

Predoms: A Computational Toolkit Developed to Perform Prediction of Domains from Sequence

Sualeha Khan, Anna Shahid, Sahar Fazal and Atif Jilani

Department of Bioinformatics,
Muhammad Ali Jinnah University, Islamabad, Pakistan

Abstract: Function of proteins is dependent upon the domain content in them. Thus domains serve as determinants for a protein to perform its function. These domains are quite difficult to predict. PreDomS computational toolkit; predicts domains for a given sequence based on entropy parameter. In contrast to other methods for predicting domains which use 3D structural data as an input or even if they require sequence data as input but, they are based on homology searching against large, domain databases. A sole feature of PreDomS is that, along with domain prediction it provides a single platform to perform multiple operations like Transcription, Translation, Secondary Structure prediction.

Key words: Prediction • Protein • Domain • Computational toolkit

INTRODUCTION

Function of proteins plays a vital role in almost all cellular processes. Proteins have specific components in them which are called domains. They are made up of secondary structural elements [1]. Each domain performs a specific function. A protein can have multiple domains in it. Thus, function of a protein is dependent upon the type of domain content it has.

Several computational techniques have been proposed for the prediction of domains. Some of them require 3D structural data as input [2]. But the problem arises with some orphan proteins that have no 3D structural data determined yet. Thus, the ease of prediction gets limited if 3D structural data for a protein is not available. Another prediction method which requires sequence data as input seems to be quite easy to be dealt with. But this method is based on homology searching. It searches the given sequence's similarity against large databases. In this case as well if some unknown sequence is given as input then, probability for occurrence of a domain cannot be predicted [3-5].

Our toolkit named as PreDomS predicts domains when sequence data is available. Using entropy parameter the probability for the occurrence of domain in a given sequence is predicted. Through studies, it was found that amino acids present in side chain have high entropy and this is compensated by low entropy values of amino acids

which fold themselves into some confirmation [6]. In this method, conformational entropy is considered as the number of degrees of freedom on the angles Φ , Ψ and χ for each amino acid. It is also observed that hydrophilic amino acids appear to be present in side chains while, hydrophobic amino acids are involved in forming the domain or in other words they appear to be involved in attaining some confirmation.

Other than domain prediction, PreDomS allows its users to perform transcription, translation and secondary structure prediction on the same single platform to avoid the hectic task of switching between different tools. It handles three types of input (cDNA, mRNA, amino acid) sequence for Transcription, Translation and Secondary Structure prediction. Domain prediction is made, based on amino acid sequence.

PreDomS' users can maintain the records of their performed operations in the database so that if needed they can be used for future research. Each user has a unique account to ensure privacy. Users can even use the valid sequences saved in the toolkit database or provide their own input sequences by using browse option.

MATERIAL AND METHODS

Transcription and Translation: Transcription and translation are performed only on cDNA or mRNA sequence to ensure only correct output.

CONCLUSION

PreDomS employs a simple and user friendly approach for domain prediction. It has user friendly interface and provides opportunity for maintaining record of activities carried out through it as well. Instead of seeking multiple tools, users can work on a single platform while dealing with any of the four operations performed through PreDomS.

This figure shows that PreDomS can manage all four operations that are performed in it. Other than this it manages its users as well and keeps the record of operations performed by each user.

This figure shows that multiple operations like transcription, translation, secondary structure prediction and domain prediction can be performed on the single platform of PreDomS, using any of the three sequences DNA, RNA or AA.

REFERENCES

1. Gewehr, J.E. and Ralf Zimmer, 2006. SSEP-Domain: Protein Domain Prediction by Alignment of Secondary Structure Elements and Profiles. *Bioinformatics*, 22(2): 181-187.
2. ToolKit: DHcL Domain Hierarchy and closed Loop (<http://sitron.bccs.uib.no/dhcl/>).
3. Vinayagam, A., J. Shi, G. Pugalenthi, B. Meenakshi, T.L. Blundell and R. Sowdhamini, 2003. DDBASE2.0: Updated Domain Database with Improved Methods for the Identification of Structural Domains. *Bioinformatics*, 19: 1760-1764.
4. Novatchkova, M., G. Schneider R. Fritz, F. Eisenhaber and A. Schleiffer, 2006. DOUT finder-Identification of Distant Domain Outliers Using Subsignificant Sequence Similarity. *Nucleic Acids Res.*, 1: 34 (Web Server issue): W214-8.
5. George R.A. and J. Heringa, 2002. Protein Domain Identification and Improved Sequence similarity Searching Using PSI-BLAST *Proteins* 1, 48(4): 672-81.
6. Oxana V.G. and S.M. Bogdan, 2003. Prediction of Protein Domain Boundaries from Sequence Alone. *Protein Sci.*, 12(4): 696-701.
7. Galzitskaya, O.V., N.V. Dovidchenko, M. Yu, Lobanov, M. Yu and S.O. Garbuzynskiy, 2009. Prediction of Protein Domain Boundaries from Statistics of Appearance of Amino Acid Residues. *Biochemistry (Moscow)*, 74(8): 888-897.