

Finding the beta-beta residue pairing for protein structure prediction

Prediction of the tertiary structure of proteins from their amino acid sequences is of great importance and of urgent demand in structural bioinformatics and biophysics, especially considering the increasing gap between the numbers of sequences and structures of proteins determined experimentally. Knowledge of native contacts between amino acid residues can greatly facilitate the protein structure prediction, and therefore the prediction of protein residue contacts from amino acid sequences has attracted more and more attention. In principle, native residue contacts that are essential for protein structure or function could be inferred from correlated mutations of residue pairs in evolution. With sequence data accumulated at an unprecedented speed, extraction of such coevolution information from multiple sequence alignment has become more and more practicable. Unfortunately, even the residue contact maps predicted from the state-of-the-art algorithms frequently contain many errors, which impairs the accuracy of structural models generated by contact-assisted protein structure prediction. However, information of residue pairing in β strands could be rather reliably inferred from a noisy contact map, due to the presence of characteristic contact patterns in β - β interactions. Such β - β contact information may benefit the tertiary structure prediction of mainly β proteins, a group of protein targets that have complex topologies and thus are highly challenging in protein structure prediction.

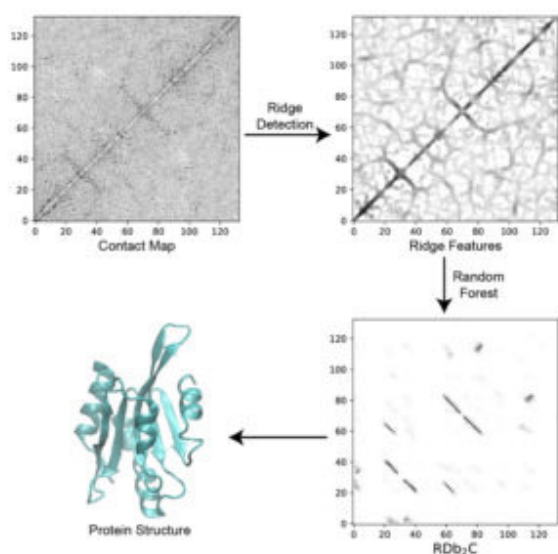


Fig. 1.

This work introduces a novel ridge-detection-based β - β contact predictor, RDb₂C, to identify residue pairing in β strands from any predicted residue contact map. Theoretically, consecutive residue pairs from interacting β strands should present continuous contacts in the diagonal or off-diagonal directions on a native contact map, which may be disguised by noises in a predicted contact map. Here, we adopted the ridge detection, a traditional computer image processing technique, to capture the characteristic pattern of β - β interactions from predicted contact maps with high levels of noises. We then designed a novel multi-stage random-forest

framework, which reads in the ridge information as well as additional features and predicts the residue pairing between all β strands. Our method could effectively improve the prediction on β - β contacts and thus the structural modeling of mainly β proteins.

RDb₂C could refine the β - β contact prediction for any predicted contact maps. Starting from the contact maps predicted by CCMpred, a popular sequence-alignment-based contact predictor, RDb₂C reaches F1-scores of ~62% and ~76% at the residue level and strand level respectively, when evaluated on two conventional test sets of β proteins, BetaSheet916 and BetaSheet1452. Even in this case, our method remarkably outperforms all state-of-the-art β - β contact prediction methods, including the famous bbcontacts, which exhibits F1-scores of ~56% and ~72% at the residue level and strand level respectively. Moreover, when taking the contact maps predicted by the more advanced RaptorX-Contact as input, RDb₂C achieves impressively higher performance, with F1-scores reaching ~76% and ~86% at the residue level and strand level, respectively. Enhancement in the accuracy of β - β contact prediction can further improve the contact-assisted structural prediction of mainly β proteins in a test of structural modeling using the top L (L is the protein length) predicted contacts as constraints. For 61 mainly β proteins that contain $\geq 50\%$ β residues, the quality of predicted structural models (as evaluated by TM-score) is 0.442 on average, when using the raw RaptorX-Contact prediction. In contrast, when using the improved prediction by RDb₂C, the number increases to 0.506, indicating that the predicted models have the correct fold (TM-score > 0.5) on average.

The original codes, source data/results and the training protocol for practical usage are available at <http://structpred.life.tsinghua.edu.cn/Software.html> or at the GitHub address of <https://github.com/wzmao/RDb2C>.

Wenzhi Mao, Haipeng Gong
*School of Life Sciences and Beijing Advanced Innovation Center for Structural Biology,
Tsinghua University, Beijing 100084, China*

Publication

[Identification of residue pairing in interacting \$\beta\$ -strands from a predicted residue contact map.](#)

Mao W, Wang T, Zhang W, Gong H
BMC Bioinformatics. 2018 Apr 19