International Journal of Information & Computation Technology.
ISSN 0974-2239 Volume 3, Number 1 (2013), pp. 33-43

http://www. irphouse.com

Literature Survey on the Prediction of Secondary Structure of Proteins Using Radial Basis Function Neural Networks (RBFNN) and Support Vector Machines (SVM)

Dr. K. Meenaand Dr. M. Manimekalai***

***Vice Chancellor, Bharathidasan University,*
Tiruchirappalli, Tamil Nadu, India
***Director, Department of MCA, Shrimati Indira Gandhi College,*
Tiruchirappalli, Tamil Nadu, India

Abstract

The Protein structure prediction has been an active research area for the last 40 years or so. The technical progress in computational Molecular Biology during the last decades has contributed significantly to the progress we see today. The major goal of predicting Protein structures underpins the correct assumption that three dimensional structures confer protein function. The linear Amino Acid sequences must transform to nonlinear Secondary Structures and then to Tertiary and Quaternary Structures that are responsible for biological functions. Biological functions may remain similar or change in the related organisms through the evolutionary process. By considering the importance of the prediction of secondary structure of protein a detailed literature study of the same using Radial Basis Function Neural Networks (RBFNN) AND Support Vector Machines (SVM) has been reviewed in this paper.

Keywords: Radial Basis Function Neural Networks, Support Vector Machines, Secondary Structures, Tertiary Structure and Quaternary Structures

1. INTRODUCTION

Methods predicting Protein secondary structure have improved substantially in the 90's through using evolutionary information taken from the divergence of Proteins in the same structural family. Recently, the evolutionary information resulting from

improved searches and large databases has again boosted prediction accuracy by more than four percentage points to its current height around 76% of all residues predicted correctly in one of the three states Helix, Strand, other.

2. REVIEW OF LITERATURE OF SECONDARY STRUCTURE PREDICTION USING VAROIUS METHODS

David T. Jones, [1] proposed two-stage neural network to predict protein secondary structure on the basis of Position Specific Scoring Matrices (PSSM) generated by PSI-BLAST. PSI-BLAST is a very powerful sequence searching method. This generates sequence profiles as part of the search process, and here the idea of intermediate PSI-BLAST is explored, as a direct input to a secondary structure prediction method rather than extracting the sequences, and producing an explicit multiple sequence alignment as a separate step. BLAST Programs offer the following programs,

- xblastn: Search a Nucleotide database with a Nucleotide query
- blastp: Search Protein database with a Protein query
- tblastn: Search a six-frame dynamic translation of a Nucleotide database with a protein query
- blastx: Search a Protein database with a six-frame translation of a Nucleotide query sequence.
- tblastx: Search a six-frame translation of a Nucleotide database with a six-frame translation of a Nucleotide query sequence.

Wotton and Federhen [2] stated the iterative nature of the PSI-BLAST algorithm which increases the sensitivity of the data base by detecting homologous matches with relatively low sequence identity. In a PSI-BLAST search, a PSSM generates a scoring system that is specific to the group of matches detected using the initial query sequence. The subsequent PSI-BLAST iteration using a customized matrix extends to the result that allow the detection of more distantly related homology.

In order to maximize the effectiveness of PSI-BLAST, Stephen F. Altschul et al., Thomas L. Madden [3] deliberately have the idea that, BLAST programs are widely used tools for searching Protein and DNA databases for sequence similarities.

The original BLAST program seeks short word pairs whose aligned score is at least T. The consumption of the processing time is high for the default T value. A new 'two-hit' method by two non-overlapping word pairs on the same diagonal, and within a distance of one another can be created.

In order to achieve comparable sensitivity, the threshold parameter T must be lowered. Then it generates gapped alignments, finally concludes by BLAST searches and may be iterated, with a PSSM generated from significant alignments. To compare the performance of the new gapped version of BLAST and its PSI-BLAST extension to that of the Smith-Waterman algorithm, it can be suggested that PSI-BLAST iteration still runs faster than the original BLAST, and 40 times faster than Smith–Waterman, but in many cases seems to be more sensitive. But by multiple iterations it can yield better results.

Leong Lee et al., [4] gave an approach on the basis of Protein Secondary Structure Prediction by Using Parallelized Rule Induction from Coverings.

The recent breakthrough in 3D structure prediction is to combine the multiple sequence alignment with artificial intelligence to predict the Protein structure Accuracy, Q3 which is a major factor of this method, here we have only 75%, but in previous analysis the accuracy Q3 was rated as 80.3% in RT-RICO (Relaxed Threshold Rule Induction from COvering).

In Parallelized RT-RICO Q3 ranges 74.6%, which is higher than the consensus prediction accuracy of 72.9. Rost [5] suggests that although protein 3D structure is not achieved fully, then also the study is continuing in the same area. It is found that Secondary structure prediction, have accuracy of 70% threshold for all residues of a protein. This is done by combining multiple sequence alignment information with artificial intelligence algorithms.

The new thought raised by W. Kabsh and C. Sander [6] who performed the test by some prediction methods using proteins that is not in use for the development of the algorithms, found the accuracy for most of those methods decreased by 7 to 27%.

Taner Z. Sen et al., [7] put fourth about GOR V web server for Protein secondary structure prediction. This algorithm combines information theory, Bayesian statistics and evolutionary information. Here the accuracy, Q3 rates 73.5%.

Garnier et al., [8] gave an intro for GOR (Garnier–Osguthorpe– Robson). It uses both information theory and Bayesian statistics for predicting the secondary structure of Proteins.

Sung-Joon Park [9] had a look on Fragment-Based Protein Structure Prediction: these computational methods are categorized into two steps, Template-Based Prediction methods, such as homology modeling or fold recognition method, which have fast and effective analysis of Protein structure and function.

Fragment Assembly (FA) acts as backbone for de novo Protein structure prediction. In contrast, GA has feasibility to consider two conformations as a target and a template.

Probabilistic Fragment Selection (PRO) [10] [11] cannot be fed in the prediction of new folds. This study employed the framework of GA.

W.R. Pearson and D.J. Lipman [12] proposed a Protein structure prediction by pair wise comparison. This method plays a vital role in the application of bioinformatics in the biological sciences, since they allow making predictions about a protein's function, structure, and evolution.

HHsearch is a typical program for protein sequence searching that is free for non-commercial usage. HHpred is a free server for protein structure prediction based on the HHsearch approach by Soding et al. [13]. HHpred/HHsearch technique is among the most popular methods for protein structure prediction and the detection of distantly related sequences, having been cited over 450 times.

HHpred/HHsearch denotes query and database proteins by profile Hidden Markov Models (HMMs), an extension of sequence profiles which also record position-specific amino acid insertion and deletion frequencies.

The output of HHpred and HHsearch is a ranked list of database matches and the pairwise query-database sequence alignments. In the CASP7[14] blind protein

structure prediction experiment, HHpred5 was ranked 2nd out of 68 automatic structure prediction servers, while being more than 50 times faster than the best 20 servers.

Rost and Wang [15] explore these problems and suggest some alternative measures of predictive success based on secondary structure segment overlap.

In Structural class prediction, if the structural class of a protein (alpha, beta, alpha / beta, or alpha + beta) is known then the secondary structure prediction problem is simplified

Brocchieri L. Karlin S. [16] gave a new thought about Forces Controlling Protein Structure of Hydrogen Bonding Polypeptides contains numerous proton donors and acceptors both in their backbone and in the R-groups of the amino acids.

Daniel C. Berwick stated three methods for Protein prediction such as 1D, 2D and 3D. 1D includes in secondary structure and it is solvent accessibility, which residues are exposed to water, which are buried and transmembrane helices. Prediction in 2D includes inter-residue/strand contacts. Then finally prediction in 3D, means homology modeling, fold recognition, molecular dynamics, fragment assembly and ab initio prediction. In 3D modeling technique, in homology modeling a query sequence Q, a database of protein structures find protein P such that structure of P has high sequence similar to Q and return P's structure as an approximation to Q's structure.

In fold recognition (threading) a query sequence Q, with a database of known folds which find fold F such that Q can be aligned with F in a highly compatible manner and return F as an approximation to Q's structure. Next one is fragment assembly. Here also a query sequence Q is stated, a database of structure fragments continues to find a set of fragments that Q can be aligned with in a highly compatible manner and return fragment assembly as an approximation to Q's structure and then finally molecular dynamics have a query sequence Q which use laws of Physics to simulate folding of Q.

Comparative modeling makes use of the fact that evolutionarily related proteins with similar sequences have similar structures. Sequence similarity is calculated by the percentage of identical residues at each position based on an optimal structural superposition. The structure similarity is very high in the so-called core regions, which characteristically consist of secondary structure elements such as α-helices and sheets. These secondary structures are connected by the Loop regions such as β turns.

Threading methods are very much similar to comparative modeling. Threading produces a list of scores by comparing a target sequence against a library of structural templates. The ab initio technique is a mixture of science and engineering. In most of the ab initio techniques, the two components are coupled together in such a way that a search function drives, and is driven by, the scoring function to find native-like structures.

K. Henrick in [17] discusses about the Quaternary Structure of Proteins. Many proteins contain 2 or more different polypeptide chains that are held in association by the same non-covalent forces that stabilize the tertiary structures of proteins.

Fischer et al [18] in 2001 gave a thought to blind and manual evaluation of structure prediction methods; mainly the plethora of protein prediction methods developed by different research groups all over the world is notified. The challenge is

to identify suitable methods that are truly superior in accurately predicting the protein structures of interest.

Protein structure prediction is wide-spread approach, and has been previously addressed using evolutionary algorithms, such as the Simple genetic algorithm (GA), messy GA (mga), fast messy GA (fmGA), and Linkage Learning GA (LLGA). However, past research used offs the shelf software such as GENOCOP, GENESIS, and mGA. Ruth Pachter in [19] reported the results of a modified fmGA, which is found to be "good" at finding semi-optimal solutions in a reasonable time.

2.2.1 Survey on Prediction Techniques using RBFNN Approach

Jing et al. in [20] putforth a technique for predicting the secondary structure of Proteins using Radial Basis Function Neural Networks (RBFNN). Their research proposed a new method based on RBFNNs for prediction of Protein secondary structure. In order to make their proposed algorithm comparable to other secondary structure prediction methods, they used the benchmark evaluation data set of 126 protein chains in their approach. In addition, they also analyzed how to use evolutionary information to improve the prediction accuracy.

P.K. Dash et al. in [21] proposed that Radial basis functional neural networks (RBFNN) provide an outstanding possibility for generating rules for solving pattern classification problems. One of the most important factors in RBFNN is finding out the center and spread. The work examines rules extracted from RBF networks trained by Particle swarm Optimization (PSO). The selection of the RBFNN centers, spreads and the network weights can be viewed as a system identification problem. The Simulation results using Radial Basis Functional Neural Networks (RBFNN) was applied to the PAT, WBC and IRIS data sets as a classification problem to illustrate the new knowledge extraction technique.

A Radial Basis Functional neural network (RBFNN) is trained to perform a mapping from an m-dimensional input space to an n-dimensional output space. RBFNN's can be used for discrete pattern classification, function approximation, signal processing, control, or any other application, which requires a mapping from an input space to an output space. Many recent developments of RBFNN and its applications can be found in Neuro computing special issues on RBFNN.

It is also demonstrated that the RBFNN network training is yet another fruitful application of Particle Swarm Optimization (PSO). The simulation is done using MATLAB v6.5 verifies that initialization of the centers through Particle Swarm Optimization provides better performance. Further research could focus on the application of Particle Swarm Optimization (PSO) training to RBFNN networks with alternative forms of the generator function.

Nicolaos B. Karayiannis [22] proposed an axiomatic approach for constructing radial basis function (RBF) neural networks. The approach results in a broad variety of admissible RBF models, including those employing Gaussian RBFs. The form of the RBFs is determined by a generator function. New RBF models can be developed according to the proposed approach by selecting generator functions other than exponential ones, which lead to Gaussian RBFs.

The performance of an RBF network is based on the number and positions of the RBFs, their shape, and the method used for learning the input-output mapping. The existing learning strategies for RBF neural networks can be categorized as,

- Strategies selecting the RBF centers randomly from the training data.
- Strategies employing unsupervised procedures for selecting the RBF centers.
- Strategies employing supervised procedures for selecting the RBF centers.

The success of a neural-network model depends rather strongly on its association with an attractive learning algorithm. For example, the popularity of feed forward neural networks with sigmoidal hidden units was to a larger extent due to the error in back propagation algorithm. On the other hand, the potential of RBF neural models for classification and function approximation was downgraded by the lack of effective and reliable learning algorithms for such models.

According to the axiomatic approach proposed for reformulating RBF neural networks, the development of admissible RBF models reduces to the selection of admissible generator functions that determine the form and properties of the RBF's. The reformulated RBF networks developed by linear and exponential generator functions are trained by gradient descent and perform considerably better than conventional RBF networks. Moreover, training reformulated RBF neural networks by gradient descent is not necessarily slow.

Susan C. White [23] described a methodology used to construct forecasts for the subset of 11 time series in the NN3 competition with Radial Basis Function Networks.

Chris Bishop [24] proposed that an important feature of RBFNNs is the existence of a fast, linear learning algorithm in a network capable of representing complex nonlinear mappings. Reasonable generalization in these networks requires that the network mapping be satisfactorily smooth.

John Tait et al in [25] described that Radial basis neural (RBF) networks provide an excellent solution to many pattern recognition and classification problems. RBF networks are a localist representation technique that enables the easy conversion of the hidden units into symbolic rules. They examined the quality and comprehensibility of rules extracted from RBF networks.

Radial basis function networks are a localist type of learning technique. Local learning systems generally contain elements that are responsive to only a limited section of the input space. This is quite different from the distributed approach of multi-layer perceptron networks (MLP). The local nature of RBF net-works makes them an ideal platform for performing rule extraction. The ability of rule extraction algorithms is examined to extract meaningful rules that describe the overall performance of a particular network.

J. Park et al. in [26] proved that RBF networks having one hidden layer are capable of universal approximation. They emphasized on the typical RBF networks, and the results show that a certain class of RBF networks with the same smoothing factor in each kernel node is broad enough for universal approximation.

2.2.2. Survey On Prediction Techniques Using SVM

Prediction of Protein structure began in 1960s when the first protein crystal structures

were available for detailed study. Protrero (1966) reported certain amino acids could be used to predict helices in myoglobin and hemoglobin. Schiffer and Edmunson in (1967) developed the helical wheel both to predict helical and, if a helix is present, to indicate the presence of a hydrophobic region.

Prediction of Protein structure has been a successful technique in the data mining approach. Several researchers have exploited this prediction process using many techniques like radial basis function Jing.

Learning methods like SVMs are used for the prediction technique which is based on the SCOP, in which protein domains are classified based on known structures and the evolutionary relationships and the principles that govern their 3-D structure [27].

Data mining concept is used in the prediction of Protein structure (Shi Liu et al.) which is constructed based on the rough set algorithms.

Bioinformatics techniques to Protein Secondary Structure (PSS) prediction is mainly based on the information available in amino acid sequences which uses the SVMs [28]. This shows higher accuracy of the prediction technique compared to other traditional methods.

Hu et al. in [29] proposed an approach of predicting the secondary structure of Proteins using SVM. The authors proposed the encoding scheme for training the SVM known as the PSSM.

A new approach for predicting the Protein structural class was proposed by Kurgan et al. in [30]. They proposed a new sequence representation, which is based on PSI-BLAST profile based collocation of AA pairs.

Kristin et al. in [31] projected an approach for Protein structure prediction. The protein energy landscape theory is used to achieve optimal energy functions for Protein structure prediction by means of simulated annealing. The analysis in this research takes advantage of a more complete statistical characterization of the Protein energy landscape and thereby improves on previous approximations.

Adam Krzyzak in [32] proposed a fast SVM training algorithm under the decomposition framework of SVM's algorithm by successfully integrating kernel caching, digest and shrinking policies and stopping conditions. Experiments on MNIST handwritten digit database have been conducted and the results show that the proposed approach is much faster than Keerthi et al.'s improved SMO, about 9 times. The promising scalability of the proposed technique can make it possible to apply SVM to a wide variety of problems in engineering.

Protein secondary structure carries data regarding local structural arrangements. Most of successful methods for predicting the secondary structure are based on multiple sequence alignment. However, the multiple alignments fail to achieve significant results when a protein sequence is characterized by low homology. To overcome this, Chao Chen et al in [33] proposed a novel method for prediction of secondary structure content through comprehensive sequence representation. This approach is featured by employing a SVM regressing system and adopting a different pseudo amino acid composition (PseAAC), which can partially take into account the sequence-order effects to represent Protein samples. It was observed from both the self-consistency test and the independent-dataset test that the trained SVM has remarkable power in grasping the relationship between the PseAAC and the content

of Protein secondary structural elements, including α-helix, β-strand, bridge, turn, bend and the rest random coil. Results obtained indicate that the present method may at least serve as an alternative to the existing predictors in this area.

Decoste et al. [34] found that the number of candidate support vectors during the early stage of training is much greater than that of final support vectors while many experimental results seem to show that SMO's time complexity can be approximated to about O(L:N), where n is the size of training set and L is the average number of candidate support vectors during iterations.

Effectively reducing L will have an important impact on the performance of SMO. Therefore, Decoste introduced "digest" idea to avoid this inefficiency and jump out of the full SMO iteration early as the candidate support vectors grow by a large number and switch SMO into "inbound" iteration to "digest" these candidate SV sets. The number of kernel re-evaluations is reduced by the heuristics. However, DeCoste et al.'s heuristics contain a lot of ad-hoc parameters and caching the entire rows of kernel matrix is still inefficient.

3. CONCLUSION

Problems remain with buried α-helices that comprise short runs of conserved hydrophobic amino acids. These often look like potential β-strands and can mislead both automatic and manual predictive methods, evolutionarily conserved residues and prediction. The improvements in the accuracy of secondary structure prediction that are seen when multiple alignments are used from the observation that positions in an alignment where the identity of the amino acid residue varies slowly during the course of evolution and are important to the stability of the fold or the protein function. Patterns of conservation can be discerned by eye, but ideally automatic protocols should be used to improve objectivity.

The earlier secondary structure prediction techniques use local information of a single sequence. These approaches suffered from the following drawbacks.

The three-state pre-residue accuracy (Q3) was about 65% which is very low.

Sheets were predicted at levels of 28-48% which is slightly better than random.

The predicted secondary structure segments were only half as long as the observed segments on average. Most of the available methods for prediction of the secondary structure are based on multiple sequence alignment. However, multiple alignments produce very low accurate results when a sequence comes from low homology.

Moreover, majority of the previous research have ignored the influence of residue conformational preference on structure prediction of Proteins. Furthermore, function of a Protein is determined by its three dimensional structure. Determination of protein structure required lab experiments which are time and cost unaffordable and complicated. At the same time as some experiments can take months or years to achieve and for some Proteins like membrane Proteins current experimental techniques are not proficient for structure determination.

REFERENCE

[1] David T. Jones, "Protein Secondary Structure Prediction Based on Position-specific Scoring Matrices, " University of Warwick, Coventry CV4 7AL, United Kingdom, 1999.

[2] Wootton, J.C. and Federhen, S., "Statistics of local complexity in amino acid sequences and sequence databases", Computers and Chemistry 17, 149-163 1993.

[3] Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schäffer1, Jinghui Zhang, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Received June 20, 1997; Revised and Accepted July 16

[4] L., Leopold, J.L., Frank, R.L. and Maglia, A.M., "Protein Secondary Structure Prediction Using Rule Induction from Coverings", Proceedings of IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (part of IEEE Symposium Series on Computational Intelligence 2009), Nashville, Tennessee, USA, pp. 79-86, 2009.

[5] Rost, B., "Rising accuracy of protein secondary structure prediction", D.Chasman, Ed., "Protein structure determination, analysis, and modeling for drug discovery", New York: Dekker, pp. 207–249, 2003.

[6] Kabsh, W. and Sander, C., "How good are predictions of protein secondary structure, " FEBS Letters, 155, pp. 179-182, 1983.

[7] Kloczkowski, A., Taner Z. Sen, "Combining the GOR V algorithm with evolutionary information for protein secondary structure prediction from amino acid sequence", Proteins, 49, 154–166, 2002.

[8] Garnier, J.J., Gibrat, J.F. and Robson, "GOR method for predicting protein secondary structure from amino acid sequence, Methods Enzymol", 266, 540–550, 1996.

[9] Sung-Joon Park, "A Study of Fragment-Based Protein Structure Prediction: Biased Fragment Replacement for Searching Low-Energy Conformation", 104-115, 2005.

[10] Karplus, K., Karchin, R., Draper, J., Casper, J., Mandel-Gutfreund, Y., Diekhans, M. and Hughey, R., "Combining local-structure, fold-recognition, and new fold methods for protein structure prediction", Proteins, 53:491-496, 2003

[11] Kolodny, R., Koehl, R., Guibas, L. and Levitt, M., "Small libraries of protein fragments model native protein structures accurately", J. Mol. Biol., 323:297-307, 2002.

[12] Pearson, W.R. and Lipman, D.J. Improved tools for biological sequence comparison proc. Natl. Acad. Sci. U.S.A., 85, 2444-2448, 1998.

[13] Söding, J, Biegert, A. and Lupas, A.N., "The HHpred interactive server for protein homology detection and structure prediction, " Nucleic Acids Research 33 ((Web Server issue)): W244–248, 2005.

[14] Battey, J.N., Kopp, J., Bordoli, L., Read, R.J., Clarke, N.D. and Schwede, T., "Automated server predictions in CASP7, " Proteins 69 (S): 68–82, 2007.

[15] Wang Z-X. "Assessing the accuracy of secondary structure". NatStruct Biol., 3:145–146. 1994.
[16] Luciano Brocchieri and Samuel Karlin, "How are close residues of protein structures distributed in primary sequence", Proc. Natl. Acad. Sci. USA Vol. 92, 12136-12140, December 1995.
[17] Henrick, K. and Thornton, J.M., "PQS: a protein quaternary structure file server", Trends. Biochem. Sci. 23:358-361, 1998.
[18] Fischer, D., Rychlewski, L., Elofsson, A., Pazos, F., Valencia, A., Rost, B., Ortiz, A.R. and Dunbrack, R.L.J., "Proteins", Suppl. 5, 171-183, 2001.
[19] Richard O. Day, Gary B. Lamont and Ruth Pachter, "Protein Structure Prediction by Applying an Evolutionary Algorithm", in December 2001.
[20] Zhen Zhang and Nan Jing, "Radial basis function method for prediction of protein secondary structure, " International Conference on Machine Learning and Cybernetics, vol. 3, pp. 1379-1383, 2008.
[21] Senapati, M.R., Vijaya, I. and Dash, P.K., "Rule Extraction from Radial Basis Functional Neural Networks by Using Particle Swarm Optimization", Journal of Computer Science, 3(8): 592-599, 2007.
[22] Karayiannis, N.B. and Mi, G.W., "Growing radial basis neural networks: Merging supervised and unsupervised learning with network growth techniques, " IEEE Trans. Neural Networks, Vol. 8, 1492–1506, 1997.
[23] Susan C. White, "NN3 Time Series Forecasting with Radial Basis Function Networks".
[24] Bishop, C., "Improving the Generalization Properties of Radial Basis Function Neural Networks, " Neural Computation, 3, 579 – 588, 1991.
[25] Kenneth J. McGarry, John Tait, Stefan Wermter and John Macintyre, "Rule-Extraction from Radial Basis Function Networks", IEE, 613-618, 1999.
[26] Park, J. and Sandberg, I.W., "Approximation and radial-basis-function networks, " Neural Comput., Vol. 5, 305–316, 1993.
[27] Cai, Y.D., Liu, X.J., Xu, X.B. and Zhou, G.P., "Support vector machines for predicting protein structural class", BMC Bioinformatics, 15, 2001.
[28] Minh Ngoc Nguyen and Jagath C. Rajapakse, "Two-stage support vector machines for protein secondary structure prediction, " Neural, Parallel & Scientific Computations, vol. 11, No. 2, 1-18, 2003.
[29] Hu, H.-J. P.C. Tai, J. He, R. Harrison, and Y. Pan, "Protein secondary structure prediction using support vector machine with a PSSM profile and an advanced tertiary classifier, " IEEE Computational Systems Bioinformatics Conference. 213-214, 2005.
[30] Lukasz A. Kurgan, Mandana Rahbari and Leila Homaeian, "Impact of the Predicted Protein Structural Content on Prediction of Structural Classes for the Twilight Zone Proteins", Proceedings of the 5th International Conference on Machine Learning and Applications, IEEE.180-186, 2006.
[31] Kristin K. Koretke, Zaida Luthey-Schulten and Peter G. Wolynes, "Self-consistently optimized energy functions for protein structure prediction by molecular dynamics, " Journal List, Biophysics, Vol. 95, No. 6, 2932-2937, 1998.

[32] Jian-xiong Dong, Adam Krzyzak and Ching Y. Suen, "A Fast SVM Training Algorithm", S.-W. Lee and A. Verri (Eds.): SVM 2002, LNCS 2388. 53–67, 2002.

[33] Chao Chen, Lixuan Chen, Xiaoyong Zou and Peixiang Cai, "Prediction of Protein Secondary Structure Content by Using the Concept of Chou's Pseudo Amino Acid Composition and Support Vector Machine", Protein & Peptide Letters, 16, 27-31, 2009.D.

[34] DeCoste and Scholkopf, B., "Training invariant support vector machines", Machine Learning, 46(1-3):161–190, 2002.