Int J Pharm Bio Sci 2015 July; 6(3): (B) 1309 - 1321

Research Article **Bioinformatics**

International Journal of Pharma and Bio Sciences **ISSN 0975-6299**

INSIGHT INTO SEQUENCE, SRTUCTURE AND HOMOLOGY MODELLING OF MITOCHONDRIAL CITRATE SYNTHASE OF *Homo sapience*.

CHITTRAN ROY

Department of Biotechnology, Visva-Bharati, West Bengal, India

ABSTRACT

Citrate Synthase is an enzyme and most often responsible for catalyzing the first reaction of the citric acid cycle: the condensation of acetyle-CoA and oxaloacetate to form citrate. The protein sequence of mitochondrial citrate synthase having Uniprot Id O75390, has no 3D structure in protein data bank (PDB) till now. We know protein structure play an important role in their function. Our present work is based on primary and secondary structure analysis and production of high quality 3D structure of mitochondrial citrate synthase and its salt-bridge analysis. The protein was retrieved from UniProt date base. Physicochemical property has been calculated with the help of PHYSICO software. Homology modeling was performed by using MODELLER 9v11 and final model was then undergoing NAMD minimization and successfully passes through PROCHEAK, ERRAT, VERIFY-3D, Pro-Q and RAMPAGE. This model is successfully submitted in Protein Model Database having PMDB id PM0080060.

KEYWORDS: **mt-Citrate Synthase, MODELLER 9v11, PHYSICO, NAMD Minimization, SBION2.**

CHITTRAN ROY
Department of Biotechnology, Visva-Bharati, West Bengal, India

***Corresponding author**

INTRODUCTION

mtCitrate Synthase is one of the most important enzyme and exists as a homodimer of a single amino acid chain monomer. Each identical subunit consists of a large and a small domain[1]. The protein sequence of mt-citrate-synthase (human) with Uniprot[2] Id O75390 has no structure in protein data bank (PDB) till now. So structure prediction is necessary. In bioinformatics approach, structure of a protein is produced by Homology Modelling. Homology Modelling is a process by which we can produce 3D structure of protein from its amino acids sequence. Primary structure analysis is performed by Physico software[3]. With help of this software we can easily calculate GRAVY, Isoelectric P^I, Mw etc. Every enzyme has an ability to cut protein sequence at different site. Peptide cutter[4] is an important web tool which can give a list of an enzyme and their cleavage site for input protein sequence. Secondary structure analysis is performed by using SOPMA web tool[5]. Template is identified from $BLAST_P$[6] analysis of query sequence, keeping pdb as database templates was identified. Homology Modelling was performed by using MODELLER 9v11[7,8,9,10]. Trajectory analysis is performed by using VegaZZ software[11]. Model is generated from 4cts template and this model is has then undergone NAMD Minimization[12]. After energy minimization this model has been successfully passed through PROCHECK[13], ERRAT[15], VERIFY-3D[16], ProQ[17] and RAMPAGE server and analysis. PROCHECK checks the stereochemical quality of a protein structure by analyzing residue-by-residue geometry and overall structure geometry. ERRAT analyzes the statistics of non-bonded interactions between different atom types and plots the value of the error function versus position of a 9-residue sliding window, calculated by a comparison with statistics from highly refined structures. This model was successfully submitted in Protein Model Database (PMDB)[18] having PMDB id PM0080060. Structural superimposition is performed by STRAP interface. Active site prediction is performed by CASTp server(http://stsfw.bioengr.uic.edu/castp/calculation.php). CASTp server uses the weighted Delaunay triangulation and the alpha complex for shape measurement. It provides identification and measurements of surface accessible pocket as well as interior inaccessible cavities, for protein and other molecule

MATERIALS AND METHODS

(i) sequence Retrieval

The amino acid sequence of Citrate-synthase of mitochondria (human) has retrieved from UniProt public protein database (http://www.uniprot.org/). UniProt has scanned with the searching string “mitochondrial citrate synthase, human”. The search result gives only one sequence (UniProt id: O75390). The amino acid sequence was retrived in FASTA format for further analysis.

(ii) Primary Structure Analysis

(a)Analysis of Amino acid composition

The amino acid composition of this protein sequence is run through PHYSICO software. The percentage of the amino acid composition is presented in Table 1. The molecular weight (MW), theoretical isoelectric P^I, total number of positive and negative residue, Instability index, aliphatic index, grand average hydropathy or GRAVY were calculated by using PHYSICO software (Table 2).

(b)Cleavage Site Prediction by PeptideCutter

Peptide-Cutter (http://web.expasy.org/peptide_cutter/) is an important web tool to predict the potential cleavage sites cleaved by proteases or chemicals in a given protein sequence. Different enzyme has an ability to cut a particular position of protein sequence. Peptide-Cutter can easily identify the cleavage site of the input sequence. The different cleavage sites of citrate synthase protein sequence are presented in Table 3.

(iii) Secondary Structure Analysis

The secondary structure is predicted using SOPMA web tool. It is an important web tool used for secondary structure analysis of protein. The secondary structure analysis present in Table 4.

(iv) Identification of Template

The amino acid sequence of citrate synthase has retrieved from UniProt protein sequence database. Then BLASTp (http://blast.ncbi.nlm.nih.gov/) has performed keeping pdb as database, for this query sequence and appropriate templates for model building has selected based on the high score, lower evaluate and maximum sequence identity. Finally one template is used for model production. The template 4CTS_A is from the A chain of Crystal structure of the Cathelin-like Domain Of Human Cathelicidin L1-37 (hcld).

(v) Model Production

The protein sequence of citrate synthase with Uniprot Id: O75390 has no three dimensional structure in protein data bank (PDB). The templates with PDB Id 4CTS_A is used form model production. Prediction of 3D structure of this query protein was performed by widely used and most popular homology modelling software package MODELLER 9v11. Alignments between template and target sequences were performed by Clustal Omega. A few manual improvements were given in alignments. Finally protein model is produced by model single.py script. At least five models are produced from each modelling procedure. The best model is then selected, out of five models. Final Model is produced from template 4CTS_A.

(vi) Energy Minimization

The model structures are separately subjected for refinement via energy minimization. Model has then undergone NAMD minimization under explicit water box condition (fig 5). After 5000 steps of NAMD minimization, trajectory analysis is performed by using VegaZZ software. The best fame is selected and saved it in pdb format.

(vii) Model Evalution and Submission

a. Structural evaluation, validation and stereochemical analysis of model is performed using various bioinformatics tools and software package i.e. PROCHECK, RAMPAGE, ERRAT, VERIFY 3D. New model is successfully passed each criterion. Model then successfully submitted in Protein Model Database (PMDB) having PMDB Id PM0080060.

b. Extraction of salt bridges and evaluation of net energy ($\Delta\Delta G_{net}$) for template (4cts.pdb) and target (Model_A.pdb, Id=o75390) is performed using SBION2[18] and ADSBET[19] respectively.

(viii) Structural Superimposition

It is very much effective tool in bioinformatics for comparison of two structures. Structural superimposition was performed from STRAP interface by using Superimpose3D_TM_align method.

(ix) Active Site prediction

CASTp (i.e. Computed Atlast of Surface Topography of Proteins) is a web tool, used to predict active sites with their respective volume and area. Inside the home page(http://stsfw.bioengr.uic.edu/castp/calculation.php), the model structures are separately uploaded and then Submit button option was clicked. In the next page result appeared.

Table1

Amino acid composition of mt citrate synthase(human). UniProt Id : O75390

Amino Acid	Codon	% Percent of Protein
C	1	0.9
S	6	7.1
T	4	5.6
P	4	4.7
A	4	8.6
G	4	7.7
N	2	4.3
D	2	4.3
E	2	5.2
Q	2	3.6
H	2	3.2
R	6	4.5
K	2	5.6
M	1	3.3
I	3	3.6
L	6	12.7
V	4	6.4
F	2	3.0
Y	2	3.9
W	1	1.9

Table2

Different physicochemical parameters of mt citrate synthase(Uniprot Id O75390)

Accession Number	M.W	P^I	II	AI	GRAVY
O75390	51712.4	8.45	22.4	90.86	-0.165

Table3

Different cleavage site of different enzymes in citrate synthase seq.

Name of enzyme	No. of cleavages	Positions of cleavage sites
CNBr	15	1 67 68 72 75 154 165 203 243 293 373 414 417 452 458
Chymotrypsin-high specificity (C-term to [FYW], not before P)	38	51 69 80 93 96 120 121 136 141 181 185 194 195 199 217 221 236 240 246 251 258 286 289 311 331 333 345 357 363 374 381 411 412 419 420 424 438 460
Formic acid	20	35 39 66 86 88 156 201 204 231 235 248 264 284 322 325 330 354 371 402 462
Hydroxylamine	1	294
Pepsin (pH1.3)	111	2 3 3 4 10 10 18 19 20 21 32 33 37 39 50 77 84 93 95 103 118 119 119 120 121 122 122 123 136 137 147 154 155 158 161 167 168 174 175 180 181 196 197 204 205 209 219 233 234 239 243 244 250 253 254 257 259 276 277 281 282 287 288 289 295 296 299 302 308 309 311 312 314 315 327 328 335 336 350 362 364 365 374 375 376 379 380 387 388 388 389 407 408 408 409 422 423 423 424 429 432 433 435 436 441 444 445 456 457 459 460
Staphylococcal peptidase I	22	44 81 89 100 112 117 131 140 178 196 200 223 253 266 307 318 326 362 390 416 447 455
Trypsin	46	9 14 24 34 43 47 49 52 57 73 76 92 94 103 107 109 139 143 144 183 191 193 208 215 218 222 256 317 321 327 329 340 351 352 356 361 366 375 382 393 395 428 440 450 459 464

Table4

Secondary structure analysis of mt citrate synthase(Uniprot Id O75390)

Accession Number	Secondary Structure				
	Alpha Helix	Beta Turn	Random Coil	Others	Extended Strand
O75390	41.20%	50%	9.23%	0.0%	13.09

RESULTS AND DISCUSSION

(i) Primary Structure Analysis

The results of primary structure analysis suggest that the protein citrate synthase is hydrophilic in nature due to the presence of high content of polar residues. The kye-dollite hydrophobicity of this protein sequence is calculated by using online ProtScale web server (Fig. 1) (http://web.expasy.org/protscale/). The average molecular weight of citrate synthase is about 51712.40. Isoelectric point (pI) is the pH at which the net charge of the protein is zero. Sequence analysis of this protein shows that this protein is alkali in nature due to pI = 8.45. The estimated half life (not shown in table) of citrate synthase is 30 is greater than 20 hour in case of yeast (in vivo), and greater than 10 hours in in E.Coli (in vivo). The instability index is about 22.40 that mean this protein is stable in nature.

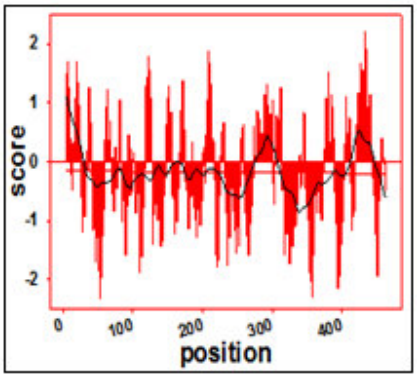

Figure1

Kye-dollite-hydrophobicity analysis of mtcitrate synthase sequence (Uniprot Id: O75390)

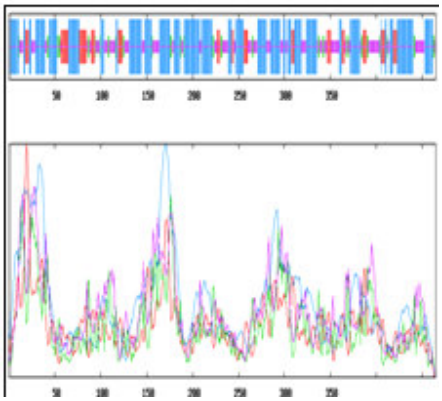

Figure 2

Secondary structure analysis of Using SOPMA Tool (Uniprot Id: O75390)

The aliphatic index of a protein is a measure of the relative volume occupied by an aliphatic side chain of the following amino acids: alanine, valine, leucine and isoleucine. An increase in the aliphatic index increases the thermostability of globular proteins. The aliphatic index is calculated by the following formula- Aliphaticindex[23]= X(Ala)+a*X(Val)+b*X(Leu)+b*(X)Ile. Here *X(Ala)*, *X(Val)*, *X(Ile)* and *X(Leu)* are the amino acid compositional fractions. The constants a and b are the relative volume of valine (a=2.9) and leucine/isoleucine (b=3.9) side chains compared to the side chain of alanine. The aliphatic index of citrate synthase is about 90.86. The Grand Average Hydrophaty or

GRAVY of citrate synthase is about -0.165. The GRAVY value for a peptide or protein is calculated as the sum of hydropathy value of all the amino acids, divided by the number of residues in the sequence. The peptide cutter web server shows that many enzymes have an ability to cut different site within the protein sequence. Table 3. Secondary Structure analysis by using SOPMA web tool shows that (Table 4) the secondary structure of this protein sequence contain 41.20% α-helix, 9.23% random coil and 0.0% consist of other part. The graphical representation of SCOPMA is shown in Fig.2. The Intrinsic Protein Disorder, Domain and globularity prediction were performed from GLOBPLOT 2.3 web server (http://globplot.embl.de/)Fig.3

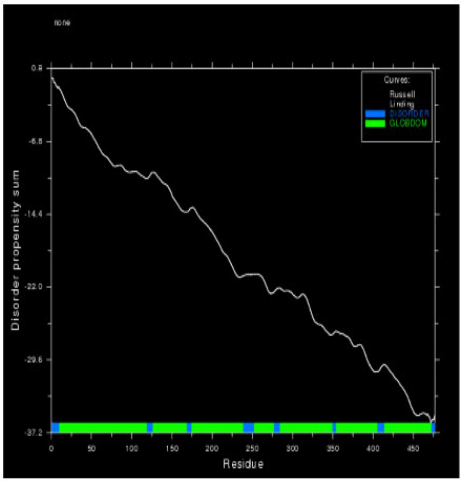

Figure3
Disorder Propensity Sum of citrate synthase (Uniprot Id O75390, modeled protein** PM0080060) **calculatedfromGLOBPLOT 2.3 Web Server

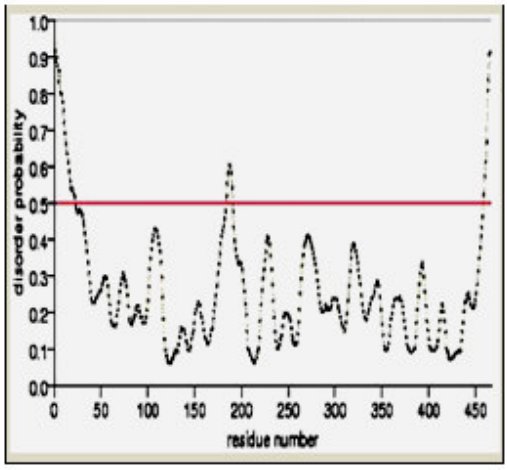

Figure4
Disorder Probability of the entire Citrate synthase (Uniprot Id O75390, modeled protein** PM0080060) **Sequence calculated from PrDOS Web Server

(ii)Homology Modelling, NAMD Minimization

Homology Modelling of Citrate synthase of human was performed by using one of the popular homology modelling software package MODELLER 9v11. After PrDOS (http://prdos.hgc.jp/cgi-

bin/top.cgi) is a server to predict natively disordered regions of a protein chain from its amino acid sequence. PrDOS returns disorder probability of each residue as prediction results Fig. 4.

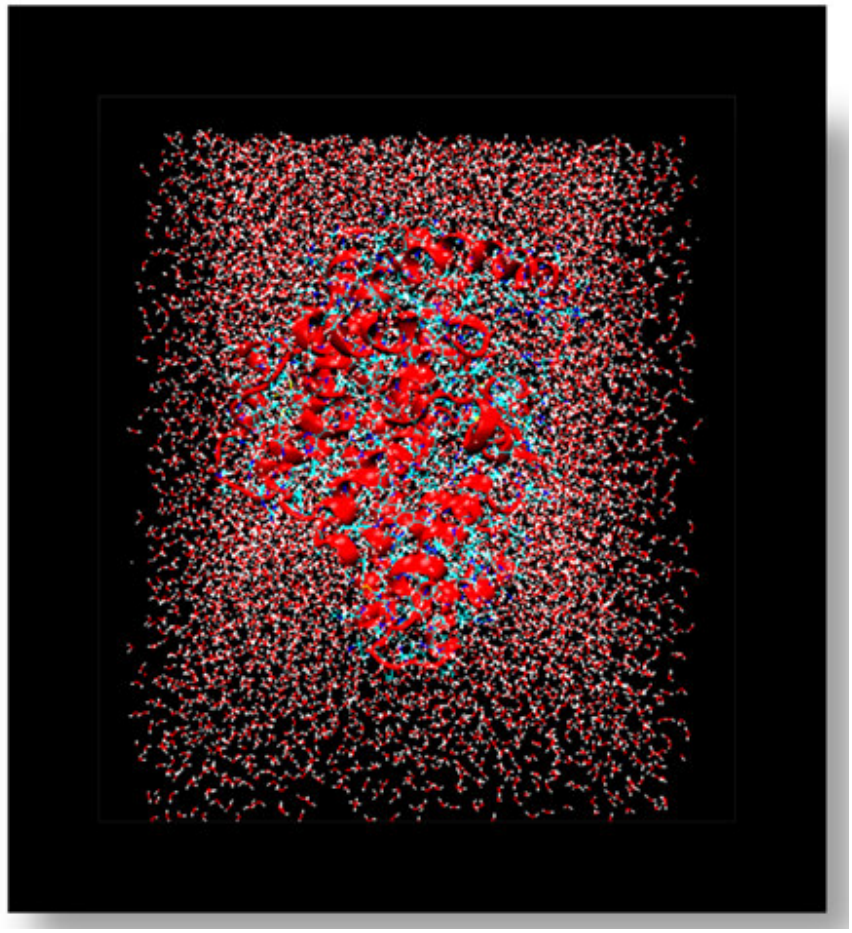

Figure 5
Modelled Protein (*PM0080060****) under in water box***

Completion of modeling model was then separately undergoing NAMD Minimization under explicit Water-Box condition (Fig 5). After energy minimization best model was selected from each experiment. Finally best model was produced from selected template i.e. 4CTS_A. The 3D structure of a generated model is shown in Fig.6. Structural evaluation, validation and stereochemical analysis were performed using various online available evolution tool such as

PROCHECK, ERRAT, VERIFY 3D, and ProQ.

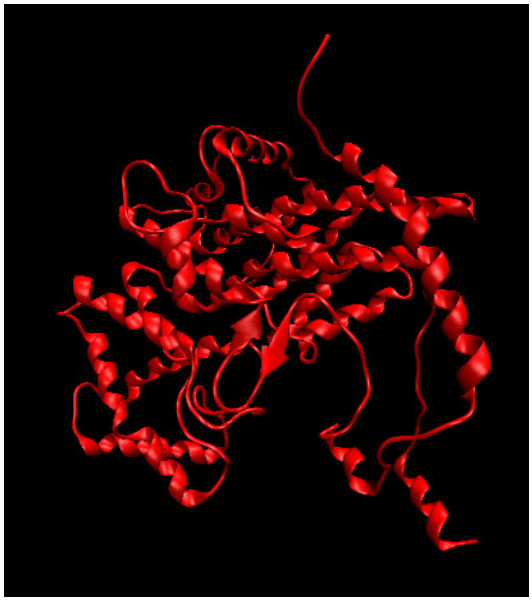

Figure 6.1
3D Modelled Structure of citrate synthase(Uniprot Id O75390)

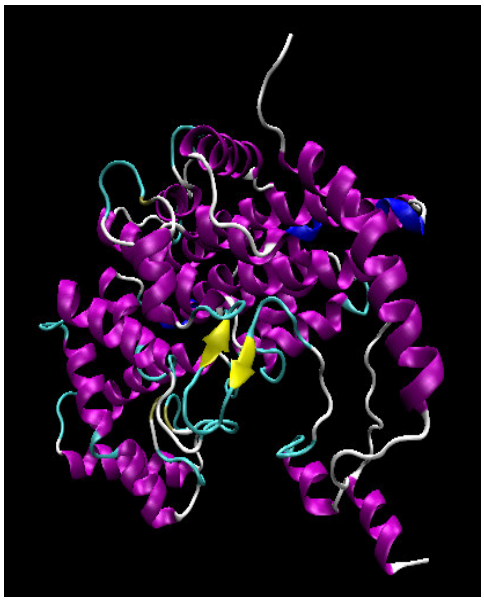

Figure 6.2
Modelled Structure of citrate synthase (Uniprot Id O75390).

All mention web tool suggest that this model was good model according to their criteria (Table 5). The PROCHECK analyses of this model are shown in (Fig.7). The ERRAT analysis is shown in Fig.8.Structural Superimpositions were performed from STRAP interface (Fig.9).

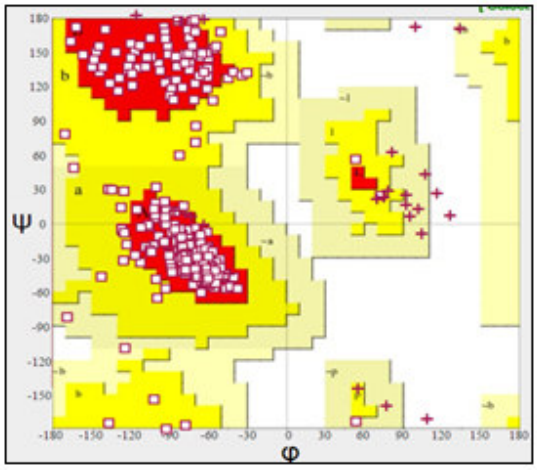

Figure 7
Procheck Analysis of model protein

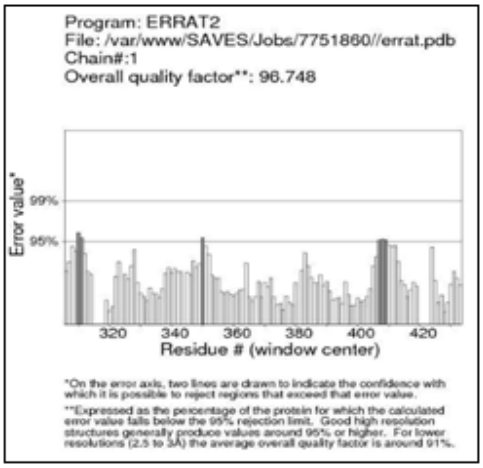

Figure 8
ERRAT analysis of model protein

TABLE 5
Validation result of model-Model_A (Human mitochondrial Citrate Synthase) of Citrate Cynthase

Model name	Template use	ERRAT	VERIFY 3D	RAMPAGE	ProQ
PM0080060	4CTS__A	96.74	92.43	96.9	LG Scor-3.49

Salt bridge evaluation of the developed model is compared with the template (4CTS). it is interesting to note that i] salt bridge specificity of the model remains almost similar as template, ii**]** net energy of each salt bridge is more stabilizing than that of the template and iii] excess salt bridges (three in number) are formed in case of the target indicating their crucial structural roles(table 6).

TABLE6

Result of ion-pair distribution and energy calculation of model (Model_A) and template (4CTS) protein

ION-PAIR	$\Delta\Delta G_{net}^{Model\_A}$ (1)	$\Delta\Delta G_{net}^{4CTS}$ (2)	(1) – (2)
LYS290_GLU291	-2.22	-1.66	**-0.56**
LYS181_ASP177	-8.13	-1.46	**-6.67**
LYS188_ASP129	-6.24	-3.54	**-2.7**
LYS166_GLU169	-1.55	-1.46	**-0.09**
LYS339_GLU85	-5.96	-1.36	**-4.6**
LYS325_ASP61	-1.94	0.62	**-2.56**
ARG302_GLU363	-3.81	-1.9	**-1.91**
ARG229_GLU226	-13.66	-3.27	**-10.39**
ARG117_GLU113	-18.98	-14.44	**-4.54**
LYS7_GLU173	-8.69	-6.4	**-2.29**
ARG117_GLU173	-18.59	-11.94	**-6.65**
ARG229_ASP208	-10.88	-8.11	**-2.77**
ARG117_ASP177	-18.84	-11.61	**-7.23**
ARG329_ASP327	-4.43	3.65	**-8.08**
HIS246_GLU420	-0.28	1.08	**-1.36**
LYS116_GLU113	-8.17	-8.09	**-0.08**
ARG20_GLU17	-4.64	-1.18	**-3.46**
LYS16_ASP12	**-2.9**		
HIS123_GLU151	**-1.17**		
ARG401_ASP237	**-0.18**		

(iii) Active Site Prediction

The active site is a small part of enzyme where substrate molecules bind and undergo chemical reaction. The active site (26aa to 36aa) of the final model (Model_A) is calculated by using CASTp web server (Fig 10).

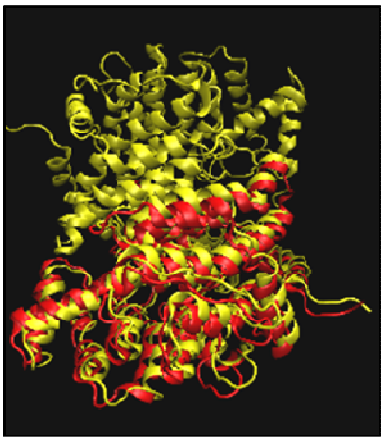

Figure 9

Structural Superimpositions between model(yellow) and template(red).

Figure 10
Active site of citrate synthase (Human),Red colour indicate active site.

CONCLUSION

Citrate synthase is a cytosolic protein that participate in the first step of TCA cycle. Primary structure and its physico-chemical properties studies show that protein is hydroplilic in nature with PI value at alkaline range. Aliphatic index shows protein possesses thermal stability. Highly stable homology model of the protein is developed for the first time (deposition Id PM0080060) as verified by steriochemical and structural criteria. Relative to template the model gain extra stability by forming excess stabilized salt bridges. Overall the studies demonstrate detail of sequence and structural properties of human citrate synthase.

ACKNOWLEDGEMENT

Authors are grateful to the Department of Biotechnology (DBT), Govt. of India for the computational laboratory facility.

REFERENCES

1. W.Georg, and S.J.Remington (1986). "Citrate Synthase: Structure, Control, and Mechanism." Ann. rev. Biophys. Biophys. Chem. Pg 98
2. Update on activities at the Universal Protein Resource (UniProt) in 2013 Nucleic Acids Res. 41: D43-D47 (2013).
3. Sen Gupta.PS, Banerjee.S, Mondal.S, Islam,RNU, Mondal.B, Bandyopadhyay.AK. PHYSICO: An UNIX based Standalone Procedure for Computation of Individual and Group Properties of Protein Sequences. Bioinformation. 2014, 10 (2), 105
4. Barrett. A, Rawlings.N.D, Woessner.J.F. *Handbook of proteolytic enzymes*. Academic Press, (1998)
5. Geourjon C, Deleage G .SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci, 11(6):681-684(1995).
6. Altschul S.F., Gish W., Miller W., Myers E.W. and Lipman D.J. Basic local alignment search tool.J. Mol. Biol, 215: 403-410. (1990)
7. Eswar.N, Marti-Renom.MA, Webb.B, Madhusudhan.MS, Eramian.D, Shen.M, Pieper.U, Sali.A Comparative Protein

Structure Modeling With MODELLER. Current Protocols in Bioinformatics, John Wiley & Sons, Inc., Supplement 15, 5.6.1-5.6.30, (2006)

8. Marti-Renom.MA, Stuart.A, Fiser, R. Sánchez, Melo.F, Sali.A, Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. 29, 291-325, (2000).
9. Sali.A & Blundell.TL. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779-815, (1993).
10. A. Fiser, R.K. Do, & A. Sali. Modeling of loops in protein structures, Protein Science 9. 1753-1773, (2000).
11. Pedretti.A,Villa.L,Vistoli.G ,Vega - An Open Platform To Develop Chemo-bio-informatics application, using plug-in architecture and script programming. J.C.A.M.D., Vol. 18, 167-173, (2004).
12. James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kale, and Klaus Schulten. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry,* 26:1781-1802, (2005).
13. Laskowski. R. A., Macarthur. M. W, Moss. D. S, Thornton. J. M, PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, 26: 283-291 (1993).
14. Rob W. W. Hooft, Gert Vriend, Chris Sander, Enrique E. Abola. Errors in protein structures. *Nature*, 381:272-272, (1996)
15. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. 2, 1511-1519 ,(1993).
16. Luthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. Nature 356, 83-85,(1992).
17. Björn Wallner and Arne Elofsson. Can correct protein models be identified? Protein Sci. 12(5):1073-1086, (2003).
18. Sen Gupta PS, Nayek A, Banerjee S, Seth P, Das S, Sur VP, Roy C, Bandyopadhyay AK. SBION2: Analyses of Salt Bridges from Multiple Structure Files, Version 2. Bioinformation 2015;11:39-42
19. Sen Gupta PS, Nayek A, Banerjee S, Seth P, Das S, Sur VP, Bandyopadhyay AK. ADSBET: Automated determination of salt-bridge energy term
20. Tiziana Castrignanò, Paolo D'Onorio De Meo, Domenico Cozzetto1Ivano ,Giuseppe Talamo1 and Anna Tramontano. The PMDB Protein Model Database, 34: D306-D309, (2006)
21. Sen Gupta P S, Mondal B and Bandyopadhyay A K. In silico characterization of human tyrosinase using computational tools and servers IJPBS. (2013), 4(3), 181
22. Sen Gupta PS, Mondal B and Bandyopadhyay A K. In silico characterization of human cyclooxygenase using computational tools and servers. IJIPLS. (2013), 3(6), 124.