BIOINFORMATICS APPLICATIONS NOTE Vol. 18 no. 3 2002
Pages 498–499

BetaTPred: prediction of β-TURNS in a protein using statistical algorithms

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

Institute of Microbial Technology, Sector-39A, Chandigarh, India

Received on July 22, 2001; revised on September 19, 2001; accepted on October 29, 2001

ABSTRACT

Motivation: β-turns play an important role from a structural and functional point of view. β-turns are the most common type of non-repetitive structures in proteins and comprise on average, 25% of the residues. In the past numerous methods have been developed to predict β-turns in a protein. Most of these prediction methods are based on statistical approaches. In order to utilize the full potential of these methods, there is a need to develop a web server.

Results: This paper describes a web server called BetaTPred, developed for predicting β-TURNS in a protein from its amino acid sequence. BetaTPred allows the user to predict turns in a protein using existing statistical algorithms. It also allows to predict different types of β-TURNS e.g. type I, I′, II, II′, VI, VIII and non-specific. This server assists the users in predicting the consensus β-TURNS in a protein.

Availability: The server is accessible from http://imtech.res.in/raghava/betatpred/

Contact: raghava@imtech.res.in

Supplementary information: http://imtech.res.in/raghava/betatpred/intro.html

INTRODUCTION

Protein secondary structure prediction is an intermediate step in tertiary structure prediction. In the past, a number of methods have been developed for predicting protein secondary structure (Chou and Fasman, 1974; http://PredictionCenter.llnl.gov/casp4/). These methods only predict helix and beta sheet in protein and the rest they define as coil. These coiled regions include tight turns, bulges and random coil structures (Chou, 2000). One of the tight turns is the β-turn that consists of four consecutive residues defined by positions i, $i+1$, $i+2$ and $i+3$, which are not present in an α-helix; the distance between $C_\alpha(i)$ and $C_\alpha(i+3)$ is less than 7 Å and the turn leads to a reversal in the protein chain.

The prediction of β-turns was a part of some of the earliest methods of secondary structure prediction (Lewis *et al.*, 1973; Chou and Fasman, 1974; Garnier *et al.*, 1978). To predict β-turns Wilmot and Thornton (1988) derived the parameters from their frequency distributions. All these methods are based on the site—independent model (Chou and Fasman, 1979). In 1997, models 1–4 and 2–3 correlation and sequence coupled were proposed by taking into account, the correlations between the pairing of residues (Chou, 1997; Zhang and Chou, 1997; Chou and Blinn, 1997). Recently, a method as well as a web server based on neural network, was developed to predict the β-turns (Shepherd *et al.*, 1999).

This paper describes a web server developed for predicting β-turns in a protein from its amino acid sequence using existing statistical methods.

ALGORITHMS

The BetaTPred server implements the following statistical algorithms for predicting β-turns. The following is a brief description.

Chou–Fasman algorithm

In this method, 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 (Chou and Fasman, 1974).

Thornton's positional frequencies

Wilmot and Thornton (1988), developed a prediction program by using a dataset of 59 proteins. The conformational parameters for turn types (P_t) I and II were calculated and a β-turn is predicted if the calculated value of P_t is greater than a cut off value.

1–4 and 2–3 correlation model

In this algorithm, the coupling effect between the 1st and 4th residue and that between the 2nd and 3rd is given a special consideration. This model is based on the first-order Markov chain involving conditional probabilities $P3(X3|X2)$ and $P4(X4|X1)$. On the basis of these probabilities, an attribute function Ø is calculated and a β-turn is predicted if the discriminant function Δ is positive where, $\Delta = Ø - \lambda$ and λ is the threshold value determined by an optimization procedure (Zhang and Chou, 1997).

*To whom correspondence should be addressed.

Sequence coupled model

This residue coupled model is based on first order Markov chain to predict β-turns in proteins. Using this model, the turns as well as their type were predicted, as described below.

Turn prediction. 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 Ψ which can be defined as:

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

where $g = 10^4$ is the amplifying factor, $P_i^+(R_i)$ is the probability of amino acid R_i occurring at sub-site i in the β-turn tetrapeptide set S^+. $P_{i+1}^+(R_{i+1}|R_i)$ is the probability of amino acid R_{i+1} occurring at the subsite $i+1$ given that R_{i+1} has occurred at position $i+1$ and so forth. Similarly, for the non-β-turn set, the attribute function Ψ^- will be calculated as:

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

The discriminant function Δ can be defined as

$$\Delta(R_i R_{i+1} R_{i+2} R_{i+3}) = w^+ \Psi^+(R_i R_{i+1} R_{i+2} R_{i+3}) - w^- \Psi^-(R_i R_{i+1} R_{i+2} R_{i+3})$$

where w^+ and w^- are the weight factors for the probabilities derived from the β-turn and non-β-turn training datasets respectively. Thus, a β-turn is predicted if $\Delta > 0$.

Turn type prediction. The residue-coupled model was extended to make it able to predict different β-turn types as well (Chou and Blinn, 1997). Here, seven different sets S^1, $S^{1\prime}$, S^2, $S^{2\prime}$, S^6, S^8 and S^- are used to represent the tetrapeptide sets of turn types I, I′, II, II′, VI, VIII and non-β-turn respectively. Given a tetrapeptide, its attribution to the sets S^1, $S^{1\prime}$, S^2, $S^{2\prime}$, S^6, S^8 and S^- are expressed by Ψ^1, $\Psi^{1\prime}$, Ψ^2, $\Psi^{2\prime}$, Ψ^6, Ψ^8 and Ψ^-, respectively. By taking into consideration the coupling effect between the adjacent residues, the attribute functions for different β-turn types are expressed by equations similar to (1). The tetrapeptide is predicted to be the structural type for which Ψ has the maximum value.

GORBTURN (v3.0)

The program GORBTURN (v 3.0) is a user-friendly piece of software written in Fortran77, and is a new version of BTURNPRED (Wilmot and Thornton, 1988). The program uses Thornton 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.

Consensus

A consensus method was developed by combining all the algorithms. A consensus β-turn will be that segment of sequence that is predicted as β-turn by all the methods.

DESCRIPTION OF THE SERVER

The server has two main options called default and advanced options. In the default option the server predicts the residues forming the β-turn. The server performs analysis of protein sequences using each method independently and presents the result on a single screen, which will assist the user in rapid visualization of consensus prediction of β-turns. The advanced option allows to predict the different turn types (I, I′, II, II′, VI, VIII and non-specific) using each method and to select the cut-off or threshold values.

BetaTPred provides three options to present the results: (i) graphical frame; (ii) graphical non-frame; and (iii) text form. The text output is in the form of a table which will give the predicted β-turn with the corresponding position number of the residues forming the turn. In the graphical frame option, the output is presented in four frames, where β-turns are indicated by stars. This option is very useful in identifying the overlapping β-turns from display. In the non-frame option, the predicted turns are presented by a single frame. In brief, the server is intended to predict β-turns in a protein using various statistical approaches.

REFERENCES

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

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

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.

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

Garnier,J., Osguthorpe,D.J. and Robson,B. (1978) Analysis and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.*, **120**, 97–120.

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.

Lewis,P.N., Momany,F.A. and Scheraga,H.A. (1973) Chain reversals in proteins. *Biochim. Biophys. Acta*, **303**, 211–229.

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

Wilmot,C.M. and Thornton,J.M. (1988) Analysis and prediction of the different types of β-turns in proteins. *J. Mol. Biol.*, **115**, 135–175.

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