

In silico approach to develop structure and functional analysis of response receiver regulator protein of the strain pseudomonas fulva 12x

Pipasha Biswas

Department of Biochemistry and Bio-technology University of Science and Technology Chittagong

Abstract: *In total world population, most of the people depend on rice as a staple food because it gives proper amount of nutrition, which help to people in leading healthy life mainly in the developing countries of Asia. This valuable plant is attacked by various types of micro-organism like bacteria. Under the pseudomonas species, pseudomonas fulva 12-x is one of the strain, contain receiver domain include regulator protein, which ultimately help them to maintain signal transduction pathway.so the aim is to find out new target site in this hazardous bacteria to develop a potential medicine that help in saving our crops without using fertilizer because fertilizer causes various types of environmental problem. Very little amount of information is known about the structural biology of pseudomonas fulva 12-x,so using homology modeling of bioinformatics approach the tertiary structure of response receiver regulator protein of pseudomonas fulva 12-x has determined. The model building was done by swiss model workspace. Then the model was evaluated and validated using PROCHEK, Verify 3d,PROSA and ERRAT in these the verify 3d value was 0.70 suggest surrounding profile is good an (PROSA) determined the disallowed region contain 0.0% and allowed region contain 93.5% and most favored region contain 6.5% amino acid, which implies perfect stereochemical quality of the model. Then physico-chemical characterization, secondary structure and functional site prediction were determined using Expasys protparam server, GOR4, SOPMA and Scan prosite tools. Finally, the active site was analyzed using CASTp server, which predicted target region and help in developing medicine for various rice diseases like bacterial PANICLE BLIGHT disease in further lab test.*

I. Introduction

Bacterial panicle blight is a seed borne diseases of rice and gradually making a big problem in Arkansas since 1995 when it hugely destroyed Bengal and cypress cultivators' state wide. Rod shaped bacteria of soil cause hampers of these rice grain and stops crop production by up to 60%. Various fertilization process has been suggested to do but no use give appropriate result [1]. The bacterium *Pseudomonas fulva* is a member of *pseudomonas pudina* group, which is regarded as a soil bacterium group. *pseudomonas fulva* are the main bacterium of causing all kind of rice diseases, because they were collected mainly from rice grains [2][3]. In the cellular control process, phosphorylation of protein is a very important event [4][5] because it mediate most of the organism in signal transducing system like bacteria, microbial eukaryote & higher plants [7][8] protein phosphorylation is a central to the function of two large protein families, one is response regulator protein & another is haloacid dehalogenases superfamily of hydrolase.[6]

Two component regulatory signaling system is the most common process in higher plants such as dicoid plant, but a slight knowledge has been gathered about the response receiver regulator protein of monocoid plant like rice. Cytokinin is a plant hormone and mediates in plant growth and development. It acts as signaling system like bacterial two component signaling system [9][10][11]. Generally using the regulatory system bacteria & plant both adapt itself in any environmental condition [12]. Reasons behind the name of two component regulatory system are the presence of input & output system. Hisdine kinase which act as input system for receiving all kind of signal, which mediate bacteria to adapt in any type of environment feature like osmotic condition, pH, temperature or to resist any complex hazardous ionic compound dissolved in that soil condition. This inputting system also sucks various macro and micro nutrient, so that remain balance in the concentrations. Another name of it is sensor kinase that holds some important characteristics like presence of transmembrane helices and continuous settlement of sensor system at the N-terminal input domain form in the cytoplasmic membrane layer of the protein.

When a signaling component binds with environmental stimulus, it evolves stimulation and this signal is passed from N-terminal input domain to c-terminal output domain. In this middle time, signal which coming from input domain cause intramolecular structure alteration, that lead to the activation of cytoplasmic carrier domain [13][14][15][16]. This carrier activates nearest the receiver domain of response regulator protein. Response regulator is output system where phosphorylation related reaction occurs. Using a divalent cation both phosphorylation & dephosphorylation reaction occur in receiver domain. So phosphorylation in the active site of

this receiver domain, an altered form of stimulation from the expressed target gene is come out from regulatory system that control various expression according to its needs. [17]

In this study, we determined the 3d structure of the response receiver regulator protein that is signaling protein for pseudomonas fulva strain 12-x and identified the active site for potential medicine target using in silico approach.

II. Material and Method

2.1. Computational Methods

The computational process of 3d model develops included template identification, alignment of template with the target, model build up and creation of the structure. The sequence of response regulator receiver protein of pseudomonas fulva 12-x (Accession ID:-YP_004474011.1) was obtained from NCBI database.

2.2. Template identification and sequence alignment

Basic local alignment tool (Blast) [18] that was modified and maintained in Adam Godzik,s laboratory at the Burnham institute perform to identify the best template where it hits against the PDB(Brookhaven protein Databank)(19) database. In PSI-BLAST, a feature retrieved from nr protein database for proteins containing similarity to the unknown structure was used to find out PDB database.

2.3. Model Build Up

The model build up of response receiver regulator protein of pseudomonas fulva 12-x as a PDB file was done using CPH model 3.0 [20] server program which is homology modeling method related, then it was submitted to swiss model workspace server [21][22] to gain the 3D homology structure.

2.4. Evolution and validation of model

The Program PROCHEK [23] by ramachandan plot analysis is used to define the stereo chemical quality of the generated model and it was validated by verify 3D [24][25] and ERRAT programs [26]. To check the 3D structure of the generated protein model PROSA tool was used [27].

2.5. Physical and chemical characterization

Expasys protparam server is widely used in determining the physico-chemical characteristics of the generated protein model [32].

2.6. Secondary structure and functional site determination

Two self-optimized determination methods SOPMA [36] and GOR IV [35] Program were mediated to develop secondary structure of the protein. The functional site present in response receiver regulator protein of pseudomonas fulva 12-x was calculated using Expasy's scan prosite tool [33].

2.7. Active site prediction

To identify the ligand binding capacity with the determined model, CASTp server [37] was used. The predicted active site in generated model will help in further work to study on docking site.

III. Result

3.1 Template choice with sequence alignment

The structure of the response receiver regulator protein of pseudomonas fulva 12-x was performed based on the availability of the template sequence of the signaling protein from helicobacter pylori (PDB CODE:-3H1G) [9].

3.2 3d Structure

The 3D model of response receiver regulator protein of pseudomonas fulva 12-x was generated by homology modeling on the template, which is collected from protein data bank [figure 1]

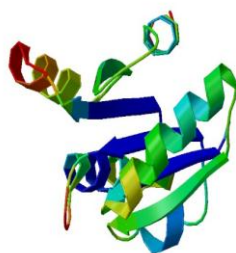
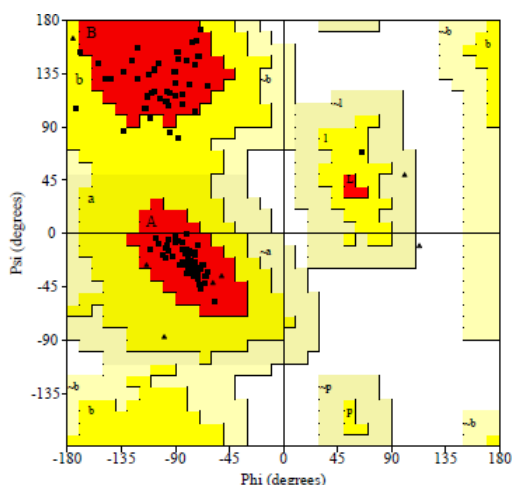


Fig 1: 3D structure of the response receiver regulator protein of Pseudomonas fulva strain 12-x

3.3 Evolution of model



Residues in most favoured regions [A,B,L]	100	93.5%
Residues in additional allowed regions [a,b,l,p]	7	6.5%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	107	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	7	
Number of proline residues	4	
Total number of residues	120	

Figure 3: Ramachandran plot analysis of response receiver regulator protein of Table 1:

Pseudomonas fulva 12-x was calculated using PROCHEK

3.4 Verification of 3D model

The profile score above zero in the verify 3D graph suggest the approvable environment profile of the model [figure 4]. ERRAT [26] analyzes the statistics of non-bonded interactions between different atom types and plots the value of the error function versus position of a residue [figure 5]. The results PROSA program [27] used for the display of scores and energy plots that highlight potential problems spotted in the modeled protein structure are shown in [figure 6].



Figure 4: verify 3D GRAPH of response receiver regulator protein of Pseudomonas fulva 12-x obtained using modeling tool.

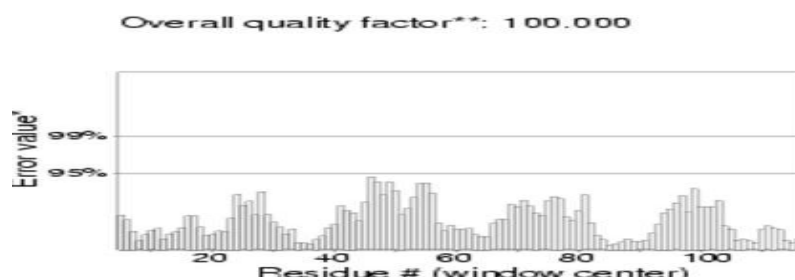


Figure 5: plot statistics value of response receiver regulator protein of *Pseudomonas fulva* 12-x obtained using (ERRAT) modeling tool. Expressed as the percentage of the protein for which the calculate error value falls below the 95% rejection limit. Good high resolution structures generally produce values around 95% or higher. For lower resolution (2.5 to 3Å) the average overall quality factor is around 91%.

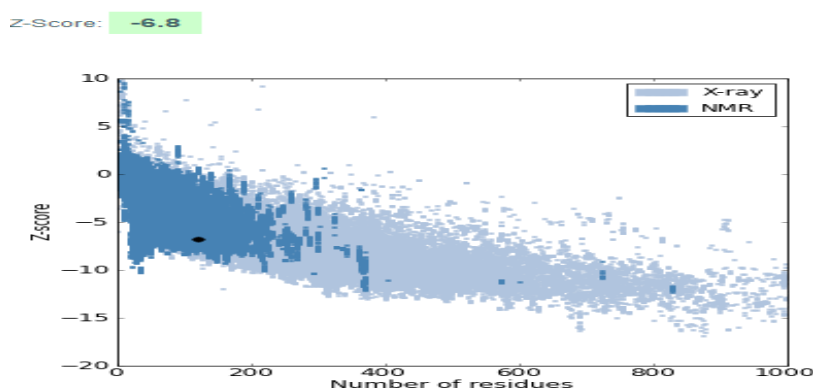


Figure 6: Z score for the response receiver regulator protein of *Pseudomonas fulva* 12-x obtained from modeling tool PROSA

3.5 Physical and chemical characterization:

The physico-chemical characteristics is used to know number of amino acids, molecular weights, total number of negatively and positively charged residues, Theoretical pI (isoelectric point), Extinction coefficient (EI) [28], Instability index(II) [29][30], grand average of hydropathicity (GRAVY) and aliphatic index (AI) [31]. These features are determined using Expasys protparam server, which are presented in table 2

Number of amino acids	124
Molecular weight	13882.1
Theoretical pI	5.45
Total number of negatively charged residues(Asp+Glu)	17
Total number of positively charged residues(Arg+Lys)	16
Extinction co efficients	6990
Instability index	26.93
Aliphatic index	103.79
Grand average of hydropathicity(GRAVY)	-0.001

Table2: Physico-chemical features value of response receiver regulator protein of *pseudomonas fulva* 12-x obtained using modeling tool.

3.6 Secondary structure and functional site prediction

The secondary structure analysis is used to know the composition of alpha helix strand and beta coil in the modeled protein. Secondary structural characteristics are obtained using SOPMA and GOR IV is presented in table 3 and 4. Using fault parameter of SOPMA in which window width

Secondary structure	(%)
Alpha helix	46.58
3_{10} helix	0.00
Pi helix	0.00
Beta bridge	17.74
Extended strand	10.48
Beta turn	0.00
Bend region	25.00
Random coil	0.00
Ambiguous states	0.00
Other states	0.00

TABLE 3: Computed secondary structural elements of response receiver regulator protein of pseudomonas fulva 12-X by SOPMA.

is 17, similarity threshold is 8 and number of states is 4 and GOR IV, the secondary structure is determined. The output of using Scanprosite tool that is recorded in terms of the length of amino residues of modeled protein with specific profile and pattern is represented in table 3.

Secondary structure	(%)
Alpha helix	47.58
3_{10} helix	0.00
Pi helix	0.00
Beta bridge	13.71
Extended strand	0.00
Beta turn	0.00
Bend region	38.71
Random coil	0.00
Ambiguous states	0.00
Other states	0.00

Table 4: Computed secondary structure of response receiver regulator protein of pseudomonas fulva 12-x by GOR IV

Length of amino residues of modeled protein with specific profiles & pattern	Functional region position
<p>MKILIVDDFSTMRRRIKNLLRDLGFTNTAEADDGVTALPMLQSGNFDLVTWNPGMTGIDLLRA</p> <p>VRADERLKQLPVLMTAEAKRDQIEAAQAGVNGYVVKPFTAQVLKEKIEKIFERVNG</p>	2-119 amino acid

Table 5: Functional site present in response receiver regulator protein of Pseudomonas fulva 12-x obtained using Scan Prosite tool in Expasys

3.7 Active site prediction

CASTp server is used for active site determination of response receiver regulator protein of Pseudomonas Fulva 12-x (figure 6), then also reported the best active site area and the position of amino acid and other amino acid involved in it [figure 7]

Chain 0
1- MKILIVDDFS TMRRIKNLL RDLGFTNTAE ADDGVTALPM LQSGNFDL V
51- TDWNMPGMTG IDLLRAVRAD ERLKQLPVLMTAEAKRDQIEAAQAGVNG
101- YVVKPFTAQVLKEKIEKIFE

Figure 6: Active site prediction by CASTp server, where Green color shows the active site position from 5 to 106 with the Beta sheet in between them.

	ID	Area	Vol
<input type="checkbox"/>	8	98.7	85
<input checked="" type="checkbox"/>	7	126.6	82.4
<input type="checkbox"/>	6	57.5	45.6
<input type="checkbox"/>	5	27.7	22.4
<input type="checkbox"/>	4	16.5	10.9
<input type="checkbox"/>	3	28	12.3
<input type="checkbox"/>	2	9.1	5.7
<input type="checkbox"/>	1	24.7	11.5

Figure 7: The table shows the area and the volume

For different active sites in response receiver regulator protein Of *Pseudomonas fulva* 12-x and the best active site area exist in an area of 126.6 and a volume of 82.4.

IV. Discussion

In the field of bioinformatics, homology modeling algorithm is regarded as an interesting site for any biological experiment and in other silico work planning. Homology modeling is the appropriate method to estimate structure related protein molecule and functional information. For 3D structure generation, need to homology search which is done from PDB to identify template structure then alignment of target sequence and template structure from nr protein Database searched by BLASTp. Various protein model designing software are developing day by day, with raising number of known protein structure. In this study, Model build up is done using different program which suggest that the proteins contains alpha helix, beta strand and loop. The program CPH model 3.0 server was used to save query sequence as PDB file, then submit to another web based software ,name swiss –prot workspace that reduce the problem of downloading and installing wide range of program packages and databases [40]. PROCHEK program check the Stereo chemical quality and perfection of the developed models that shows amino acid percentage in the favorable region was 93.5% and in the disallowed region was 0.0%. In total 100% of the amino acid residues present in favored and allowed region that was illustrated in the ramachandan plot (figure2) with statics(table 1). Verify 3D server show the compatibility of an atomic model (3D) with its own amino acid sequence(1D) by assigned a structural class depend on its position and environmental profile (Alpha sheet, beta loop, polar, non-polar etc.) and determine the better qualityful structure.

Another validation program ERRAT shows the overall quality factor for the modeled protein is 100% which indicate good resolution structure (fig 3).

The verify 3D graph shown the value 0.70 (fig 4), where 3D-1D score greater then 0.2 suggest the environmental condition for side chain are acceptable. Modeled protein was also examined by X-ray-NMR studies, which shown the point of the structure within a range. Energy plot for modeled protein implies using “PROSA” tool [27], where Z score is -6.8 and represent native conformation quality is good (fig 5).

The Physico-chemical characteristics of the modeled protein is calculated using Expasy’s protparam tool(Table 2) [39], where Isoelectric point value less then 7 suggest the protein as acidic in nature and suitable for developing buffer system to purify using isoelectric focusing method. The protein also contain almost nearest number of positively and negatively charged residues. On the other side, EXPASY’S protparam computes the extinction coefficient for this protein at 280nm wavelength, because this wavelength helps protein to absorb strongly, while other substances can’t absorb like that in the protein solution. Extinction coefficient value is 6990M⁻¹cm⁻¹ at 280nm measured in water which also suggests the protein can’t be analyzed using X-ray and NMR spectral method but not using UV spectral methods. Another parameter is instability index determine the stability of this modeled protein in a test tube. Molecular weight value of this protein help to calculate an instability index. Naturally,if the instability index value less then 40 is indicated as stable, a value above 40 indicate as unstable. This instability index value for this protein is 26.93, the result suggest the response receiver regulator as stable protein. On the basis of Aliphatic index Expasy’s protparam, indicate highest value of response receiver regulator protein that predicates the protein remain stable for a large range of temperature. Aliphatic index for the response receiver regulator receiver of *pseudomonas fulva* 12-x is 103.79.The Grand average hydropathy (GRAVY) value for a protein is calculated as the sum of hydropathy values of all the amino acids divided by the number of residues in the sequences. This parameter gives the value -0.01 that suggest the possibility of better interaction of the modeled protein with water at lower range value [39]. The secondary structure is determined using GOR IV and sopma shown in table 3 and 4 from which it is clear that random coil,alpha helix and extended strand ,all are present in predominant amount that is the best site of a good structure. The output of scan prosite tool [34]was calculated in association of the length of amino residues of

protein with specific profile and pattern that guidance in molecular function determination. Important information can be gathered, if the binding site of protein is determined, which help to suggest in experimental technique that predict 46.58% alpha helix, 25 % random coil, 17.74% extended strand according to SOPMA and 47.58% alpha helix, 38.71% random coil, 13.71% extended strand according to GOR IV tools, presence in the complex modeled protein structure.[38]To calculate pockets and voids on 3D structure of the modeled protein and indicate the active site use CASTp program that is online tool. The program predicts best active site has an area of 126.6 and volume 82.4(figure 7). The best active site position of the modeled protein as well as the amino acid in the active site shown in blue color starting from amino acid 5 to some interaction with amino acid 106.The main coverage of amino acid present in building cavities between 5-7.

V. Conclusion

The knowledge of homology modeling & functional region searching tool act as a fundamental way that will give various possible application of modeled protein of pseudomonas fulva 12-x help of enzyme inhibition assay in further experiment. In this study, the native conformation create purpose of modeled protein, give the answer about ligand binding affinity and show effort in designing mutagenesis experiment. All the information collect from above studies will assist to explore the search on perfect inhibitor of response receiver regulator protein from pseudomonas fulva 12-x,that can be used in future plant medicine development project.

Reference:-

- [1]. Mary Hightower. Resistance against bacterial panicle blight in rice. Delta farma press. 2011.
- [2]. Renault D, Deniel F, Benizri F, Sohier D, Barbier G, Rey P. Characterization of Bacillus and pseudomonas strains with suppressive traits isolated from tomato hydroponic-slow filtration unit. Can J Microbiol. 2007;53:784-97.
- [3]. Uchino M, Shida O, Uchimura T, Komagata K. Recharacterization of Pseudomonas fulva Iizuka and Komagata 1963, and proposals Pseudomonas parafulva sp. nov. and Pseudomonas cremoricolorata sp. nov. J Gen Appl Microbiol. 2001;47:247-261.
- [4]. Post RL, Kume S. Evidence for an aspartyl phosphate residue at the active site of sodium and potassium ion transport adenosine triphosphate. J Biol Chem. 1973; 248:6993-7000.
- [5]. V Weiss, B Magasanik. Phosphorylation of nitrogen regulator I (NRI) of Escherichia coli. Proc Natl Acad Sci U S A. 1988;85:8919-8923.
- [6]. Hoch JA, Silhavy TJ. Two component signal transduction (Am. Soc. Microbiol., Washington, DC) 1995
- [7]. Djordjevic S and A.M Stock. Structural analysis of bacterial chemotaxis proteins: components of a dynamic signaling system. J. Struct Biol. 1998;124:189-200.
- [8]. Falke JJ, Bass RB, Butler SL, Chervitz SA, Danielson MA. The Two-component signaling pathway of bacterial chemotaxis: a molecular view of signal transduction by receptors, kinases and adaptation enzymes. Annu Rev Cell Dev Biol. 1997;13:457-512.
- [9]. Mukesh Jain, Akhilesh K Tyagi, Jitendra P Khurana. Molecular characterization and differential expression of Cytokinin-responsive type-A response regulators in rice (Oryza sativa). BMC Plant Biology. 2006;6:1.
- [10]. Mok DW, Mok MC. Cytokinin Metabolism and Action. Annu Rev Plant Physiol Plant Biol. 2001;52:89-118.
- [11]. Davies PJ. Plant hormone, Biosynthesis, signal transduction, Action. The Netherlands: Kluwer Academic press; 2004.
- [12]. West AH, Stock A.M. Histidine kinases and response regulator proteins in two component signaling system. 2001. Trends Biochem Sci; 26:369-376.
- [13]. Hoch, J.A., and T.J. Silhavy (ed.). Two component signal transduction. ASM press, Washington, D.C. 1995.
- [14]. Inoué M and R. Dutta (ed.). Histidine kinases in signal transduction. Academic press, San Diego, Calif. 2003.
- [15]. Wolanin, P.M., P.A. Thomason and J.B. Stock. Histidine protein kinases: key signal transducers outside the animal kingdom. Genome Biol. 2002; REVIEWS301.
- [16]. Schultz, J., F. Milpetz, P. Bork and C.P. Ponting. SMART, a simple modular architecture research tool: identification of signaling domains. Proc. Natl. Acad. Sci. USA. 1998;95:5857-5864.
- [17]. Bourret RB. Receiver domain structure and function in response regulator protein. Curr Opin Microbiol. 2010;13:142-149
- [18]. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403-10 [pubmed]
- [19]. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic acids Res. 2000;28:235-42. [pub med]
- [20]. Nielsen M, Lundergaard C, Lund O, Petersen TN. CPH models-3.0. Remote homology modeling using structure guided sequence profiles. Nucleic Acids Res. 2010;38(Web server issue) w576-81.
- [21]. Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T. Protein structure homology modeling using SWISS-MODEL workspace. Nat protoc. 2009;4:1-13
- [22]. Amod K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics. 2006;22:195-201.
- [23]. Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR programs for checking the quality of protein structures solved by NMR. J Biomol NMR. 1996;8:477-86.
- [24]. Bowie JU, Lüthy R, Eisenberg D. A method to identify protein sequence that fold into a known three-dimensional structure. Science. 1991;253:164-70.
- [25]. Lüthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. Nature. 1992;356:83-5
- [26]. Colovos VC and Yeast TO. Verification of protein structures: pattern of nonbonded atomic interactions. Protein sci. 1993;2:1511 [pubmed]
- [27]. Sippl MJ. Recognition of errors in three-dimensional structures of proteins. Proteins. 1993;17:355 [pubmed]
- [28]. Gill SC & von Hippel PH. Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem. 1989;182:319.
- [29]. Guruprasad K et al. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in-vivo stability of a protein from its primary sequence. Protein Eng. 1990;4:155.
- [30]. Ikai A. Thermostability and aliphatic index of globular proteins. J Biochem. 1980;88:1895.
- [31]. Kyte J & Doolittle RF. A simple method for displaying the hydrophobic character of a protein. J Mol Biol. 1982;157:105.

- [32]. Wilkins MR et al. Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol*. 1999;112:531.
- [33]. Gattiker A, Gasteiger E and Bairoch A. ScanProsite: a reference implementation of a PROSITE scanning tool. *Appl Bioinformatics*. 2002;1:107-108.
- [34]. ED castro, et al. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res*. 2006;34:w362-w365.
- [35]. Garnier J, Gibrat J-F, Robson B. GOR secondary structure prediction method version IV. *Methods in Enzymology*. 1996;266:540-553.
- [36]. Geourjon C, Deleage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments comput. *Appl Biosci*. 1995;11:681-4.
- [37]. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpoz Y and Liang J. CASTp: Computed atlas of surface topography of proteins with structure and topographical mapping of functionally annotated residues. *Nucleic Acids Res*. 2006;34(web server issue):w116-8.
- [38]. Neelmathi E, Vasumathi E, Bagyalakshmi S, Kannam R. In silico prediction of structure and functional aspects of a hypothetical protein of *Neurospora crassa*. *J Cell Tissue Res*. 2009;9:1989-94.
- [39]. Suseep Roy et al. Structure prediction and functional characterization of secondary metabolite proteins of *Ocimum*. *Bioinformatics*. 2011;6:315-319.
- [40]. M ullah, J Hira et al. A bioinformatics approach for homology modeling and binding site identification of triosephosphate isomerase from *Plasmodium falciparum* 3D7. *J Young Pharm*. 2012;4:261-266.