Computers in Biology and Medicine 104 (2019) 241–249

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

BetaDL: A protein beta-sheet predictor utilizing a deep learning model and independent set solution

Toktam Dehghani, Mahmoud Naghibzadeh*, Mahdie Eghdami

Department of Computer Engineering, Ferdowsi University of Mashhad, Mashhad, Iran

ARTICLE INFO

Keywords:
Protein structure prediction
Deep learning model
Contact map
Beta-sheet structure
Graph search algorithm
Independent set

ABSTRACT

The sequence-based prediction of beta-residue contacts and beta-sheet structures contain key information for protein structure prediction. However, the determination of beta-sheet structures poses numerous challenges due to long-range beta-residue interactions and the huge number of possible beta-sheet structures. Recently gaining attention has been the prediction of residue contacts based on deep learning models whose results have led to improvement in protein structure prediction. In addition, to reduce the computational complexity of determining beta-sheet structures, it has been suggested that this problem be transformed into graph-based solutions. Consequently, the current work proposes BetaDL, a combination of a deep learning and a graph-based beta-sheet structure predictor. BetaDL adopts deep learning models to capture beta-residue contacts and improve beta-sheet structure predictions. In addition, a graph-based approach is presented to model the beta-sheets conformational space and a new score function is introduced to evaluate beta-sheets. Furthermore, the present study demonstrates that the beta-sheet structure can be predicted within an acceptable computational time by the utilization of a heuristic maximum weight independent set solution. When compared to state-of-the-art methods, experimental results from BetaSheet916 and BetaSheet1452 datasets indicate that BetaDL improves the accuracy of beta-residue contact and beta-sheet structure prediction. Using BetaDL, beta-sheet structures are predicted with a 4% and 6% improvement in the F1-score at the residue and strand levels, respectively. BetaDL's source code and data are available at http://kerg.um.ac.ir/index.php/datasets/#BetaDL.

1. Introduction

Prediction of a protein's 3D-structure from its sequence is a long-standing issue in structural biology [1]. According to the statistical data of NCBI's reference sequence database [2], out of the 47 million protein sequences collected until 2018, only the structure of 3% of these proteins have been deposited in the Protein Data Bank (PDB) archive [3]. As a result, there is a wide gap between the number of known protein sequences and the number of determined structures. This disparity indicates that the computational methods of protein structure prediction (PSP), e.g. homology-based and ab-initio approaches, can be considered as powerful complements to existing experimental techniques. Without any homologous proteins, ab-initio methods are the main solutions for the PSP problem. However, dealing with the computational complexity of these methods remains an open issue in research. In order to resolve this issue, the prediction of protein structures is simplified by applying the divide and conquer technique [4], by which the protein structure is decomposed into smaller sub-structures known as protein secondary structural elements, e.g. alpha-helices and beta-sheets. These elements are regarded as bridges between the sequence and the structure of proteins. Therefore, the accurate determination of the structure of these elements can significantly reduce the search space of the PSP problem. According to recent studies, the overall accuracy of secondary structure prediction methods is about 82% [5,6]. However, when compared to alpha-helices, the determination of beta-sheet structures is reported to be more challenging and the accurate prediction of their structures remains to be further explored [7,8].

The prediction of beta-sheet structures plays a critical role in a variety of applications, such as in designing new proteins [9], exploring energy landscapes [10], and understanding protein folding paths [11]. Furthermore, beta-sheet interactions have been implicated in the formation of protein aggregations observed in many human diseases, such as AIDS, cancer, and Alzheimer's [12,13]. Owing to the importance of sheets, the present work concentrates on the prediction of beta-sheet structures.

Beta-sheets are composed of strand pairs which are held together in parallel and anti-parallel forms by interactions between their beta-residues, known as beta-residue contact maps [14]. The major bottleneck

* Corresponding author.
E-mail address: naghibzadeh@um.ac.ir (M. Naghibzadeh).

https://doi.org/10.1016/j.compbiomed.2018.11.021
Received 27 August 2018; Received in revised form 23 November 2018; Accepted 27 November 2018

in sheet structure prediction is the difficulty of capturing long-range interactions between discontinued beta-strands, which are sequentially distant but spatially close in the protein structure [15,16]. Moreover, the other fundamental issue of the beta-sheet structure prediction is the huge number of possible sheet structures, which renders the accurate determination of the arrangement of strands within sheets more challenging [17,18]. To deal with these issues, efficient methods for the beta-sheet structure prediction are in demand.

In recent years, deep learning neural networks have demonstrated superior performance in several computational biology and chemistry applications, such as protein-protein interaction prediction [19], secondary structure prediction [20,21], protein fold recognition [22], and protein contact map prediction [23–26]. However, to the best of the current study's knowledge, deep learning models have not been directly applied to the field of beta-residue contact map and beta-sheet structure prediction. Hence, the motivation for the present work is to focus on employing deep learning models in the determination of sheet structures.

The current paper presents a novel beta-sheet structure predictor, BetaDL, which integrates deep learning models and graph-search algorithms. BetaDL predicts protein sheet structures in a four-step framework. First, to predict beta-residue contact maps, BetaDL takes advantage of the deep residue neural networks proposed by RaptorX-Contact. Second, a dynamic programming algorithm computes optimal strand pairwise alignments. Third, the new graph-based model for beta-sheets conformational space extracts high-scoring beta-sheets. Fourth, to manage the computational complexity of the beta-sheet structure prediction problem, a heuristic maximum weight independent set algorithm combines high-scoring beta-sheets to construct the final structure. Furthermore, to improve the accuracy of the predicted beta-sheets, a new score function evaluates sheet structures. The present work's approach compares well with that of state-of-the-art methods and BetaDL demonstrates better performance. The overall view of BetaDL is illustrated in Fig. 1.

2. Literature review

To predict protein beta-sheet structures, various approaches have been proposed. The predictors can be classified as homology-based [27–29] and ab-initio [17,18,30–35] methods, depending on whether or not these predictors incorporate the knowledge of sheet structures of homologous proteins. The present paper concentrates on ab-initio methods for beta-sheet structure prediction. As a result, the existing methods in this category are discussed in the following paragraphs.

More than a decade ago, Cheng and Baldi established a method called BetaPro for the beta-sheet structure prediction problem [30]. BetaPro consists of three steps: deriving beta-residue contact maps, determining strand pairwise alignments, and obtaining strand arrangements within sheets. According to this framework, two main challenges arise in sheet structure prediction. First, based on the Critical Assessment of Structure Prediction (CASP) experiments, the accurate prediction of long-range contacts between residues remains as a matter for further exploration [36]. As a result, there is a need for a precise method of beta-residue contact map prediction [37]. Second, the search space of possible sheet structures exponentially grows in proteins with a large number of strands [32]. Thus, it is essential to consider strategies for reducing search space of this problem.

The enhancement of the beta-residue contact map prediction is the first step towards achieving accurate determination of sheet structures. Among the existing residues in a protein, it should be noted that the prediction of long-range contacts, e.g. beta-residue contacts, is one of the most challenging tasks [38]. As a result, a great variety of methods have been developed to improve beta-residue contact prediction. Lippi and Frasconi suggested Markov logic networks (MLNS) which incorporate the structural constraints of sheets in the learning process [39]. In addition, Burkoff et al. introduced maximum entropy-based correlated mutation measurements (CMM) for predicting beta-residue contacts [40]. Moreover, Jones et al. put forward a method (PSICOV) which employs a sparse inverse covariance matrix to predict residue contacts [41]. This approach was adopted by other beta-sheet predictors, such as [33,34]. For instance, BCov utilizes PSICOV's residue contact estimations to define strand pair scores and applies integer

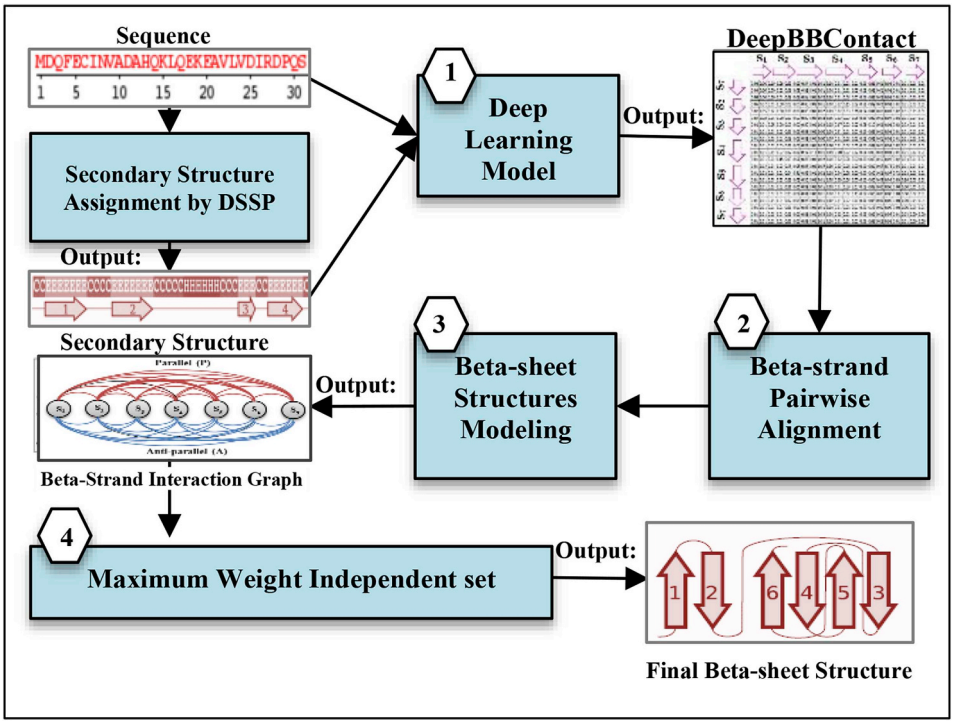

Fig. 1. A four-step framework of BetaDL for the protein beta-sheet structure prediction.

linear optimization to predict sheet structures. Furthermore, CCMpred employs a pseudo-likelihood maximization to predict contact maps [42]. This approach was followed by bbcontacts [35] which introduced two hidden Markov models to detect parallel and anti-parallel strand pairs from the residue contact maps predicted by CCMpred. In spite of all the efforts made, without many homologs sequences, the overall accuracy of predicted beta-residue contacts falls lower than that is required for the protein structure prediction [37]. Therefore, further development and new methods for beta-residue contact prediction are still needed.

As mentioned, the second significant challenge in beta-sheet structure prediction is the management of the exponential growth of number of possible sheet structures with respect to the number of protein strands. To deal with this issue, BetaPro utilizes a simple greedy algorithm to determine strand pairs and their interactions [30]. However, more accurate algorithms remain to be further explored. Some of the current methods, such as [18,31,32], present strategies to prune the search space. However, these methods are applicable for proteins with a limited number of strands, i.e. up to ten strands. For proteins with a large number of strands, the most recent method, Top-DBS, enhances the accuracy of sheet structure predictions by modeling potential sheet structures as a graph and applying a path cover algorithm [17]. However, the experimental results of the existing methods indicate that employing more recent contact map prediction methods and score functions can improve the accuracy of predictions.

3. Materials and methodology

3.1. Problem statement

The prediction of a protein sheet structure is addressed as the determination of the optimal assignment of strands to the independent sheets and the arrangement of strands in each sheet. To be more specific, the determination of the arrangement of strands in each sheet includes three parts: the order of strand pairs, their interaction types, i.e. parallel and anti-parallel, and their beta-residue contacts. For example, Fig. 2 shows the beta-sheet structure of protein 1GMX with seven strands where the numbers representing the sequential indices of strands. These strands are assigned to two sheets, $< s_1, s_6, s_5, s_2, s_4 >$ and $< s_3, s_7 >$, whose interaction types are parallel and anti-parallel, respectively.

The present work concentrates on proteins with open beta-sheets, i.e. no cycles, which are the most common beta-sheet types in cellular proteins [43]. As is the case for the most widely used methods [17,18,32,33], two constraints for the open sheets are considered. First, each strand can be paired with, at the most, two other strands in the sheet. Second, in multiple-sheet proteins, sheets must have no common strands, i.e. independent sheets. It is worth noting that these conditions are satisfied in about 80% of proteins in well-known beta-sheet datasets [30,33]. With these constraints in mind, the following sections present a graph-based method to predict beta-sheets.

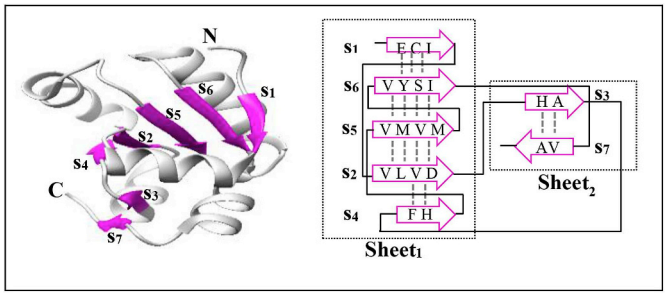

Fig. 2. Left: Structure of protein 1GMX. Right: parallel and anti-parallel beta-sheets. Vertical dash lines indicate beta-residue contacts.

3.2. General approach

The main goal of the proposed method is to improve the performance of beta-sheet structure prediction. The key to reaching this aim is the accurate prediction of beta-residue contact maps. Furthermore, for proteins with a large number of strands, it is extremely time-consuming to seek the native structure among the huge conformational space. Thus, devising an efficient strategy to explore conformation space is essential. Towards this end, BetaDL predicts protein sheet structures in a four-step framework (see Fig. 1) as follows:

(1) Predicting beta-residue contact probabilities via a deep learning model.
(2) Computing strand pairwise alignment scores by a dynamic programming algorithm.
(3) Modeling beta-sheets conformational space as a multi-edge graph.
(4) Determining the final beta-sheet structure with a maximum weight independent set algorithm.

3.3. Datasets

In order to predict beta-residue contact maps, the training set is a PDB25 subset extracted by Ref. [23]. This set includes 6767 protein chains whose structures were obtained by X-ray diffraction with a resolution of ≤ 2.5 Å in which any two proteins share less than 25% sequence identity.

For the sake of comparison with previous methods, the current study's test data are obtained from two well-established datasets: BetaSheet916 [30] and BetaSheet1452 [33]. These datasets have been utilized by many recent methods, such as [17,18,35]. The BetaSheet916 dataset was introduced as a benchmark for beta-sheet structure prediction methods and it includes 916 protein chains. The BetaSheet1452 dataset was presented as a complementary dataset and it was built from more recently deposited protein chains. The redundancy between the training and test datasets is filtered by excluding test proteins which share more than 25% sequence identity with any proteins within the training set.

In the present study, the Define Secondary Structure of Proteins (DSSP) tool assigns secondary structures [14]. The beta-residues are defined based on the DSSP's assignments and both beta-bridge and beta-strand residues (labeled B and E in DSSP) are considered as beta-residues.

3.4. Beta-residue contact map prediction using a deep learning model

The first step towards the beta-sheet structure prediction is the determination of beta-residue contact maps. Generally, the contact map is a 2D-representation of the protein structure indicating which residues in the sequence are close in the protein's 3D-structure. According to the CASP experiments, two residues are defined as being in contact if the Euclidean distance between their C_β atoms (C_α in the case of Glycine) is less than 8.0 Å [1]. The present work concentrates on beta-residue contact map prediction. To determine their contacts, the residue contact probabilities are obtained from one of the most successful contact predictors, known as RaptorX-Contact [23]. In this method, two deep residual neural networks are presented to integrate sequence conservation information and direct evolutionary coupling. These two networks are explained as follows:

(1) The first residual neural network learns the sequential features of a residue. It consists of a series of 1D-convolutional transformations of the sequential features, such as the protein sequence profile, the secondary structure, and solvent accessibility.
(2) The second residual neural network learns the contact occurrence patterns and the 2D-context of residue pairs. It conducts a series of 2D-convolutional transformations of pairwise features, such as

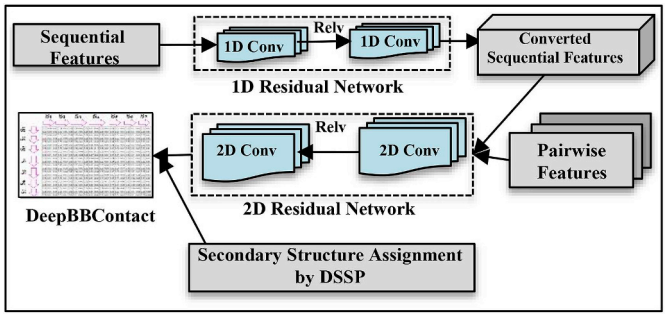

Fig. 3. Overall view of the deep beta-residue contact map prediction.

mutual information, pairwise potential, direct evolutionary information, and the converted output of the previous level. To achieve residue contact probabilities, a logistic regression is applied to the output of the second network.

The probabilities of beta-residue contacts (DeepBBContact) are determined using the residue contact probabilities predicted by the deep learning model and the DSSP-defined beta-strand regions. Fig. 3 illustrates the overall view of the deep learning model for DeepBBContact prediction. The next section uses these probabilities to infer beta-strand pairwise alignments in parallel and anti-parallel forms.

3.5. Beta-strand pairwise alignment using a dynamic programming algorithm

Prior to the prediction of sheet structures, the propensity of pairing of each two strands must be determined. Hence, this section computes the pairwise alignment of each two strands. Later, according to these alignment scores, the final sheet structure is predicted. The present study determines the optimal strand pairwise alignments in parallel and anti-parallel forms by applying a dynamic programming algorithm, i.e. Needleman-Wunsch [44]. This algorithm consists of three phases: initialization, forward, and backtracking. In the initialization and forward phases, the alignment matrix (G) for two strands, X and Y, is computed as follows:

$$G(0,\ 0) = 0,\quad G(i,\ 0) = i \times d,\quad G(0,\ j) = j \times d$$

$$G(i,j) = max\begin{cases} G(i-1,j-1) + DeepBBContact(X_i, Y_j) \\ G(i-1,j) + d \\ G(i,j-1) + d \end{cases} \quad (1)$$

where DeepBBContact(X_i,Y_j) represents the probability of the ith beta-residue of strand X and the jth beta-residue of strand Y making contact. It is worth mentioning that the beta-residue contact probabilities are computed in Section 3.4 and the strand regions are derived from the DSSP-assigned secondary structures. Finally, the optimal alignment is obtained by performing the backtracking phase and the alignment score, $Score_{pair}$, is found in the bottom right-hand corner of the alignment matrix. Note that this algorithm determines the optimal alignment in the parallel interaction. To compute the anti-parallel alignment, the same algorithm must be applied, with the exception that one of the strands is reversed by placing its amino acids in the opposite order. The next section constructs a model for sheet structures in the target protein using the computed alignment scores.

3.6. Beta-strand interaction graph

The current work introduces a new graph, named the beta-strand interaction graph, to comprehensively model the conformational space of the possible sheet structures of a protein. Note that pruning interactions before the enumeration of the possible arrangements of strand pairs in sheets can lead to overlooking some important structures. Therefore, in recent methods, such as [17,33], all possible forms of strand pairs are taken into account. Following the approach of these

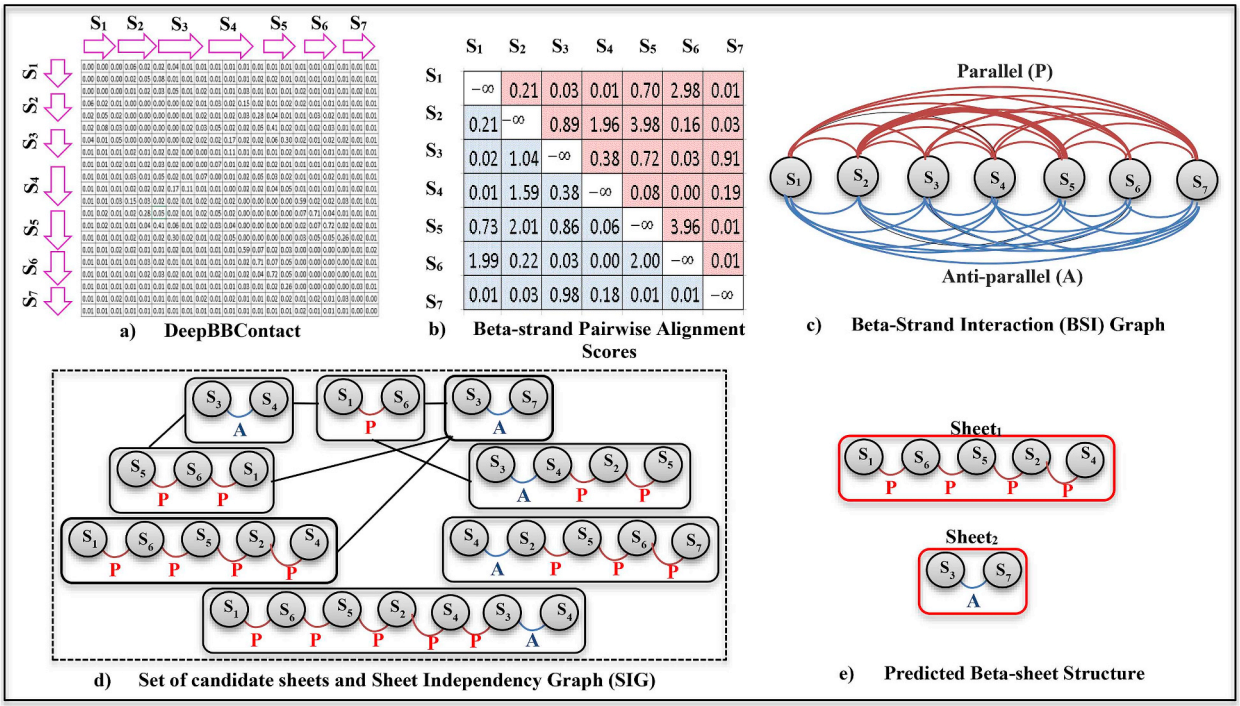

Fig. 4. BetaDL's steps for the beta-sheet structure prediction of protein 1GMX with seven strands. a) Predicted beta-residue contact (DeepBBContact) by the deep learning model. b) Beta-strand pairwise alignment scores (W), where the upper (red) and lower (blue) diagonal parts indicate parallel and anti-parallel alignment scores, respectively. c) The BSI graph, where parallel and anti-parallel interactions are shown in red and blue arcs, respectively. The weight of edges is not indicated for simplicity. d) The set of candidate sheets generated by APSP is illustrated as a sheet independency graph (SIG). Sheets with no common strand are connected by edges. Note that weight of sheets and some of the generated sheets are not shown for simplicity. e) The maximum weight independent subset of sheets determined by MWIS as a predicted beta-sheet structure. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

methods, the present work considers the parallel and anti-parallel forms of strand pairs. In addition, to reduce the computational complexity, a new representation for strand interactions is presented.

The Beta-Strand Interaction (BSI) graph is an undirected multi-edge graph that consists of four-tuples<V, E, W, L>. In this graph, the set of vertices, V= $\{s_1, ..., s_n\}$, represents strands, s_i, where $1 \leq i \leq n$ and n is the number of strands in the protein derived from the DSSP-assigned secondary structures. In addition, the set of edges, E, represents all types of strand pairwise interactions. In order to encode parallel (P) and anti-parallel (A) interactions, this graph is considered to be multi-edge, i.e. with exactly two edges between each pair of strands (Fig. 4c). In this graph, the interaction types, parallel and anti-parallel, are indicated by labels: L= {P, A}. Hence, set E is defined as follows:

$$E = \{(s_i, s_j, l) | s_i, s_j \in V \ \wedge \ l \in L\} \tag{2}$$

In the BSI graph, the weight matrix, W, indicates the alignment scores, $Score_{pair}$, between all strand pairs (s_i,s_j) in parallel and anti-parallel interactions, as computed in Section 3.5. This 2D-matrix includes (n× n) entries and the upper and lower diagonal parts of W indicate the parallel and anti-parallel alignment scores of strand pairs, respectively (Fig. 4b). Matrix W is defined as:

$$W_{i,j} = \begin{cases} Score_{pair}(s_i, s_j, P), & i < j \\ -\infty \quad , & i = j \\ Score_{pair}(s_i, s_j, A), & i > j \end{cases} \tag{3}$$

3.7. Beta-sheet structure prediction

In the previous section, all possible interactions between strands are modeled as a BSI graph. Note that there is a one-to-one correspondence between open beta-sheets in the target protein and simple paths, i.e. with no cycle, in the BSI graph [17]. Moreover, as mentioned in Section 3.1, a beta-sheet structure is defined as a set of independent sheets with no common strands. Consequently, the correct beta-sheet structure can be predicted by determining the maximum weight independent set of paths in the BSI graph. By means of pruning the search space, the current work solves the maximum weight independent set in two steps by:

(1) Producing a set of candidate sheets.
(2) Determining the maximum weight independent subset of sheets in the above set.

In order to achieve a reliable prediction, it is essential to provide an accurate score function to evaluate the generated paths. Hence, first, the next subsection describes the proposed score function and later, the details of the above-mentioned steps are discussed.

3.7.1. Score function

A critical issue that impacts the performance of the sheet structure prediction is the accuracy of score functions. In the literature, two types of score functions are utilized: pairing and topology score functions [32]. The pairing score function was employed by Refs. [17,18,30,33], while the topology score function was applied by Refs. [15,31]. Furthermore, methods, such as [32,34], utilized combinations of these two score functions. In the pairing score function, scores of strand pairwise interactions are calculated by finding their strand alignments. Later, the score of sheet structures is assigned by taking the average or sum of their strand pairwise interactions. This type of score function suffers from its monotonic nature and tendency to assign all strands into one sheet. On the other hand, in the topology score function, scores of strand pairwise interactions and sheets are determined by assigning probabilities based on several topological features. However, the discriminative power of this type of score function is significantly reduced by the growth of the number of strands in the proteins [15]. Therefore, providing an accurate score function for beta-sheet structure prediction remains an open problem.

The present paper concentrates on the pairing score function and attempts to overcome its deficiencies. For this reason, by considering the natural protein beta-sheet conformations, a new score function is introduced. This score function evaluates the sheet structures based on the strand pairwise alignment scores and a probability model for the sheet sizes, i.e. the number of strands in the sheets. To generate this probability model, proteins with open beta-sheets are extracted from the training dataset. In these proteins, the probabilities of different sheet sizes are derived with respect to the total number of strands in the proteins. Later, according to these probabilities, for each protein with n strands, the probability of sheet formation having n_S strands or more, denoted by $\overline{F_x}(n_S|n) = 1 - P(x < n_S|n)$, is computed. Therefore, in the target protein, for each sheet, $S = < s_1, \ ..., s_{n_S} >$, its score is determined as follows:

$$Score_{sheet}(S) = \left(\sum_{i=1}^{n_S} Score_{pair}\left(s_i, s_{i+1}, l\right) \right) \times \bar{F}_x(n_S|n) \tag{4}$$

where $Score_{pair}$ are computed in Section 3.5. Based on the proposed score function, the following subsection generates a set of maximum-score sheets.

Step 1: Producing a set of candidate beta-sheets

To generate a set of paths with the highest scores, the present study utilizes the all-pairs shortest paths (APSP) solution. To determine APSP in the BSI graph and convert the longest paths into the shortest ones, the negative of the original weights of edges, $-W_{BSI}$, is considered. One of the promising algorithms for finding APSP in the presence of negative weights is the Floyd-Warshall algorithm [45]. Hence, due to the existence of negative cycles in the BSI graph, a modified version of this algorithm is introduced (see Fig. 5).

The APSP algorithm begins by assuming no intermediate vertex in the paths between pairs. Later, in each step, an intermediate vertex, r, is selected and the path between each pair of vertices, Path[u,v], is updated if the new path is shorter than its previous estimation. In addition, before merging any two paths, the common strands of two paths, Path[u,r] and Path[r,v], are extracted and stored in the set Com. Later, their intersections are omitted from one of the paths and the two paths are merged. In this algorithm, in the merging phase, the score of Path [u,v] is updated based on two paths, Path[u,r] and Path[r,v], as follows:

$$Score_{sheet}(Path[u, r] + Path[r, v]) = \left(\sum_{i=1}^{|Path[u,r]+Path[r,v]|} - W_{s_i, s_{i+1}} \right) \times \overline{F_x}(|Path[u, r] + Path[r, v]||n) \tag{5}$$

Step 2: Determining the final beta-sheet structure

Generating a set of beta-sheets

```
1  for all vertices u ∈ V_BSI
2    for all vertices v ∈ V_BSI
3      for all vertices r ∈ V_BSI
4        If (Score_sheet(Path[u, v]) >
               Score_sheet(Path[u, r] + Path[r, v])) then
5          Extract their common vertices
6            Com=Path[u, r] ∩ Path[r, v] − {r}
7          Omit their intersection
8            Remove c_k ∈ Com from one of the path
9          Merge two paths
10           Path[u,v]=Merge(Path[u,r],Path[r,v])
```

Fig. 5. An overview of the APSP algorithm.

In the previous section, between each two strands in the target protein, the strand arrangement with the highest score is determined and a set of candidate sheets is produced. To infer the final beta-sheet structure, first, the independency between sheets in the generated set is indicated as a new graph and later, in this graph, the maximum weight independent set is determined.

The Sheet Independency Graph (SIG) is defined as a weighted graph, where vertices represent sheets and their weights signify their corresponding sheet scores, i.e. $Score_{sheet}/n_s$, where n_s is the number of strands in the sheet and $Score_{sheet}$ is computed by Eq. (4). Moreover, edges indicate the independency of sheets and two vertices are connected by an edge if their corresponding sheets have no common strands. The overall view of this graph is illustrated in Fig. 4. d.

The current paper presents an algorithm, named MWIS, to determine the Maximum Weight Independent Set in SIG. To find MWIS, a subset of vertices must be selected, in which every two vertices are adjacent to each other. In addition, the sum of weights of selected vertices must be maximal. Consequently, in the MWIS algorithm, the independent subset with the maximum weight is determined by finding the largest weight independent subset for each vertex in SIG and selecting the maximum weight among these. Fig. 6 outlines this algorithm in detail.

In the MWIS algorithm, vertices are sorted in a descending order of their weights and the maximum degree of each vertex is restricted by considering its n-top-score neighbours, where n is the number of strands in the protein (Line 3–4 in Fig. 6). Furthermore, during the search for an independent subset containing a given vertex, this algorithm prunes vertices unable to form an independent subset with a weight larger than the weight of the current maximum subset, Max (Line 10 in Fig. 6). After all vertices in SIG are considered, the MWIS algorithm is terminated and the maximum weight independent set, Max-Set, is presented as the final beta-sheet structure.

4. Experimental results

The present method, BetaDL, is compared with recent beta-sheet structure predictors in the literature, which are applicable for proteins with any number of strands and available for downloading as standalone packages, such as BetaPro [30], BCov [33], bbcontacts [35], and Top-DBS [17].

Determining the final beta-sheet strucure

```
1  Initialization:
2     Max=0 , Max-Set=∅
3     Sort vertices in SIG according their weights.
4     Select n-top-score neighbours of each vertex and omit others.
5  Main routine:
6        for each vertex, vi , in SIG
7           U=Neighbours of vi with the larger indices
8           W=Weight of vi
9           V= vi
10          if (W+∑e∈U we )>Max
11             MWIS (U,W,V)
12 end
13 MWIS routine (U,W,V):
14       if U=∅ and W>Max
15          Max=W, Max-Set=V
16          Return
17       while U≠ ∅
18          Omit a vertex , u , with the maximum score from U
19          U=Intersection of U and neighbours of u
20          W=Sum of W and weight of u
21          V=V ∪ u
22          MWIS (U,W,V)
23 End
```

Fig. 6. An overview of the MWIS algorithm.

4.1. Performance measurements

In this study, the performance of beta-sheet structure predictors is evaluated at two levels: beta-residue (contact map) and strand (strand pairing and interaction types) levels. At these levels, the following performance measurements are computed: Precision, Recall, F1-score, and Matthews Correlation Coefficient (MCC), where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively:

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \tag{6}$$

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \tag{7}$$

$$\text{F1} - \text{score} = (2 \times \text{Precision} \times \text{Recall})/(\text{Precision} + \text{Recall}) \tag{8}$$

$$\text{MCC} = \frac{(\text{TP} \times \text{TN} - \text{FP} \times \text{FN})}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}} \tag{9}$$

4.2. Experiments

In order to analyse BetaDL's performance, the present study conducts eight experiments. On two well-known datasets, the prediction power of BetaDL is compared with other methods at the residue and strand levels. There is also a discussion on BetaDL's efficacy in the prediction of beta-residue contact maps and this is compared to that of the RaptorX-Contact server. In addition, without considering the knowledge of secondary structures, the current work evaluates the capability of BetaDL in the prediction of beta-sheet structures. Furthermore, the impact of score functions on the performance of BetaDL is discussed. Finally, an analysis is provided on the computational time of BetaDL in proteins with a variant number of strands.

4.2.1. Performance comparison at the beta-residue level

One of the main goals of existing methods is improving the performance of beta-sheet predictors at the residue level. Hence, in the current study, BetaDL's performance is reported at the beta-residue level on the BetaSheet916 dataset and compared with present methods that predict beta-residue contact maps, i.e. BetaPro, BCov, and bbcontacts. As seen in Table 1, BetaDL outperforms the others in almost all performance measurements i.e. its F1-score is 4% higher than the best results of other studies. However, an exception should be noted in that bbcontacts performs slightly higher in terms of precision.

4.2.2. Performance comparison at the beta-strand level

In this experiment, the performance of the beta-sheet predictors is discussed at the strand level. Table 2 compares BetaDL's performance to that of state-of-the-art methods at the strand level on the BetaSheet916 dataset. Results show that BetaDL surpasses most performance measurements, i.e. the proposed method's F1-score is 6% higher than the best results of previous studies (Top-DBS). In addition, this evaluation

Table 1
Performance comparison of beta-sheet predictors at the residue level on the BetaSheet916 dataset.

Method	Performance Measurments		
	Precision	Recall	F1-score
BetaDL	71%	73%	72%
bbcontacts	72%	65%	68%
BCov	41%	42%	42%
BetaPro	38%	44%	41%

Table 2
Performance comparison of beta-sheet predictors at the strand level on the BetaSheet916 dataset.

Method	Performance Measurments			
	Precision	Recall	F1-score	F1 ≥ 70
BetaDL	83%	80%	82%	81%
bbcontacts	84%	58%	68%	56%
Top-DBS	75%	77%	76%	62%
BCov	59%	62%	61%	44%
BetaPro	53%	60%	56%	32%

employs the qualitative measure, F1 ≥ 70, introduced by Ref. [39]. This measure indicates the percentage of proteins in the dataset for which strand pairs are predicted with an F1-score equal to or more than 70%. BetaDL achieves a higher F1 ≥ 70 than that of previous research, i.e. a 19% improvement compared to the best results of previous works.

4.2.3. Performance comparison on BetaSheet1452

In order to study BetaDL's performance on a larger and newer dataset, the current study conducts a comparison of methods on the BetaSheet1452 dataset. Table 3 indicates that BetaDL improves the performance measurements at the residue level. In addition, as seen in Table 4, BetaDL's performance is higher than that of previous works at the strand level, i.e. its F1-score is 10% higher than the best results of other studies. A comparison of the reported results in Tables 1 and 2 with those in Tables 3 and 4 clearly shows that BetaDL's performance is unaffected by a dataset change.

4.2.4. Performance comparison between BetaDL and RaptorX-Contact

The current study discusses the effect of BetaDL on beta-residue contact map prediction. As described in Section 3.4, BetaDL predicts beta-residue contact maps and beta-sheet structures based on residue contact probabilities determined by one of the most successful general-purpose contact map predictors, the RaptorX-Contact server [23]. At the beta-residue level, the present work compares the performance of BetaDL with that of RaptorX-Contact on the BetaSheet916 dataset. As shown in Table 5, BetaDL improves the performance measurements of RaptorX-Contact, i.e. with 13%, and 15% increments in the F1-score and MCC, respectively.

4.2.5. Impact of deep learning model on BetaDL's performance

In this experiment, the effect of incorporating the deep learning model into beta-sheet structure prediction is discussed by comparing two versions of BetaDL. One version employs beta-residue contact probabilities in DeepBBContact and the second version utilizes the beta-residue contact probabilities determined by BetaPro [30]. At the beta-residue level, Table 6 presents the performance measurements of these two versions of BetaDL on the BetaSheet916 dataset. It is evident that the BetaDL version utilizing DeepBBContact surpasses the other package in the whole performance measurements, as highlighted by the 5% increments in both the F1-score and MCC.

Table 3
Residue-level performance comparision on BetaSheet1452.

Method	Performance Measurments		
	Precision	Recall	F1-score
BetaDL	76%	78%	76%
bbcontacts	73%	65%	69%
BCov	42%	45%	43%

Table 4
Strand-level performance comparision on BetaSheet1452.

Method	Performance Measurments		
	Precision	Recall	F1-score
BetaDL	83%	82%	82%
bbcontacts	88%	61%	72%
Top-DBS	74%	70%	71%
BCov	59%	63%	61%

Table 5
Performance comparison between BetaDL and RaptorX- Contact at the beta-residue level on the BetaSheet916 dataset.

Method	Performance Measurments			
	Precision	Recall	F1-score	MCC
BetaDL	71%	73%	72%	71%
RaptorX-Contact	62%	59%	59%	56%

*Note that the measurements of RaptorX-Contact are reported at its best performance, i.e. highest F1-score.

Table 6
Effect of beta-residue contact probabilities on BetaDL's performance.

Method	Performance Measurments			
	Precision	Recall	F1-score	MCC
BetaDL (using DeepBBContact)	71%	73%	72%	71%
BetaDL (using BetaPro's contacts)	67%	69%	67%	66%

4.2.6. Impact of incorporating predicted secondary structures on BetaDL's performance

This experiment verifies BetaDL's capability in the prediction of beta-sheet structures without the knowledge of exact secondary structures. BetaDL's performance in two scenarios is analyzed: (1) using the exact secondary structures assigned by DSSP [14] and (2) employing the predicted secondary structures determined by DeepCNF, one of the most successful secondary structure predictors [21]. At the residue-level, Table 7 illustrates BetaDL's performance measurements in these two cases on the BetaSheet916 dataset and compares these with the results of bbcontacts. As expected, the BetaDL and bbcontacts versions based on DSSP-assigned secondary structures remarkably enhance performance measurements when compared to the other versions. Furthermore, as shown in Table 7, without employing exact secondary structures, BetaDL achieves higher performance measurements than bbcontacts, i.e. an 8% improvement in the F1-score.

4.2.7. Impact of score functions on BetaDL's performance

The current study discusses BetaDL's performance by employing different score functions. As mentioned in Section 3.7, the present paper attempts to improve the pairing score function. Hence, the effect

Table 7
Effect of secondary structure assignment on BetaDL's performance at the residue level.

Method	Secondary Structure Assignment	Performance Measurments		
		Precision	Recall	F1-score
BetaDL	DSSP	71%	73%	72%
	DeepCNF	58%	61%	59%
bbcontacts	DSSP	72%	65%	68%
	PSIPRED	47%	55%	51%

*The values of bbcontacts are taken from the corresponding article [35].

Table 8
Effect of score function modification on BetaDL's performance at the strand level.

Method	Performance Measurments			
	Precision	Recall	F1-score	MCC
BetaDL	83%	80%	82%	77%
BetaDL (using BetaPro's scores)	76%	75%	76%	70%

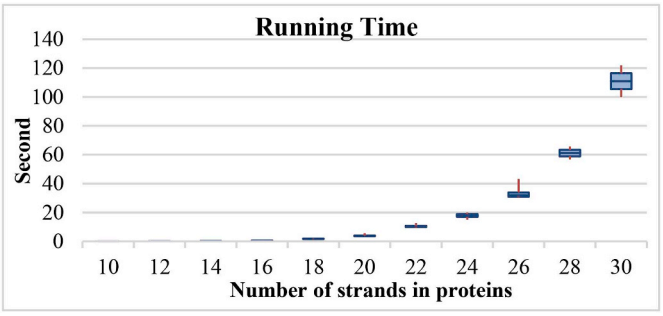

Fig. 7. BetaDL's runtime with respect to the total number strands in proteins.

of the proposed score function on beta-sheet structure prediction is evaluated by comparing two versions of the BetaDL algorithm: one employing the score function proposed in the current study and the second using the standard score function presented by BetaPro [30]. As shown in Table 8, at the strand level, the proposed score function improves prediction results, i.e. a 6% and 7% improvement in the F1-score and MCC, respectively.

4.2.8. Runtime of BetaDL

The runtime of BetaDL is evaluated on a workstation with Microsoft Windows 10 Enterprise, an Intel core i7 ×64 processor, 2.5 GHZ CPU, and 16 GB RAM. With respect to the number of strands in the proteins, the plot box of BetaDL's runtime is illustrated in Fig. 7. It is evident that runtime exponentially increases with the number of strands. However, for 98% of the proteins in the datasets, i.e. proteins with less than 27 strands, BetaDL determines beta-sheet structures in less than 1 min. Furthermore, the runtime of BetaDL is acceptable for proteins with a large number of strands.

5. Comparison and discussion

This section provides an overall comparison between the performance of the current work and that of existing methods in beta-sheet structure prediction. The present study conducts several experiments and, from these experimental results, concludes that BetaDL offers some advantages over other methods. It is evident that the current method achieves higher performance at the residue and strand levels when compared to previous beta-sheet structure predictors in literature (Tables 1 and 2). BetaDL outperforms bbcontacts [35] in the recall and F1-score. However, the precision of bbcontacts is slightly better when compared to that of the proposed method, which may be due to filtering low-scoring beta-sheet structures in the post-processing phase of bbcontacts. Moreover, BetaDL's performance on two well-known beta-sheet datasets is studied, the results of which indicate that the current work's performance is unaffected by dataset changes (Tables 3 and 4). In addition, it is noteworthy that BetaDL utilizes the residue contact probabilities determined by RaptorX-Contact [23]. However, compared to RaptorX-Contact, BetaDL significantly improves performance measurements in beta-residue contact predictions (Table 5). Furthermore, the current method's performance is studied using predicted secondary structures as input data, instead of exact structures. As expected, at the residue level, performance measurements decrease. However, compared to bbcontacts, an improvement in performance is observed (Table 7). Therefore, BetaDL enhances the prediction of beta-sheet structures when the exact secondary structures are not available. In any case, providing more accurate secondary structure predictions can improve the performance.

6. Conclusion

The current study develops a new method, BetaDL, which integrates deep learning models and graph search algorithms to refine beta-sheet structure prediction. BetaDL outperforms previous methods due to its four main contributions. First, deep residual neural networks are employed for the beta-residue contact map prediction. Second, in contrast with most existing methods, the proposed approach comprehensively models the protein beta-sheet conformational space as a multi-edge graph. Third, a new score function is presented which determines the probability of a beta-sheet formation with a specific number of beta-strands and leads to improvements in prediction results. Fourth, beta-sheet structures are predicted by transforming this problem into a computational solution, a maximum weight independent set, and by applying heuristic algorithms.

The current method demonstrates that deep-learning models exhibit better performance in beta-residue contact predictions. Furthermore, the BetaDL's higher performance and ability to explore conformation space in an acceptable computational time renders the proposed method suitable for beta-sheet structure prediction. In the future, the authors plan to extend BetaDL to predict various types of beta-sheets, such as circular beta-sheets and beta-sheets which contain strands with more than two partners.

Acknowledgments

The authors would like to express their gratitude to Dr. Castrense Savojardo, Dr. Jessica Andreani, Dr. Johannes Soding, and Dr. Jinbo Xu for sharing their research results.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2018.11.021.

References

[1] C.-H. Tai, H. Bai, T.J. Taylor, B. Lee, Assessment of template-free modeling in CASP10 and ROLL, Proteins Struct. Funct. Bioinforma. 82 (2014) 57–83, https://doi.org/10.1002/prot.24470.
[2] K.D. Pruitt, T. Tatusova, G.R. Brown, D.R. Maglott, NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy, Nucleic Acids Res. 40 (2012) D130–D135, https://doi.org/10.1093/nar/gkr1079.
[3] H.M. Berman, The protein Data Bank, Nucleic Acids Res. 28 (2000) 235–242, https://doi.org/10.1093/nar/28.1.235.
[4] J. Lee, S.-Y. Kim, J. Lee, Protein structure prediction based on fragment assembly and parameter optimization, Biophys. Chem. 115 (2005) 209–214, https://doi.org/10.1016/j.bpc.2004.12.046.
[5] Y. Yang, J. Gao, J. Wang, R. Heffernan, J. Hanson, K. Paliwal, Y. Zhou, Sixty-five years of the long march in protein secondary structure prediction: the final stretch? Briefings Bioinf. (2016) bbw129, https://doi.org/10.1093/bib/bbw129.
[6] C.N. Magnan, P. Baldi, SSpro/ACCpro 5: almost perfect prediction of protein secondary structure and relative solvent accessibility using profiles, machine learning and structural similarity, Bioinformatics 30 (2014) 2592–2597, https://doi.org/10.1093/bioinformatics/btu352.
[7] W. Mao, T. Wang, W. Zhang, H. Gong, Identification of residue pairing in interacting β-strands from a predicted residue contact map, BMC Bioinf. 19 (2018) 146, https://doi.org/10.1186/s12859-018-2150-1.
[8] W. Qu, H. Sui, B. Yang, W. Qian, Improving protein secondary structure prediction using a multi-modal BP method, Comput. Biol. Med. 41 (2011) 946–959, https://doi.org/10.1016/J.COMPBIOMED.2011.08.005.
[9] C.K. Smith, L. Regan, Guidelines for protein design: the energetics of beta sheet side chain interactions, Science 270 (1995) 980–982 http://www.ncbi.nlm.nih.gov/pubmed/7481801.
[10] N.S. Burkoff, C. Várnai, S.A. Wells, D.L. Wild, Exploring the energy landscapes of protein folding simulations with Bayesian computation, Biophys. J. 102 (2012)

878–886, https://doi.org/10.1016/j.bpj.2011.12.053.
[11] S. Shenker, C.W. O'Donnell, S. Devadas, B. Berger, J. Waldispühl, Efficient traversal of beta-sheet protein folding pathways using ensemble models, J. Comput. Biol. 18 (2011) 1635–1647, https://doi.org/10.1089/cmb.2011.0176.
[12] E. Koh, T. Kim, H. s. Cho, Mean curvature as a major determinant of β-sheet propensity, Bioinformatics 22 (2006) 297–302, https://doi.org/10.1093/bioinformatics/bti775.
[13] B.L. Kagan, J. Thundimadathil, Amyloid peptide pores and the beta sheet conformation, Adv. Exp. Med. Biol. 677 (2010) 150–167, https://doi.org/10.1007/978-1-4419-6327-7_13.
[14] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, Biopolymers 22 (1983) 2577–2637, https://doi.org/10.1002/bip.360221211.
[15] I. Ruczinski, C. Kooperberg, R. Bonneau, D. Baker, Distributions of beta sheets in proteins with application to structure prediction, Proteins Struct. Funct. Genet. 48 (2002) 85–97, https://doi.org/10.1002/prot.10123.
[16] D. Kihara, The effect of long-range interactions on the secondary structure formation of proteins, Protein Sci. 14 (2005) 1955–1963, https://doi.org/10.1110/ps.051479505.
[17] T. Dehghani, M. Naghibzadeh, J. Sadri, Enhancement of protein β-sheet topology prediction using maximum weight disjoint path cover, IEEE ACM Trans. Comput. Biol. Bioinf (2018), https://doi.org/10.1109/TCBB.2018.2837753 1–1.
[18] M. Sabzekar, M. Naghibzadeh, M. Eghdami, Z. Aydin, Protein β-sheet prediction using an efficient dynamic programming algorithm, Comput. Biol. Chem. 70 (2017) 142–155, https://doi.org/10.1016/j.compbiolchem.2017.08.011.
[19] T. Sun, B. Zhou, L. Lai, J. Pei, Sequence-based prediction of protein protein interaction using a deep-learning algorithm, BMC Bioinf. 18 (2017) 277, https://doi.org/10.1186/s12859-017-1700-2.
[20] R. Heffernan, K. Paliwal, J. Lyons, A. Dehzangi, A. Sharma, J. Wang, A. Sattar, Y. Yang, Y. Zhou, Improving prediction of secondary structure, local backbone angles and solvent accessible surface area of proteins by iterative deep learning, Sci. Rep. 5 (2015) 11476, https://doi.org/10.1038/srep11476.
[21] S. Wang, J. Peng, J. Ma, J. Xu, Protein secondary structure prediction using deep convolutional neural fields, Sci. Rep. 6 (2016) 18962, https://doi.org/10.1038/srep18962.
[22] J. Hou, B. Adhikari, J. Cheng, DeepSF: deep convolutional neural network for mapping protein sequences to folds, Bioinformatics 34 (2018) 1295–1303, https://doi.org/10.1093/bioinformatics/btx780.
[23] S. Wang, S. Sun, Z. Li, R. Zhang, J. Xu, Accurate de novo prediction of protein contact map by ultra-deep learning model, PLoS Comput. Biol. 13 (2017) e1005324, , https://doi.org/10.1371/journal.pcbi.1005324.
[24] B. Adhikari, J. Hou, J. Cheng, DNCON2: improved protein contact prediction using two-level deep convolutional neural networks, Bioinformatics 34 (2018), https://doi.org/10.1093/bioinformatics/btx781.
[25] D. Xiong, J. Zeng, H. Gong, A deep learning framework for improving long-range residue–residue contact prediction using a hierarchical strategy, Bioinformatics 33 (2017) 2675–2683, https://doi.org/10.1093/bioinformatics/btx296.
[26] K. Stahl, M. Schneider, O. Brock, EPSILON-CP: using deep learning to combine information from multiple sources for protein contact prediction, BMC Bioinf. 18 (2017) 303, https://doi.org/10.1186/s12859-017-1713-x.
[27] M. Menke, B. Berger, L. Cowen, Markov random fields reveal an N-terminal double beta-propeller motif as part of a bacterial hybrid two-component sensor system, Proc. Natl. Acad. Sci. U. S. A 107 (2010) 4069–4074, https://doi.org/10.1073/pnas.0909950107.
[28] N.M. Daniels, R. Hosur, B. Berger, L.J. Cowen, SMURFLite: combining simplified Markov random fields with simulated evolution improves remote homology detection for beta-structural proteins into the twilight zone, Bioinformatics 28 (2012) 1216–1222, https://doi.org/10.1093/bioinformatics/bts110.
[29] N.M. Daniels, A. Gallant, N. Ramsey, L.J. Cowen, MRFy: remote homology detection for beta-structural proteins using Markov random fields and stochastic search, IEEE ACM Trans. Comput. Biol. Bioinf 12 (2015) 4–16, https://doi.org/10.1109/TCBB.2014.2344682.
[30] J. Cheng, P. Baldi, Three-stage prediction of protein β-sheets by neural networks, alignments and graph algorithms, Bioinformatics 21 (2005) i75–i84, https://doi.org/10.1093/bioinformatics/bti1004.
[31] Z. Aydin, Y. Altunbasak, H. Erdogan, Bayesian models and algorithms for protein β-sheet prediction, IEEE ACM Trans. Comput. Biol. Bioinf 8 (2011) 395–409, https://doi.org/10.1109/TCBB.2008.140.
[32] R. Fonseca, G. Helles, P. Winter, Ranking beta sheet topologies with applications to protein structure prediction, J. Math. Model. Algorithm. 10 (2011) 357–369, https://doi.org/10.1007/s10852-011-9162-4.
[33] C. Savojardo, P. Fariselli, P.L. Martelli, R. Casadio, BCov: a method for predicting β-sheet topology using sparse inverse covariance estimation and integer programming, Bioinformatics 29 (2013) 3151–3157, https://doi.org/10.1093/bioinformatics/btt555.
[34] M. Eghdami, T. Dehghani, M. Naghibzadeh, BetaProbe, A probability based method for predicting beta sheet topology using integer programming, 2015 5th Int. Conf. Comput. Knowl. Eng. ICCKE 2015, 2015, https://doi.org/10.1109/ICCKE.2015.7365819.
[35] J. Andreani, J. Söding, bbcontacts: prediction of β-strand pairing from direct coupling patterns, Bioinformatics 31 (2015) 1729–1737, https://doi.org/10.1093/bioinformatics/btv041.
[36] B. Monastyrskyy, D. D'Andrea, K. Fidelis, A. Tramontano, A. Kryshtafovych, New encouraging developments in contact prediction: assessment of the CASP11 results, Proteins Struct. Funct. Bioinforma. 84 (2016) 131–144, https://doi.org/10.1002/prot.24943.
[37] Q. Wuyun, W. Zheng, Z. Peng, J. Yang, A large-scale comparative assessment of methods for residue–residue contact prediction, Briefings Bioinf. (2016) bbw106, https://doi.org/10.1093/bib/bbw106.
[38] P. Di Lena, K. Nagata, P. Baldi, Deep architectures for protein contact map prediction, Bioinformatics 28 (2012) 2449–2457, https://doi.org/10.1093/bioinformatics/bts475.
[39] M. Lippi, P. Frasconi, Prediction of protein β-residue contacts by Markov logic networks with grounding-specific weights, Bioinformatics 25 (2009) 2326–2333, https://doi.org/10.1093/bioinformatics/btp421.
[40] N.S. Burkoff, C. Várnai, D.L. Wild, Predicting protein β-sheet contacts using a maximum entropy-based correlated mutation measure, Bioinformatics 29 (2013) 580–587, https://doi.org/10.1093/bioinformatics/btt005.
[41] D.T. Jones, D.W.A. Buchan, D. Cozzetto, M. Pontil, PSICOV: precise structural contact prediction using sparse inverse covariance estimation on large multiple sequence alignments, Bioinformatics 28 (2012) 184–190, https://doi.org/10.1093/bioinformatics/btr638.
[42] S. Seemayer, M. Gruber, J. Söding, CCMpred:fast and precise prediction of protein residue–residue contacts from correlated mutations, Bioinformatics 30 (2014) 3128–3130, https://doi.org/10.1093/bioinformatics/btu500.
[43] A. Subramani, C. a. Floudas, β-sheet Topology Prediction with High Precision and Recall for β and Mixed α/β Proteins, PLoS One 7 (2012) e32461, , https://doi.org/10.1371/journal.pone.0032461.
[44] S.B. Needleman, C.D. Wunsch, A general method applicable to the search for similarities in the amino acid sequence of two proteins, J. Mol. Biol. 48 (1970) 443–453, https://doi.org/10.1016/0022-2836(70)90057-4.
[45] R.W. Floyd, Algorithm 97: shortest path, Commun. ACM 5 (1962) 345, https://doi.org/10.1145/367766.368168.

Toktam Dehghani received her B.S. and M.S. degree in Computer Engineering, with a concentration in Evolutionary Algorithms and Data Mining, in 2005 and 2010, respectively. She is currently earning a Ph.D. degree in the field of Computational Biology at the Department of Computer Engineering, Ferdowsi University of Mashhad, Iran. She has published papers in several international conferences and journals. Her current research interests are computational biology, bioinformatics algorithms, and data mining algorithms.

Mahmoud Naghibzadeh has received MS and PhD degrees in Computer Science and Computer Engineering, respectively, from University of Southern California, USC, USA. He has taught Undergraduate and Graduate courses at USC and University of South Florida, USA and is now a full professor at Ferdowsi University of Mashhad, Iran where he is also the director of Knowledge Engineering Research Group (KERG) laboratory. He was a visiting professor at University of California-Irvine, UCI, USA, in 1991 and a visiting professor at Monash University, Australia, in 2003–2004. His research interests include scheduling aspects of real-time systems, Grid, Cloud, Multiprocessors, Multicores, and GPGPUs and also Bioinformatics algorithms, especially Genomics and Proteomics. He has published numerous papers as well as eight books and has been the chairman of two international conferences and technical chair of many others.

Mahdie Eghdami has received the B.S. and M.S. degrees in computer engineering from the Ferdowsi University of Mashhad, Mashhad, Iran, in 2012 and 2015, respectively. She is currently a Ph.D. student at the Ferdowsi University of Mashhad. So far, she has been the coauthor of several papers in the field of protein β-sheet structure prediction. Her main research interests are NGS data analysis, protein structure prediction and pattern recognition.