

Characterization of the Leucine-Responsive Transcription Factor from Pathogenic *Vibrio cholerae* Using Molecular Modelling and Molecular Dynamics Simulations

Angshuman Bagchi*

Department of Biochemistry and Biophysics, University of Kalyani, Kalyani, Nadia, India

*Corresponding author: Angshuman Bagchi, Department of Biochemistry and Biophysics, University of Kalyani, Kalyani, Nadia, India, Tel:-+91-9051948843; E-mail: angshumanb@gmail.com

Received date: April 23, 2014; Accepted date: June 16, 2014; Published date: June 20, 2014

Copyright: © 2014 Bagchi A. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Vibrio cholerae is the causative agent of cholera. It is responsible behind a variety of diseases specifically the multidrug-resistant nosocomial infections and chronic lung infections in cystic fibrosis patients. One of the vital genes of the organism responsible for its multidrug resistant behaviour is the gene PA3523 which codes for the multidrug efflux transporter. The expression of PA3523 is regulated by the dimeric transcription factor Lrp having Helix-Turn-Helix DNA binding motif. So far there have been no previous reports that depict the characterization of Lrp protein from *Vibrio cholerae* from a structural point of view. In the present work, an attempt has been made to characterize Lrp protein by structural bioinformatics approach. The dimeric structure of Lrp was built by comparative modelling technique. The dimeric model of Lrp was then docked on to the corresponding promoter region of the PA3523 gene encoding the multidrug efflux transporter. The docked complex of promoter DNA with Lrp protein was subjected to molecular dynamics simulations to identify the mode of DNA protein interactions. So far, this is the first report that depicts the mechanistic details of gene regulation by Lrp protein. This work may therefore be useful to illuminate the still obscure molecular mechanism behind disease propagation by *Vibrio cholerae*.

Keywords: Transcriptional Regulation; Molecular Modelling; DNA-Protein Docking; Molecular Dynamics Simulations; Protein-DNA interactions

Introduction

Vibrio cholerae is a gram negative bacterium found mostly in soil, water and inside a number of host organisms. This organism is an opportunistic pathogen particularly known as the causative agent of multidrug-resistant nosocomial infections. It is also the main cause behind chronic lung infections in patients diagnosed with cystic fibrosis. The effect of *Vibrio cholerae* in cystic fibrosis patients is so severe that its establishment in the patients' lungs is likely to be a permanent fixture of the patient's life [1-10]. Recently, a *Vibrio cholerae* strain with increased capability of antibiotic resistance was identified [7]. The most important characteristic of *Vibrio cholerae* is the resistance against multiple drugs. One of the vital genes of the organism responsible for the multidrug resistant behavior is the gene PA3523 which codes for the multidrug efflux transporter [11]. The efflux transporters are proteins bound to the bacterial cell membranes. These proteins function as pumps to efflux out the toxic substances from bacterial cells and thereby confer single- and multiple-drug resistances. There are basically three different transport systems; the active transport, involving ATP hydrolysis for the entry or efflux of substances from cells; then there is phosphoenolpyruvate-dependent phosphotransferase system (PTS) in which a solute is phosphorylated to be transported across the membrane; the third type utilizes ion gradients as the energy-mode for transportation process. An interesting property of these efflux transporters is that they intrinsically harbour multiple substrates thereby providing the advantage of resistance to multiple structurally-different drugs. These

efflux transporters reduce the accumulation of antibiotics inside bacterial cells. However, the process of the antibiotic efflux is very slow. Thus it provides sufficient time for the bacterium to adapt to the antibiotics and become resistant through mutations or alteration of antibiotic targets. Such multi-drug efflux transporters in bacterial pathogens are considered to be good targets for inhibitors, as reducing the efflux of multiple drug molecules from the bacterial cells may restore the clinical efficacy of older drugs and prevent emergence of drug-resistant variants. [12-14]. The expression of PA3523 is controlled by the transcription factor Lrp. Lrp is a transcriptional regulator having Helix-Turn-Helix DNA binding motif and it acts as a dimer in presence of metal ions like Cu [15]. However, till date no reports regarding the molecular mechanism of regulation of PA3523 gene by Lrp protein is available. In the present scenario, an attempt has been made to explore the molecular mechanism of the involvements of Lrp in the regulation of PA3523 gene expression. The computational model of Lrp protein has been built as no structures of Lrp protein from *Vibrio cholerae* was available. The Dimeric model of the Lrp protein has been built and then docked onto the known promoter DNA region of the PA3523 gene. The interactions between the DNA and the Lrp protein were analyzed after molecular dynamics simulation of the DNA-protein complex. This computational analysis of the interactions between the promoter DNA and Lrp protein is the first report that analyzes the involvements of the amino acid residues of Lrp protein in the regulation of bacterial gene expression. This report would therefore pave the pathway for future genetic studies to identify the roles of the individual amino acids of Lrp protein in bacterial gene regulation. The interaction analysis would also be useful to develop new drugs aiming to attack the important amino acid residues of Lrp protein involved in binding to promoter DNA regions during the PA3223 gene expression.

Materials and Methods

Sequence analysis and homology modeling of Lrp

The amino acid sequence of Lrp protein from *Vibrio cholerae* was obtained from GenBank (Refseq id: YP_001217437.1) and was used to search the Brookhaven Protein Data Bank (PDB) [15] using the software BLAST [16] to find suitable template(s) for homology modeling. For the BLAST search the default parameters were used (BLOSUM62 matrix, Word size: 3, Expect threshold: 100, Gap costs: Existence: 11, Extension: 1). The best match from the BLAST search result was found to be the X-ray crystal structure of the leucine responsive regulatory protein from *E. coli* (pdb code: 2GQQ, Chain A) with 89% sequence identity, 99% query coverage and an e-value of 2×10^{-106} . The other results obtained from the BLAST search were 2L4A, Chain A (NMR structure of the DNA-Binding Domain of *E. coli* Lrp with 40% query coverage, 94% identity and e-value of 8×10^{-38}), 2P5V, Chain A (Crystal structure of transcriptional regulator Nmb0573 from *Neisseria Meningitidis*, with 92% query coverage, 41% identity and e-value of 2×10^{-36}). A homology model of the Lrp protein was built using the software Modeller [18] with the A chain of 2GQQ as template. 5 models were built with high optimization levels. The loops generated in the models were refined by loop refinement technique as employed in the MODELLER software tool. This was done in order to avoid generation of erroneous loop structures [18]. The modeled structure of the Lrp protein was then superimposed on to the crystal template without altering the coordinate system of atomic positions in the template. The root mean squared deviation (RMSD) for the superimposition was 0.5 Å. The model of the Lrp protein was then energy minimized in two steps. In the first step the modelled structure was minimized without restraints. In the second step, the energy minimization was done by fixing the backbone of the modeled protein to ensure proper interactions. All energy minimizations were done using conjugate gradient (CG) with CHARMM force fields [19] until the structures reached the final derivative of 0.001 kcal/mole. For the energy minimization, 200 cycles each of steepest descent and conjugate gradient methods were used. The non-bonded list radius was kept at 14 Å (default in Discovery Studio package). The electrostatic interactions were evaluated with a Spherical Cutoff.

Validation of the model of Lrp protein

Regarding the main chain properties of the modeled protein, no considerable bad contacts nor Ca tetrahedron distortions nor hydrogen bond (H-bond) energy problems were found. There were no side chain distortions as observed by measuring the side chain torsion angles. The Z-scores calculated using PROSA [https://prosa.services.came.sbg.ac.at/prosa.php] web server showed that the predicted homology model was well inside the range of typical native structures. The residue profiles of the three dimensional model were further checked by VERIFY3D [20]. PROCHECK [21] analyses were performed in order to assess the stereo-chemical qualities of the models and Ramachandran plots [22] were drawn. No residues were found to be present in the disallowed regions of the Ramachandran plots.

Building the model of the promoter DNA: In order to find the interactions between promoter DNA and the Lrp protein, the nucleotide sequence of the promoter DNA region for the gene VC0395_A2223 from *Vibrio cholerae* was extracted. For this purpose

the database CollecTF [15] was used. The database CollecTF showed the following nucleotide sequence to be the promoter DNA sequence:

3'-CAGTATAAAAAACTA-5'

The nucleotide sequence was used to build the model of the corresponding DNA region using the CHARMM software tool and then subjected to energy minimizations. The resulting energy minimized structure was used for docking studies.

Molecular docking simulation

It is known that Lrp binds to DNA as a homo-dimer [15]. Thus a homo-dimeric model of Lrp was built by superimposing the models of Lrp on to the two chains its crystal template 2GQQ. The homo-dimeric model of Lrp so obtained was subjected to energy minimization as per the protocol previously mentioned in the section Sequence analysis and homology modeling of Lrp. In order to elucidate the mode of binding between the promoter DNA and Lrp, the models of the protein and the DNA were docked using the software patchdock [23]. Patchdock uses surface information about proteins and docks the proteins on the basis of their surface complementarity. The final docked complexes are ranked on the basis of their geometric shape complementarity scores [23]. The highest scoring docked structure of the DNA-protein complex that yielded the best score was selected and analyzed visually using DS modeling software suite. The docked complex was then energy minimized as per the protocol previously mentioned in the section Sequence analysis and homology modeling of Lrp.

Molecular dynamics (MD) simulation

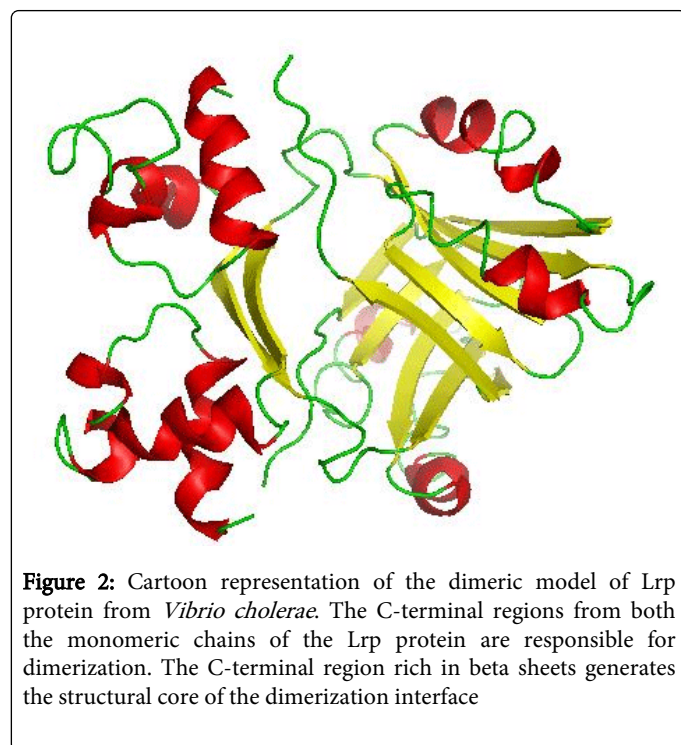
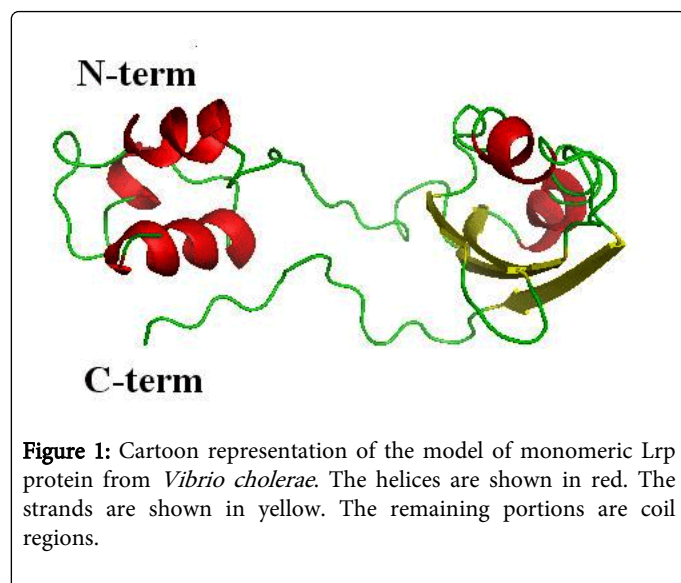
The MD simulation of the DNA-protein complex was performed using the CHARMM module of DS modeling software suite. The initial coordinates were extracted from the energy minimized structure of the promoter DNA-Lrp docked complex. The complex was placed in an orthorhombic box having dimensions preventing self-interactions and periodic boundary conditions were employed. The system was solvated with adequate water molecules at the typical density of water at 298K and 1.0 atm pressure utilizing single point charge (SPC) model. The whole system was energy minimized keeping the temperature constant to the body temperature of 310K using NPT dynamics protocol. The LINCS [24] method was applied to constrain the covalent bond lengths to keep the integration steps to 2fs. The particle mesh Ewald [25] was applied to calculate the electrostatic interactions. A 100 ns dynamics run was then performed for the DNA-protein complex. The modes of binding interactions between Lrp and the corresponding promoter DNA was then analyzed using DS modeling software suite.

Results and Discussions

Structure of Lrp from *Vibrio cholerae*

The structure of Lrp was built using 2GQQ, Chain A as the template. The amino sequences of Lrp and 2GQQ, Chain A showed 89% sequence identity. The modelled structure of Lrp monomer was presented in figure 1. The protein has two domains; the N-terminal DNA binding domain and the C-terminal domain consisting of anti-parallel beta sheets flanked by alpha helices. The two domains are connected via a linker loop region. The DNA binding region of the protein was present at the N-terminus and comprised of long helix-

turn-helix motif. The DNA binding domain consists of three alpha helices joined by loop regions. The third helix is the DNA recognition helix and binds to the major groove of DNA. The dimerization of the Lrp proteins is mediated by the C-terminal region. The beta strands present in the C-terminal region serves as the structural core of the dimerization interface between the Lrp monomers (figure 2). The Lrp protein from *Vibrio cholerae* is known to have metal binding sites [12]. Sequence alignment of Lrp protein from *Vibrio cholerae* with 2GQQ, Chain A, revealed the presence of the metal ion binding region in the Lrp protein from *Vibrio cholerae*. The sequence alignment was presented in Figure 3. The metal ions were found to be bound to the Lrp protein from *Vibrio cholerae* by getting coordinated to the Ser30 and Glu32. The metal ion binding region is present in the HTH DNA binding motif of the Lrp protein from *Vibrio cholerae*.



Interaction with promoter DNA

It is known that the transcription factor Lrp interacts with DNA as a homo-dimer. In order to analyze the interactions between Lrp and its promoter DNA region, a homo-dimeric model of Lrp was generated. The resulting model was docked onto the model of the promoter DNA region to find the DNA-protein interactions. The docked complex was subjected to MD simulation after energy minimization. The progress and completion of the MD simulation process were monitored by plotting a graph of RMSD of the backbone atoms of the docked complex vs time periods of the MD simulation runs (Figure 4). Interestingly, the protein conformation does not undergo an appreciable change upon DNA binding.

```
Lrp 2 VDSYKPKPSKDLDRIDRNILNELQKDGRISNVELSKRVGLSPTPCLERVRRRLERQGFITGY 61
2GQQ 1 VDSKKRPGKDLDRIDRNILNELQKDGRISNVELSKRVGLSPTPCLERVRRRLERQGFIQGY 60

Lrp 62 TALLNPQYLDASLLVFVEITLNRGAPDVFEQFNAAVQKLLDDIQECHLVSGDFDYLLKTRV 121
2GQQ 61 TALLNPHYLDASLLVFVEITLNRGAPDVFEQFNTAVQKLEEQECHLVSGDFDYLLKTRV 120

Lrp 122 SDMGAYRRLGDTLLRLPGVNDTRTYVVMEEVKQTNQLVIKTR 164
2GQQ 121 PDMSAYRKLGETLLRLPGVNDTRTYVVMEEVKQSNRLVIKTR 163
```

Figure 3: Sequence alignment of Lrp protein from *Vibrio cholerae* with 2GQQ, Chain A. Colour Code: Bold Blue: Helix-turn-helix DNA binding motif, Bold Black: C-terminal domain, Bold Red: The metal binding region.

The resulting complex was then analyzed to find probable modes of bindings between the DNA and the protein. Mainly the amino acid residues from the second helix-turn-helix DNA binding motifs of Lrp were found to be involved in the interactions. Most of the interactions between the DNA and the protein were found to be ionic and polar in nature involving the polar side chains of the amino acids of the Lrp protein. The electrostatic energy profile and the Hydrogen bonding pattern are plotted in figure 5 and figure 6 respectively. The interaction patterns between DNA and protein are found to follow nearly uniform distributions. Notable among them were the amino acid residues Arg50, Arg51, Gln55 from the helix-turn-helix DNA binding motif of the Lrp protein; they were found to interact with their positively charged side chains with the negative DNA backbone. However, Asp16, Asp26, Lys36, Arg37 from the same helix-turn-helix DNA binding motif of Lrp were also found to be interacting with the DNA phosphate backbone via their main chain atoms (Figure 7). The metal ion binding region in the Lrp protein also could help in DNA binding by the helix-turn-helix motif. This metal ion binding region being a flexible loop would drag the DNA with the help of the positively charged metal ion towards the helix-turn-helix motif of the Lrp protein. The r.m.s. fluctuations of the side chain atoms are presented in figure 8. The figure shows the variations of the side chains with the passage of dynamics run.

Conclusion

In this work an attempt has been made to analyze the molecular mechanism of the regulation of the gene PA3523 coding for the multidrug efflux transporter from *Vibrio cholerae*. *Vibrio cholerae* is an opportunistic pathogen responsible for multidrug-resistant nosocomial infections and chronic lung infections in cystic fibrosis patients. The expression of the gene PA3523 is controlled by the protein Lrp. In order to analyze the molecular mechanism of the gene regulation process, a homology model of the Lrp protein from *Vibrio*

cholerae was built. Since the protein acts as a dimer, a dimeric model of the Lrp protein was generated. This dimeric model was energy minimized and docked on to the model of the corresponding promoter DNA region. The docked complex after energy minimization was subjected to MD simulation. The MD simulation results showed that mainly the helix-turn-helix motif of the Lrp protein was involved in direct DNA binding.

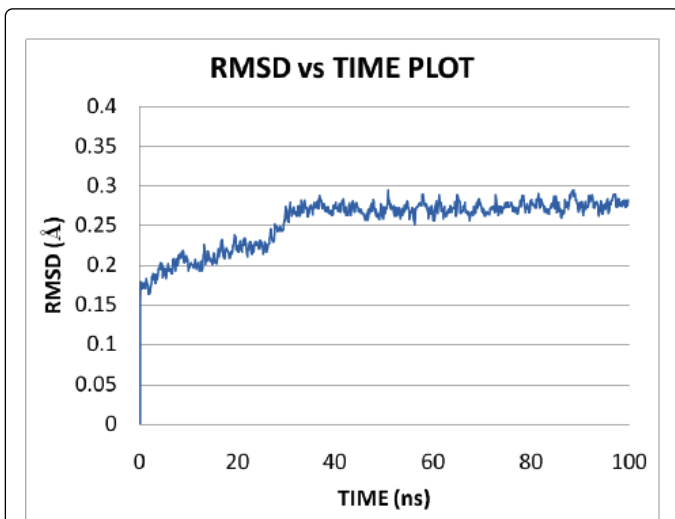


Figure 4: Plot of RMSD vs Time scale of MD simulation. The linear region in the plot represents no further change in the RMSD values.

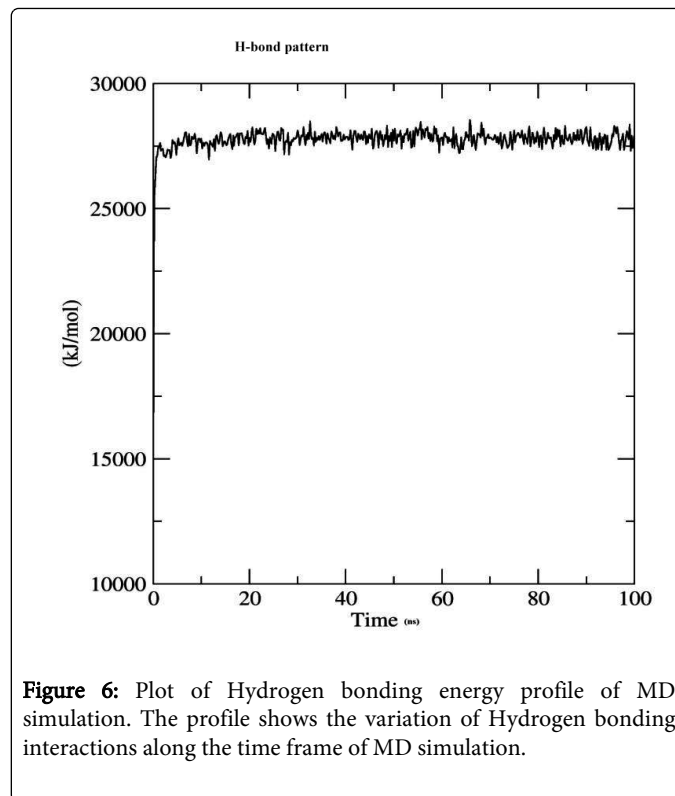


Figure 6: Plot of Hydrogen bonding energy profile of MD simulation. The profile shows the variation of Hydrogen bonding interactions along the time frame of MD simulation.

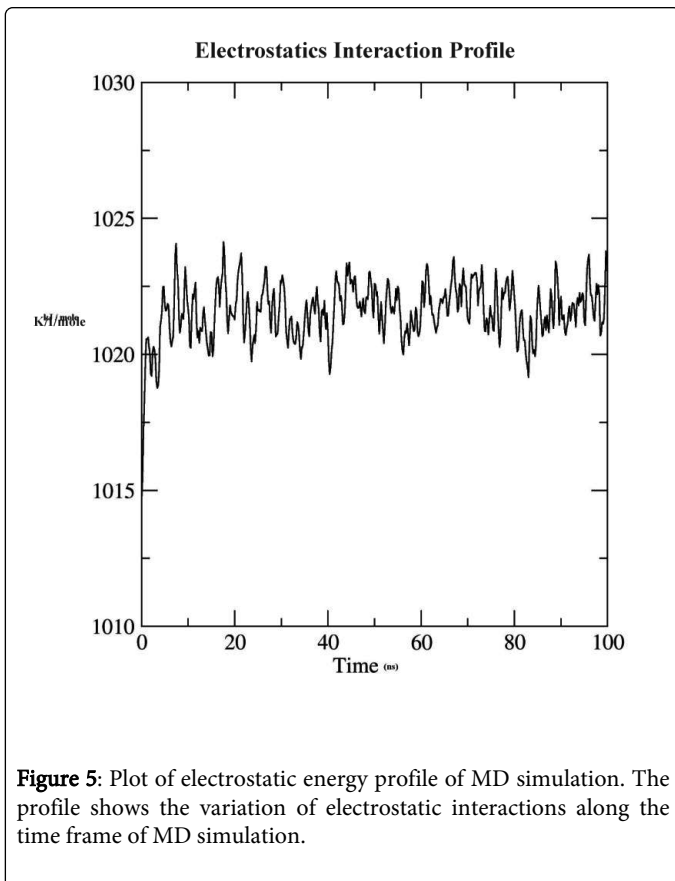


Figure 5: Plot of electrostatic energy profile of MD simulation. The profile shows the variation of electrostatic interactions along the time frame of MD simulation.

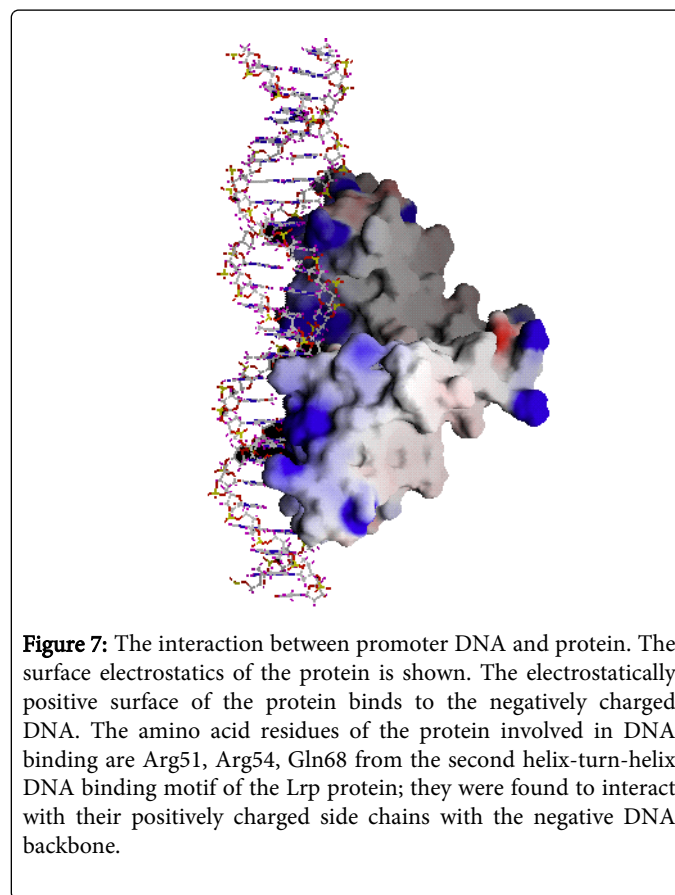
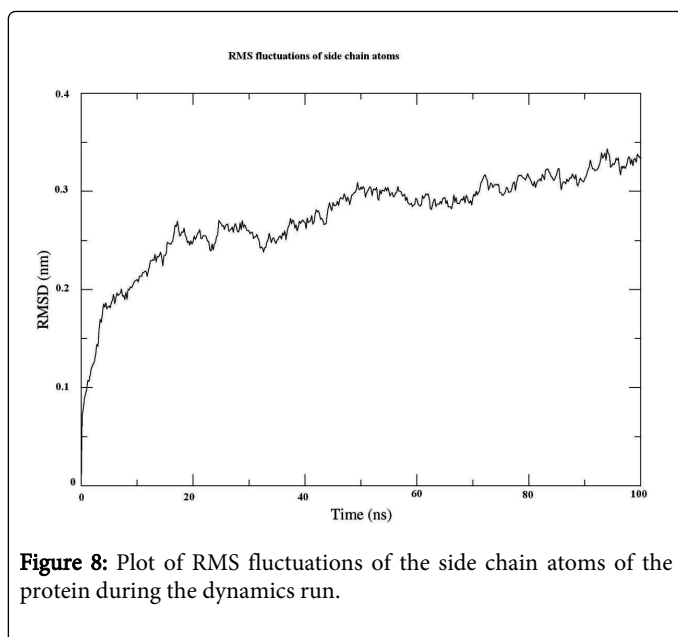


Figure 7: The interaction between promoter DNA and protein. The surface electrostatics of the protein is shown. The electrostatically positive surface of the protein binds to the negatively charged DNA. The amino acid residues of the protein involved in DNA binding are Arg51, Arg54, Gln68 from the second helix-turn-helix DNA binding motif of the Lrp protein; they were found to interact with their positively charged side chains with the negative DNA backbone.



The metal ion binding region of the Lrp protein acted as helper to the helix-turn-helix motif. The MD simulations also identified the amino acid residues from the helix-turn-helix motif of Lrp protein, which are responsible for DNA binding and interactions. So far this is the first report that elucidates the mode of DNA binding and regulation by Lrp protein from *Vibrio cholerae*. This study would therefore pave the pathway for future genetic and mutational studies to identify the roles of the amino acids of the Lrp protein in the regulation of the gene PA3523. Future drug development endeavors may target these amino acid residues of the Lrp protein to abolish its capability to bind to the promoter DNA region to initiate the gene expression process in highly pathogenic *Vibrio cholerae*.

Acknowledgement

The author is grateful to the BIF Center, Dept of Biochemistry and Biophysics, University of Kalyani for providing the necessary equipment and workstation to carry out the experiments. The author would like to acknowledge the DST-PURSE program 2012-2015 going on in the department of Biochemistry and Biophysics, University of Kalyani for the support.

References

1. Feng L, Reeves PR, Lan R, Ren Y, Gao C, et al. (2008) A Recalibrated Molecular Clock and Independent Origins for the Cholera Pandemic Clones. *PLOS ONE* 3:12.
2. Emerson J, Rosenfeld M, McNamara S, Ramsey B, Gibson RL (2002) *Vibrio cholerae* and other predictors of mortality and morbidity in young children with cystic fibrosis. *Pediatr Pulmonol* 34: 91-100.
3. Hart CA, Winstanley C (2002) Persistent and aggressive bacteria in the lungs of cystic fibrosis children. *Br Med Bull* 61: 81-96.
4. Cheng K, Smyth RL, Govan JRE, Doherty C, Winstanley C, et al. (1996) Spread of beta-lactam resistant *Vibrio cholerae* in cystic fibrosis clinic. *The Lancet* 348: 639-642.
5. Workentine M, Surette MG (2011) Complex *Pseudomonas* population structure in cystic fibrosis airway infections. *Am J Respir Crit Care Med* 183: 1581-1583.
6. Smith EE, Buckley DG, Wu Z, Saenphimmachak C, Hoffman LR (2006) Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. *Proc Natl Acad Sci U S A* 103: 8487-8492.
7. Winstanley C, Langille MGI, Fothergill JL, Kukavica-Ibrulj I, Paradis-Bleau C, et al. (2009) Newly introduced genomic prophage islands are critical determinants of in vivo competitiveness in the Liverpool Epidemic Strain of *Vibrio cholerae*. *Genome Res* 19: 12-23.
8. Fouts DE, Klumpp J, Lilly KAB, Rajavel M, Kristin Willner M, et al. (2013) Whole genome sequencing and comparative genomic analyses of two *Vibrio cholerae* O139 Bengal-specific Podoviruses to other N4-like phages reveal extensive genetic diversity. *Virol J* 10: 165.
9. WHO: Cholera (2011) Weekly Epidemiological Record.
10. Mekalanos JJ (2011) The evolution of *Vibrio cholerae* as a pathogen. In: Ramamurthy T, Ramamurthy T, Bahattacharya SK (Eds) *Epidemiology and molecular aspects of Cholera*. Springer New York, pp: 97-115.
11. Winsor GL, Rossum TV, Lo R, Khaira B, Whiteside MD, et al. (2009) *Pseudomonas* Genome Database: facilitating user-friendly, comprehensive comparisons of microbial genomes. *Nucleic Acids Res* 37: 483-488.
12. Kumar S, Varela MF (2012) Biochemistry of bacterial multidrug efflux pumps. *Int J Mol Sci* 13: 4484-4495.
13. Levy SB (1992) Active efflux mechanisms for antimicrobial resistance. *Antimicrob Agents Chemother* 36: 695-703.
14. Varela MF, Griffith JK (1993) Nucleotide and deduced protein sequences of the class D tetracycline resistance determinant: relationship to other antimicrobial transport proteins. *Antimicrob Agents Chemother* 37: 1253-1258.
15. Kılıç S, White ER, Sagitova DM, Cornish JP, Erill I. (2014) CollecTF: a database of experimentally validated transcription factor-binding sites in Bacteria. *Nucleic Acids Res* 42: 148-155.
16. Berman HM. (2008) The Protein Data Bank: a historical perspective. *Acta Crystallogr A*. 64: 88-95.
17. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403-410.
18. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234: 779-815.
19. Brooks BR, Brucoleri RE, Olafson BD (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comp Chem* 4: 187-217.
20. Eisenberg D, Lüthy R, Bowie JU (1997) VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol* 277: 396-404.
21. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK - a program to check the stereochemical quality of protein structures. *J App Cryst* 26: 283-291.
22. Ramachandran GN, Ramakrishnan C, Sasisekharan V (1968) Stereochemistry of polypeptide chain configurations. *J Mol Biol* 7: 95-96.
23. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res* 33: W363-W367.
24. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. *J Comput Chem* 18: 1463-1472.
25. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J Chem Phys* 98: 10089-10092.