Journal of Computational Intelligence in Bioinformatics.
ISSN: 0973-385X Volume 4 Number 1 (2011) pp. 79—86

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

Homology Modeling based Protein Structure Prediction of Iron-regulated Peptidyl-Prolyl cis-trans Isomerase A from *Mycobacterium tuberculosis* Strains H37Rv

Harshada H. Pilley* and Nitin M. Atre

P.G. Department of Bioinformatics, Shri Shivaji Science College, Amravati, India
**Corresponding Author E-mail: harshadapilley15@gmail.com*

Abstract

The complete genome sequence of the best-characterized strain of *Mycobacterium tuberculosis* H37Rv has been determined and analyzed in order to improve our understanding of the biology of this slow-growing pathogen and to help the conception of new prophylactic and therapeutic interventions. The *Mycobacterium tuberculosis* strains H37Rv and H37Ra are the most commonly used controls for *Mycobacterium tuberculosis* identification in the clinical and research laboratory setting. PPIASES accelerate the folding of proteins. The process by which a protein folds into the structure is called protein folding. In order to carry out their function, they must take on a particular shape, also known as a "fold". Proteins what are synthesized have to fold correctly so that they would acquire functional conformation. Despite of its importance 3D structure of this protein is not reported yet.

Homology modeling, also known as comparative modeling of protein refers to constructing an atomic-resolution model of the "*target*" protein from its amino acid sequence and an experimental three-dimensional structure of a related homologous protein (the "*template*"). For the modeling, template protein was obtained by Geno3D server, template protein pdb|1W74| chain A having identity 100%, E value 2e-103 and alignment score 370. By comparing the template protein a rough model was constructed for the target protein using Modeller9v7, a program for comparative modeling. The structure of iron-regulated peptidyl- prolyl cis-trans isomerase A of *Mycobacterium tuberculosis* strain H37Rv was found to resemble the X-ray structure of Peptidyl – Prolyl cis-trans isomerase A, PPIA, RV0009, from *Mycobacterium tuberculosis*. From Ramachandran plot analysis it was found that the portion

of residues falling into the most favored regions was 89.7%. Structure validation by Verify3D gives 93.99% of the residues. The predicted 3-D model may be further characterized and analyzed using other techniques.

Keywords: iron-regulated peptidyl-prolyl cis-trans isomerase A, Homology Modeling, Mycobacterium tuberculosis, H37Rv.

Introduction

Countless millions of people have died from tuberculosis, a chronic infectious disease caused by the tubercle bacillus [1][3]. The complete genome sequence of the best characterized strain of *Mycobacterium tuberculosis* H37Rv has been determined and analyzed in order to improve our understanding of the biology of this slow growing pathogen and to help the conception of new prophylactic and therapeutic interventions [1]. The *Mycobacterium tuberculosis* strains H37Rv and H37Ra are the most commonly used controls for *Mycobacterium tuberculosis* identification in the clinical and research laboratory setting. Iron-regulated peptidyl-prolyl cis-trans isomerase A PPIASES accerelate the folding of proteins [CATALYTIC ACTIVITY: Cis-Trans Isomerization of Proline Imidic Peptice Bonds In Oligopeptides] [1].

Homology modeling relies on the identification of one or more known protein structures likely to resemble the structure of the query sequence and on the production of an alignment that maps residues in the query sequence to residues in the template sequence [2]. Modeller9v7 is a computer program that models three-dimensional structures of proteins and their assemblies by satisfaction of spatial restraints [15]. Modeller9v7 is most frequently used for homology or comparative protein structure modeling: The user provides an alignment of a sequence to be modeled with known related structures and Modeller9v7 will automatically calculate a model with all non-hydrogen atoms [15].

Since there was no structure reported of iron-regulated peptidyl-prolyl cis-trans isomerase A of *Mycobacterium tuberculosis* strains H37Rv, the main aim of this study is to predict three dimensional structure of this important protein and inspects its various aspects using different techniques.

Material and Method

Retrieval of Target Sequence

The amino acid sequence of the iron-regulated peptidyl- prolyl cis-trans isomerase A, was obtained from the sequence database of NCBI (Accession No: NP_214523.1 and GeneID: 15607151) [12]. It was ascertained that the three dimensional structure of the protein was not available in Protein Data Bank. Hence the present exercise of developing the 3D model of the protein of *Mycobacterium tuberculosis* strains H37Rv was undertaken. The protein consists of 182 amino acids.

Secondary Structure Prediction

The amino acid sequence of the iron-regulated peptidyl- prolyl cis-trans isomerase A,

was subject to predict the secondary structure by GOR-IV (secondary structure prediction method) [7].

```
         10        20        30        40        50        60        70
          |         |         |         |         |         |         |
MADCDSVTNSPLATATATLHTNRGDIKIALFGNHAPKTVANFVGLAQGTKDYSTQNASGGPSGPFYDGAV
ccccccccccccchhhhhhccccceeeeeecccccceeeceeeeccccccccccccccccccccccce
FHRVIQGFMIQGGDPTGTGRGGPGYKFADEFHPELQFDKPYLLAMANAGPGTNGSQFFITVGKTPHLNRR
ecceeeeeeeeccccccccccccccccccccccccccchhhhhhhccccccccceeeeeeccccccccc
HTIFGEVIDAESQRVVEAISKTATDGNDRPTDPVVIESITIS
ceeeeeeechhhhhhhhhhhccccccccccccceeeeeeeec
Sequence length:   182

GOR4:
   Alpha helix      (Hh) :    24 is  13.19%
   3₁₀ helix        (Gg) :     0 is   0.00%
   Pi helix         (Ii) :     0 is   0.00%
   Beta bridge      (Bb) :     0 is   0.00%
   Extended strand  (Ee) :    45 is  24.73%
   Beta turn        (Tt) :     0 is   0.00%
   Bend region      (Ss) :     0 is   0.00%
   Random coil      (Cc) :   113 is  62.09%
   Ambigous states  (?)  :     0 is   0.00%
   Other states          :     0 is   0.00%
```

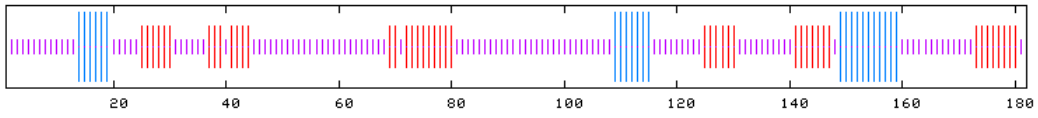

Figure 1: Predicted secondary structure of iron-regulated peptidyl-prolyl cis-trans isomerase A protein of *Mycobacterium tuberculosis strains H37Rv*

Template Searching

An attempt was made to find a suitable template protein for the modeling of the target protein. The template protein was searched through Geno3D server [5], (an online tool for searching similar sequences, based on sequence and structure-wise similarity). From the homology searching X-ray structure of peptidyl-prolyl cis-trans isomerase A from the *Mycobacterium tuberculosis* H37Rv [1W74: A] Chain A was selected as template protein.

Sequence Alignment

Amino acid sequence alignment of target and template proteins was derived using the Swiss-PdbViewer package [6]. Default parameters were applied and the aligned sequences were inspected and adjusted manually to minimize the number of gaps and insertions.

Rough Model

Rough 3-D models (20 models) were constructed from the sequence alignment

between iron-regulated peptidyl-prolyl cis-trans isomerase A protein and the template proteins using MODELLER 9v7 [15].

Figure 2: Predicted Model of iron-regulated peptidyl-prolyl cis-trans isomerase A protein of *Mycobacterium tuberculosis* strains H37Rv

Evaluation of Refined Model

In the last step of HM the refined structure of the model was subjected to a series of tests for testing its internal consistency and reliability [2][4][10]. Backbone conformation was evaluated by the inspection of the Psi/Phi Ramachandran plot obtained from PROCHECK [9] analysis. The Swiss-PdbViewer energy minimization test was applied to check for energy criteria in comparison with the potential of mean force derived from a large set of known protein structures.

Result and Discussions

In general, 30% sequence homology is required for generating useful models. In our study, based on the results obtained from Geno3D server, the X-ray structure of the Chain A, pdb|1W74|A, that is a high resolution structure of iron regulated peptidyl-prolyl cis-trans isomerase A from the *Mycobacterium tuberculosis* H37Rv was selected as template. By using BLAST, for template protein |1W74:A| E-value 2e-103, identity 100% and alignment score 370 was obtained. Modeller9v7 was used for building the model and global energy minimization. With the help of Modeller9v7 software 20 models were generated, stereochemical parameters of the proteins like main and side chains data of iron-regulated peptidyl-prolyl cis-trans isomerase A, was considered for determining the quality of the model which were generated by using PROCHECK 3.0 [9]. The main chain parameters like Ramachandran plot quality; peptide bond planarity, C-alpha Chirality, over-all G factor and the bad contacts per 131 residues are found to be within the limits for the model. The side chain parameters are in better range and within the limits for iron regulated peptidyl-prolyl cis-trans isomerase a from the *Mycobacterium tuberculosis* H37Rv. After comparing all parameters best model was obtained. Best model was analyzed by different protein

analysis programs including RAMPAGE Software for the evaluation of the Ramachandran plot quality, and WHATIF for the calculation of packing quality.

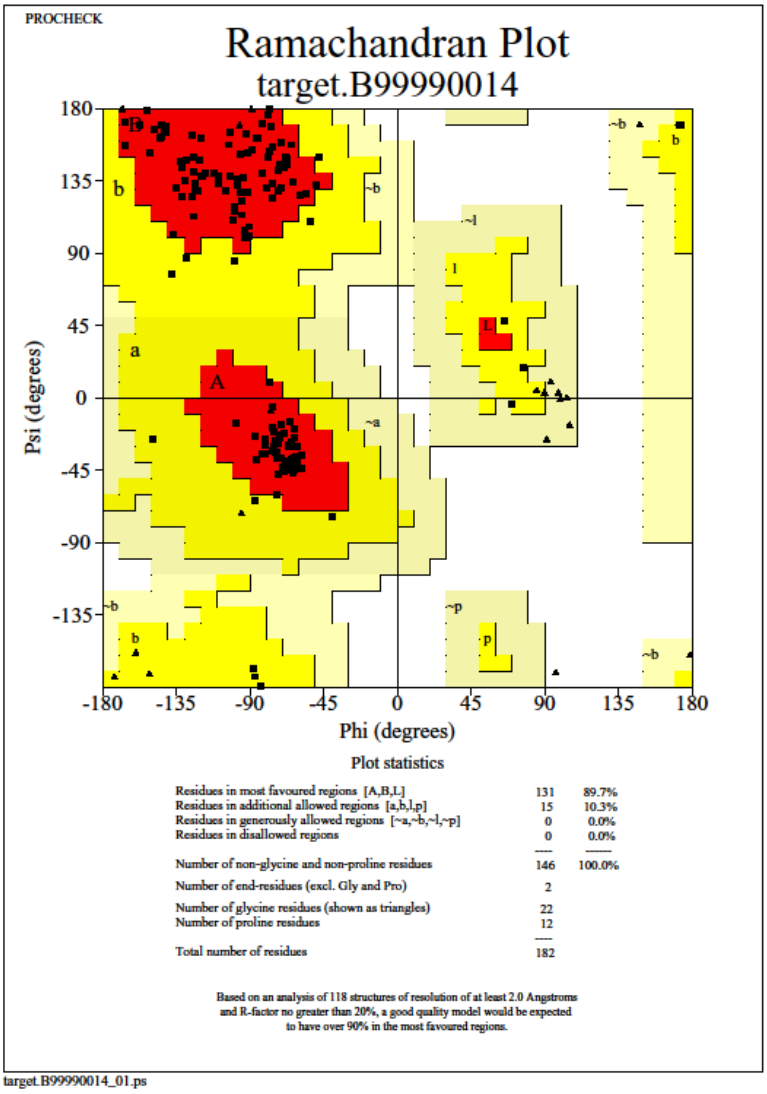

Figure 3: The Ramachandran plot for iron-regulated peptidyl-prolyl cis-trans isomerase A, using PROCHECK software

The Ramachandran plot for iron-regulated peptidyl- prolyl cis-trans isomerase A, using PROCHECK software revealed that among the 182 residues, 131 (89.7%) were in favoured region, 15 (10.3%) were in allowed region, 0 (0.0%) were in generously allowed regions and 0 (0.0%) were in disallowed region proving that the predicted model is acceptable (Fig. 3). Ramachandran plot for general, glycine, pre-proline and proline was also done and it showed the glycine, pre-pro and proline of iron-regulated peptidyl-prolyl cis-trans isomerase A falling under allowed regions. Packing quality

for this model was found in normal range with the help of WHATIF server. From the tools VERIFY_3D & ERRAT which is available at Structural Analysis and Verification Server (SAVES) it was shown that 89.7% of the residues had an averaged 3D-1D score > 0.2 and Overall quality factor 87.861.

The overall results provided the evidences that the predicted 3-Dimensional structure of for iron-regulated peptidyl- prolyl cis-trans isomerase A is acceptable and of good quality. This structure (Fig 4: see PMDB ID - PM0077224 for the corresponding coordinates in pdb format). With the development of techniques in molecular biology that allow rapid identification, isolation, and sequencing of genes, we are now able to infer the sequences of many proteins. However, it is still a time-consuming task to obtain the three-dimensional structures of these proteins. A major goal of structural biology is to predict the three-dimensional structure from the sequence, a pursuit that has not yet been realized. Thus, alternative strategies are being applied to develop models of protein structure when the constraints from X-ray diffraction or NMR are not yet available. Insights into the three-dimensional (3D) structure of a protein are of great assistance when planning experiments aimed at the understanding of protein function and during the drug design process. This model is successfully submitted in Protein Model Database (PMDB) [16] the PMDB ID of the structure is – PM0077224.

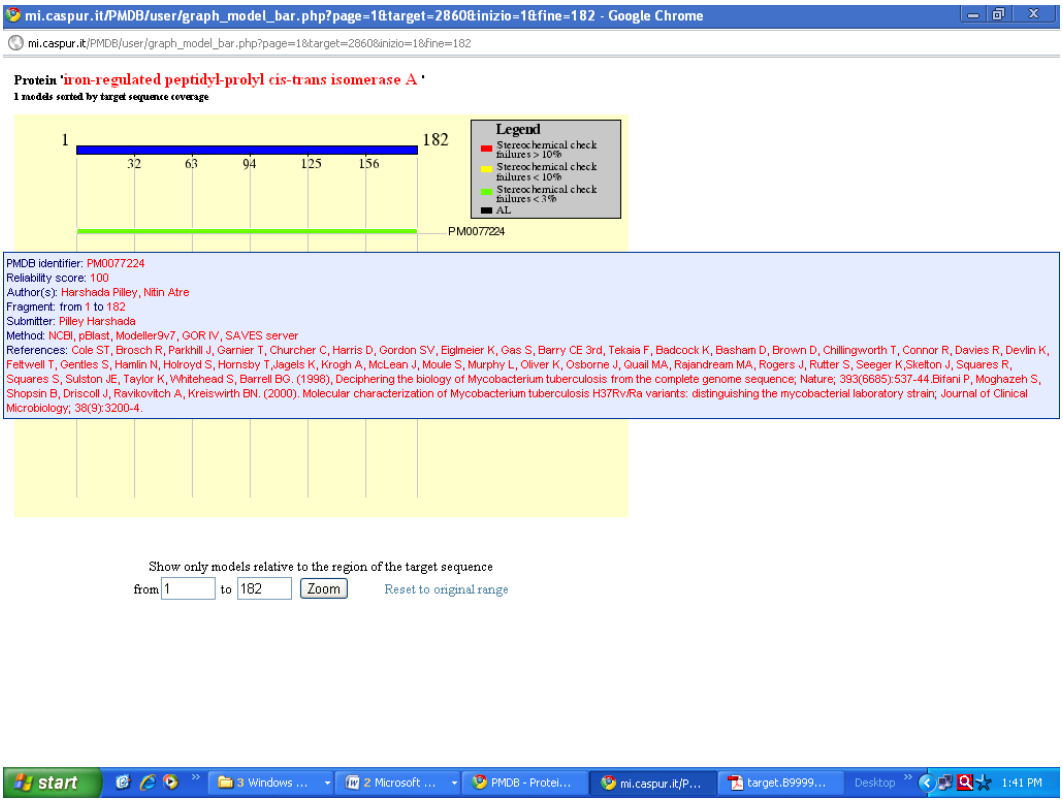

Figure 4: Predicted Structure of iron-regulated peptidyl-prolyl cis-trans isomerase A protein of *Mycobacterium tuberculosis strains H37Rv* Submitted to Protein Model DataBase.

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; *Journal of Clinical Microbiology*; 38(9):3200-4.

[2] Bowie JU, Lüthy R, Eisenberg D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. Science; 253(5016):164-70.

[3] Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T,Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K,Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG. (1998), Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence; *Nature*; 393(6685):537-44.

[4] Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor Colovos C., Yeates T.O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. 2(9):1511- 1519.

[5] Combet C, Jambon M, Deléage G, Geourjon C. (2002). Geno3D: Automatic comparative molecular modelling of protein. Bioinformatics.18:213-214.

[6] Guex N,Peitsch M.C. (1997). Swiss model and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* 18:2714- 2723.

[7] J. Garnier, J.-F. Gibrat, B. Robson; GOR secondary structure prediction method version IV; Methods in Enzymology,R.F. Doolittle Ed., vol 266, 540-553

[8] Larkin M A, Blackshields G, Brown N P, Chenna R, McGettigan P A, McWilliam H, Valentin F, Wallace I M Wilm A, Lopez R, Thompson J D, Gibson T J, Higgins D.G.(2007). ClustalW and ClustalX version 2. *Bioinformatics*. 23(21):2947-2948.

[9] Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. 26, 283-291.

[10] Martí-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, Sali A. (2000). Comparative protein structure modeling of genes and genomes. Annu Rev. Biophys Biomol Struct. 29:291- 325.

[11] McGuffin, L. J. and Jones, D. T. (2003). Improvement of the GenTHREADER method for genomic fold recognition. Bioinformatics 19:874-881.

[12] Pruitt KD, Tatusova T, Klimke W, Maglott DR; NCBI Reference Sequences: current status, policy and new initiatives.; Nucleic Acids Res. 2009 Jan;37(Database issue): D32-6

[13] R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh Sali A, Blundell T.L. (1993). Comparative protein modeling by satisfaction of spatial restraints. J Mol Biol. 234(3):779-815.

[14] Sali A, Blundell T.L. (1993). Comparative protein modeling by satisfaction of spatial restraints. J Mol Biol. 234(3):779-815.

[15] Sanchez R, Sali A (1997): Evaluation of comparative protein structure modeling by MODELLER-3. *Proteins* (Suppl. 1):50–8.

[16] Tiziana Castrignano`, Paolo D'Onorio De Meo, Domenico Cozzetto,Ivano Giuseppe Talamo, Anna Tramontano. (2006). The PMDB Protein Model Database. Nucleic Acids Research. 34: D306–D309.