Volume : 2 | Issue : 8 | August 2013 • ISSN No 2277 - 8179

Research Paper

Homology Modeling and Structural Analysis of DNA Binding Response Regulator of *Bacillus anthracis*

Bioinformatics

KEYWORDS : Homology Modeling, Model Evaluation, Energy Minimization, Structural Superposition

Mr. Amajala Krishna Chaitanya	Assistant Professor, Department of Biochemistry/Bioinformatics, GITAM Institute of Science, GITAM University, Visakhapatnam-530045, Andhra Pradesh, India
Dr. I. Bhaskar Reddy	Professor, Department of Biochemistry/Bioinformatics, GITAM Institute of Science, GITAM University, Visakhapatnam-530045, Andhra Pradesh, India
Ms. P. Minakshi Kumari	Student, Department of Biochemistry/Bioinformatics, GITAM Institute of Science, GITAM University, Visakhapatnam-530045, Andhra Pradesh, India
Mr. Kunal Zaveri	Research Scholar, Department of Biochemistry/Bioinformatics, GITAM Institute of Science, GITAM University, Visakhapatnam-530045, Andhra Pradesh, India
Dr. DSVGK. Kaladhar	Assistant Professor, Department of Biochemistry/Bioinformatics, GITAM Institute of Science, GITAM University, Visakhapatnam-530045, Andhra Pradesh, India

ABSTRACT

DNA binding response regulator plays a vital role in the signal transduction and various cellular functions. It is widely expressed in various beings such as Mycobacterium tuberculosis, Streptococcus pneumoniae as well in Bacillus anthracis. As it is very essential protein involving in the two-component regulatory system metabolism, it can be considered as one of the drug target for the Anthrax disease. As the three dimensional structure is not resolved for B. anthracis DNA Binding response regulator, we made an attempt to model the protein structure by selecting appropriate template structure through Modeller software. The obtained homology modeled structure is subjected to secondary structure and transmembrane prediction analysis. The modeled protein is further subjected to energy minimization & model evaluation studies with Swiss-Pdb Viewer & Procheck, Errat respectively. The energy minimized modeled protein is further subjected to structure superposition studies in which the target & template protein structures are aligned by computing the RMSD (Root Mean Square Deviation) value.

1. Introduction

DNA-binding response regulator is a phosphoprotein involved in various biological processes such as two-component regulatory system, regulation of transcription and DNA-dependent intra cellular signal transduction. Usually the two-component regulatory proteins are involved in a system responding to environmental changes characterized usually by a sensor kinase in the cell membrane that phosphorylates itself in response to a signal and a response regulator to which the phosphoryl group is transferred. The responder is typically a DNA-binding protein that regulates transcription. Several of these systems are quite complex, involving many proteins in a signaling cascade or contributing to several responses simultaneously. They are involved in a variety processes such as chemotaxis, osmoregulation, magnesium transport, pH tolerance, sporulation, or response of virulent species to host cell's environments [1]. As this protein is involved in the sporulation as well other vital cellular processes it can be considered as key drug candidate for the organisms such as Bacillus anthracis, Mycobacterium tuberculosis, etc. This protein three dimensional structure is not yet resolved for the B. anthracis. Hence our present study involves the modeling of the B. anthracis DNA Binding Response Regulator protein and its structural analysis.

2. Methodology

2.1. Retrieval of Target Protein Sequence using SwissProt

SwissProt is a curated protein sequence database which provides a high level of annotations such as the description of the function of a protein, its domains structure, post-translational modifications, variants, binary interactions, catalytic/active sites, structural details, etc with a minimal level of redundancy and high level of integration with other databases. The protein sequence of DNA binding response regulator (Target Protein) is retrieved with the accession number C3LB48 which consists of 229 amino acids.

2.2. Template Selection by PDB-BLASTP Analysis

The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches [2]. By selecting the PDB as database name, BLASTP is performed to the target protein sequence with the default parameters. The maximum sequence alignment identity or highest query coverage obtained in the PDB-BLASTP result is selected as template PDB structure to model the target protein.

2.3. Comparative Secondary Structure & TMHMM Prediction

The template and target protein sequences were retrieved and secondary structure and transmembrane helices were predicted using GOR IV and TMHMM version 2.0 respectively.

2.4. Target Protein Modeling using MODELLER 9.10

Three Dimensional Homology model was built using the crystal structure coordinates of template protein. All the steps of homology modeling were carried out using MODELLER software (VER: 9.10).

2.5. Energy Minimization Studies Using SWISS-PDB Viewer

Energy minimization method is used to generate lower energy conformations of a protein structure. The lower energy of the chemical system (protein structure) infers the greater stable conformation. During modeling of the protein structures, some of the faulty bonded and non-bonded interactions lead to the errors in the geometry of the structure. To optimize the errors, energy minimization process is essential. The steepest descent algorithm is employed for the geometry optimization process.

2.6. Comparative Model Evaluation Before and After Energy Minimization

An important part of the homology model building is the evaluation of the predicted model; this mainly involves the analysis of geometry distribution in the models. The reliability of the predicted structure was tested using the evaluation tools such as PROCHECK and ERRAT.

2.6.1. Procheck

PROCHECK checks the stereochemical quality of a protein structure, producing a number of PostScript plots analyzing its overall and residue-by-residue geometry. It includes PROCHECK-NMR for checking the quality of structures solved by NMR [3].

Volume : 2 | Issue : 8 | August 2013 • ISSN No 2277 - 8179

2.6.2. Errat
ERRAT is a protein structure verification algorithm that is especially well-suited for evaluating the progress of crystallographic model building and refinement. The program works by analyzing the statistics of non-bonded interactions between different atom types. A single output plot is produced that gives the value of the error function vs. position of a 9-residue sliding window [4].

2.7. Structural Superimposition of Both Target & Template Structures using Swiss-Pdb Viewer
By using Swiss-Pdb Viewer the proteins can be superimposed in order to deduce structural alignments and compare their active sites or any other relevant parts. The term structural superposition refers to rotations and translations performed on one molecular structure to make it match another structure or structures. The structural similarity can be computed in the terms of RMSD [8] (Root Mean Square Deviation. Lower the RMSD values, higher the structural similarity.

3. Results & Discussion
3.1. Target-Template Sequence Alignment
DNA binding response regulator (target protein) shares 55% sequence identity with the Chain A, Micarec Ph7.0 from Streptococcus pneumoniae, (PDB Id: 1NXO) with the maximum score of 117 and quiet less sequence coverage of 50%. As the query sequence coverage is less, the maximum query coverage is considered in selecting the template. The maximum query coverage sharing is 99% by Chain A, The structure of the response regulator Regx3 from Mycobacterium tuberculosis (PDB Id: 2OQR) with the highest score of 171 and sequence identity of 42%. Another factor to select the 2OQR PDB structure as template is it is composed of 230 amino acids.

3.2. Protein Secondary Structure & Transmembrane Helices Prediction
The secondary structure features for the template structure were predicted as: α helices 44.98%, β strands 0.00% and coils 37.12%. For Target structure, the features predicted are: α helices is 33.48%, β strands 0.00% and coils 44.35 % (Table 1). The percentage values of secondary structural elements between target & template are not varying in larger range. Hence the template 3D structure is considered for the modeling of the target 3D structure. No sign of transmembrane helices were predicted for both target & template protein sequences using TMHMM.

3.3. Homology Model of the Target Protein (DNA Binding Response Regulator)
As the target protein, DNA binding response regulator doesn't have the 3D structure in PDB, the molecule is modelled using the Modeller 9.10 with the help of template retrieved from PDB-BLAST result. Homology Modelling of the target protein give rise to five different conformational structures (Table 2). Among the five different energy based conformations the third modelled structure which possessing the least DOPE (Discrete Optimized Protein Energy) value of -23223.36 is further subjected to energy minimization & structural evaluation studies. (Fig 1)

3.4. Energy Minimization & Model Evaluation Studies
The reliability of the predicted structure (homology modeled) was tested using the energy evaluation tools like PROCHECK and ERRAT. The model is evaluated before energy minimization and after energy minimization process. In before energy minimization step, the modeled structure possesses the molecular energy of 527.288 kj/mol. After the energy minimization step, the modeled structure possesses the molecular energy of -8202.371 kj/mol which is the lowest energy with comparison to the before energy minimized modeled structure (Table 3c).

3.5. Quality of Proposed Model (before & after energy minimization)
Using different validation tools, quality of model was assessed. According to the results of Procheck for before energy minimization of the modeled protein, there are 90.4% residues in most favored region (core), 6.7% allowed regions, 1.4% generously allowed regions and 1.4% disallowed region, whereas the results for after energy minimization of the modeled protein, are 88.0% residues in most favored region (core), 9.6% allowed regions, 1.0% generously allowed regions and 1.4% disallowed regions (Table 3a, b). From the Errat results, the overall quality factor of the modeled protein is analyzed.

3.6. Structural Superposition Studies
The structural superposition refers to the structural alignment of two or more protein structures. RMSD (Root mean square deviation) was calculated by superposition of Cα traces and backbones onto the template crystal structure using Swiss-Pdb Viewer. The RMSD between target homology model and template calculated for Cα traces is 3.75A°. (Fig 2)

4. Conclusion
The three-dimensional model of DNA Binding Response Regulator protein has been generated based on the homology of Crystal Structure of the response regulator Regx3 from Mycobacterium tuberculosis (PDB Id: 2OQR chain A) with the highest score of 171 and sequence identity of 42%. The secondary structural analysis also revealed that the model has a common percentage range of helix and sheets regions. The energy minimization process resulted in generation of stable conformation of the modeled protein. The RMSD values suggest that the model contains a similar topology comparatively to its template. The present study would aid in studying the active site, protein-protein interactions, ligand-docking studies.

Acknowledgement:
We would like to thank GITAM University for providing Bioinformatics lab facility to carry out the research work.

Tables
Table 1. Comparative Secondary Structure Prediction

Secondary Structure	Template	Target
α helices	44.98%	33.4%
β sheets	0.00%	0.00%
Coils	37.12%	44.35%

Table 2. Target structure modeled with five different conformations

Model	Zdope	DOPE
1	-0.18	-22930.91
2	-0.16	-22866.13
3	-0.27	-23223.36
4	-0.22	-23079.81
5	-0.23	-23113.95

Table 3a. PROCHECK Statistics*

	No. of Residues		%	
Most favoured region [A,B,L]	188	183	90.4	88.0
Additional allowed regions [a,b,l]	14	20	6.7	9.6
Generously allowed regions [~a, ~b, ~l, ~p]	3	2	1.4	1.0
Disallowed regions [XX]	3	3	1.4	1.4
Non glycine and non proline residues	208	208	100.0	100.0
End residues (excl,gly & pro)	2	3	-	-
Glycine residues	11	11	-	-
Proline residues	7	7	-	-
Total number of residues	228	229	-	-

Table 3b. G-Factors*

Parameter	Score		Avg. Score	
Dihedral angles	-	-	-	-
Phi-psi distribution	0.04	2.64	-	-
Chi1-chi2 distribution	-0.11	-0.14	-	-
Chi 1 only	0.16	0.01	-	-
Chi3 & chi4	0.34	0.29	-	-
Omega	-0.21	-0.39	-0.02	0.69
Main chain covalent forces	-	-	-	-
Main-main bond lengths	-0.14	0.38	-	-
Main-main bond angles	-0.23	-0.06	-0.19	0.12
Overall coverage	-	-	-0.08	0.49

***The before Energy Minimization values are indicated in bold and after minimization values are indicated in normal.**

Table 3c. Energy Minimization Studies using Swiss-Pdb Viewer

Before Energy Minimization						
Bonds	Angles	Torsion	Improper	Nonbonded	Electrostatic	Total
1613.299	1493.965	1003.378	454.04	1422.04	-5460.19	527.288
After Energy Minimization						
406.453	962.453	1083.622	397.604	-4761.81	-6290.69	-8202.37

Fig 1. Homology Modeled Protein Structure of DNA Binding Response Regulator

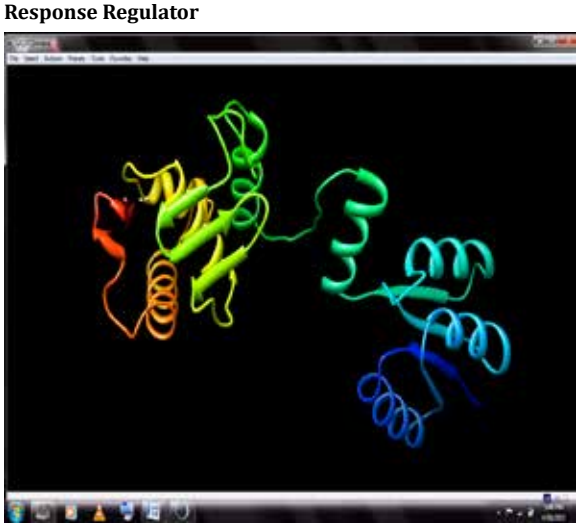
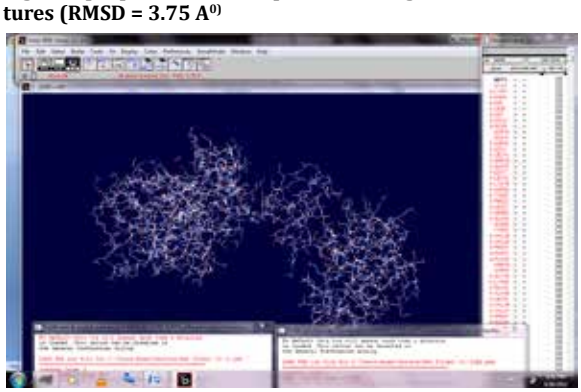


Fig 2. Superposition of Template and Target Protein Structures (RMSD = 3.75 A$^{0)}$



REFERENCE

1. http://www.uniprot.org/keywords/KW-0902 | 2. http://blast.ncbi.nlm.nih.gov/Blast.cgi | 3. http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/ | 4. http://nihserver.mbi.ucla.edu/ERRATv2/ | 5. Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ. (1990). Basic local alignment search tool. J Mol Biol, 215(3):403-10. | 6. Sali A & Blundell TL, (1993). Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol, 234(3):779-815. | 7. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, and Thornton JM. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR. Dec;8(4):477-86. | 8. Zhang Y & Skolnick J. (2004) Scoring function for automated assessment of protein structure template quality. Proteins, Dec 1;57(4):702-10. |