IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 5 Issue 3, March 2018
ISSN (Online) 2348 – 7968
www.ijiset.com

In- Silico Structural & Molecular Characterization Of Clgr From Mycobacterium Tuberculosis

Yogesh N.Joshi and Rupali D. Salunkhe
Department of PG studies and Research in Bioinformatics, Walchand Centre for Biotechnology,
Walchand College of Arts & Science, Solapur, Maharashtra, India
email:rupali7salunkhe@gmail.com

ABSTRACT

Tuberculosis (TB), caused by Mycobacterium tuberculosis (Mtb), is the leading cause of death from an infectious disease worldwide. Chaperone and protease systems play essential roles in cellular homeostasis and have vital functions in controlling the abundance of specific cellular proteins involved in processes such as transcription, replication, metabolism and virulence. Bacteria have evolved accurate regulatory systems to control the expression and function of chaperones and potentially destructive proteases. ClgR performs different functions during stress response and is important for the pathogenicity of Mtb. Clp protease plays an unusual and essential role in Mtb and may serve as an ideal target for antimycobacterial therapy. Mtb encodes two ClpP genes, ClpP1 and ClpP2, which associate together to form a single proteolytic complex, referred to as ClpP1P2. ClgR regulation of chaperone and protease system is essential for mycobacterium tuberculosis parasitism of the microphages. Clp protease complex is an attractive novel target for antitubercular drugs because it is essential for growth and virulence. The present investigation includes retrieval of amino acid information and sequence analysis of ClgR from major protein sequence database and tools. The Physiochemical parameters like amino acid propensity, molecular weight, isoelectric point, aliphatic index, hydropathicity were determined. The secondary structure of clgR was predicted using SOPMA secondary structure method. The 3D structure was predicted using SWISS-MODEL server and model was further validated using PROCHECK analysis tool. The function of clgR was identified by Pfam domain database.

Keywords: Mycobacterium tuberculosis, ClgR, in silico, caseinolytic protease (Clp)etc.

INTRODUCTION

Tuberculosis (TB) is a major global public health problem. It is an infectious disease caused by *Mycobacterium tuberculosis* (*M.tb*), typically affects the lungs but can also affect other sites and spreads when a person with TB expels the bacteria into the air while coughing or sneezing. TB is one of the top 10 causes of death worldwide and the leading cause of death from an

infectious disease. World Health Organization estimates that, in 2016, there were 10.4 million new TB cases (including 1.2 million among HIV-infected individuals) worldwide. 90% of cases were in adults and 10% in children with a male female ratio of 1.6:1. 5–15% of the estimated two to three billion people infected with *M.tb* will develop TB disease during their lifetime, the probability increasing among people with HIV [1]. Tuberculosis is a chronic granulomatous infectious disease. Infection occurs via aerosol, and inhalation of a few droplets containing *M. tuberculosis* bacilli. After infection, *M. tuberculosis* pathogenesis occurs in two stages. The first stage is an asymptomatic state that can persist for many years in the host, called latent TB. When the immune system is weak, the bacteria begin replicating and cause characteristic symptoms such as cough, chest pain, fatigue and unexplained weight loss. If left untreated, the disease eventually culminates in death [2]. Proteases and their associated chaperones carry out an essential homeostatic role in all living cells. They catalyze the degradation and recycling of truncated, damaged or aggregated proteins. Protease systems also have a direct regulatory role in the physiological state of the cell by controlling the abundance of specific proteins involved in a variety of fundamental cellular processes such as transcription, DNA replication and cell division [3]. Rv2745c, the *M. tuberculosis* ClgR homologue, binds upstream of *clpP1P2* and activates transcription in *M. tuberculosis* [4].Deletion of ClgR results in a reduced capacity to replicate in macrophages [5]. Moreover ClgR, as well as the two ClpP proteases, are involved in the reaeration response: they have been found to be induced during the transition of *M. tuberculosis* from bacteriostasis to growth [4].The caseinolytic protease gene regulator ClgR is an activator of the clpP1P2 genes and also suggested that this transcription factor may be a substrate of the protease. Interestingly, accumulation of ClgR appears to be toxic for bacilli, suggesting a mechanism for how pharmacological inhibition of ClpP1P2 protease activity by bortezomib translates into whole-cell antibacterial activity. The mycobacterial caseinolytic protease (Clp) is, similarly to the human proteasome, a degradative protease machine with a role in proteome housekeeping [6,7]. The *M. tuberculosis* Clp protease complex is an attractive novel target for antitubercular drugs because it is essential for growth and virulence [8]. The present study focused on the clgR Transcriptional regulator protein as novel target in *Mycobacterium Tuberculosis.* The sequence and structural analysis of clgR which detailed for structure based drug designing approach.

MATERIALS AND METHODS

1) Retrieval of ClgR (CLGR_MYCTU Transcriptional regulator ClgR) from Protein database

The Transcriptional regulator protein sequence (ClgR) from *Mycobacterium tuberculosis* was retrieved from UniprotKB database [9]. UniProtKB is public protein database which contains the amino acid sequences of proteins. The sequence was retrieved & saved in FASTA file format

with its Accession ID.

ii) Physicochemical analysis of ClgR by ProtParam tool

Physicochemical properties of ClgR were performed by using ProtParam analysis tool which on ExPASy server [10]. It allows the computation of various physical and chemical parameters for a given protein. The computed parameter includes amino acid composition, molecular weight, theoretical pI, Instability index, Grand average of hydropathicity.

iii) Identification of functional domain in ClgR from Pfam database

Domain is the most important factor governing the protein folding into the structure. The domain of the ClgR protein was predicted from the Pfam domain database which contains the information about protein families & domains [11].

iv) Secondary structure prediction and analysis of ClgR

The secondary structure of ClgR was predicted by SOPMA secondary structure prediction method [12]. SOPMA stands for self-optimized prediction method with alignment for the prediction of helix, strands and coils of the protein sequence.

v) Prediction, Validation & Visualization of 3D structure of ClgR

The 3D structure of ClgR was predicted by using Swiss-model server [13]. The selection of template was accomplished by protein BLAST using PDB database having identity more than 30%. The evaluation and validation of generated model was performed with PROCHECK server on PDBSum database [13] and predicted model was visualized by Rasmol visualization tool [15].

RESULTS & DISCUSSION

i) Retrieval of ClgR (CLGR_MYCTU Transcriptional regulator ClgR) from Protein database

Transcriptional regulator ClgR [Uniprot ID: P9WMH7] sequence from *Mycobacterium tuberculosis* was retrieved from UniProtKB database with its 112 amino acids and saved in FASTA format which shown as below,

>sp|P9WMH7|CLGR_MYCTU Transcriptional regulator ClgR OS=Mycobacterium tuberculosis (strain ATCC 25618 / H37Rv) GN=clgR PE=1 SV=1
MAALVREVVGDVLRGARMSQGRTLREVSDSARVSLGYLSEIERGRKEPSSELLSAICTA
LQLPLSVVLIDAGERMARQERLARATPAGRATGATIDASTKVVIAPVVSLAVA

ii) Physicochemical analysis of ClgR by ProtParam tool

Physicochemical composition of ClgR was analyzed by using ProtParam analysis tool on ExPASy server. The physiochemical parameters were tabulated in Table no. 1. As per table instability index is 46.56 classifies the ClgR is unstable, on the basis theoretical pI the ClgR is basic in nature, as there are presence of positively and negatively charged amino acids and the ClgR is cationic in nature.

Table 1: Physicochemical parameters of ClgR**.**

Sr.No.	***Parameters***	***Values***
1.	Number of amino acids	112
2.	Molecular weight	11785.67
3.	Theoretical pI	9.75
4.	Instability index	46.56
5.	Grand average of hydopathicity	0.202
6.	No. of positively charged amino acids	15
7.	No. of negatively charged amino acids	12

iii) Identification of functional domain in ClgR from Pfam database

The functional domain of ClgR was predicted by Pfam domain database which shows only one domain HTH_3 showed in Fig no. 1 and the region was shown in Table no 2. The main function of the helix-turn-helix (HTH) is a major structural motif capable of binding DNA. It is composed of two α helices joined by a short strand of amino acids and is found in many proteins that regulate gene expression. It should not be confused with the helix-loop-helix domain.

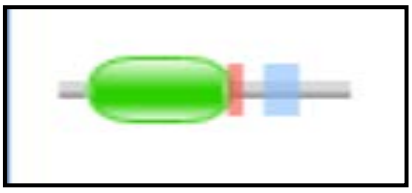

Figure 1: Domains of ClgR predicted in Pfam database.

Table 2: Starting & ending regions of ClgR **domains**.

Sr. no.	***Domains***	***Start***	***End***
1	HTH_3	13	67

iv) Secondary structure prediction and analysis of ClgR

The secondary structure of ClgR was predicted by SOPMA Secondary Structure Prediction method. Secondary structural elements Alpha helices, strands & coils were enlisted in following

Table no.3 and Fig no. 2. The table shows the ClgR has more number of coils that is 33.04% followed by alpha helices 52.62% and strands 14.29%.

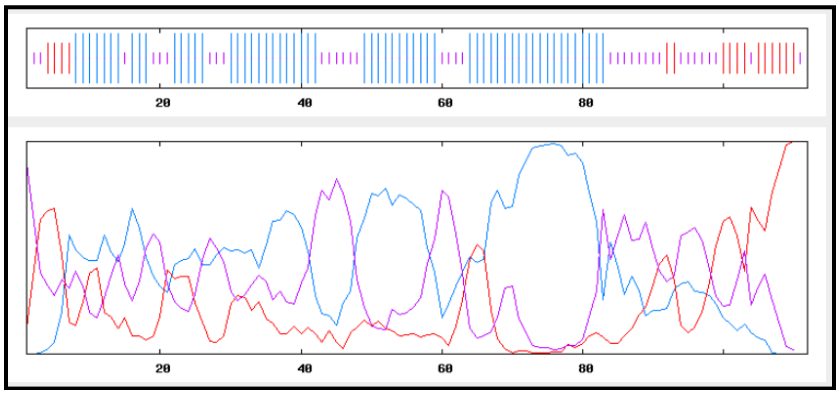

Figure 2: Secondary structure of ClgR protein using SOPMA

Table 3: Secondary structure of ClgR**.**

Sr.No .	***Secondary structure***	***No. of residues***	***Percentage***
1.	Alpha helices	59	52.68
2.	Coils	37	33.04
3	Strands	16	14.29

v) Prediction, Validation & Visualization of 3D structure of ClgR

The homology modeling of ClgR from *Mycobacterium tuberculosis* was obtained through SWISS MODEL server using resolution 1.8 A˚ structure of ClgR from *Mycobacterium smegmatis* (PDB Id:5woq.2.A Chain A) as a template. The evaluation and validation of generated model were executed with PROCHECK server on PDBSum database which is shown in Fig.no.2. Validation of the predicted ClgR from *Mycobacterium tuberculosis* by PROCHECK analysis showed that 97.9% of the residues of model were present in the most favoured region followed by 97.9% in the allowed region, 0% in generously allowed region and disallowed region respectively of Ramachandran plot which are shown in Fig. no.3 and Table no.4. Further the predicted structure was visualized by Rasmol viewer which is shown in Fig.no.4.

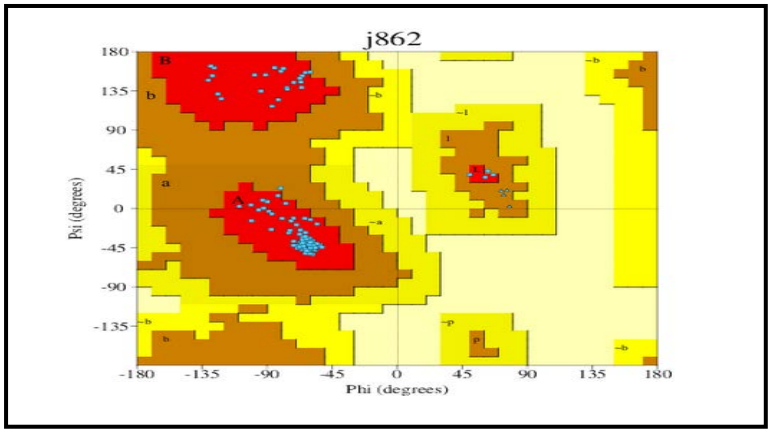

Figure 3: Ramchandran plot of Predicted 3D structure of ClgR**.**

Table 4: Residue numbers & its % in different regions of Ramchandran plot of ClgR**.**

Sr. no.	***Regions***	***Residue no.***	***Percentage***
1.	Most favored regions [A,B,L]	143	97.9
2.	Additional allowed regions [a, b, l, p]	3	2.1
3.	Generously allowed regions [~a, ~b, ~l, ~p]	0	0
4.	Disallowed regions	0	0

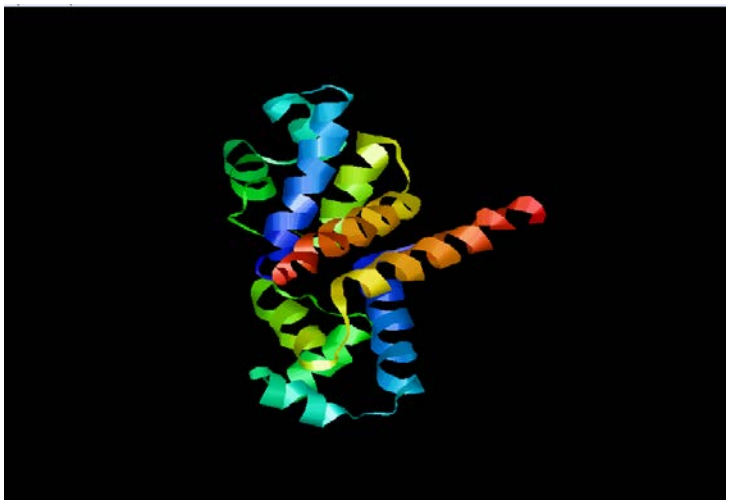

Figure 4: Predicted 3D structure of ClgR **viewed in RASMOL.**

CONCLUSION

The present preliminary investigation mainly leads to understand the basic primary, secondary structure and tertiary structure of ClgR from *Mycobacterium tuberculosis* using various *in-silico* tools and techniques. The primary structure illustrates that the ClgR is an enzyme with its 112 amino acids. The physicochemical properties depict that the ClgR is basic, unstable and cationic protein. The secondary structure reveals that ClgR consist of a helix, a sheet and random coil structure within its short stretch of residues. The 3D structure predicted by SWISS-MODEL was validated using PROCHECK, the percentage of most favorable region was 97.9%. This present study put molecular insight into the further studies to find the structural and functional properties of this ClgR to find or design the novel inhibitors of *Mycobacterium tuberculosis* by Structure based drug design approach.

REFERENCES

1. World Health Organization. *Global Tuberculosis Report 2016*. World Health Organization (2017).
2. Kaufman S, Introduction, Seminars in Immunology, Volume 26, Issue 6, Pages 429-430
3. Darwin, K. H. (2009). Prokaryotic ubiquitin-like protein (Pup), proteasomes and pathogenesis. Nat Rev Microbiol 7,

485–491.

4. Sherrid AM, Rustad TR, Cangelosi GA, Sherman DR (2010) Characterization of a Clp protease gene regulator and the reaeration response in *Mycobacterium tuberculosis.* PLoS One 5: e11622.
5. Estorninho M, Smith H, Thole J, Harders-Westerveen J, Kierzek A, et al. (2010) ClgR regulation of chaperone and protease systems is essential for *Mycobacterium tuberculosis* parasitism of the macrophage. Microbiology 156: 3445–3455.
6. Raju RM, Unnikrishnan M, Rubin DH, Krishnamoorthy V, Kandror O, Akopian TN, Goldberg AL, Rubin EJ. 2012. Mycobacterium tuberculosis ClpP1 and ClpP2 function together in protein degradation and are required for viability in vitro and during infection. PLoS Pathog 8:e1002511.
7. Raju RM, Jedrychowski MP, Wei JR, Pinkham JT, Park AS, O'Brien K, Rehren G, Schnappinger D, Gygi SP, Rubin EJ. 2014. Post-translational regulation via Clp protease is critical for survival of Mycobacterium tuberculosis. PLoS Pathog 10:e1003994.
8. Sassetti, C.M., Boyd, D.H., and Rubin, E.J. (2001) Comprehensive identification of conditionally essential genes in mycobacteria. *Proc Natl Acad Sci USA* 98: 12712–12717.
9. Magrane M. And the UniProt consortium. UniProt Knowledgebase: a hub of integrated protein data Database, 2011, bar009.
10. Gasteiger E, Hooglan C, Gattiker A, Duvaud S, Wilkins MR, Appel RD *et al*. The Proteomics Protocols Handbook, Human Press, 2005, 571-607.
11. Erik LL, Sonnhammer, Eddy SR, Durbin R. Pfam: comprehensive database of protein domain families based on seed alignment. PROTEINS: structure, functions and Genetics 1997; 28:405-420
12. Geourjon C, Deléage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci 1995; 11(6):681-4
13. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISSMODEL Workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* 2006; 22:195-201
14. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK - a program to check the stereochemical quality of protein structures. J App. Cryst. 1993; 26:283-291.
15. Sayle RA, Milner-White EJ. (RASMOL: biomolecular graphics for all. Trends Biochem Sci. 1995; 20(9):374