



e-ISSN: 2319-8753 | p-ISSN: 2320-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 9, Issue 12, December 2020

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.512

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Homology Modeling and Structure Prediction of DNA-processing protein *DprA* isolated from *Frankia sp.*

Minisha Bagaria, Santosh Kumar*, Kuldip Dwivedi

Student, Department of Life Sciences, ITM University, Gwalior, M.P., India

Asst. Professor, Department of Life Sciences, ITM University, Gwalior, M.P., India*

Professor, Department of Environmental Science, Amity University, M.P., Gwalior, India

ABSTRACT: *Frankia sp. is* filamentous bacteria that grow by hyphal branching and tip extension. They produce three cell types during growth: vegetative hyphae, sporangiospores and the lipid-enveloped cellular structure known as “diazo-vesicles”. *Frankia sp.* has the ability to form symbiotic nitrogen-fixing root nodules on certain woody angiosperms termed “actinorhizal plants”. The vesicle cell type develops during N- starvation and contains the oxygen- labile nitrogenase. Actinorhizal plants have become increasingly important as climate changes threaten to remark the global landscape over the next decades. Most actinorhizal plants are capable of high rates of nitrogen fixation due to their capacity to establish root nodule. Symbiosis with nitrogen-fixing *Frankia* strains. One of these mechanisms is the induction of defense-related event’s, whose precise role in a symbiotic interaction remains to be established. The physicochemical characterization of the response regulator protein isolated from *Frankia sp.* involved in the nitrogen fixation was done to interpret properties like Isoelectric point (pI), EC, AI, GRAVY and instability index. Homology modelling was performed to generate good quality models. The assessment of generated three-dimensional structure against structure verification tools PROCHECK and WHAT IF showed that the model generated by the SWISS MODEL was more acceptable to that by Geno3D. The structure was stable and could be used for further studies in environmental eradication.

KEYWORDS: *Frankia sp.*, Actinorhizal, GRAVY, PROCHECK, Geno3D

I. INTRODUCTION

Frankia involved gram-positive and gram-variable actinomycetes which grow in hyphal form. The way like cyanobacteria and in divergence to most rhizobia, *Frankia* strains can fix N₂ in the free-living state also. Under fixed-N limitation and aerobic conditions, *Frankia* strains form special organs (spherical vesicles) for N₂-fixation. These spherical vesicles have hyphae or short side hyphae at the ends. Vesicles are enveloped by multiple layers of hopanoids. In plants, the shape, septation, and sub cellular localization of *Frankia* vesicles are decided by the host plant, which do can direct bacterial differentiation. *Frankia* can also form multilocular sporangia ^[1]. *Frankia alni* ACN14a strain was isolated from green alder (*Alnus crispa*) growing in Tadoussac, Canada. It comes under a member of the medium host range "Alnus" strains. This bacterium is able to facultative establish a nitrogen-fixing symbiosis with alder (*Alnus* spp.) and myrtle (*Myrica* spp.), as well as other plants in the *Betulaceae* and *Myricaceae*. This strain of *Frankia alni* ACN14a can be isolated from soils on all continents except Australia and Antarctica. It causes root hair deformation: it penetrates the cortical cells and induces the formation of nodules which resemble those induced by *Rhizobium* in legumes. These nodules are then colonized by vegetative hyphae (mycelium filaments) which differentiate into diazo-vesicles. It contains spherical, thick-walled cells that perform reductive nitrogen fixation and also protected from molecular oxygen by numerous layers of tightly stacked hopanoid lipids ^[2]. *DprA* and *Smf* bind combinely to single-stranded DNA and then those proteins both self-interact and after that they interact with *RecA*. *DprA-RecA-ssDNA* filaments are formed and then these filaments catalyze the homology-dependent formation of combined molecules. *Escherichia coli* ssDNA-binding protein (SSB) limits action of conversion of *RecA* to ssDNA, *DprA* hinders this barrier. *DprA* is a member of the recombination-mediator protein family, useful in natural bacterial transformation ^[3].

Genetic recombination in *Helicobacter pylori* involved in host adaptation of this gastric pathogen and uptake of DNA by natural transformation can make changes in virulence factors and antigenic variation as well. To annotate the mechanisms played in natural transformation we have to test two genes with homology to known competence genes (dprA and traG). Insertion mutants in these genes were produced in two different *H. pylori* strains and their competence by natural transformation can be compared to the wild-type. Mutation of the traG homolog will not reduce competence. Mutation of the dprA gene will impaired natural transformation in both the plasmid and the chromosomal DNA. Some data indicated that dprA and comb3 are very important parts in pathway of chromosomal and plasmid transformation^[4].

II. MATERIALS AND METHODS

- **Retrieval of target sequence-** The amino acid sequence of DNA-processing protein DprA isolated from *Frankia sp.* of 95 amino acids was obtained from the sequence database of NCBI (<http://www.ncbi.nlm.nih.gov/protein/>). It was ascertained that the three-dimensional structure of the protein was not available in Protein Data Bank^[5]. Hence the present work for developing the 3D model of the protein from *Frankia sp.* was undertaken.
- **Physico-chemical characterization-** The values of theoretical isoelectric point (pI), molecular weight, total number of positive and negative residues, extinction coefficient, instability index, aliphatic index and grand average hydropathicity (GRAVY)^[6] (<https://web.expasy.org/protparam/>) were computed using the ExPASy's ProtParam server^[7] (<http://us.expasy.org/tools/protparam.html>).
- **Secondary structure prediction-** Secondary structure^[8] has been predicted using PHYRE2^[9] software (<http://www.sbg.bio.ic.ac.uk/phyre2>) where the FASTA format^[10] of the sequence was given as input. SOPMA^[11] was employed for calculating the secondary structural features of the protein sequence considered for this study.
- **Model building and quality assessment-** The modeling of the three dimensional structure of the protein was done using two homology modeling programs, Geno3D^[12] and SWISS MODEL^[13]. The overall stereo chemical property of the protein was assessed by Ramachandran plot analysis^[14]. The evaluation of structure models obtained from the two software tools was performed by using PROCHECK^[15]. The models were further checked with WHAT IF^[16] for standard bond length and bond angle determination.

III. RESULTS AND DISCUSSION

The physicochemical parameters computed using ExPASy's ProtParam tool for the DNA-processing protein DprA isolated from *Frankia sp.* had the molecular weight 10132.71D theoretical pI 11.01, a number of negatively charged residues (Asp+Glu) equal to 7, extinction coefficient 12615, instability index 33.32, aliphatic index 86.46 and grand average of hydropathicity (GRAVY) -0.40 which characterizes the physicochemical properties of the protein represented in Table 1. A protein having instability index smaller than 40 is predicted as stable, on the other hand, a value above 40 predicts that the protein may be unstable^[17]. Instability index of 33.32 indicates the stable nature of the protein. The high extinction coefficient (12615) indicates the presence of a high concentration of Cystine, Tryptophan and Tyrosine. The aliphatic index is generally considered as a positive factor for the increase in thermal stability. High aliphatic index (86.46) of the protein suggests that the protein may be stable for a wide temperature range. The Grand Average hydropathicity (GRAVY) value is low (-0.40) which increases the possibility of interaction with water.

Table 1-Physicochemical parameters computed using ExPASy's ProtParam tool

S.No.	Parameters	Values
1.	Number of amino acids	95
2.	Molecular weight	10132.71
3.	Theoretical Pi	11.01
4.	Total number of negatively charged residues(Asp+Glu)	7
5.	Total number of negatively charged residues(Arg+Lys)	12
6.	Extinction coefficient	12615
7.	Extinction coefficient*	12490
8.	Instability index	33.32
9.	Aliphatic index	86.46
10.	Grand average of hydropathicity	-0.040

The secondary structure of DNA-processing protein DprA was predicted by two software namely SOPMA (Self Optimized Prediction Method with Alignment) and Phyre 2. SOPMA predicts around 69.5% of amino acids correctly to describe secondary structure prediction. Table 2 represents the results of SOPMA. These results shows higher number of random coils in comparison to other secondary structure elements (alpha helix, extended strand and beta turns) (Fig. 6). Default parameters (Window width: 17, similarity threshold: 8 and number of states: 4) were set by SOPMA software for secondary structure prediction. Secondary structure and disorder prediction was made using phyre 2 which is shown in figure 7.

Table 2- Calculated secondary structure elements by SOPMA

S.No.	Parameters	Value
1.	Alpha helix	38.95%
2.	3_{10} helix	0.00%
3.	Pi helix	0.00%
4.	Beta bridge	0.00%
5.	Extended strand	18.95%
6.	Beta turn	12.63%
7.	Bend region	0.00%
8.	Random coil	29.47%
9.	Ambiguous state	0.00%
10.	Other state	0.00%

Three dimensional structures of proteins are predicted due to unavailability of such data. There is no experimental structure found for the protein considered. The three dimensional structure prediction of the protein was done using two programs Geno3D and Swiss Model. The results obtained from these programs were compared in table 3. This comparison suggests that the model generated by Swiss model (Fig. 1) is much better than the model generated by Geno3D due to presence of maximum residues (68.00%) in the most favoured region. Finally chosen model was visualized by Rasmol. The evaluation of predicted structure generated by SWISS MODEL for the stereochemical quality was done using Ramchandran plot (Fig. 2-Fig. 5), calculations done with the PROCHECK. The 38.95% of residues were found in the core right handed alpha helices(A), beta sheets (B)and left-handed alpha helix(L) region. 10.7% of residues were found in the allowed right- handed alpha helix (a), beta sheets (b) and left -handed alpha helices regions. The 0.7% of the residues were found in the generously allowed alpha helices (~a), beta sheets (~b), left handed alpha helices (~l) and epsilon (~p) regions. The 0.0% of the residues was found to be localized at the disallowed regions. Above results indicate to a good quality of predicted model. The analysis of modeled structure of DNA-processing protein DprA by structure verification server WHATIF revealed such RMS Z-score (0.7888) which was nearly meeting the required score for a high quality model.

Table 3 - Ramachandran plot calculation and Comparative analysis of the models from Geno3D and SWISS -model computed with the PROCHECK program

Server	Parameters	Value (%)
Geno 3D	Residues in the most Favoured Region	68.00%
	Residues in additionally allowed region	30.00%
	Residues in generously allowed region	2.00%
	Residues in disallowed region	0.00%
Swiss Model	Residues in the most Favoured Region	95.58%
	Residues in additionally allowed region	55.32%
	Residues in generously allowed region	1.53%
	Residues in disallowed region	0.08%

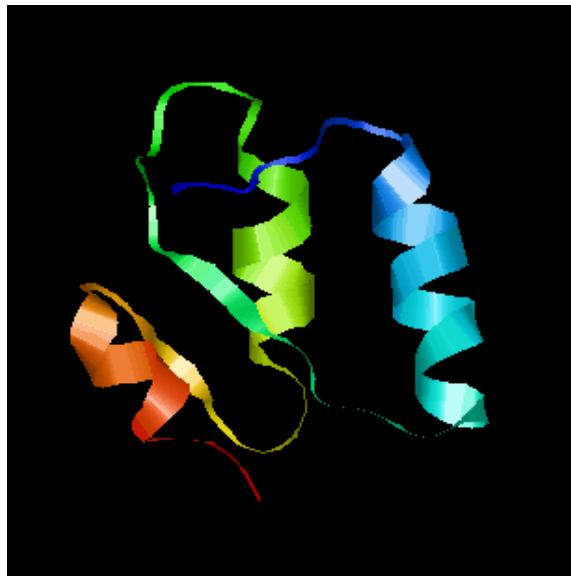


Fig 1: Structure generated by SWISS MODEL software

& Structure generated by Geno 3D software

Ramachandran Plot generated by Geno 3D

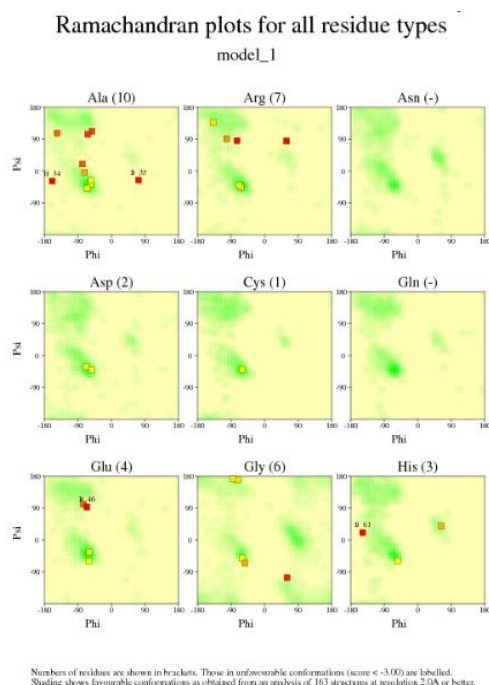
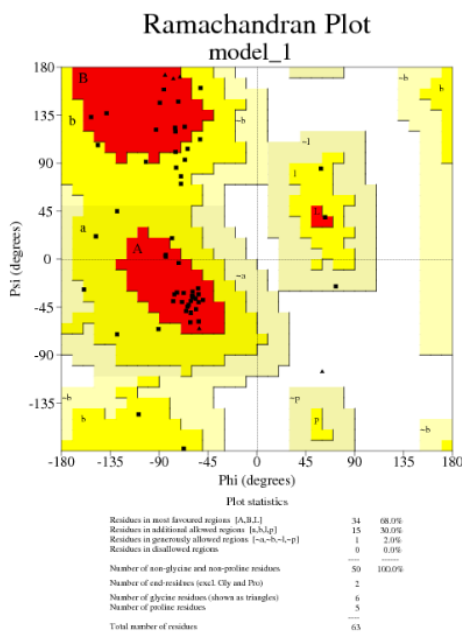


Fig 2: Ramachandran plot for all residues

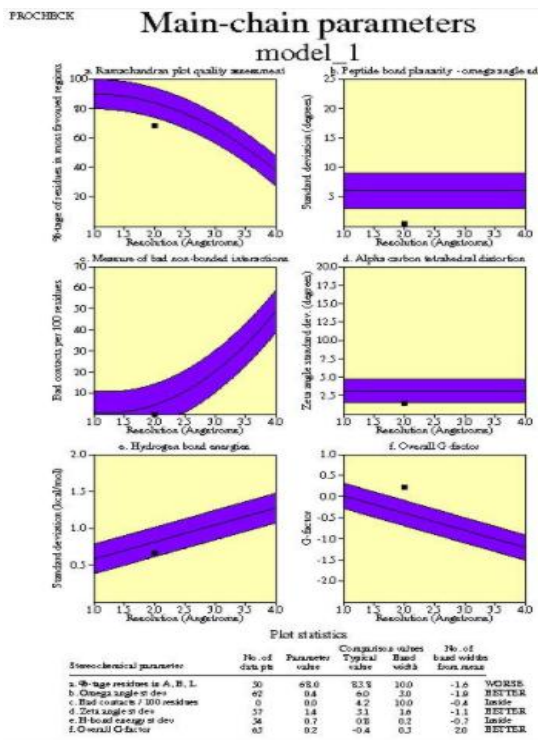


Fig 3: Main chain parameters

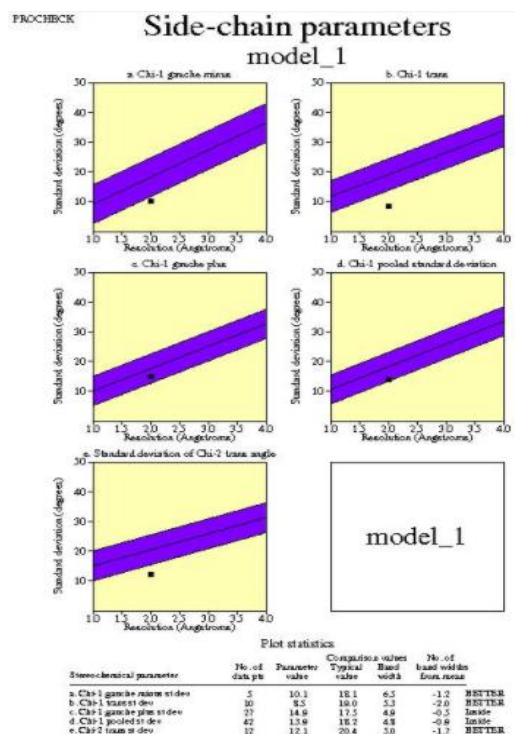


Fig 4: Graph for side chain residues

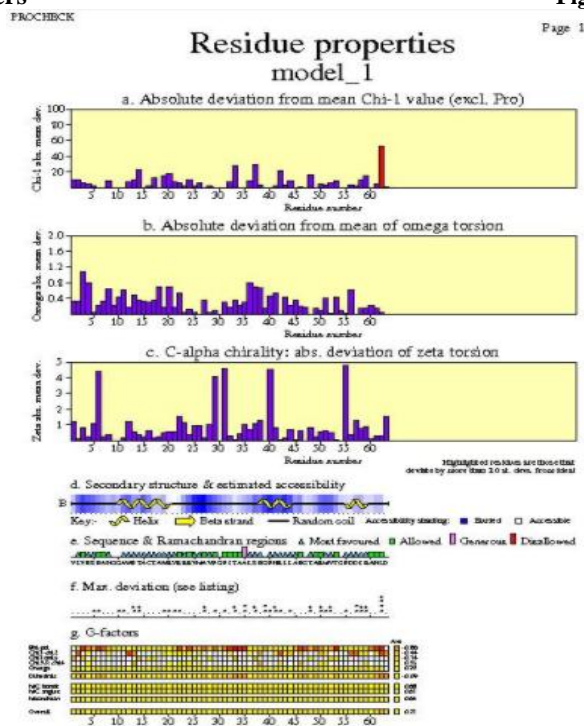


Fig 5: Graph for residual properties

SOPMA RESULTS

Significant improvement in protein secondary structure prediction by consensus prediction from multiple alignments.

10 20 30 40 50 60 70

| | | | | | |

MVCVTAPGTARHGLHALARARTLAVLTRGTVLVESPAHGGAWETA
CTAWRYRRRVMAVPGPITAALSEGP

eeeecttcccthhhhhhhhhhhhhtceeeeeeccctcchhhhhhhhhhtceeeccccchhhccc

HRLLAEGTARMVTGPDDIRAHLDRT

hhehttcceeeccchhhhhhhht

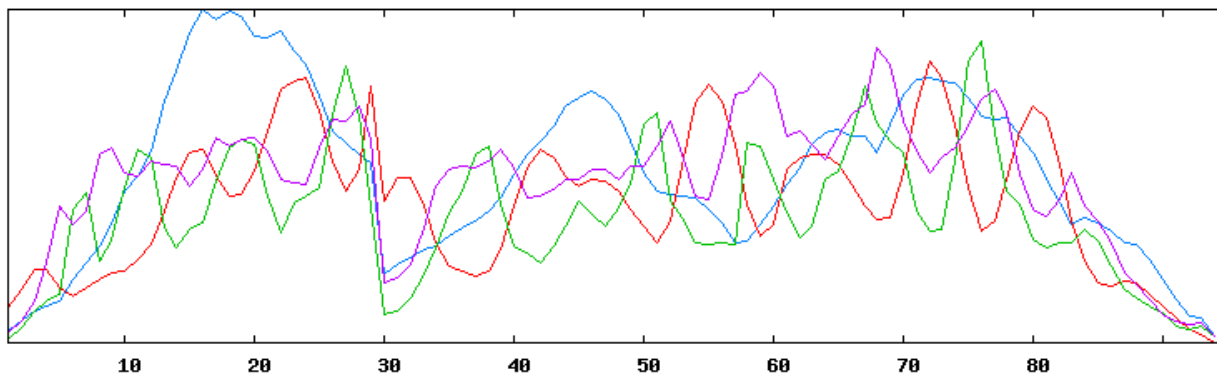


Fig. 6: SOPMA results showing deviations, coils, helix and sheet

PHYRE 2 Result:

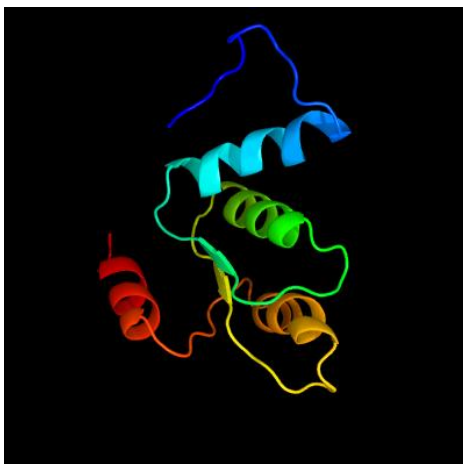


Fig. 7: Secondary structure generated by Phyre 2 software



IV. CONCLUSION

On the basis of various structural and physiochemical parameters assessment, this can be concluded that the predicted three dimensional structure of DNA-processing protein DprA is stable. The Ramchandran plot data showed that nearly 68% of the residues for each protein lies in the core region. The maximum percentage of the residues lying in the core region represents the best stereo chemical quality of the protein structure which is further supported by its nearly 68% residues lying in the most favoured and allowed regions. 0.00 % residues lie in the disallowed regions which generally involve steric hindrance between the side chain C-beta methylene group and main chain atoms. Since the percentage residues lying in the disallowed region is much lower than its presence in core and allowed regions the structures generated show best stereochemical quality and stability. The three dimensional structures generated using SWISS Model software and validated by PROCHECK software showed the best identity, lowest energy in Kcal/mol and root mean square deviation (RMSD) were considered as the best optimized and superimposed model.

V. ACKNOWLEDGEMENTS

We would like to acknowledge ITM University, Gwalior for its timely support and also the groups who supported us directly or indirectly in making this research a success. We would also thank the software without whose support this work would have not been completed.

REFERENCES

1. Katharina Pawlowski, Birgitta Bergman, Plant Symbioses with Frankia and Cyanobacteria in Biology of the Nitrogen Cycle, 2007.
2. Normand P., Lapierre P., Tisa L.S., Gogarten J.P., Alloisio N., Bagnarol E., Bassi C.A., Berry A.M., Bickhart D.M., Choisine N., Couloux A., Cournoyer B., Cruveiller S., Daubin V., Demange N., Francino M.P., Goltsman E., "Genome characteristics of facultatively symbiotic Frankia sp. strains reflect host range and host plant biogeography." *Genome Res.* 17:7-15(2007).
3. Isabelle Mortier-Barrière, Marion Velten, Pauline Dupaigne, Nicolas Mirouze, Olivier Piétrement, Stephen McGovern, Gwennaële Fichant, Bernard Martin, Philippe Noirot, Eric Le Cam, Patrice Polard, Jean-Pierre Claverys "A key presynaptic role in transformation for a widespread bacterial protein: DprA conveys incoming ssDNA to RecA".
4. L C Smeets, J J Bijlsma, E J Kuipers, C M Vandenbroucke-Grauls, J G Kusters. "The dprA gene is required for natural transformation of *Helicobacter pylori*".
5. RCSB. Biological Macromolecular Structures Enabling Breakthroughs in Research and Education.
6. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* 157: 105- 132.
7. Gasteiger E (2005) Protein Identification and Analysis Tools on the Expasy Server. In: John M. Walker, The Proteomics Protocols Handbook, Humana Press 571-607.
8. Geourjon C, Deléage G (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl. Biosci.* 11: 681-684.
9. Lawrence A Kelley, Stefans Mezulis, Christopher M Yates, Mark N Wass & Michael J E Sternberg. The Phyre2 web portal for protein modeling, prediction and analysis 2015.
10. William R. Pearson, Finding Protein and Nucleotide Similarities with FASTA 2016 Mar 24.
11. C. Geourjon, G. Deléage SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments *Bioinformatics*, Volume 11, Issue 6, December 1995, Pages 681–684.
12. Combet C, Jambon M, Deleage G, Geourjon C (2002) Geno3D: Automatic comparative molecular modelling of protein. *Bioinformatics* 18: 213-214.
13. Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22: 195-201.
14. Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology* 7: 95-99.
15. Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *Journal of Biomol. NMR* 8: 477-486.
16. Vriend G (1990) WHAT IF: A molecular modeling and drug design program. *J. Mol. Graph* 8: 52-56.
17. Ken A. Dill, S. Banu Ozkan, M. Scott Shell, and Thomas R. Weikl The Protein Folding Problem 2008 Jul 3



INNO SPACE
SJIF Scientific Journal Impact Factor

**Impact Factor:
7.512**

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details