Journal of Computational Intelligence in Bioinformatics
ISSN: 0973-385X Volume 4 Number 1 (2011) pp. 1–5

http://www.ripublication.com/jcib.htm

Insilico Studies and the Homology Modeling of a Hypothetical Protein in Mycobacterium tuberculosis H37Rv

S. Amala

Lecturer, Department of Biotechnology & Bioinformatics
Dhanalakshmi Srinivasan College of Arts & Science for Women
Perambalur-621212, Tamil Nadu, India
E-mail: amala.santhanam@gmail.com

Abstract

Mycobacterium tuberculosis H37Rv strain was first isolated in the year 1905. It was pathogenic and most widely used in the tuberculosis research. An *in-silico* technique was initiated to characterize a hypothetical protein to deduce its structural and functional information. The hypothetical protein analyzed in the present study showed two conserved domains Patatin like Phopsholipase and cNMP family of transcription factors. The modeled protein revealed the presence of maximum number of random coils in the secondary structural elements. The existence of two domains indicated its role in enzymatic activity of a lipid acyl hydrolase and cNMP binding proteins. The 3D modeling of the protein showed that 98% of the residues were in allowed regions.

Keywords: H37Rv, Tuberculosis, Hypothetical, Mycobacterium.

Introduction

Mycobacterium is a genus within the order Actinomycetales that comprises a large number of well characterized species. Several of which are associated with human & animal diseases such as tuberculosis & leprosy

Mycobacterium tuberculosis is pathological; bacterial species and it's this causative agent of most cases of tuberculosis. It was first discovered by Robert Koch in the year 1882.Mycobacterium tuberculosis is highly aerobic and require high level of oxygen. This infect the lungs which cause tuberculosis.This genome was sequenced in the year 1998.

Mycobacterium tuberculosis H37Rv strain was first isolated in 1905 has remained pathogenic & most widely used strain in Tuberculosis research. The complete sequence and annotation of the H37Rv strain was published in the year 1998. Its size is about 4 million base pairs .The original sequence and annotation of Mycobacterium tuberculosis H37Rv identified 3974 genes [Cole et al.,] .This 3924 genes thought to encode protein & 50 encoding stable RNA. Among these genes 40% of the gene functions are characterized with possible function postulated for another 44%. The genome has 6 pseudo genes.

The reannotation included 82 additional genes. All new genes encode protein and no change in the number of RNA molecules .The reannotaion of genome gave many changes in the functional classification of predicted proteins. The unknown protein number has been decreased from 606 to 272 function of 2058 proteins was predicted and more than 150 was experimentally proven in mycobacterial research. The number of conserved hypothetical protein changes from 910 to 1051(Jean christophe *et al.*, 2002)

The Mycobacterium tuberculosis H37Rv genome contains 250 genes involved in fatty acid metabolism with 39 involved in polyketide metabolism generating the waxy coat 10% of the coding capacity is taken up by 2 clustered gene families that encode acidic glycine rich proteins

Materials and methods

To analyze the hypothetical protein & assign its functional and structural roles various tools & softwares were used .The primary sequence of the hypothetical protein was obtained form Genbank at NCBI. The sequence was then compared for detecting the homologues sequence found in database using BLASTP.Protparam is the tool which was used for the calculation of the physicochemical properties. The secondary structure of the protein was predicted by tools like GOR .The conserved domains in the proteins were detected from the BLAST analysis. The Family of the protein was recognized by using Pfam database. Since only the primary sequence information was available from NCBI, no structure in the X-ray crystallographic data was available from PDB.Hence the modeling of the protein has to be done to deduce the 3D structure of protein .Here Homology modeling was done by using Swiss model & the structure was validated by PSVS server. Then the 3D co-ordinate file was visualized in the Rasmol

Results

The similarity search for the sequence was carried out with the help of BLAST tool. The results indicated two conserved domains Patatin like phospholipase domain containing protein 6, protein 7 & CAP family of transcription factors (Fig 1.). Members include CAP (or) CAMP receptor protein (CRP). The first domain identified in the protein sequence was positioned at 42 to 128 and the second domain at 309 to 469. The Super families of these domains were cNMP-binding & Patatin (Table .1) . The Physicochemical properties of the hypothetical protein showed that

the number of amino acids are 563 with a Molecular weight of 621237 & isoelectric point as 5.89.The maximum number of the amino acid present in the sequence was found to be alanine (14.1%) & the least was found to be cysteine (0.5%). Total number of positively charged residues were 59 and negatively charged residues were 67.The instability index of the protein was computed to be 42.71 which classified the protein as unstable. The grand hydropathicity was calculated to be 0.032 .The secondary structural analysis of protein revealed that the random coil (47.71%) was found to be the most frequent followed by alpha helix (34.65%) .Extended strand was found to be least frequent (18.18%) (Fig. 2). The structure of the hypothetical protein was deduced by homology modeling .The homology modeling was performed using Swiss model workspace (Fig.3). The structure was validated by PSVS tool. The Ramachandran plot revealed that 98% of the residues are in the allowed regions (Fig.4)

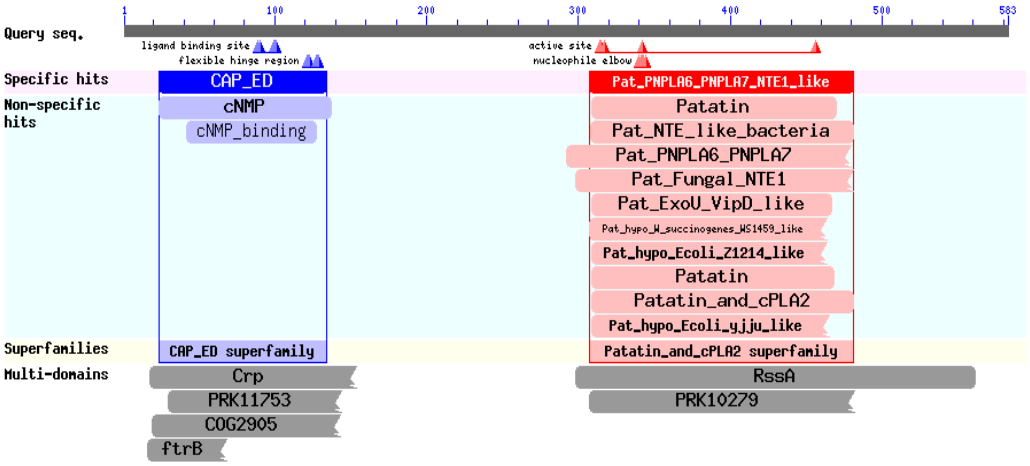

Figure 1: Graphical Representation of conserved domain in the hypothetical protein.

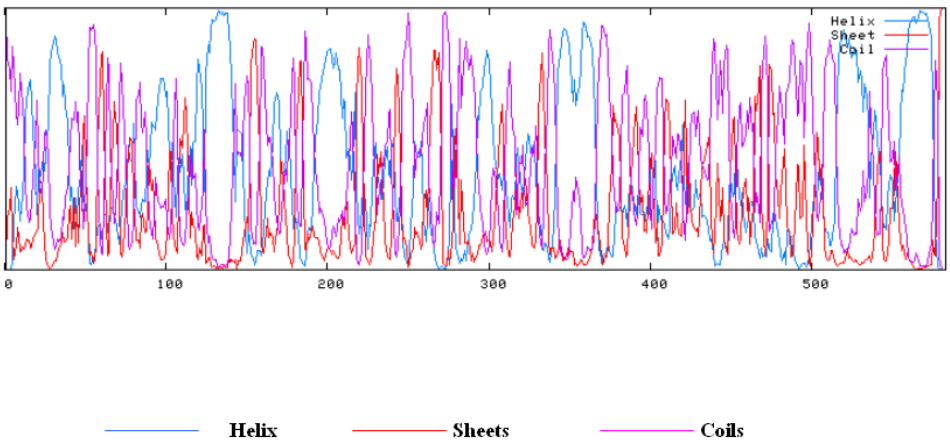

Figure 2: Graphical Representation of Secondary Elements in hypothetical Protein.

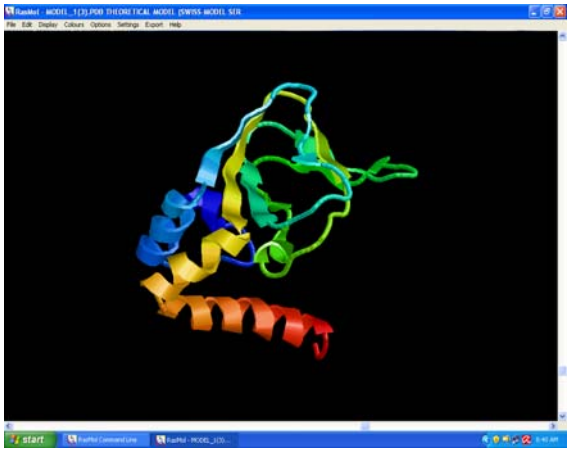

Figure 3: Modeled 3D structure of the hypothetical protein.

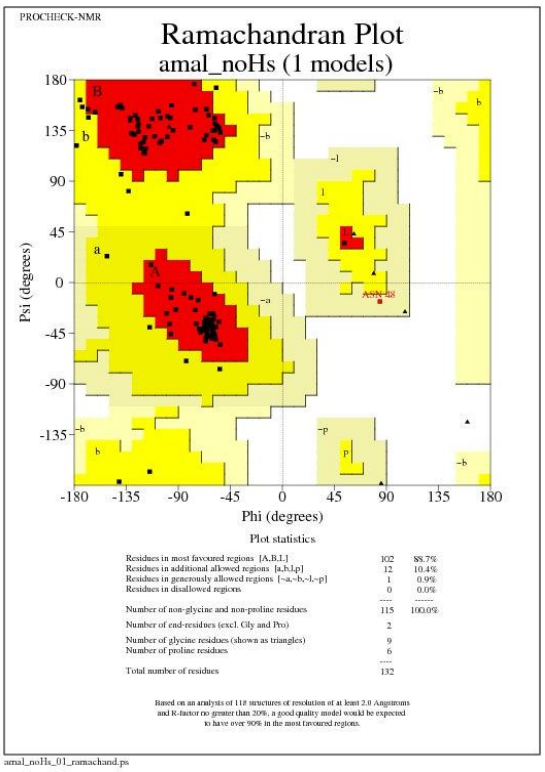

Figure 4: Graphical representation of Ramachandran plot by PSVS.

Table 1: Pfam results of the hypothetical protein.

Family	Description	Entry type	Envelope	
			Start	End
cNMP_binding	Cyclic nucleotide-binding domain	Domain	42	128
Patatin	Patatin-like phospholipase	Family	309	469

Conclusion

In the present study the sequence and structure analysis of the hypothetical protein was done by various tools and softwares.The analysis of the hypothetical protein showed sequence similarity mostly to the Patatin like phospholipase domain containing protein 6, protein 7 & CAP family of transcription factors. The domains identified are Patatin and cNMP binding domain. Patatin is a group of plant storage glycoprotein that show lipid acyl hydrolase activity. cNMP binding domain is a domain present in ion channels and cNMP –dependent kinases. The dominance of the coiled regions indicated high level of conservation and stability of protein. Further research involving development of appropriate strategies for studying these domains in the hypothetical protein and also help in annotating the function of the protein in Mycobacterium tuberculosis H37Rv.

References

[1] Bifani P, Moghazeh S, Shopsin B, Driscoll J, Ravikovitch A, Kreiswirth BN., 2000 “Molecular characterization of Mycobacterium tuberculosis H37Rv/Ra variants: distinguishing the mycobacterial laboratory strain” , J Clin Microbiol. 38(9), pp. 3200-3204.

[2] Camus JC, Pryor MJ, Medigue C, Cole ST.,2002 “ Re-annotation of the genome sequence of Mycobacterium tuberculosis H37Rv”, Microbiology., 148(10) pp.2967-2973

[3] Barbara H Miller and Thomas M Shinnick., 2001 “Identification of two *Mycobacterium tuberculosis* H37Rv ORFs involved in resistance to killing by human macrophages” , BMC Microbiol. 1: 26.

[4] Joanna C. Betts, Paul Dodson, Selwyn Quan, Alan P. Lewis, Pam J. Thomas, Ken Duncan., 2000 “Comparison of the proteome of Mycobacterium tuberculosis strain H37Rv with clinical isolate CDC 1551 ”, Microbiology 146 .pp. 3205-3216.

[5] Fleischmann, R. D. D. Alland,. A. Eisen,. Carpenter,. White,J. Peterson, R. DeBoy, R. Dodson, M. Gwinn, D. Haft, E. Hickey, J. F. Kolonay, W. C. Nelson, L. A. Umayam, M. Ermolaeva, S. L. Salzberg, A. Delcher, T. Utterback, J. Weidman,H. Khouri, J. Gill,A. Mikula, W. Bishai, W. R. Jacobs, Jr., J. C. Venter,and C. M. Fraser., 2002 “Whole-Genome Comparison of Mycobacterium tuberculosis Clinical and Laboratory Strains” J Bacteriol 184(19) pp. 5479–5490.

[6] Nuria Andreua, Isidre Gibertab., 2008 “Cell population heterogeneity in Mycobacterium tuberculosis H37Rv”. Tuberculosis 88(6), pp.553-559.