

# Three dimensional (3D) structure prediction and substrate-protein interaction study of the chitin binding protein CBP24 from *B. thuringiensis*

Ujala Sehar<sup>1</sup>, Muhammad Aamer Mehmood<sup>1\*</sup>, Salman Nawaz<sup>2</sup>, Shahid Nadeem<sup>1,3</sup>, Khadim Hussain<sup>1</sup>, Iqra Sohail<sup>1</sup>, Muhammad Rizwan Tabassum<sup>1</sup>, Saba Shahid Gill<sup>1</sup> & Anam Saqib<sup>1</sup>

<sup>1</sup>Department of Bioinformatics and Biotechnology, Faculty of Science & Technology, Government College University, Faisalabad, Pakistan; <sup>2</sup>Department of Biosciences, COMSATS Institute of Information Technology, Islamabad, Pakistan; <sup>3</sup>Nuclear Institute for Agriculture and Biology, Faisalabad, Pakistan; Muhammad Aamer Mehmood - Email: draamer@gcuf.edu.pk; \*Corresponding author

Received July 24, 2013; Accepted July 25, 2013; Published August 07, 2013

## Abstract:

*Bacillus thuringiensis* is an insecticidal bacterium whose chitinolytic system has been exploited to improve insect resistance in crops. In the present study, we studied the CBP24 from *B. thuringiensis* using homology modeling and molecular docking. The primary and secondary structure analyses showed CBP24 is a positively charged protein and contains single domain that belongs to family CBM33. The 3D model after refinement was used to explore the chitin binding characteristics of CBP24 using AUTODOCK. The docking analyses have shown that the surface exposed hydrophilic amino acid residues Thr-103, Lys-112 and Ser-162 interact with substrate through H-bonding. While, the amino acids residues Glu-39, Tyr-46, Ser-104 and Asn-109 were shown to have polar interactions with the substrate. The binding energy values evaluation of docking depicts a stable intermolecular conformation of the docked complex. The functional characterization of the CBP24 will elucidate the substrate-interaction pathway of the protein in specific and the carbohydrate binding proteins in general leading towards the exploration and exploitation of the prokaryotic substrate utilization pathways.

**Keywords:** *B. thuringiensis*, CBP24, homology modeling, Molecular docking

## Background:

Chitin is a ubiquitous biopolymer in ecosystem but is not being accumulated despite of its recalcitrant nature that refers to existence of an efficient chitin degradation system. So, it is not astonishing that a variety of proteins have evolved with indistinct variations in structure and organization, to degrade the plentiful biopolymer on Earth. Microorganisms such as, *Serratia marcescens*, *Serratia proteamaculans*, *Streptomyces tendae* Tü109, *Streptomyces olivaceoviridis*, *Streptomyces reticuli*, *Bacillus amyloliquefaciens* and *Bacillus thuringiensis* produce chitin binding protein(s) [1-9] with the several chitinases [1, 10, 11]. The family 33 chitin binding proteins (CBM33) are believed to interact with chitin in crystalline form, making it easily available for degradation by chitinases [12], where some of

them can specifically bind to  $\alpha$ -chitin [13] while others prefer  $\beta$ -chitin. The ChbB from *B. amyloliquefaciens* [7] and CBP21 from *S. marcescens* [13] preferably bind  $\beta$ -chitin. On the other hand, some chitin binding proteins have shown synergistic action with chitinases either specifically or non-specifically [14]. Three chitin binding proteins (CBPs) from *Artemia parthenogenetica* (a shrimp) have been found involved in oviparous development. Moreover, it was found that these proteins are involved in the embryonic cuticular layer and all three CBPs have shown carbohydrate-binding activities [15]. Nonetheless, chitin binding proteins have been found involved in adhesion strategy of some pathogenic bacteria, including *E. coli*, *Listeria monocytogenes* and *Vibrio cholerae* during the intestinal inflammation in humans [16]. So, it seems that chitin binding

proteins are important not only for biomass turnover, but they have crucial roles to play in different metabolic pathways.

The computational packages are frequently being used these days for the sequence analyses and characterization of proteins. Various structural and physicochemical properties of proteins can be better exploited by using computational tools. Although precise and accurate structure of proteins can be guaranteed by experimental methods only yet they have the disadvantage of being expensive, time consuming and large amount of purified protein is required for this purpose. Computational methods are an excellent and cost effective alternative, in this context. Despite of the fact that they are not as much reliable as experimental ones, still they can provide us nearly exact structure of proteins besides the deep understanding of structure-function relationship and substrate-protein interaction of protein at almost no cost. By far the most authentic and precise solution to the problem of protein structure prediction is template-based modeling [17]. Extensive expertise in the structural biology and the use of extremely particularized computer programs are the prerequisite of modeling of protein structures for each of the individual steps of the modeling process [18].

In the present study, the chitin binding protein CBP24 *Bacillus thuringiensis* serovar *konkukian* was characterized in an *in-silico* study. The protein was subjected to several online and desktop based bioinformatics tools to study physico-chemical properties. The structure reflects function, so the 3D structure of the protein was established using homology modeling approach. The predicted structure was subjected to structure evaluation and validation to authenticate the results. Later, the protein was docked using chitin hexamer as a ligand to study the substrate-protein interaction sites. The study will lead us to better understanding of the substrate-protein interaction principles and mechanism.

## Methodology:

The *in-silico* analyses of CBP24 were carried out using HP630 i-5 workstation having CentOS as operating system. The whole study involves various desktop based applications for comparative modeling and docking including MODELLER (v9.9) [19], Chimera [20]; an extensible molecular modeling system, Autodock (v4.2) [21] and PyMol. Amino acid sequence of the target protein was retrieved from Uniprot [22] (accession number Q6HHR5) in FASTA format. SignalP4.0 server was used to identify the cleavage site of extra cellular transport signal site. The physico-chemical parameters of the protein sequence that includes amino-acid and atomic compositions, molecular weight and isoelectric point (*pI*) were computed by ProtParam program available at ExPASy ([www.expasy.org](http://www.expasy.org)). Secondary structure analyses of the query protein were performed by PSIPRED Server [23]. The sequence is then submitted to Pfam database and also to Conserved Domains Database at NCBI for domain prediction and analyses. The presence of particular motifs that reflects the specific functions of the proteins was searched by Motif Search Library.

## Molecular modeling and validation

The sequence of predicted domains was searched for their structural similarity with the query protein by running PSI-

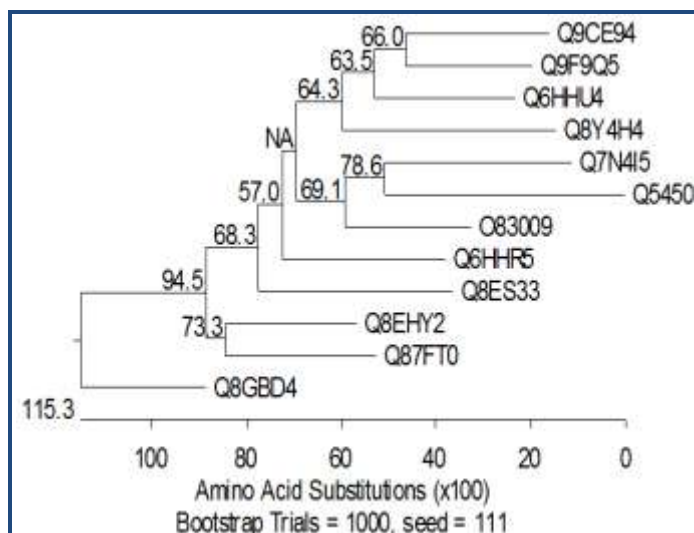
BLAST against Protein Data Bank (PDB) [24]. The resulting hits were short listed on four criteria by eliminating hits with e-values greater than 0.01, alignment length shorter than 85% of target sequence or lower functional similarity and origin of protein. The template with maximum score and least e-value (smaller the e-value, greater is the confidence) were selected as templates for homology modeling.

After the selection of template, the alignment between template and target sequence was generated by "align2d" function of MODELLER. The align2d implements global dynamic programming with an affine gap penalty function and is preferred for aligning a sequence with structures because it tends to place gaps in a better structural context [19]. Domain wise multiple sequence alignment of the chitin binding protein CBP24 was performed to see the conserved residues in predicted domain using multiple templates retrieved from family (CBM33) members. Once a target-template alignment was constructed, MODELLER calculated 3-D models of the target completely automatically, by using its automodel class. The Lowest Objective Function is used to select the best model by the smallest value of normalized Discrete Optimized Molecule Energy (DOPE) score. Loop optimization function of MODELLER9.9 was used for the loop modeling of the CBP24 based on satisfaction of spatial restraints. The high-resolution X-ray rotamers library of CHIMERA was used to model side chains that would have least stereochemical clashes.

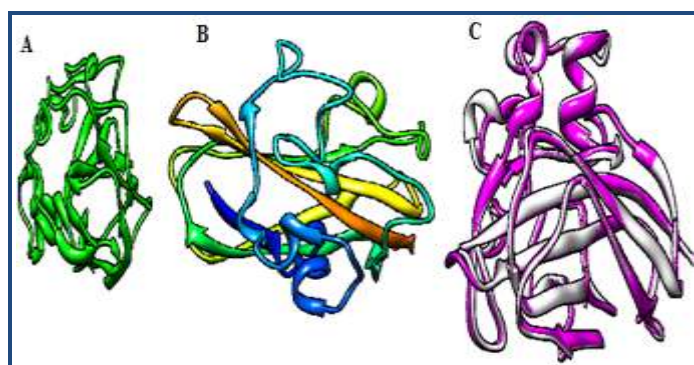
Structures were analyzed for their stereochemical properties through MolProbity and NIH server. Molprobity server computed Ramachandarn values for dihedral angles, poor rotameric conformations, C<sub>β</sub> deviation, bad angles and bond lengths of all the residues. NIH server embeds various evaluation tools like PROCHECK [25], ERRAT [26] and VERIFY 3D [27]. Structure-structure alignment of models with their respective templates was assessed by FATCAT server. The outputs of a structural alignment are superimpositions of the atomic coordinate sets and a minimal root mean square deviation (RMSD) between tertiary structures. CHIMERA's structure comparison function also gives the RMSD value validates the model. Energy minimization was done by Chimera's built in function AMBER.

## Ligand selection, active site prediction and molecular Docking

The CBP24 is known to have a carbohydrate binding domain which primarily hydrolyzes oligomeres of chitin. Acetylated hexamer of chitin, *N*-acetyl-chitohehexose was selected as ligand for docking. The 3D structure of the ligand molecule was obtained from a small molecule database ChemIDplus (<http://chem.sis.nlm.nih.gov/chemidplus/>). To predict the substrate binding pocket, 3D model of CBP24 was submitted to POCKET-FINDER server (<http://www.modelling.leeds.ac.uk/pocketfinder/help.html>). Auto Dock Tools (ADT) is a graphical front end for AutoDock (v4.2) and was used for flexible protein-ligand docking. Polar hydrogen was added to protein molecule and all atoms had assigned Gasteiger charges. Ligand bonds were set non-rotatable. A grid box of size (x= 98 Å, y= 82 Å and z= 106 Å) was generated having centroid (x=10.919 Å, y= 31.108 Å, z= 9.782 Å). A blind docking was then launched using Lamarckian genetic algorithm with population size 150, maximum numbers of generations as 27000 for 100 docked conformations.



**Figure 1:** Phylogenetic dendrogram based upon alignments with other domain family (CBM33) members. The phylogenetic dendrogram was built using the CBM33 family members having same domains as in CBP24. The bacterial sequences used are *Lactococcus lactis subsp. lactis*, (Q9CE94), *B. amyloliquefaciens* (Q9F9Q5), *B. thuringiensis serovar konkukian* (Q6HHU4), *L. monocytogenes* (Q8Y4H4), *Photorhabdus luminescens subsp. laumondii* (Q7N4I5), *S. marcescens* (Q83009), *Oceanobacillus iheyensis* (Q8ES33), *Yersinia enterocolitica* (Q8GBD4), *Shewanella oneidensis* (Q8EHY2), *Vibrio parahaemolyticus* (Q87FTO), *S. olivaceoviridis* (Q54501). The numbers at nodes indicate bootstrap confidence values. The position of the sequence originating from the *B. thuringiensis serovar konkukian* (Present study) is represented by Q6HHR5.

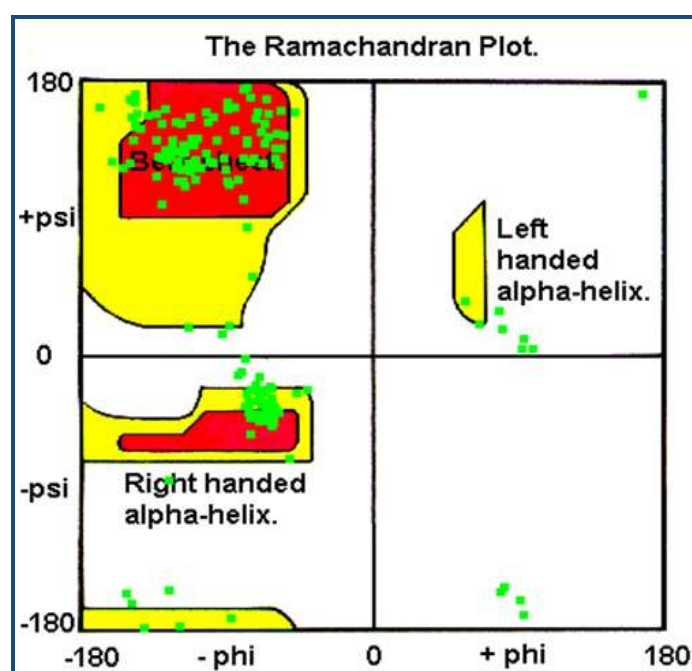


**Figure 2:** A) Template (2BEM) model; B) Homology model of CBP21; C) Structure-Structure alignment between target protein CBP24 (white) and template 2BEM (magenta)

## Results & Discussion:

The present study comprises sequence, structural and docking analysis of the chitin binding protein CBP24 from *B. thuringiensis serovar konkukian*. Previously, we have characterized the modular chitin binding protein CBP50 and two chitinases (Chi74, Chi39) [9, 28, 29]. It was shown that the C-terminal carbohydrate binding domain of the CBP50 has a critical role in substrate binding activity of the CBP50 [9]. ProtParam was used for primary sequences analyses of the target protein, the CBP24 contains 221 amino acid, with a molecular weight of 24082.3 Dalton and an isoelectric point (*pI*) of 9.11, while the total number of negatively and positively charged residues were 17 and 22, respectively. Secondary

structure analysis was performed using PSIPRED and the protein was shown to contain several  $\beta$ -strands and  $\alpha$ -helices, where the predicted percentage of  $\beta$ -strands was higher than the  $\alpha$ -helices. Domain analyses of the protein have shown a single CBM33 domain, whereas no functional motif was found in the query protein. Inconsistent to our findings, CBP21 has three chitin binding domains. The multiple sequence alignment of CBP24 has shown its association with other members of CBM33 as well, particularly with template CBP21, where both the proteins have shown conserved residues on pair wise alignment (Figure 1). The SignalP4.0 HMM showed the maximum cleavage site probability (0.289) of the CBP24 between amino acid residues at position 34 and 35. Subsequently, 35 residues were thus removed from the N-terminus of the protein sequence. The signal is recognizable by gram-positive bacteria only, which is in accordance with previous findings [30].



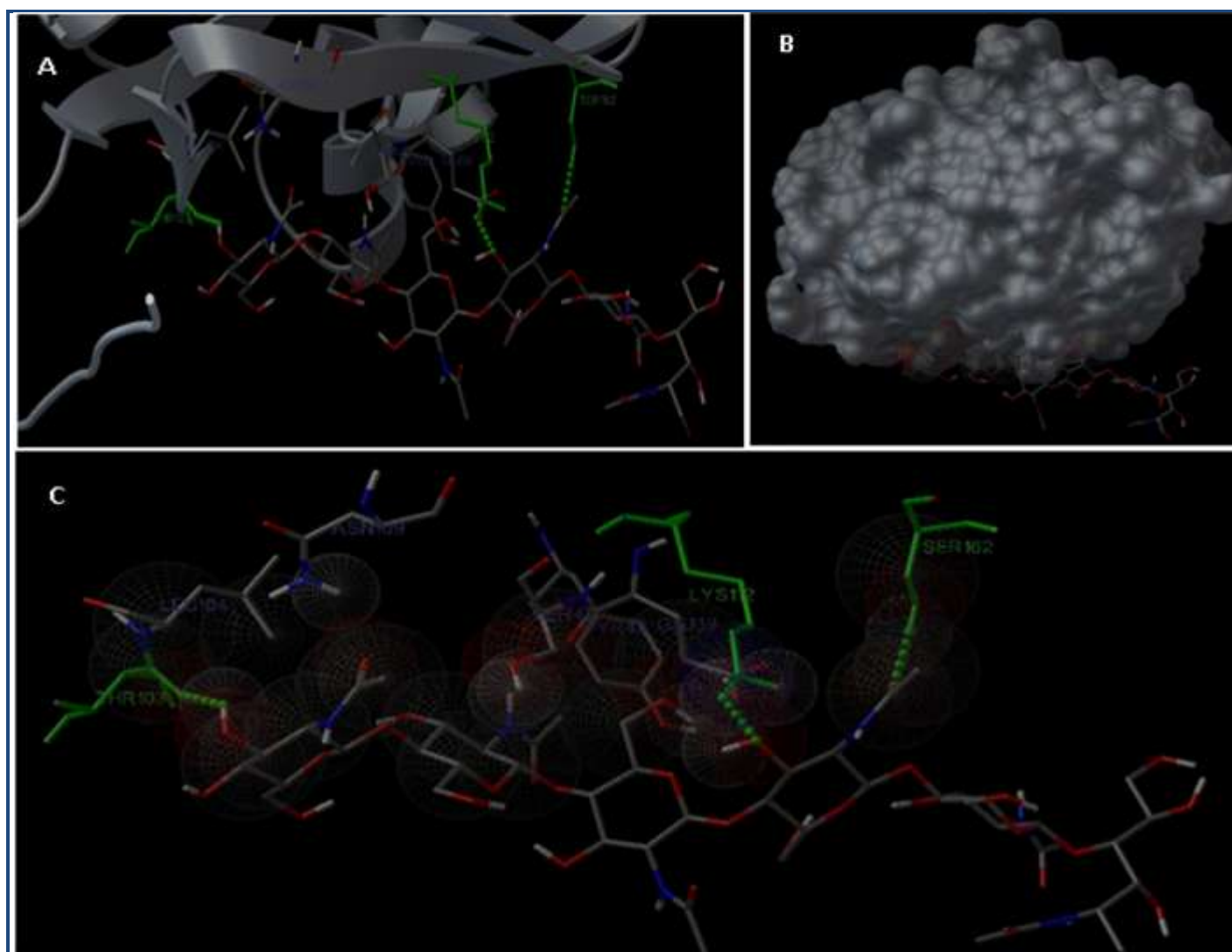
**Figure 3:** Ramachandran Plot of the best selected model showing ~ 95% residues in favored region.

## Homology modeling; template selection, model generation and validation

Advance search feature of the PDB PSI-BLAST produced thirteen homologues for CBP24. The template selection was made based on sequence identity greater than 30% and e-value. For the chitin binding domain of CBP24, the crystal structure of the *S. marcescens* chitin binding protein CBP21, (2BEM; PDB entry name) was selected as template with 62% homology. Although the pair wise alignment of the CBP24 and CBP21 has shown the presence of conserve residues yet the CBP24 appeared on a separate branch as compared to the other CBM33 domains, when a protein sequence based phylogenetic tree was constructed (Figure 1). Out of thirty generated models, the one with lowest objective function was selected for further refinement. Loop optimization side chain modeling produced considerable improvements in model (Figure 2). Several structure assessment methods including Ramachandran plots, verify 3D, RMSD and Errat quality factor were used to check the reliability of 3D model. Structure superimposition with

2BEM finally evaluated the model showing a low RMSD value of 0.161 Å ( $\text{Å}^{\circ} = 10^{-8} \text{ cm}$ ) which is acceptable based on sequence homology between CBP21 and selected template 2BEM. The verify 3D showed that amino acid residues of the template and model were 97.66% and 76.58% in favored regions, respectively. Similarly Errat quality factor was computed to assess the quality of the model, for 2BEM (template) its value was 98.765 and for 3D model it was 77.838. The Ramachandran plot was obtained for template and model and shows 99.01% and 95.89% residues in the most favored region respectively. The comparable RMSD values, verify 3D values, Errat quality factor and Ramachandran characteristics confirm the quality of the homology model of CBP24 (**Figure 3**).

Flexible protein-ligand docking was carried out to assess the binding mode of the CBP24. Blind docking for whole protein showed the similar pattern of substrate binding at exposed binding surface as that of CBP21. Out of 100 docked conformations, model 37 having cluster size as 16, showed the lowest binding free energy i.e -0.5 kcal/mol. Final intermolecular forces, van der Waal's force, H-bond energies, desolvation energies and electrostatic energies of this complex were computed as -8.34 kcal/mol, -8.20 kcal/mol, -4.55 kcal/mol respectively. This depicts a stable intermolecular conformation of the docked complex. These results are in accordance with the previous work on homology modeling and docking [31].



**Figure 4:** Substrate binding mode of CBP24. **(A)** Chitin oligomer bound at surface of protein is shown. Molecular surface representation is used for protein while ligand in sticks model is shown; **(B)** Residues involved in substrate interaction are shown with their exposed side chains, H bonded Thr-103, Lys-112 and Ser-162 are highlighted in green while green dotted lines represent H-bonds. Glu-39, Ser-41, Tyr-46, Leu-104 and Asn-109 are marked as residues involved in polar interactions; **(C)** Close atomic view of interaction shows all the interactions involved including polar contacts as well as H-bonds, wireframe circles are shown around the atoms involved in interaction.

Detailed molecular analysis of binding site prediction has shown that several residues (Glu-39, Tyr-46, Thr-103, Leu-104, Lys-112, Ser-162) are involved in substrate-protein interaction. Among them surface-exposed hydrophilic residues Thr-103, Lys-112 and Ser-162 were found to have strong H-bonding with

substrate. Whereas, amino acids residues Glu-39, Tyr-46, Leu-104 and Asn-109, shown to involve polar interactions with the substrate (**Figure 4**). Whereas, the surface-exposed polar residues Tyr-54, Glu-55, Glu-60, His-114, Asp-182, and Asn-185 were predicted to be involved in the substrate-binding activity

of the CBP21 in a homology modeling based study [1, 14]. It is interesting to note that Tyr-54, which is Tyr-46 at corresponding position in our CBP24, had previously been found critical in a mutational study of a homologous CBP from *Streptomyces olivaceoviridis* [5]. In this way, the CBP24 is presenting novel substrate binding sites. Moreover, it remains unconfirmed that the CBP24 may have role(s) other than chitin binding. This finding is in accordance with the smCBP28 [14]. In CBMs surface exposed aromatic residues are known to interact with substrates but the template CBP21 lacks an aromatic surface region, and the mode of CBP21 substrate binding is not clear [12]. The analysis had showed that CBP21 exerts this effect through specific polar interactions, which are not only important for binding, but also for alteration of the substrate structure. Using mutation analysis, Asn-185 has proved to be one of such residue [32].

Instead of being top of functional homolog of the CBP21, the CBP24 is presenting a different mechanism of interaction with the substrate. This may be confirmed through site-directed mutagenesis of the predicted amino-acids residues. So, the functional characterization of the CBP24 will elucidate the substrate-interaction pathway of the protein in specific and the carbohydrate binding proteins in general leading towards better understanding of substrate utilization pathways by prokaryotes.

## References:

- [1] Vaaje-Kolstad G *et al.* *J Biol Chem.* 2005 **280**: 28492 [PMID: 15929981]
- [2] Bormann C *et al.* *J Bacteriol.* 1999 **181**: 7421 [PMID: 10601197]
- [3] Schnellmann J *et al.* *Mol Microbiol.* 1994 **13**: 807 [PMID: 7815940]
- [4] Zeltins A & Schrempf H, *Anal Biochem.* 1995 **231**: 287 [PMID: 8594975]
- [5] Zeltins A & Schrempf H, *Eur J Biochem.* 1997 **246**: 557 [PMID: 9208950]
- [6] Kolbe S *et al.* *Microbiol.* 1998 **144**: 1291 [PMID: 9611804]
- [7] Chu HH *et al.* *Microbiol.* 2001 **147**: 1793 [PMID: 11429457]
- [8] Mehmood MA *et al.* *World J Microbiol Biotechnol.* 2009 **25**: 1955 [PMID: NA]
- [9] Mehmood MA *et al.* *Antonie Van Leeuwenhoek* 2011 **100**: 445 [PMID: 21647612]
- [10] Saito A *et al.* *Biosci biotechnol biochem.* 1999 **63**: 710 [PMID: 10361684]
- [11] Pleban S *et al.* *Lett Appl Microbiol.* 1997 **25**: 284 [PMID: 9351279]
- [12] Vaaje-Kolstad G *et al.* *J Biol Chem.* 2005 **280**: 11313 [PMID: 15590674]
- [13] Schrempf H, *Exs* 1999 **87**: 99 [PMID: 10906954]
- [14] Purushotham P *et al.* *PLoS One.* 2012 **7**: e36714 [PMID: 22590591]
- [15] Ma WM *et al.* *Biochem J.* 2013 **449**: 285 [PMID: 23013449]
- [16] Tran HT *et al.* *Histol Histopathol.* 2011 **26**: 1453 [PMID: 21938682]
- [17] Wu S *et al.* *Bioinfo.* 2008 **24**: 924 [PMID: 18296462]
- [18] Tramontano A *et al.* *Proteins.* 2001 **5**: 22 [PMID: 11835479]
- [19] Eswar N *et al.* *Curr Protoc Bioinformatics.* 2006 **5**: 5 [PMID: 18428767]
- [20] Pettersen EF *et al.* *J Comput Chem.* 2004 **25**: 1605 [PMID: 15264254]
- [21] Goodsell DS *et al.* *J Mol Recognit.* 1996 **9**: 1 [PMID: 8723313]
- [22] Bairoch A *et al.* *Nucleic Acids Res.* 2005 **33**: 154 [PMID: 15608167]
- [23] McGuffin LJ *et al.* *Bioinfo.* 2000 **16**: 404 [PMID: 10869041]
- [24] Bernstein FC *et al.* *J Mol Biol.* 1977 **112**: 535 [PMID: 875032]
- [25] Laskowski RA *et al.* *J Biomol NMR.* 1996 **8**: 477 [PMID: 9008363]
- [26] Colovos C & Yeates TO, *Protein Sci.* 1993 **2**: 1511 [PMID: 8401235]
- [27] Luthy R *et al.* *Nature.* 1992 **356**: 83 [PMID: 1538787]
- [28] Mehmood MA *et al.* *World J Microbiol Biotechnol.* 2010 **26**: 2171 [PMID: NA]
- [29] Mehmood MA *et al.* *Pak J Life Soc Sci.* 2012 **10**: 116 [PMID: NA]
- [30] Ruiz-Sanchez A *et al.* *Mol Biotechnol.* 2005 **31**: 103 [PMID: 16170210]
- [31] Batool M *et al.* *Bioinformation.* 2011 **7**: 384 [PMID: 22347779]
- [32] Vaaje-Kolstad G *et al.* *Science.* 2010 **330**: 219 [PMID: 20929773]

Edited by P Kanguane

Citation: Sehar *et al.* *Bioinformation* 9(14): 725-729 (2013)

**License statement:** This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited