Available online at www.ijpab.com

DOI: http://dx.doi.org/10.18782/2320-7051.2255

ISSN: 2320 – 7051

Int. J. Pure App. Biosci. **4 (2):** 71-77 (2016)

International Journal of Pure & Applied Bioscience

Research Article

OPEN ACCESS

In-Silico Structural Modelling of Transaldolase from *Helicobacter pylori* (Strain G27) A Class I Transaldolase

Rabiu Salihu[1,2]*, Ismail Haruna[1,3], Hassana Abubakar[1,4], Mohd Shahir Shamsir[1] and Sepideh Parvizpour[1]

[1]Faculty of Biosciences and Medical Engineering, University Technology Malaysia, 81310 Skudai, Johor, Malaysia

[2]Department of Biological Sciences, Federal University Dutse, PMB 7156 Dutse Jigawa State, Nigeria

[3]Department of Microbiology, Bauchi State University Gadau, PMB 65 Gadau Bauchi State, Nigeria

[4]Department of Biochemistry, Ibrahim Badamasi Babangida University Lapai, PMB 11 Niger State, Nigeria

*Corresponding Author E-mail: salihu.r@fud.edu.ng

Received: 29.03.2016 | Revised: 10.04.2016 | Accepted: 12.04.2016

ABSTRACT

Numerous biotechnological applications of enzymes including rational drug design, turbidity reducing in the brewing industry and animal feeds digestibility enhancement have been recorded. An enzyme Transaldolase from Helicobacter pylori G27 (an important causative agent for gastric ulcers worldwide) plays an important role in NADPH synthesis regulation. However, an experimental structure of this enzyme is yet to be known. In this paper, we modelled an in-silico 3-Dimentional structure of a Transaldolase from H. pylori G27 using several Bioinformatics tools. Reliability and precision of the modelled structure were also evaluated. This work is anticipated to be found useful in reviewing structure/function relationship as well as rational drug design for the treatment of gastric ulcers.

Key words: *In-silico, Transaldolase, Helicobacter pylori*

INTRODUCTION

Helicobacter pylori, a gram-negative pathogen causing different gastric infirmities including gastritis, ulcers, and gastric carcinoma. It colonizes the stomachs of over half the world's population[4]. *H. pylori* has long been categorized as “Class I carcinogen” by the World Health Organisation (WHO) as prolong colonization by this organism can aggravate chronic gastritis leading to gastric atrophy, metaplasia, and dysplasia[14]. The organism is one of the most important causative agent for gastric cancer and the second leading cause of cancer death worldwide[18].

Transaldolase from *H. pylori*, Uniprot ID (B5Z9B4) is an enzyme belonging to the class I transaldolases (EC: 2.2.1.2) that plays an important role in non-oxidative pentose phosphate pathway (PPP) of central carbon metabolism by employing a Schiff-base mechanism for stabilization of the reaction intermediates[16]. Pentose phosphate pathway for glucose-6-phosphate metabolism was described to contain three distinct enzyme systems.

Cite this article: Salihu, R., Haruna, I., Abubakar, H., Shamsir, M.S. and Parvizpour, S., In-Silico Structural Modelling of Transaldolase from *Helicobacter pylori* (Strain G27) A Class I Transaldolase, *Int. J. Pure App. Biosci.* **4(2):** 71-77 (2016). doi: http://dx.doi.org/10.18782/2320-7051.2255

Members of the class I aldolases are the ubiquitous transaldolases, which catalyze the reversible transfer of a dihydroxyacetone moiety, derived from fructose-6-phosphate to erythrose-4-phosphate yielding sedoheptulose-7-phosphate and glyceraldehyde-3-phosphate. The enzyme is found in the non-oxidative branch of the pentose phosphate pathway, which is important for generation of NADPH, the reducing power in the cell[24]. As other class I aldolases, transaldolases fold into a β_8 barrel structure. A shared feature in these aldolases is a charged aliphatic amino acid residue (Lysine) located on strand β_6 of the barrel, forming a Schiff base reaction with substrate in the active site[23]. In transaldolases however, this lysine residue is positioned on strand β_4 of the α/β barrel and transaldolases appear to be related to the other class I aldolases by a circular permutation. Replacement of this lysine residue by site-directed mutagenesis in yeast and human transaldolase resulted in loss of catalytic activity[3].

3-Dimentional structures of transaldolases from different organisms were said to be available as reported by Soderberg and Alver[21]. However, to our knowledge no 3-Dimentional structure is available for this enzyme from *Helicobacter pylori*. In this paper therefore, we present an in-silico 3D structure of a transaldolase from *H. pylori* strain G27 as well as validatory results from different servers.

MATERIALS AND METHODS

Sequence retrieval and Analysis

The amino acid sequence (316 residues) of transaldolase from *Helicobacter pylori* strain G27 with accession number: B5Z9B4 was retrieved from UniProtKB[4] in a FASTA format. Running a BLASTp using NCBI[1] shows a very low identity of (<40%) with many other entries in the PDB.

Homology modeling

The amino acid sequence was submitted to I-TASSER[25], Swiss model[2], Phyre2[11] and RaptorX[10] servers for prediction of a 3D model. PDB output file of all the predicted models from the above servers were generated together with some potential templates. Structural alignment and superimposition for templates and the predicted models was performed using Pymol[6].

RESULTS

Sequence analysis

To analyze the *H. pylori* (G27) sequence obtained from the Uniprot database, NCBI Blastp was used to run a similarity search which shows a similarity of (<97%) and low identity of (<40%) with Transaldolase from *Neiserria gonorrhoeae* (3CLM) in PDB database[13]. Blastp also shows the specific hits, super family hits as well as the putative conserved domains[15]. Running a multiple sequence alignment between the query, 3CLM, 3R5E, 3HJZ and 1WXO sequences using Clustal Omega[20] confirms the putative conserved residues as well as the active site residues (Figure 1).

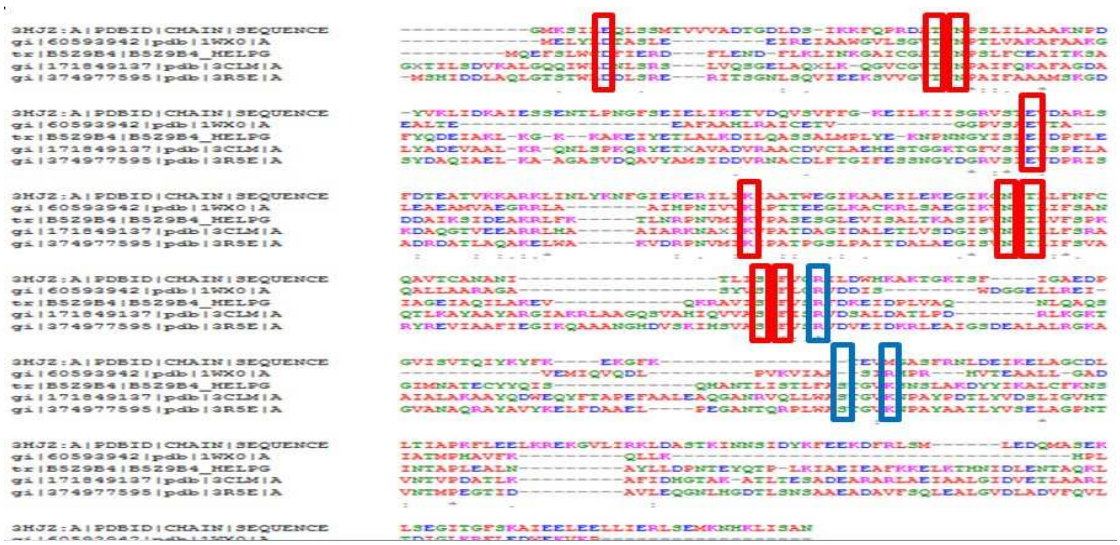


Fig. 1: Multiple sequence alignment between the query, 3CLM, 3R5E, 3HJZ and 1WXO sequences showing the conserved residues including active site (Red) and phosphate binding site (Blue).

Template selection

3CLM appeared to be the suitable template from the NCBI result with ~97% similarity. To further validate the template selection, the sequence was submitted to I-TASSER[25], HHpred[22], Phyre2[11], mGenthreader[9] and PSI-Blast[1] tools and different closest templates were obtained, among which the suitable (3CLM) was selected based on similarity and e-value as well as its commonest appearance in all the 5 folding libraries as in (Table 1).

Table 1 Showing templates from different threading tools among which 3CLM was selected based on its low identity and e-value as well as commonest appearance in all the threading tools

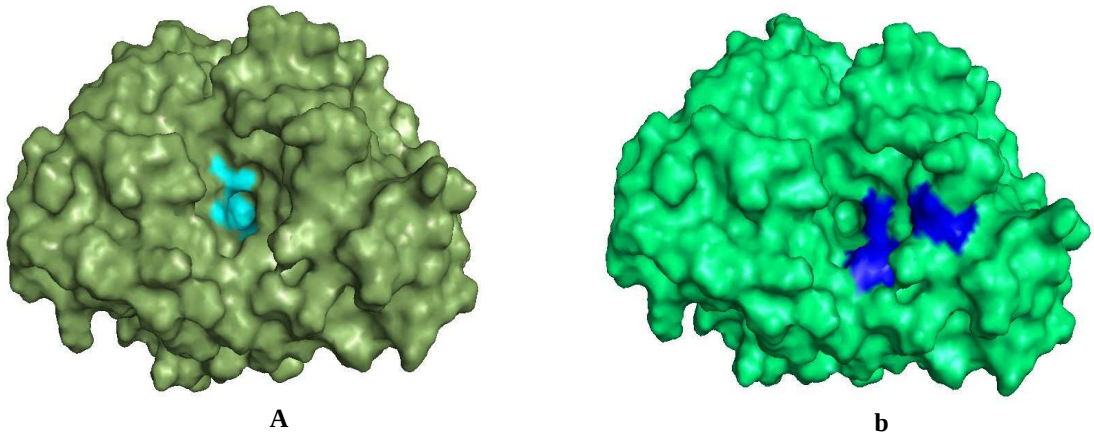
Server name	Template PDB ID	Protein name	Organism	Identity	e-value
Phyre2	3R5E	Transaldolase	*C. glutamicum*	31%	NA
	3CLM			32%	NA
	3HJZ			20%	NA
mGenthreader	3CLM	Transaldolase	*N. gonorrhoeae*	NA	
	3R5E			NA	
	3CWN			NA	
I-TASSER	3CLM			NA	NA
	3R5E	Transaldolase B	*E. coli*	NA	NA
	3HJZ			NA	NA
HHpred	3CLM			32%	1.8e-85
	3R5E			32%	1.3e-84
	3HJZ	Transaldolase B	*P. marinus*	21%	1e-69
PSI-Blast	3CLM			31%	3e-45
	3R5E			31%	2e-39
	1WXO	Transaldolase	*T. thermophiles*	28%	4e-05

Homology modelling

To build a 3D model for the query sequence, the sequence was submitted to I-TASSER, RaptorX, Swiss-model and Phyre2. The predicted models were also submitted to RAMPAGE for Ramachandran plot (Table 2). The 3D model with highest Ramachandran plot score (RaptorX) was selected among the three (Figure 2). Multiple sequence alignment was performed using Clustal Omega[20] to further ratify conserved regions (active site and phosphate binding site residues). Secondary structure alignment was also done using Phyre2 to compare the structural motifs[11].

Table 2 Models from different servers compared based on Ramachandran plot analysis

Server name	No. of residues in favored region (%)	No. of residues in allowed region (%)	No. of residues in outlier region (%)
RAPTORX	298 (94.9%)	10 (3.2%)	6 (1.9%)
SWISS - MODEL	287 (92.9%)	14 (4.5%)	8 (2.6%)
PHYRE2	292 (94.8%)	11 (3.6%)	5 (1.6%)
YASARA	305 (97.1%)	8 (2.5%)	1 (0.3%)



A **b**

Fig. 2: 3D surface structure of Transaldolase showing (a) the active site pocket (Cyan color), (b) the phosphate biding site (Blue color)

Model validation

RaptorX predicted model with highest Ramachandran score (number of residues in allowed region) was selected. To further refine its quality the selected model was subjected to YASARA energy minimization server. The final model from YASARA was resubmitted to RAMPAGE for Ramachandran[12] analysis to check the model quality improvement when the energy has been reduced. VERIFY3D[7] and ERRAT[5] servers were used to further validate the predicted model (Figure 3).

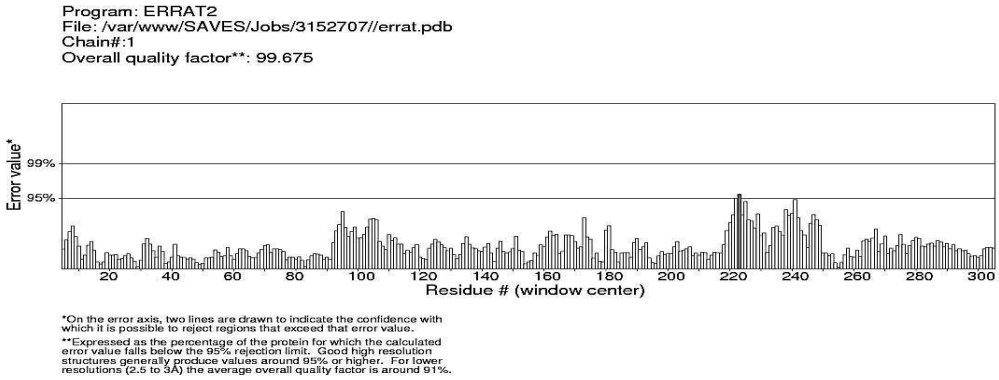
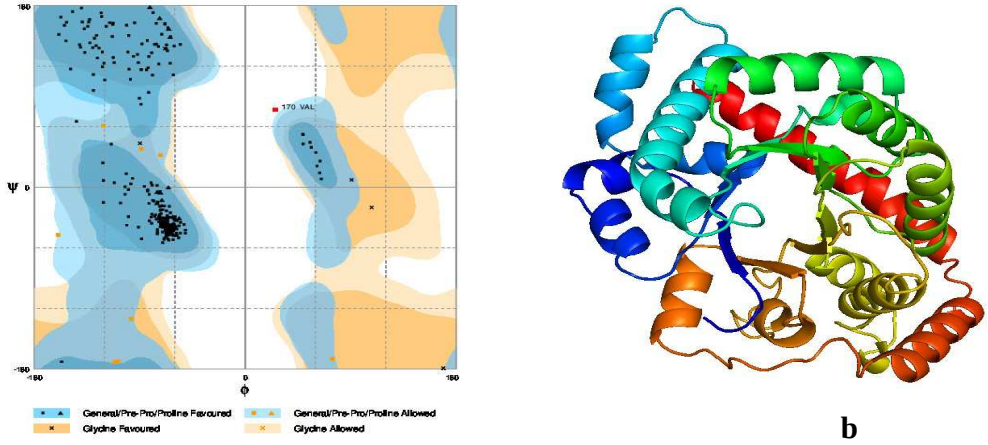


Fig. 3: ERRAT result showing the overall quality factor of the model

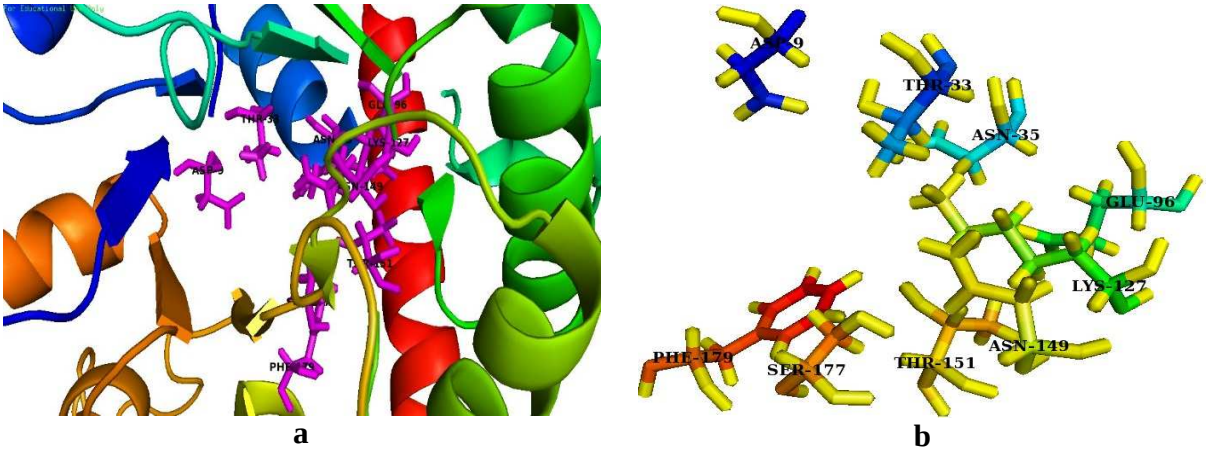


a **b**

Fig. 4: (a) Ramachandran plot showing the number of residues in favoured, allowed and outlier regions (97.1%, 2.5% and 0.3%) respectively and (b) A predicted 3D strucuture of a Transaldolase from *H. pylori* strain (GH27)

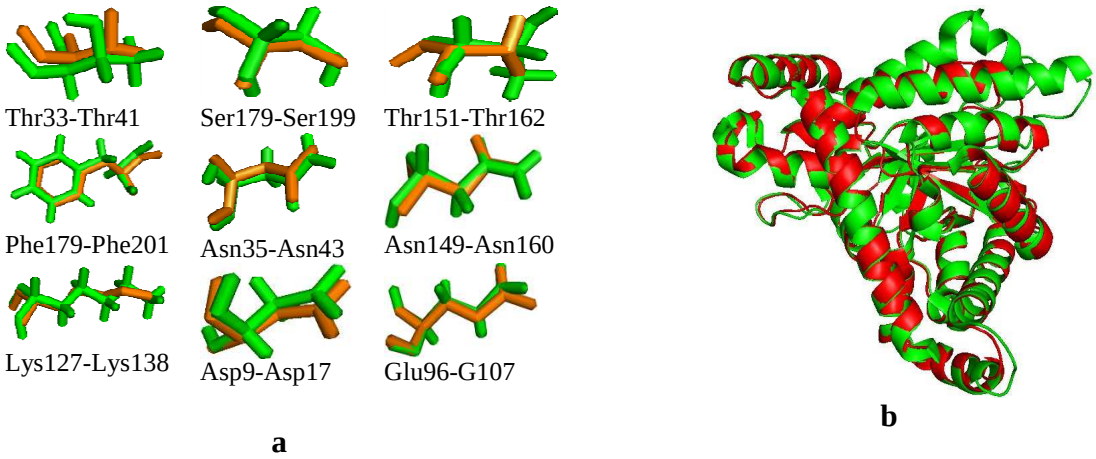
DISCUSSION

From the results presented above, the similarity search using NCBI shows that the query sequence belongs to the family of Transaldolase I. Members of the family plays a role in central carbon metabolism specifically the non-oxidative pentose phosphate pathway. Although there were paucity of literature on the template transaldolase from *Neiseria gonnrrhoeae* (3CLM) about the active site residues, transaldolase from *E. coli* and *Corynebacterium glutamicum* were used to identify the conserved residues since members of this family have well conserved sequences. But when superimposition of the model and template (3CLM) was made using Pymol, all the conserved residues were found to be well aligned. Lysine acting as nucleophillic residue to attack the carbonyl group of the fructose-6-phosphate is highly conserved in this family[23] and Miosga, *et al.*,[17]. The active site residues for the transaldolase from *H. pylori*, and (*Neisseria gonorrhoeae*) in bracket were Asp9 (Asp17), Thr33 (Thr41), Asn35 (Asn43), Glu96 (Glu107), Lys127 (Lys138), Asn149 (Asn160), Thr151, (Thr162), Ser177 (Ser199) and Phe179 (Phe201). The phosphate binding site residues were also Arg182 (Arg204), Ser225 (Ser259), but Argenine and Lysine found in *E. coli* and *C. glutamicum* respectively were absent in both our model and the template.



a **b**

Fig. 5: Showing the active site residues (Asp9, Thr33, Asn35, Glu96, Lys127, Asn149, Thr151, Ser177 and Phe179)



a **b**

Fig. 6: (a) superimposition of catalytic residues for model (green) and template (orange) **(b)** 3D structure of superimposed model (red) and template (green)

Fig. 7: Secondary structure alignment between the model and template (3CLM) predicted by Phyre2

The structural motif of this protein was found to be in consistent with $(\beta/\alpha)_8$-barrel fold, as it is well conserved between the different subfamilies though different degrees of oligomerization. The active site conservation, the phosphate binding site as well as the position of an important catalytic residue i.e. Lysine that was positioned on strand β_4 of the α/β barrel were also in consistent with the family members[19]. Pylogenetic analysis of the sequence shows the closeness of TalHp with the template (3CLM) in evolutionary relationship (Figure 8) though there is a slight divergence in the tree, but they all belongs to class I transaldolase family[23].

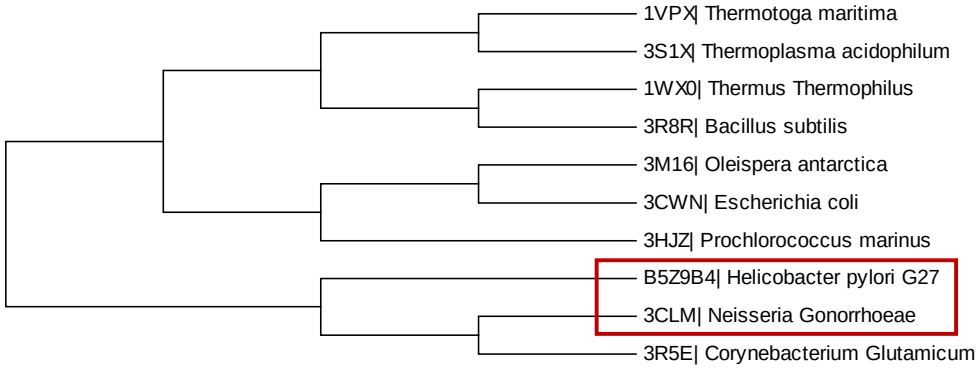


Fig. 8: Pylogenetic tree showing the closest family members

Assessment of quality profiles of a good including geometry of the backbone conformation, energy profiles and correctness of residues interactions was done using Ramachandran plot, YASARA, ERRAT and VERIFY3D. Ramachandran plot analysis done to assess the stereochemical properties of the predicted model was very interesting with almost all the residues falling in favoured and allowed region (97.1% and 2.5%) respectively. This in essence shows a very good arrangement. As most of proteins in nature are known to have lower energy profile for flexibility and function[8], the predicted model was also submitted to YASARA for energy minimization to simulate the natural conformation. The end model from YASARA was found to have less energy than the start model i.e. energy reduces drastically (-25232.4 to -171784.1 kJ/mol). ERRAT analysis for the backbone conformation of the model was extreemly high. The accepted score range is >50% and the higher the score the higher the

quality of a modelled structure[5]. The predicted model has a score of 99.675% which shows that the structural conformation is very efficient.

To further check the efficiency of the model, it was subjected to VERIFY3D server. The result shows that 95.89% of the residues have an average score of 3D/1D>=0.2. Since atleast 85% of the residues in the structure must have a score of >=0.2[7], our result indicates very high score and this implies the structural accuracy of the predicted model.

CONCLUSION

In-silico prediction of three-dimensional structure (3D) structure of a class I transaldolase using different bioinformatics approachs will in future provide to proteomics community a useful information that can serve as tool for rational designing of proteins from the same or different family sharing the same structural motifs. Interestingly, the predicted structure validation using some validation servers also yields a very encouraging results signifying the closeness of the predicted sructure to the crystallogarphic type.

REFERENCES

1. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. *Nucleic acids research.*, **25:** 3389-3402 (1997).
2. Arnold, K., Bordoli, L., Kopp, J. and Schwede, T., *Bioinformatics*., **22:** 195-201 (2006).
3. Banki, K. and Perl, A., *FEBS letters*., **378:** 161-165 (1996).
4. Baltrus, D. A., Amieva, M. R., Covacci, A., Lowe, T. M., Merrell, D. S., Ottemann, K. M., Stein, M., Salama, N. R. and Guillemin, K., *Journal of bacteriology*., **191:** 447-448 (2009).
5. Colovos, C. and Yeates, T. O., *Protein science.*, **2:** 1511 (1993).
6. DeLano, W. L., The PyMOL molecular graphics system (2002).
7. Eisenberg, D., Lüthy, R. and Bowie, J. U., *Methods in enzymology*.,**277:** 396-404 (1997).
8. Hendlich, M., Lackner, P., Weitckus, S., Floeckner, H., Froschauer, R., Gottsbacher, K., Casari, G. and Sippl, M. J., *Journal of molecular biology.*,**216:** 167-180 (1990).
9. Jones, D. T., *Journal of molecular biology*., **287:** 797-815 (1999).
10. Källberg, M., Margaryan, G., Wang, S., Ma, J. and Xu, J., *Protein Structure Prediction.*, 17-27 (2014).
11. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. and Sternberg, M. J., *Nature protocols*., **10:** 845-858 (2015).
12. Kleywegt, G. J. and Jones, T. A., *Structure*., **4:** 1395-1400 (1996).
13. Kouranov, A., Xie, L., de la Cruz, J., Chen, L., Westbrook, J., Bourne, P. E. and Berman, H. M., *Nucleic acids research*., **34:** 302-305 (2006).
14. Lee, Y.-C., Chen, T. H.-H., Chiu, H.-M., Shun, C.-T., Chiang, H., Liu, T.-Y., Wu, M.-S. and Lin, J.-T., *Gut*., **62:** 67-682 (2013).
15. Marchler-Bauer, A., Anderson, J. B., Cherukuri, P. F., DeWeese-Scott, C., Geer, L. Y., Gwadz, M., He, S., Hurwitz, D. I., Jackson, J. D. and Ke, Z., *Nucleic acids research*., **33:** 192-196 (2005).
16. Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., Fong, J. H., Geer, L. Y., Geer, R. C. and Gonzales, N. R., *Nucleic acids research*., **39:** 225-229 (2011).
17. Miosga, T., Schaaff-Gerstenschläger, I., Franken, E. and Zimmermann, F. K., *Yeast*., **9:** 1241-1249 (1993).
18. Parkin, D. M., Bray, F., Ferlay, J. and Pisani, P., *CA*., **55:** 74-108 (2005).
19. Samland, A. K., Baier, S., Schürmann, M., Inoue, T., Huf, S., Schneider, G., Sprenger, G. A. and Sandalova, T., *FEBS Journal*., **279:** 766-778 (2012).
20. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M. and Söding, J., *Molecular systems biology*., **7:** 539 (2011).
21. Soderberg, T. and Alver, R. C., *Archaea*., **1:** 255-262 (2004).
22. Söding, J., Biegert, A. and Lupas, A. N., *Nucleic acids research*., **33:** 244-248 (2005).
23. Thorell, S., Schürmann, M., Sprenger, G. A. and Schneider, G., *Journal of molecular biology.*,**319:** 161-171 (2002).
24. Wamelink, M., Struys, E. and Jakobs, C., *Journal of inherited metabolic disease*., **31:** 703-717 (2008).
25. Zhang, Y., *BMC bioinformatics.*,**9:** 40 (2008).