PEPstr: A *de novo* Method for Tertiary Structure Prediction of Small Bioactive Peptides

Harpreet Kaur, Aarti Garg and G.P.S. Raghava*

Bioinformatics Center, Institute of Microbial Technology, Chandigarh, India

Abstract: Among secondary structure elements, β-turns are ubiquitous and major feature of bioactive peptides. We analyzed 77 biologically active peptides with length varying from 9 to 20 residues. Out of 77 peptides, 58 peptides were found to contain at least one β-turn. Further, at the residue level, 34.9% of total peptide residues were found to be in β-turns, higher than the number of helical (32.3%) and β-sheet residues (6.9%). So, we utilized the predicted β-turns information to develop an improved method for predicting the three-dimensional (3D) structure of small peptides. In principle, we built four different structural models for each peptide. The first 'model I' was built by assigning all the peptide residues an extended conformation ($\phi = \Psi = 180^o$). Second 'model II' was built using the information of regular secondary structures (helices, β-strands and coil) predicted from PSIPRED. In third 'model III', secondary structure information including β-turn types predicted from BetaTurns method was used. The fourth 'model IV' had main-chain ϕ, Ψ angles of model III and side chain angles assigned using standard Dunbrack backbone dependent rotamer library. These models were further refined using AMBER package and the resultant C^{α} rmsd values were calculated. It was found that adding the β-turns to the regular secondary structures greatly reduces the rmsd values both before and after the energy minimization. Hence, the results indicate that regular and irregular secondary structures, particularly β-turns information can provide valuable and vital information in the tertiary structure prediction of small bioactive peptides. Based on the above study, a web server PEPstr (http://www.imtech.res.in/raghava/pepstr/) was developed for predicting the tertiary structure of small bioactive peptides.

Keywords: Amber, beta-turn, protein secondary structure, protein tertiary structure, webserver, bioactive peptides.

1. INTRODUCTION

Peptides have the capability to control important functions of the organism, such as cell reproduction, appetite, euphoria, sleep, learning, immune response etc. There is a plethora of bioactive peptides, which act as hormones, neurotransmitters, antioxidants, toxins and antibiotics. Due to the importance of bioactive peptides, extensive studies have been carried out directed at their structure determination with a goal to understand function and to design clinically and diagnostically useful compounds.

Each role assumed by a bioactive peptide typically corresponds to a unique three-dimensional (3D) structure. Moreover, to design biologically active peptide requires a detailed knowledge of the 3D structure and is generally focused towards the modification of secondary structure elements. Also, the secondary structure rather than the tertiary structure is the dominant factor affecting the binding characteristics of the peptides [1]. However, smaller a polypeptide, the lesser well defined is its structure [2]. NMR is the widely used technique for determining the structure of polypeptides and proteins up to 100 residues or so [3-4]. But, the technique is time consuming and unable to keep up with the ongoing sequencing projects. Therefore, in past, few in-silico methods such as Robetta [5] and PepLook [6] were developed to predict the 3D structure of peptides. Robetta is based on building models of protein domains using both a template-based and de novo approaches [7-8]. PepLook is a recent method based on random combinations of ϕ-Ψ values and minimizing the structures by an iterative Boltzmann-Stochastic procedure [6].

Our strategy for predicting tertiary structure of small peptides is based on the observation that β-turn is an important and consistent feature of small peptides in addition to the regular secondary structures. A large body of evidence points to the significant occurrence of β-turns in bioactive peptides. It has also been shown that β-turns occur frequently among the conformationally active forms of the various linear and cyclic peptides [9-10]. For instance, in many antimicrobial peptides the structure that is responsible for bactericidal activity contains β-turn [11]. Moreover, the introduction of non-peptide bond mimics of the β-turn motif provides greater potential therapeutic value [12-13]. Also, in past, a number of methods have been developed for prediction of β-turns and their types in proteins [14-17]. Taken together, it should be possible, given the sequences of peptides, to make accurate predictions about their structure using both the regular (helices and β-strands) and irregular (β-turns) secondary structure information. Energy minimization and molecular dynamic simulations can be used to further refine the structure.

Hence, in present study, an attempt has been made to predict the tertiary structure of bioactive peptides using regular predicted secondary structures and β-turn types predicted from BetaTurns [17] method. Four models were generated

*Address correspondence to this author at the Bioinformatics Center, Institute of Microbial Technology, Chandigarh, India; Tel: +91-172-2690557; Fax: +91-172-2690632; E-mail: raghava@imtech.res.in

0929-8665/07 $50.00+.00

for each peptide, which include (i) model with all peptide residues in extended conformation ($\phi = \Psi = 180^o$), (ii) second model with backbone torsion angles corresponding to predicted secondary structure states, (iii) third model with backbone torsion angles corresponding to predicted secondary structure states and β-turn types predicted by BetaTurns and (iv) fourth model with side chain χ angles assigned using standard Dunbrack backbone dependent rotamer library [18] in addition to main-chain ϕ, Ψ angles of model III.

All these models were then subjected to energy minimization using Assisted Model Building and Energy Refinement (AMBER) *version6* computer package [19]. In order to measure the performance of the predicted structure of all peptides using these models, root mean square deviations (rmsd) were calculated after superimposing the predicted structures with the actual known structure. Both before and after energy minimization, superimposing the peptide with its actual structure after incorporating β-turns information obtained a lower C^α rmsd in comparison to models having extended conformation and regular secondary structure information alone.

2. METHODS

The Dataset

A representative data set comprising of 3D structures of 77 biologically active peptides was selected from PDB [20] and other databases such as PSST (http://pranag.physics.iisc.ernet.in/psst) [21] and PRF (http://www.genome.ad.jp/) [22]. The data set was restricted to those biologically active peptides that consist of only natural amino acids and are linear with length varying between 9-20 residues. For NMR structures, the first model was considered. Further, we excluded the peptides stabilized by a disulfide bridge and the remaining 42 were used for model generation and simulation studies.

Assignment of Secondary Structure

The secondary structure assigned by Database of Secondary Structure in Proteins (DSSP) was used to determine the distribution of different secondary structure states in the data set. It provides eight states classification of secondary structure [23].

Assignment of β-Turns

The details of the location and types of β-turns in the dataset were determined using 'PROMOTIF' program [24]. Promotif defines a β-turn as a stretch of four consecutive residues (denoted by *i*, *i*+1, *i*+2 and *i*+3), where the distance between the C^α atom of residue *i* and the C^α atom of residue *i*+3 is less than 7Å and the two central residues are not helical. The present study used the β-turn types classification scheme proposed by Hutchinson and Thornton [25] which categorizes nine types of β-turn types I, II, I', II', IV, VIa1, VIa2, VIb and VIII. The program lists the residue number, one-letter amino acid code of residues *i*, *i*+1, *i*+2 and *i*+3; β-turn type and dihedral angles ϕ(*i*+1), Ψ(*i*+1), ϕ(*i*+2), Ψ(*i*+2) of residues *i*+1 and *i*+2.

Prediction of Secondary Structure

The secondary structure of bioactive peptides was predicted using PSIPRED, which provides three states – helices, β-sheets and coil [26]. It is a neural network based secondary structure prediction method, which uses multiple alignment information of the target sequence obtained from PSI-BLAST [27].

Prediction of β-Turn Types

The β-turn types were predicted using BetaTurns (http://www.imtech.res.in/raghava/betaturns/) [17]. It is a neural network based method, which predicts β-turn types I, II, IV, VIII and non-specific using multiple sequence alignment information. It uses the position specific matrix obtained from PSI-BLAST [27] and secondary structure information obtained from PSIPRED [26].

Side-Chain Torsion Angles

All amino acids except Ala and Gly have a side-chain with one or more angles of rotation. For a specified ϕ, Ψ angles of main-chain, the side-chain χ angles were obtained from standard backbone dependent Dunbrack rotamer library [18] of May, 2002.

Energy Minimization and Molecular Dynamics Simulations (MD)

In the present study, energy minimization and dynamic simulations was carried out using SANDER module of Amber *version 6.0*. Following protocol was used:

a) Building of the Models

The Terminal Leap (Tleap) module of Amber v6.0, which is the non-graphical and command-line interface was used for constructing different models for each peptide. To establish baseline performance, the first model I was constructed by considering all the peptide residues in the extended conformation ($\phi = \Psi = 180^o$). The second model II was built by assigning the peptide residues ϕ, Ψ angles of the regular secondary structure states predicted by PSIPRED. In third model III, we specified the main-chain conformation the ϕ, Ψ values corresponding to PSIPRED predicted states: helix (-60, -40), β-strand (-120, 120) and ϕ, Ψ angles corresponding to β-turn types predicted by BetaTurns. The fourth model IV had main-chain ϕ,Ψ of model III and side-chain χ angles obtained from Dunbrack library. To achieve the maximum performance of the method, two other models were built up using observed secondary structure states. Thus, the fifth model was generated using ϕ, Ψ angles of observed secondary structure states as assigned by DSSP. The sixth model VI was built using ϕ, Ψ angles of observed secondary structure states assigned by DSSP and β-turn types assigned by Promotif. The models V and VI were used to establish the upper level performance of the prediction method. Further, for all these models, TLeap was used to prepare the coordinate and topology files, which were used as input to the energy minimization and dynamics.

b) Energy Minimization and Dynamics

Energy minimization and MD calculations were carried out using SANDER module with Amber force field, distance-dependent dielectric constant and the non-bonded cut-off value of 8Å. It was performed for all the six types of models for 42 peptide studied. Each round consisted of few initial cycles of steepest descent minimization followed by dynamics. The system was equilibrated and the constant pressure (NTP) simulations were carried out for 25ps at 300K using 1-fs time steps. Finally, this was followed by minimization using a combination of steepest descent and conjugate gradient algorithms. The final low energy conformations were saved in PDB format.

Performance Measure of Predicted Structures

After doing the MD simulations, it was necessary to inspect the quality of the obtained structure and see how well it actually fit the experimental structure. The carnal module of AMBER was used to calculate the backbone rmsd by superimposing the respective C^{α} atoms. It calculates the backbone rmsd as

$$\text{RMSD} = \sqrt{\sum_{n=1}^{N} \frac{(q_n - q'_n)^2}{N}}$$

where, q_n is the coordinates of the nth residue in one peptide and q'_n is the coordinates of the nth residue in the other peptide. The more similar the peptides are, the lower is the RMSD value.

3. RESULTS

A data set of 77 experimentally determined 3D structures of bioactive peptides was used. Only few of them had their 3D structures solved by X-ray crystallography otherwise most of them were have NMR solved structures. Most of the peptides were antibacterial, conotoxins and binding peptides. From these 77 peptides, we excluded 35 peptides stabilized by disulfide bridges. Table **1** lists the PDB codes and the solvent types of the remaining 42 peptides studied.

Distribution of Secondary Structure States

The regular (helices, β-sheets) and irregular (β-turns) secondary structure states were assigned in the data set using DSSP and Promotif respectively. The analysis of the distribution of secondary structure states clearly indicates that both regular and irregular secondary structures occur in bioactive peptides, with β-turns being the most prominent (Table **2**). A significant percentage of β-turns has been found in the peptides in comparison to helices and β-sheets. Out of 77 peptides, 58 have been found to contain at least one β-turn, comprising 75.3% of total peptides. It has been followed by helices forming 60%. The least occurring state is the β-strand, present in just 13% of the total peptides studied. At residue level, nearly 35% of total peptide residues are found in β-turn conformation.

Table 1. PDB Codes and Solvent Type of the 42 Bioactive Peptides

1qcm	1egs (H_2O/D_2O)
1odp (SDS)	1myu ($DPC/D_2O/H_2O$)
1m02 (SDS)	1qfa (TFE/Water)
1sol (SDS)	1rpv (TFE/Water)
1jav (SDS)	1id6 (Dimethyl sulfoxide/D_2O)
1g89 (Dodecyl phosphate)	1e0q (H_2O/D_2O)
1d7n (SDS/ D_2O/H_2O)	1kzv (Methanol chloroform)
1c98 (SDS)	1b03A
1dn3 (SDS/Na_2PO_4)	2bta
1p0o (SDS)	1lcx (NaN_3)
1p0j (SDS)	1gjf (5mm peptide)
1p5k (SDS)	1in3 (5mm peptide)
1p0l (SDS)	1l3q (Na_2PO_4)
1d6x (SDS)	1a13
2bp4 (TFE/Water)	1hu6 (Sodium phosphate)
1nkf (H_2O/D_2O)	1hu7 (Sodium phosphate)
1d9j ($DPC/H_2O/D_2O$)	1hu5 (Na_2PO_4 buffer)
1d9o ($DPC/D_2O/H_2O$)	1pef
1d9l (DPC/D_2O)	1du1
1d9m (DPC/D_2O)	1l2y (Peptide)
1d9p (DPC/D_2O)	1niz (NaAcetate buffer)

Distribution of β-Turn Types

The 58 peptides consist of 141 numbers of β-turns, which have been identified and categorized, in different types as shown in Fig. (**1**). The distribution of different types of β-

Table 2. Distribution of DSSP and Promotif Assigned Secondary Structure States in 77 Bioactive Peptides

Secondary structure state	No. of peptides	% of total peptide residues
Helices	46	32.3
β-sheets	10	6.9
β-turns	58	34.9

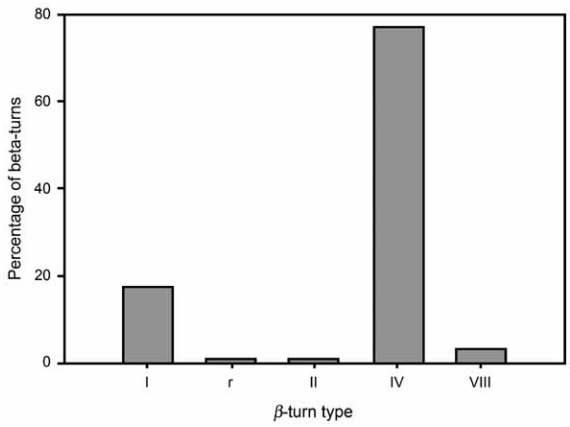

Figure 1. Distribution of different β-turn types in 77 bioactive peptides.

turns is highly asymmetric. Type IV is found to be most frequently occurring turn type (around 77.3%) followed by Type I β-turn (17.7%). The percentage occurrence of turn types I', II and VIII in bioactive peptides is negligible. Moreover Types II' and VI does not occur even once. Further, it has also been found that out of 141 β-turns identified, only 15 β-turns is part of β-hairpin structure. The distribution of β-turn types in bioactive peptides is in similar to that in proteins, with Type IV and I being the numerous occurring β-turn types [28].

Peptide Tertiary Structure Prediction

We studied four models for each of 42 peptides, whose brief description is given below:

Model I: with all peptide residues in extended conformation.

Model II: with ϕ, Ψ angles of regular secondary structure states

Model III: with ϕ, Ψ angles of secondary structure states including β-turns

Model IV: with ϕ, Ψ angles of model III and side chain χ angles from rotamer library

These models were compared with the actual three-dimensional structure and averaged backbone rmsds were calculated to assess the prediction performance (Table **3**). It is clear from the table that the model III is the most closest to the experimental structure with average C^{α} rmsd of 4.7Å. The first model with extended conformation has the largest backbone deviation from the actual structure. The model II with only the information of regular secondary structures obtained rmsd value of 5.7Å, intermediate of models I and III. Adding the side chain χ angles to model III reduces the rmsd a little to 4.5Å. Thus, using β-turns, a difference of 1.0Å is achieved in comparison to regular secondary states containing model II.

An important step used in later stages of structure determination is energy minimization. This can be improved by MD, which is a means of simulating the motion expected in a molecule. For all the three different models for each peptide, energy minimization and dynamics simulations were performed to further refine the models and the final averaged rmsd values obtained are 7.1Å, 4.4Å, 4.1Å and 4.0Å for models I, II, III and IV respectively (Table **3**). As expected, the models III and IV with β-turn information are found to be close to the actual structure in comparison to other models. It is clearly evident from the results that incorporation of β-turn information has minimized the backbone root mean deviation and contributes significantly to the overall tertiary structure prediction.

Limits of Method

Ideally, one should be able to achieve 100% prediction accuracy, however it is not possible. For instance, with models III and IV, the final rmsd value is 4.1 and 4.0Å respectively, which is still higher. In order to assess the limit of these models, further two models were built using DSSP

Table 3. Averaged Backbone Root Mean Deviation of 42 Peptides Before and After Energy Minimization and Dynamics Simulations

Models	Averaged backbone root mean deviation (Å)	
	before EM & DS[a]	***after EM & DS***
Models built using predicted secondary structure information		
I (extended conformation)	**10.0**	**7.1**
II (regular secondary states)	**5.7**	**4.4**
III (regular states + β-turns)	**4.7**	**4.1**
IV (III model + χ angles)	**4.5**	**4.0**
Models built using observed secondary structure information		
V (DSSP)	**5.6**	**4.7**
VI (DSSP + Promotif)	**4.9**	**4.0**

[a] EM and DS denote energy minimization and dynamics simulations respectively.

assigned regular secondary structures (model V) and Promotif assigned β-turns with regular secondary structures (model VI). These models are then used to establish a higher-level accuracy or the maximum accuracy that one can achieve. The results of model V and VI can be compared with that of models II and III respectively. As expected, the minimum rmsd values have been achieved with these models (Table **3**). The model V has rmsd value of 5.6Å, which is slightly lower (~0.1Å) than model II. However, with both the DSSP and Promotif information, model VI has the least rmsd of 4.0Å.

Pepstr Server

Based on the study, we have developed a web server that allows the user to predict the tertiary structure of small peptides. A flow chart of our procedure for tertiary structure prediction of peptides is shown in Fig. (**2**). The method is a

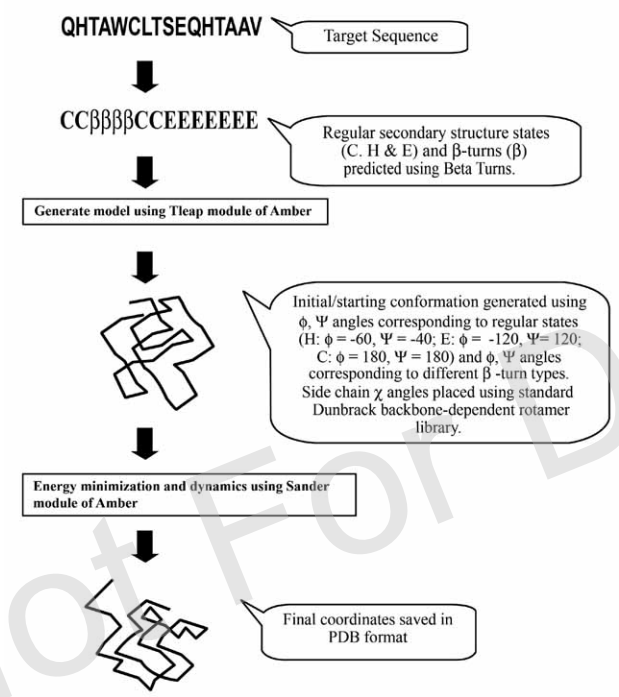

Figure 2. Flowchart of tertiary structure prediction procedure of Pepstr.

de novo protocol, which involves many steps- In first step, the regular secondary structure states (helix, β-strand and coil) and β-turn types are predicted using BetaTurns method. Secondly, a conformation is generated using ϕ, Ψ angles corresponding to secondary structures. Then, the side-chain torsion angles are assigned using standard Dunbrack backbone-dependent rotamer library. Further, the structure is refined using energy minimization and MD simulations. There is also an option for the user to specify the environment (vacuum, hydrophilic or hydrophobic) for the simulations. Finally, the server provides the final coordinates in PDB format.

The web server is available free for academic or non-profit users. Users can enter the input sequence in a single line using the one-letter amino acid notation. The output consists of predicted tertiary structure with final coordinates in PDB format.

Comparison with Existing Methods

It is important to compare the newly developed method with other existing methods. Recently, Thomas *et al.* [29] compared our Pepstr method with other prediction methods such as Robetta [5] and PepLook [6] on five peptides with lengths shorter than 30 residues [29]. The performance is measured more quantitatively using secondary structures, rmsd and mean force potential (MFP) energy scores. The results showed that Robetta which is more dedicated to proteins has some limitations for small peptides. However, Pepstr and PepLook yielded better predictions. For instance, rmsd values obtained for peptides Magainin 2 (2MAG) and Transportan (1SMZ) are 7.3 Å, 1.3 Å, 3.3 Å and 8.7 Å, 6.5 Å, 4.3 Å for methods Robetta, PepLook and Pepstr respectively. Moreover, it is also shown that the MFP energy scores of Pepstr are comparable to the NMR structures, which clearly suggests that Pepstr predicted structures are close to NMR structures. It is also to be noted that Pepstr predicts structure faster than PepLook in many instances.

CONCLUSIONS

To conclude, the present work is an attempt to improve the prediction performance of tertiary structure prediction of bioactive peptides using β-turns information along with the regular secondary structure states. The suggested approach is open for further improvement especially in the view of the further growth of the structural database of bioactive peptides, which will probably clarify their biological roles, and occurrence of tight turns in bioactive peptides.

ACKNOWLEDGEMENTS

We are thankful to Dr. Balvinder Singh for assisting us in the energy minimization and dynamics simulations. The work was supported by financial assistance from Council of Scientific and Industrial Research (CSIR) and Department of Biotechnology (DBT).

LIST OF ABBREVIATIONS

3D	=	Three-dimensional
AMBER	=	Assisted Model Building with Energy Refinement
RMSD	=	Root mean square deviation
PDB	=	Protein data bank
PSST	=	Protein sequence search tool
PRF	=	Protein research foundation
DSSP	=	Database of secondary structure in proteins
Tleap	=	Terminal Leap
MD	=	Molecular dynamics

REFERENCES

[1] Kaiser, E.T. (**1987**) *Trends Biotechnol., 12*, 305.
[2] Milner-White, E.J. (**1989**) *Trends Pharmacol. Sci., 10*, 70.
[3] Clore, G.M. and Gronenborn, A.M. (**1987**) *Protein Eng., 1*, 275.

[4] Campbell, I.D. and Sheard, B. (**1987**) *Trends Biotechnol*., *5*, 302.
[5] Kim, D.E., Chivian, D. and Baker, D. (**2004**) *Nucleic Acids Res*., *32*, W526.
[6] http://www.biosiris.com/en/Online_order/PepLook/PepLook_order
[7] Simons, K.T., Kooperberg, C., Huang, E. and Baker, D. (**1997**) *J. Mol. Biol*., *268*, 209.
[8] Kim, D.E., Chivian, D. and Baker, D. (**2004**) *Nucleic Acids Res*., *32*, W526.
[9] Kessler, H. (**1982**) *Angew. Chem. Int. Ed. Engl*., *21*, 512.
[10] Smith, J.A. and Pease, L.G. (**1980**) *Crit. Rev. Biochem*., *8*, 315.
[11] Kawano, K., Yoneya, T., Miyata, T., Yoshikawa, K., Tokunaga, F., Teranda, Y. and Iwanega, S. (**1990**) *J. Biol. Chem*., *265*, 15365.
[12] Kahn, M., Wilke, S., Chen, B., Fujita, K., Lee, Y.H. and Johnson, M.E. (**1988**) *J. Mol. Recognit*., *1*, 75.
[13] Pavone, V., Lombardi, A., Saviano, M., Nastri, F., Fattorusso, R., Maglio, O., Isernia, C., Paolillo, L. and Pedone, C. (**1994**) *Biopolymers*, *34*, 1517.
[14] Shepherd, A.J., Gorse, D. and Thornton, J.M. (**1999**) *Protein Sci*., *8*, 1045.
[15] Kaur, H. and Raghava, G.P.S. (**2002**) *Bioinformatics*, *18*, 1508.
[16] Kaur, H. and Raghava, G.P.S. (**2003**) *Protein Sci.*, *12*, 627.
[17] Kaur, H. and Raghava, G.P.S. (**2004**) *Bioinformatics*, *20*, 2751.
[18] Dunbrack, R.L. and Jr.Karplus, M. (**1994**) *Nat. Struct. Biol.*, *1*, 334.
[19] Case, D.A., Pearlman, D.A., Caldwell, J.W., Cheatham III, T.E., Ross, W.S., Simmerling, C.L., Darden, T.A., Merz, K.M., Stanton, R.V., Cheng, A.L., Vincent, J.J., Crowley, M., Tsui, V., Radmer, R.J., Duan, Y., Pitera, J., Massova, J., Seibel, G.L., Singh, U.C., Weiner, P.K. and Kollman, P.A. (**1999**) *AMBER 6*, University of California, San Francisco.
[20] Bernstein, F.C., Koetzle, T.F., Williams, G., Mayer, E.F., Bryce, M.D., Rodgers, J.R., Kennard, O., Simanouchi, T. and Tasumi, M. (**1977**) *J. Mol. Biol*., *112*, 535.
[21] Saravanan, S., Khan, A.A. and Sekar, K. (**2000**) *Current Sci*., *78*, 550.
[22] Kanehisa, M. (**1997**) *Trends Biochem. Sci*., *22*, 442.
[23] Kabsch, W. and Sander, C. (**1983**) *Biopolymers*, *22*, 2577.
[24] Hutchinson, E.G. and Thornton, J.M. (**1996**) *Protein Sci.*, *5*, 212.
[25] Hutchinson, E.G. and Thornton, J.M. (**1994**) *Protein Sci.*, *3*, 2207.
[26] Jones, D.T. (**1999**) *J. Mol. Biol.*, *292*, 195.
[27] Altschul, S.F., Madden, T.L., Alejandro, A.S., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (**1997**) *Nucleic Acids Res.*, *25*, 3389.
[28] Guruprasad, K. and Rajkumar, S. (**2000**) *J. Biosci.*, *25*, 143.
[29] Thomas, A., Deshayes, S., Decaffmeyer, M., Van Eyck, M.H., Charloteaux, B. and Brasseur, R. (**2006**) *Proteins*, *65*, 889.

Received: April 22, 2007 **Accepted: May 02, 2007**