DNCON2: improved protein contact prediction using two-level deep convolutional neural networks

Badri Adhikari¹, Jie Hou¹ and Jianlin Cheng¹,²,*

¹Department of Mathematics and Computer Science, University of Missouri-St. Louis, St. Louis, MO 63121, USA and ²Informatics Institute, University of Missouri, Columbia, MO 65211, USA

*To whom correspondence should be addressed.

Abstract

Motivation: Significant improvements in the prediction of protein residue–residue contacts are observed in the recent years. These contacts, predicted using a variety of coevolution-based and machine learning methods, are the key contributors to the recent progress in ab initio protein structure prediction, as demonstrated in the recent CASP experiments. Continuing the development of new methods to reliably predict contact maps is essential to further improve ab initio structure prediction.

Results: In this paper we discuss DNCON2, an improved protein contact map predictor based on two-level deep convolutional neural networks. It consists of six convolutional neural networks—the first five predict contacts at 6, 7.5, 8, 8.5 and 10 Å distance thresholds, and the last one uses these five predictions as additional features to predict final contact maps. On the free-modeling datasets in CASP10, 11 and 12 experiments, DNCON2 achieves mean precisions of 35, 50 and 53.4%, respectively, higher than 30.6% by MetaPSICOV on CASP10 dataset, 34% by MetaPSICOV on CASP11 dataset and 46.3% by Raptor-X on CASP12 dataset, when top L/5 long-range contacts are evaluated. We attribute the improved performance of DNCON2 to the inclusion of short- and medium-range contacts into training, two-level approach to prediction, use of the state-of-the-art optimization and activation functions, and a novel deep learning architecture that allows each filter in a convolutional layer to access all the input features of a protein of arbitrary length.

Availability and implementation: The web server of DNCON2 is at http://sysbio.rnet.missouri.edu/dncon2/ where training and testing datasets as well as the predictions for CASP10, 11 and 12 free-modeling datasets can also be downloaded. Its source code is available at https://github.com/multicom-toolbox/DNCON2/.

Contact: chengji@missouri.edu

Supplementary information: Supplementary data are available at Bioinformatics online.
Rosetta (Ovchinnikov et al., 2016) and Quark (Xu and Zhang, 2012), for fragment-free methods like CONFOLD (Adhikari et al., 2015) and GFuzz3D (Pietal et al., 2015), other two types of contacts—short- and medium-range contacts—are also important. While successful contact prediction methods like DNCON (Eickholt and Cheng, 2012) find it effective to predict these separately, a more recent trend of predicting all contacts with a single machine learning architecture appears promising (Jones et al., 2014; Wang et al., 2017).

Much of the recent improvement in the performance of contact prediction is from detecting coevolving residue pairs in a multiple sequence alignment and from the machine learning techniques used to integrate these predictions as features along with other standard features. Coevolution-based contact predictors can generally predict accurate contacts in presence of at least a few hundred effective sequences in the input alignment (Michel et al., 2017b). However, recent state-of-the-art methods demonstrate that integrating these co-evolution-based predictions with other features and using a machine learning method to make final predictions, can almost always perform better than using coevolution information alone. These integrative contact predictors have used neural networks (Jones et al., 2014), random forests (Skwark et al., 2014) and convolutional neural networks (Wang et al., 2017) to combine co-evolutionary features with other common features like secondary structures and position specific scoring matrices.

As a successor of our deep belief network based contact predictor, DNCON (Eickholt and Cheng, 2012, 2013), which was ranked as the top method in the CASP10 experiment (Monastyrskyy et al., 2014), in this paper, we present our improved contact prediction method—DNCON2. The primary enhancements of DNCON2 are (i) inclusion of coevolution-based features, (ii) new deep convolutional neural networks to predict full contact maps and (iii) addition of new features at multiple distance thresholds, which further improves the performance. In DNCON2, we transform all 27 input features, e.g. scalar features like protein length, one-dimensional (1D) features like secondary structure prediction and two-dimensional (2D) features like coevolution-based predictions, into 56 two-dimensional features. As the first step of our two-level prediction approach, we train five convolutional neural networks (CNNs), which accept these 56 2D features as input to predict contact maps at distance thresholds of 6, 7.5, 8, 8.5 and 10 Å. In the second level, a separate CNN is trained with these five sets of predictions as additional 2D features, to make final short-, medium- and long-range predictions in one contact map all at once. Finally, we test our method using the free-modeling datasets of CASP10, 11 and 12 and compare it with other state-of-the-art methods, and, also discuss how the various training hyper-parameters influence the performance.

2 Materials and methods
2.1 Datasets and evaluation metrics
We used the original DNCON dataset consisting of 1426 proteins having length between 30 and 300 residues curated before the CASP10 experiment to train and test DNCON2. The protein structures in the dataset were obtained from the Protein Data Bank (PDB) and of 0-2 Å resolution, which were filtered by 30% sequence identity to remove redundancy. 1230 proteins from the dataset are used for training and 196 as the validation set, and the two sets have less than 25 percent sequence identity between them. In addition to the validation dataset, we benchmarked our method using (i) 37 free-modeling domains in the CASP12 experiment, (ii) 30 free-modeling domains in the CASP11 experiment (Kinch et al., 2016) and (iii) 15 free-modeling domains in the CASP10 experiment (Monastyrskyy et al., 2014). These CASP free-modeling datasets have zero or very little identity with the training dataset.

In this study, we define a pair of residues in a protein to be in contact if their carbon beta atoms (carbon alpha for glycine), are closer than 8 Å in the native structure. We consider contacts as long-range when the paring residues are separated by at least 24 residues in the protein sequence. Similarly, medium-range contacts are pairs which have sequence separation between 12 and 23 residues and short-range contacts are pairs with sequence separation between 6 and 11 residues. These definitions are consistent with the common standards used in the field (Monastyrskyy et al., 2016).

As a primary evaluation metric of contact prediction accuracy, we use the precision of top L/5 or L/2 predicted long-range contacts, where L is the length of the protein sequence. The metric has also been the main measure in the recent CASP evaluations (Monastyrskyy et al., 2011, 2014, 2016) and some recent studies (Adhikari et al., 2016). When evaluating the predictions for the proteins in the CASP datasets, we evaluate them at the domain level to be consistent with the past CASP assessments, although all predictions were made on the full target sequences without any knowledge of domains. We used the ConEVA tool to carry out our evaluations (Adhikari et al., 2016).

2.2 Input features
In addition to the existing features used in the original DNCON, we used new features derived from multiple sequence alignments, coevolution-based predictions and three-state secondary structure predictions from PSIPRED (Jones, 1999). The original DNCON feature set includes length of the protein, secondary structure and solvent accessibility predicted using the SCRATCH suite (Cheng et al., 2005), position specific scoring matrix (PSSM) based features (e.g. PSSM sums and PSSM sum cosines), Atchley factors, and several pre-computed statistical potentials. During our experiments, we found PSSM and amino acid composition from the original DNCON feature set were not very useful and hence removed them from the feature list. Besides the DNCON features, the new features include coevolutionary contact probabilities/scores predicted using CCMPred (Seemayer et al., 2014), FreeContact (Kaján et al., 2014), PSICOV (Jones et al., 2012) and alignment statistics such as number of effective sequences, Shannon entropy sum, mean contact potential, normalized mutual information and mutual information generated using the alignment statistics tool ‘alnstat’ (Jones et al., 2014). During our experiments, often, PSICOV did not converge when there are too many or too few alignments, especially if the target sequence is long. To guarantee to get some results, we set a time limit of 24h, and run PSICOV with three convergence parameters (‘d = 0.03’, ‘r = 0.001’ and ‘r = 0.01’) in parallel. If the first prediction (with option d = 0.03) finishes within 24h, we use the prediction, and if not, we use the second prediction and so on. Using all these features above as input, we predict contact maps at 6, 7.5, 8, 8.5 and 10 Å distance thresholds at first, and then use these five contact-map predictions as additional features to make a second round of prediction. Contact predictions at lower distance thresholds are relatively sparse and include only the residue pairs that are very close in the structure, whereas, contact predictions at higher distance thresholds are denser and provide more positive cases for the deep convolution neural network to learn (see Supplementary Fig. S1 for visualization of contact maps at various distance
thresholds, and Supplementary Table S2 for a full list of all the features used).

2.3 Generating multiple sequence alignments
Generating a diverse/informative multiple sequence alignment with a sufficient number of sequences is critical for generating quality coevolution-based features for contact prediction. On one hand, having too few sequences in the alignment, even though they may be highly diverse, can lead to low contact prediction accuracy. On the other hand, having too many sequences can slow down the process of co-evolution feature generation, creating a bottleneck for an overall structure prediction pipeline. To reliably generate multiple sequence alignments, an alignment method should produce at least some sequences in alignment whenever possible, and does not generate too many more sequences than necessary. Following a similar procedure in (Ovchinnikov et al., 2016) and (Kosciolek and Jones, 2015), we first run HHblits (Remmert et al., 2011) with 60% coverage thresholds, and if a certain number of alignments are not found (usually around 2L), then we run JackHMMER (Johnson et al., 2010) with e-value thresholds of 1E−20, 1E−10, 1E−4 and 1 until we find some alignments. JackHMMER is not run if HHblits can find at least 5000 sequences. These alignments are used by the three coevolution-based methods (CCMpred, FreeContact and PSICOV) to predict contact probabilities/scores, which are used as two-dimensional features and to generate alignment statistics related features for deep convolutional neural networks.

2.4 Deep convolutional neural network architecture
Convolutional neural networks (CNNs) are widely applied to recognize images with each input image translated into an input volume such that the size of the image are length and width of the volume, and the three channels (hue, saturation and value) represent the depth. Based on such ideas, to build an input volume for each protein of any length, we translate all scalar and one-dimensional input features into two-dimensional features (channels) so that all features (including the ones already in 2D) are in two-dimension and can be viewed as separate channels. While scalar features like sequence (including the ones already in 2D) are in two-dimension and can be expanded to around 35 GB when all features are translated to 2D format. To minimize disk input/output, we translate our scalar features and 1D features into 2D, at runtime, in CPU memory. We used the Keras library (http://keras.io/) along with Tensorflow (www.tensorflow.org) to implement our deep CNN networks. Our training was conducted on Tesla K20 Nvidia GPUs each having 5 GB of GPU memory, on which, training one model took around 12 h. Finally, we use an ensemble of 20 trained deep models to make final predictions for testing on the CASP datasets.

3 Results
3.1 Using contact predictions at 6, 7.5, 8, 8.5 and 10 Å distance thresholds as features improves precision
A contact map is a binary version of the distance map of a protein structure according to a distance threshold, which usually is 8 Å. This threshold of 8 Å, although widely used, can be viewed as an arbitrary and rigid criterion to decide if a pair of residue is a contact or non-contact, considering the flexibility of protein structures. For instance, a pair separated by 8.1 Å distance is, by definition, a non-contact, but by 7.9 Å is a contact. And using one distance threshold causes the loss of some distance information. In order to account for uncertainty and flexibility in residue-residue distance, in a first round of prediction, using all the features and true contact maps at 6, 7.5, 8, 8.5 and 10 Å distance thresholds, we trained five CNN models to predict contact probabilities at these five distance thresholds. Then, in the second round of prediction, we added these predictions as new 2D features into the feature list and trained a sixth CNN model to predict contacts at 8 Å distance threshold.

On the 196 proteins in the validation dataset, the CNN model in the second level achieves a precision of up to 73.5% higher than 70.7% in the first level, when top 1/5 long-range contacts are evaluated. To verify if the improvement comes from the predictions at different distance thresholds or from the iterative two-level training, in a separate experiment, we trained a second level model with only the contact prediction at 8 Å distance threshold as additional feature. In this case, a precision of 72.2% is achieved, higher than 70.7% of using one level prediction, but lower than 73.5% of using both two-level prediction and multiple thresholds, indicating that both two-level training and multiple thresholds contribute to the improvement. The results summarized in Figure 2 show similar results when top 1/2 contacts are evaluated. In addition to these experiments, we tested adding more predictions at higher distance thresholds of 12, 14, 16 and 18 Å as features, and found that they did not significantly improve the performance. As an additional validation, we used the ensemble of the models trained with five distance thresholds in the first level to predict the contacts for the proteins in the validation dataset, similar to a traditional neural network ensemble in (Jones et al., 2014). Such a multi-distance ensemble has a precision of 72.8% slightly higher than the 72.6% precision achieved by an ensemble of all five models trained at the same 8 Å distance threshold, but lower than 73.5% of using predictions of multiple thresholds with two-level networks in DNCON2.
3.2 Comparison between deep belief networks in DNCON 1.0 and deep convolutional neural networks in DNCON2

DNCON 1.0 used an ensemble of deep belief networks (DBN) trained with windows of seven different fixed sizes and boosting to predict contacts and achieved an accuracy of 34% on the 196 proteins in the validation dataset. For a fair comparison with DNCON, we trained one CNN using the same features that DNCON 1.0 used (excluding new coevolution-based features). Different from the DNCON 1.0 of using different networks to predict contacts at different ranges, DNCON2 uses a CNN network to predict short-, medium- and long-range contacts of a protein of arbitrary length. With the same features as input, a CNN network trained with all contacts and non-contacts achieves a slightly better precision of 35.4% on top L/5 long-range contacts than DNCON 1.0. So, a single CNN model performs better than a boosted and ensembled deep belief networks, suggesting that the deep convolutional neural network (CNN) is more suitable for contact prediction than the deep belief network (DBN). Moreover, it is more convenient to train and test CNN than DBN because CNN can take a full input matrix of arbitrary size as input to predict full contact maps without generating the features for each pair of residues, separating contacts at different ranges, and balancing the ratio of contacts and non-contacts as required by DBN based on fixed size windows.

3.3 Performance of DNCON2 on the validation and CASP datasets

On the 196 proteins in the validation dataset, DNCON2 yields a mean precision of 74%, when top L/5 long-range contacts are evaluated. As summarized in Table 1, the average length, number of sequences in the alignment, and the number of effective sequences for these proteins are 190, 5351 and 1718 respectively. On this dataset, the three-individual coevolution-based features generated by CCMpred, FreeContact and PSICOV can predict contacts with precisions of 51.0, 43.1, 42.1% respectively, for top L/5 long-range contacts, which is much lower than 74% of DNCON2. And for 96% of these proteins, DNCON2 performs better than any of the individual coevolution-based features (see Supplementary Fig. S3). These results indicate that integrating all the 2D coevolution-based features with the other features can drastically improve the accuracy of contact prediction.

Since predicted contacts are most useful for ab initio folding of proteins whose structures cannot be predicted by template-based modeling, we evaluated our method on the free-modeling protein datasets in the CASP10, 11 and 12 experiments and compared it with top CASP methods and a standard coevolution-based method MetaPSICOV (Jones et al., 2014) (see Table 2). Since our training and validation datasets were curated before the CASP10 experiment, the CASP datasets are independent test data. For evaluating
Table 1. Performance of DNCON2 on the 196 proteins in the validation dataset when top L/5 and top L/2 long-range contacts are evaluated

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>N</th>
<th>N eff</th>
<th>P L/5</th>
<th>P L/2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>190</td>
<td>5351</td>
<td>1718</td>
<td>74.0%</td>
<td>64.2%</td>
</tr>
<tr>
<td>Median</td>
<td>188</td>
<td>1607</td>
<td>412</td>
<td>88.2%</td>
<td>74.4%</td>
</tr>
<tr>
<td>Maximum</td>
<td>299</td>
<td>62889</td>
<td>29547</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Minimum</td>
<td>50</td>
<td>1</td>
<td>1</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

Notes: L, N and N eff stand for length of a protein, number of sequences in the alignment, and the number of effective sequences in the alignment. P L/5 and P L/2 are the precisions of top L/5 and L/2 long-range contacts.

Table 2. Summary of the performance of DNCON2 on the CASP10, CASP11 and CASP12 free-modeling (FM) datasets, measured using the precision of top L/5 long-range contacts

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Domain Count</th>
<th>Top CASP Group</th>
<th>Meta PSICOV</th>
<th>DNCON 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>CASP10</td>
<td>15</td>
<td>18.1 (DNCON 1.0)</td>
<td>30.6</td>
<td>35.0</td>
</tr>
<tr>
<td>CASP11</td>
<td>30</td>
<td>29.7 (CONSIP2)</td>
<td>34.4</td>
<td>50.0</td>
</tr>
<tr>
<td>CASP12</td>
<td>37</td>
<td>46.3 (Raptor-X)</td>
<td>42.9</td>
<td>53.4</td>
</tr>
</tbody>
</table>

Note: The precision of the top method in each CASP experiment and a standard method MetaPSICOV (run locally) is also included as a reference.

Our method on the most recent CASP12 dataset, we generated all features using all programs and databases released before the CASP12 experiment started, making our results not influenced by the new releases of protein structures and sequences thereafter. On the 37 free-modeling datasets for which native structures were available for us to perform the evaluation, DNCON2 outperforms all the top methods participating in the CASP12 experiment such as Raptor-X (Wang et al., 2017) (Table 2), MetaPSICOV (Jones et al., 2014) (Table 2), iFold1, and our own method MULTICOM-CONSTRUCT as well as the baseline method DNCON 1.0 (see Supplementary Table S4). Specifically, when top L/5 long-range contacts are evaluated, DNCON2 achieves an average precision of 53.4% compared to 46.3, 42.9 and 45.7% by Raptor-X, MetaPSICOV and iFold1, respectively. A similar performance is observed when top L/2 contacts are evaluated instead of L/5 (see Supplementary Table S4). The 24.9% precision of DNCON 1.0, which does not use any coevolution-based features, is a benchmark for the other methods, and the difference between its accuracy and the other methods highlights the improvement gained from the inclusion of the coevolution-based features into the input. The full evaluation on the complete set of CASP12 domains is presented in Supplementary Table S5.

For evaluating our method on CASP11 and CASP10 free-modeling datasets, we ran MetaPSICOV locally as a benchmark. For a fair comparison, we use the same sequence databases for DNCON2 and MetaPSICOV. For completeness, we also compared DNCON2 with the best performing groups in the two CASP experiments—CONSIP2 in the CASP11 experiment (Monastyrskyy et al., 2016), and DNCON 1.0 in the CASP10 experiment (Monastyrskyy et al., 2014) (Table 2). On the 30 free-modeling domains in the CASP11 experiment, DNCON2 achieves an average precision of 50% compared to 34.4% by MetaPSICOV and 29.7% by the best performing method CONSIP2 (Kosiolek and Jones, 2015) in CASP11, when top L/5 long-range contacts are evaluated. Similarly, on the 15 free-modeling structural domains in the CASP10 experiment, DNCON2 achieves a mean precision of 35%, compared to 30.6% by MetaPSICOV and 18.1% by the best performing method DNCON. For both datasets, similar results are observed when medium-range and short-range contacts are evaluated. Detailed evaluations of top L/5 and top L/2 contacts, are presented in Supplementary Tables S6 and S7.

3.4 Hyper-parameters optimization

To obtain the best performance on the validation dataset, we fine-tuned our network by investigating a range of values/options for the following hyper-parameters: (i) depth of the network, (ii) filter sizes in each layer, (iii) number of filters in each layer, (iv) batch normalization, (v) batch size, (vi) optimization function and (vii) activation function. After many rounds of iterative hyper-parameter selection, we found that the optimal parameters for number of layers was seven, filter size was five, number of filters was 16, batch size was 30, and chosen ReLU as the activation function in hidden layers, applied batch normalization at each layer, and used NAdam as the optimization function. With the performance of the CNN in a setting as a reference, we tuned each parameter, one-by-one, to study how they influenced the performance on the training data and validation data. For the depth of the network we tested networks with two to nine layers. Similarly, for filter size and number of filters, we tested filter sizes of 1, 3, 5, 7, 9 and 11, and number of filters as 1, 4, 8, 16 and 24. On the validation dataset, the networks with filter sizes greater than 3, with >8 filters and >5 hidden layers deliver around the top performance (see Supplementary Fig. S8). When the filter size is increased beyond 9, or the number of filters is increased beyond 24, or the depth of the network is increased beyond 9, the training was very slow and often failed because of insufficient GPU memory. In addition, through trials, we found that keeping the filter size in all seven layers and number of filters in the six hidden layers the same performs better than having different filter sizes or numbers of filters in different layers.

Batch normalization is important for training deep CNN to deal with the covariate shift problem (Ioffe and Szegedy, 2015). To test how batch normalization affects the training performance, we tried applying batch normalization after each layer (a), after every alternate layer (b), or only on the first layer (c) and not using batch-normalization at all (d). We found that applying batch normalization at each layer delivers the best performance compared to any of the three other settings. While the full batch normalization applied after each layer delivers a mean precision of 70.8% and the batch normalization at every alternate layer results in a mean precision of 68.7%, for top L/5 predicted long-range contacts on the validation dataset. When batch normalization is not used at all, or is applied only to the first layer, the precision drops to 65.7%. Similarly, after testing various batch sizes, we found batch sizes of around 30 delivered the best performance on the validation dataset. For optimization methods, we tested (a) ADADELTA, (b) Adagrad, (c) Adam, (d) Nesterov Adam, (e) RMSprop and (f) stochastic gradient descent optimizers. The results show that three optimization functions Adam, Nesterov Adam and RMSprop deliver better performance than the others, with Nesterov Adam performing best among all (see Supplementary Fig. S9). Finally, the activation functions sigmoid, tanh and ReLU can achieve the precisions of 70.4, 69.4 and 70.9% respectively, when top L/5 long-range contacts are evaluated.

Besides the machine learning hyper-parameters, we also tested if training using only long-range contacts improves the precision of long-range contact prediction. Interestingly, we find that including medium-range contact and short-range contacts into training
improves the performance even when only long-range contacts are evaluated. However, when all the local contacts are included, i.e. contacts with sequence separation less than five, we observed a slight decrease in performance (see Supplementary Fig. S10). In summary, including all except the local contacts during training yields better precision. In addition, to test how ensembling improves the performance, we first ranked the trained models using the average precision on the validation dataset. Then, we calculated precision of the averaged predictions of top $x$ models, where $x$ is an integer in the range $[1, 50]$. The precision of ensembling increases initially and then saturates after more than four models are used.

3.5 Importance of features

We removed one or more features at a time and trained the CNN using the remaining features, to study the contribution of the removed features towards the overall performance of DNCON2. We tested by removing (i) multiple sequence alignment (MSA) statistics related features comprising of Shannon entropy sum, mean contact potential, normalized mutual information and mutual information, (ii) CCMpred coevolution feature, (iii) FreeContact coevolution feature, (iv) PSICOV coevolution feature, (v) several pre-computed statistical potentials, (vi) number of sequences in the alignment and the number of effective sequences in the alignment, (vii) PSIPRED and PSISOLV predictions of secondary structures and solvent accessibility, (viii) PSSM related features comprising of PSSM sums and PSSM sum cosines, (ix) SCRATCH secondary structure and solvent accessibility predictions, (x) relative counts of helical, strand and buried residues, also do not contribute much to the performance either. Secondary structure predictions from both methods SCARTCH and PSIPRED are useful, and complement each other to improve the overall performance.

4 Conclusion

We developed DNCON2—a new two-level deep convolutional neural network method—to predict the contact map of a protein of any length by integrating both residue–residue coevolution features and other features such as secondary structures, solvent accessibility and pairwise contact potentials. The method can predict all the contacts in a protein at once from the entire input information of a protein, which is more effective and easier to train and use than local fixed window-based approaches such as deep belief networks. By including new coevolution features, using CNNs of multiple-distance thresholds, integrating all the features of all the residue pairs of a protein through 2D-convolution in a two-level architecture, and adopting the latest optimization and training techniques, DNCON2’s accuracy is more than double of that of DNCON 1.0 on the same validation dataset. On the three independent CASP datasets, DNCON2 outperforms the top methods in CASP10, CASP11 and CASP12 experiments. The results demonstrate DNCON2 and its deep convolutional neural network architecture are useful for protein contact prediction.

Acknowledgements

Some of the computation for this work was performed on the high-performance computing infrastructure provided by Research Computing Support Services and in part by the National Science Foundation under grant number CNS-1429294 at the University of Missouri, Columbia MO. We would like to thank the RCSS team for their infrastructure and technical support. We would also like to thank Dr. Andrew Ng for his freely accessible lectures on Machine Learning at coursera.org.
Funding
This work was supported by the National Institutes of Health (NIH) R01 grant (grant number: R01GM093123) to JC.

Conflict of Interest: none declared.

References


Kaján,L. et al. (2014) FreeContact: fast and free software for protein contact prediction from residue co-evolution. BMC Bioinformatics, 15, 85.


