COMPARATIVE MODELING OF GTP CYCLOHYDROLASE I IN *TOXOPLASMA GONDII*

Budhayash Gautam* **and Satendra Singh**
Department of Computational Biology and Bioinformatics, Sam Higginbottom Institute of Agricultural, Technology and Sciences, Allahabad 211007, U.P., India
***Email:** budhayashgautam@gmail.com

ABSTRACT

The completion of genome sequence of many species marks a major milestone in molecular biology. Currently, the genome sequence of more than twenty-five organisms has been completed and dozens more are expected to be completed in next couple of years. The availability of these sequences presents opportunities to expand our understanding of biological processes at molecular level. Efforts to study the functions of newly sequenced genes are already underway. One of the most wide used methods in initial steps of function identification of genes is sequence comparison of gene products. By finding homology between a new gene and a gene of known function, one can draw some inferences about the function of the new gene. A better understanding of a protein function can be learned from its 3D structure. The function of a protein is highly correlated with its structure. In this regard, a major effort is underway to develop strategies of high throughput structure determination of proteins. Such structure genomics initiative proposes to determine the structures of proteins in an entire organism. So far, none of these projects has produced large numbers of new structures, as the difficulties of protein expression, purification, and crystallization still remain unsolved. In this study, GTP cyclohydrolase I, (GTP 7, 8-8, 9-dihydrolase) of *Toxoplasma gondii* (which is an enzyme that is a part of the folate and biopterin biosynthesis pathways and is responsible for the hydrolysis of guanosine triphosphate (GTP) to form 7, 8-dihydroneopterin 3'-triphosphate) has been modeled. The best three dimensional model is having most of the residues in favorable region on the ramachandran plot and also having lowest minimum energy. This model could be used in future as a potential drug target by the design of suitable inhibitors for the treatment of toxoplasmosis disease.

Keywords : *Toxoplasma gondii*, Toxoplasmosis, Modeller, Homology Modelling.

INTRODUCTION

Protein three-dimensional (3D) structure determines protein function. The hypothesis that structure is uniquely determined by the specificity of the sequence has been verified for many proteins. While it is now known that particular proteins (chaperones) often play a role in the folding pathway, and in correcting misfolds, it is still generally assumed that the final structure is at the free-energy minimum. Thus, all information about the native structure of a protein is coded in the amino acid sequence, plus its native solution environment. In principle, the code could by deciphered from physicochemical principles using, for example, molecular dynamics methods. In practice, however, such approaches are frustrated by two principal obstacles. First, energy differences between native and unfolded proteins are extremely small (order of 1 kcal/mol.). Second, the high complexity (i.e., cooperativity) of protein folding requires several orders of magnitudes more computing time than we anticipate having over the next decades[1]. Thus, the inaccuracy in experimentally determining the basic parameters and the limited computing resources become fatal for predicting protein structure from first principles. The only successful structure prediction tools are knowledge-based, using a combination of statistical theory and empirical rules[2].

Protein structure prediction is the prediction of the three-dimensional structure of a protein from its amino acid sequence that is, the prediction of a protein's tertiary structure from its primary structure. It is one of the most important goals pursued by Bioinformatics and Theoretical Chemistry. Protein structure prediction is of high importance in medicine (for example, in drug design) and biotechnology (for example, in the design of novel enzymes).The practical role of protein structure prediction is now more important than ever. Massive amounts of protein sequence data are produced by modern large-scale DNA sequencing efforts such as the Human Genome Project. Despite community-wide efforts in structural genomics, the output of experimentally determined protein structure typically by time-consuming and relatively expensive X-ray crystallography or NMR spectroscopy is lagging far behind the output of protein sequences. A number of factors exists that make protein structure prediction a very difficult task. The two main problems are that the number of possible protein structures is extremely large, and that the physical basis of protein structural stability is not fully understood. As a result, any protein structure prediction method needs a way to explore the space of possible structures efficiently (a search strategy), and a way to identify the most plausible structure (an energy function)[3].

In comparative structure prediction (also called homology modeling), the search space is pruned by the assumption that the protein in question adopts a structure that is reasonably close to the structure of at least one known protein. In de novo or ab initio structure prediction, no such assumption is made, which results in a much harder search problem. In both cases, an energy function is needed to recognize the native structure, and to guide the search for the native structure[1-2].

Toxoplasma gondii is a species of parasitic protozoa in the genus Toxoplasma. *T. gondi* cause Toxoplasmosis, which is usually minor and self limiting but can have serious or even fatal effects on a fetus whose mother first contracts the disease during pregnancy or on an immunocompromised human or cat. *T. gondii* was first described in 1908 in Tunis by Nicolle and Manceaux within the tissues of the gundi (Ctenodactylus gundi) [4,5].

In the present work structure prediction of GTP cyclohydrolase I (GTP 7, 8-8, 9-dihydrolase) will be done with the help of in silico approaches. GTP cyclohydrolase I (GTP 7, 8-8, 9-dihydrolase) is an enzyme that is part of the folate and biopterin biosynthesis pathways. It is responsible for the hydrolysis of guanosine triphosphate (GTP) to form 7, 8-dihydroneopterin 3'-triphosphate. This gene encodes a member of the GTP cyclohydrolase family. The encoded protein is the first and rate-limiting enzyme in tetrahydrobiopterin (BH4) biosynthesis, catalyzing the conversion of GTP into 7, 8-dihydroneopterin triphosphate. BH4 is an essential cofactor required by aromatic amino acid hydroxylases as well as nitric oxide synthases. Mutations in this gene are associated with malignant hyperphenylalaninemia and dopa-responsive dystonia[6].

METHODOLOGY

Sequence Retrieval

Amino acid sequences of GTP cyclohydrolase I of Toxoplasma gondii is retrieved from swissprot/ uniprot, **(**http://www.uniprot.org)**.**

Template Identification

Structurally homologous subsets of the experimentally determined 3D structure of the GTP cyclohydrolase I were retrieved from PDB database, by running pairwise sequence similarity search using BLAST[7] against PDB entries under default parameters.

Comparative Modeling

Tertiary structure of a protein is build by packing of its secondary structure elements to form discrete domains or autonomous folding units. Comparative modeling to build 3D structure of GTP cyclohydrolase I was generated based on the experimentally solved structural homologues using by Geno3d server[8] and Swiss-PDB viewer[9].

Structure Validation PROCHECK and Verify_3D

The hypothetical protein models generated were analyzed online by submitting to NIH MBI Laboratory for Structural Genomics and Proteomics' SAVES server. Validity reports generated by PROCHECK[10] and Verify_3D[11] judged accuracy of the protein models.

DISCUSSION

Amino acid sequences retrieved from swissprot/uniprot, provided descriptions of a non-redundant set of proteins including their function, domain structure, post-translational modifications and variants. GTP cyclohydrolase I sequence from *Toxoplasma gondii* is retrieved as query sequence with total length of 262 amino acids and the Accession No is ABB53417 from http://www.uniprot.org. This database merges all proteins in single entry coded by one gene so as to minimize redundancy and improve reliability with fully featured information.

Figure 1: Predicted 3-Dimensional Structural model of GTP cyclohydrolase I (Model 2)

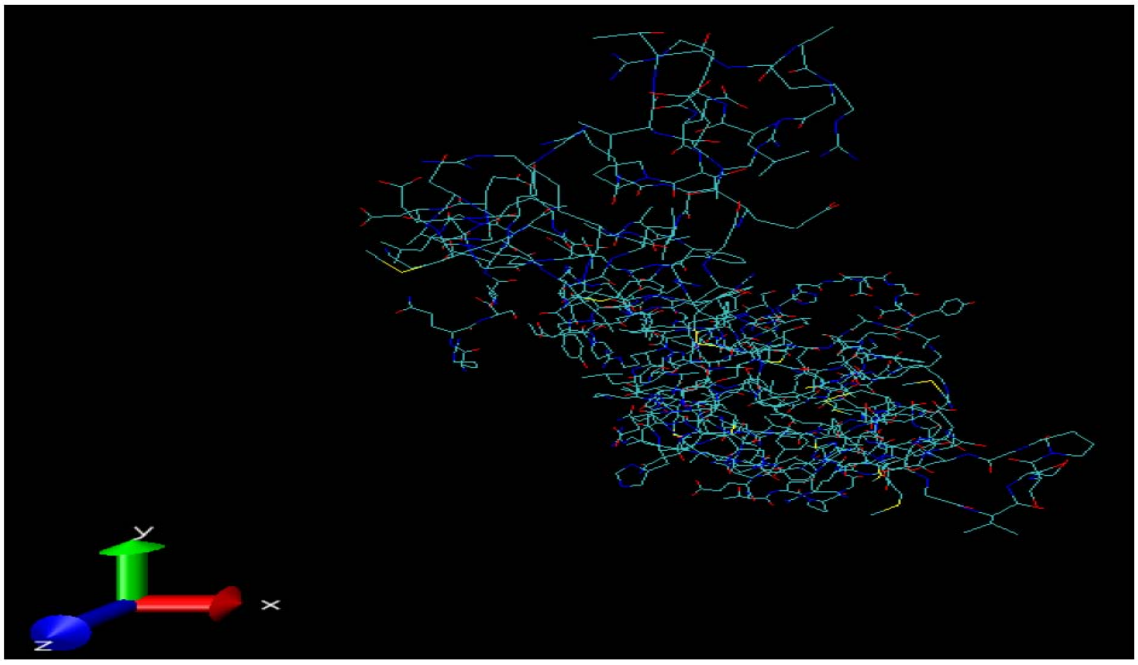

Figure 2: Ramachandran Plot of the predicted 3-Dimensional Structural model of GTP cyclohydrolase I. It can be seen that only one Arg residue is present in the disallowed region of the plot and most of the residues are present in the allowed region of the ramachandran plot and few of them are generously allowed in the ramachandran plot.

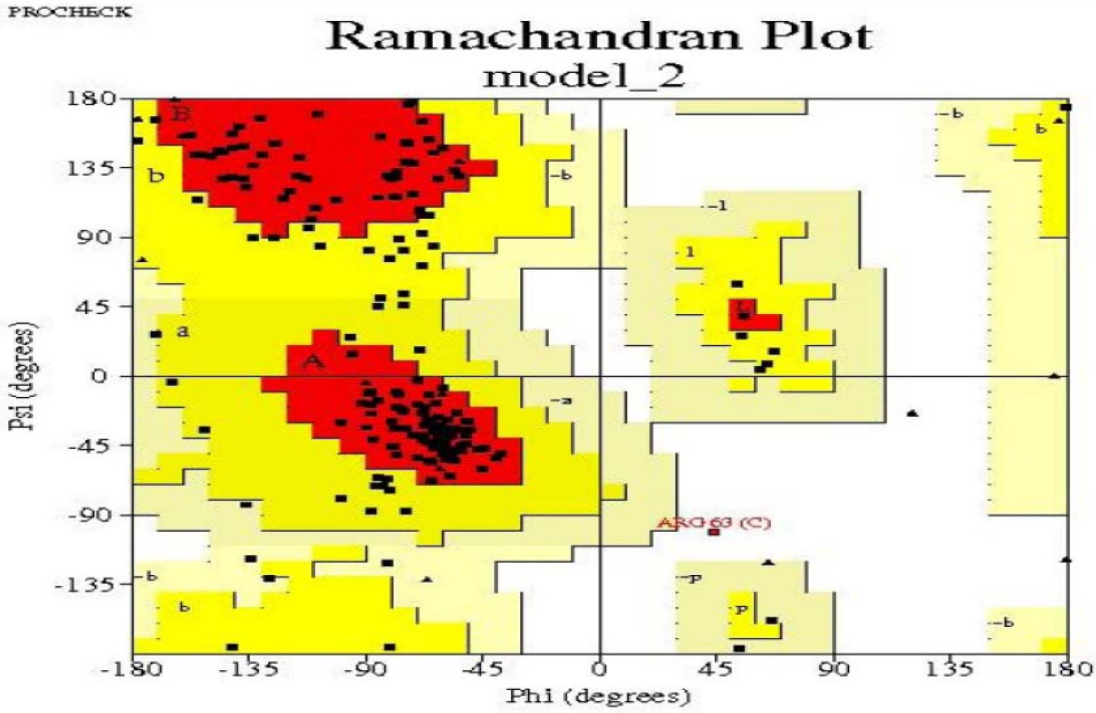

Figure 3: Ramachandran Plot for all residues types in the predicted 3-Dimensional Structural model of GTP cyclohydrolase I. It can be seen that only one Arg residue is present in the disallowed region of the plot.

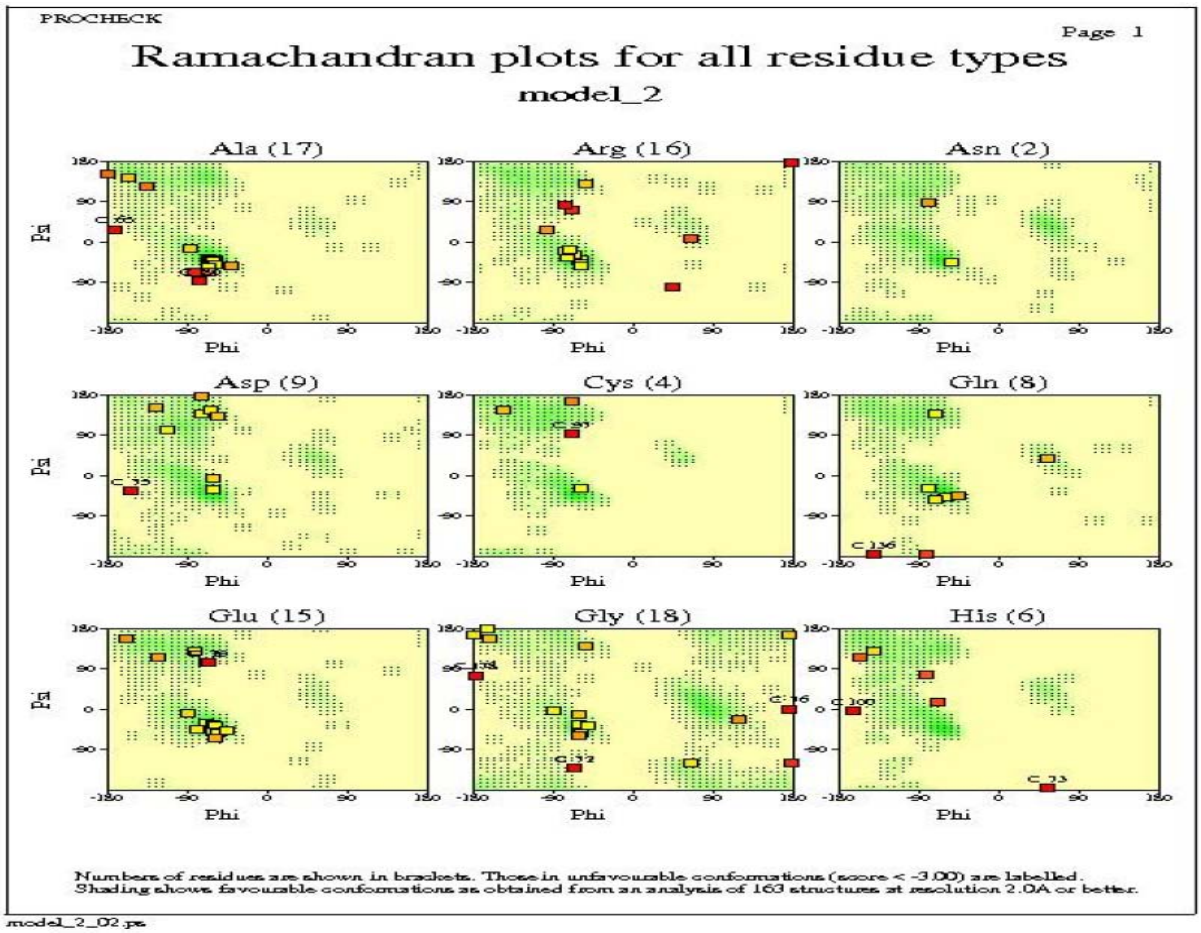

Table 1: Ramachandran plot statistics of GTP cyclohydrolase I computed by PROCHECK.

Residues in quadrangular regions	Scattered residues	
	Number	Percentage
Most favored regions (A,B,L)	131	73.6%
Additional allowed regions (a,b,l,p)	39	21.9%
Generously allowed regions (~a,~b,~l,~p)	7	3.9%
Disallowed regions (XX)	1	0.6%
Non-glycine and non-proline residues	178	100%

Sequence similarity search using BLAST against PDB database resulted in several hits with the maximum sequence identity of 49.7%. Three templates were selected pdb1fb1A_0, pdb1is7A_0 and pdb1wm9A_0 with percentage sequence identity with the target sequence, 49%, 49.7% and 48.6% over the length of 262 residues respectively.

Comparative modeling to build 3D structure of **GTP cyclohydrolase I** was generated based on the experimentally solved structural homologues using by Geno3d server and Swiss-PDB viewer. Three models were predicted,which were named as model 1, model 2 and model 3, with model energies -8578.62 kcal/mol., -8791.01 kcal/mol. and -8621.45 kcal/mol. Respectively.

A comparison of the results obtained from PROCHECK and Verify_3D validation tools, showed that one of the models (i. e. model two, **Figure 1**) generated by Geno3d server and Swiss-PDB viewer is more acceptable in comparison to the others, as it contains less number of residues in disallowed regions of ramachandran plot. The PROCHECK parameters plotted for the main chain are Ramachandran plot quality, peptide bond planarity, Bad non-bonded interactions, main chain hydrogen bond energy, C-alpha chirality's and over-all G factor. Procheck validation reports for model 2 are shown in **Figure 2.** Ramachandran plot statistics for model two is shown in **Table 1.** Ramachandran plot for all redidue types in the predicted model is also shown in **Figure 3**.

CONCLUSION

Development of drug resistance in *Toxoplasma gondii* can be better augmented when structure is available from the present strain. Enzyme structure modeling undertaken in present work could be a tool to study better structural characteristics of GTP cyclohydrolase I for drug design community. Functional part from structure after recognition can be set in structure based drug design codes. Modeling studies manifested good stereochemical placement of main chain parameters. Bond angles and bond lengths are under confined limits although side chain modeling introduced some levels of displacement of residues beyond most favored regions. Catalytic site in enzyme structure can be examined after screening catalytic databases available. More efforts in structural analysis in concern with mutational studies can provide better insight towards development of drug resistance profiles of this organism. Thus present studies of modeling of GTP cyclohydrolase I have brought future prospective to fight against toxoplasmosis and provide better health standards for community.

Acknowledgment: The authors would like to acknowledge the support and facilities provided by the Department of Computational Biology and Bioinformatics, Sam Higginbottom Institute of Agricultural, Technology and Sciences, Allahabad, U.P., India.

REFERENCES

1. P. R. Graves & T. A. Haystead, *Microbiol Mol Biol Rev.* (2002) 66:39-63 [PMID: 11875127].
2. B. S. Cherian & A. S. Nair, *Biochem Biophys Res Commun.* (2010) 22;1670-1674 [PMID: 20036215]
3. D. Petrey *et al., Proteins* (2003) 53:430-435. [PMID: 14579332]
4. Francart Aurelie *et al.* , Vet Parasitol (2008) 157:128-132 [PMID: 18707811]
5. Y. Sukthana ,*TRENDS in Parisitology* (2006) 22 :137 [PMID: 16446116]
6. R. Ernst et al., J Biol Chem. (2002) 22:10129-10133 [PMID: 11799107]
7. Stephen Altschul *et al., J. Mol. Biol.* (1990) 215:403-410 [PMID: 2231712]
8. C. Combet *et al., Bioinformatics* (2002) 18: 213-214 [PMID: 11836238]
9. N. Guex & M. C. Peitsch *Electrophoresis* (1997) 18: 2714-2723. [PMID: 9504803]
10. R. Laskowski *et al., J. Appl. Cryst. (1993)* 26:283-291
11. D. Eisenberg *et al., Methods. Enzymol.* (1997) 277:396-404 [PMID: 9379925]

****