BIOINFORMATICS

Vol. 18 no. 11 2002
Pages 1508–1514

An evaluation of β-turn prediction methods

*Harpreet Kaur and G. P. S. Raghava**

Institute of Microbial Technology, Chandigarh, India

Received on March 13, 2002; revised on May 13, 2002; accepted on May 24, 2002

ABSTRACT

Motivation: β-turn is an important element of protein structure. In the past three decades, numerous β-turn prediction methods have been developed based on various strategies. For a detailed discussion about the importance of β-turns and a systematic introduction of the existing prediction algorithms for β-turns and their types, please see a recent review (Chou, *Analytical Biochemistry*, **286**, 1–16, 2000). However at present, it is still difficult to say which method is better than the other. This is because of the fact that these methods were developed on different sets of data. Thus, it is important to evaluate the performance of β-turn prediction methods.

Results: We have evaluated the performance of six methods of β-turn prediction. All the methods have been tested on a set of 426 non-homologous protein chains. It has been observed that the performance of the neural network based method, BTPRED, is significantly better than the statistical methods. One of the reasons for its better performance is that it utilizes the predicted secondary structure information. We have also trained, tested and evaluated the performance of all methods except BTPRED and GORBTURN, on new data set using a 7-fold cross-validation technique. There is a significant improvement in performance of all the methods when secondary structure information is incorporated. Moreover, after incorporating secondary structure information, the Sequence Coupled Model has yielded better results in predicting β-turns as compared with other methods. In this study, both threshold dependent and independent (ROC) measures have been used for evaluation.

Contact: raghava@imtech.res.in

Supplementary information: http://imtech.res.in/raghava/betatpred/eval/.

INTRODUCTION

Protein secondary structure prediction is an intermediate step in the prediction of tertiary structure from amino acid sequence. Numerous methods have been developed in the past for the secondary structure prediction (http://PredictionCenter.llnl.gov/casp4/). Except for a few of the earlier methods, all the secondary structure prediction methods predict only three states in a protein—helices, β-strand and coil. The coil region in a protein includes tight turns (Chou, 2000), bulges and random coil structures. One of the tight turns is a β-turn, which plays a vital role in protein folding and stability (Takano *et al.*, 2000). The present secondary structure prediction methods do not provide any information about β-turns in proteins; despite the fact that β-turns is the most common type of non-repetitive structure. On average, the β-turns constitute about 25% of the residues in the globular proteins (Kabsch and Sander, 1983). In contrast to a vast number of methods for secondary structure prediction, only a few methods have been reported for β-turn prediction (Chou and Fasman, 1979; Chou, 1997; Shepherd *et al.*, 1999). Most of the β-turn prediction methods are statistical except BTPRED, which is based on a neural network. Recently, we have developed a web server for predicting β-turns in a protein using existing statistical algorithms (Kaur and Raghava, 2002; http://imtech.res.in/raghava/betatpred/).

Although, there are a number of worldwide experiments to assess the performance of protein structure prediction methods (e.g. CASP, CAFASP and EVA), however, there is no report on evaluation of β-turn prediction methods. Thus, there is a need to assess the quality of these methods. The developers have reported accuracy on different data sets that make it difficult to have an objective comparison of methods. In this paper, an attempt has been made to evaluate the performance of different β-turn prediction methods on a uniform data set.

MATERIALS AND METHODS

Data set

Our data set consists of 426 non-homologous protein chains, as described by Guruprasad and Rajkumar (2000). In this data set, no two protein chains have more than 25% sequence identity. The structure of these protein chains is determined by X-ray crystallography at 2.0 Å resolution or better. The PROMOTIF program was used to assign the β-turns in proteins (Hutchinson and Thornton, 1996).

Cross validation technique

In order to evaluate a prediction method, it is necessary to have different data sets for training and testing. The

*To whom correspondence should be addressed.

jackknife test is the most objective and rigorous cross validation method compared with the independent data set test and sub-data set test (Chou and Zhang, 1995). In a full jack-knife test of N proteins, one protein is removed from the set, the training is done on the remaining $N-1$ proteins and the testing is done on the removed protein. This process is repeated N times by removing each protein in turn. Since this training technique is very time consuming, particularly for methods that take a long time in training (e.g. neural networks), a more limited cross-validation or sub-data set test is often performed. In sub-data set test, the set of proteins is divided into M equally balanced subsets. Parameters are developed on $(M-1)N/M$ proteins and then tested on the remaining N/M proteins. This process is repeated M times, once for each subset. In this report, a 7-fold cross-validation technique is used where the data is divided into 7 subsets.

Inclusion of secondary structure information in prediction

BTPRED (Shepherd *et al.*, 1999) is a neural network based method that incorporates secondary structure information. In order to have an objective comparison between BTPRED and other statistical approaches, the performance of different statistical methods has also been assessed by including secondary structure information. In the first step, the secondary structures of all the 426 proteins have been predicted by the PROF method (Quali and King, 2000). In the second step, the turns are predicted only for those residues that are in the predicted coil region, i.e. eliminating the helix- and strand-forming residues from β-turn prediction.

Measures of prediction accuracy

The measure used in this study can be divided in following two categories.

Threshold dependent measures Four parameters were used to measure the performance of prediction methods as described by Shepherd *et al.* (1999). Following is the brief description of these parameters: (i) **Qtotal** (or prediction accuracy), is the percentage of correctly classified residues; (ii) **Matthews Correlation Coefficient (MCC)**, accounts for both over- and under-predictions; (iii) **Qpredict** is the percentage of correct prediction of β-turns (or probability of correct prediction); and (iv) **Qobserved** is the percentage of observed β-turns that are correctly predicted (or percent coverage). These parameters can be calculated by following equations:

$$\boldsymbol{Qtotal} = \left(\frac{p+n}{t}\right) \times 100$$

$$\boldsymbol{MCC} = \frac{pn - ou}{\sqrt{(p+o)(p+u)(n+o)(n+u)}}$$

$$\boldsymbol{Qpredicted} = \left(\frac{p}{p+o}\right) \times 100$$

$$\boldsymbol{Qobserved} = \left(\frac{p}{p+u}\right) \times 100$$

where p and n are number of correctly classified β-turn and non-β-turn residues, respectively. where o and u are number of incorrectly classified β-turn and non-β-turn residues, respectively.

Threshold independent measures—ROC One problem with the threshold dependent measures is that they measure the performance on a given threshold. They fail to use all the information provided by a method for evaluation. The Receiver Operating Characteristic (ROC) is a threshold independent measure that has been developed as a signal processing technique (Deleo, 1993). The area under the ROC curve measures discrimination, the ability of a method to correctly classify β-turns and non-turn residues.

Prediction methods analyzed

Six different β-turn prediction methods have been evaluated. Each is briefly described here.

Chou–Fasman algorithm The Chou–Fasman algorithm depends on assigning a set of prediction values to each of the residues and determining the conformational parameters and the positional frequencies (Chou and Fasman, 1974). The conformational parameters for each amino acid are calculated by considering the relative frequency of a given type of secondary structure, and the fraction of residues occurring in that type of structure.

Thornton's algorithm Wilmot and Thornton (1988) has developed a prediction program by using a data set of 59 proteins based on the statistical method as employed by Chou and Fasman (1974). Initially the absolute amino acid occurrence for each of the four positions in the β-turn types I and II are calculated. These are then normalized to give positional frequencies $f(i)$, $f(i+1)$, $f(i+2)$ and $f(i+3)$. Conformational parameters for turn categories, P_t, are calculated for each amino acid. The conformational parameters for helix, P_a and β-sheet, P_b are taken from (Chou and Fasman, 1979).

1–4 & 2–3 Correlation model In this model, the coupling effect between the 1st and 4th residue and that between the 2nd and 3rd is given a special consideration. When a tetrapeptide folds into a β-turn, the interaction between its 1st and 4th residue and between its 2nd and 3rd residue play an important role. Particularly, a hydrogen bond may be formed between the backbone $C=O$ of the 1st residue and the backbone NH of the 4th residue. This model is based on the first-order Markov chain involv-

Table 1. Results of β-turn prediction methods at original and new thresholds using original parameters

Methods	*Qtotal* Original	New	*Qpredicted* Original	New	*Qobserved* Original	New	MCC Original	New
BTPRED	74.4	–	48.3	–	57.3	–	0.35	–
Chou–Fasman	65.2	72.4	37.6	43.6	63.5	43.6	0.26	0.25
Thornton	68.0	71.5	38.6	41.3	52.4	39.2	0.23	0.22
1–4 & 2–3 Correlation model	59.1	69.8	32.4	37.9	61.9	36.4	0.17	0.17
Sequence coupled model	53.3	69.2	32.4	36.9	72.8	35.8	0.17	0.16
GORBTURN	70.5	–	39.3	–	37.3	–	0.19	–

ing conditional probabilities $P_3(X_3|X_2)$ and $P_4(X_4|X_1)$. On the basis of these probabilities, an attribute function ϕ is calculated and a β-turn is predicted if the discriminant function Δ is positive where, $\Delta = \phi - \lambda$ and λ is the threshold value determined by an optimization procedure (Zhang and Chou, 1997).

Sequence coupled model Chou (1997) proposed a residue coupled model based on first order Markov chain to predict β-turns in proteins. Given a tetrapeptide, its attribute to the β-turn set S^+ or the non-β-turn set S^- is expressed, respectively by an attribute function (which can be defined as

$$\Psi^+(R_i R_{i+1} R_{i+2} R_{i+3}) = g P_i^+(R_i) P_{i+1}^+(R_{i+1}|R_i) \times P_{i+2}^+(R_{i+2}|R_{i+1}) P_{i+3}^+(R_{i+3}|R_{i+2})$$

where $g = 10^4$ is the amplifying factor used for making the data in a range easier to handle. $P_i^+(R_i)$ is the probability of amino acid R_i occurring at sub site i in the β-turn tetrapeptide set S^+, and it is independent of the other subsites because R_i is located at the first position of the four subsite sequence. $P_{i+1}^+(R_{i+1}|R_i)$ is the probability of amino acid R_{i+1} occurring at the subsite $i+1$ given that R_i has occurred at position i and so forth. Similarly, for the non-β-turn set, the attribute function Ψ can be defined.

Later, the conditional probabilities have been calculated for different types of β-turns to enable the residue-coupled model to predict different β-turn types as well (Chou and Blinn, 1997).

GORBTURN (v3.0) The program GORBTURN (v 3.0), a new version of BTURNPRED (Wilmot and Thornton, 1990) is a user-friendly piece of software written in Fortran77. The program uses Thornton's type I and type II positional frequencies and the directional parameters in combination with equivalent parameters (Gibrat *et al.*, 1987) to eliminate potential helix- and strand-forming residues from the β-turn prediction.

BTPRED This is a neural network-based method (Shepherd *et al.*, 1999) developed on a set of 300 non-homologous protein domains with resolution 2.0 Å or better. A filtering network is used to improve the accuracy and the individual turn type is predicted using a separate neural network for each turn type to be predicted. It uses secondary structure information obtained from PHDsec program (Rost and Sander, 1993) about each amino acid.

RESULTS

Evaluation of methods using original/default parameters

Default threshold We have predicted the β-turns in all the proteins in our data set using different methods. The parameters and thresholds have been used as described by their authors. As shown in Table 1, BTPRED is significantly better than any other method. The overall prediction accuracy of BTPRED is 4–21% higher than the other methods. Taking the Matthew's correlation coefficient as one of the performance metric, BTPRED has an MCC value of 0.35 compared with the 0.26 achieved by Chou–Fasman, 0.23 achieved by Thornton, 0.17 of 1–4 & 2–3 correlation model and sequence coupled model and 0.19 of GORBTURN. Out of all the statistical methods, GORBTURN has the highest prediction accuracy. However, its MCC value is less than Chou–Fasman and Thornton's algorithms. Moreover, Chou–Fasman and Thornton's algorithms perform better than the correlation models.

New threshold Except BTPRED and GORBTURN, the performance of all other methods is threshold dependent. Thus, we have determined the new threshold value for each method at which Qbserved and Qpredict are nearly same. At this new threshold, evaluation has been performed (see Table 1). As shown in Table 1, there is an improvement in prediction accuracy of all the methods at new threshold values. The performance measure of different methods follows the same trend as the original threshold. Chou–Fasman method performs better than all other statistical methods. Chou–Fasman and Thornton's algorithm have higher prediction accuracy and MCC values as compared to correlation models. At

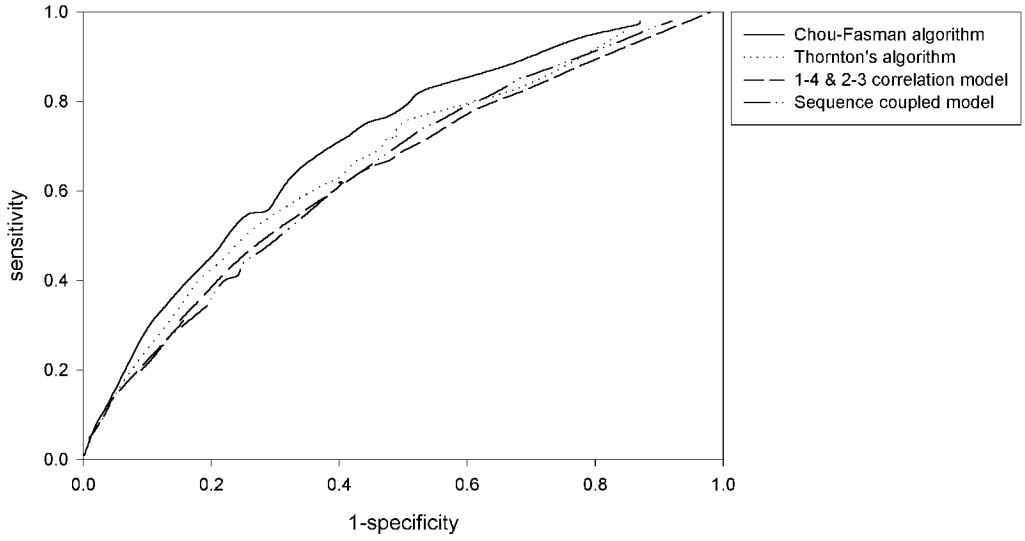

Fig. 1. ROC plot without cross-validation.

Table 2. Results of 7-fold cross-validation at original and new threshold values

Methods	*Qtotal*		*Qpredicted*		*Qobserved*		MCC	
	Original	New	Original	New	Original	New	Original	New
Chou–Fasman	74.9	69.3	46.1	36.9	16.9	35.3	0.16	0.16
Thornton	74.5	70.1	44.0	36.7	16.7	30.5	0.15	0.14
1–4 & 2–3 Correlation model	63.2	71.1	35.3	40.8	60.4	40.3	0.21	0.21
Sequence coupled model	50.6	72.7	31.7	43.9	88.4	41.0	0.23	0.25

new threshold values, the sequence coupled model and 1–4 & 2–3 correlation model show quite similar results although the percentage coverage ($Qobserved$) of the former is better. The prediction abilities of these methods appear to be of the same level.

Threshold independent It can be seen from Table 1 that some methods have higher Qtotal and lower MCC value or vice-versa. It is obviously not possible to compare the methods objectively. Therefore, a single threshold independent measure of performance, ROC, has been used to assess the performance of the methods. In order to have an entire range of ROC plot from 0 to 1, helix and sheet's conformational parameters have been lowered in the case of Chou–Fasman and Thornton's algorithms. It is clear from the ROC plot (Figure 1) and ROC values (see *Supplementary information*) that the Chou–Fasman algorithm performs better than all other methods. Its ROC value, which is indicative of better performance as compared to other methods, is in agreement with the higher prediction accuracy and MCC value as achieved by the Chou–Fasman method (Table 1). Both 1–4 & 2–3 Correlation models and Sequence Coupled models have ROC value equal to 0.64 and perform equally.

Testing, training and evaluation using 7-fold cross validation

We have trained, tested and evaluated the performance of methods on a new set of proteins using 7-fold cross-validation for all the methods except BTPRED and GORBTURN. The results for each algorithm that represent the average of 7 runs (with each run performed on a different training/test set pair) at original and new threshold values are presented in Table 2.

Original threshold After cross-validation, the Chou–Fasman algorithm performs slightly better than Thornton's algorithm. For Chou–Fasman and Thornton's methods, there is an approximately 10% increase in overall accuracy (Qtotal) and probability of correct prediction (Qpredicted) but there is a significant decrease in the percentage coverage of turns (Qobserved) and MCC value after cross validation (Table 2). However, for site-coupled models (1–4 & 2–3 correlation model and sequence coupled model), there is a 4–5% improvement in MCC value.

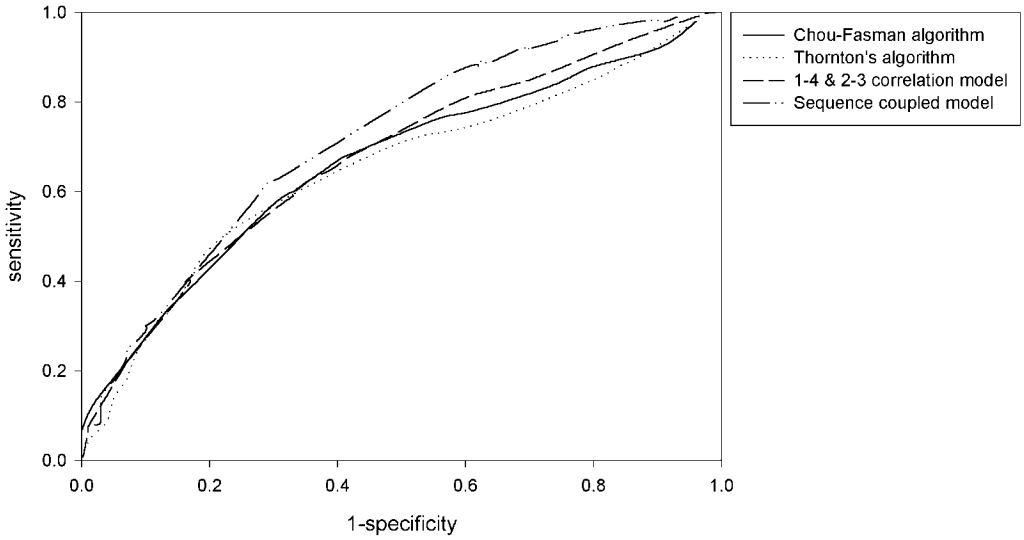

Fig. 2. ROC plot with cross-validation.

Table 3. Effect of secondary structure information on performance of β-turn prediction methods

Methods	*Qtotal*		*Qpredicted*		*Qobserved*		MCC	
	Original	New	Original	New	Original	New	Original	New
BTPRED	75.3	–	49.7	–	63.4	–	0.39	–
Chou–Fasman	74.3	75.3	47.7	49.6	54.3	47.5	0.34	0.32
Thornton	75.2	75.2	49.3	49.3	44.9	44.9	0.31	0.31
1–4 & 2–3 Correlation model	73.4	74.8	46.2	48.0	51.5	39.8	0.31	0.28
Sequence coupled model	72.2	75.4	45.0	49.6	60.0	40.0	0.33	0.28
GORBTURN	75.4	–	49.6	–	37.7	–	0.28	–

The MCC value averaged over 7 testing sets is maximum for the sequence coupled model and shows a substantial improvement as compared to other methods.

New threshold Prediction at new threshold values does not improve the results for Chou–Fasman and Thornton's method; however, there is a significant improvement in overall prediction accuracy and MCC value in case of correlation models (Table 2).

ROC It is clear from the ROC values (see *Supplementary information*) and ROC plot (Figure 2) that sequence coupled model has the highest ROC value and performs better than other statistical methods after cross-validation. 1–4 & 2–3 correlation model yields better results than Chou–Fasman and Thornton's algorithms.

Effect of secondary structure on performance of β-turn prediction

An increase in prediction accuracy of statistical methods can be expected by including secondary structure information. The results are summarized in Table 3. The first conclusion that can be drawn from the results obtained is that at both original and new thresholds, there is a significant improvement in prediction performance by inclusion of secondary structure information for all the methods. After incorporating secondary structure information, site- independent algorithms show a 5–9% increase in percentage accuracy whereas correlation models show a significant improvement of 14–19% in percentage accuracy. MCC values so obtained are in the range 0.28–0.39, the least being of GORBTURN and the maximum of BTPRED. Similarly, the probability of correct prediction (Qpredicted) has improved significantly.

Effect of accuracy of secondary structure prediction on BTPRED

Table 4 shows the performance of BTPRED by incorporating the secondary structure information predicted from different methods. It is evident that the accuracy of BTPRED is dependent on the accuracy of secondary struc-

Table 4. BTPRED results using secondary structure information from different methods

Methods	*Qtotal*	*Qpredicted*	*Qobserved*	MCC
PSIPRED (single)	74.2	47.5	46.4	0.30
PSIPRED (multiple)	76.0	50.9	63.0	0.40
PROF	75.3	49.7	63.4	0.39
PHD	74.4	48.3	57.3	0.35

ture prediction. The secondary structure prediction performance of PROF and PSIPRED(multiple) is better than PHD and PSIPRED(single). Thus, the percentage accuracy(Qtotal) and MCC value of BTPRED is higher for these methods. BTPRED shows a 2% improvement in prediction accuracy when multiple protein sequence data is included. This may be because the secondary structure prediction from a multiple alignment of protein sequences rather than a single sequence improves accuracy of secondary structure prediction and so does the β-turn prediction accuracy.

DISCUSSION

The assessment of performance of a method (or technique or process or machine) plays a vital role in the development of any field of science. In this study, we have assessed the methods for β-turn prediction in proteins. Both from structural and functional point of view, β-turns play important biological roles as reflected from the following facts: (i) a polypeptide chain cannot fold into a compact globular fold without β-turns; (ii) β-turns usually occur on the exposed surface of a protein and hence here likely involved in molecular recognition processes; and (iii) also play an important role in protein folding and stability. Thus, β-turn is an important component of protein structure whose prediction can provide enormous information to the researchers working in the field of protein structure prediction. The prediction of β-turns would not only aid in overall tertiary structure prediction but also assists in fold recognition studies.

It is apparent from the analysis that when β-turns are predicted by all the methods by using their respective original parameters and threshold values, BTPRED has an overall higher prediction accuracy than the statistical programs. The better performance of BTPRED is due to the neural network-based learning algorithms and inclusion of secondary structure information as compared to that of statistical methods, which use simple positional preferences and conformational parameters. Chou–Fasman and Thornton's algorithms perform better than the correlation methods as evident from the ROC values. Prediction at new threshold results in improvement in prediction accuracy of all the methods. Again, at new thresholds, Chou–Fasman and Thornton's methods have better results than the correlation models. At new thresholds, 1–4 & 2–3 correlation model and sequence coupled model have similar prediction abilities in terms of prediction measures. The reason for their similar behavior may be that both methods take into consideration the coupling among the residues in a β-turn sequence during prediction.

There has been significant improvement in the performance of all methods when trained and tested on a new data set using 7-fold cross-validation. It is because the earlier statistical methods have suffered from a lack of sufficient data. For example, Chou–Fasman algorithms have used parameters based on the information from a small set of 29 proteins. Moreover, an early problem in prediction has been the inclusion of structures used to derive parameters in the set of structures to assess the accuracy of the method. Different methods can be ranked depending on the values of *Qtotal*, *Qpredicted*, *Qobserved* and MCC values. In terms of *Qtotal* and *Qpredicted*, the Chou–Fasman algorithm performs better as compared to other algorithms. However, for *Qobserved* and MCC values, the performance of the Chou–Fasman and Thornton's algorithms fall significantly. In case of Sequence Coupled Model, there is a notable increase in *Qobserved* and MCC values. Different methods have been compared by ROC values. Sequence coupled model has achieved a higher ROC value as compared to other methods which indicate its better prediction ability.

BTPRED uses secondary structure information for predicting β-turns. Therefore, the prediction of β-turns has been carried out by statistical methods by incorporating secondary structure information. As inferred from the results, there is an increase in all prediction accuracy for all methods.

The performance of BTPRED is evaluated by using the secondary structure output from different methods. The difference in predictive accuracy of BTPRED can be primarily attributed to the difference in secondary structure prediction accuracy of these methods. Therefore, an increase in β-turn predictive accuracy is possible as more effective protein secondary structure prediction methods are developed.

CONCLUSION

This paper provides a benchmarking of different β-turn prediction methods. BTPRED, which is a neural network-based method, performs better than statistical method. The performance of statistical methods increases significantly when their statistical parameters were derived from the recent/large data set of proteins. The accuracy of these statistical methods has improved further when predicted

secondary structure information is incorporated. This study demonstrates that even the older statistical methods can achieve a level of accuracy equal to current day neural network methods, by careful refinement and consideration of three state secondary structure prediction. The results also indicate that the improvement in accuracy of β-turn prediction over the years is mainly due to increase in size of protein data set used for training rather than due to improvement in technique. For example the older statistical method of Chou and Fasman developed in 1974 can be brought up to the level of accuracy shown by modern day methods such as neural networks, by using new parameters and predicted secondary structure.

A combination of neural network and statistical approaches may provide substantially better results than either one alone. A possibility for future research includes combining a statistical method with a neural network method. For example, the prediction resulting from a statistical method may be used as one of the inputs to the network. In this way, the network can be used to 'fine tune' the results from the statistical method and improve prediction accuracy.

REFERENCES

Chou,K.C. (1997) Prediction of β-turns. *J. Pept. Res.*, **49**, 120–144.

Chou,K.C. (2000) Prediction of tight turns and their types in proteins. *Analytical Biochem.*, **286**, 1–16.

Chou,K.C. and Zhang,C.T. (1995) Prediction of protein structural classes. *Critical Reviews in Biochem. and Mol. Biol.*, **30**, 275–349.

Chou,K.C. and Blinn,J.R. (1997) Classification and prediction of β-turn types. *J. Protein Chem.*, **16**, 575–595.

Chou,P.Y. and Fasman,G.D. (1974) Conformational parameters for amino acids in helical, β-sheet and random coil regions calculated from proteins. *Biochemistry*, **13**, 211–222.

Chou,P.Y. and Fasman,G.D. (1979) Prediction of β-turns. *Biophys. J.*, **26**, 367–384.

Deleo,J.M. (1993) Proceedings of the Second International Symposium on Uncertainty Modelling and Analysis. *IEEE*. Computer Society Press, College Park, MD, pp. 318–325.

Gibrat,J.-F., Garnier,J. and Robson,B. (1987) Further development of protein secondary structure prediction using information theory. New parameters and consideration of residue pairs. *J. Mol. Biol.*, **198**, 425–433.

Guruprasad,K. and Rajkumar,S. (2000) β- and γ-turns in proteins revisited: a new set of amino acid dependent positional preferences and potential. *J. Biosci.*, **25**, 143–156.

Hutchinson,E.G. and Thornton,J.M. (1996) PROMOTIF—a program to identify and analyze structural motifs in proteins. *Protein Sci.*, **5**, 212–220.

Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

Kaur,H. and Raghava,G.P.S. (2002) BetaTPred-prediction of β-turns in a protein using statistical algorithms. *Bioinformatics*, **18**, 498–499.

Quali,M. and King,R.D. (2000) Cascaded multiple classifiers for secondary structure prediction. *Protein Sci.*, **9**, 1162–1176.

Rost,B. and Sander,C. (1993) Improved prediction of protein secondary structure by use of sequence profiles and neural networks. *Proc. Natl Acad. Sci. USA*, **90**, 7558–7562.

Shepherd,A.J., Gorse,D. and Thornton,J.M. (1999) Prediction of the location and type of β-turns in proteins using neural networks. *Protein Sci.*, **8**, 1045–1055.

Takano,K., Yamagata,Y. and Yutani,K. (2000) Role of amino acid residues at turns in the conformational stability and folding of human lysozyme. *Biochemistry*, **39**, 8655–8665.

Wilmot,C.M. and Thornton,J.M. (1988) Analysis and prediction of the different types of β-turns in proteins. *J. Mol. Biol.*, **203**, 221–232.

Wilmot,C.M. and Thornton,J.M. (1990) β-turns and their distortion: a proposed new nomenclature. *Protein Engg.*, **3**, 479–493.

Zhang,C.T. and Chou,K.C. (1997) Prediction of β-turns in proteins by 1–4 & 2–3 Correlation Model. *Biopolymers*, **41**, 673–702.