ISSN: 0975-833X

Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 6, Issue, 02, pp.5290-5293, February, 2014

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

RESEARCH ARTICLE

HOMOLOGY MODELING OF SALMONELLA TYPHIMURIUM CRP PROTEIN

***Ghorpade Prabhakar B. and Bhaskar Sharma**

Div. of Animal Biochemistry, Indian Veterinary Research Institute, India

ARTICLE INFO

Article History:
Received 18th November, 2013
Received in revised form
10th December, 2013
Accepted 15th January, 2014
Published online 28th February, 2014

Key words:
ProSA, crp salmonella,
Qmean and Anolea.

ABSTRACT

Understanding protein modeling is key to protein structure analysis. The current research includes crp salmonella typhimurium protein for modeling. The main criteria in model generation is generation of model based on available structures and then knowing model quality generated in process. Swiss prot gives automated model with model quality estimation by Qmean and Anolea. The other dedicated model quality estimation server used is ProSA. The protein model determination is helping in various protein- protein interaction studies, drug-taget interaction. The removing fear of protein model can give better insight into problems based on interaction study. This step by step research would will promote Research into protein study for biologist who are not well aqucuiented with protein modeling.

INTRODUCTION

The modeling of protein helps to understand protein three dimensional shape. If any raw amino acid sequence is there we can determine three dimensional shape of the molecule based on homology modeling. The homology refers to similarity of structure between two structures. The sequences having same amino acid sequences mostly gives similar 3D structure. The currently it is said that three dimensional structure are conserved than sequence conservation. Current research is for giving understanding of available online resources for protein model prediction and validation. The swiss modeling provides platform for model determination by homology modeling Benkert and Biasini (2011). The cAMP receptor Protein (CRP) is catabolic activator protein. CRP is sensitive to cAMP concentration in body. Upon binding by cAMP the CRP tightly binds to promoters of gene. This activate protein transcription. The main role of CRP is in energy metabolism in body (Busby Ebright (1999). We have taken Salmonella typhimurium bacterial crp protein sequence for model generation. Salmonella typhimurium causes typhoid like disease in mice and major gastrointestinal disturbances in animals as well as humans.

MATERIAL AND METHODS

NCBI Search of Protein sequence

The NCBI protein sequence selected is Crp (cAMP Receptor Protein) (NCBI Reference Sequence: NP_462369.1). The sequence available in standard fasta format is selected for modeling (http://www.ncbi.nlm.nih.gov/protein/NP_462369.1?report=fasta). The downloaded FASTA sequence can be opened in Text editor. Crp is 210 amino acid sequence.

**Corresponding author: Ghorpade Prabhakar B.*
Div. of Animal Biochemistry, Indian Veterinary Research Institute, India.

Swiss Modelling

In Swiss Model (http://swissmodel.expasy.org/) there are four major modeling mode Automated mode, Alignment mode, Project mode, Beta-swiss model. For this study we have created myWorkspace (Arnold K., Bordoli L., (2006)) and submitted automated mode protein model. Swiss Model workspace gives personalized space to monitor modeling projects.

ProFunc

Profunc identifies biochemical function of protein from its three-dimensional structure. The parameters that could be studied include fold matching, residue conservation, surface cleft analysis, and functional 3D templates and identifying active site or possible homologues in PDB.

ProSA

It refers to protein structure analysis. In this model quality of protein is estimated. It uses only c-alpha atoms as input structure. It mainly studies low resolution structures, early structures.

Model validation

Results

Swiss prot

The Swiss prot Model report shows that modeled reidue range for given protein is 2 to 208. The model generated is based on template 1i5zB with 1.90A^{o} resolution. The sequence identity

with available protein sequence is 99.52%. The quality QMEAN Z-Score is -0.33. The swiss prot model for crp is shown in Fig.1. QMEAN for crp is shown in Fig.2 and Fig.3, Table 1. Homology Model generation is not without error and the error is represented in Fig.4, Fig.5.

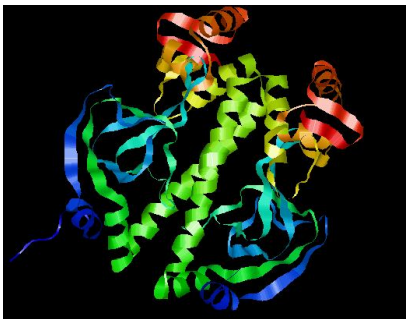

Fig.1. Actual model generated in Swiss prot with coloring of chains

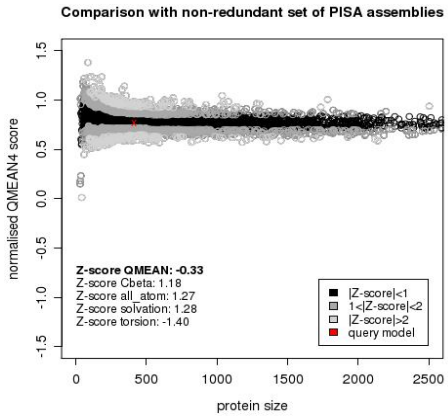

Fig. 2. Swiss Prot Protein model quality estimation

Table 1. Swiss Prot QMEAN4 global scores Chart

The QMEAN4 score is a composite score consisting of a linear combination of 4 statistical potential terms (estimated model reliability between 0-1). The pseudo-energies of the contributing terms are given below together with their Z-scores with respect to scores obtained for high-resolution experimental structures of similar size solved by X-ray crystallography:

Scoring function term	Raw score	Z-score
C_beta interaction energy	-233.28	1.18
All-atom pairwise energy	-15250.05	1.27
Solvation energy	-62.13	1.28
Torsion angle energy	-71.72	-1.40
QMEAN4 score	0.764	-0.33

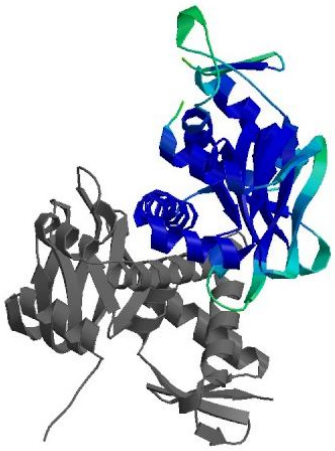

Fig. 4. Swiss Prot Protein model with coloring by residue error

ProSA:- ProSA is Protein Structure analysis model validation software. The ProSA Z score for our protein of interest is -7.28. The Figure

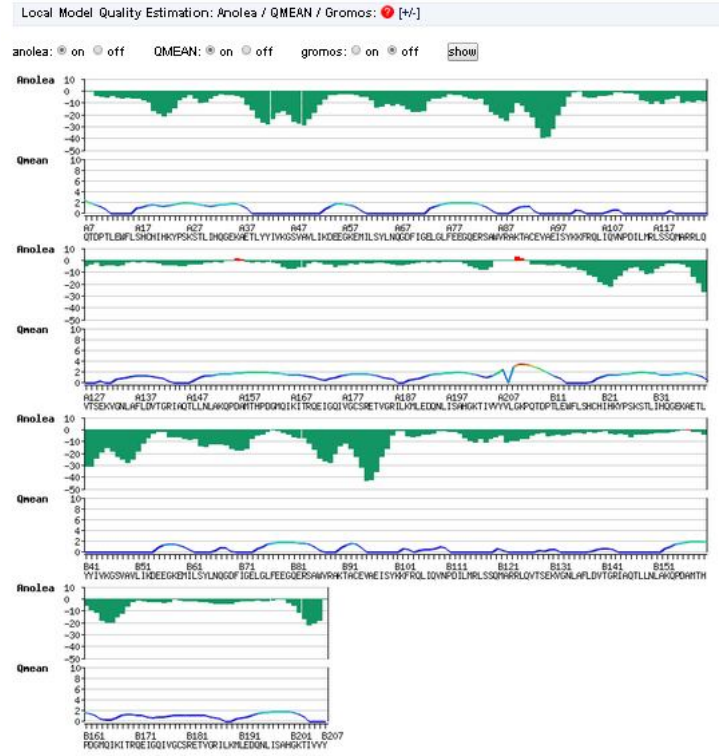

Fig.3. Swiss Prot Model Estimation by QMEAN and Anolea

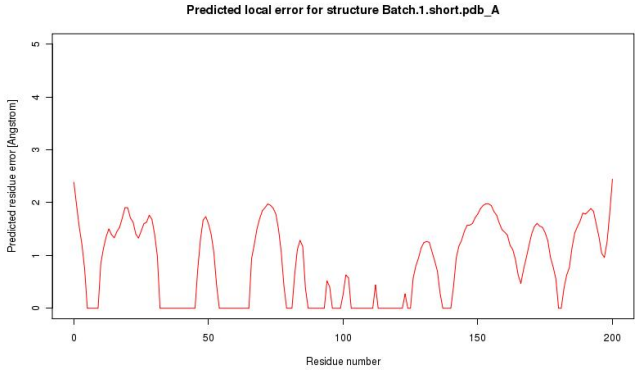

Fig.5. Error Plot for residues

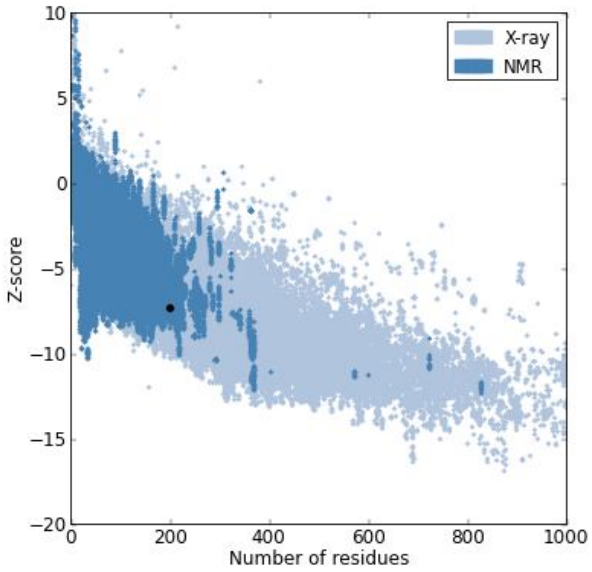

Fig.6. Overall Model Quality estimation for crp by ProSA

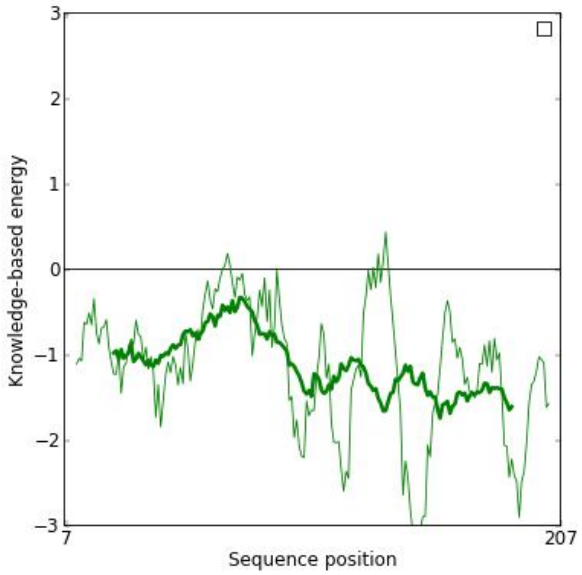

Fig.7. Local Model quality estimation for crp by PrpSA

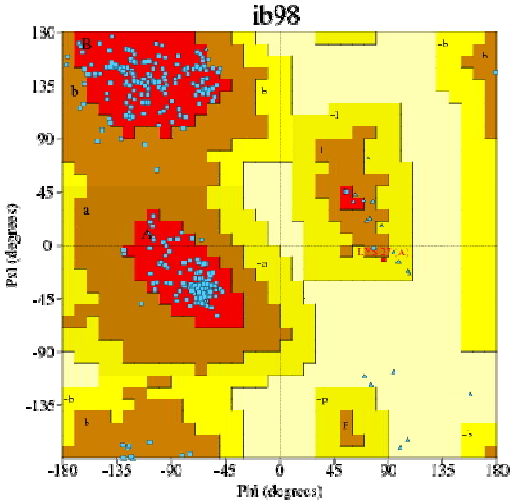

Fig.8. Ramachandran Plot for crp Protein

PROCHECK statistics

Table 2. Ramachandran Plot statistics

S. No.	Parameter	No. of residues	%-tage
1.	Most favoured regions [A,B,L]	334	91.8%
2.	Additional allowed regions [a,b,l,p]	29	8.0%
3.	Generously allowed regions [~a,~b,~l,~p]	1	0.3%
4.	Disallowed regions [XX]	0	0.0%
5.	Non-glycine and non-proline residues	364	100.0%
6.	End-residues (excl. Gly and Pro)	3	
7.	Glycine residues	30	
8.	Proline residues	11	
	Total number of residues	408	

This analysis was based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and *R*-factor no greater than 20.0a good quality model would be expected to have over 90% in the most favoured regions [A,B,L].

Table 3. G-Factors

Parameter	Average Score	Score
Dihedral angles:-		
Phi-psi distribution	-0.03	
Chi1-chi2 distribution	0.33	
Chi1 only	0.37	
Chi3 & chi4	0.56	
Omega	-0.12	
Main-chain covalent forces:-		
Main-chain bond lengths	0.49	
Main-chain bond angles	0.37	
	0.42	
OVERALL AVERAGE	0.23	

G-factors provide a measure of how **unusual**, or out-of-the-ordinary, a property is.
Values below -0.5* - unusual
Values below **-1.0**** - highly unusual

Important note

The main-chain bond-lengths and bond angles are compared with the Engh and Huber (1991) ideal values derived from

small-molecule data. Therefore, structures refined using different restraints may show apparently large deviations from normality.

DISCUSSION

For any structure determination true structure is one which is determined by experimental methods. The X-ray differaction, NMR (Nuclear Magnetic Resonance) gives true experimental structure. If three dimensional model is not available then prediction of model is done. 3D Prediction is based Homology modeling, Threading or Ab-initio. The three dimensional structure by homology modeling is created based on known three dimensional structure. In threading full experimental structure is not available so based on available folds for matching regions of sequences structure is determined. Other is Ab-initio structure in this structure of protein is determined based on molecular energy calculation by Schrodinger, not using experimental parameters. In Swiss Prot Model (Schwede Kopp (2003) quality is estimated by Qmean and Anolena. In Our case QMEANscore4 is 0.76. Absolute model quality is -0.33. Score component are shown for protein from Negative -4 to positive +4 for our protein based on CB interaction, all atom interaction, salvation and torsion. Coloring by Residue color Z-score = -0.56. The error prediction for generated model is shown on error generation plot. Z-Score is statistical representation of scores relationship to mean score. A 0 Z score indicate it is same as mean. The Z score can be negative or positive. In our ProSA Wiederstein and Sippl (2007), Sippl, (1993) we have got -7.28 Z score, indicating that it is below mean. Z-Score helps to identify overall model quality. Z score helps to give understanding as where model protein fall in comparison to experimental structure. The experimentally determined protein are shown by different color. In Our case light blue color indicate X-Ray determined structure. Whearas dark blue indicate X-Ray determined structure. ProSA local model quality indicate energy as function of amino acid sequence position. The amino acids are plotted on X-axis and knowledge based energy is used on Y axis to indicate each residue position energy estimation.

From Ramachandran Plot Analysis (Fig.8, Table 2 and Table 3) we know that our score for most favoured residue is 91.8% which is greater than 90% so our model shows good model quality. In current research we have selected cAMP receptor protein as modeling molecular sequence. The cAMP protein is important in virulence determination of bacteria and taken as exemplary sequence for current analysis. The overall average G factor less than 0.5 indicating unsual properties for model bond angle.

Conclusion

This research will help to give clear and precise understanding of amino acid to protein three dimensional structure determination to biologist.

REFERENCES

Arnold K., Bordoli L., Kopp J., and Schwede T. 2006. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modeling. *Bioinformatics*, 22,195-201.

Benkert P, Biasini M, Schwede T. 2011. "Toward the estimation of the absolute quality of individual protein structure models." *Bioinformatics*, 27(3):343-50.

Busby S., Ebright RH. 1999. Transcription activation by catabolite activator protein (CAP). *J. Mol. Biol*. 293: 199–213

Guex, N. and Peitsch, M. C. 1997. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis 18: 2714-2723.

Schwede T, Kopp J, Guex N, and Peitsch MC 2003. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Research 31: 3381-3385.

Sippl, M.J. 1993. Recognition of Errors in Three-Dimensional Structures of Proteins. Proteins 17, 355-362.

Wiederstein and Sippl 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Research 35, W407-W410.

****