BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 11(6)

Implication from the predicted docked interaction of sigma H and exploration of its interaction with RNA polymerase in *Mycobacterium tuberculosis*

Aayatti Mallick Gupta[1], Simanti Bhattacharya[2], Angshuman Bagchi[2] & Sukhendu Mandal[1]*

[1]Department of Microbiology, University of Calcutta, 35, Ballygunge Circular Road, Kolkata, 700019, India; [2]Department of Biochemistry and Biophysics, University of Kalyani, Kalyani, 741235, Nadia, India; Sukhendu Mandal – Email: sukhendu1@hotmail.com; smmicrobio@caluniv.ac.in; *Corresponding author

Received February 27, 2015; Accepted March 16, 2015; Published June 30, 2015

Abstract:
M. tuberculosis is adapted to remain active in the extreme environmental condition due to the presence of atypical sigma factors commonly called extra cytoplasmic function (ECF) sigma factors. Among the 13 sigma factors of *M. tuberculosis*, 10 are regarded as the ECF sigma factor that exerts their attributes in various stress response. Therefore it is of interest to describe the structural prediction of one of the ECF sigma factors, sigma H (SigH), involved in oxidative and heat stress having interaction with the β′ subunit of *M. tuberculosis*. RNA polymerase (Mtb-RNAP). The model of Mtb-SigH was build using the commercial package of Discovery Studio version 2.5 from Accelerys (San Diego, CA, USA) containing the inbuilt MODELER module and that of β′ subunit of Mtb-RNAP using Phyre Server. Further, the protein models were docked using the fully automated web tool ClusPro (cluspro.bu.edu/login.php). Mtb-SigH is a triple helical structure having a putative DNA-binding site and the β′ subunit of Mtb-RNAP consists of 18-beta sheets and 22 helices. The SigH-Mtb-RNAP β′ interaction studies showed that Arg26, Gln19 andAsp18, residues of SigH protein are involved in binding with Arg137, Gln140, Arg152, Asn133 and Asp144 of β′ subunit of Mtb-RNAP. The predicted model helps to explore the molecular mechanism in the control of gene regulation with a novel unique target for potential new generation inhibitor.

Keywords: transcription, *Mycobacterium* sigma factor, transcription regulator, homology modelling, SigH-RNAP interaction.

Background:
Bacterial gene expression is primarily regulated at the level of transcription. Sigma factor plays a pivotal role in binding with the core RNAP and subsequently in promoter recognition **[1]**. In the last few years, after the publication of the *M. tuberculosis* whole genome sequence, the 13 sigma factors of *M. tuberculosis* have become an important subject of investigation of which 10 are regarded as the ECF sigma factors. ECF sigma factors play important role in bacterial pathogenecity and stress, among them SigH is involved in oxidative stress and heat stress **[2, 3]**. Understanding the structural-functional relationship of these atypical sigma factors will help to understand the physiology and virulence mechanisms of *M. tuberculosis* and will help to design new strategies to fight against the deadly pathogen. These sigma factors are called as atypical due to the fact that they consist of two structural domain instead of four that is usually found in the housekeeping sigma factor SigA or Sig70 (in case of *E. coli*) **[4, 5]**. Domain 2 and domain 4 recognises the -10 and -35 promoter-binding element, respectively **[6, 7]**. Biochemical studies on bacterial RNA polymerase typically describe that domain 2 region of sigma factor binds with the β′ subunit of RNAP **[1]**. The sigma binding locus on the β′ subunit is quite conserved throughout the bacterial kingdom. The present study delineates the structural understanding of Mtb-SigH and β′ subunit of Mtb-RNAP and thereby implicates the predicted interaction between them as one of the vital target for gene expression.

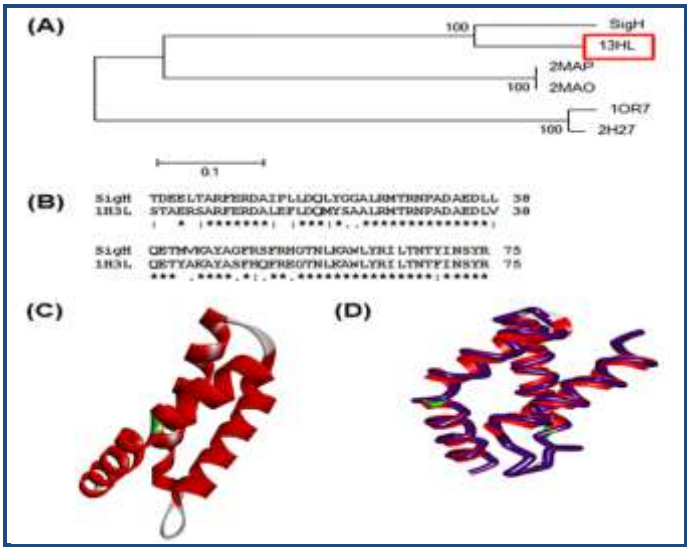

Figure 1: A) Phylogenetic relationship between Mtb-SigH with its template SigR from *Streptomyces coelicolor* (**PDB ID: 1H3L**) which was built using MEGA 5.2 **[21]**. 2MAP signifies solution structure of the complex formed by the region 2 of *E. coli* sigma E and its cognate -10 promoter element non template strand TGTCAAA, 2MAO denotes the NMR structure of region 2 of *E. coli* sigma E, 1OR7 is the crystal structure of *E. coli* sigma E with the cytoplasmic domain of its anti-sigma RseA, 2H27 is the crystal structure of *E. coli* sigma E region 4 bound to its -35 element DNA. **B)** Pair-wise alignment of Mtb-SigH with the template showing 58.9% sequence identity and 67.4% sequence similarity. **C)** Model of domain 2; built using *S. coelicolor* SigR crystal structure as template. The model predominantly shows the triple helical structure that distinctly have DNA binding feature. **D)** Superimposition of SigH (Red) on template (Indigo). RMSD value being 0.264Å.

Methodology:

Homology search

The sequence of Mtb-SigH (Accession Number CCP46040.1) and β′ subunit of Mtb-RNAP (Accession Number CCP43411.1) were obtained from NCBI database (http://www.ncbi.nlm.nih.gov). The protein sequences retrieved were used in BLAST (National Library of Medicine, USA) search against PDB databases **[8]** in order to obtain a suitable template for homology modeling. Mtb-SigH and β′ subunit of Mtb-RNAP were modeled using 75 and 974 amino acid residues, respectively.

Model building

The homology models of Mtb-SigH and β′ subunit of Mtb-RNAP were developed using the N-terminal fragment of SigR from *Streptomyces coelicolor* **(PDB ID 1H3L)** and the crystal structure of *Thermus thermophilus* (Tth) transcription initiation complex containing 2 nucleotide of nascent RNA (**PDB ID: 4G7O**) as template respectively. The sequence similarities of the templates with the targets were verified through multiple sequence alignment. The model of Mtb-SigH was build with MODELER module using Discovery Studio version 2.5 from Accelerys (San Diego, CA, USA) and that of β′ subunit of Mtb-RNAP using Phyre Server **[9]**. Mtb-SigH showed 58.9% sequence identity and 67.4% sequence similarity with its template and β′ subunit of Mtb-RNAP showed 51% homology with the template.

Model evaluation and refinement

Structural evaluation of the built model was done using SAVES program (http://nihserver.mbi.ucla.edu/SAVES) (which at a time runs PROCHECK, Verify3D, What_Check, ERRAT). All these together are important to clarify the overall fold or structure, errors over localized regions and stereochemical parameters such as bond lengths and angles.

Analysis of the binding site of Mtb-SigH

Active site prediction computes the cavities of a given protein. The active site of Mtb-SigH was predicted using the server from supercomputing facility for bioinformatics and computational biology, IIT, Delhi (http://www.scfbio-iitd.res.in/) **[10]**. Top cavity is analysed to find out the amino acid residues involved in binding with the core Mtb-RNAP.

Docking Studies

Protein-protein interaction study between Mtb-SigH and β′ subunit of Mtb-RNAP were carried out using the fully automated web based program ClusPro (cluspro.bu.edu/login.php) **[11, 14]**. It is a web-server that uses the algorithm of Global rigid search: FFT using DOT or ZDOCK program. The docked structures and interface residues were analyzed using Discovery Studio version 4 from Accelerys (San Diego, CA, USA). Accessible surface area of both Mtb-SigH and β′ subunit of Mtb-RNAP were calculated using Surface racer version 5 **[15].**

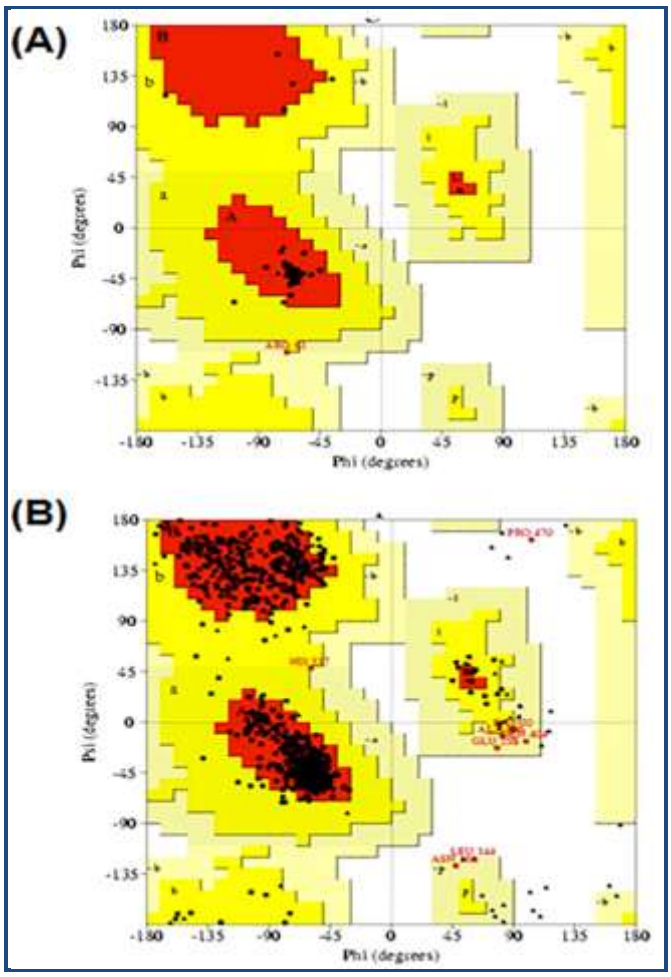

Figure 2: A) Ramachandran plot of Mtb-SigH protein model. The phi-psi torsion angle of the protein is shown in the

Ramachandran plot which depicts that 92.9% of the amino acid residues are in the core region (red), 6% residues in allowed (yellow) and 1.5% in generously allowed region. **B)** Ramachndran plot analysis for β′ subunit of Mtb-RNAP. Containing total 1019 residues among which 92.9% residues in the core region (red); 6.3 % residues in allowed (yellow) and 0.8% in generously allowed (light yellow) region. None of the residues are present in the disallowed region.

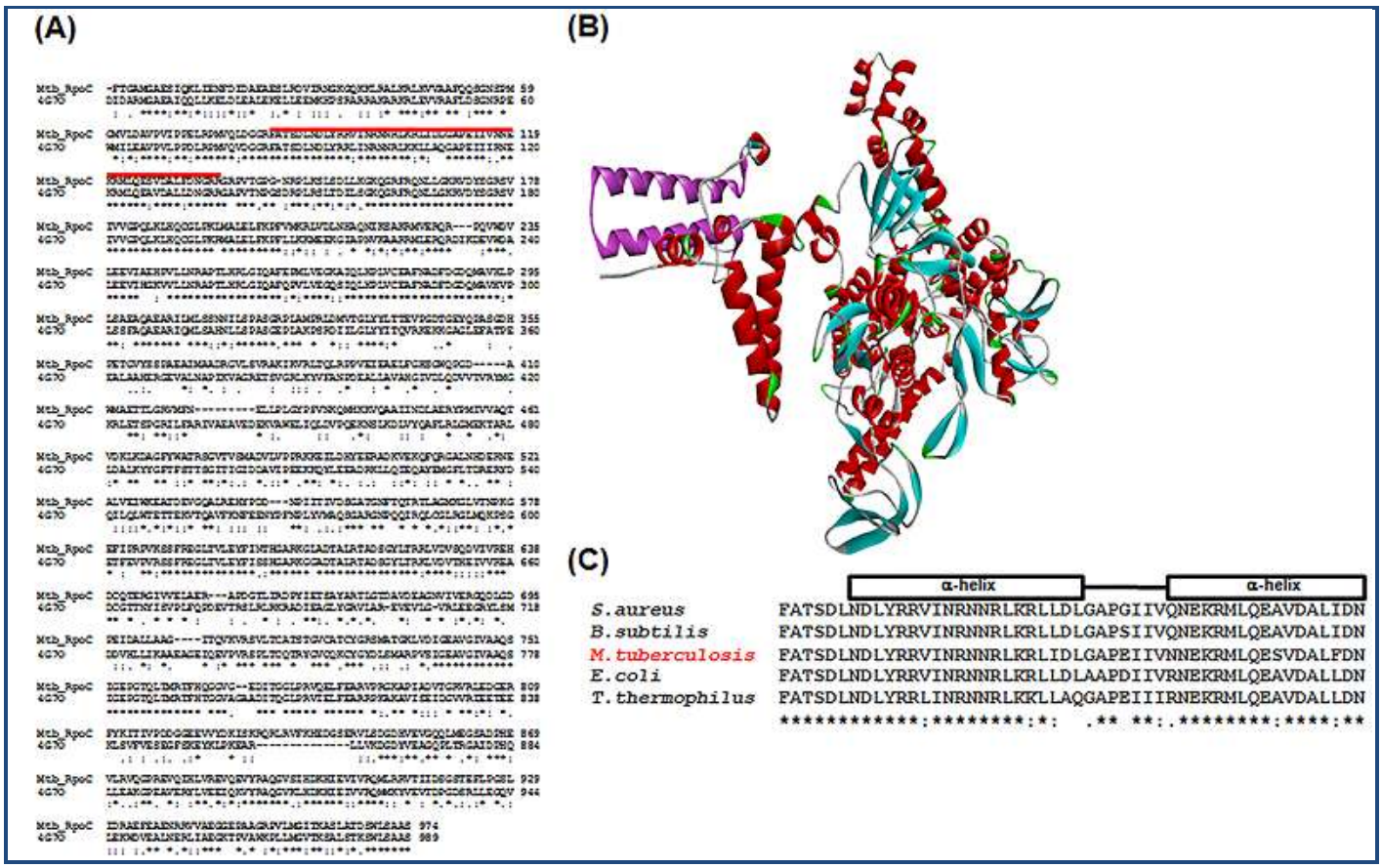

Figure3: A) Pairwise sequence alignment of Mtb-β′ subunit with the template, crystal structure of Tth-RNAP (4G7O). The target sequence showed 51% homology with the template. The Mtb-SigH binding region is highlighted with red bar. **B)** 3D Structure of β′ subunit of Mtb-RNAP. Mtb-SigH binding region is marked with purple colour. The modeled structure consists of 18 beta sheet and 22 helices. **C)** Sequence alignment of sigma domain 2-binding region in various β′ homolog. The binding region of the sigma factor with the core RNAP is quite conserved throughout the bacterial RNAP.

Results & Discussion:

The present study reports the 3D structure of two globular proteins; Mtb-SigH and β′ subunit of Mtb-RNAP. Mtb-SigH, an ECF sigma factor is 216 amino acids long, expressed under specific condition of oxidative stress, hypoxia, heat shock etc. **[16-18]**. The conserved domain search from NCBI tools predicted that the protein consists of two domains; domain 2 and domain 4 recognizing -10 and -35 promoter binding element, respectively. The present model of Mtb-SigH is built using the chain A of N-terminal fragment of SigR from *Streptomyces coelicolor* (**PDB ID: 1H3L**) as template based on the BLAST search **(Figure 1A**) against PDB using Discovery Studio version 2.5. Pairwise sequence alignment between the template and the target molecule showed 58.9% sequence identity and 67.4% sequence similarity **(Figure 1B)**. The model is constructed using 75 amino acid residues consisting domain 2. The best-selected model was then energy minimized keeping their backbone fixed to ensure proper interactions. 1130 cycles of conjugate gradient energy minimization steps were applied with CHARMm force field until the structure reached a final derivative of 0.01 Kcal/mole. The refined model after energy minimization was subjected for PROCHECK analysis in order to analyze the stereochemical property of the model. Ramachandran plot retrieved from PROCHECK analysis showed 92.5% of the residues in the most favored region and none in the disallowed region **(Figure 2A)** that confers Mtb-SigH to be a good model. The overall structural quality of the model is validated through verify 3D graph. The predicted model is the triple helical structure **(Figure 1C)** having a probable DNA-binding role. The built model was superimposed with the template **(Figure 1D)** and the root mean square deviation (RMSD) value between template and predicted model is 0.264Å, indicative of a good model.

β′ subunit of Mtb-RNAP (*rpoC*) is 1316 amino acid residues long and the present model is constructed with 974 amino acid residues **(Figure 3A)** using Phyre server. The model is built using the crystal structure of *Thermus thermophilus* (Tth) transcription initiation complex containing 2 nucleotide of nascent RNA (**PDB ID: 4G7O**) as template. The target sequence showed 51% sequence identity with the template. The phi-psi torsion angle of the protein is shown in the

Ramachandran plot that depicts that 92.8% of the amino acid residues are in the core region and none of these are in the disallowed region **(Figure 2B)** which predicts the overall stereochemical quality of the protein is significantly high. The modeled structure consists of 18 beta sheet and 22 alpha helices **(Figure 3B)**.

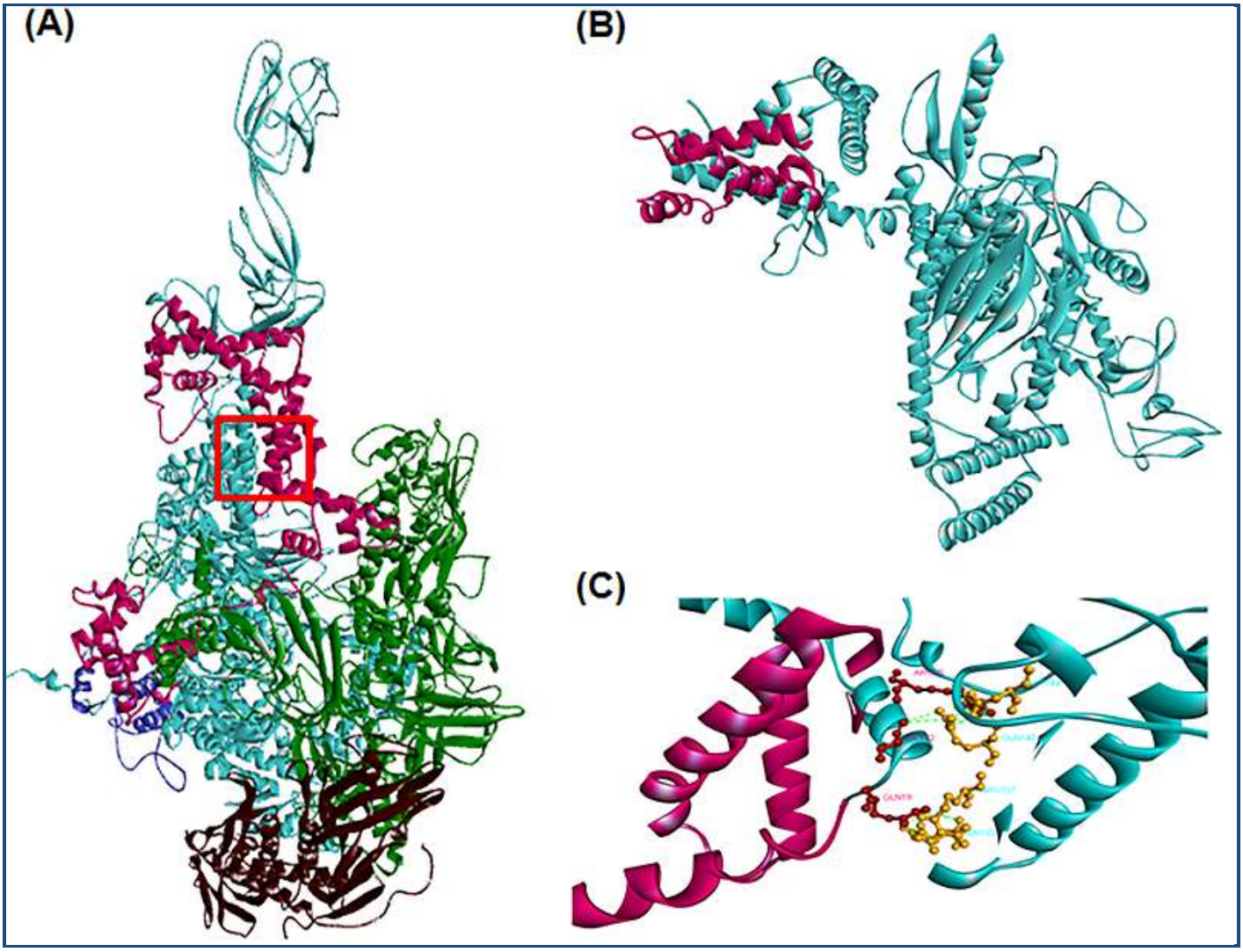

Figure 4: A) Generalised view of bacterial RNAP (PDB ID:3DXJ). σ70-β′ binding interface is marked in red box. **B)** Magnified view of Mtb-SigH and β′ subunit of Mtb-RNAP binding interface. The binding site of Mtb-SigH is denoted in cyan. **C)** Close up view of Mtb-SigH and β′ subunit of RNAP complex. Mtb-SigH is marked in magenta and Mtb-β′ in cyan. The green dotted line shows the H-bonding between Arg26, Gln19 and Gly22 of SigH and Arg137, Gln140, Arg152, Asn133 and Asp144 of β′ subunit of RNAP.

The top cavity is used to analyze the result using the parameters like shape complementarities, solvation energy, and electrostatistics. Amino acid residues like Asp18, Gln19 and Arg26 found in the active site are also found in the binding interface of Mtb-SigH and β′ subunit of Mtb-RNAP. Moreover the sequence alignment of sigma domain 2-binding region in various β′ homolog is found to be quite conserved throughout the bacterial RNAP **(Figure 3C)**. The study reflects how the atypical sigma factor interacts with the core RNAP in order to initiate the promoter recognition. In consideration of the various experimental data on bacterial RNAP it is well known that generally the domain 2 of the sigma factor binds with the α helical coil-coiled region of β′ subunit of Mtb-RNAP **(Figure 4A)**. From the present study it has also been predicted that Mtb-SigH and β′ subunit of Mtb-RNAP interaction takes place through α helical coil-coiled region **(Figure 4B)**. The report confers that Arg26, Gln19, and Asp18 of Mtb-SigH and Arg137, Gln140, Arg152, Asn133 and Asp144 of β′ subunit of Mtb-RNAP are probably found at the binding interface **(Figure 4C)**. The result elicited that the predicted amino acid residues present in the binding sites are also present in the binding cavities.

The binding energy of the docked complex is calculated using the forcefield CHARMm which shows the potential energy of the complex to be -33177.633 Kcal/mol. Most of the interface residues form conventional intermolecular H-bonding within 3.4 Å, while a salt bridge is found between Asp18 of Mtb-SigH and β′ subunit of Mtb-RNAP. The interface residue shows an electrostatic mode of interaction in between Arg26 of Mtb-SigH (electropositive) and Asp144 of β′ subunit of Mtb-RNAP (electronegative). No potential van der Waal clashes are noted within 0.7 Å of the binding interface **Table 1 (see supplementary material)**. The unfavorable interactions

(including bumps) are remarkably not found in the interacting site of the two proteins. Accessible surface area (ASA) of both Mtb-SigH and β′ subunit of Mtb-RNAP is calculated using the van der Waal radii from Richards 1977 taking probe radius 1.4 Å. The output of which typically showed that the interacting residues of both the subunit have much higher ASA value than those of the non-interacting ones **Table 2 (see supplementary matrerial)**. Thus the result clearly predicts the fact that Arg26, Gln19, and Asp18 of Mtb-SigH and Arg137, Gln140, Arg152, Asn133 and Asp144 of β′ subunit of Mtb-RNAP have a significant ASA making them potent as the interface residues. The binding region of the sigma factor with the core RNAP is quite conserved throughout the bacterial RNAP. Bacteriological studies with Tth-RNAP and Ec-RNAP showed similar amino acid residues in the binding interface **[19-20].**

The fine structural model of the undertaken molecule (Mtb-RNAP and Mtb-SigH) could be used to understand the function of other transcription associated protein in much detail. This work will further assist and deliver clue to perform precise experimental work with such macromolecules. The developed model will help to explore the interacting inhibitors by means of chemi-informatics. Thus *in-silico* understanding could attribute to develop inhibitor molecule for sigma factors and Mtb-RNAP. The structural understanding could further delineate the regulation mechanism by the ECF sigma factors. The region of interaction is quiet conserved in the prokaryotic RNAP and hence targeting such stress responsive atypical sigma factor is worth to explore the possible control of gene expression. There are very limited reports on the gene loci that are transcribed by Mtb-SigH which in turn could be an excellent material to explore the putative promoter element which can be recognized by Mtb-SigH. Moreover, such studies will enlighten the future research of drug discovery where such sigma factors could be treated as the target molecules.

Acknowledgement:
Authors highly acknowledge the generous help of the bioinformatics infrastructural facility of Kalyani University. We are thankful for the assistance of Dr. Prithidipa Sahoo of Visva Bharati University for proofreading the manuscript. Authors are thankful to anonymous reviewer for his/her comments for the betterment of the manuscript.

References:

[1] Murakami KS & Darst SA, *Curr Opin Struct Biol*. 2003 **13:** 31 [PMID: 12581657]
[2] Cole ST *et al*. *Nature* 1998 **393:** 537 [PMID: 9634230]
[3] Tekaia F *et al*. *Tuber Lung Dis*. 1999 **79:** 329 [PMID: 10694977]
[4] Sachdeva P *et al*. *FEBS J*. 2010 **277:** 605 [PMID: 19951358]
[5] Rodrigue S *et al*. *FEMS Microbiol Rev*. 2006 **30:** 926 [PMID: 17064287]
[6] Gross CA *et al*. *Cold Spring Harb Symp Quant Biol*. 1998 **63:** 141 [PMID: 10384278]
[7] Campbell EA *et al*. *Mol cell*. 2002 **9:** 527 [PMID: 11931761]
[8] Berman H *et al*. *Nucleic Acids Res*. 2007 **35**: D301 [PMID: 17142228]
[9] Kelley LA *et al*. *Nat Protoc*. 2009 **4:** 363 [PMID: 19247286]
[10] Singh T *et al*. *J Chem Inf Model*. 2011 **51:** 2515 [PMID: 21877713]
[11] Kozakov D *et al*. *Proteins*. 2013 **81**: 2159 [PMID: 23996272]
[12] Kozakov D *et al*. *Proteins*. 2006 **65**: 392 [PMID: 16933295]
[13] Comeau SR *et al*. *Bioinformatics* 2004 **20:** 45 [PMID: 14693807]
[14] Comeau SR *et al*. *Nucleic Acids Res*. 2004 **32:** W96 [PMID: 15215358]
[15] Tsodikov OV *et al*. *J Comput Chem*. 2002 **23:** 600 [PMID: 11939594]
[16] Raman S *et al*. *J Bacteriol*. 2001 **183:** 6119 [PMID: 11567012]
[17] Park ST *et al*. *Proc Natl Acad Sci U S A*. 2008 **105:** 13105 [PMID: 18728196]
[18] Dutta NK *et al*. *PLoS One* 2012 **7:** e28958 [PMID: 22235255]
[19] Vassylyev DG *et al*. *Nature* 2002 **417:** 712[PMID: 12000971]
[20] Burgess RR & Anthony L, *Current opinion in Microbiol*. 2001 **4:** 126 [PMID: 11282466]
[21] Tamura K *et al*. *Mol Biol Evol*. 2011 **28:** 2731 [PMID: 21546353]

Edited by P Kangueane
Citation: **Gupta *et al.*** Bioinformation 11(6): 296-301 (2015)

BIOINFORMATION

open access

Supplementary material:

Table 1 Intermolecular interaction of the binding interface in which the amino acid residues of Mtb-SigH are marked in bold

Interacting molecules	Bond Type	Bond Distance
Gln19:HE21-:Asn133:OD1	H-bond	2.04369
Arg26:HH22-:Gln140:O	H-bond	1.9025
Arg26:HE-:Asp144:OD1	H-bond	2.2018
Arg26:NH1-:Asp144:OD2	Electrostatic	5.31545
Arg26:NH2-:Asp144:OD1	Electrostatic	2.72004
Arg137:HE-:**Gln19:OE1**	H-bond	1.99359
Arg137:HH21-:**Gln19:OE1**	H-bond	1.83375
Arg152:HH12-:**Asp18:OD1**	Salt bridge	1.77648
Arg152:HH22-:**Asp18:OD2**	Salt bridge	1.78815

Table 2: Accesible surface area (ASA) of the protein subunits in which the interacting residues are demarcated in bold

Protein subunit	Residue name	ASA
Mtb-SigH	**Asp18**	116.20
	Gln19	99.02
	Arg26	209.23
	Thr28	1.50
	Gly22	27.73
	Leu37	5.12
β' subunit of Mtb-RNAP	**Arg137**	134.51
	Gln140	73.29
	Arg152	127.58
	Asn133	44.35
	Asp144	25.84
	Glu135	4.01
	Leu146	3.71
	Ala145	4.42