International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No.02, February 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

Computational Comparative Homology based 3D-Structure Modeling of the cytochrome c oxidase subunit III

Sonu Mishra[1], Virendra S. Gomase[2]

[1,2] *Department of Biotechnology, MewarUniversity, Chittorgarh, (India)*

ABSTRACT

Homology modeling techniques is the computational based modelling approach which allows to model the 3Dstructure of the protein by use of the experimentally determined three dimensional structure of related homologues protein as template. In the current investigation, we have taken cytochrome c oxidase subunit III protein from Dracunculus medinensis and performed the 3-imensional structure prediction through Swiss-Model by using the crystal structure as the template. Template search of the desired protein were conducted with Blast against the SMTL and identified the highest sequence identity, comparatively, above the other obtained template sequences. The homology modelling was performed and the model protein was evaluated by using the PSVS(The protein structure validation software suite).

Keyword: ***cytochrome c oxidase subunit III (mitochondrion) [Dracunculus medinensis], Homology Modeling, Protein structure, Molecular modeling***

I. INTRODUCTION

The computational based approach is the one of the reliable methodology to generate amino acid sequence into 3-D structure models [1] and routinely a wide range of the approaches were applied for such prediction for many biological applications. Homology modeling is based on the sensible assumption of two homologous proteins shares very standardized alike structures. The name that is homology modelling itself conveys exactly what this procedure is all about; modeling a structure using homologous model as template(which is usually an exact X-ray or NMR-determined structure).In homology modeling it is important that modeller finds a template structure with the highest possible sequence-identity. If the identity between the input sequence and the template structure falls below 40%, the output model is likely to be implausible. Earlier study reveals that about the conservation of the protein structures than protein sequences in the midst of homologues [2]. In this investigation we have taken the cytochrome c oxidase III for homology modelling through the PSVS(The protein structure validation software suite) .this protein from GWD (Guinea worm disease) consists of 255 amino acid residues. The previous study suggest that this is the only species of *Dracunculus* genus which has been found so far which infects human [3-6] , and widely known as "Guinea worm disease (GWD)". Whereas, the oother species from this genus found as to resides in the internal

tissues and body cavities of non-human mammals and reptiles (snake and turtles) [7].This parasite well known for its unusual life cycle .The incubation period last for about one an half years approximately with six developmental stages . This is also counted as the one of the most neglected tropical parasites which bears clinical importance and needs to be eradicated after small pox [8].After maturation stage, these worms copulate and an adult female produces millions of eggs in its uterus whereas mail dies after the copulation process .Later on, the female worm release the larvae which induces a painful blister (1 to 6cm diameter) on the skin of lower limbs (predominantly localized in the lower extremities(80-90%) in most of the reported cases).The infected person develops slight fever , local skin-redness , swelling and severe pruritus around the blister . Other symptoms include: diarrhea, nausea, vomiting and dizziness. The blister burst within three days and female worms one or more slowly comes out from the wounds which causes an excoriating burning sensation and pain [9].Immersing or pouring water over the blister provides pain relief. But this the moment that adult female is exposed to the external environment [10]. During emergence of the limbs in open water sources it recognizes the temperature difference and releases the milky white liquid in the water which contains millions of immature larvae, when larvae released in water are ingested by copepods where they mount twice and become infective larvae within two weeks [11].The protein sequence of cytochrome c oxidase subunit III were obtained from the NCBI database and and 3-D structure prediction and modelling were performed This protein is one of main transmembrane subunits of cytochrome c oxidase. MTCO3(mitochondrial Cytochrome c oxidase subunit III) is encoded by the guanine-rich heavy (H) strand of the mtDNA which is located between and 9990 nucleotide pairs (nps) 9207 [12]. Giles et al., and Case and Wallace suggest that, it is maternally inherited along with the mtDNA [13]. Mutations in mtDNA-encoded cytochrome c oxidase subunit genes have been observed to be associated with isolated myopathy or severe encephalomyopathy[14].

II. METHODOLOGY

1. Retrieval of the protein sequence

For the present study cytochrome c oxidase subunit III (mitochondrion) [Dracunculus medinensis] the protein sequence was retrieved from the NCBI having 255 aa amino acids sequence Accession: YP_004857905.1 GI: 347600384[Table1][15].

MKHNYHLLSYSGYPFMVFCSVMGLSSSLVIFLKYGVIFGVFFGVFCLFCVVMVWCKDIFMEGLSGYHNFF VMNGFKYGMVFFIFSEFMFFFGVFWVFFDSSLVPNSELGMSWCPLGIGLINPLGVPLLNTLILLSSAVTV TWCHNSMLCNYNSFYGLFFTCVLALFFLVFQMLEYDESGFSMSDGIYGSIFYLSTGFHGMHVFFGMIFLF VNLFRLYMDHFNSDHHLGLEFSIVYWHFVDLIWLFLFVFVYWWSF

Table-1: Primary amino acid sequences of target protein for which template were searched for models building.

2. Template search and selection

For the target protein, the templates weresearch and identified by performing BLAST and HHBlits against the SWISS-MODEL template library (SMTL) [16,17].A total of 96 templates were opted .For each identified template, the quality of the template'swas predicted from features of the target-template alignment.Among of them; 1qle.1chain c with highest sequence identity (Sequence identity: 88.45), and lowest two sequence similarity found are 1qle.1.C chain c with sequence identity: 43.28value had then selected as template for building model.

3. Molecular Modelingof Protein

The methodological strategy used in the model building was based on the target –template alignment using Promote-II. The template were selected based on the maximum sequence similarity .The 1qle.1.C found to have the maximum sequence similarity and taken as template and alignment were performed between the both the targets and template. The coordinates which are found as conserved between the targets and templates are copied from the template to the model. Using a fragment library , insertion and deletion are remodeled and the rebuilding of the side chains were performed. Finally, the geometry of the resulting model is regularized by using a force field. The obtained model is visualized with DeepViewer.The global and per-residue model quality has been assessed using the QMEAN4 scoring function [18].

4. Ligand modeling

The number of the ligands which were found to be present in the template structure (Ligands such as SO4 ,BEF-ADP,ADP) are reassigned by homology in order of modelling, where the ligands shouldn't be clashing with the target protein and the residue which are in contact with the ligand were conserved among the target and the template.

5. The protein model Statistical Assessment

The protein model statistical Assessment were performed via the protein structure validation software suite (PSVS).PSVS were used for those protein structures which are generated from NMR, X-Ray crystallographic and homology modeling methods.PSVS incorporates analyses from numerous widely-used structure quality evaluation tools, including RPF,PROCHECK,MolProbity, Verified 3D,Prosa II, the PDB validation software and various structure validation tools [19].PSVS provides a standard constraint analyses, statistics on the PDB validation goodness-of-fit between structures and experimental data, Z-score values and knowledge- based structure quality scores in a standardized format suitable for database integration.

6. Visualization of 3D Model

The generated model was visualizedin 3D using the RasMOl molecular 3D viewer. The RasMol generated model information regarding chain, atoms, groups and bonds of 3D model were extracted and analysed [20].

International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No.02, February 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

III. RESULTS AND INTERPRETATION

For the recognized 3-D protein structure model, the complete assessment were performed , analyzed and described into the following leads:

1. Template Validation

For each indentified template, the template's quality has been predicted from the features of the target-template alignment as shown in table-2.The templates with the highest quality have been selected for the model building. Template 1qle.1.C validations for the generated model of cytochrome c oxidase subunit III protein was analyzed and were employed structure analysis for template chain. The multiple sequence alignment result of target protein sequence and the template 1qle.1 chain c is found to be 43.28% identity by BLASTas shown in the table-3.

Template	Seq Identity	Oligo-State	Found by	Method	Resolution	Sequence similarity	Range	Coverage	Description
1qle.1.C	43.28	Hetero-oligomer	BLAST	X-RAY DIFFRACTION	3.00 Å	0.42	54-254	0.79	CYTOCHROME C OXIDASE POLYPEPTIDE III

Table-2: 1qle.1 chain C with highest sequence similarity choosen as the template for the protein modeling.

Table3: Cytochrome c oxidase subunit III protein sequence and template 1qle.1chain C Multiple sequence alignment result with 43.28% of identity obtained through BLAST.

2. Model Analysis

Cytochrome c oxidase subunit III protein model structure was prepared for the target-template alignment, because the template 1qle.1.C has the highest quality and alignment with the target [table-4], therefore selected for modeling as shown in Figure-1 and Figure-2: is the secondary structure plot for 1qle (C).

International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No.02, February 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

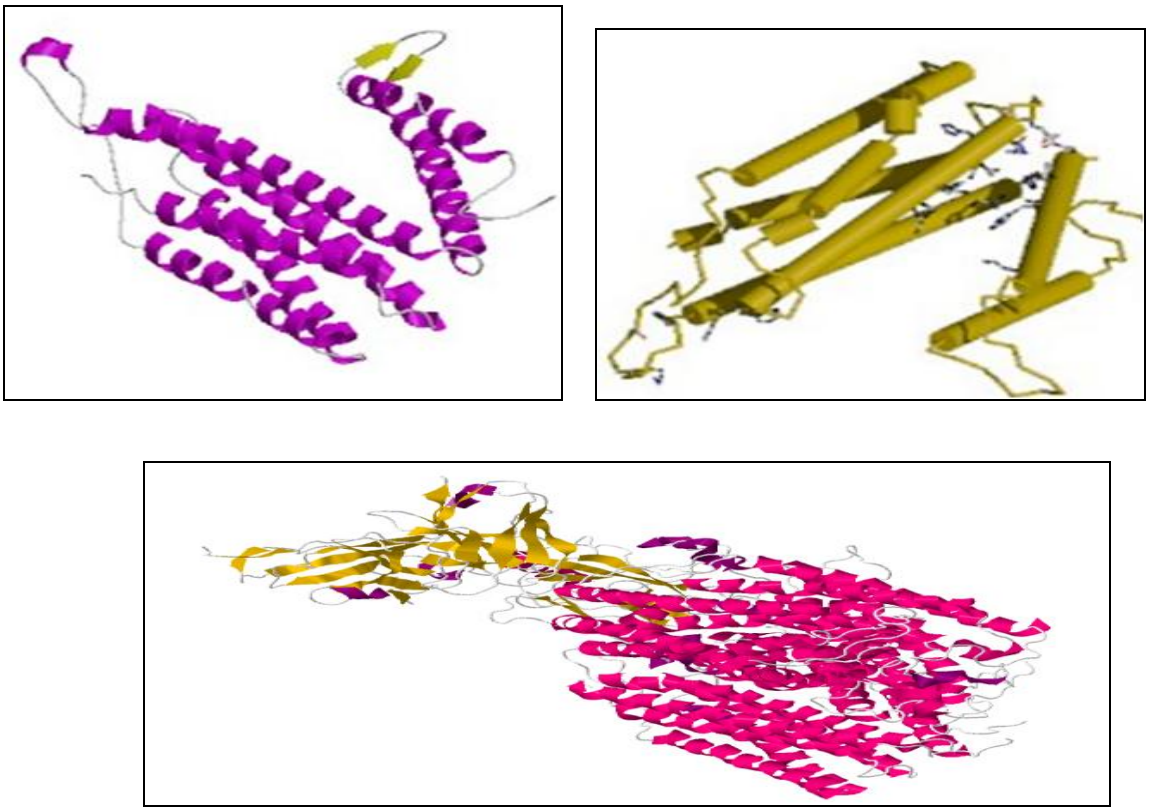
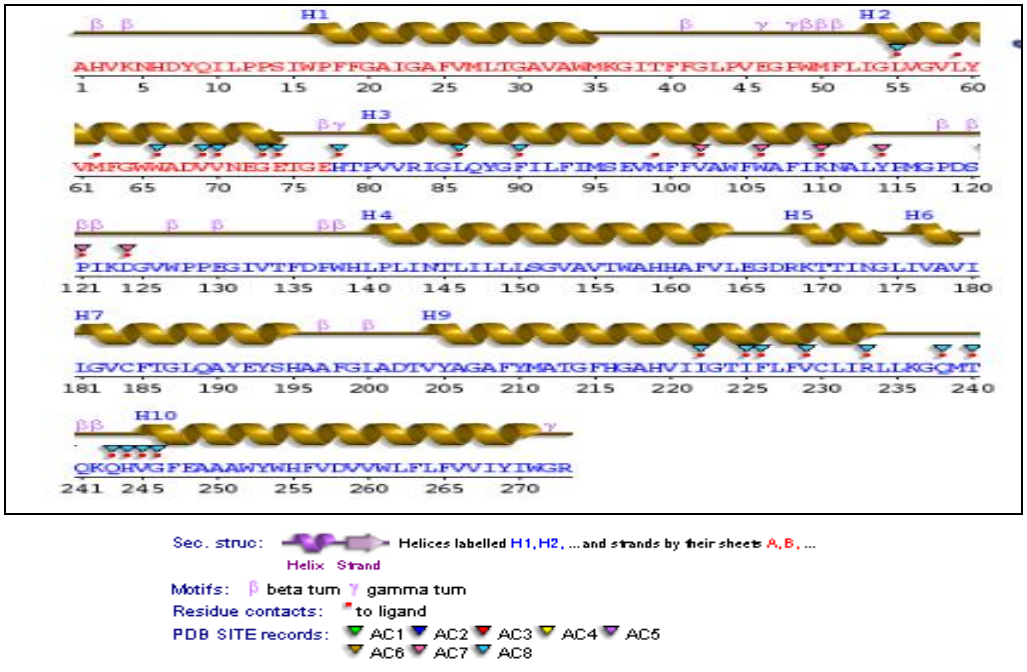


Figure-1: The template 1qle.1.C **chain CPDB structure in ribbon form.(Chain B-coloured as 'Red' and chain A as 'Purple')**



Sec. struc: Helices labelled H1, H2, ... and strands by their sheets A, B, ...
Helix Strand
Motifs: β beta turn γ gamma turn
Residue contacts: to ligand
PDB SITE records: AC1 AC2 AC3 AC4 AC5 AC6 AC7 AC8

Figure-2: secondary structure plot for 1qle (C)

The template model important information are showed in table-4

Model Properties	Data
Number of Atoms	1333
Number of Bonds	148644

The results of PSVS provides the stereo- chemical property of the model. The molecular weight of the model is 148644.

RMS deviation of the bond angle is 1.5 °. Number of close contacts (within 2.2 Å):0 and the bond lengths is 0.011 Å. With respect to mean and standard deviation for a set of 252 X-ray structures < 500 residues, of resolution <= 1.80 Å, R-factor <= 0.25 and R-free <= 0.28; a positive value indicates a 'better' score Selected residues: -1273F-835B.The model protein 3D structure is shown in the figure -3.

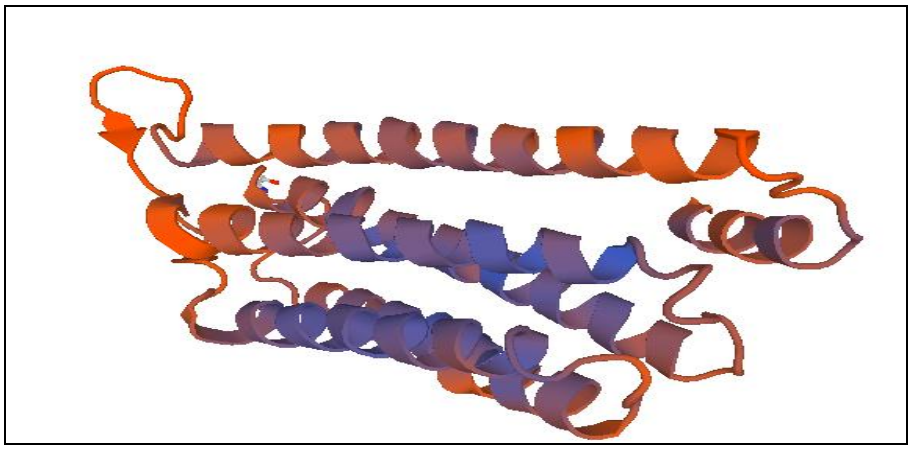


Figure-3: The model of target protein with template

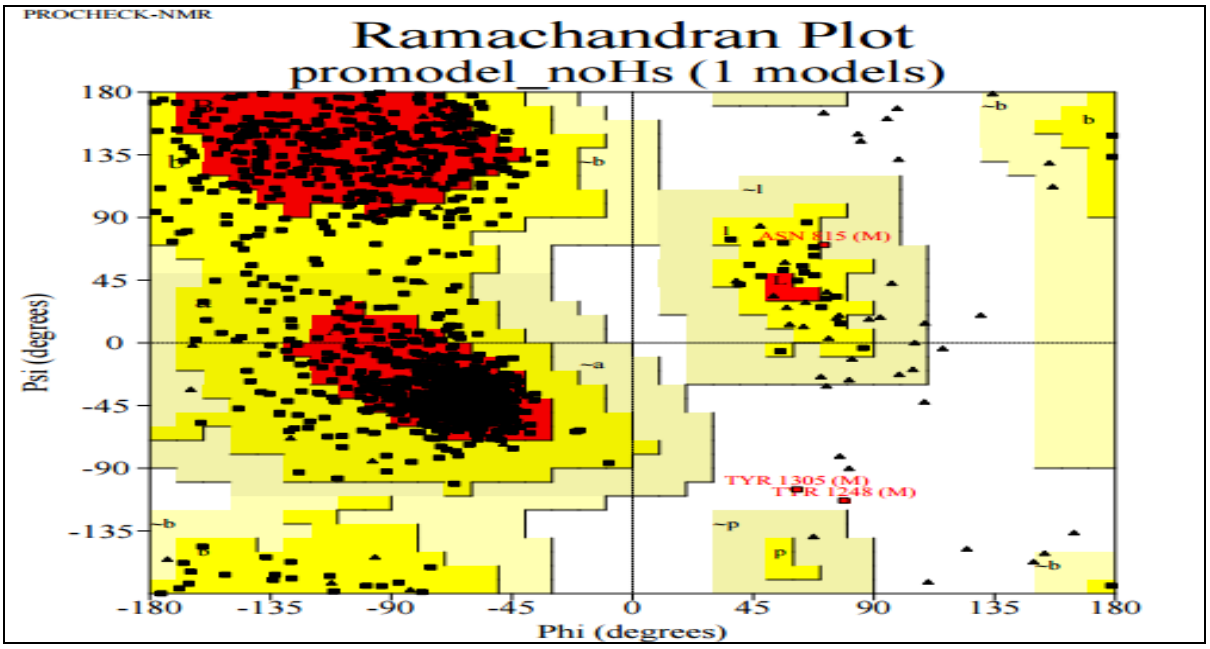
3. Ramachandran Plot statistics

The Ramachandran plot displays the psi and phi backbone conformational angles for the each residue in the target protein Cytochrome c oxidase III as shown in the figure -4.The displayed darkest region as ‘red’ in color and correspond to the “core” region and represent the most favourable combination of phi-psi values. Few residues found in allowed region. The percentage of residue in the core regions are described as follows in table-5. Figure :5 shows the residue plot of Ramachandran analysis (Based on data from Richardson Lab’s Molprobility).

Table-5: The percentage of residue in the core regions.

	From PROCHECK	From Richardson Lab's Molprobity % target	
Most favoured regions	78.6%	85.3%	

Additional allowed regions	21.2%	13.6%	
Generously allowed region	0.1%	N/A	
Disallowed regions	0.2%	1.1 %	
Structure quality-Factors-overall statistics	**Mean Score**	**SD**	**Z-Scores**
Procheck G -factor (phi/psi only)	-0.66	N/A	-2.28
Procheck G -factor (phi/psi only)	-0.66	N/A	-3.90
MollProbityClashscore	52.04	0.0000	-7.40



Plot statistics

Residues in most favoured regions [A,B,L]	887	78.6%
Residues in additional allowed regions [a,b,l,p]	239	21.2%
Residues in generously allowed regions [~a,~b,~l,~p]	1	0.1%
Residues in disallowed regions	2	0.2%
Number of non-glycine and non-proline residues	1129	100.0%
Number of end-residues (excl. Gly and Pro)	11	
Number of glycine residues (shown as triangles)	123	
Number of proline residues	70	
Total number of residues	1333	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Figure-4: Showing the Ramachandran plot of modeled protein Cytochrome c oxidase III.

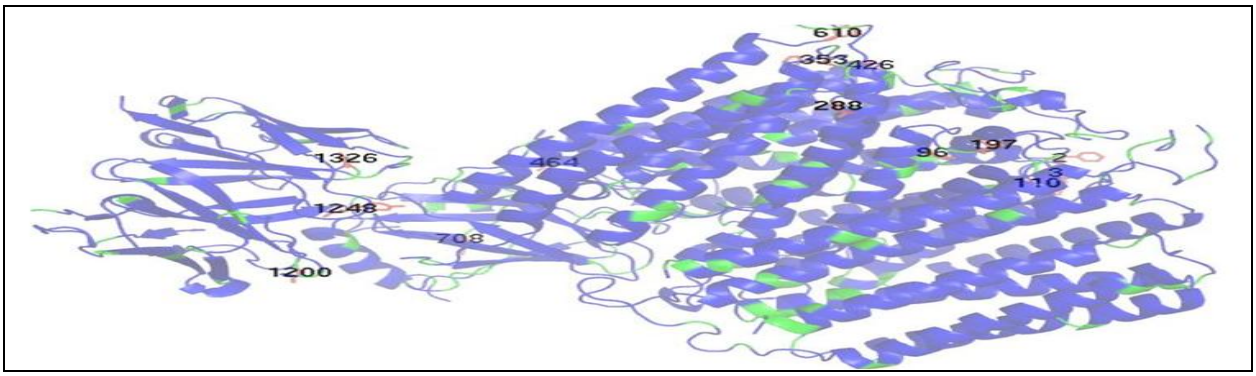


Figure-5: Residue Plot of Ramachandran analysis (Based on data from Richardson Lab's Molprobility)

PROCHECKOutput

The protein sterochemical quality were analysed .The PDB structure of Cytochrome c oxidase III protein was examined by PROCHECK tool .Procheck G-factor evaluated (figure:6) probability of dihedral angles of a residue types to be within a given range as below -

(a) Procheck G factor for phi-psi for ordered residue overall is -0.66

(b) Procheck g-factor for all dihedral angles for ordered residue overall is -0.66

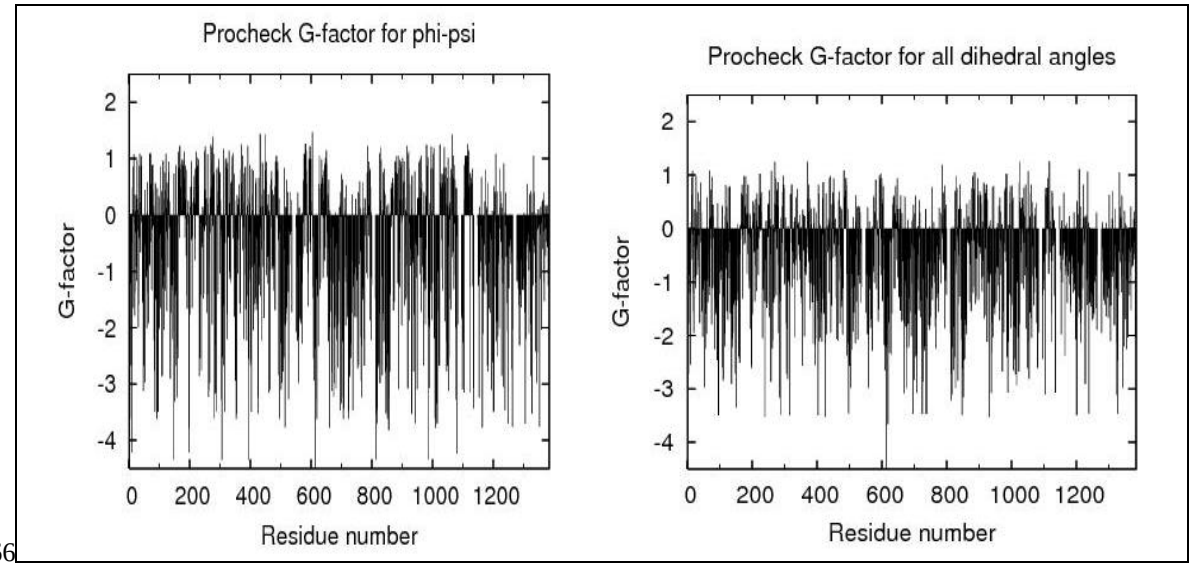


Figure-6: Procheck G-factor evaluated probability of dihedral angels of a residue.

Output from MolProbity

The MolProbity server (in PSVS server) is a one of the worthy structure validation tool in the final stage of structure refinement.VDW violations from MAGE calculate MAGE VdWclash score: Mean 52.04;SD: 0.0000 . MolProbitycalsh score and visualize atomic overlap and beta position deviations (Figure-7).

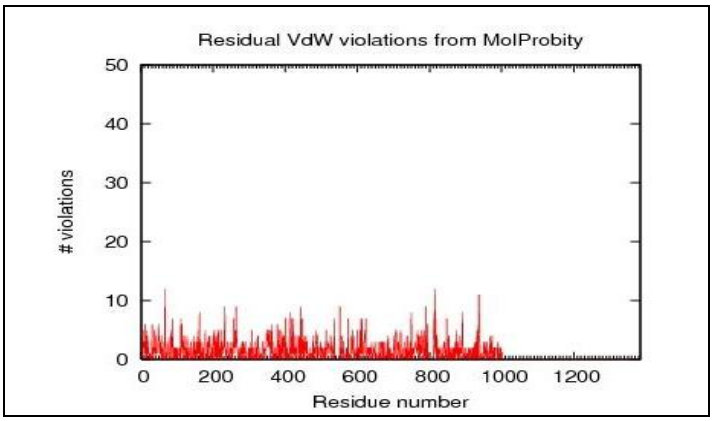


Figure-7: Showing VDW violations from MAGE calculate MolProbity clash score

PDB validation SoftwareOutput

After the PDB validation software analysis 3.5 Angstroms are consider for hydrogen bonding in the asymmetric unit and 2.2 Angstroms are considered as close contact for heavy atoms in same asymmetric unit. Distances smaller than 2.2 Angstroms are considered as close contacts. The RMS deviation for covalent bonds relative to the standard dictionary is 0.011 Angstroms.

The RMS deviation for covalent bonds relative to the standard dictionary is 0.011 Angstroms

Table -6 :The following table contains a list of the covalent bonds greater than 6.0*RMSD.

Deviation	Name of Residue	Sequence No.	Chain ID	AT1-AT2	Bond Distance	Dictionary Value
0.091	ARG	113	A	CB-CG	1.611	1.52
0.11	MET	18	B	SD-CE	1.901	1.791
0.115	TRP	252	C	CB-CG	1.613	1.498
-0.096	MET	105	E	SD-CE	1.695	1.791

Covalent Angle Values:

The RMS deviation for covalent angles relative to the standard dictionary is 1.5 degrees.

Table-7:The following table contains a list of the covalent bond angles greater than 6.0*RMSD.

Deviation	Name of Residue	Sequence No.	Chain ID	AT1-AT2 - AT3	Bonad Angle	Dictionary Value
8.9	MET	82	A	CG-SD-CE	109.8	100.9
-8.9	MET	100	A	CB-CG-SD	103.8	112.7
-9.3	MET	311	A	CB-CG-SD	103.4	112.7
9.1	MET	35	C	CG-SD-CE	110	100.9
10.8	LEU	30	D	CA-CB-CG	127.1	116.3

TORSION ANGLES

The torsion angle distributions have been checked.

CHIRALITY

The chirality checking were performed and it was found that, that there are no incorrect carbon chiral centers.

IV. CONCLUSION

In this current study, we have modeled a 3D structure of the cytochrome c oxidase III protein by homology modeling and visualized with the help of online computational tools. 3-D models of cytochrome c oxidase III protein shows significant amino acid sequence similarity with the target sequence .Homology modeling suggested the similarity between targe:template sequences. In this model, a template is a homologous protein that can be identified by a sequence similarity with target, and 43.28% identity was identified. Protein validation prediction indicates about the region where residues are present. Ramachandran Plot analysis from PROCHECK which indicated the maximum of the residues present in most favoured region , i.e81.40%of the residues were found in the most favoured region and form Richardson's lab Molprobity present 85.8% favoured regions for the selected residues.3D structure of nematode protein -cytochrome c oxidase III protein structure reported extremely in any cases in GWD. In the present current study we used GWD –cytochrome c oxidase III protein sequence as target with 1qle.1.C template for the modeling of the target protein sequence, because at the point of the template selection procedure, we found that the target protein template provides more sequence identity in comparison to other founded template sequences. This study may provide the future prospects to illustrate consideration towards a computational approach for 3D molecular modeling and computer generated model are expected to be the most accurate model but

it can't be a substitute of a crystal structure. The cytochrome c oxidase III 3-D model representation will prove to be a useful model for exploiting in the dracunculiasis disease outbreak database and residue or derivatives relationship and experimental verification in future and could be helpful in drug designing and development.

Abbreviations

GWD: Guinea worm Disease

HLA-DR: Human Leukocyte Antigen - antigen D Related

MHC: Major Histocompatibility Complex

Cytochrome c oxidase III: Heat Shock Protein 70

BLAST: Basic Local Alignment Search Tool

HHBlits : lightning-fast iterative protein sequence searching by HMM-HMM alignment

SMTL: SWISS-MODEL template library

QMEAN: Qualitative Model Energy Analysis

PSVS: The protein structure validation software suite

PDB: Protein Data Bank

Conflict of Interest – None

REFERENCES

[1] Bagchi P (2014) "Ayurinformatics : Selecting the best ligand for schizophrenic ProteinD(3) Dopamine Receptor (DRD3) from the Compund of Valerianawallichii Asparagus racemosus and Acoruscalamus".Int.J Pharm.Bio.Sci.Vol:1;Issue;1.

[2] Chothia, C;Lesk,Am (1986). "The relation between the divergence of sequence and structure in proteins". EMBO J 5(4):823-6.

[3] Muller ,R.(1971). Dracunculus and dracuneuliasis. Adv. Parasitol., 9,73.

[4] Muller,R.(1979). Guinea worm disease: epidemiology, control and treatment .Buel.WHO,57.683.1979.

[5] Jones,h.I., and Mulder E.(2007). Dracunculusmulbusn.sp.(nematode:Spirurida) from water Python Liasisfuscus(sperpentes:Boidae) in northern .Australia .Syst. Parasitol., 66.195.

[6] Moravee,F, and Santos C.P.(2009). -Dracunculusbrasiliensissp.n.(Nematoda:Dracunculidae)from the anaconda.Eunectesmurinus (Ophidia: Boidae).Parasitol.Res., 104,589,2009.

[7] Bimi,L., et al.,(2005). Ginea worm infection of urinary bladder manifesting as obstructureuropathy in rural Maharastra.Trop.DOct ., 35,242.

[8] Greenaway C. (2004).Dracunculiasis (Guinea worm disease). CMAJ 170(9), 495-500.

[9] Miillner A, HelferA ,KotlyarD.Oswald J, EfferthT. (2011).Chemistry and pharmacology of neglected helminthic disease.Curr Med Chem.,18(5), 767-789.

[10] Ruiz-Tiben E, Hopkins DR.(2006).Dracunculiasis (Guinea worm disease) eradication .AdvParasitol., 61:275-309.

[11] IriemenamNC,Oyibo WA, Fagbenro-BeyiokuAF.(2008).Dracunculiasis – The saddle is virtually ended. Parasitol Res., 102(3): 343-347.

[12] Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R., Young, I. G.,(1981). Sequence and organization of the human mitochondrial genome. Nature 290: 457-465.

[13] Giles, R. E., Blanc, H., Cann, H. M., Wallace, D. C.,(1980).Maternal inheritance of human mitochondrial DNA. Proc. Nat. Acad. Sci. 77: 6715-6719.

[14] Horváth, R; Schoser, BG; Müller-Höcker, J; Völpel, M; Jaksch, M; Lochmüller, H (2005). "Mutations in mtDNA-encoded cytochrome c oxidase subunit genes causing isolated myopathy or severe encephalomyopathy". Neuromuscular disorders : NMD 15 (12): 851–7.

[15] http://www.ncbi.nlm.nih.gov/

[16] http://swissmodel.expasy.org/ accessed on 25.01.2018.

[17] Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. &Lipman, D.J. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res. 25:3389-3402.

[18] Benkert p.,Biasini m. and Schwede T.(2011). Towards the estimation of the absolute quality of individual protein structure models. Bioinformatics.27,343-350.

[19] http://psvs-1_5-dev.nesg.org/(accesses on 25.01.2018).

[20] http://rasmol.org/(accessed on 25.01.2018)