Learning Deep Architectures for Protein Structure Prediction

Kyungim Baek
Department of Information and Computer Sciences, University of Hawai`i at Mānoa
1680 East-West Road, Honolulu, HI 96822, U.S.A.
E-mail: kyungim@hawaii.edu; Tel: +1.808.956.8560; Fax: +1.808.956.3548

Abstract

Protein structure prediction is an important and fundamental problem for which machine learning techniques have been widely used in bioinformatics and computational biology. Recently, deep learning has emerged as a new active area of research in machine learning, showing great success in diverse areas of signal and information processing studies. In this article, we provide a brief review on recent development and application of deep learning methods for protein structure prediction. The objective of this review is to motivate and facilitate deep learning studies for addressing problems in bioinformatics and computational biology where interesting avenues of research can emerge.

1 Introduction

Protein structure prediction is a long-established problem in bioinformatics and computational biology. The goal of protein structure prediction is to determine the 3D structure of a protein from its amino acid sequence. Although understanding protein structure is highly important in various areas of research (e.g. medicine and biotechnology), experimental determination of the complete protein structure is an expensive, time-consuming and laborious task with many technical challenges [41]. Moreover, experimental methods are extremely inefficient in this new era of data deluge prompted by the advancement of high-throughput sequencing technology and information sharing [7]. Therefore, effective and efficient computational methods for protein structure prediction are needed to mitigate these problems and help advance research in protein structure analysis and related fields of study.

Predicting protein 3D structure directly from its amino acid sequence is a daunting challenge, especially when there are no known template structures. For this reason, much research efforts have been made on predicting protein properties such as secondary structure, residue-residue contacts, and disorder regions that can be used to facilitate 3D structure prediction. Among the proposed approaches, machine learning methods that can extract important information out of abundant data have been proven to be useful for such prediction tasks [7].

Recently, deep learning architectures has emerged and has been evolving rapidly in the field of machine learning. The key aspect of learning deep architectures is the multiple levels of hierarchical non-linear information processing, which enables learning of successively higher and more abstract representations of the data. The effective learning of hierarchical representations using complex, non-linear functions became possible owing to the rapid increase of available data for training and the advance of hardware acceleration technologies (e.g., general purpose GPUs and custom FPGA-based accelerators) [10]. The powerful modeling and representational capability of deep architectures has the great potential to efficiently and effectively describe the highly non-linear and complex interactions and/or structures, which often arise in many biological problems.

However, compared to the vast amount of literature on theory and applications of deep learning in diverse areas of information processing, relatively few studies of deep learning applications in bioinformatics or computational biology have been presented during the past few years. These studies illustrate how various deep architectures and learning algorithms can be utilized to deal with complex relationships in the biological data, which is a critical factor in addressing important biological questions. In this article, we focus our review on the studies of learning deep architectures for protein structure prediction, a problem domain where majority of deep learning approaches have been used in bioinformatics. Our goal is to present to the bioinformatics and computational biology research communities the recent, state-of-the-art trend in machine learning and the current state of its application in the fields so as to motivate and facilitate research in this promising direction.

2 Deep Learning Background

Deep learning (also known as representation learning or feature learning) refers to machine learning techniques that attempt to learn models consisting of many layers of non-linear information processing for hierarchical feature representation. The concept of deep models has been around almost 30 years (e.g., multilayer perceptron [37]). However, deep learning became one of the key areas of research in machine learning and has been successfully applied in diverse areas of study after 2006 when an efficient, unsupervised learning algorithm to train deep models was introduced [22, 23]. In this section, we briefly describe three models of deep learning widely used in various applications. These models are also the basis of the approaches for protein structure prediction described in the next section.

2.1 Deep Belief Networks

A deep belief network (DBN) is a probabilistic generative model formed by stacking restricted Boltzmann machines (RBMs) [22]. An RBM consists of one layer of visible units and one layer of hidden units, where the two layers have symmetric connections between them but there are no connections between units within the same layer. RBMs can be trained by maximizing the likelihood of the training data using learning algorithms such as contrastive divergence [21] or stochastic maximum likelihood [40]. The stacking procedure is done by using activation probabilities of the hidden units of a learned RBM as the training data for another RBM in one layer up. It is shown that the likelihood of the training data improves as each layer is added to the DBN [22]. When applied to classification, a final layer of units associated with class labels can be added and the entire network can be further fine-tuned using a gradient-based, supervised learning.

2.2 Stacked Auto-encoder

An auto-encoder is a neural network where the target values are set to be the same as the inputs. It has one or more hidden layers that encode compressed representations of the input if their dimension is smaller than the input dimension. When the dimension of the hidden layer is larger than the input layer, it maps the input to a higher dimensional space where interesting structure in the data can be captured. In this case, a sparsity constraint is imposed on the hidden layer to prevent the auto-encoder from learning trivial identity mapping.

A stacked, or deep, auto-encoder consists of multiple layers of (sparse) auto-encoders, which can be trained using the greedy layer-wise approach [2]. In the layer-wise training, the first auto-encoder with single hidden layer is trained on the raw input data. Then, the learned feature representation (i.e. activations of the hidden units) is used as the input to the auto-encoder on the second layer. This learning process is repeated for subsequent layers and, when the layer-wise pre-training is complete, fine-tuning of the resulting network can be performed using back-propagation to improve its performance.

2.3 Convolutional Neural Networks

Convolutional neural network (CNN) is a discriminative deep architecture composed of multiple convolutional and pooling layers. The architecture of CNN is inspired by the simple and complex cell model of visual cortex. In the convolutional layer, the input is convolved with trainable filters and biases to compute local features like simple cells. Then, similar to complex cells, the pooling layer combines the neighboring features through weighted sum and a nonlinear transformation, producing a more abstract set of features of reduced dimension. The convolutional and pooling layers can be alternated multiple times in a CNN, forming a hierarchical deep architecture. The outputs of the final pooling layer become the input to a conventional neural network that produces the CNN's final output. Training of CNNs is typically performed by gradient-based supervised learning [26, 27]. Convolutional structure is has been also applied to DBN [28].

3 Deep Learning Applications to Protein Structure Prediction

3.1 Secondary Structure Prediction

Protein secondary structures are local structural conformations, such as α-helices, β-pleated sheets and loops, resulted from hydrogen bonding. These secondary structures are the intermediate substructures that can lead to higher level of protein structures with greater complexity. Therefore, accurate prediction of protein secondary structure can yield great benefit to understanding of the 3D structure (tertiary structure) of protein and even the larger protein complex (quaternary structure). It may also provide important information for the study of protein functions.

Since the early attempt at predicting secondary structures using a Bayesian classifier [20] machine learning approaches have been proven to be effective for the task, especially for prediction with no known template structure or homology models (i.e. *ab initio* prediction). Many of the successful machine learning approaches are based on the Artificial Neural Networks (ANNs) (e.g., [9, 25, 33, 36]). In the line of ANN-based approach, studies utilizing deep hierarchical architectures and deep learning methods have been proposed recently in attempts to further improve the prediction performance.

In [35], Qi et al. present a unified architecture combining a deep neural network with multi-task learning for simultaneously predicting various local protein properties including secondary structures. The architecture consists of feature extraction layers followed by multiple classic neural network layers, where all layers are shared across different tasks except the last layer which comprises multiple task-specific neural networks. In this study, eight tasks are learned simultaneously by multi-task learning using back-propagation, which captures and utilizes intrinsic relationship of the related tasks [5].

The experimental results show that learning related tasks in parallel can improve generalization performance of the predictor and the performance improvement is significant especially when very little training data are

available. For example, the performance on secondary structure prediction task improves much more for the small CB513 dataset [8] compared to the lager dataset formed from the DSSP database [24]. The deep, multifaceted architecture proposed by Qi et al. provides a flexible model for prediction system as it can be easily adapted to or incorporate additional tasks.

More recently, Zhou and Troyanskaya [44] introduce a novel deep learning approach for local secondary structure prediction, which utilizes supervised generative stochastic network (GSN) [3]. The two main aspects incorporated by the proposed approach are *i*) construction of supervised GSN to infer the distribution for reconstruction (i.e. conditional distribution of the output label given the input data), and *ii*) the use of convolutional architecture for GSN for efficient learning on high-dimensional, large-sized data to build deep hierarchical representations. By using multi-layer convolutional GSN, the proposed approach is able to capture global information of the input sequences.

For secondary structure prediction, Zhou and Troyanskaya apply the convolutional GSN to 8-state problem, which is more challenging task than the 3-state prediction problem addressed by most of previous studies. Their model predicts the secondary structure of four major, highly represented states (α-helix, β-strand, loop/irregular, and β-turn) with high accuracy. The predictive performance is low for the rarely appearing structures (bend, 3_{10}-helix, β-bridge, and π-helix) mainly due to the severely unbalanced labels caused by their limited presence in the data. The problem with unbalanced dataset should be addressed in the future to improve the prediction on 8-states problem. Results on CB513 benchmark dataset [8] also show that the convolutional GNS model achieves the state-of-the-art performance on 8-state prediction.

Spencer et al. [39] present rigorous experimental evaluation of using DBNs for *ab initio* secondary structure prediction, which provides valuable insights on application of deep learning to the secondary structure prediction. The deep network employed in this study is the standard DBN consisting of multiple RBM layers and a conventional neural network added at the last layer. A semi-supervised learning algorithm is used to train the DBN, where unsupervised learning is first applied to initialize the network's weighs, followed by a supervised learning for fine-tuning of the network. A classic back-propagation algorithm is used for the supervised fine-tuning process. Two DBNs are trained independently to make preliminary predictions that are subsequently fed into the third DBN for final, refined prediction. The preliminary predictions are the inputs used to train the third DBN.

In the evaluation, several experiments are performed to investigate different aspects involved in the DBN approach such as the number of hidden layers in the DBN and different combinations of input features. The three input features used in the experiments are amino acid residues, the position-specific scoring matrix (PSSM) profile calculated by PSI-BLAST, and the Atchley factors [1]. The results illustrate that the use of deeper network does improve the prediction performance, and the combination of PSSM profile and the Atchley factors produces the most accurate prediction among all combinations of features evaluated. Furthermore, comparative studies are performed with exiting machine learning based secondary structure predictors. The results reveal that the DBN model makes more consistent predictions and suggest that it may provide more useful information for 3D structure predictions as it produces better score of segment overlap measure [43] than the other predictors.

3.2 Residue-Residue Contact Prediction

Predicting which residues in a protean sequence are in contact with each other in the 3D structure is an important task with significant impacts on elucidating the protein folding process and understanding 3D structure as well as a functional mechanism of a protein. However, *ab initio* residue-residue contact prediction is a challenging problem and the average accuracies of sequence-based predictors are in about 20-35% range for long-range contacts, which is far from reliable performance to be useful for 3D structure prediction.

Most of the advanced contact predictors are based on well-known machine learning methods such as support vector machines [6], hidden Markov models [4], and neural networks [19, 34, 38, 42]. Since deep learning has emerged as a new active area of research in machine learning recently, deep architectures have been applied to the residue-residue contact prediction problem to stimulate further progress in addressing the problem.

One of the two major approaches using deep architectures is presented by Di Lena et al. [11, 12]. In [11], they introduce a deep spatio-temporal neural network (DST-NN) architecture that utilizes both spatial and temporal features to predict a protein residue-residue contact map. The DST-NN architecture consists of multiple levels of a set of neural networks. All networks in the architecture share the same configuration having a single hidden layer and a single output unit for prediction of the contact probability for a residue-residue pair. The weights are also shared among all the networks in the same level.

The novel aspect of DST-NN is the use of temporal features for training in addition to the spatial features commonly used in contact map prediction. Temporal features consist of the contact probabilities estimated by the neural networks at the previous level within a neighborhood of a residue pair under consideration. These features essentially capture the spatial correlation between

residue contacts, which has not been exploited before. Later, they also extend the spatial features to include coarse contact orientation features and the alignment predictions between contacting secondary structural elements [12].

The DST-NN is trained level-wise by a supervised learning where the training process moves upward in the network hierarchy. During training, the spatial features are static in the sense that they are identical across all levels while the temporal features are formed from the outputs of the neural networks at the previous level. By integrating the contact map predictions made in the previous level through temporal features, the proposed DST-NN keeps refining the prediction as it moves up to the higher levels. The experimental results show significant improvement by the DST-NN over plain three-layer neural network and the 2D recurrent neural network models, two of the most widely used machine learning approaches for contact prediction showing the state-of-the-art performance.

Besides the improved performance in contact prediction, a strong point of the DST-NN architecture is its generality/flexibility in that the basic learning component – the neural network – can be replaced by other learning models for prediction. Also, the architecture can be easily applied to other types of problems where both spatial and temporal components play an important role.

Another application of a deep architecture to contact prediction is presented by Eickholt and Cheng [14]. They train multiple deep networks (DNs) in semi-supervised fashion, where layer-wise weight initialization by unsupervised learning is followed by fine-tuning of the entire network using supervised learning. Then, using a modified version of AdaBoost, the trained DNs are combined to construct boosted ensembles. The final prediction produced by a boosted ensemble is the sum of the weighted outputs of the component DNs. The large boosted ensembles of DNs are trained efficiently using GPUs and CUDA technology. The resulting sequence-based contact predictor is called DNcon.

DNcon is shown to be competitive with the two best sequence-based contact predictors (SVMcon and ProC_S3) in the ninth round of Critical Assessment of protein Structure Prediction (CASP) experiments [32]. In their successive study, Eickholt and Cheng [15] performed further analysis and evaluation of DNcon in comparison with the sequence-based, template-free residue-residue contact predictors from CASP10 [31] experiments. Overall, DNcon is among the best predictors under the standard residue level assessment metric used in the CASP. Furthermore, they propose novel, coarser level assessment metrics that consider the neighborhood of predicted contact to evaluate how well the predictions indicate areas of interaction. DNcon performs the best under the new metrics with 66% accuracy considering two residue neighborhoods for long-range contacts. The results explain that, although the residue level accuracy is low, the predicted contacts can still be useful for protein structure prediction. DNcon is also shown to be robust under the changes of its underlying architecture.

3.3 Other Applications

Deep neural networks have been exploited to address other problems related to protein structure prediction. One of the problems is to predict protein disorder regions, where a stable 3D structure is absent in the protein's native, functional state. Eickholt and Cheng [16] present a novel sequence-based protein disorder predictor, called DNdisorder. Like the DNcon described in the previous section, DNdisorder consists of a number of boosted ensembles of deep networks (DNs). RBMs are used in each deep, multi-layer neural network to initialize the weights of each layer, and then the entire DN is fine-tuned using the back-propagation algorithm minimizing the cross-entropy error. The boosted ensembles are produced by training a series of DNs using a modified version of AdaBoost. The use of CUDA and GPU enables training of very large, deep networks within reasonable amount of time. Although DNdisorder performs comparably and competitively against the existing state-of-the-art disorder predictors, it is not suitable for genomic scale analysis due to the excessive computational cost caused by the construction of sequence profiles using PSI-BLAST.

More recently, another application of deep neural network has been reported on the prediction of protein backbone inter-residue angles by Lyons et al. [30]. The protein backbone structure provides useful information for *ab initio* protein structure prediction. The two inter-residue angles used to represent the backbone structure in their study are the angle between the three consecutive Cα atoms (θ) and a dihedral angle rotated about the neighboring Cα bond (τ). These angles reflect the local conformation of the backbone over three to four neighboring residues while the torsion angles about the N-Cα bond (φ) and the Cα-C bond (ψ) are limited to a single residue. Since the secondary structure involves more than three residues θ and τ angles can provide local structure information complementary to φ and ψ angles and the secondary structures.

The work of Lyons et al. is the first application of a deep learning approach to sequence-based θ and τ prediction. Their approach utilizes a stacked sparse auto-encoder deep neural network consisting of 651 input units, three hidden layers with 150 units in each layer, and the four output units for sin(θ), cos(θ), sin(τ), and cos(τ). The network is trained through unsupervised weight initialization followed by supervised learning for fine-tuning. The results of a series of experimental analysis illustrate that the direct prediction of θ and τ angles shows better accuracy than the indirect estimation based on the φ

Table 1: Summary of Deep Learning Approaches to Protein Structure Prediction

Task	Approach	Learning Algorithm
Secondary Structure Prediction	Feature extraction layers connected to a deep stack of conventional neural networks [35]	Multi-task learning using back-propagation
	Deep belief network [39]	Unsupervised pre-training + supervised fine-tuning using back-propagation
	Multi-layer convolutional generative stochastic network [44]	Multi-task learning using back-propagation
Residue-Residue Contacts Prediction	Multi-level stacking of a set of neural networks [11, 12]	Level-wise, hierarchical supervised leaning using back-propagation
	Boosted ensembles of deep networks with restricted Boltzmann machines [14, 15]	Unsupervised pre-training + supervised fine-tuning + boosting
Disordered Region Prediction	Boosted ensembles of deep networks [16]	Unsupervised pre-training + supervised fine-tuning + boosting
Backbone Dihedral Angle Prediction	Stacked sparse auto-encoder [30]	Unsupervised pre-training + supervised fine-tuning

and ψ angles predicted by SPINE-X [18], a secondary structure predictor which also estimates the backbone torsion angles. The local fragment structure constructed from the predicted angles appears to be quite close to the corresponding native fragment structure. Therefore, it can be a good alternative to the homolog-based approaches for fragment structure generation. Also, the predicted θ and τ angles can provide useful information for template-free structure prediction as well as template-based prediction and model validation.

4 Conclusion

In this brief review, we have tried to give an overview of deep learning approaches applied to protein structure prediction problems summarized in Table 1. Overall, the predictive performance of deep learning methods is either improved upon or competitive with the existing state-of-the-art predictors. Also, deep learning approaches often provide more robust and reliable predictions.

Application of deep learning in bioinformatics and computational biology is in its early stage, continuously demonstrating its potential in addressing diverse range of problems – e.g., gene cluster detection [13], cancer type analysis [17], and splicing pattern prediction [29] – in additions to the structure prediction problems discussed in this article. As the available biological data becomes more abundant than it is today, generating great challenges and opportunities, powerful modeling and representational capability of deep learning approaches will continue to pave the road for interesting avenues of research in the field of bioinformatics and computational biology.

References

[1] William Atchley, Jieping Zhao, Andrew Fernandes, and Tanja Drüke. Solving the protein sequence metric problem. *Proceedings of the National Academy of Sciences of the United States of America*, 102(18):6395-6400, 2005.

[2] Yoshua Bengio, Pascal Lamblin, Dan Popovici, and Hugo Larochelle. Greedy layer-wise training of deep networks. *Advances in Neural Information Processing Systems (NIPS) 19*, pages 153-160, 2006.

[3] Yoshua Bengio, Eric Thibodeau-Laufer, Guillaume Alain, and Jason Yosinski. Deep generative stochastic networks trainable by backprop. *Journal of Machine Learning Research*: *W&CP*, 32(1):226-234, 2014.

[4] Patrik Björkholm, Pawel Daniluk, Andriy Kryshtafovych, Krzysztof Fidelis, Robin Andersson, and Torgeir R. Hvidsten. Using multi-data hidden Markov models trained on local neighborhoods of protein structure to predict residue-residue contacts. *Bioinformatics*, 25(10):1264-1270, 2009.

[5] Rich Caruana. Multitask learning. *Machine Learning*, 28:48-75, 1997.

[6] Jianlin Cheng and Pierre Baldi. Improved residue contact prediction using support vector machines and a large feature set. *BMC Bioinformatics*, 8:113, 2007.

[7] Jianlin Cheng, Allison N. Tegge, and Pierre Baldi. Machine learning methods for protein structure prediction. *IEEE Reviews in Biomedical Engineering*, 1:41-49, 2008.

[8] James Cuff and Geoffrey Barton. Evaluation and improvement of multiple sequence methods for protein secondary structure prediction. *Proteins*, 34:508–519, 1999.

[9] James Cuff and Geoffrey Barton. Application of multiple sequence alignment profiles to improve protein secondary structure prediction. *Proteins*, 40(3):502-511, 2000.

[10] Li Deng and Dong Yu. Deep learning: methods and applications. *Foundations and Trends in Signal Processing*, 7(3-4):197-387, 2013.

[11] Pietro Di Lena, Ken Nagata, and Pierre Baldi. Deep spatio-temporal architectures and learning for protein structure prediction. *Advances in Neural Information Processing Systems (NIPS) 25*, pages 521-529, Lake Tahoe, Nevada, December 3 – 6, 2012.

[12] Pietro Di Lena, Ken Nagata, and Pierre Baldi. Deep architectures for protein contact map prediction. *Bioinformatics*, 28(19):2449-2457, 2012.

[13] Nan Du, Xiaoyi Li, Yuan Zhang, and Aidong Zhang. Detecting mutual functional gene clusters from multiple related diseases. In *Proceedings of IEEE International Conference on Bioinformatics and Biomedicine*, pages 159-6164, Shanghai, China, 2013.

[14] Jesse Eickholt and Jianlin Cheng. Predicting protein residue-residue contacts using deep networks and boosting. *Bioinformatics*, 28(23):3066-3072, 2012.

[15] Jesse Eickholt and Jianlin Cheng. A study and benchmark of DNcon: a method for protein residue-residue contact prediction using deep networks. *BMC Bioinformatics*, 14(Suppl 14):S12, 2013.

[16] Jesse Eickholt and Jianlin Cheng. DNdisorder: predicting protein disorder using boosting and deep networks. *BMC Bioinformatics*, 14:88, 2013.

[17] Rasool Fakoor, Faisal Ladhak, Azade Nazi, and Manfred Huber. Using deep learning to enhance cancer diagnosis and classification. In *Proceedings of the 30th International Conference on Machine Learning*, Atlanta, Georgia, USA, 2013.

[18] Eshel Faraggi, Tuo Xhang, Yuedong Yang, Lukasz Kurgan, and Yaoqi Zhou. SPINE X: Improving protein secondary structure prediction by multi-step learning coupled with prediction of solvent accessible surface area and backbone torsion angles. *Journal of Computational Chemistry*, 33(3):259-267, 2012.

[19] Piero Fariselli, Osvaldo Olmea, Alfonso Valencia, and Rita Casadio. Prediction of contact maps with neural networks and correlated mutations. *Protein Engineering Design & Selection*, 14(11):835-843, 2001.

[20] Jean Garnier, David J. Osguthorpe, and Barry Robson. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *Journal of Molecular Biology*, 120(1):97-120, 1978.

[21] Geoffrey E. Hinton. Training products of experts by minimizing contrastive divergence. *Neural Computation*, 14:1771-1800, 2002.

[22] Geoffrey E. Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural Computation*, 18:1527-1554, 2006.

[23] Geoffrey E. Hinton and Ruslan R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504-507, 2006.

[24] Wolfgang Kabsch and Christian Sander. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577-2637, 1983.

[25] Sol Katzman, Christian Barrett, Grant Thiltgen, Rachel Karchin, and Kevin Karplus. PREDICT-2ND: a tool for generalized protein local structure prediction. *Bioinformatics*, 24(21):2453-2459, 2008.

[26] Yann LeCun, Bernhard Boser, John S. Denker, Donnie Henderson, Richard E. Howard, Wayne Hubbard, and Lawrence D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541-551, 1989.

[27] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278-2324, 1998.

[28] Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Y. Ng. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In *Proceedings of the 26th International Conference on Machine Learning*, pages 609-616, Montréal, Canada, 2009.

[29] Michael K. K. Leung, Hui Yuan Xiong, Leo J. Lee, and Brendan J. Frey. Deep learning of the tissue-regulated splicing code. *Bioinformatics*, 30:i121-i129, 2014.

[30] James Lyons, Abdollah Dehzangi, Rhys Heffernan, Alok Sharma, Kuldip Paliwal, Abdul Sattar, Yaoqi Zhou, and Yuedong Yang. Predicting backbone Cα angles and dihedrals from protein sequences by stacked sparse auto-encoder deep neural network. *Journal of Computational Chemistry*, 35(28):2040-2046, 2014.

[31] Bohdan Monastyrskyy, Daniel D'Andrea, Krzysztof Fidelis, Anna Tramontano, and Andriy Kryshtafovych. Evaluation of residue-residue contact prediction in CASP10. *Proteins*, 82 (Suppl 2):138-153, 2014.

[32] Bohdan Monastyrskyy, Krzysztof Fidelis, Anna Tramontano, and Andriy Kryshtafovych. Evaluation of residue-residue contact predictions in CASP9. *Proteins*, 79(Suppl 10):119-125, 2011.

[33] Thomas Nordahl Petersen, Claus Lundegaard, Morten Nielsen, Henrik Bohr, Jakob Bohr, Søren Brunak, Garry P. Gippert, and Ole Lund. Prediction of protein secondary structure at 80% accuracy. *Proteins*, 41:17-20, 2000.

[34] Gianluca Pollastri and Pierre Baldi. Prediction of contact maps by GIOHMMs and recurrent neural networks using lateral propagation from all four cardinal corners. *Bioinformatics*, 18(Suppl 1):S62-S70, 2002.

[35] Yanjun Qi, Merja Oja, Jason Weston, and William Stafford Noble. A unified multitask architecture for predicting local protein properties. *PLoS ONE*, 7(3):e32235, 2012.

[36] Ning Qian and Terrence J. Sejnowski. Predicting the secondary structure of globular proteins using neural network models. *Journal of Molecular Biology*, 202(4):865-884, 1988.

[37] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(9):533-536, 1986.

[38] George Shackelford and kevin Karplus. Contact prediction using mutual information and neural nets. *Proteins*, 69(Suppl 8):159-164, 2007.

[39] Matt Spencer, Jesse Eickholt, and Jianlin Cheng. A deep learning network approach to *ab initio* protein secondary structure prediction. *IEEE Transactions on Computational Biology and Bioinformatics*, TCBB-2014-03-0071, 2014.

[40] Tijmen Tieleman. Training restricted Boltzmann machines using approximations to the likelihood gradient. In *Proceedings of the 25th International Conference on Machine Learning*, pages 1064-1071, Helsinki, Finland, 2008.

[41] Sitao Wu and Yang Zhang. Protein structure prediction. *Bioinformatics: Tools and Applications*, D. Edwards, D. Hansen and J. Stajich (Eds.), pages 225-242, Springer-London, 2009.

[42] Bin Xue, Eshel Faraggi, and Yaoqi Zhou. Predicting residue-residue contact maps by a two-layer, integrated neural-network method. *Proteins*, 76:176-183, 2009.

[43] Adam Zemla, Česlovas Venclovas, Krzysztof Fidelis, and Burkhard Rost. A modified definition of Sov, a segment-based measure for protein secondary structure prediction assessment. *Proteins*, 34(2):220-223, 1999.

[44] Jian Zhou and Olga G. Troyanskaya. Deep supervised and convolutional generative stochastic network for protein secondary structure prediction. *Journal of Machine Learning Research*: *W&CP*, 32(1):754-762, 2014.