International Journal of Drug Discovery, ISSN: 0975–4423, Volume 2, Issue 2, 2010, pp-11-16

Comparative modeling of CDP-diacylglycerol-serine O-phosphatidyl-transferase in *Clostridium botulinum:* A potent target of Botulism

Koteswara Reddy G.[1*], Mohan Kalyan Reddy K.[2], Nagamalleswara Rao K.[3] and Gyana R.Satpathy[4]
[1*]Department of Biotechnology, Bapatla Engineering College, Bapatla-522 101, A.P, India, kotireddy.nit@gmail.com, Ph: +91 9908594414
[2]Department of Biotechnology, Bapatla Engineering College, Bapatla-522 101, A.P, India
[3]Department of Chemical Engineering, Bapatla Engineering College, Bapatla-522 101, A.P, India
[4]Department of Biotechnology & Medical Engineering, National Institute of Technology, Rourkela-769008, Orissa, India, gyansatpathy@gmail.com , Ph: +919437579091

Abstract –The bacteria *Clostridium botulinum* causing botulism. Botulism also known as botulinus intoxication is a rare but serious paralytic illness caused by botulinum toxin, which is produced by the bacterium *Clostridium botulinum.CDP-diacylglycerol-serine O-phosphatidyltransferase* is an enzyme that catalyzes the *CDP-diacylglycerol, L-serine* and produce *Cytidine monophosphate (CMP), (3-sn-phosphatidyl)-L-serine* in *Clostridium botulinum.* CMP is a nucleotide that is helpful in RNA synthesis in *Clostridium botulinum*.The objective of enzyme inhibition needs structural characterization from its 3-D structure in rational structure based drug design. Homology modeling can produce high-quality structural models when the target and templates are closely related. MODELLER is the program used for homology modeling which provides accurate and efficient models to build loops and side chains found non-identical in sequence. It implements comparative protein structure modeling by satisfaction of spatial restraints. The stereo chemical quality of a best model is validated by PROCHECK server with 86.2% residues under favored region from Ramachandran plot. The CASTp is used for locating, delineating and measuring concave surface regions on three-dimensional structures of proteins. The residues are identified in burial cavity of the best model as *Leu (131,141,156), Ala (132), Ile (138,142), Asn (145), Thr (149,153), and Gly (151)* with hydrophobic nature. These include pocket located on protein surfaces and voids buried in the interior of proteins that are frequently associated with binding events. Thus present studies of modeling of CDP-diacylglycerol-serine O-phosphatidyltransferas enzyme has brought future prospective to fight against botulism disease and provide better health standaaards for community.

Key Words - *Clostridium botulinum F strain*, comparative modeling, homology modeling, Ramachandran plot, Drug design.

Background

The bacteria *Clostridium botulinum* causing botulism. Botulism also known as botulinus intoxication is a rare but serious paralytic illness caused by botulinum toxin, which is produced by the bacterium *Clostridium botulinum* [1]. Botulism (Latin, *botulus*, "sausage") also known as botulinus intoxication is a rare but serious paralytic illness caused by botulinum toxin, which is produced by the bacterium *Clostridium botulinum* [2]. The toxin enters the body in one of four ways: by colonization of the digestive tract by the bacterium in children (infant botulism) or adults (adult intestinal toxemia), by ingestion of toxin from foodstuffs (food borne botulism) or by contamination of a wound by the bacterium (wound botulism).All forms lead to paralysis that typically starts with the muscles of the face and then spreads towards the limbs. In severe forms, it leads to paralysis of the breathing muscles and causes respiratory failure. In view of this life-threatening complication, all suspected cases of botulism are treated as medical emergencies, and public health officials are usually involved to prevent further cases from the same source [3]. *Clostridium botulinum* would normally be harmless to humans, but it can become infected by a virus. The viral DNA, integrated into the bacterial genome, causes the host to produce toxins [4]. Neurotoxin production is the unifying feature of the species *C. botulinum*. Seven types of toxins have been identified and allocated a letter (A-G). Most strains produce one type of neurotoxin but strains producing multiple toxins have been described. *Clostridium botulinum* producing B and F toxin types have been isolated from human botulism cases in New Mexico and California [5].

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*"). 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 [6]. Homology modeling is usually the method of choice when a clear relationship of homology between the sequence of target protein and query protein sequence at least one known structure is found. This approach would give reasonable results based on the assumption that the tertiary structures of two proteins will be similar if their sequences are related and three-dimensional structure of proteins is better conserved during evolution than its sequence [7].

The quality of the homology model is dependent on the quality of the sequence alignment and template structure. The approach can be complicated by the presence of alignment gaps (commonly called indels) that indicate a structural region present in the target but not in the template, and by structure gaps in the template that arise from poor resolution in the experimental procedure (usually X-ray crystallography) used to solve the structure. Model quality declines with decreasing sequence identity; a typical model has ~1-2 Å root mean square deviation between the matched C^{α} atoms at 70% sequence identity but only 2-4 Å agreement at 25% sequence identity. However, the errors are significantly higher in the loop regions, where the amino acid sequences of the target and template proteins may be completely different [8]. Homology modeling can produce high-quality structural models when the target and template are closely related, which has inspired the formation of a structural genomics consortium dedicated to the production of representative experimental structures for all classes of protein folds[9]. The chief inaccuracies in homology modeling, which worsen with lower sequence identity, derive from errors in the initial sequence alignment and from improper template selection [10]. The stereo chemical quality of a protein structure is validated by ramachandran plot.

Materials and methods:

Sequence retrieval

CDP-diacylglycerol-serine O-phosphatidyltransferase amino acid sequence retrieved from swissprot/uniprot. It is retrieved as query sequence with a total length of 173 amino acids and molecular weight of 18940.the database accession number is A7G9V1. Swiss-Prot is a manually curated biological database of protein sequences. Swiss-Prot was created in 1986 by Amos Bairoch during his PhD and developed by the Swiss Institute of Bioinformatics and the European Bioinformatics Institute[11,12].Swiss-Prot strives to provide reliable protein sequences associated with a high level of annotation (such as the description of the function of a protein, its domains structure, post-translational modifications, variants, etc.), a minimal level of redundancy and high level of integration with other databases (http://us.expasy.org/sprot).

Comparative modeling

Building a homology model comprises four main steps: *identification of structural template(s*), *alignment of target sequence* and *template structure(s*), *model building*, and *model quality* evaluation. These steps can be repeated until a satisfying modeling result is achieved. The MODELLER is used for homology or comparative modeling of protein three-dimensional structure prediction [13, 14].

1. Identification of structural template(s)

The query sequence CDP-diacylglycerol-serine protein was searched to find out the related protein sequences as a template by the BLAST program against the protein data bank (PDB). Four templates were identified with PDB entry IDs 1bp6A, 1m38A, 2bptA and 2bkuA.The PDB entry ID 2bptA The sequence that showed maximum identity 29% with high score 27.7, better resolution 1.99, better crystallographic R-factor 0.174 and less E-value. The comparison in the figure.1 shows that *2bpt: A* is the best one because of its low resolution and high similarity), which is produced by the MODELLER9v7. The *2bpt: A* was aligned and used as a reference structure to build a 3D model for target protein.

2. Alignment of target sequence and template structure

Structure similarity searching was performed by standalone *blastp* search was performed for finding similar structures entry in PDB database from ftp download available on *ftp://ftp.ncbi.nih.gov/blast/db/FASTA/pdbaa.gz.results* from blastp show 29% identity and 46% similarity with above query from which is a structure of Nuclear Transport protein from Saccharomyces Cerevisiae (Baker's Yeast) with a PDB (protein data bank) entry ID is 2BPT:A [15] of template which is X-ray crystallized structure at 1.99 A^{0} and selected for backbone alignment with A chain identified in secondary structure studies determined by Stewart, M. et al available on http://www.pdb.org/pdb/explore/explore.do?structureId=2BPT. Best template is also identified by using template proteins clustering based on a distance matrix in the MODELLER 9V7 program from Fig. (1).

3. Model building

The model was constructed by using the program Modeller9v7 under Windows. Modeller is a comparative protein structure modeling software. It is based on spatial restraints derived from the alignment and Probability Density Functions (PDFs) [16].The 3D model of a protein is obtained by optimization of the molecular PDFs such that the model violates the input restraints as little as possible.

4. Model quality evaluation

After building the protein 3D structure, in order to assess the overall stereo chemical quality of the modeled protein, Ramachandran plot analysis was performed using the program PROCHECK [17,18,19].Further evaluation of modeled structure was done by VERIFY3D [20], ERRAT is a protein structure verification algorithm that is especially well-suited for evaluating the progress of crystallographic model building and refinement [21]and Prove[22].Through structure analysis and verification server (SAVS): (http://nihserver.mbi.ucla.edu/SAVS/). The GNU plot supports many types of plots in either 2D or 3D. It can draw using lines, points,

boxes, contours, vector fields, surfaces, and various associated text. The GNU plot is available at http://www.gnuplot.info/. The ProSA (Protein Structure Analysis) is a web database is used to test the local and overall quality of the developed models from MODELLER. The ProSA is available at https://prosa.services.came.sbg.ac.at/prosa.php

Active Site Identification:

Sites of activity in proteins usually lie in cavities. The size and shape of protein cavities dictates the three-dimensional geometry of ligands that must fit like a hand in glove. The binding of a substrate typically serves as a mechanism for chemical modification or conformational change of protein [23].Binding sites are often targeted by various ligands in attempts to interrupt related molecular processes. Active sites of the target protein were predicted using The CASTp, pocket finder and Q-site finder active site prediction tools. Active sites of a protein are a key factor for the flexible docking. Active sites of the target protein CDP-diacylglycerol-serine O-phosphatidyltransferase were predicted by using tool CASTp (computed atlas of surface topography of proteins) [24].

Results and discussion

The objective of enzyme inhibition needs structural characterization from its 3-D structure in rational structure based drug design. Homology modeling produced high-quality structural models when the target and templates are closely related. Enzyme structure modeling undertaken in present work could be a tool to study better structural characteristics of *CDP-diacylglycerol-serine O-phosphatidyltransferas* enzyme for drug design community. The modeled enzyme 3-D structure was shown in Fig. (2) with ribbon and hydrophobic view. Modeling studies manifested good stereo chemical 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 was examined after screening catalytic database available. More efforts in structural analysis in concern with mutational studies can provide better insight towards development of drug resistance profiles of this *Clostridium botulinum*. Thus present studies of modeling of CDP-diacylglycerol-serine O-phosphatidyltransferas enzyme has brought future prospective to fight against botulism disease and provide better health standards for community.

The modeled protein is validated with the SAVES server. The results of the PROCHECK analysis indicate that a relatively low percentage of residues have phi/psi angles in the disallowed ranges, the quality of Ramachandran plots is acceptable. The percentage of residues in the "core" region of modeled was found to be 86.2%. The stereo chemical quality of the model was found to be satisfactory. The Ramachandran plot of the modeled protein is shown in Fig. (3).

DOPE score was obtained for template and modeled protein with different file formats.GNU plot is used to draw the graph between template and modeled protein DOPE score was shown in Fig.(4).The green color and red color indicates that the template residues and modeled protein residue in the plot. The PDB codes for template chain and modeled protein was submitted to ProSA web database and local and overall quality was checked. ProSA was used to test for overall model quality and local model quality of protein CDP-diacylglycerol-serine O-phosphatidyltransferase with chain blank (173 AA) and Z-score:-1.09 was shown in Fig. (5).Errat analyzes the statistics of non-bonded interactions between different atom types and plots the value of the error function versus position of a residue. Errat is showing an overall quality factor of 55.758. Verify_3D determines the compatibility of an atomic model (3D) with its own amino acid sequence (1D) by assigned a structural class based on its location and environment (alpha, beta, loop, polar, non-polar etc) and comparing the results to good structures.

Active Site Prediction: Active sites of the target protein were predicted using The CASTp, pocket finder and Q-site finder active site prediction tools. The feasible catalytic site residues are screened based on consensus ranked among three methods (CASTp, pocket finder and Q-site finder).The residues in catalytic site are summarized as residue name and corresponding residue numbers in the 3-D CDP-diacylglycerol-serine O-phosphatidyltransferase enzyme structure. The residues are identified in burial cavity of enzyme as *Leu (131,141,156), Ala (132), Ile (138,142), Asn (145), Thr (149,153), and Gly (151)* with hydrophobic nature.

References

[1] Bengston I.A. (1924) *Hyg. Lab. Bull.,* 136:101.

[2] Suen J.C., Hathaway C.L., Steigerwalt A.G. and Brenner D. J. (1988) *Int. J. Sys. Bacteriol*. 38:375–381.

[3] Tomb J.F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G. and Fleischmann R.D. (1997) *Nature,* 388:539-47.

[4] Doyle and Michael P. (2007) *Food Microbiology: Fundamentals and Frontiers-ASM Press.*

[5] Hathaway C.L. and McCroskey L.M. (1987) *J. Clin. Microbiol.,* 25:2334–2338.

[6] Marti-Renom M.A., Stuart A.C., Fiser A., Sanchez R., Melo F. and Sali A.

(2000) *Annu Rev. Biophys. Biomol. Struct.* 29: 291-325.

[7] Kroemer R.T., Doughty S.W., Robinson A.J. and Richard W.G. (1996) *Protein Engineering,* 9, 493–498.

[8] Chung S.Y. and Subbiah S. (1996) *Structure,* 4: 1123–27.

[9] Williamson A.R. (2000) *Nat. Struc.t Biol.,* 7 S1 (11s):953.

[10] Venclovas C. and Margeleviĉius M. (2005) *Proteins,* 61(S7):99-105.

[11] Bairoch Amos S. (2000) *Bioinformatics,* **16**: 48–64.

[12] Séverine Altairac (2006) *Protéines à la Une* ISSN 1660-9824.

[13] Eswar N., Marti-Renom M. A., Webb B., Madhusudhan M. S., Eramian D., Shen M., Pieper U. and Sali U. (2006) *John Wiley & Sons, Inc., Supplement*, 15, 5.6.1-5.6.30.

[14] Marti-Renom M.A., Stuart A., Fiser A., Sánchez R., Melo F. and Sali A. (2000) *Annu. Rev. Biophys. Biomol. Struct.,* 29, 291-325.

[15] Liu S.M. and Stewart M. (2005) *J.Mol.Biol.* 349**:** 515

[16] Sali A. and Blundell T. L. (1993) *Journal of Molecular Biology*, 234, 779-815.A

[17] Fiser R.K. and Sali A. (2000*)* *Protein Science*, 9. 1753-1773.

[18] Laskowski R.A., MacArthur M.W., Moss D.S. and Thornton J.M. (1993) *Journal of Applied Crystallography*, 26, 283-291.

[19] Morris A.L., MacArthur M.W., Hutchinson E.G. and Thornton J.M. (1992) *Proteins: Structure, Function, and Bioinformatics*, 12, 345-364.

[20] Ramachandran G.N., Ramakrishnan C. and Sasisekharan V. (1963) *Journal of Molecular Biology*, 7, 95-99.

[21] Eisenberg D., Luthy R. and Bowie J. U. (1997) *Methods in Enzymology*, 277, 396-404.

[22] Colovos C. and Yeates T.O. (1993) *Protein Science*, 2, 1511-1519.

[23] Vriend G. (1990) *Journal of Molecular Graphics*, 8, 52-56.

[24] Binkowski T.A., Naghibzadeh S. and Liang J. (2003*)* *Nucleic Acids Research*, 31, 3352-3355.

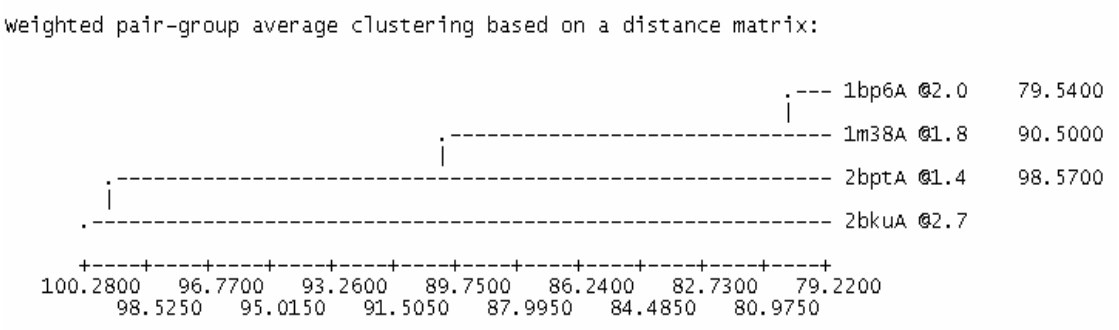

Fig. 1. -*Template proteins clustering based on a distance matrix*

Table 1- *Results of modeling target protein CDP-diacylglycerol-serine O- phosphatidyltransferase with 6 models.*

The enzyme CDP-diacylglycerol-serineOphosphatidyltransferase model 2 was top ranked based on consensus validated among four structure validation servers with 86.2% residues under favored region from Ramachandran plot.

S. No	Protein model	Procheck (Ramachandran plot %)	Verify 3D(% of the residues had an averaged 3D-ID score > 0.2)	Errat (Overall quality factor)	Molpdf score
1	CDP-diacylglycerol-serine O-phosphatidyltransferase1	85.5	9.20	47.273	1104.97
2	CDP-diacylglycerol-serine O-phosphatidyltransferase2	86.2	6.90	55.758	1145.73
3	CDP-diacylglycerol-serine O-phosphatidyltransferase3	85.5	2.87	29.878	1100.17
4	CDP-diacylglycerol-serine O-phosphatidyltransferase4	86.8	6.32	51.220	1109.73
5	CDP-diacylglycerol-serine O-phosphatidyltransferase5	85.5	7.47	57.317	1101.37
6	CDP-diacylglycerol-serine O-phosphatidyltransferase6	86.2	4.02	50.610	1025.87

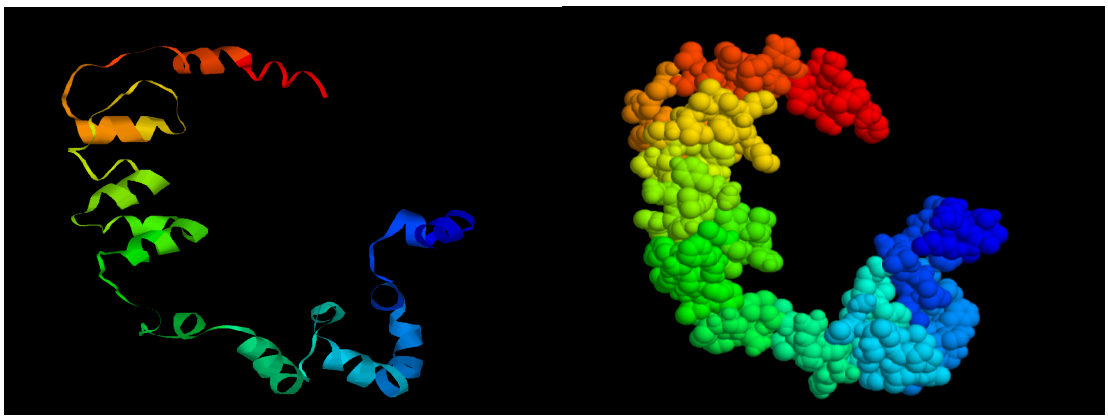

Fig. 2.- Predicted 3-D structure of protein CDP-diacylglycerol-serine O-phosphatidyltransferase

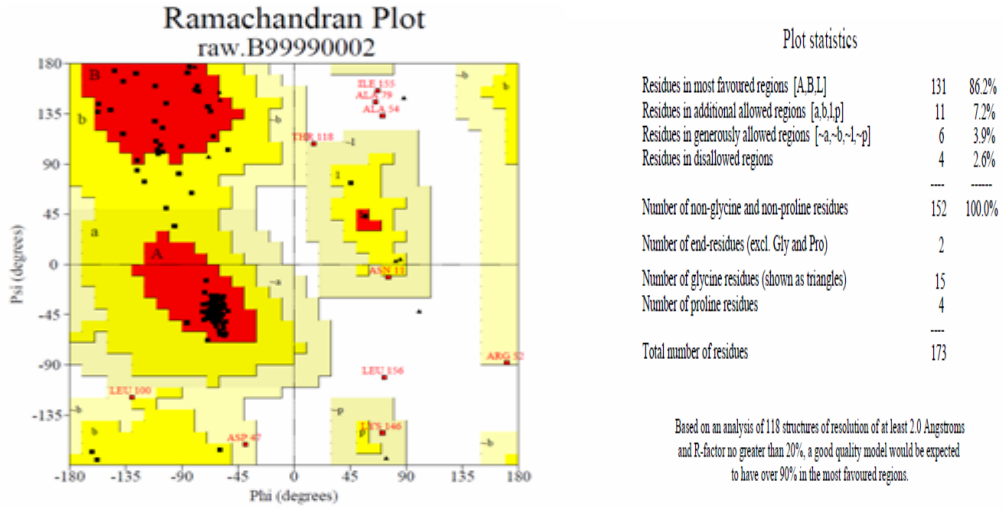

Fig. 3- Ramachandran plot of protein CDP-diacylglycerol-serine O-phosphatidyltransferase from PROCHECK. Most favored regions are colored red, additional allowed, generously allowed and disallowed regions are indicated as yellow, light yellow and white fields, respectively.

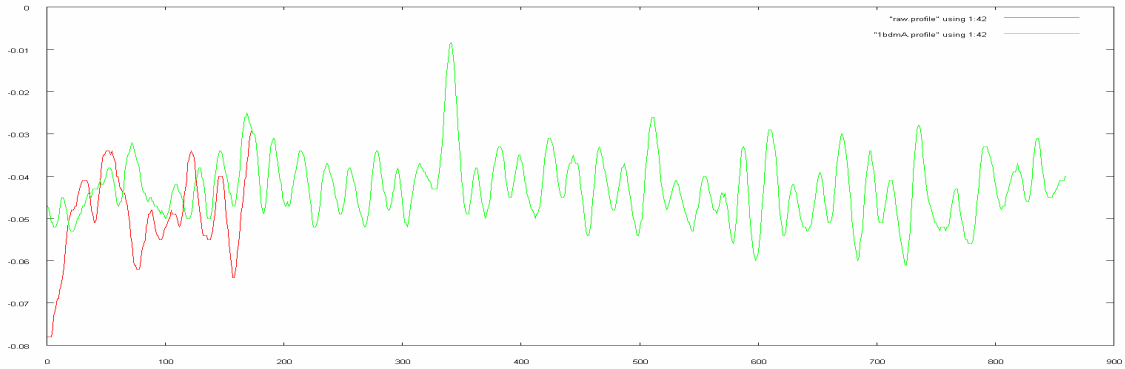

Fig. 4- DOPE score profile for single template model (raw) and template 2bptA.

Overall model quality **Local model quality**

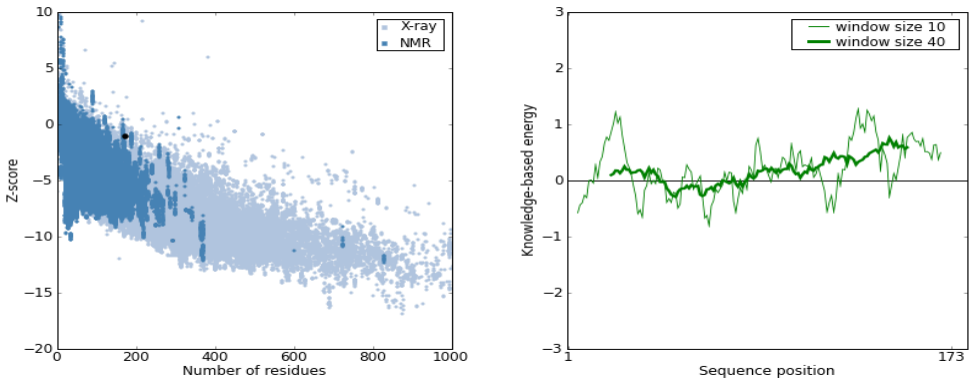

Fig. 5- Plots for overall model quantity and local model quality of protein CDP-diacylglycerol-serine O-phosphatidyltransferase with chain blank (173 AA) and Z-score:-1.09 from ProSA (Protein Structure Analysis).