

Performance Analysis of Deep Learning Methods for Protein Contact Prediction in CASP13

Romina Valdez, Khevin Roig

Universidad Nacional de Asuncion, Facultad Politecnica,
San Lorenzo, Paraguay, 2160,
{valdezromina, maxiroig}@fpuna.edu.py

and

Diego P. Pinto-Roa, Jose Colbes

Universidad Nacional de Asuncion, Facultad Politecnica,
San Lorenzo, Paraguay, 2160,
{dpinto, jcolbes}@pol.una.py

Abstract

Protein structure prediction is one of the most important problems in Computational Biology; and consists of determining the 3D structure of a protein given its amino acid sequence. A key component that has allowed considerable improvements in recent decades is the prediction of contacts in a protein, since it provides fundamental information about its three-dimensional structure. In the 13th edition of the CASP (*Critical Assessment of protein Structure Prediction*), a notable progress has been evidenced for both problems with the use of deep learning algorithms. For the contact prediction category, the best methods in CASP13 achieved an average precision of 70%. In the present work, the performance of these methods is analyzed using a larger data set, with 483 proteins from four families according to the structural classification of the SCOP database (*Structural Classification of Proteins*). The selected methods were evaluated using the CASP metrics, and their results indicate an average contact prediction precision greater than 90%. SPOT-Contact was the method with the best overall performance, and one of the methods with the best performance for each SCOP class. The set of proteins used for the experiments and the implementations made for the analysis are publicly available.

Keywords: protein contact prediction, protein structure prediction, deep learning, CASP, SCOP.

1 Introduction

Proteins are one of the most biologically important macromolecules and have a wide variety of functions. Since the function of a protein is closely related to its structure [1], one of the most relevant open problems in Computational Biology is the prediction of the three-dimensional structure of a protein given its amino acid sequence using computational methods [2].

The protein structure prediction problem is divided into different subproblems, one of which is the *residue-residue*¹ *contact prediction* [3, 4]. The atoms of the residues in contact are considered to have direct interactions within the protein; and two residues are defined to be in *contact* if the Euclidean distance between their C_β atoms (C_α in the case of glycine) is less than 8 Å (angstroms) [5]. This can be seen in Figure 1.

The input for this subproblem is the sequence of L residues of the protein, and the output is a symmetric $L \times L$ matrix called *contact map*, which represents the contacts between all their pairs of residues. An element i, j of the contact map is equal to 1 if there is a contact between the corresponding residues i and j , and 0 otherwise. However, it is common to present the results in the form of a *contact list*, where each line

¹residue of amino acid.

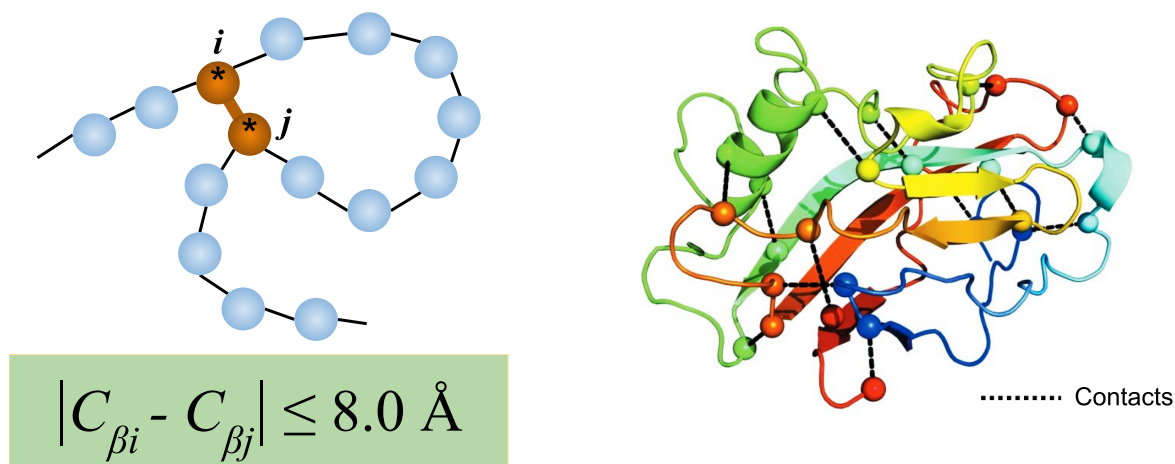


Figure 1: Contacts in a protein [6]. The left figure shows the formal definition of a contact. The right figure shows the contacts between residues in proteins

indicates the pair of residues i, j possibly in contact, associating a probability of occurrence to each predicted contact. To adapt this to a binary contact, a threshold probability of 0.5 is usually considered [5].

The contact map of a protein is a two-dimensional representation of its three-dimensional structure; therefore, it defines the global topology of the structure of such protein. The contact map includes information that can be used as distance restrictions to guide the search process for the protein structure prediction [7], especially in those that lack homologous templates in the PDB (Protein Data Bank) [8]. The most relevant contacts² for this process are the *long-range* ones, which exist between residues separated by at least 24 positions in the protein sequence [5]. Figure 2 shows an example of a contact map based on a protein structure.

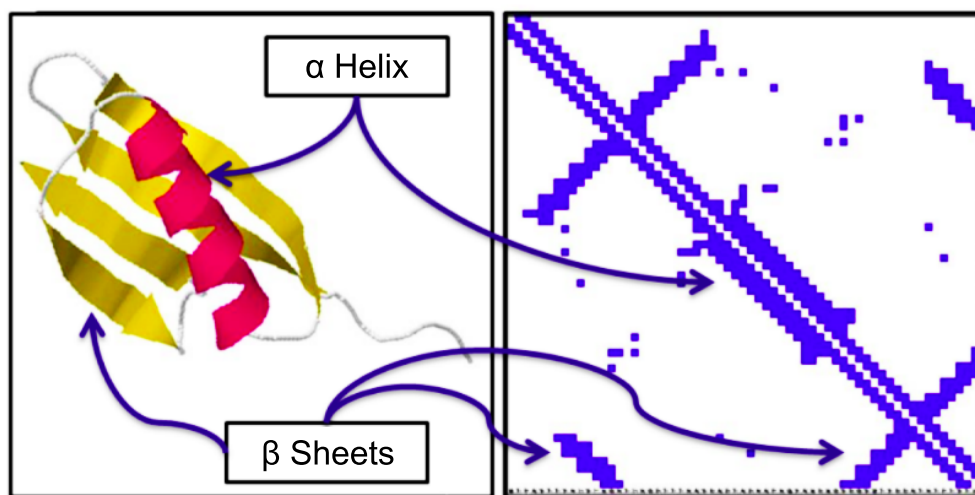


Figure 2: **Left:** A 3D rendered view of a protein. **Right:** The corresponding contact map with the contact threshold of 8 Å [9]

The CASP³ (*Critical Assessment of Protein Structure Prediction*) is a scientific community event in which various research groups contribute to determine and advance the state of the art for the protein structure prediction problem. It has been held every two years since 1994; and in each CASP, participants have a limited period to submit models for a set of proteins whose experimental structures are not yet public. Once this period ends, the independent evaluators compare them with the 3D structures obtained experimentally (by X-ray crystallography, nuclear magnetic resonance, among others). The results and evaluations are

²A detailed discussion on this point can be found in this CASP forum: <http://predictioncenter.org/forcasp/viewtopic.php?f=30&t=6481&sid=a21b38a96099a26c9b451e0efa48f16e>

³<http://predictioncenter.org/>

published in a special issue of the journal PROTEINS⁴.

The most important performance metric for protein contact prediction is *precision*, which basically indicates the success rate of the contacts predicted by the evaluated method. It has been shown that contact predictions must have a minimum precision of 22% to have a positive effect on the *ab initio* protein structure prediction [10]. In the 13th edition of CASP (CASP13) carried out in December 2018, a set of 32 test proteins was considered; and among the 46 participants, the best method obtained an average precision of 70%⁵ [11]. This result can be considered quite good; however, there is still a significant difference with the optimal precision. In the most recent edition of the CASP (CASP14), whose results⁶ were released at the end of November 2020, the best method for protein contact prediction has not exceeded the precision achieved in CASP13.

Contribution: In this work, we measured and compared the performance of the best state-of-the-art methods that participated in the protein contact prediction category of CASP13, using a larger input data set (483 proteins), which also considers four types of proteins according to the SCOP classification [12, 13]. This analysis allowed us to determine the difference of performance between the data sets, the variation in the ranking of best methods, and their behavior for different protein classes.

1.1 Current approaches for protein contact prediction

A very important element in the state-of-the-art methods for contact prediction is the MSA (*Multiple Sequence Alignment*); consisting of an alignment of three or more biological sequences (usually proteins, DNA, or RNA). It is often used to assess the conservation of protein domains, tertiary and secondary structures, and even individual amino acids or nucleotides. State-of-the-art methods for protein contact prediction use the information provided by the MSA on mutations of homologous sequences [7], where the pairs of residues in contact occur in a joint pattern in the course of evolution [8].

Current methods for predicting protein contact maps can be classified into two distinct groups: Evolutionary Coupling Analysis (ECA) and Machine Learning (ML) techniques. The ECA methods use MSA [14] to identify the correlation in the change (co-evolution) of pairs of residues, assuming that residues in close proximity mutate synchronously with the functional and structural evolutionary requirements of a protein; that is, if one residue mutates, other neighboring residues in the structural space need to mutate accordingly to maintain the structure and biological function of the protein [7]. Popular ECA methods include: CCMPred [15], FreeContact [16], GREMLIN [17], PlmDCA [18], and PSICOV [19]. While these methods are useful for predicting long-range contacts in proteins with a high number of sequence homologs, their precision is poor if the number of homologs is low [20].

The second approach is through machine learning, which trains on the contact maps of known structures on sequence-based features such as sequence profile, solvent accessibility, secondary structure, residue type, and residue separation [21, 22, 23]. These methods have been successful because of their ability to learn, given a labeled data set, the underlying relationships present in sequence-based features; and they have been especially effective in predicting proteins with few homologues. Machine learning methods include support vector machines (SVM) [24, 21, 25] and deep artificial neural network (DNN) [26, 27] in various forms. For example, in ResPRE [8], DNCON2 [28], MapPred [7], DeepCov [29], RaptorX-Contact, [30], DMP [31], SPOT-Contact [3], among others [32, 21].

Furthermore, there are the so-called complementary methods, which can be combined in the form of metapredictors, where a single network combines the outputs of several other classifiers. Examples of this architecture are MetaPSICOV [33] and NeBcon [10, 3].

The remainder of this paper is organized as follows. Section 2 describes the selected methods for protein contact prediction, the data set used for the experiments, the use mode and execution details of the selected methods, the implemented evaluation framework, and the metrics for contact prediction. Section 3 presents the analysis of the results from our experiments, as well as the comparison with the ones of CASP13. Finally, some concluding remarks are given in Section 4.

2 Materials and Methods

2.1 Protein contact prediction algorithms

As a first step, we made a selection of the 10 groups with the best performance⁷ in the CASP13 contact prediction category, which are the following: TripletRes, RaptorX-Contact, TripletRes_AT, ResTriplet, DMP,

⁴<https://onlinelibrary.wiley.com/toc/10970134/2019/87/12>

⁵http://predictioncenter.org/casp13/rrc_avrg_results.cgi

⁶https://predictioncenter.org/casp14/rrc_avrg_results.cgi

⁷http://predictioncenter.org/casp13/zscores_rrc.cgi

Table 1: Protein contact predictions methods selected for this work. The names of CASP13 groups/methods can be found at: https://predictioncenter.org/casp13/doc/CASP13_Abstracts.pdf

Method	Group name in CASP13	Availability
RaptorX-Contact	RaptorX-Contact	Web server
MapPred	Yang-Server, RRMD-plus	Web server
SPOT-Contact	ZHOU-Contact	Web server
ResPRE	TripletRes, ResTriplet	Standalone code, Web server
DMP (DeepMetaPSICOV)	DMP	Standalone code
NeBcon	Zhang_Contact	Standalone code

Zhang.Contact, ZHOU-Contact, ResTriplet_AT, Yang-Server and RRMD-plus. Subsequently, considering the public availability of their methods (either in a web server or independent code), the list was reduced to six methods, which are listed at Table 1.

The NeBcon2 server allows the reception of protein sequences and prediction of contacts in a sequential manner (a list of amino acid sequences is provided) and does not allow the execution of protein lists in parallel; furthermore, the respective standalone package code has not yet been published. For this reason, we decided to use the code of the independent package NeBcon (named NN-BAYES in the contact prediction category at CASP12⁸, and also from the Zhang_Contact group) for the execution of the contact prediction tests.

The TripletRes and ResTriplet servers perform the predictions sequentially and do not allow the execution of protein lists in parallel. For this reason, and based on the author’s recommendation, we used the independent package code of the ResPRE method for the execution of the contact prediction tests. This method is also available on a server.

DMP does not have a server for predictions, but its implementation is publicly available at the GitHub repository. Therefore, we used that implementation for testing with this method.

RaptorX-Contact does not provide an independent package of its implementation, but it is available on a web server, which was employed for our tests.

RRMD-plus, RRMD and Yang-Server share the same deep model in CASP13; hence, and based on the recommendation of the author of MapPred, we decided to use MapPred in representation of the Yang-Server and RRMD-plus methods. A web server is provided for this method.

The independent package code of SPOT-Contact is not publicly available, but this method has a web server that was used for the prediction tests performed in our work.

2.2 Selected methods for protein contact prediction

In this subsection we briefly describe the general characteristics and training techniques used by each of the state-of-the-art methods for predicting contacts in proteins selected for this work. It should be noted that the algorithms of the selected methods were previously trained by their respective authors, so we proceeded directly to the performance evaluation of these methods with the test set proposed in this work.

2.2.1 DMP

DeepMetaPSICOV (DMP)⁹ [31] is a deep learning-based contact prediction tool that uses the sequence profile, predicted secondary structure, and solvent accessibility, among others, as input. The model of this method consists of a completely convolutional *residual neural network* (ResNet), which is known to have high performance in tasks such as image recognition [34] and contact prediction [30] [35].

Each target sequence is first used as input in the MSA alignment generation and contact prediction steps to generate an initial contact list. Using HHblits [36], the target sequence is scanned against a database of 70 proteins. Regions of the sequence that do not match a PDB template and are at least 30 residues long are removed, and the alignment generation and contact prediction steps are performed again on the remaining domain sequence. The predicted contact scores for such domains are copied back to the relevant regions of the initial contact list to produce the final prediction.

⁸The names of CASP12 groups/methods can be found at: https://predictioncenter.org/casp12/doc/CASP12_Abstracts.pdf

⁹DMP: <https://github.com/psipred/DeepMetaPSICOV>

2.2.2 RESPRE

This method¹⁰ [8] uses an inverse covariance matrix (or precision matrix) of multiple sequence alignments (MSA) through deep convolutional residual neural network training. ResPRE consists of three steps: MSA generation, precision matrix-based feature collection, and deep ResNet training.

First, a precision matrix estimator is used to evaluate the conditional relationships between different types of residues at different positions derived from the MSA. The potentials at each pair of positions are used as training features, which are combined with deep fully ResNets [34] for the final modeling of the contact map.

The main advantage of ResPRE lies in the use of the precision matrix that helps to rule out transition noises from contact maps.

2.2.3 RAPTORX-CONTACT

This method¹¹ [30] predicts contacts by integrating evolutionary coupling (EC) and sequence conservation information, through an ultra-deep neural network made up of two deep ResNets, each one being a module of the method.

The first module performs a series of one-dimensional (1D) convolutional transformations of sequential features (sequence profile, predicted secondary structure, and solvent accessibility). The output of this 1D convolutional network is converted to a two-dimensional (2D) matrix by external concatenation, and then fed to the second module along with pairwise features (i.e., coevolution information, pairwise contact, and distance potential). The second module is a 2D residual network that performs a series of 2D convolutional transformations on its input. Finally, the output of the 2D convolutional network is entered into a logistic regression, which predicts the probability that any two residues form a contact. Furthermore, each convolutional layer is also preceded by a simple nonlinear transformation called a rectified linear unit [37].

Mathematically, the output of the residual 1D network is just a 2D matrix with dimension $L \times m$, where m is the number of new features (or hidden neurons) generated by the last convolutional layer of the net. Biologically, this 1D residual network learns the sequential context of a residue. By stacking multiple layers of convolution, the network can learn information in a very large sequential context. The output of a 2D convolutional layer has a dimension $L \times L \times n$, where n is the number of new features (or hidden neurons) generated by this layer for a pair of residues. The 2D residual network primarily learns patterns of contact or correlation occurrence from high-order residues (i.e. the 2D context of a pair of residues). The number of hidden neurons can vary in each layer.

2.2.4 SPOT-CONTACT

SPOT-Contact (*Sequence-based Prediction Online Tools for Contact map prediction*)¹² [3] is a web application that enables the prediction of contacts of multiple protein sequences (up to 100 for a single submission). This method adopts a deep hybrid network: prepared inputs are fed into a ResNet model, and the outputs are subsequently processed by a 2D Bidirectional-ResLSTM model (*Residual Long Short-Term Memory*) (2D-BRLSTM) [38]. The base model is divided into four separate segments: Input Preparation, ResNet, 2D-BRLSTM, and *Fully-Connected* (FC).

Inputs to the SPOT-Contact model include one-dimensional (that is, of the primary sequence) and two-dimensional (that is, pairs of residues) features. The one-dimensional features consist of the Position Specific Scoring Matrix (PSSM) profile, the HHblits [36] HMM (*Hidden Markov Models*) profile, and various structural probabilities predicted by SPIDER3 [39]. The 2D features consist of the output from CCMPred [15] and two outputs (direct and mutual coupling information) from DCA [40], resulting in three pairwise features for concatenation with the output of the first section of the network.

2.2.5 MAPRED

This method¹³ [7] for predicting contacts and distances between protein residues consists of two component methods: DeepMSA and DeepMeta, both trained with ResNets. For each sequence of residues, it constructs a multiple sequence alignment (MSA). ResNets are used as a driving force for training and prediction, with covariance features derived from the MSA.

¹⁰ResPRE: <https://zhanglab.ccmb.med.umich.edu/ResPRE/>

¹¹RaptorX-Contact: <http://raptorx.uchicago.edu/ContactMap/>

¹²SPOT-Contact: <http://sparks-lab.org/jack/server/SPOT-Contact/>

¹³MapPred: <http://yanglab.nankai.edu.cn/mappred/>

The success of MapPred is attributed to three factors: the robustness of the MSA from metagenome sequence data, the improved feature design with DeepMSA, and the optimized training with ResNets. The output of the method includes the predicted contact map, distance maps, and distance distribution.

2.2.6 NEBCON

NeBcon (*Neural-network and Bayes-classifier based contact prediction*)¹⁴ [10] consists of a pipeline that uses the *Naive Bayes classifier* (NBC) theorem to combine eight contact prediction methods that are built from co-evolution and machine learning approaches. The posterior probabilities of the NBC model are trained with intrinsic structural features through neural network learning for the final prediction of the contact map.

NeBcon consists of two steps. The query sequence is first fed into a set of eight representative (contact map) predictors, including three machine learning-based methods, three coevolution-based methods, and two meta-server-based methods. A set of posterior probability scores is then calculated from the eight predictors using the NBC. In the second step, six inherent structural features are extracted from the query sequence, which are trained along with the NBC probabilities using the neural network to generate the final contact map.

This is the only considered method not based on deep learning.

2.3 Protein data set for the experiments

To evaluate the performance of the six CASP13 methods selected for this work, a data set of 721 proteins was initially considered; where each sequence is made up of natural and synthetic residues, and can be complete or with missing residues. The proteins that are part of this set were determined by X-ray crystallography, they have a single chain with a length between 40 to 400 residues, a maximum R factor of 20%, a resolution better than 2 Å and a maximum sequence identity of 25%.

Furthermore, the considered proteins have a single domain under the SCOP 1.75 classification. SCOP (*Structural Classification Of Proteins* [12]) is a database for the structural classification of proteins, and classifies protein domains according to their evolutionary and structural relationship. SCOP provides classification of almost all super-families and families with representatives in the PDB (*Protein Data Bank*), where the most important classes are the following:

- A (All- α): proteins whose secondary structures are essentially composed of α -helices.
- B (All- β): proteins whose secondary structures are essentially composed of β -sheets.
- C (α/β): proteins whose secondary structures are composed of α -helices and β -sheets (not separable).
- D ($\alpha+\beta$): proteins whose secondary structures are composed of α -helices and β -sheets (separable).
- Multi-Domain: proteins with different fold domains and for which no homologues are currently known. Other classes have been assigned for peptides, small proteins, theoretical models, nucleic acids, and carbohydrates [13].

For this research work we performed the tests with single domain proteins; therefore, we did not use the SCOP 1.75 Multi-Domain classification.

Next, a filter was applied to the set of 721 proteins, conserving only those whose sequences contain exclusively natural residues; thus leaving a total of 555 proteins. Then another filter was applied, eliminating the proteins with a percentage of missing residues greater than 10%; thus resulting in a data set of 486 proteins. In the next step, a protein was removed from the set (with PDBID 2G7O) because it did not have long-range contacts in its respective residue sequence. Finally, two more proteins were eliminated (PDBID 2IW1 and 1Q0R) due to prediction errors obtained with the RaptorX-Contact server. Consequently, a total of **483 proteins** were considered as a data set of input cases for our experiments and the subsequent analysis. The list with the PDB identifiers of the 483 proteins can be found inside the .zip file at the following link: <https://doi.org/10.6084/m9.figshare.12150318>.

2.4 Use mode and response time of predictors

MapPred allows the contact prediction in parallel for up to 10 protein sequences; and for our tests, the author allowed parallel predictions in batches of 600 proteins. RaptorX-Contact admits parallel prediction for up to 20 protein sequences in a batch, allowing multiple batches that are processed sequentially; accepting a total

¹⁴NeBcon: <https://zhanglab.ccmb.med.umich.edu/NeBcon/>

of 500 protein sequences at a time. SPOT-Contact allows sequential predictions of up to three proteins in a batch, allowing batches to be processed in parallel.

The MapPred, RaptorX-Contact and SPOT-Contact servers showed an average prediction response time of 10, 40 and 20 minutes, respectively. Depending on the protein size and the number of available servers, the response time could vary up to a few hours for prediction.

The local execution of the NeBcon, ResPRE and DMP methods yielded an average prediction response time of 60, 5 and 30 minutes, respectively. Given that NeBcon combine eight contact prediction methods for its own predictions, it is not surprising that it has a higher average execution time than the other two methods based on deep learning.

However, the execution time was not considered as a performance measure, since the executions of the different methods were carried out in different computational environments.

2.5 Execution environment for protein contact prediction algorithms

The specifications of the local environment used are as follows: CPU Intel Core i7 7700HQ, GPU Nvidia GeForce GTX 1050Ti 4 GB, SO Ubuntu 16.04.6 LTS, RAM 16 GB, Python 3.6.9, Miniconda 4.7.11, Numba 0.35, Pytorch 0.4.0, CCMpred v0.1.0, FreeContact 1.0.21, Legacy BLAST 2.2.26. The dependencies included within the NeBcon package are: SVMSEQ, PSICOV, CCMpred, Freecontact, STRUCTCH, MetaPSICOV, libgsl.so.0 library.

The databases used were: UniClust 30_2018_08 and HH-suite 3 for ResPRE 1.0; Uniprot20, Uniref90 and nr for NeBcon 1.0; UniClust 30 for DMP 1.1.0.

The specifications of the server used are as follows: SO Ubuntu 16.04.5 LTS, CPU 2x Intel Xeon E52640 10-core 2.4 Ghz., GPU Nvidia Tesla K40, RAM 128 GB.

2.6 Performance evaluation framework for protein contact prediction methods in CASP13

A method for protein contact prediction can be evaluated *in silico*, comparing its list of *predicted contacts* for a protein with another one of *real contacts*, which can be generated from the experimental structure data of such protein in the PDB.

Therefore, the performance analysis in this work consists of two stages of data processing that are represented by the flow diagram in Figure 3. The first stage (left section of the diagram in Figure 3) processes the *real contact map*, obtained from the input PDB files corresponding to the selected data set of proteins; and the second stage (right section of the diagram in Figure 3) processes the contact maps resulting from each evaluated contact prediction algorithm. The framework implementation is available at the following link: <https://doi.org/10.6084/m9.figshare.12150318>.

2.6.1 ProCMAP-R - Real Contact Map for Proteins

For the first stage (corresponding to the left section of the diagram in Figure 3), a tool to calculate the list or map of real contacts in a protein is necessary. As we did not find a publicly available tool that could calculate the real contacts for a set of N proteins at the same time, in this work we propose **ProCMAP-R**: a calculator and viewer of real protein contact maps developed with BioPython, Python 3 and Matplotlib. This tool allows to obtain the real contact maps of N proteins giving their respective PDB files as input.

ProCMAP-R uses the *contact* definition of CASP, which considers that two residues are in contact if the 3D euclidean distance between their C_β atoms (C_α for GLY) is less than the threshold value of 8 \AA . Therefore, the *contact map* for a protein sequence with N residues is a binary symmetric matrix $N \times N$. It should be noted that the maximum number of contacts for a sequence of N residuals is equal to: $\frac{N \times (N-1)}{2}$.

The ProCMAP-R implementation can be found at the following link: <https://doi.org/10.6084/m9.figshare.12150471>.

2.6.2 Residues identifier mapping

Given the real contact maps, we perform the *mapping* of residue identifiers obtained from the PDB file with each pair of contacts $C_{i,j}$ of the real contact map.

The mapping consisted of locating residues, both from the structure and from the list of missing residues, and mapping them according to the position in the list of all residues, contained in the SEQRES of the PDB of each evaluated protein. We then filter the long-range contacts from the real contact maps.

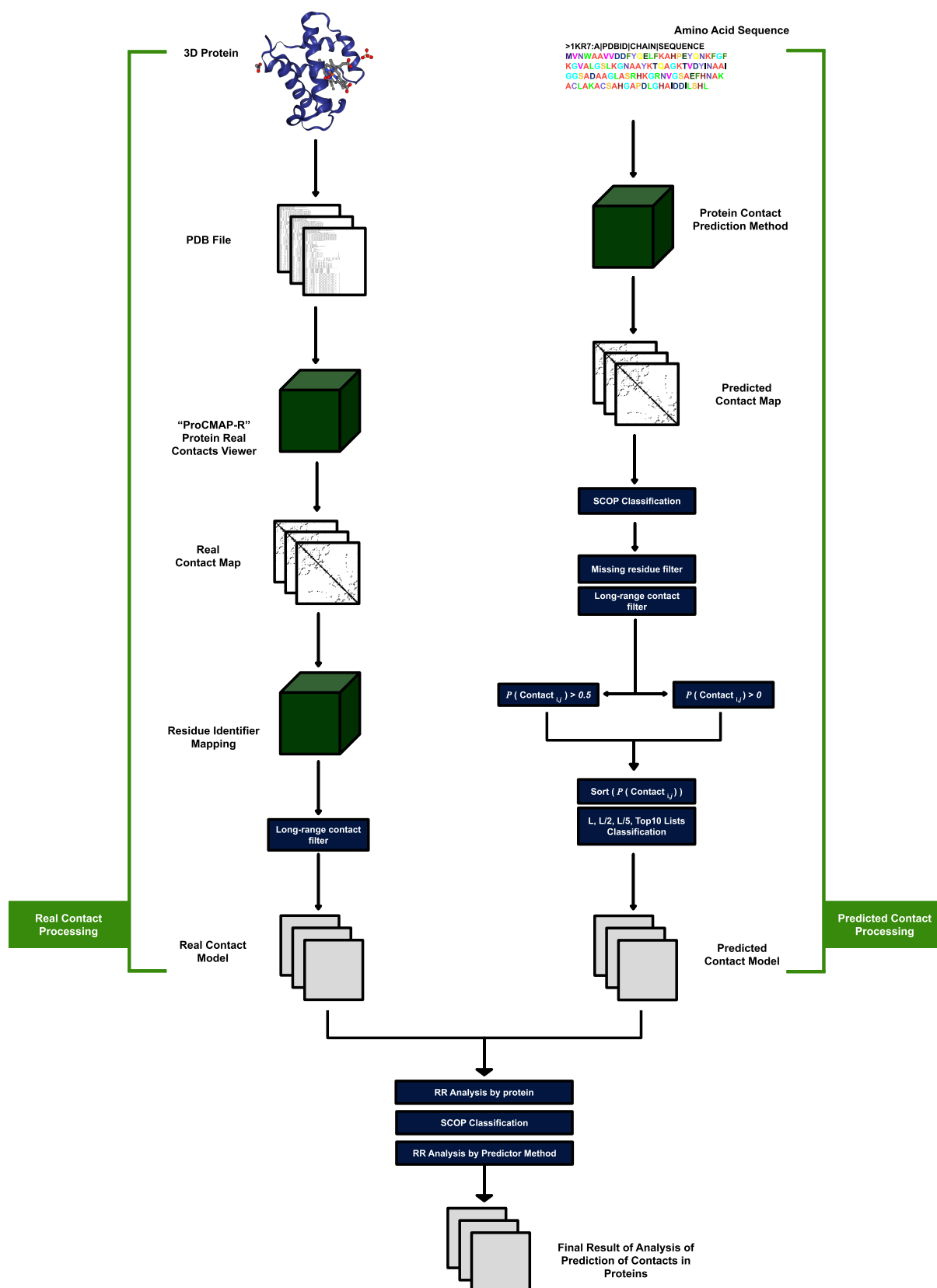


Figure 3: Framework proposed in this work for performance analysis of protein contact prediction methods. This figure is available at: <https://doi.org/10.6084/m9.figshare.12212903>

2.6.3 Elimination of missing residues in contact predictions

The second stage of the analysis (corresponding to the right section of the diagram in Figure 3) consisted of processing the contact lists generated by each predictor method for each input case. Thus, for each output generated, the contact predictions considering missing residues in the corresponding structure were eliminated, and then the long-range contacts were filtered.

Depending on the specific application, users may have different requirements about the performance of a protein contact prediction algorithm. For example, some might be more interested in a small number of long-range precise contacts; while others strive for larger lists of medium and long-range contacts that will tolerate a greater number of false positives within the prediction [5]. As mentioned above, our evaluation concentrates on long-range contacts due to their relevance in predicting protein structures. In addition, it is used in the evaluation of CASP contact predictions.

2.6.4 Classification of predictions by contact probabilities

The analysis of the resulting contact map predictions was divided into two groups: (i) one that evaluates contacts with *contact probability* greater than 0; and (ii) another that evaluates contacts with *contact probability* greater than 0.5, which allow higher prediction quality for each algorithm.

Subsequently, for each group, we ordered the contacts in each list from highest to lowest contact probability. Next, we generated *reduced lists* with the first L , $L/2$, $L/5$ and $Top10^{15}$ contacts; where L is the length of the protein sequence. This is usually done in CASP to evaluate the participating methods [4, 5].

2.6.5 Contact prediction performance metrics for a single protein

The performance metrics, which are obtained for each protein and for each predictor, are: (i) Precision, (ii) Sensitivity, and (iii) F1 score. The *precision* (P) represents the percentage of relevant results; that is, the percentage of correctly predicted contacts out of all contacts predicted as such. The *sensitivity* (S) is defined as the total percentage of relevant results (contacts) correctly classified by the evaluated algorithm. Finally, $F1$ indicates the harmonic average between precision and sensitivity. These values are calculated as follows:

$$P = TP / (TP + FP) \quad (1)$$

$$S = TP / (TP + FN) \quad (2)$$

$$F1 = 2 \times (P \times S) / (P + S) \quad (3)$$

Where TP indicates the number of correctly predicted contacts (true positives), FP is the number of incorrectly predicted contacts (false positives), and is FN the number of unpredicted real contacts (false negatives).

Due to the fact that these measures are highly correlated in the *reduced lists* of contacts, we focus the analysis on the *precision* because it is the most intuitive measure and, considering the application of contact prediction to the protein structure prediction, it is important to keep FP as low as possible. However, it should be noted that precision is not an appropriate measure for the entire contact list, as it does not provide information on the fraction of true contacts that has been predicted. Therefore, a more appropriate measure in this case is the F1 metric, which takes into account the precision of the predicted contacts and the fraction of the set of real contacts that was predicted [5]. Since the percentage of contacts in a protein is very small (<2%), the *protein contact prediction* problem is extremely unbalanced [30]. For this reason, *specificity* is not considered as a performance metric in this work.

2.6.6 Contact prediction performance metrics for a protein set

For each predictor method and considering the $L/5$ reduced lists, we calculated the average values of precision, sensitivity and F1 score. According to previous works [5, 4], the $L/5$ list is one of the most important for the structure prediction process. For this reason, it is used for the evaluation of the participating methods of the CASP.

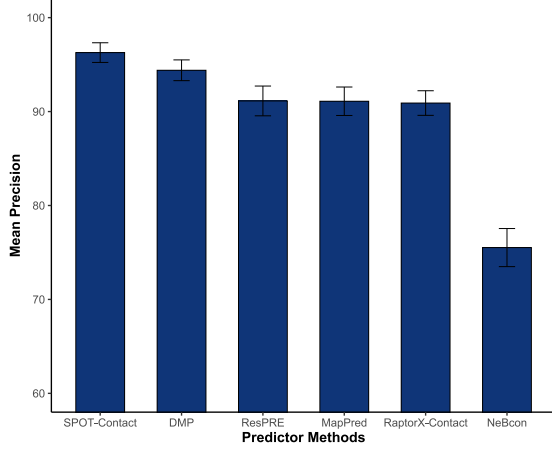
The performance of each predictor method for each SCOP class was evaluated using:

$$\bar{P}_{jk} = \frac{\sum_{i \rightarrow K(i)=k} P_{ij}}{N_k} \quad (4)$$

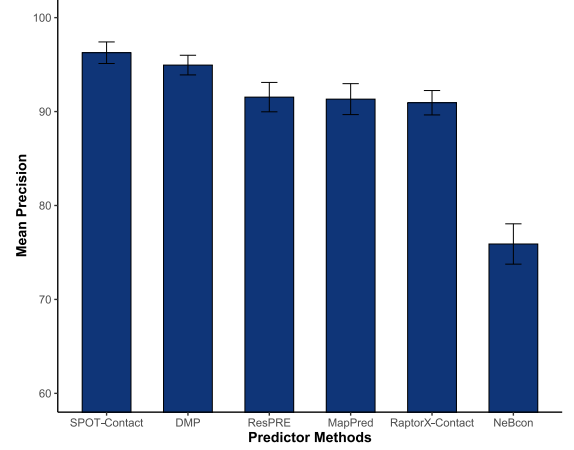
¹⁵the 10 contacts with highest probability in the contact list.

Table 2: General performance of protein contact prediction methods (considering the entire data set)

Methods	P(Contact) > 0			P(Contact) > 0.5		
	F1	Precision	Sensitivity	F1	Precision	Sensitivity
MapPred	24.84	91.10	14.85	24.17	91.32	14.31
DMP	26.15	94.40	15.79	25.71	94.95	15.28
SPOT-Contact	26.68	96.28	16.07	26.33	96.27	15.68
ResPRE	24.75	91.13	14.77	24.57	91.54	14.52
NeBcon	20.28	75.52	12.05	19.26	75.90	11.32
RaptorX-Contact	25.05	90.91	15.07	25.03	90.94	15.01



(a) Contact probability greater than 0



(b) Contact probability greater than 0.5

Figure 4: General average precision for protein contact prediction of the evaluated methods (considering the entire data set)

$$\bar{S}_{jk} = \frac{\sum_{i \rightarrow K(i)=k} S_{ij}}{N_k} \quad (5)$$

$$\bar{F1}_{jk} = \frac{\sum_{i \rightarrow K(i)=k} F1_{ij}}{N_k} \quad (6)$$

Where j indicates the predictor method, k the SCOP classification, i the evaluated protein, N the total number of proteins, N_k the total number of proteins with class k according to SCOP, and $K(i)$ is a function that returns the class k to which the protein i belongs. To calculate the general metrics (considering the entire data set), it can be assumed that all the proteins in the set belong to the same class k_0 .

3 Results and Discussion

3.1 Comparison of contact prediction metrics among CASP12 and CASP13 methods

Table 2 shows the results for each of the six selected contact prediction methods from CASP12 and CASP13, using the data set of 483 proteins.

We analyzed the performance of the predictors in reduced lists of size $L/5$, considering only long-range contacts. For the performance evaluation of each predictive method, we used the precision as a comparative metric, for the reasons explained in subsection 2.6.5. In addition, the sensitivity may decrease when considering reduced lists, since the number of false negatives could increase (due to the initial predicted contacts that could have been eliminated).

In addition, the statistical tests *ANOVA* and *permutation t-test* (as *Post Hoc*) were performed [41, 42] with the precisions obtained by the evaluated methods, considering in both tests a *significance level* of 0.05. Statistical tests were applied on the precisions for the complete protein data set (general precisions), as well as on the precisions for the protein sets classified according to SCOP 1.75 (precisions by classes: A, B, C and D). All of the above was performed for the evaluations that followed two rules for considering contacts

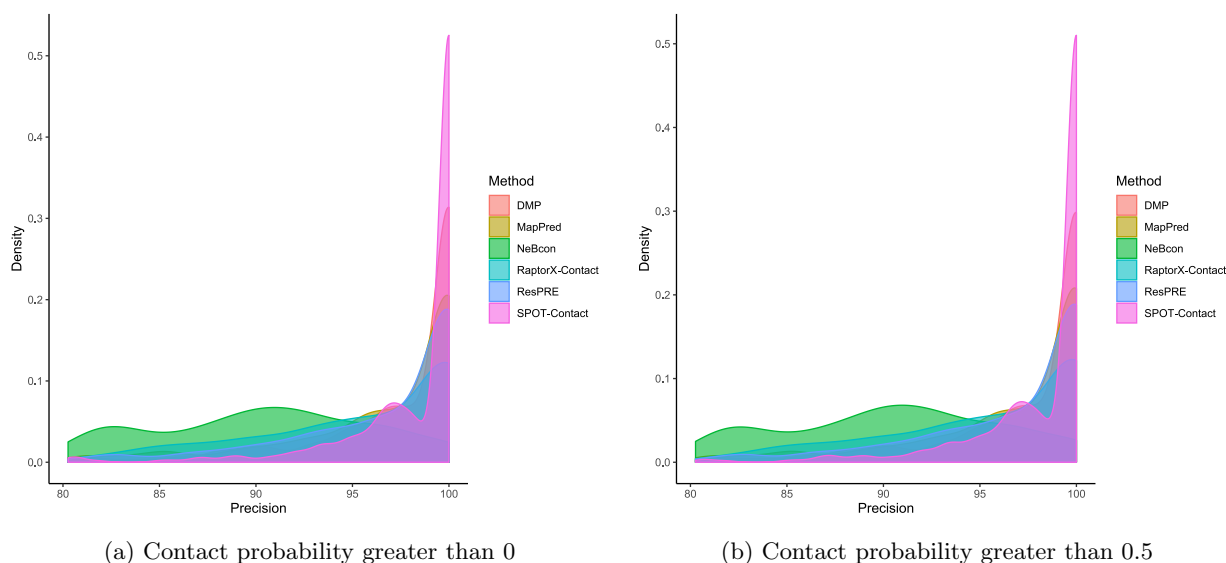


Figure 5: Precision distribution for each evaluated method considering the complete protein data set

according to their probabilities ($P(C)$): (i) greater than 0 and (ii) greater than 0.5. Figure 4 illustrates the relative performance of the six considered methods for these two rules; and Figure 5 shows the precision distribution for each evaluated protein contact prediction method, considering the complete data set.

These statistical tests were performed in order to determine if the differences between the precisions obtained by the methods considered are statistically significant (*ANOVA* test); and to determine which is the best method (*Post Hoc* test).

3.1.1 Performance of evaluated methods for contacts with probability greater than 0

Figure 4(a) shows the obtained precision results. SPOT-Contact (96,28%) is the method with the highest precision in contact prediction; slightly improving the precision of DMP (94.40%) and significantly the precision of NeBcon (75.52%). The methods ResPRE, MapPred and RaptorX Contact have an average precision of 91.13%, 91.10 % 90.91%, respectively.

The difference in precision results between SPOT-Contact and DMP is statistically significant; therefore, we conclude that SPOT-Contact is the method with the best performance in predicting contacts for the complete protein data set, considering probabilities of contacts greater than 0.

3.1.2 Performance of evaluated methods for contacts with probability greater than 0.5

In this case (see Figure 4(b)), the precisions obtained by SPOT-Contact (96,27%) and DMP (94,95%) were the highest, slightly exceeding the precisions of ResPRE (91,54%) and MapPred (91,32%), and widely to the one of NeBcon (75.90%).

The precisions reached by both SPOT-Contact and DMP do not present a significant difference between them, and for this reason it is concluded that SPOT-Contact and DMP are the methods with the best performance in predicting contacts for the complete protein data set, considering probabilities of contacts greater than 0.5.

3.2 Comparison among CASP12 and CASP13 methods according to the SCOP classification of proteins

Tables 3 and 4 summarize the contact prediction performance of the six evaluated methods from CASP12 and CASP13, considering the SCOP classification of proteins in the data set and the contact predictions with probabilities greater than 0 and greater than 0.5, respectively. We can also visualize, in Figures 6, 7, 8 and 9; the resulting precision plots for each SCOP class for each evaluated method.

3.2.1 Performance of prediction methods for contacts with probability greater than 0

For the protein sets corresponding to SCOP classes A, B, C and D, SPOT-Contact achieved precisions of 90.94%, 97.08%, 98.97% and 95.75%, respectively; which are slightly higher than the precisions obtained with DMP (87.86%, 95.75%, 97.34% and 93.87%, respectively) and considerably higher than the precisions

Table 3: Performance of protein contact prediction methods according to SCOP classification ($P(C)>0$)

Methods	A Class			B Class			C Class			D Class		
	F1	P	S	F1	P	S	F1	P	S	F1	P	S
MapPred	31.27	79.84	20.40	23.12	93.07	13.58	21.93	96.78	12.41	25.78	89.68	15.33
DMP	35.00	87.86	23.05	23.94	95.75	14.34	22.07	97.34	12.48	27.39	93.87	16.39
SPOT-Contact	36.18	90.94	23.84	24.20	97.08	14.35	22.43	98.97	12.69	27.90	95.75	16.64
ResPRE	30.57	79.52	19.77	23.54	95.04	13.86	21.45	94.88	12.13	26.21	90.84	15.60
NeBcon	23.20	59.48	15.10	18.49	75.08	10.79	19.42	85.69	10.99	21.04	73.77	12.47
RaptorX-Contact	33.41	84.50	21.98	22.62	91.34	13.37	21.62	95.37	12.23	25.93	89.02	15.50

Table 4: Performance of protein contact prediction methods according to SCOP classification ($P(C)>0.5$)

Methods	A Class			B Class			C Class			D Class		
	F1	P	S	F1	P	S	F1	P	S	F1	P	S
MapPred	29.92	80.86	19.18	22.52	93.59	13.02	21.86	97.02	12.36	24.68	88.97	14.60
DMP	33.59	89.29	21.59	23.53	95.76	13.62	22.00	97.44	12.44	27.14	94.95	16.15
SPOT-Contact	34.84	90.36	22.40	24.16	98.15	14.03	22.41	99.03	12.68	27.57	94.98	16.44
ResPRE	30.12	82.38	19.06	23.34	94.95	13.50	21.45	94.89	12.13	25.98	90.39	15.46
NeBcon	20.61	60.65	13.08	17.32	75.59	9.99	19.21	85.52	10.86	20.23	74.21	11.97
RaptorX-Contact	33.23	84.44	21.69	22.75	91.47	13.37	21.62	95.37	12.23	25.87	89.07	15.45

of NeBcon (59.48%, 75.08%, 85.70% and 73.77%, respectively). Thus, SPOT-Contact and DMP were the methods with the best performance in predicting contacts for SCOP classes A, B, C, and D, with contact probability greater than 0; considering as *best performance* the highest average precision. Each of the six methods achieved its highest average precision in classes B and C.

Comparing the precision distributions of SPOT-Contact and DMP for the C class, a statistically significant difference was found between them; which did not happen for the other classes. Therefore, this analysis indicates that the method with the best performance for class C was SPOT-Contact; and for classes A, B and D, the methods with the best performance were SPOT-Contact and DMP.

3.2.2 Performance of prediction methods for contacts with probability greater than 0.5

For the protein sets corresponding to classes A, B, C and D, SPOT-Contact (90.36%, 98.15%, 99.03% and 94.98%, respectively) and DMP (89.29%, 95.76%, 97.44% and 94.95%, respectively) achieved again the highest precisions. The lowest precisions (60.65%, 75.59%, 85.52% and 74.21%, respectively) corresponds to the NeBcon method.

The six evaluated methods reached their highest precision for the protein sets of classes B and C. The methods with the highest precision in the four SCOP classes, SPOT-Contact and DMP, had a statistically significant difference for classes B and C, but not for classes A and D. Therefore, this analysis indicates that SPOT-Contact had the best performance in classes B and C; while the methods with the best performance for classes A and D were SPOT-Contact and DMP.

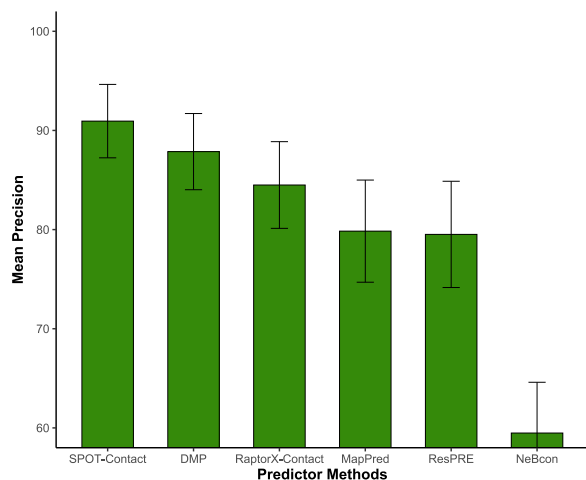
3.3 Comparison between the CASP13 Ranking and the Experimental Ranking

Using a larger and varied protein data set, we can see that the relative positions of the methods in the *Experimental Ranking* obtained in this work are not the same as in the *CASP13 Ranking* (Table 5). Considering long-range contacts, reduced lists $L/5$ and probability of contact greater than 0, the CASP13 Ranking for the contact prediction category places RaptorX-Contact in 1st place of performance, with a precision of 70%; while it obtains the 5th place in the Experimental Ranking, with a precision of 91%.

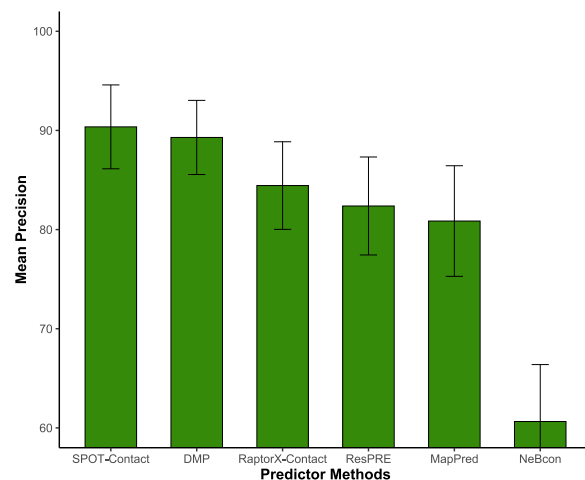
SPOT-Contact and DMP reached 1st and 2nd place in the Experimental Ranking with precisions of 96% and 95%, respectively; while in the CASP13 Ranking, DMP obtained a precision of 61% and SPOT-Contact a precision of 59%, thus remaining in 4th and 7th place, respectively.

3.4 Comparison among approaches of selected methods

It is important to note that the protein contact prediction methods based on *deep learning* algorithms (see subsection 2.2) performed markedly better compared to the one based only on other *machine learning* algorithms (NeBcon), since there was a performance difference of approximately 15% to 20% between them. On the other hand, a maximum precision difference of approximately 5% is observed between the methods based on deep learning.

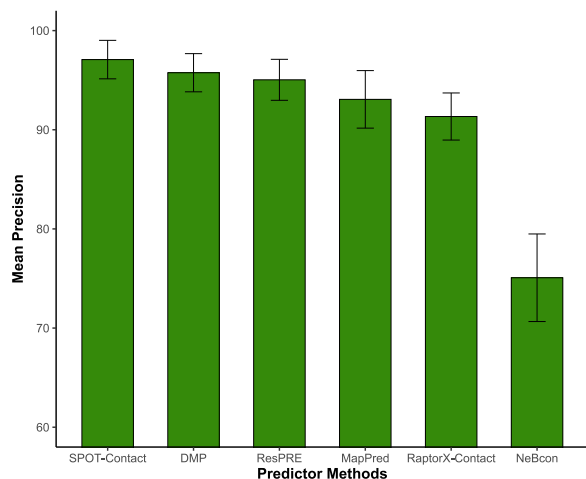


(a) Contact probability greater than 0

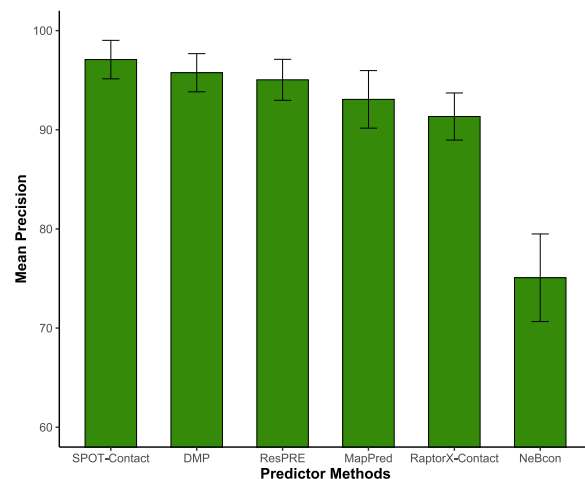


(b) Contact probability greater than 0.5

Figure 6: Precision of methods for class A. In both cases, SPOT-Contact and DMP are statistically the best methods for this class

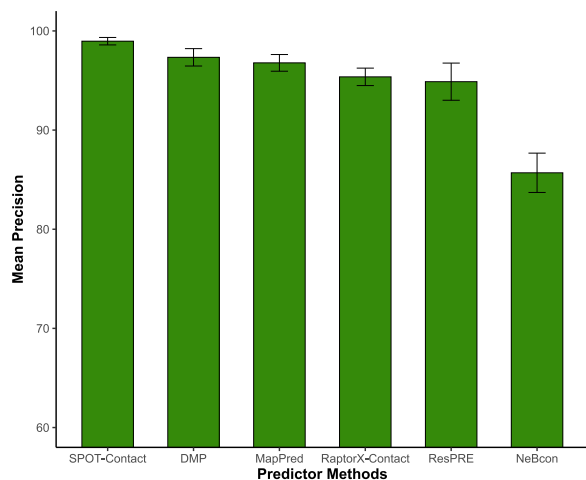


(a) Contact probability greater than 0

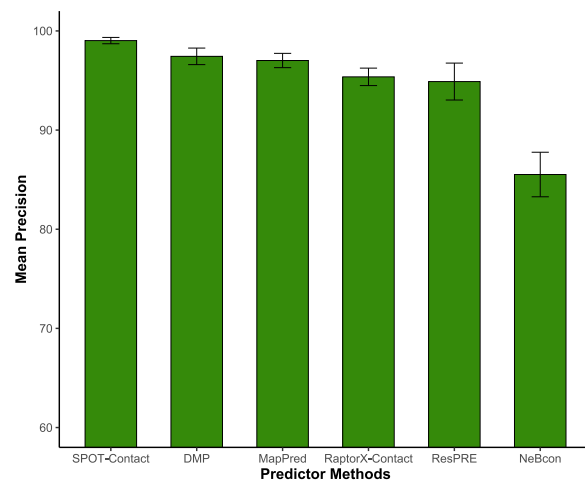


(b) Contact probability greater than 0.5

Figure 7: Precision of methods for class B. The results are very similar to the ones presented in Figure 6, so it can be concluded that SPOT-Contact and DMP are also statistically the best methods for this class



(a) Contact probability greater than 0



(b) Contact probability greater than 0.5

Figure 8: Precision of methods for class C. In both cases, SPOT-Contact is the statistically the best method for this class

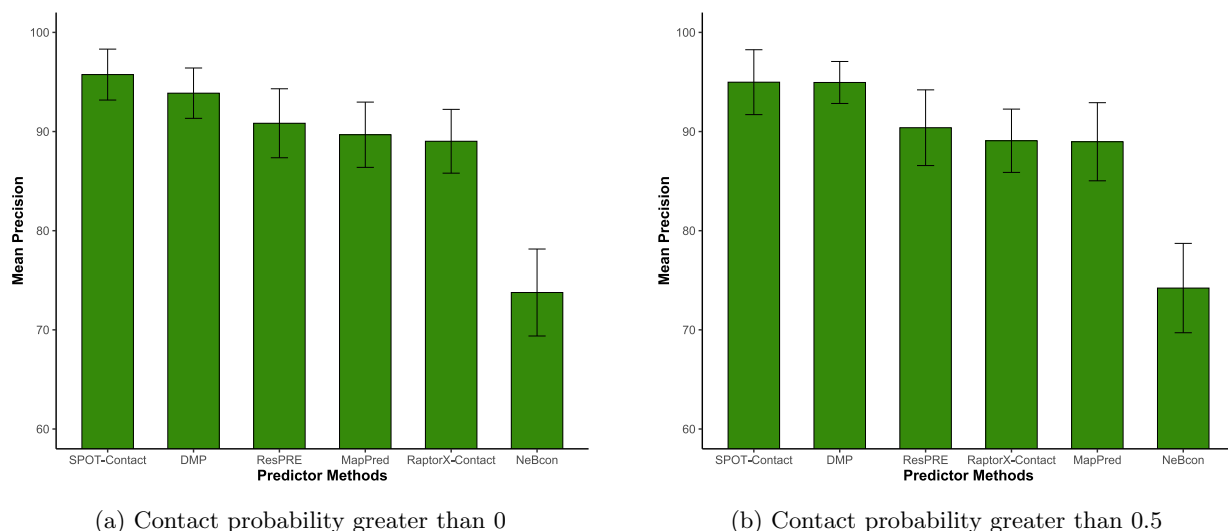


Figure 9: Precision of methods for class D. In both cases, SPOT-Contact and DMP are statistically the best methods for this class

Table 5: Comparison among protein contact prediction algorithms with $L/5$ reduced lists, long-range contacts, and $P(C)>0$

CASP13 Ranking			Experimental Ranking		
Position	Method	Precision	Position	Method	Precision
1	RaptorX-Contact	70%	1	SPOT-Contact	96%
2	TripletRes	66%	2	DMP	95%
4	DMP	61%	3	ResPRE	92%
6	MapPred	60%	4	MapPred	91%
7	SPOT-Contact	59%	5	RaptorX-Contact	91%
9	NeBcon2	57%	6	NeBcon	76%

4 Conclusions

We tested six of the top ten methods in the CASP13 contact prediction category, carried out in December 2018. A data set of 483 proteins, whose primary sequences are entirely composed by natural residues and with a limit of missing residues, were provided as input for the evaluated methods. The DMP, ResPRE and NeBcon methods were tested locally, using our own resources as well as resources provided by the Polytechnic Faculty of the National University of Asunción; while the RaptorX-Contact, MapPred and SPOT-Contact methods were tested using their corresponding web servers.

Once the prediction results of each of the selected methods had been obtained, the proteins were classified into classes A, B, C and D following the SCOP 1.75 classification, in order to group the proteins according to their structural characteristics. Although there are more classes within this classification, only the proteins within the four mentioned classes were considered because they are the most abundant in nature.

At the same time, we developed a tool for obtaining the real contact maps of the proteins from their corresponding PDB file. To assess the results obtained by the selected CASP13 methods, we also developed a tool to perform the analysis of the prediction results of each evaluated method.

The evaluation metrics used were the same as in CASP13: precision, sensitivity and the F1 measure. In order to assess the performance of a particular protein contact prediction method, we used: (i) the prediction files obtained by the method for the protein data set, where a prediction file consists of a list of residues in contact in RR (residue-residue) format with a probability of occurrence; and (ii) the output of the developed tool for obtaining the list of real contacts of each protein in the data set. Based on these data, the list of real contacts of a protein and the corresponding $L/5$ reduced list of predicted contacts were compared to

determine the precision, sensitivity and F1 measure. Then, we calculated the average for each metric; and they represent the final precision, sensitivity and F1 measure for the considered method.

The obtained performance results show a high precision achieved by the evaluated state-of-the-art methods. However, the sensitivity is low considering the reduced lists of $L/5$ size, which suggests that although contacts are correctly predicted as such, they are still few; thus ignoring several other real contacts, resulting in a high number of false negatives compared to the number of true positives.

From the results for the complete data set of 483 proteins, it can be noted that most of the selected methods obtain precisions above 90%; while the best method in the CASP13 obtained a precision of 70% in this competition. This difference may be due to the fact that the proteins in the data set are found in the PDB, and these structures (or similar ones) could be used in the learning process of such methods. Through statistical tests, it was determined that the method obtaining the best results, both for the complete set of proteins and for each selected SCOP class, was SPOT-Contact [3]; followed by DMP [31].

Considering the results for the classes according to SCOP 1.75 [12], a greater difficulty of prediction can be noticed for classes A and D; so we can infer that the presence of α -helices makes the protein contact prediction problem more difficult.

Clearly, there is still a long way to go in the area of contact prediction, but the CASP13 results suggest that the right steps have been taken by using machine learning and deep learning as a means to increase the precision of predictions. More research is needed to explore in greater depth the use of these approaches, in conjunction with other potential tools, to achieve the next major advance in contact prediction, and thus for protein structure prediction.

Acknowledgements

We thank PhDs. Jinbo Xu, Jianyi Yang, Andrew W. Senior, Qi Wu, Chengxin Zhang, Yang Li, Andriy Kryshchak, Christophe N. Magnan, and Alexey Murzin, for the comments, explanations and discussions given in the development of this research work; and biologist PhD. Pastor Enmanuel Pérez Estigarribia for the support provided in Statistics; and classmates Hyun Min (Daniel Min), Daniel Wohlgemuth, Christian Chilavert, Tadashi Akagi, and Isaac Amarilla; and Professor PhD. Cristian Cappo, for the technical software collaboration provided to this research.

We also thank the Technological Research and Development Nucleus (NIDTEC) of the Polytechnic College of the National University of Asuncion (FPUNA) for providing the computing power and necessary storage for the execution of the tests performed for the analysis of this work.

References

- [1] G. Gatto, J. Berg, L. Stryer, and J. Tymoczko, *Biochemistry*, 9th ed. W. H. Freeman, 2019, 1120 pp.
- [2] A. W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Žídek, A. W. Nelson, A. Bridgland *et al.*, “Improved protein structure prediction using potentials from deep learning,” *Nature*, vol. 577, no. 7792, pp. 706–710, 2020.
- [3] J. Hanson, K. Paliwal, T. Litfin, Y. Yang, and Y. Zhou, “Accurate prediction of protein contact maps by coupling residual two-dimensional bidirectional long short-term memory with convolutional neural networks,” *Bioinformatics*, vol. 34, no. 23, pp. 4039–4045, 2018.
- [4] J. Moult, K. Fidelis, A. Kryshchak, T. Schwede, and A. Tramontano, “Critical assessment of methods of protein structure prediction (casp)—round xii,” *Proteins: Structure, Function, and Bioinformatics*, vol. 86, pp. 7–15, 2018.
- [5] J. Schaarschmidt, B. Monastyrskyy, A. Kryshchak, and A. M. Bonvin, “Assessment of contact predictions in casp12: co-evolution and deep learning coming of age,” *Proteins: Structure, Function, and Bioinformatics*, vol. 86, pp. 51–66, 2018.
- [6] T. J. Taylor, H. Bai, C.-H. Tai, and B. Lee, “Assessment of casp10 contact-assisted predictions,” *Proteins: Structure, Function, and Bioinformatics*, vol. 82, pp. 84–97, 2014.
- [7] Q. Wu, Z. Peng, I. Anishchenko, Q. Cong, D. Baker, and J. Yang, “Protein contact prediction using metagenome sequence data and residual neural networks,” *Bioinformatics*, vol. 36, no. 1, pp. 41–48, 2020.

- [8] Y. Li, J. Hu, C. Zhang, D.-J. Yu, and Y. Zhang, “Respre: high-accuracy protein contact prediction by coupling precision matrix with deep residual neural networks,” *Bioinformatics*, vol. 35, no. 22, pp. 4647–4655, 2019.
- [9] C. E. S. Toca, *Predicción de mapas de contactos de proteínas mediante multclasificadores*. Editorial Universitaria, 2014.
- [10] B. He, S. Mortuza, Y. Wang, H.-B. Shen, and Y. Zhang, “Nebcon: protein contact map prediction using neural network training coupled with naïve bayes classifiers,” *Bioinformatics*, vol. 33, no. 15, pp. 2296–2306, 2017.
- [11] A. Kryshchuk, T. Schwede, M. Topf, K. Fidelis, and J. Moult, “Critical assessment of methods of protein structure prediction (casp)—round xiii,” *Proteins: Structure, Function, and Bioinformatics*, vol. 87, no. 12, pp. 1011–1020, 2019.
- [12] A. Andreeva, E. Kulesha, J. Gough, and A. G. Murzin, “The scop database in 2020: expanded classification of representative family and superfamily domains of known protein structures,” *Nucleic Acids Research*, 11 2019, gkz1064. [Online]. Available: <https://doi.org/10.1093/nar/gkz1064>
- [13] A. Amanullah and S. Naheed, “Structural bioinformatics: Computational software and databases for the evaluation of protein structure,” *RADS Journal of Biological Research & Applied Sciences*, vol. 9, no. 2, pp. 94–101, 2018.
- [14] S. de Oliveira and C. Deane, “Co-evolution techniques are reshaping the way we do structural bioinformatics,” *F1000Research*, vol. 6, 2017.
- [15] S. Seemayer, M. Gruber, and J. Söding, “Ccmpred—fast and precise prediction of protein residue–residue contacts from correlated mutations,” *Bioinformatics*, vol. 30, no. 21, pp. 3128–3130, 2014.
- [16] L. Kaján, T. A. Hopf, M. Kalaš, D. S. Marks, and B. Rost, “Freecontact: fast and free software for protein contact prediction from residue co-evolution,” *BMC bioinformatics*, vol. 15, no. 1, p. 85, 2014.
- [17] H. Kamisetty, S. Ovchinnikov, and D. Baker, “Assessing the utility of coevolution-based residue–residue contact predictions in a sequence-and structure-rich era,” *Proceedings of the National Academy of Sciences*, vol. 110, no. 39, pp. 15 674–15 679, 2013.
- [18] M. Ekeberg, T. Hartonen, and E. Aurell, “Fast pseudolikelihood maximization for direct-coupling analysis of protein structure from many homologous amino-acid sequences,” *Journal of Computational Physics*, vol. 276, pp. 341–356, 2014.
- [19] D. T. Jones, D. W. Buchan, D. Cozzetto, and M. Pontil, “Psicov: precise structural contact prediction using sparse inverse covariance estimation on large multiple sequence alignments,” *Bioinformatics*, vol. 28, no. 2, pp. 184–190, 2011.
- [20] J. Ma, S. Wang, Z. Wang, and J. Xu, “Protein contact prediction by integrating joint evolutionary coupling analysis and supervised learning,” *Bioinformatics*, vol. 31, no. 21, pp. 3506–3513, 2015.
- [21] J. Luttrell, T. Liu, C. Zhang, and Z. Wang, “Predicting protein residue-residue contacts using random forests and deep networks,” *BMC bioinformatics*, vol. 20, no. 2, pp. 115–127, 2019.
- [22] J. Xie, W. Ding, L. Chen, Q. Guo, and W. Zhang, “Advances in protein contact map prediction based on machine learning,” *Medicinal Chemistry*, vol. 11, no. 3, pp. 265–270, 2015.
- [23] B. Adhikari and J. Cheng, “Protein residue contacts and prediction methods,” in *Data Mining Techniques for the Life Sciences*. Springer, 2016, pp. 463–476.
- [24] C. E. Santiesteban-Toca and J. S. Aguilar-Ruiz, “A new multiple classifier system for the prediction of protein’s contacts map,” *Information Processing Letters*, vol. 115, no. 12, pp. 983–990, 2015.
- [25] V. Vapnik and R. Izmailov, “Rethinking statistical learning theory: learning using statistical invariants,” *Machine Learning*, vol. 108, no. 3, pp. 381–423, 2019.
- [26] N. Kriegeskorte, “Deep neural networks: a new framework for modeling biological vision and brain information processing,” *Annual review of vision science*, vol. 1, pp. 417–446, 2015.
- [27] M. Torrisi, G. Pollastri, and Q. Le, “Deep learning methods in protein structure prediction,” *Computational and Structural Biotechnology Journal*, 2020.

- [28] B. Adhikari, J. Hou, and J. Cheng, “Dncon2: improved protein contact prediction using two-level deep convolutional neural networks,” *Bioinformatics*, vol. 34, no. 9, pp. 1466–1472, 2018.
- [29] D. T. Jones and S. M. Kandathil, “High precision in protein contact prediction using fully convolutional neural networks and minimal sequence features,” *Bioinformatics*, vol. 34, no. 19, pp. 3308–3315, 2018.
- [30] S. Wang, S. Sun, Z. Li, R. Zhang, and J. Xu, “Accurate de novo prediction of protein contact map by ultra-deep learning model,” *PLoS computational biology*, vol. 13, no. 1, p. e1005324, 2017.
- [31] S. M. Kandathil, J. G. Greener, and D. T. Jones, “Prediction of interresidue contacts with deepmetapsicov in casp13,” *Proteins: Structure, Function, and Bioinformatics*, vol. 87, no. 12, pp. 1092–1099, 2019.
- [32] H. Fukuda and K. Tomii, “Deepeca: an end-to-end learning framework for protein contact prediction from a multiple sequence alignment,” *BMC bioinformatics*, vol. 21, no. 1, pp. 1–15, 2020.
- [33] D. W. Buchan and D. T. Jones, “Improved protein contact predictions with the metapsicov2 server in casp12,” *PROTEINS: structure, function, and bioinformatics*, vol. 86, pp. 78–83, 2018.
- [34] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [35] J. Zhu, S. Wang, D. Bu, and J. Xu, “Protein threading using residue co-variation and deep learning,” *Bioinformatics*, vol. 34, no. 13, pp. i263–i273, 2018.
- [36] M. Remmert, A. Biegert, A. Hauser, and J. Söding, “Hhblits: lightning-fast iterative protein sequence searching by hmm-hmm alignment,” *Nature methods*, vol. 9, no. 2, p. 173, 2012.
- [37] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in *ICML*, 2010.
- [38] W. Ding, W. Mao, D. Shao, W. Zhang, and H. Gong, “Deepconpred2: An improved method for the prediction of protein residue contacts,” *Computational and Structural Biotechnology Journal*, vol. 16, pp. 503–510, 2018.
- [39] R. Heffernan, Y. Yang, K. Paliwal, and Y. Zhou, “Capturing non-local interactions by long short-term memory bidirectional recurrent neural networks for improving prediction of protein secondary structure, backbone angles, contact numbers and solvent accessibility,” *Bioinformatics*, vol. 33, no. 18, pp. 2842–2849, 2017.
- [40] F. Morcos, T. Hwa, J. N. Onuchic, and M. Weigt, “Direct coupling analysis for protein contact prediction,” in *Protein structure prediction*. Springer, 2014, pp. 55–70.
- [41] R. Witte and J. Witte, *Statistics*. Wiley, 2017.
- [42] K. J. Berry, J. E. Johnston, and J. W. Mielke, *A primer of permutation statistical methods*. Springer, 2019.

Appendix

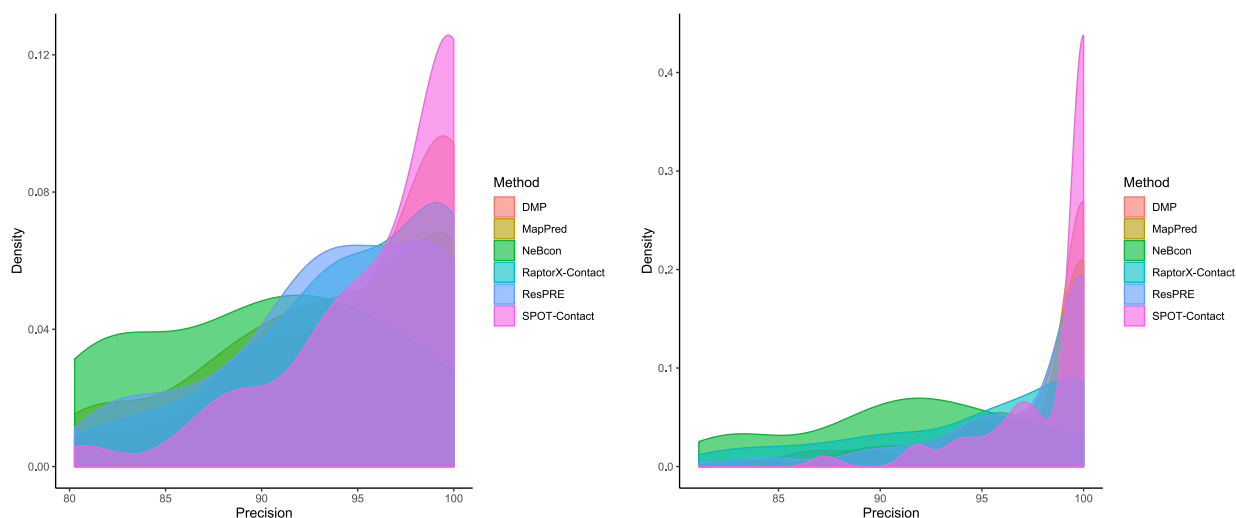


Figure 10: General distributions of the precisions for protein contact prediction with contact probability greater than 0.5, for each of the evaluated methods. The left figure corresponds to the SCOP class A and the right figure corresponds to the SCOP class B

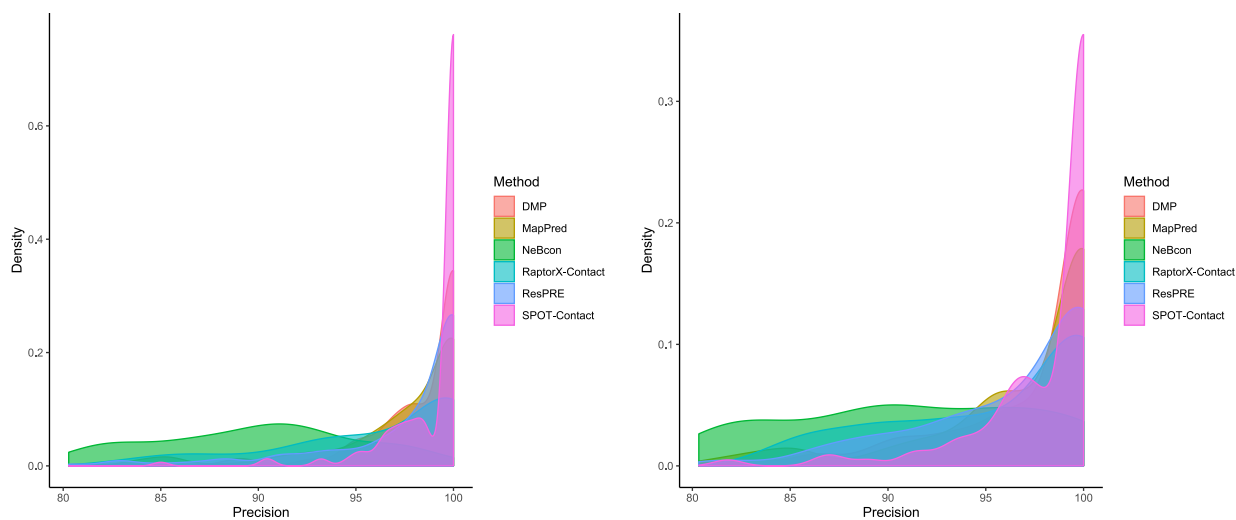


Figure 11: General distributions of the precisions for protein contact prediction with contact probability greater than 0.5, for each of the evaluated methods. The left figure corresponds to the SCOP class C and the right figure corresponds to the SCOP class D