelling results for a large number of cases, including longer and/or multi-domain proteins. Co-evolution has broadened the applicability of models produced in the absence of a template, leading to more consistent and reliable predictions. Models generated *ab initio* in conjunction with co-evolution constraints have been shown to improve the success of molecular replacement[55,56]. Co-evolution has also been used to characterise the order in which macromolecular complexes self-assemble, complementing existing experimental data for those complexes[57].

Expanding the applicability of co-evolution via metagenomics

The precision of co-evolution methods is known to be dependent on the number of non-redundant sequences used in the MSA[13–15,19,30]. Insufficient sequence information constitutes the main limitation for co-evolution techniques. In the absence of a minimal number of non-redundant sequences, the inferred evolutionary couplings are unlikely to suit any of the purposes mentioned thus far. The usefulness of the predictions is therefore restricted to protein families for which a sufficient number of non-redundant sequences is available. It was previously reported that approximately 25% of the protein families on Pfam[58] would have a sufficient number of sequences for reliable co-evolution inference[17]. Metagenomics data have been used as a source for additional sequences, thus expanding the applicability of co-evolution[59]. Ovchinnikov and colleagues used metagenomics to increase the number of MSA sequences and subsequently to predict the protein structure of an additional 614 protein families, 140 with no members with known structure. Metagenomics provides a wealth of sequence information that is yet to be explored in other applications of co-evolution techniques, such as protein–protein interaction prediction and functional characterisation.

Functional characterisation and fitness estimation

A common application in bioinformatics is to predict the effect of a particular mutation on a phenotype. Given that co-evolution aims to capture correlated mutations, it can be used to quantify how likely a mutation is to be compensated for by a second mutation in another residue. This, in turn, provides a means of estimating the fitness cost for a particular mutation considering its effect based on co-evolving residues. A recent method, EVmutation, uses co-evolution to quantify the effects of multiple mutations on the phenotype[60]. Though the method can be generalised for any organism, it was tested for 34 cases to identify deleterious mutations in humans, showing comparable results to state-of-the-art supervised methods.

Maximum entropy models, which serve as a basis for several co-evolution methods, can also provide insights on the fitness landscape of a particular protein family[61–63]. These methods can estimate an energy for a target sequence that can be interpreted as the compatibility of this sequence to the fitness landscape of its family.

Another application of co-evolution-based fitness estimation relates to bioengineering. Co-evolution can be used to identify pairs of residues which, if mutated, can alter a protein's stability and/or function. This is particularly important when selecting hotspots for enzyme engineering[64]. As an example, co-evolving site mutagenesis was used to improve protein thermostability of alpha-amylase[65]. There are also examples where mutations in co-evolving positions were shown to be de-stabilising[64]. Evolutionary couplings can also highlight residue interactions that are not known either because a structure is unavailable or because such relationships are not evident from structural data (e.g. unresolved residues). This provides additional insights into protein folding, stability, and function that can be explored by synthetic biology/bioengineering.

Conclusions

The advent of precise methods for the identification of co-evolving residues has led to progress in many areas of structural bioinformatics. The limited applicability of these methods, usually constrained by the amount of sequence information available, may be offset by metagenomics efforts and the exponential growth in sequence information. This paves the way for co-evolution to become as pivotal to bioinformatics analyses as sequence alignments themselves. The functional relationships that can be derived from these predictions provide a source of additional data that goes beyond the realm of structural prediction, translating an abundant source of information (sequence) into biological signal. Though still in their infancy, many of the alternative applications of co-evolution show great promise, and we can expect to see many advances and new techniques in these areas over the coming years.

References

F1000 recommended

1. Göbel U, Sander C, Schneider R, *et al.*: **Correlated mutations and residue contacts in proteins.** *Proteins.* 1994; **18**(4): 309–17.
   PubMed Abstract | **Publisher Full Text**

2. Fariselli P, Olmea O, Valencia A, *et al.*: **Prediction of contact maps with neural networks and correlated mutations.** *Protein Eng.* 2001; **14**(11): 835–43.
   PubMed Abstract | **Publisher Full Text**

3. Fariselli P, Olmea O, Valencia A, *et al.*: **Progress in predicting inter-residue contacts of proteins with neural networks and correlated mutations.** *Proteins.* 2001; **45**(Suppl 5): 157–62.
   PubMed Abstract | **Publisher Full Text**

4. Olmea O, Valencia A: **Improving contact predictions by the combination of correlated mutations and other sources of sequence information.** *Fold Des.* 1997; **2**(3): S25–32.
   PubMed Abstract | **Publisher Full Text**

5. Pazos F, Olmea O, Valencia A: **A graphical interface for correlated mutations and other protein structure prediction methods.** *Comput Appl Biosci.* 1997; **13**(3): 319–21.
   PubMed Abstract

6. Shindyalov IN, Kolchanov NA, Sander C: **Can three-dimensional contacts in protein structures be predicted by analysis of correlated mutations?** *Protein Eng.* 1994; **7**(3): 349–58.
   PubMed Abstract | **Publisher Full Text**

7. Cline MS, Karplus K, Lathrop RH, *et al.*: **Information-theoretic dissection of pairwise contact potentials.** *Proteins.* 2002; **49**(1): 7–14.
   PubMed Abstract | **Publisher Full Text** | **F1000 Recommendation**

8. Liu Y, Bahar I: **Sequence evolution correlates with structural dynamics.** *Mol Biol Evol.* 2012; **29**(9): 2253–63.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

9. Shackelford G, Karplus K: **Contact prediction using mutual information and neural nets.** *Proteins.* 2007; **69 Suppl 8**: 159–64.
   PubMed Abstract | **Publisher Full Text**

10. Horner DS, Pirovano W, Pesole G: **Correlated substitution analysis and the prediction of amino acid structural contacts.** *Brief Bioinform.* 2008; **9**(1): 46–56.
    PubMed Abstract | **Publisher Full Text**

11. Lapedes AS, Giraud B, Liu L, *et al.*: **Correlated mutations in models of protein sequences: phylogenetic and structural effects.** In: Seillier-Moiseiwitsch F, editor. *Statistics in molecular biology and genetics.* Hayward, CA: Institute of Mathematical Statistics; 1999; **33**: 236–256.
    Publisher Full Text

12. Balakrishnan S, Kamisetty H, Carbonell JG, *et al.*: **Learning generative models for protein fold families.** *Proteins.* 2011; **79**(4): 1061–78.
    PubMed Abstract | **Publisher Full Text**

13. 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**

14. 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**

15. 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**

16. 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): 12707.
    PubMed Abstract | **Publisher Full Text** | **F1000 Recommendation**

17. 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**

18. Seemayer S, Gruber M, Söding J: **CCMpred--fast and precise prediction of protein residue-residue contacts from correlated mutations.** *Bioinformatics.* 2014; **30**(21): 3128–30.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

19. Kaján L, Hopf TA, Kalaš M, *et al.*: **FreeContact: fast and free software for protein contact prediction from residue co-evolution.** *BMC Bioinformatics.* 2014; **15**: 85.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

20. Skwark MJ, Abdel-Rehim A, Elofsson A: **PconsC: combination of direct information methods and alignments improves contact prediction.** *Bioinformatics.* 2013; **29**(14): 1815–6.
    PubMed Abstract | **Publisher Full Text**

21. Skwark MJ, Raimondi D, Michel M, *et al.*: **Improved contact predictions using the recognition of protein like contact patterns.** *PLoS Comput Biol.* 2014; **10**(11): e1003889.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

22. [F] Jones DT, Singh T, Kosciolek T, *et al.*: **MetaPSICOV: combining coevolution methods for accurate prediction of contacts and long range hydrogen bonding in proteins.** *Bioinformatics.* 2015; **31**(7): 999–1006.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

23. de Oliveira SH, Shi J, Deane CM: **Comparing co-evolution methods and their application to template-free protein structure prediction.** *Bioinformatics.* 2017; **33**(3): 373–81.
PubMed Abstract | Publisher Full Text

24. Yang J, Jang R, Zhang Y, *et al.*: **High-accuracy prediction of transmembrane inter-helix contacts and application to GPCR 3D structure modeling.** *Bioinformatics.* 2013; **29**(20): 2579–87.
PubMed Abstract | Publisher Full Text | Free Full Text

25. [F] Zhang H, Huang Q, Bei Z, *et al.*: **COMSAT: Residue contact prediction of transmembrane proteins based on support vector machines and mixed integer linear programming.** *Proteins.* 2016; **84**(3): 332–48.
PubMed Abstract | Publisher Full Text | F1000 Recommendation

26. [F] Zhang L, Wang H, Yan L, *et al.*: **OMPcontact: An Outer Membrane Protein Inter-Barrel Residue Contact Prediction Method.** *J Comput Biol.* 2017; **24**(3): 217–28.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

27. [F] Andreani J, Söding J: **bbcontacts: prediction of β-strand pairing from direct coupling patterns.** *Bioinformatics.* 2015; **31**(11): 1729–37.
PubMed Abstract | Publisher Full Text | F1000 Recommendation

28. Schneider M, Brock O: **Combining physicochemical and evolutionary information for protein contact prediction.** *PLoS One.* 2014; **9**(10): e108438.
PubMed Abstract | Publisher Full Text | Free Full Text

29. [F] Wang S, Sun S, Li Z, *et al.*: **Accurate *De Novo* Prediction of Protein Contact Map by Ultra-Deep Learning Model.** *PLoS Comput Biol.* 2017; **13**(1): e1005324.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

30. 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

31. [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

32. [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

33. [F] Moult J, Fidelis K, Kryshtafovych A, *et al.*: **Critical assessment of methods of protein structure prediction: Progress and new directions in round XI.** *Proteins.* 2016; **84 Suppl 1**: 4–14.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

34. [F] Ovchinnikov S, Kinch L, Park H, *et al.*: **Large-scale determination of previously unsolved protein structures using evolutionary information.** *eLife.* 2015; **4**: e09248.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

35. [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

36. [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

37. [F] Yu J, Andreani J, Ochsenbein F, *et al.*: **Lessons from (co-)evolution in the docking of proteins and peptides for CAPRI Rounds 28–35.** *Proteins.* 2017; **85**(3): 378–90.
PubMed Abstract | Publisher Full Text | F1000 Recommendation

38. Andreani J, Faure G, Guerois R: **InterEvScore: a novel coarse-grained interface scoring function using a multi-body statistical potential coupled to evolution.** *Bioinformatics.* 2013; **29**(14): 1742–9.
PubMed Abstract | Publisher Full Text

39. [F] Feinauer C, Szurmant H, Weigt M, *et al.*: **Inter-Protein Sequence Co-Evolution Predicts Known Physical Interactions in Bacterial Ribosomes and the Trp Operon.** *PLoS One.* 2016; **11**(2): e0149166.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

40. [F] Gueudré T, Baldassi C, Zamparo M, *et al.*: **Simultaneous identification of specifically interacting paralogs and interprotein contacts by direct coupling analysis.** *Proc Natl Acad Sci U S A.* 2016; **113**(43): 12186–91.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

41. [F] Bitbol A, Dwyer RS, Colwell LJ, *et al.*: **Inferring interaction partners from protein sequences.** *Proc Natl Acad Sci U S A.* 2016; **113**(43): 12180–5.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

42. [F] 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 | F1000 Recommendation

43. [F] Malinverni D, Marsili S, Barducci A, *et al.*: **Large-Scale Conformational Transitions and Dimerization Are Encoded in the Amino-Acid Sequences of Hsp70 Chaperones.** *PLoS Comput Biol.* 2015; **11**(6): e1004262.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

44. [F] Ovchinnikov S, Kim DE, Wang RY, *et al.*: **Improved *de novo* structure prediction in CASP11 by incorporating coevolution information into Rosetta.** *Proteins.* 2016; **84 Suppl 1**: 67–75.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

45. Siddiqui AS, Barton GJ: **Continuous and discontinuous domains: an algorithm for the automatic generation of reliable protein domain definitions.** *Protein Sci.* 1995; **4**: 872–84.
PubMed Abstract | Publisher Full Text | Free Full Text

46. Swindells MB: **A procedure for detecting structural domains in proteins.** *Protein Sci.* 1995; **4**(1): 103–12.
PubMed Abstract | Publisher Full Text | Free Full Text

47. Rigden DJ: **Use of covariance analysis for the prediction of structural domain boundaries from multiple protein sequence alignments.** *Protein Eng.* 2002; **15**(2): 65–77.
PubMed Abstract | Publisher Full Text

48. Sadowski MI: **Prediction of protein domain boundaries from inverse covariances.** *Proteins.* 2013; **81**(2): 253–60.
PubMed Abstract | Publisher Full Text | Free Full Text

49. [F] Sutto L, Marsili S, Valencia A, *et al.*: **From residue coevolution to protein conformational ensembles and functional dynamics.** *Proc Natl Acad Sci U S A.* 2015; **112**(44): 13567–72.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

50. [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

51. 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

52. [F] Toth-Petroczy A, Palmedo P, Ingraham J, *et al.*: **Structured States of Disordered Proteins from Genomic Sequences.** *Cell.* 2016; **167**(1): 158–170.e12.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

53. [F] Sfriso P, Duran-Frigola M, Mosca R, *et al.*: **Residues Coevolution Guides the Systematic Identification of Alternative Functional Conformations in Proteins.** *Structure.* 2016; **24**(1): 116–26.
PubMed Abstract | Publisher Full Text | F1000 Recommendation

54. [F] Lakhani B, Thayer KM, Hingorani MM, *et al.*: **Evolutionary Covariance Combined with Molecular Dynamics Predicts a Framework for Allostery in the MutS DNA Mismatch Repair Protein.** *J Phys Chem B.* 2017; **121**(9): 2049–61.
PubMed Abstract | Publisher Full Text | F1000 Recommendation

55. [F] Simkovic F, Thomas JM, Keegan RM, *et al.*: **Residue contacts predicted by evolutionary covariance extend the application of *ab initio* molecular replacement to larger and more challenging protein folds.** *IUCrJ.* 2016; **3**(Pt 4): 259–70.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

56. [F] Simkovic F, Ovchinnikov S, Baker D, *et al.*: **Applications of contact predictions to structural biology.** *IUCrJ.* 2017; **4**(Pt 3): 291–300.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

57. [F] Mallik S, Kundu S: **Coevolutionary constraints in the sequence-space of macromolecular complexes reflect their self-assembly pathways.** *Proteins.* 2017; **85**(7): 1183–9.
PubMed Abstract | Publisher Full Text | F1000 Recommendation

58. Finn RD, Bateman A, Clements J, *et al.*: **Pfam: the protein families database.** *Nucleic Acids Res.* 2014; **42**(Database issue): D222–30.
PubMed Abstract | Publisher Full Text | Free Full Text

59. [F] Ovchinnikov S, Park H, Varghese N, *et al.*: **Protein structure determination using metagenome sequence data.** *Science.* 2017; **355**(6322): 294–8.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

60. [F] Hopf TA, Ingraham JB, Poelwijk FJ, *et al.*: **Mutation effects predicted from sequence co-variation.** *Nat Biotechnol.* 2017; **35**(2): 128–35.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

61. Mann JK, Barton JP, Ferguson AL, *et al.*: **The fitness landscape of HIV-1 gag: advanced modeling approaches and validation of model predictions by *in vitro* testing.** *PLoS Comput Biol.* 2014; **10**(8): e1003776.
PubMed Abstract | Publisher Full Text | Free Full Text

62. [F] Rawi R, Kunji K, Haoudi A, *et al.*: **Coevolution Analysis of HIV-1 Envelope Glycoprotein Complex.** *PLoS One.* 2015; **10**(11): e0143245.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

63. [F] Figliuzzi M, Jacquier H, Schug A, *et al.*: **Coevolutionary Landscape Inference and the Context-Dependence of Mutations in Beta-Lactamase TEM-1.** *Mol Biol Evol.* 2016; **33**(1): 268–80.
PubMed Abstract | Publisher Full Text | Free Full Text | F1000 Recommendation

64. [F] Franceus J, Verhaeghe T, Desmet T: **Correlated positions in protein evolution and engineering.** *J Ind Microbiol Biotechnol.* 2017; **44**(4–5): 687–95.
PubMed Abstract | Publisher Full Text | F1000 Recommendation

65. Wang C, Huang R, He B, *et al.*: **Improving the thermostability of alpha-amylase by combinatorial coevolving-site saturation mutagenesis.** *BMC Bioinformatics.* 2012; **13**: 263.
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 **Johannes Söding** Research Group Quantitative and Computational Biology, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany
Competing Interests: No competing interests were disclosed.

1 **David Baker** Department of Biochemistry, Howard Hughes Medical Institute, University of Washington, Seattle, USA
Competing Interests: No competing interests were disclosed.