Chapter 9

Assessing Predicted Contacts for Building Protein Three-Dimensional Models

Badri Adhikari, Debswapna Bhattacharya, Renzhi Cao, and Jianlin Cheng

Abstract

Recent successes of contact-guided protein structure prediction methods have revived interest in solving the long-standing problem of ab initio protein structure prediction. With homology modeling failing for many protein sequences that do not have templates, contact-guided structure prediction has shown promise, and consequently, contact prediction has gained a lot of interest recently. Although a few dozen contact prediction tools are already currently available as web servers and downloadables, not enough research has been done towards using existing measures like precision and recall to evaluate these contacts with the goal of building three-dimensional models. Moreover, when we do not have a native structure for a set of predicted contacts, the only analysis we can perform is a simple contact map visualization of the predicted contacts. A wider and more rigorous assessment of the predicted contacts is needed, in order to build tertiary structure models. This chapter discusses instructions and protocols for using tools and applying techniques in order to assess predicted contacts for building three-dimensional models.

Key words Protein contact assessment, Contact-guided ab initio prediction

1 Introduction

In the last few years, prediction of protein residue contacts has shown improvement in the field of ab initio protein structure prediction [1–4]. Tertiary structure predictions can benefit from the use of predicted contacts for many reasons. One of the most crucial values of contact-guided protein structure prediction has to do with contact connection information that can give us a better look at the mechanism which causes proteins to fold. For successful ab initio modeling using contacts, the quality of predicted contacts is the most important consideration because for almost all proteins, accurate contact predictions result in correct folds. Since the field of contact prediction is still developing, the question of how the predicted contacts can be appropriately assessed so that we can use them to build three-dimensional models is still subject to discussion, debate and much more research. Given a set or sets of

Yaoqi Zhou et al. (eds.), *Prediction of Protein Secondary Structure*, Methods in Molecular Biology, vol. 1484,
DOI 10.1007/978-1-4939-6406-2_9, © Springer Science+Business Media New York 2017

predicted contacts for a protein sequence, we are exploring novel and potentially transformative techniques to utilize these contacts for building tertiary structure models for proteins. Current techniques include visualization using contact maps, and evaluation using various measures like precision and coverage.

Those researchers exploring the task of building tertiary structure models like Rosetta [5], I-Tasser [6], and RBO-alph [7]—all have started to incorporate contacts to aid their methods. Those focusing on building 3D models primarily using predicted contacts have developed new methods like FRAGFOLD [2], EVFOLD [3], and CONFOLD [5]. For existing structure prediction systems like Rosetta and I-Tasser, a few predicted contacts can be used as additional information to guide the ab initio folding process. On the other hand, it is also important to have a decent number of contacts (for example, those ranging from L/2 to L, where L is the length of the protein) to guide the modeling process to predict protein folds to facilitate tools which build models from scratch, like EVFOLD and CONFOLD. This second group of modeling tools dedicated to building models from scratch is ideal for studying the quality of predicted contacts because they solely rely on contacts to build models, and the results are not biased by other prediction information such as the availability of good fragments.

Whether or not a native structure exists for a set of predicted contacts, a good way to evaluate the predicted contacts is to directly build three-dimensional models using them and observe the 3D models. In this chapter, we will discuss the protocols for using one such method, CONFOLD, available at http://protein.rnet.missouri.edu/confold/. We will also discuss the available tools and techniques for precision and coverage calculations, including improved contact map visualizations. For convenience, we have built a web server, CONASSESS, available at http://cactus.rnet.missouri.edu/conassess/.

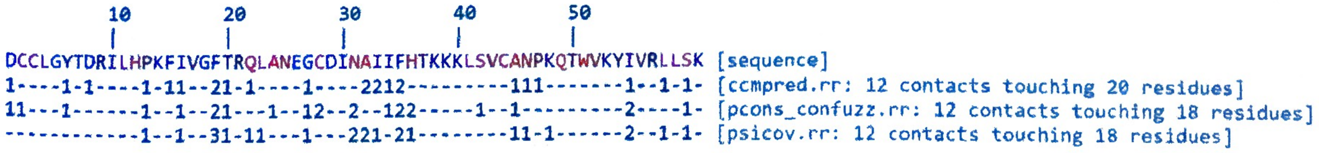
2 Materials

When the true structure of a protein is known, there are widely used tools to evaluate predicted contacts. When no true structure exists, the only analysis we can perform is visualizations to check the proportion of contact types and ensure a good coverage. The three contact types—short, medium, and long-range—are defined using sequence minimum sequence separation of at least 6, 12, and 24 residues, respectively. For instance, contacts with residue sequence separation more than 11 and less than 24 are defined as medium range contacts. Among these three contact types, long-range contacts are the most important for folding purposes and are also the most difficult to predict [8, 9] (*see* **Note 2**). To help study the coverage of predicted contacts we introduce 1D visualization of the contact coordination number, and to check the proportion of contact types, we discuss improved contact map visualization. When the 3D structure of the sequence is known we can simply calculate precision and coverage of a certain number of selected top contacts as a primary evaluation. Besides precision and coverage, other measures like spread [3], mean error, and *Xd* [10–12] are important to obtain a highly accurate three-dimensional fold of a protein. We will discuss these techniques in the following sections. All evaluation methods, including precision, *Xd*, coverage, and both 1D and 2D visualization techniques are implemented in our CONASSESS web server.

2.1 Contact Visualization

Visualizing three-dimensional information in lower dimensions is challenging, but as long as we are interested in a particular aspect of the data, simpler visualizations in lower dimensions can be easy and yet effective. A simple technique for 1D representation of predicted residue contacts is to assign numbers to each residue so that the numbers represent the number of contacts that the residue is involved in, also known as the coordination number. For a 1D visualization by showing a single character decimal number below the sequence, residues that are involved in less than nine contacts can be assigned numbers from 1 to 9, and the residues that are involved in more than nine contacts may be assigned a special character like "*." This visualization technique can show if contacts are clustered in a specific region or spread around evenly, and it is effective when we have fewer contacts to analyze, for example L/10, L/5, L/2, L, or even 2L contacts, where L is the length of the protein. In addition, it is also convenient to compare contacts predicted by multiple sources (*see* **Note 2**). An example of a 1D visualization is shown in Fig. 1. The limitation of this visualization technique is that it becomes ineffective when dealing with residues with too many of predicted contacts because all residues will be assigned the "*" character.

Two-dimensional visualization of contacts using contact maps with the help of tools like CMview [13] has been in existence for many decades in the field of proteomics (*see* **Note 4**). A slightly different version of the existing contact maps can help us differentiate long-range contacts from others, and also compare contacts from multiple sources, *see* Fig. 2. To separate the various contact types, different colors may be used for each of the three contact types. Furthermore,



```
         10        20        30        40        50
         |         |         |         |         |
DCCLGYTDRILHPKFIVGFTRQLANEGCDINAIIFHTKKKLSVCANPKQTWVKYIVRLLSK [sequence]
1----1-1----1-11--21-1----1----2212---------111-------1--1-1- [ccmpred.rr: 12 contacts touching 20 residues]
11---1------1--1--21---1--12--2--122-----1--1---------2----1- [pcons_confuzz.rr: 12 contacts touching 18 residues]
------------1--1--31-11---1---221-21--------11-1------2--1-1- [psicov.rr: 12 contacts touching 18 residues]
```

Fig. 1 An example 1D visualization of coordination numbers for predicted contacts. Top L/5 contacts predicted for the protein 1m8a, using three sources of predicted contacts (CCMPRED, PCONS-CONFUZZ, AND PSICOV), are compared in the lines below the sequence row. The numbers below each residue represent the number of contacts that the residue is involved with, such that every contact increases this number for two residues

for each contact prediction source, separate symbols may be used. This allows us to conveniently compare specific contact types of different sources such as long range contacts predicted by two sources. An example of such a contact map visualization is shown in Fig. 2.

2.2 Contact Evaluation

Precision and coverage are two of the most established methods for evaluating predicted protein contacts against a true structure. It is necessary to measure both precision and coverage because often they complement each other (*see* **Note 6**). If we evaluate just a few top predicted contacts and observe their high precision, it does not necessarily imply high coverage. Precision, as shown in Eq. 1, is calculated as the ratio of the number of correctly predicted contacts and the total number of predicted contacts. Coverage, however, may be calculated in three ways. The simplest technique, as shown in Eq. 2, is to calculate the number of predicted contacts divided by the total number of contacts in the native structure [10, 11, 14]. Coverage calculated in this way may result in a relatively smaller value because it is fairly difficult to precisely predict all of the (often redundant) neighboring contacts in the native structure.

Precision,

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{1}$$

where TP is true positive and FP is false positive.

Coverage,

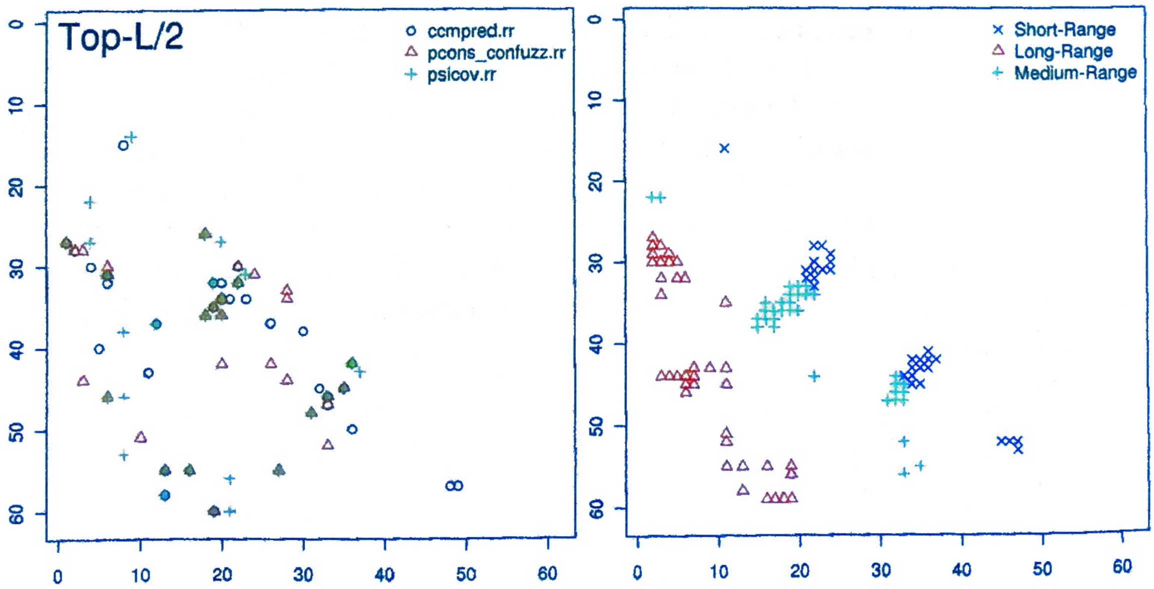


Fig. 2 Examples of contact map visualizations. *Top* L/2 contacts predicted for the protein 1m8a using three different contact prediction sources (*left*). Short-range, medium-range, and long-range true contacts in the native structure of the protein 1m8a are shown in different colors (*right*)

$$\text{Cov} = \frac{100 \times \text{TP}}{\text{no. of native contacts}} \tag{2}$$

$$\text{Mean Error} = \sum \frac{\text{error}}{\text{TP} + \text{FP}} \quad \begin{cases} \text{error} = d - T \text{ if } d > T \\ \text{error} = 0 \qquad \text{if } d \leq T \end{cases} \tag{3}$$

where *d* is the actual distance of a contact in a native structure, and *T* is the distance threshold of the predicted contact.

Distance distribution,

$$X_d = \sum_{i=1}^{15} \frac{Ppi - Pai}{15 \times di} \tag{4}$$

where *Ppi* is the fraction of predicted contacts in bin *i*, and *Pia*—the fraction of all residue pairs in bin *i*.

The second method of evaluating coverage is distance distribution: *Xd* [10–12], measures the weighted harmonic average difference between the distance distribution of predicted contacts and the all-pairs (Eq. 4). Fifteen distance bins cover the range from 0 to 60 Å. The 15 bins include ranges of distances from 0 to 4 Å, 4 to 8 Å, 8 to 12 Å, etc. This score estimates the deviation of the distribution of distances in the list of contacts from the distribution of distances in all pairs of residues in the protein (*see* **Note 5**). The Protein Structure Prediction Center sponsored by the US National Institute of General Medical Sciences (NIH/NIGMS) has been holding biannual meetings featuring preplanned Critical Assessment of protein Structure Prediction (CASP) experiments with specific goals and instructions since 1994. Their goal has been to assist in advancing the current state of the art in protein structure prediction by identifying annual progress and helping to determine where future effort should be most productively focused. CASP6 (2004) focused on precision and *Xd*, and the data from that experiment has been consistently used for contact evaluation in all the CASP competitions afterwards, including the CASP10 (2012) competition. Marks et al. introduced another method for calculating coverage by calculating the spread of contacts [3]. This is computed as the mean of the distances from every experimental (crystal structure) contact to the nearest predicted contact in the 2D contact map.

2.3 Building 3D Models Using Contacts

The emerging success of contact prediction methods demand more research towards building systems that build 3D models from contacts, and one such state-of-the-art method is CONFOLD [1], designed specifically for predicted contacts. The principal idea behind CONFOLD is to build models in two stages to detect self-conflicting contacts. In the first stage, all input contacts are used to build 3D models and the top ranking model in this stage is checked to find the contacts that are not satisfied with a looser definition of a contact. Then the unsatisfied contacts are ignored, in the second stage, as the process of building models begins again. Besides removing self-conflicting contacts in the second stage, predicted strands that are close enough are paired to form beta-sheets in order to improve the accuracy and quality of the models. CONFOLD uses an algorithm known as "distance geometry simulated annealing protocol" implemented in a customized version of a well-established structure determination tool known as the CNS suite [15, 16].

For building 3D models using predicted contacts, the CONFOLD web server may be utilized. On a benchmark data set of 150 globular proteins, contacts predicted by PSICOV [17] were used as input to build 3D models using CONFOLD, to find the Pearson correlation coefficient between the precision of top L/2 contacts and the TM-score of the best models as 0.7. This high correlation suggests that the folding method of CONFOLD is primarily contact-guided, which is ideal for studying the folding information captured in predicted contacts. Unlike many other reconstruction tools, an important feature of CONFOLD is that it can accept secondary structure information (Helix and Strand predictions) along with beta sheet pairing information. This feature may be exploited by predicting secondary structure using a variety of tools in order to obtain a pool of different secondary structures, and then using them in conjunction with the predicted contacts. For building models, CONFOLD transforms the input contacts and secondary structures into restraints for guiding the modeling. In addition, the relative weights between contact restraints and secondary structure restraints can be adjusted, giving us more control over our model building experiments.

Besides CONFOLD, other reconstruction tools may be used for using contacts to build models. Fragment-based ab initio tools like Rosetta and FRAGFOLD [2] can improve their ab initio models using a just few residue contacts. Both ROSETTA and FRAGFOLD can be downloaded and run locally. The template modeling tool, Modeller [18], also accepts secondary structures and contacts as input restraints for building 3D models even though it is not well suited for ab initio modeling [1]. Reconstruction tools like FT-COMAR [19, 20] and Reconstruct [21] have shown state-of-the art performance with true contacts and can accept predicted contacts as input. However, they are not rigorously tested with predicted contacts.

3 Methods

To build 3D models for a given input sequence, we need to decide how many contacts or determine an appropriate maximum number of contacts to consider. When reconstructing using true contacts, we know that this number must be at least 8% of the native contacts [22]. For predicted contacts, although current evaluations consider the top L/2, top L/5, and top L/10 [10, 11] (L being the length of the protein), the number of predicted contacts needed for reconstruction of a protein depends on many factors. These factors include (a) contact prediction method, (b) model building tool, (c) whether or not additional information is used for modeling, and also (d) the protein structure's reconstruct-ability. Some recent studies have considered a range of the number of contacts for building models [1–3] and the authors have suggested using up to top L contacts.

Once the number of contacts is decided, visualization techniques like 2D contact maps help to investigate the coverage and proportion of the three contact types (short-, medium-, and long-range). Upon visualization, if we observe that most of the contacts are clustered only around a specific region of the sequence, we can expect the coverage to be low. Similarly, visualization can also depict the proportion of the three contact types. For building 3D models, it is better to have a mixture of all the three contact types making sure that at least some long-range contacts are included. In addition, it may be important to observe the spread of only the long-range contact as they are considered the most important of the three. When multiple methods are used for contact prediction, visualizations also help to observe the overlaps in predicted sets of contacts. In the case that we have the true structure, however, the selected number of top predicted contacts needs to be evaluated by calculating precision and coverage. In addition, to check how much folding information is captured by the contacts, models may be built using CONFOLD. Below we present the steps for contact assessment.

1. Decide on a tool (or tools) for contact prediction. The results of searching for homologous sequences and templates may suggest whether a template-based method, a machine learning-based method like DNcon [14], NNcon [23], or SVMcon [24], a coevolution-based method like CCMpred [25], EPC-map [26], or FreeContact [27], or a hybrid contact prediction method like MetaPSICOV [28] or PconsC2 [29] is appropriate.
2. Determine the number of contacts for assessments. Typically, top L/10, top L/5, top L/2, or top L may be selected.
3. Visualize the predicted contacts using 1D and 2D methods, *see* Figs. 1 and 2. For a quick visualization submit the predicted contact to the CONASSESS web server. If contacts are predicted using multiple sources, the .RR files should be zipped into a single zip file and then be uploaded.
4. In case a native structure is available, calculate precision, coverage, *Xd*, and mean error using Eqs. 1 through 4 and the measure spread (*see* **Note 3**).
5. Build models using CONFOLD

(a) (Optional) Predict secondary structure for the input sequence. If the sequence does not have any homologous sequences, machine learning tools like SSPro [30] may be used. On the other hand, if many homologous sequences exist, sequence-based tools like Psi-blast based secondary structure prediction (PSIPRED) [31] may be considered.

(b) (Optional) Predict beta sheet pairing information using the predicted secondary structure prediction obtained in (a).

(c) Submit the input sequence, predicted contacts (obtained in **step 1**), secondary structure and beta pairing information (obtained from **steps 5a** and **5b** above) to the CONFOLD web server at http://protein.rnet.missouri.edu/confold/.

(d) Visualize the models by downloading the models from the link received in the e-mail.

4 Case Studies

One useful application of CONASSESS is to analyze predicted contacts when a native structure does not exist for the input sequence. As a case study, consider a 163 residue long CASP11 RR target T0763. The predicted contacts are available in a zip file preloaded in Set 5 of the pre-curated examples in the CONASSESS web server. Assuming that we do not have a native PDB, we may empty the "native pdb" text field. Once the job is submitted to CONASSESS, it calculates the number of long-range contacts and different numbers of top L/10 to top 2L contacts for each of the predicted contacts in the submitted set. The contact map of top L/10 contacts, shown in Fig. 3, shows the overlap in contacts predicted by the various contact prediction groups (or predictors). Upon observing the visualization of coordination numbers of the top L/10 contacts, we notice that the contacts predicted by most of the groups are well distributed over the sequence, but we may also notice some groups whose predicted contacts are clustered in 3 or 4 regions of the sequence. We may also guess that building models using such clustered contacts does not yield good models, and we may need to select the top L/5 or even the top L contacts from such predictors for model building purposes. In addition, if we plan to build models by combining the contacts predicted by all predictors, we may notice from the contact maps that the total number of contacts may be too many to efficiently work with if we select more than the top L/5 contacts.

Another important application of CONASSESS is to evaluate accuracy of predicted contacts against a known native

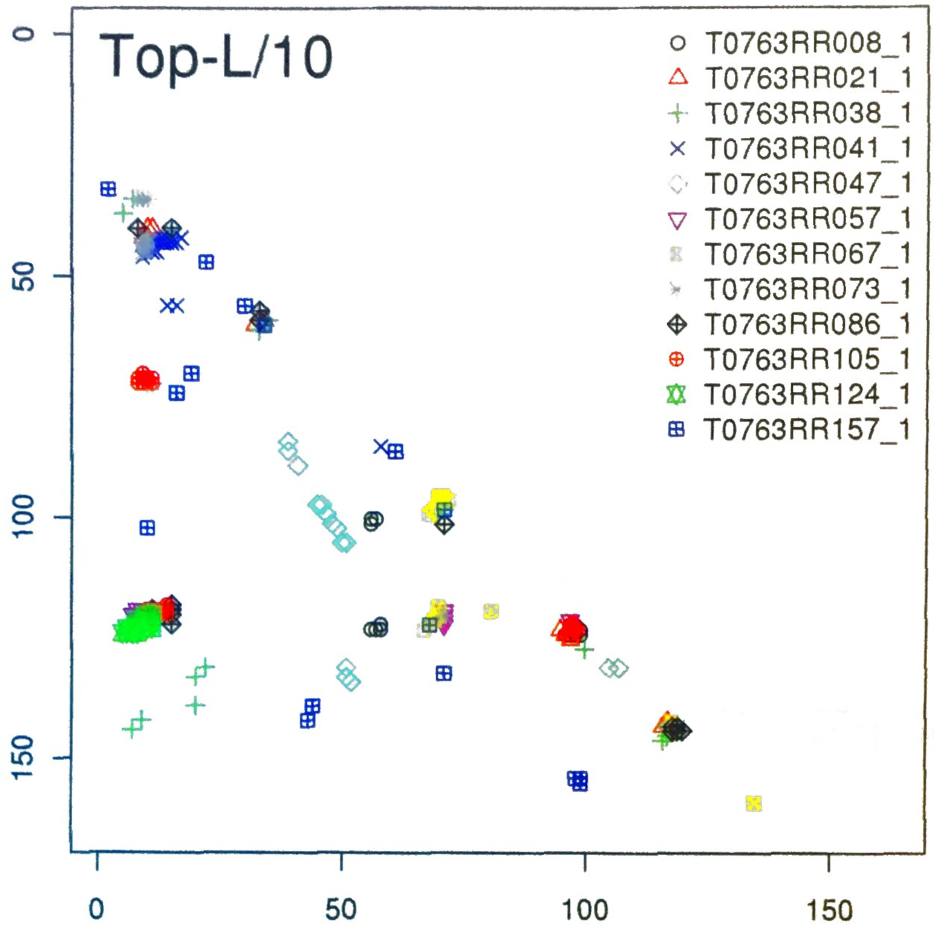


Fig. 3 An screenshot of CONASSESS server's output contact map for the top L/10 contacts predicted for the CASP11 RR target

structure from multiple, complementary perspectives. We may use contacts predicted from a diverse array of methods and readily compare them. As a second case study, let us consider a 145-residue protein (pdb id 1a3a) available in Set 2 of the examples in the CONASSESS web server. If we predict contacts from the sequence using three state-of-the-art approaches like CCMpred [25], PSICOV [17], and PconsC [32] for this protein in a pseudo-blind fashion, we can use CONASSESS web server to evaluate the accuracy of these predicted contacts using measures such as precision, mean error, coverage, distance distribution, and spread. We can then derive some interesting insights by simple visual inspections in addition to detailed, numerical data made available through CONASSESS web server in the form of tables. Fig. 4 shows a representative example for protein 1a3a. In this case the precision of the predicted contacts by PSICOV is higher compared to CCMpred or PconsC when less top ranked contacts are considered. However, CCMpred or PconsC tends to have higher precision of predicted contacts when more top ranked contacts are considered.

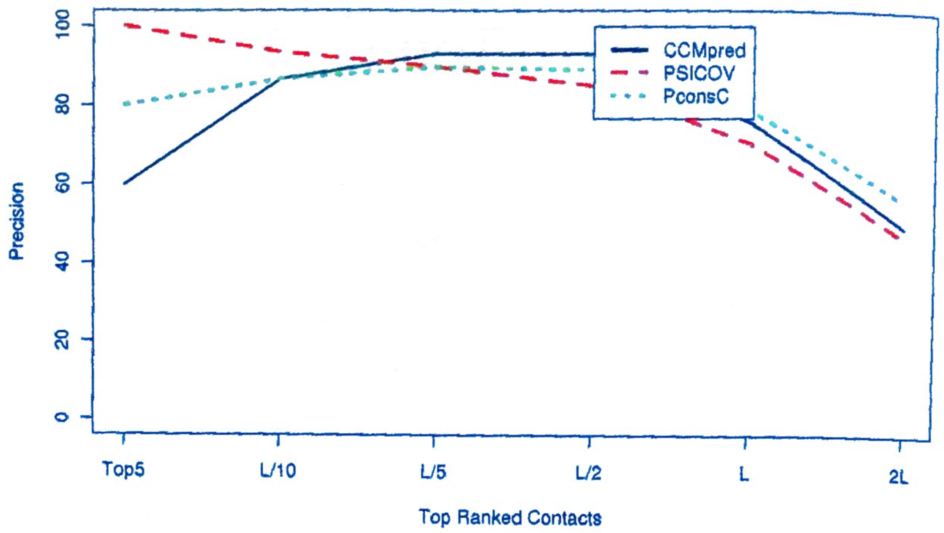


Fig. 4 Precision of predicted contact using three different methods at varied number of top ranked contacts for a representative protein (pdb id 1a3a)

5 Notes

1. Many contact prediction tools often predict many short-range contacts as the confident predictions ranked at the top. Many of these short-range contacts (contacts with small residue sequence separation, usually less than 6 residues) are not always useful if they are the only ones that are used for building models. In a set of top predicted contacts, if the proportion of short-range contacts is high compared to the proportion of long-range contacts, we may need to investigate more to find out if the 3D structure indeed has no (or too few) long-range contacts. The CONASSESS web server may be utilized to check the percentage of short-range contacts in a given set of predicted contacts.
2. Many contact prediction tools may predict contacts clustered in only one or two specific regions of the sequence/structure such as for beta-sheet proteins. Predicting secondary structure using existing tools and visualizing the coordination numbers using a simple 1D technique helps to identify this so that we are able to include more contacts to ensure good coverage.
3. Before contact assessment, make sure that the sequences of predicted contacts and the sequence of the native model are all same. Even if the sequences look similar, scan through the pdb file at least once to check (a) if the file has multiple models (b) if gaps appear in the residue numbering, (c) if residue insertions have been added, and (d) if alternate residues are being used.
4. Visual comparison of contact maps can be misleading. Two contact maps may look similar in contact maps, but the quantitative evaluations can be quite different.

5. The distance distribution score, *Xd*, can have negative values as well. This usually means that the quality of contacts is not good enough because values much higher than 0 usually refer to better contact predictions.
6. It is not surprising to observe high precision values with almost zero coverage for some predicted contacts. For instance, if we are evaluating the top five predicted contacts, and they all are correct, we will get a 100% precision score, but the coverage may be low because five contacts can be too few compared to the total number of contacts in the protein.

References

1. Adhikari B, Bhattacharya D, Cao R, Cheng J (2015) CONFOLD: residue-residue contact-guided ab initio protein folding. Proteins 83(8):1436–1449
2. Kosciolek T, Jones DT (2014) De novo structure prediction of globular proteins aided by sequence variation-derived contacts. PLoS One 9(3):e92197
3. Marks DS, Colwell LJ, Sheridan R, Hopf TA, Pagnani A, Zecchina R, Sander C (2011) Protein 3D structure computed from evolutionary sequence variation. PLoS One 6(12):e28766
4. Bhattacharya D, Cheng J (2015) De novo protein conformational sampling using a probabilistic graphical model. Sci Rep 5:16332
5. Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, Kaufman K, Renfrew PD, Smith CA, Sheffler W, Davis IW, Cooper S, Treuille A, Mandell DJ, Richter F, Ban YE, Fleishman SJ, Corn JE, Kim DE, Lyskov S, Berrondo M, Mentzer S, Popovic Z, Havranek JJ, Karanicolas J, Das R, Meiler J, Kortemme T, Gray JJ, Kuhlman B, Baker D, Bradley P (2011) ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. Methods Enzymol 487:545–574. doi:10.1016/b978-0-12-381270-4.00019-6
6. Zhang Y (2008) I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9(1):40
7. Mabrouk M, Putz I, Werner T, Schneider M, Neeb M, Bartels P, Brock O (2015) RBO Aleph: leveraging novel information sources for protein structure prediction. Nucleic Acids Res 43(W1):W343–W348. doi:10.1093/nar/gkv357
8. Chen J, Zhang L, Jing L, Wang Y, Jiang Z, Zhao D (2003) Predicting protein structure from long-range contacts. Biophys Chem 105(1):11–21
9. Gromiha MM, Selvaraj S (1999) Importance of long-range interactions in protein folding. Biophys Chem 77(1):49–68
10. Monastyrskyy B, Fidelis K, Tramontano A, Kryshtafovych A (2011) Evaluation of residue–residue contact predictions in CASP9. Proteins 79(S10):119–125
11. Monastyrskyy B, D'Andrea D, Fidelis K, Tramontano A, Kryshtafovych A (2014) Evaluation of residue–residue contact prediction in CASP10. Proteins 82(S2):138–153
12. Ezkurdia I, Grana O, Izarzugaza JM, Tress ML (2009) Assessment of domain boundary predictions and the prediction of intramolecular contacts in CASP8. Proteins 77(S9):196–209
13. Vehlow C, Stehr H, Winkelmann M, Duarte JM, Petzold L, Dinse J, Lappe M (2011) CMView: interactive contact map visualization and analysis. Bioinformatics 27(11):1573–1574
14. Eickholt J, Cheng J (2012) Predicting protein residue–residue contacts using deep networks and boosting. Bioinformatics 28(23):3066–3072
15. Brunger AT (2007) Version 1.2 of the crystallography and NMR system. Nat Protoc 2(11):2728–2733
16. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang J-S, Kuszewski J, Nilges M, Pannu NS (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 54(5):905–921
17. Jones DT, Buchan DW, Cozzetto D, Pontil M (2012) PSICOV: precise structural contact prediction using sparse inverse covariance estimation on large multiple sequence alignments. Bioinformatics 28(2):184–190
18. Eswar N, Webb B, Marti-Renom MA, Madhusudhan M, Eramian D, Shen M, Pieper U, Sali A (2006) Comparative protein struc-

ture modeling using Modeller. Curr Protoc Bioinformatics 5.6.1:5.6.32

19. Vassura M, Margara L, Di Lena P, Medri F, Fariselli P, Casadio R (2008) FT-COMAR: fault tolerant three-dimensional structure reconstruction from protein contact maps. Bioinformatics 24(10):1313–1315
20. Di Lena P, Vassura M, Margara L, Fariselli P, Casadio R (2009) On the reconstruction of three-dimensional protein structures from contact maps. Algorithms 2(1):76–92
21. Duarte JM, Sathyapriya R, Stehr H, Filippis I, Lappe M (2010) Optimal contact definition for reconstruction of contact maps. BMC Bioinformatics 11(1):283
22. Sathyapriya R, Duarte JM, Stehr H, Filippis I, Lappe M (2009) Defining an essence of structure determining residue contacts in proteins. PLoS Comput Biol 5(12):e1000584
23. Tegge AN, Wang Z, Eickholt J, Cheng J (2009) NNcon: improved protein contact map prediction using 2D-recursive neural networks. Nucleic Acids Res 37(suppl 2):W515–W518
24. Cheng J, Baldi P (2007) Improved residue contact prediction using support vector machines and a large feature set. BMC Bioinformatics 8(1):113
25. Seemayer S, Gruber M, Söding J (2014) CCMpred—fast and precise prediction of protein residue–residue contacts from correlated mutations. Bioinformatics 30(21):3128–3130
26. Schneider M, Brock O (2014) Combining physicochemical and evolutionary information for protein contact prediction. PLoS One 9(10), 10.1371/journal.pone.0108438
27. Kaján L, Hopf TA, Marks DS, Rost B (2014) FreeContact: fast and free software for protein contact prediction from residue co-evolution. BMC Bioinformatics 15(1):85
28. Jones DT, Singh T, Kosciolek T, Tetchner S (2014) MetaPSICOV: combining coevolution methods for accurate prediction of contacts and long range hydrogen bonding in proteins. Bioinformatics. btu791
29. Skwark MJ, Raimondi D, Michel M, Elofsson A (2014) Improved contact predictions using the recognition of protein like contact patterns. PLoS Comput Biol 10(11):e1003889
30. Cheng J, Randall AZ, Sweredoski MJ, Baldi P (2005) SCRATCH: a protein structure and structural feature prediction server. Nucleic Acids Res 33(suppl 2):W72–W76
31. McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. Bioinformatics 16(4):404–405
32. Skwark MJ, Abdel-Rehim A, Elofsson A (2013) PconsC: combination of direct information methods and alignments improves contact prediction. Bioinformatics 29(14):1815–1816. doi:10.1093/bioinformatics/btt259