Structural and functional annotation of citrate synthase from *Aspergillus niger* ANJ-120

Ghulam Mustafa[1]*, Rawaba Arif[2], Shazia Anwer Bukhari[1], Muhammad Ali[1], Sumaira Sharif[3] and Asia Atta[4]

[1]Department of Biochemistry, Government College University, Faisalabad, Pakistan
[2]Department of Biochemistry, University of Agriculture, Faisalabad, Pakistan
[3]College of Biosytem Engineering and Food Science, Zhejiang University, China
[4]Department of Biochemistry, Bahauddin Zakariya University, Multan, Pakistan

Abstract: Citrate synthase (CS) is involved in citric acid biosynthesis which is a well-established metabolic pathway. The condensation of acetyl-CoA with oxaloacetate is catalyzed by CS. Citric acid (CA) has a number of applications in pharmaceutical industry. CA in combination with bicarbonates is used as an effervescent in the preparations of tablets and powders. It has also been used as an anticoagulant and acidulant to form mild astringent. In current study, detailed structural and functional analyses of CS protein were carried out using various bioinformatics tools. Structural modeling was also done by building 3D model of CS from *Aspergillus niger* ANJ-120 using Modeller 9.16 software. The 3D Model was then evaluated using different online approaches. Furthermore, superimposition of query and template structures, Root Mean Squared Deviation and visualization of generated model were done through UCSF Chimera 1.5.3. Even though various roles of CS protein were already known and verified experimentally, here we presented a structural analysis of CS protein. The structural investigation of CS protein will be helpful for protein engineering strategies and understanding the interactions among proteins. Due to large number of applications, the production of citric acid by *A. niger* and its bioinformatics studies will offer substantial improvement in commercial scale intensification of this useful product.

Keywords: *Aspergillus niger,* citrate synthase, annotation, citric acid, homology modeling.

INTRODUCTION

Citrate synthase (CS, E.C. 4.1.3.7) catalyzes the condensation reaction of acetyl-CoA with oxaloacetate to produce citrate and coenzyme A. CS is involved in the biosynthesis of citric acid that is a well-established metabolic pathway (Mustafa *et al*., 2014). Due to its spacious range of applications, citric acid is extensively used in food and pharmaceutical industries (Magnuson and Lasure, 2004). Citric acid has been sold as an anhydrous or monohydrate acid in world market and 70% of its total production is used as an acidifier or antioxidant in various food and beverage industries for the preservation and/or enhancement of flavors (Lancini, 2008). About 20% of citric acid has been used in pharmaceutical industry as an antioxidant for the preservation of blood and vitamins, effervescent and pH corrector. It has also been used as an iron source in the form of iron citrate for the body as well as in tablets and balms (Max *et al*., 2010).

It was demonstrated first time by Wehmer (1983) that citric acid gets accumulated by *Penicillium* in a sugar and inorganic salts containing medium. Since then, citric acid has been found to be accumulated by various organisms. *Aspergillus niger* which is widely used in biotechnology for food ingredients, pharmaceuticals and industrial enzymes production, belongs to the family of black aspergilli. Although it is considered as "cell factory" for both citric acid and gluconic acid productions (Ruijter *et al*., 2002), but it is generally known as citric acid producer (Magnuson and Lasure, 2004). Under specific growth conditions, when glucose or sucrose is used as a carbon source in the growth medium of *A. niger*, it produces high yields of citric acid. Production of citric acid by *A. niger* is given out as model fungal fermentation process as the fungus is producing over one million metric tons of citric acid annually (Baker, 2006). The importance of *A. niger* as citric acid producer is unambiguous from the fact that it has been studied for nearly 100 years for citric acid fermentation (Currie, 1917). The daily conversion of feedstock sugar into citric acid is >90% with a yield of >200 g/L and these efficiencies are more striking as compared to other fermentation processes. As the commercial importance of citric acid is increased, a number of efforts are being employed to enhance the production of citric acid to meet its mounting demands. The current study was therefore, designed to annotate CS protein sequence structurally and functionally. Three dimensional structure of CS was also modeled and compared with the template to find out structural differences between both models.

**Corresponding author:* e-mail: gmustafa_uaf@yahoo.com

METHODS

Sequence and Physicochemical analyses

The DNA sequence of citrate synthase gene of our previous study (Mustafa *et al*., 2014) was obtained from GenBank using accession number (GenBank: KC847093.1). To find accurate coding region and amino acid sequence of CS gene, Softberry FGENESH was used (Solovyev *et al*., 2006). Physicochemical properties were calculated from ExPASy server (Gasteiger *et al*., 2003). ProtParam (Gasteiger *et al*., 2005) was also used for the prediction of physicochemical properties such as molecular weight, estimated half-life, theoretical pI, amino acid composition, atomic composition, aliphatic index, instability index and grand average of hydropathicity (GRAVY). To further confirm theoretical pI and molecular weight, Compute pI/MW was used. Hydrophobicity value was obtained through ProtScale (Bjellqvist *et al*., 1993).

Domain and Motif predictions

Interproscan (Burge *et al*., 2012) and NCBI Conserved Domain (CDD) (Marchler-Bauer *et al*., 2005) were used to predict domains in CS protein. SMART (Schultz *et al*., 1998) and Pfam (Sonnhammer *et al*., 1997) were used to further confirm and interpret detailed functions and interactions of predicted domains. Motif analysis was done using Motifscan tool (Yusim *et al*., 2003).

Prediction of secondary structure elements

Local inter-residue interactions which are facilitated by H-bonds are found in the secondary structures of proteins. Alpha helices and beta sheets are the most common secondary structure elements in proteins. Secondary structure of CS protein was predicted using PsiPred (McGuffin *et al*., 2000), and Chou & Fasman secondary structure prediction server (Chou and Fasman, 1974). The information obtained from predicted secondary structure of CS was used for the improvement of alignment between template and query protein sequences and to build 3D model of CS protein.

Homology modeling to predict 3D structure

CPH Model Server (Krieger *et al*., 2003) was used to search template. Template PDB ID 4CTS was selected from this server. Alignment was done between template and target sequence using MODELLERv 9.16 (Webb and Sali, 2014). 3D structure was predicted using sequence alignment of template and target sequences by MODELLERv 9.16. Five models for CS protein were built through python script commands. MODELLERv 9.16 was also used to optimize loops of homology model which is necessary to improve quality of predicted model. Evaluation of predicted 3D model was done using ModEval (Barkan *et al*., 2010), Dope profile (Shen and Sali, 2006), verify 3D (Luthy *et al*., 1992), ProSA-web (Wiederstein and Sippl, 2007), Z-scores, Qmean plots (Benkert *et al*., 2008) and PROCHECK Ramachandran plots (Lovell *et al*., 2002). Moreover, RMSD, superimposition of target and template 3D models and visualization of predicted models were accomplished through UCSF Chimera 1.5.3 workbench (Pettersen *et al*., 2004).

RESULTS

The present study was planned to perform sequence and structure analyses of *A. niger* citrate synthase protein.

Physicochemical properties of CS

ProtParam was used to predict physicochemical properties of CS protein from *A. niger* (table 1). The protein was found to have 460 amino acids, molecular weight of 50634.94 Daltons and isoelectric point (pI) of 8.39. pI >7 shows that protein is positively charged. The instability index (II) was found to be 26.76. This categorizes the protein as stable. The considered N-terminal sequence was T (Thr) and estimated half-life was found to be 7.2 hours (mammalian reticulocytes, *in vitro*), >20 hours (yeast, *in vivo*) and >10 hours (*Escherichia coli*, *in vivo*). The negative Grand average of hydropathicity (GRAVY) was -0.231, which shows that protein is hydrophilic and soluble in nature (Mustafa *et al*., 2017). Leucine and alanine were found in rich quantity in CS proteins.

Annotation of citrate synthase protein

The functional profile of a protein is indicated by its domains and motifs. Fig. 1 shows the visual output of domains found in citrate synthase protein predicted by InterProScan (version 4.8). Only one domain was predicted in CS and the results were in accordance with those of Pfam and SMART.

The location of predicted citrate synthase, C-terminal domain by Pfam was from 66 to 445 and of 361 amino acids Hidden Markov Model (HMM) length. ScanProsite also predicted a citrate synthase signature as an active site between amino acids 339-351. A leucine zipper pattern was also predicted in the CS protein from amino acid 277 to 298.

Other important sites predicted in citrate synthase protein have been shown in fig. 1b. A polypeptide binding site (dimer interface), active site, coenzyme A binding site (chemical binding), citrylCoA binding site (chemical binding), oxalacetate/ citrate binding site (chemical binding) and catalytic triad (active site) have been found in the CS protein.

Protein secondary structure prediction

Chou & Fasman server was accessed to predict the secondary structure of CS protein. Results showed that CS is a mixed proteins having compositions of Helices = 68%, Sheets = 62.6% and Turns = 12.6%.

Table 1: Physicochemical properties of CS protein predicted by Prot Param

Sr. No.	Property	Value
1	Molecular weight	50634.94 Dalton
2	Theoretical pI	8.39
3	Instability index	26.76
4	Aliphatic index	90.59
5	Grand average of hydropathicity (GRAVY)	-0.231

Fig. 1a: Structural and functional annotation of citrate synthase protein

Fig. 1b: Structural and functional annotation of citrate synthase protein: dimer interface (polypeptide binding); active site; coenzyme A binding site (chemical binding); citryl-CoA binding site (chemical binding); oxaloacetate/ citrate binding site (chemical binding); catalytic triad (active site)

Secondary structure of CS protein was also predicted through Garnier-Osguthorpe-Robson (GOR) method (Garnier *et al*., 1996). Values of alpha helices, beta sheets, turns and coils were given to each residue (fig. 2). The protein secondary structure analysis of CS with GOR4 online tool expressed that α-helices accounted for 39.35%, extended strand accounted for 15.43% and random coils accounted for 45.22%. These are the largest regions of structural elements of CS.

The hydrophobic residues are repelled by surrounding water molecules and therefore embedded within the protein. This trend along with 3D conditions, space and other factors eventually govern the construction of a protein folding into a 3D conformation. The results of ProtScale analysis revealed that the maximum hydrophobicity of CS was 2.122 and minimum was -2.8 (fig. 3). The amino acids in the region 77 ~ 85, 115 ~ 125, 190 ~ 220, 240 ~ 265, 410 ~ 440 for CS have strong

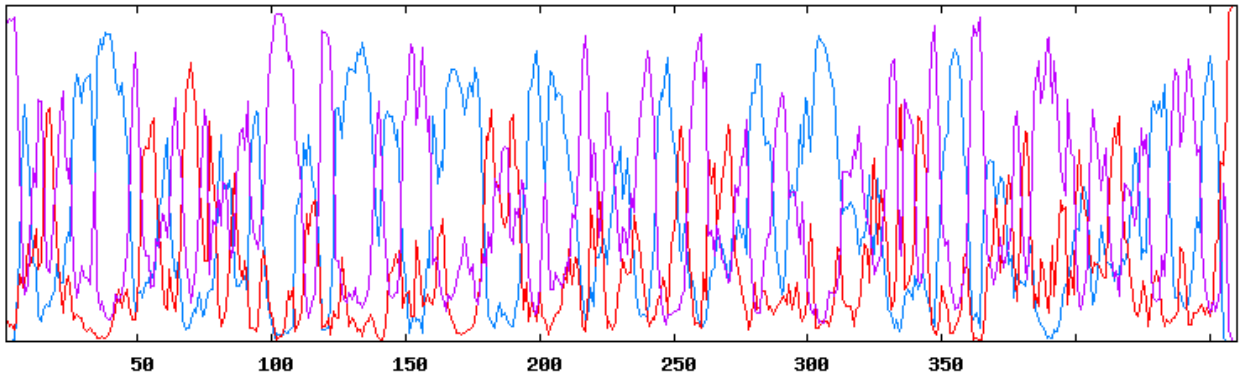

Fig. 2: Secondary structure GOR plot of citrate synthase protein

hydrophobic properties and the majority of these regions is found to be α-helices, while the corresponding random coil region is predicted to have a very low level of hydrophobicity.

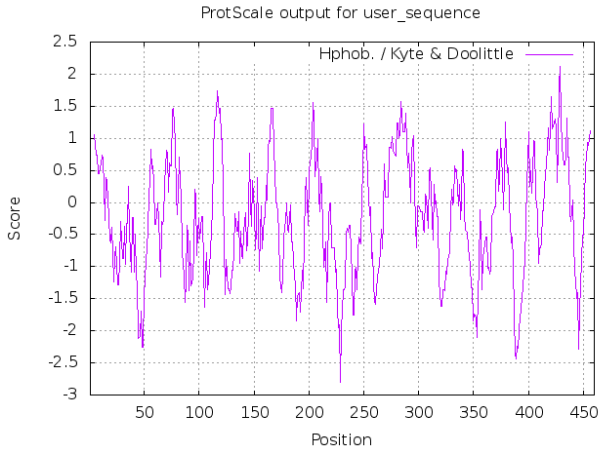

Fig. 3: Prediction of hydrophobic character of amino acid sequence of citrate synthase from *A. niger* by Hphob./Kyte & Doolittle scale

The protein sequences of CS was also submitted to the PSIPRED server in the format of single letter amino acid code and PSIPRED prediction method was selected to predict the secondary structure of CS (fig. 4).

3D structure prediction using homology approach

Tertiary structure of CS was predicted by homology modeling. In our study, the 3D structure of CS was predicted by computing modeling using Modeller9v16 software and visualized in Chimera1.6.2 software. Modeller is a python based homology modeling tool used for modeling proteins. Out of 10 models generated for CS protein, one best-fit model was selected on the basis of minimal violation of probability density function (fig. 5a).

Evaluation of protein structure quality

Predicted model of CS protein was evaluated by ModEval, Verify 3D and ProSA-web tools. According to Verify 3D, 86.21% of residues of predicted 3D structure of CS had an averaged 3D-1D score ≥0.2. The predicted 3D model of a protein is considered good if at least 80% of the amino acids have score ≥0.2 in the 3D-1D profile (Luthy *et al*., 1992).

The z-DOPE (normalized Discrete Optimized Protein Energy) of predicted model was found to be -0.879. The score of GA341 was found to be 1.000. The z-score of the predicted structure was -10.21 that is in the range of scores obtained for proteins with related sizes (fig. 5b). Energy plot of predicted 3D structure is also showing (dark green line) that all residues are at very stable position (fig. 5c).

Ramachandran's plot calculations using ProCheck was used to evaluate the stereochemistry of backbone Psi and Phi dihedral angles. Percentage of residues of CS occupying most favored regions (A,B,L) is 83.4% while 14.2% occupying additional allowed regions (a,b,l,p), 1.9% and 0.5% resides in generously allowed (~a,~b,~l,~p) and disallowed regions respectively (fig. 5d). It was ascertained on the basis of these results that the predicted model is of good quality.

A structural comparison of predicted 3D structure of CS with its template (i.e. 4CTS) was performed on the bases of structure superimposition (fig. 6). There was 63.91% identity between both superimposed structures. The score of root-mean-square deviation (RMSD) was 1.341 Å with sequence lengths of 437 (template) and 435 (target) amino acids. The values of structural distance measure (SDM) (with cutoff value of 5.0) and Q-score were found to be 26.431 and 0.814, respectively.

DISCUSSION

Citric acid has widespread applications in different fields including pharmaceutical industry. Chowdhury *et al.* (2009) used citric acid as a feed additive in broiler chicks instead of avilamycin that is an antibiotic growth promoter. They found that growth performance of broiler chicks was enhanced along with increase in their bone ash

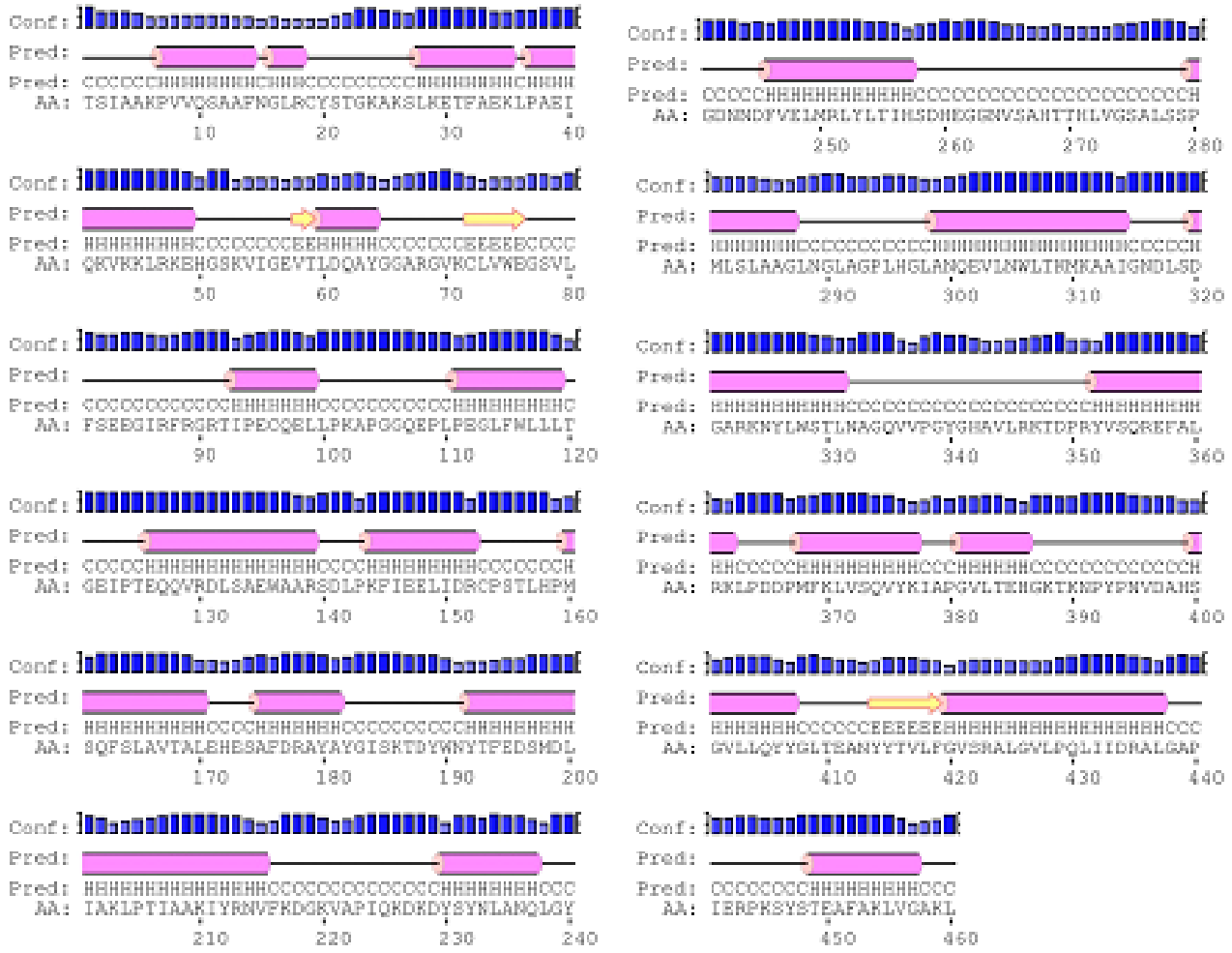

Fig. 4: PSIPRED graphical output of secondary structure elements of citrate synthase protein produced by PSIPRED View - a Java visualization tool that produces two-dimensional graphical representations of PSIPRED prediction. = Helix, = Strand, ——— = Coil, = Confidence of prediction, Pred = Predicted Secondary Structure, AA = Target Sequence

deposition. Healthy broilers were produced with higher immune status to combat against various infectious diseases and enteric pathogens. The use of citric acid as a supplementation in diet creates an acidic environment of pH 3.5 to 4.0 in the gut of animals that is a favorable environment for the growth of lactobacilli and also inhibits *Escherichia coli*, *Salmonella* and other Gram-negative bacteria to replicate in this environment (Denil *et al*., 2003).

The consensus pattern of predicted citrate synthase, C-terminal domain according to ScanProsite is G-[FYAV]-[GA]-H-x-[IV]-x(1,2)-[RKTQ]-x(2)-[DV]-[PS]-R where histidine (H) is an active site residue. The prokaryotic CS is composed of six identical subunits while eukaryotic CS is composed of two isozymes. One isozyme is present in the mitochondrial matrix and other is cytoplasmic. These isozymes seem to be dimers of identical chains (Karpusas *et al*., 1990). A leucine zipper explains how some regulatory proteins of eukaryotic genes work (Landschulz *et al*., 1988). There is a periodic repetition of leucine residues at every 7th position over eight helical turns. The consensus pattern of leucine zipper is L-x(6)-L-x(6)-L-x(6)-L according to ScanProsite. The periodic arrays of this zipper exist in α-helical conformations. A dimerization occurs because of interactions between α-helices of two polypeptides and the resultant structure is called as coiled coil. This pattern has been found in various regulatory proteins (Busch and Sassone-Corsi, 1990).

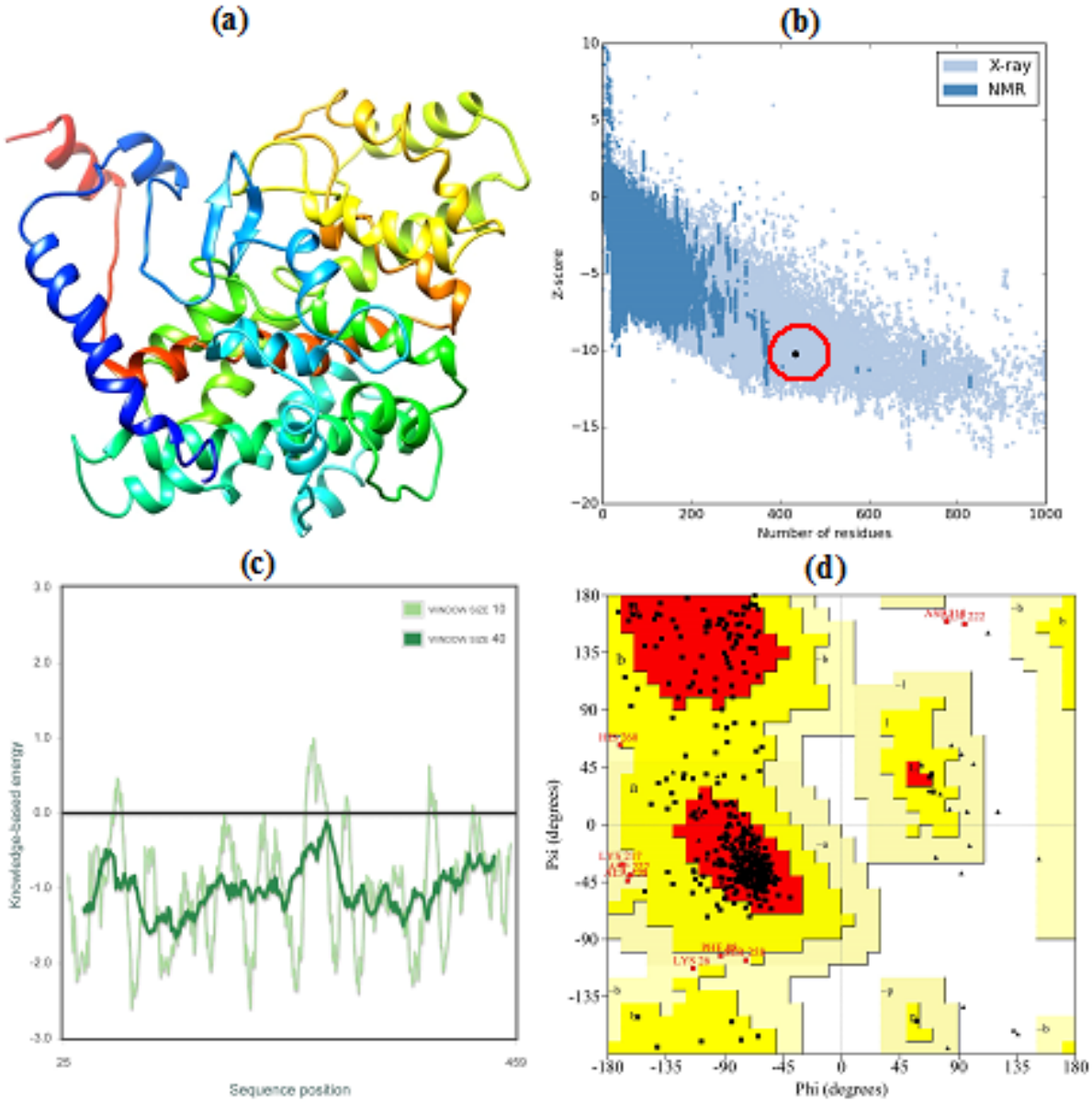

Fig. 5: Predicted 3D structure of CS and its evaluation. (a) 3D structure of citrate synthase, (b) z-score plot showing the quality of predicted model, (c) Energy plot showing the positions of all residues of predicted model, the plot is presented with window size 10 and 40, (d) Ramachandran plot analysis of the modeled structure.

A consensus sequence (i.e. Ser-Lys-Leu) was found at C-terminus that is known as C-terminal microbodies C-terminal targeting signal (CMTS). Peroxisomes, glyoxysomes and glycosomes are a class of small and single membraned organelles of microbodies. These proteins are imported post-translationally into organelle after their synthesis on free polysomes. All microbody proteins do not contain CMTS, some proteins have been found an internal CMTS-like sequence but their activity has not been studied. An N-terminal presequence has been found in few proteins that is cleaved off during their import (Gould *et al*., 1988). Two N-glycosylation sites were also predicted in the CS protein at positions 191-194 and 264-267. The consensus pattern of this site is N-{P}-[ST]-{P}. Consensus tripeptide is not enough to make a conclusion that asparagine is glycosylated. It has been found that protein folding is important for the regulation of N-glycosylation (Pless and Lennarz, 1977).

Idrees *et al*. (2012) have reported that cellular functions are mostly restricted in particular compartments. The prediction of subcellular localizations of novel proteins can therefore serve to predict functions and understanding about mechanisms. From sequencing projects, the influx of genomic data is arising and because of this the methods for efficient and automatic protein structure predictions

are becoming increasingly important. A reliable secondary structure prediction of protein is carried out by PSIPRED (McGuffin *et al*., 2000) which incorporates three methods (PSIPRED, GenTHREADER and MEMSAT 2) to predict information related to the protein using amino acid sequence. The PSIPRED is a highly accurate secondary structure prediction method and is incorporated by two feed-forward neural networks which perform the secondary structure prediction analysis on the output which is obtained from PSI-BLAST (Position Specific Iterated BLAST) (Altchul *et al*., 1997).

To predict 3D structure, homology modeling approach is used and considered as the most suitable one to build protein structure and generally a 30% sequence identity is required (Brooks *et al*., 2008). To judge the accuracy levels and suitability of the predicted model for the intended applications, the evaluation of predicted model is an essential step in the protein structure modeling. To estimate the errors in the predicted models, either the energy of the model or the resemblance of a given characteristic of model to real structures comes into account (Sippl, 1995). z-DOPE is an atomic distance-dependent statistical score and negative values of z-DOPE indicate better models (Shen and Sali, 2006). The GA341 score is used to check the reliability of a model that is derived from statistical potentials. A model is predicted to be reliable when its GA341 score is higher than a pre-specified cutoff that is equal to 0.7 (Melo *et al*., 2002). The z-score is used to check overall model quality through a plot in which groups of structures from X-ray and NMR are displayed by different colors (Wiederstein and Sippl, 2007).

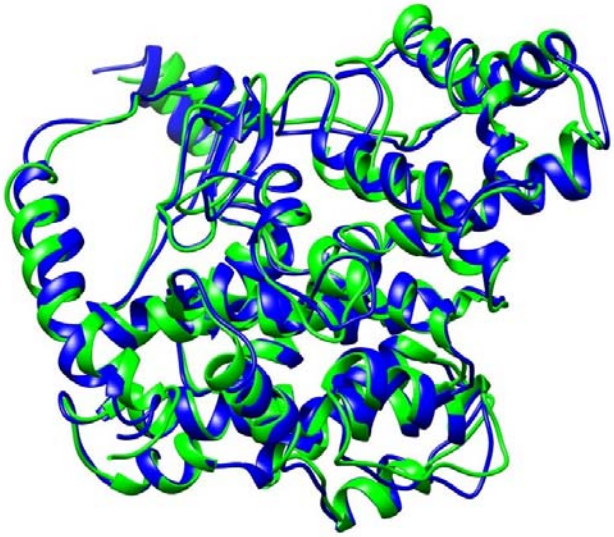

Fig. 6: Structure superimposition of CS (blue) with template (green)

Even if predicted 3D model and template structures superimposed perfectly, alone RMSD values are not meaningful as its lower values could be achieved simply by using less number of residues for calculations. Therefore, number of positions (alignment length) used for RMSD calculations should also be mentioned (Mustafa and Jamil, 2017). The values of SDM have been found linearly related with sequence-derived distances. As a rule, the value of SDM is zero for two identical structures and it gets increased with decrease in similarity (Johnson *et al*., 1990). For two well superposed structures, the Q-scores have been found to agree fairly well from various servers. Completely dissimilar or unsuperimposed structures have Q-scores of zero whereas identical structures have Q-score of 1 (Krissinel and Henrick, 2004).

CONCLUSION

An effort was made in this study to predict the 3D structure of citrate synthase from *A. niger* ANJ-120. This study provides simultaneous structural and functional annotation of CS, prediction and validation of 3D structure of CS protein from the fungus. The outcome of this study might provide a platform for simultaneous structural comparative analysis with CS proteins from other fungi and help in finding out variations in their structures to explore why *A. niger* is a good producer of citric acid. We predicted three dimensional structure of CS protein that will be useful in docking analysis which can be used to reveal various substrates and products of this enzyme. Further, the structural investigation of this protein would be helpful for protein engineering strategies and understanding the interactions between proteins in future.

ACKNOWLEDGEMENT

The work was supported by a grant from Higher Education Commission, Islamabad, Government of Pakistan.

REFERENCES

Baker SE (2006). *Aspergillus niger* genomics: Past, present and into the future. *Medical Mycology*. **44**: S17-S21.

Barkan D, Hostetter D, Mahrus S, Pieper U, Wells J, Craik C and Sali A (2010). Prediction of protease substrates using sequence and structure features. *Bioinformatics*., **26**: 1714-1722.

Benkert P, Tosatto SCE and Schomburg D (2008). QMEAN: A comprehensive scoring function for model quality assessment. *Proteins: Struct. Funct. Bioinf.* **71**(1): 261-277.

Bjellqvist B, Hughes GJ, Pasquali Ch, Paquet N, Ravier F, Sanchez JCh, Frutiger S and Hochstrasser DF (1993). The focusing positions of polypeptides in immobilized pH gradients can be predicted from their

amino acid sequences. *Electrophoresis*., **14**: 1023-1031.

Brooks WH, Daniel KG, Sung SS and Guida WC (2008). Computational validation of the importance of absolute stereochemistry in virtual screening. *J. Chem. Inf. Model*. **48**: 639-645.

Burge S, Kelly E, Lonsdale D, Mutowo-Muellenet P, McAnulla C, Mitchell A, Sangrador-Vegas A, Yong S, Mulder N and Hunter S (2012). Manual GO annotation of predictive protein signatures: the InterPro approach to GO curation.

Busch SJ and Sassone-Corsi P (1990). Dimers, leucine zippers and DNA-binding domains. *Trends Genet*. **6**: 36-40.

Chou PY and Fasman UD (1974). Prediction of protein conformation. *Biochemistry*. **13**: 211-215.

Chowdhury R, Islam KMS, Khan MJ, Karim MR, Haque MN, Khatun M and Pesti GM (2009). Effect of citric acid, avilamycin, and their combination on the performance, tibia ash, and immune status of broilers. *Poultry Sci*. **88**: 1616–1622.

Currie JN (1917). The citric acid fermentation of *A. niger*. *J. Biol. Chem*., **31**: 5.

Denil M, Okan F and Celik K (2003). Effect of dietary probiotic, organic acid and antibiotic supplementation to diets on broiler performance and carcass yield. *Pak. J. Nutr*. **2**: 89–91.

Garnier J, Gibrat JF and Robson B (1996). GOR method for predicting protein secondary structure from amino acid sequence. *Methods Enzymol*. **266**: 540-553.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD and Bairoch A (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31**(13): 3784-3788.

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD and Bairoch A (2005). Protein identification and analysis tools on the ExPASy server; (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press. pp.571-607.

Gould SJ, Keller GA and Subramani S (1988). Identification of peroxisomal targeting signals located at the carboxy terminus of four peroxisomal proteins. *J. Cell Biol*., **107**(3): 897-905.

Johnson MS, Sutcliffe MJ and Blundell TL (1990). Molecular anatomy: phyletic relationships derived from three-dimensional structures of proteins. *J. Mol. Evol*. **30**(1): 43-59.

Karpusas M, Branchaud B and Remington SJ (1990). Proposed mechanism for the condensation reaction of citrate synthase: 1.9-. ANG. structure of the ternary complex with oxaloacetate and carboxymethyl coenzyme A. *Biochemistry,* **29**(9): 2213-2219.

Krieger E, Nabuurs SB and Vriend G (2003). Homology modeling. *Methods Biochem. Anal*. **44**: 509-523.

Krissinel E and Henrick K (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr*. **60**(12): 2256-2268.

Lancini G (2008). Parte I L'uso industriale dei microrganismi. Storia e campi di applicazione. In: Donadio S; Marino G (eds.). Biotecnologie Microbiche, Casa Editrice Ambrosiana, Milan, pp.5-35.

Landschulz WH, Johnson PF and McKnight SL (1988). The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science*, **240**: 1759-1765.

Laskowski RA, MacArthur MW, Moss DS and Thornton JM (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *J. App. Cryst*., **26**: 283-291.

Lovell SC, Davis IW, Arendall WB, de Bakker PIW, Word JM, Prisant MG, Richardson JS and Richardson DC (2002). Structure validation by Calpha geometry: Phi,psi and Cbeta deviation. *Protein Struct. Funct. Genet*. **50**: 437-450.

Luthy R, Bowie JU and Eisenberg D (1992). Assessment of protein models with three-dimensional profiles. *Nature*., **356**: 83-85.

Magnuson J and Lasure L (2004). Organic acid production by filamentous fungi. *In*: Tkacz J, Lange L (eds). Advances in Fungal Biotechnology for Industry, Agriculture, and Medicine. New York: Kluwer Academic & Plenum Publishers. pp.307-340.

Marchler-Bauer A, Anderson JB, Cherukuri PF, DeWeese-Scott C, Geer LY, Gwadz M, He S, Hurwitz DI, Jackson JD and Ke Z *et al* (2005). CDD: A Conserved Domain Database for protein classification. *Nucleic Acids Res*. **33**: D192-D196.

Max B, Salgado JM, Rodríguez N, Cortés S, Converti A and Domínguez JM (2010). Biotechnological production of citric acid. *Braz. J. Microbiol*., **41**: 862-875.

McGuffin LJ, Bryson K and Jones DT (2000). The PSIPRED protein structure prediction server. *Bioinformatics*. **16**: 404-405.

Melo F, Sánchez R and Sali A (2002). Statistical potentials for fold assessment. *Protein science*. **11**: 430-448.

Mustafa G and Jamil A (2017). Comparative analyses of gene clusters and Ksα genes involved in the biosynthesis of chromomycin A_3 and mithramycin. *Indian J. Pharm. Sci*., **79**: 707-714.

Mustafa G, Iqbal MJ, Hassan M and Jamil A (2017). Bioinformatics characterization of growth differentiation factor 11 of *Oryctolagus cuniculus*. *J. Chem. Soc. Pak*. **39**: 1089-1095.

Mustafa G, Tahir A, Asgher M, Rahman MU and Jamil A (2014). Comparative sequence analysis of citrate synthase and 18S ribosomal DNA from a wild and mutant strains of *Aspergillus niger* with various fungi. *Bioinformation*. **10**(1): 1-7.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC and Ferrin TE (2004). UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem*., **25**: 1605-1612.

Pless DD and Lennarz WJ (1977). Enzymatic conversion of proteins to glycoproteins. *Proc. Natl. Acad. Sci*., **74**: 134-138.

Ruijter GJG, Kubicek CP and Visser J (2002). Production of organic acids by fungi. in The Mycota X Industrial Applications (ed. Osiewacz, H.D.) 213-230, (Springer, Berlin Heidelberg).

Schultz J, Milpetz F, Bork P and Ponting CP (1998). SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc. Natl. Acad. Sci. USA*. **95**: 5857-5864.

Shen MY and Sali A (2006). Statistical potential for assessment and prediction of protein structures. *Protein Science*. **15**: 2507-2524.

Sippl MJ (1995). Knowledge-based potentials for proteins. *Curr. Opin. Struct. Biol*., **5**: 229-235.

Solovyev V, Kosarev P, Seledsov I and Vorobyev D (2006). Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol*. **7**: S10.

Sonnhammer ELL, Eddy SR and Durbin R (1997). Pfam: A comprehensive database of protein domain families based on seed alignments. *Proteins*. **28**: 405-420.

Webb B and Sali A (2014). Comparative protein structure modeling using Modeller. Current Protocols in Bioinformatics, John Wiley & Sons, Inc., 5.6.1-5.6.32.

Wehmer C (1893). Note sur la fermentation citrique. *Bull. Soc. Chem. Fr*., **9**: 728.

Wiederstein M and Sippl MJ (2007). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res*., **35**: W407-W410.

Yusim KSJ, Honeyborne I, Calef C, Goulder PJ and Korber BT (2003). Enhanced motif scan: A tool to scan for HLA anchor residues in proteins. HIV Immunology and HIV/SIV Vaccine Databases, pp.25-36.