COMPUTATIONAL ANALYSIS AND PROTEIN STRUCTURE MODELING OF ZINC- AND IRON- DEPENDENT ALCOHOL DEHYDROGENASES FOR BIOETHANOL YIELD IN THERMOPHILES

ZAMEER, M.[1*] – RAFIQUE, A.[1] – ALYAS, S.[1] – FARRUKH, S. Y.[1] – TAHIR, U.[1] – ZAHRA, N.[1] – ALI, Q.[2*] – KHAN, M. T.[1] – MAZHAR, M.[1] – RUKHSAR, A.[1] – QURESHI, J. A.[1] – SHABBIR, A.[1] – SATTAR, H.[1] – SOHAIL, A.[1]

[1]*Institute of Molecular Biology and Biotechnology, The University of Lahore, Lahore, Pakistan*

[2]*Department of Plant Breeding and Genetics, Faculty of Agricultural Sciences, University of the Punjab, Lahore, Pakistan*

[*]*Corresponding authors*
e-mail: mariam.zameer@imbb.uol.edu.pk, saim1692@gmail.com

(Received 21st Jul 2022; accepted 11th Nov 2022)

Abstract. Climate change, dramatic fluctuations in the cost of fuel can cause an increase in the prices of basic food items. Pakistan is undergoing a serious energy crisis. Fortunately, this degrading situation is countered with the help of biomass energy (bioethanol) invention. However, the selection of the best strains with efficient enzymes for the production of bioethanol is in a continuous process of improvement. The objective of this work is to compare and analyze the structure of extremophilic strains with currently used organisms. Protein sequences of two different strains both zinc- and iron-dependent alcohol dehydrogenases (ADHs) were retrieved from Uniprot. Sequence analysis using protein sequence was conducted separately for both Zn- and Fe strains with *Saccharomycescervisae* and *Zymomonasmobilis* respectively with the help of T-coffee and Aliview. 3D protein structure modelswere built using the swiss model. The accuracy of predicted models was confirmed through online servers molprobity, ERRAT score VERIFY 3D values, and VADAR. Upon acquiring complete structural information of target strains, mutations were carried out in Fe and Zn-based organisms to check the best conformation of the enzyme for maximum bioethanol yield. The study demonstrated that the ability to withstand high-temperature ADH of thermophilic bacteria is considered a key enzyme for bioethanol production. Metal-binding residue sites His-62, Cys-38, Glu-63, and Cys-154 in Zn strains and Asp-204, His-208, His-271, and His-285 in Fe strains was conserved. Ligand interaction with residues in Fe and Zn strains was found similar in *Saccharomycescervisae* and *Zymomonasmobilis* strains. Structural and mutational analyses were performed to interpret the significance of mutated residues (A62, A38, Q63, A154 in *Sulfolobus acidoculdarius* and E204, R208, R271, R285 in *Archaeglobus fulgidus*) at the active binding site of ADH. Mutations at the active binding site in both Fe and Zn dependent ADH will impair structural stability, disturbance in catalytic pockets and bonds hence subordinating the overall enzyme-substrate complex and decreased efficiency in binding lignocellulosic biomass as substrate.
Keywords: *alcohol dehydrogenases, bioethanol, thermophiles, protein structure prediction, binding pocket*

Introduction

Hyperthermophiles are possible sources of thermostable enzymes such as alcohol dehydrogenases (ADHs). However, the metabolic mechanism and enzymes involved in these hyperthermophiles' alcohol development were unclear. ADH is a central enzyme in the metabolism of alcohol. Alcohol converts into aldehydes or ketones with the catalyzation of the enzyme ADH that is found everywhere in the world (Machielsen et al., 2006). The characterization study of ADHs from hyperthermophiles is still needed

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2102_925941

attention. It has been observed that iron and zinc-dependent ADHs are important in bioethanol production (Montella et al., 2005).

Bioethanol is the most frequently used biofuel as a fossil fuel substitute. It is primarily based on cellulose hydrolysis by the lignocellulosic biomass and sugar fermentation from different sources. Ethanol production will increase when glucose fermented using robust industrial hosts (Deng et al., 2018). Due to its high specific ethanol productivity, two strains of *Zymomonasmobilis* and *Saccharomyces cerevisiae* are used as a reference strains. Various thermophilic and hyperthermophilic bacteria and archaea have been identified as ADHs (*A. fulgidus* and *S. acidoculdarius*). The thermophilic ADH from *Sulfolobus acidoculdarius* has just been identified and is 352 amino acids long. The origin of *A. fulgidus*from submarine hot spring has Fe-dependent ADH, and it is 395 amino acids longemzyme (Radianingtyas and Wright, 2003). *S. acidoculdarius* (Zn dependent ADH) can survive in a variety of environments (volcanic area, Sulphur springs, hot springs, etc.) (Ammendola Raia et al., 2012). The iron and zinc-dependent ADHs contain 395 and 352 amino acids but few residues (Asp-204, His-208, His-285, His271 and Cys-38, His-62, Glu-63, and Cys-154, respectively) are also involved in binding pocket formation.

The structure and thermostability of a protein can be substantially altered by mutations at particular residues (Guy et al., 2003). The majority of methods for predicting mutation-induced changes in protein stability focus on a small number of mutations in a protein, often at a single site (Dehouck et al., 2009). As a result, multi-site mutations can be expected to have a larger and more complex impact on protein thermostability than single-site mutations alone (Tan et al., 2013).

The current study focuses on extremophile alcohol dehydrogenases that are zinc and iron-dependent (thermophiles & halophiles). The comparisons between zinc-based alcohol dehydrogenases in the *S. cervisiae* strain and iron-based alcohol dehydrogenases in the *zymomonas mobilis* strain were the main motivations for the computational investigation. Furthermore, certain comparisons were made to assess these strains' potential to create bioethanol. Mutations in diverse strains were carried out in later phases to confirm their functional features.

Computational methodology

Retrieval of zinc and iron-dependent ADHs sequences

Sulfolobus acidoculdarius (*S. acidoculdarius*) and *Archaeoglobus Fulgidus* (*A. Fulgidus*) protein sequences were retrieved from Uniprot Knowledgebase (UniProtKB; https://www.uniprot.org/help/uniprotkb) having accession number Q4J6S2 and PODJA2, respectively in FASTA format. Furthermore, both retrieved protein sequences were aligned with the reference sequence of *S. Cervisiae* and *Z. mobilis* to check the conservation pattern of amino acids. The alignment was done by T-Coffee (https://www.ebi.ac.uk/Tools/msa/tcoffee/) and visualized by AliView, respectively.

Protein structure predictions of wild and mutant

S. acidoculdarius and *A. fulgidus* homology modeling is the most efficient model approachto the three-dimensional structure of unknown proteins, because the structure information of these proteins is missing through crystallography data (Deng et al., 2018). The protein structure of target strains was not present in Protein Data Bank (http://www.rcsb.org/). Therefore, the homology modeling-based approach was employed during swiss modeling for the strain Zn and Fe-dependent *S. acidoculdarius*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2102_925941

and *A. fulgidus*, respectively. Out of 50 templates, two different templates (S-hydroxymethyl) glutathione dehydrogenase from *C. reinhardtii* and ADH from *T. thioreducens*) were selected based on sequence similarity (32.16%, 36.98%) having PDB IDs 7AAS and 6C76, respectively (Montella et al., 2005). Similarly, mutated models of *S. acidoculdarius* and *A. fulgidus* proteins were generated by changing residues at different positions. For, Zn-dependent ADH different substitutions (H62A, C38A, C154A, and E63Q) were induced in wild-type protein sequences and predicted models using the swiss modeling approach. Similarly, the mutated model of *A. fulgidus* was also built from the swiss-model by the substitution of residues at metal-binding sites D204E, H208R, H285R, H271R, respectively (Moon et al., 2011). The predicted models of both wild and mutant proteins further undergo structure analysis through the Molprobity webserver (http://molprobity.biochem.duke.edu/). The ProtParam tool, which assesses the physiochemical characteristics of protein sequences, has been used to predict their extinction coefficients, theoretical PIs, instability indexes, GRAVY values, and aliphatic. Moreover, the MolProbity server was also used for Ramachandran plots and values. The SAVES server was used to verify the protein structure and another server was also used for validation such as (Analysis through Ramachandran Plot, PROVE, Verify3D, and ERRAT) after structure modeling (Hassan et al., 2017).

Active site prediction of ADHs

To predict the protein-active binding sites, Residue depth (http://mspc.bii.astar.edu.sg/depth/) was employed to observe the core residues involved in the formations of binding sites. Higher peaks of generated graphs depict the amino acids that may involve in the formation of the binding pocket (http://mspc.bii.astar.edu.sg/depth/).

Mutation effect on protein stability

Multiple online tools I-Mutant (Capriotti et al., 2005), Provean (Choi and Chan, 2015), Cutoff Scanning Matrix (mCSM) (Pires et al., 2014), MUpro (Randall et al., 2006), Sorting Intolerance From Tolerance (SIFT), DUET (Pires et al., 2014) and SDM (Preissner et al., 2011) were used to check the residual effect on protein structural stability. I- Mutant is a tool for mutation analysis and uses Help Vector Machineto predict protein stability changes for point mutations automatically. SIFT predicts the alteration of amino acids which affects protein functionality. The level of sequence conservation of amino acid residue alignments produced from highly associated sequences is used to predict SIFT.

Results

Zinc and iron-based S. acidoculdarius and A. fulgidus sequence analysis

The Zn and Fe-dependent ADH from *S. acidoculdarius* and *A. fulgidus* having 352 and 395 amino acids, respectively undergoes for alignment analysis to check the conservation behavior of binding pocket motif residues are Cystine 38, Histidine 62, Glutamic acid 63, and Cystine154 in zinc strain and Aspartic acid 204, Histidine 208, Histidine 285, Histidine 271 in iron strain (*Figures 1-2*). The residues were conserved in all of the strains and appear to be in a similar arrangement, according to sequence alignment of target strains with reference *S. cervisiae* and *Z. mobilis*. Different Zinc and

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2102_925941

iron-based strains exhibit the same preserved area, confirming the depiction with *S. acidoculdarius* and *A. fulgidus*. With the help of literature, we find out that different zinc and iron-containing ADH also have the same amino acid interacting with metal ions i.e., Zinc/Iron (Ying and Ma, 2011). For example, in *S. cerevisiae*the zinc ligand interacts with amino acid residues Histidine 66, Glutamic acid 67, Cystine43, and Cystine153 (Ramaswamy et al., 2014), in *HLADH*, the three ligands are, Cystine148 and Cystine38 and Histidine61 in *SsADH* and *BaKR* the catalytic zinc bind with Histidine 148, Cystine 38, Cystine 61, and Glutamic acid 62 (Ramaswamy et al., 2014) and in *Thermococcus thioreducens* Fe-ADH, the Histidine 272, Histidine 260, Aspartic acid 193, and Histidine 197 side chains bind the Fe atom towards the second domain (Davis et al., 2007). The region that appeared to be preserved in the strain is capable of binding zinc and iron residues, according to secondary structure studies.

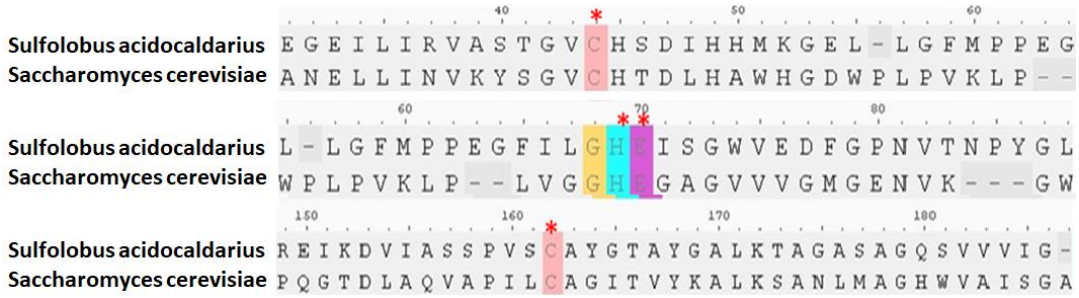

Figure 1. *The figure shows the alignment of S. acidoculdarius with S. cerevisiae. The red color * shows the residues interacting with zinc metal ion*

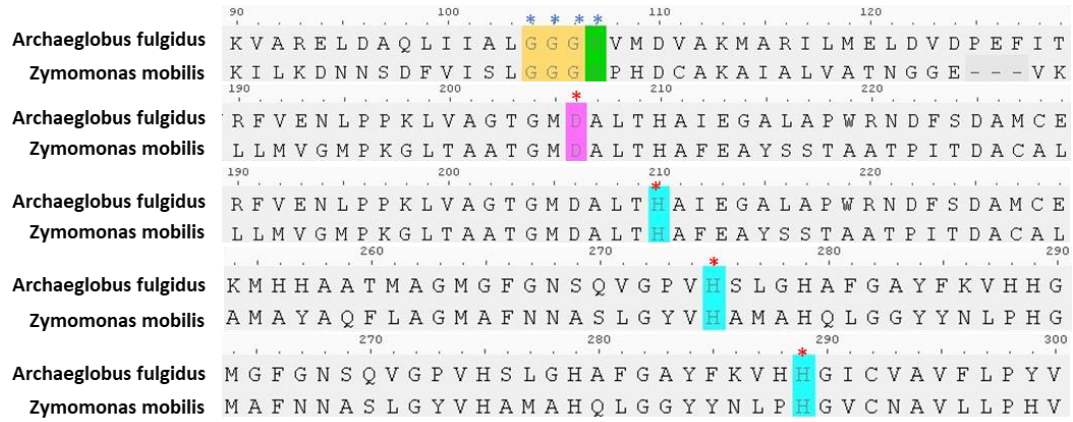

Figure 2. *Alignment of A. fulgidus with Z. mobilis. Red color * shows the conserved residues which is involved in iron binding and blue color * shows the residues involve in cofactor binding*

ADHs structure assessment

The modelled structures of ADH have been validated using Molprobity, ProtParam, VADAR, ERRAT score, 3D-verify, and Procheck, respectively. The predicted physio-chemical properties of the predicted models showed that *S. acidoculdarius* and *A. fulgidus*, exhibited molecular weight is 37298.26 (KDa) and 44237.06 (KDa) and total amino acids 352 and 395 (*Table 1*).

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2102_925941

Table 1. *Structural analysis of ADHs*

Properties	ADH, *A. fulgidus*	ADH, *S. acidocaldarius*	ADH2, *Z. mobilis*	ADH1, *S. cerevisiae*
No. of Amino Acid	395	352	383	348
Molecular weight	44237.06	37298.26	40145.20	36849.17
Isoelectric point	5.11	6.71	5.49	6.21
(-) charged Residue	61	33	41	36
(+) charged Residue	43	32	32	32
Extinction coefficient	44140	33265	13785	48860
Instability index	36.66	22.54	23.90	20.71
Aliphatic index	90.15	99.66	93.34	90.49
GRAVY	-0.082	0.107	0.121	0.030
Total no. of atoms	6219	5298	5647	5175
3D- Verify	89.84	95.99	98.69%	95.10%
ERRAT	94.2466	87.5	98.1283	96.5569
Poor rotamers	0.32%	0.35%	3.06%	5.78%
Favored rotamers	95.18%	96.45%	93.88%	82.84%
Ramachandran outliers	1.08%	2.88%	0.79%	0.87%
Ramachandran favored	94.09%	92.80%	95.26%	92.75%
Rama-distribution Z-score	-0.01 ± 0.44	-1.57 ± 0.45	-1.92 ± 0.40	-3.51 ± 0.25
Cβ deviation > 0.25Å	2.01%	4.52%	0.00%	0.50%
Bad bonds	0.00%	0.00%	0.00%	0.00%
Bad angles	0.82%	1.67%	0.00%	0.01%
Cis prolines	0.00%	0.00%	0.00%	7.69%
Tetrahedral geometry outliers	-	1	-	-

Active site prediction analysis

A binding pocket contains a specific set of binding residues which decide the position, shape in the protein, and physicochemical properties of the pocket all these help the binding pocket to work effectively. The residues located outside the binding pocket also have a great long-term impact on the properties of the binding pocket. The model of the *Z. mobilis* and *S. cerevisiae* active site residues were used as a guide for computational studies of *A. fulgidus* and *S. acidoculdarius* proteins (*Figs. 3,4*). Using Chimera, the sequence and structural analysis were done using the 3D model of the predicted active sites verified that amino acids are present in the majority of them. These are (Asp204, His208, His271, and His285) and (His62, Glu63, Cys38, and Cys154) are expected to create the active site's catalytic center. The active site prediction was done using the Residue depth server. Using this tool, we predict the active site of structural protein. Where iron and zinc ligand bind. This gives different graphs of the 3D structure's active site, the frequency of the residues establishing a binding site, it also gives the residues and their number which is forming binding site (*Figures 5-7*). These residues help in the mutagenesis at specific sites.

ERRAT plot analysis

The ERRAT graph depicts the final model of both proteins. Yellow indicates regions of the structure that can be rejected at a 95% confidence level; 5% of a decent protein structure is expected to have an error value higher than this level. Using the ERRAT tool and the Ramachandran graph, which confirms the protein structure, regions that can be rejected at a 99 percent level are displayed in redis shown in *Figure 8*.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2102_925941

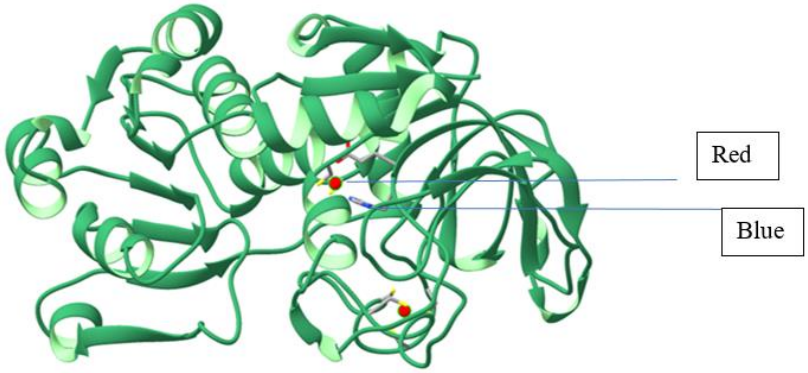

Figure 3. *Structure of ADH chain A from strain S. acidoculdarius. The red color shows the binding of iron*

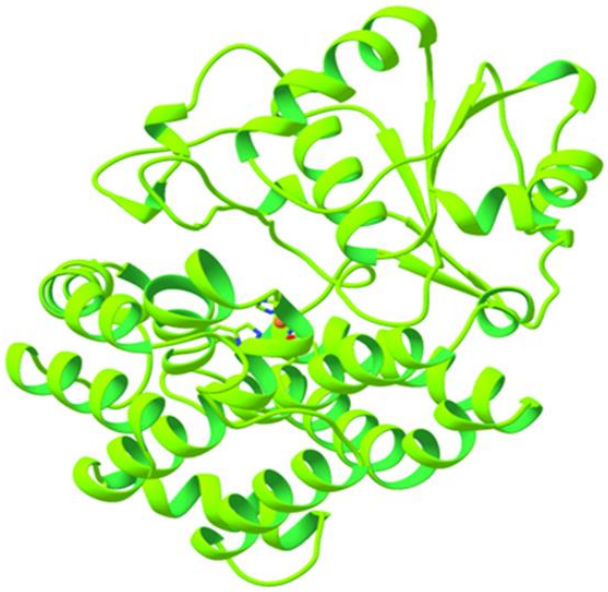

Figure 4. *Structure of ADH chain A from strain A. fulgidus. The orange color shows the binding of zinc ligand*

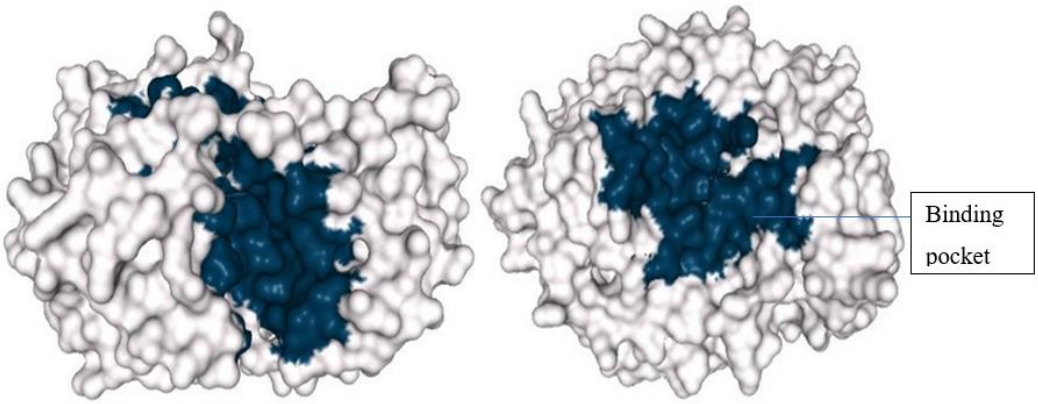

Figure 5. *Predicted 3D structure of active site of S. acidoculdarius and A. fulgidus proteins*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2102_925941

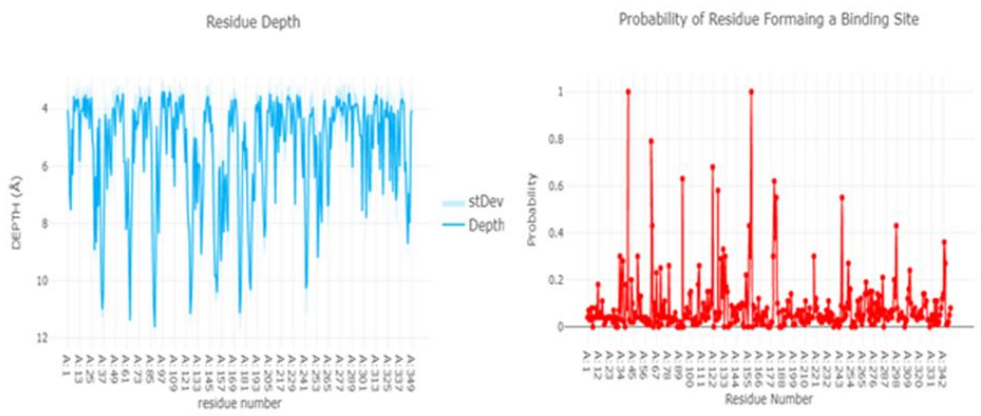

Figure 6. *The first graph shows the residue depth of the S. acidoculdarius protein structure and the Second graph shows the probability of residues forming a binding site in S. acidoculdarius protein*

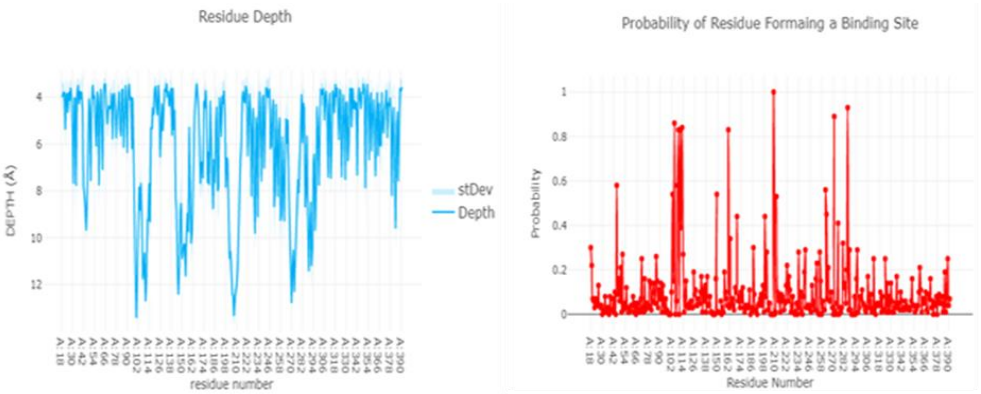

Figure 7. *The first graph shows the residue depth of protein A. fulgidus protein and the second graph shows the probability of residues forming a binding site in A. fulgidus protein*

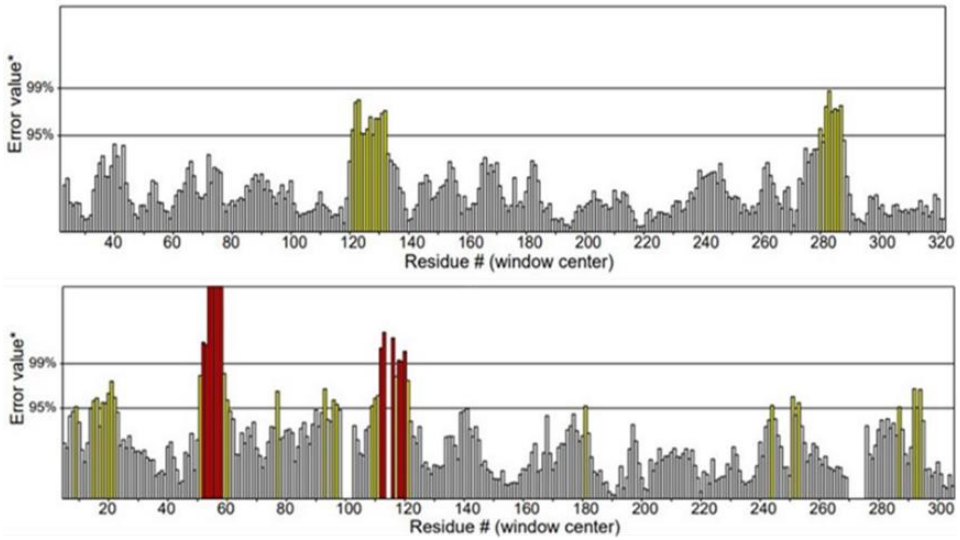

Figure 8. *ERRAT graph of A. fulgidus and S. acidoculdarius shows overall quality factor 94.247 and 87.500*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2102_925941

Ramachandran graph analysis

Ramachandran graph of *S. acidoculdarius* present amino acid residues that have shown 89.1% region in the most favored area, 9.5% amino acid residues are in the additional allowed regions, residues that have shown 0.7% region in the generously allowed area of the graph, the residue that has shown 0.7% region in the dis-allowed area, amino acid residues that have shown 100% region are total no. of non-proline and non-glycine and 349 are the totals no. of amino acid residues (*Figure 9*). Similarly, the graph of *A. fulgidus* shows that amino acid residues that have shown 90% region in the most favored area, 8.8% amino acid residues are in the additional allowed region, residues that have shown 0.8% region in the generously allowed area of the graph, the residue that has shown 0.7% region in the dis-allowed area, amino acid residues that have shown 100% region are total no. of non-proline and non-glycine and 374 are the totals no. of amino acid residue.

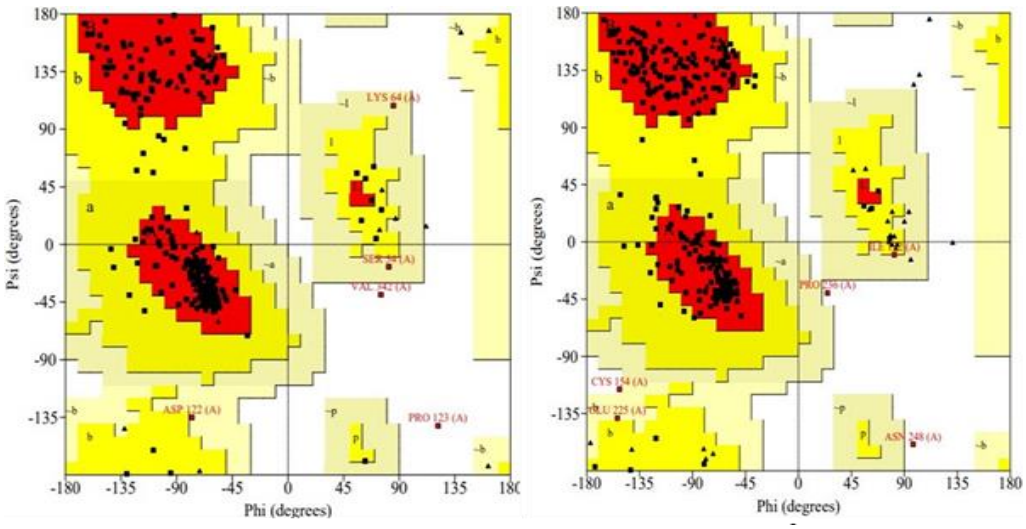

Figure 9. *Ramachandran graph of S. acidoculdarius and A. fulgidus*

Mutagenesis of S. acidoculdarius and A. fulgidus at specific sites

The conserved residues H62, C38, E63, and C154 of *S. acidoculdarius* and D204, H285, H208, H271 of *A. fulgidus*protein were selected for site-directed mutation. These residues were selected by alignment with *S. cerevisiae* and *Z. mobilis*. After visualizing the wild-type model structure in the chimera tool, the Zinc and Iron metal ion is bind in these residues which also confirms that these residues affect metal binding. These four residues were substituted with Alanine and Glutamine in Zn-dependent protein and Arginine and Glutamic acid in Fe-dependent protein. For site-directed mutagenesis, the protein sequence of *S. acidoculdarius* and *A. fulgidus* is substituted with H62A, C38A, C154A, E63Q, and H62A, C38A, C154A, E63Q through this substitution model was build using the swiss model (*Table 2*). The analysis was done using VADAR, Molprobity, save server, and ProtParam tools (*Table 4*).

Ramachandran plot analysis

Using Ramachandran plot and Procheck the structure of both mutated strains was analyzed for validation of structure. As a result of the analysis, the *S. acidoculdarius* mutated structure has residues that have shown 89.1% region in the most favored area,

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2102_925941

9.8% residues are in the additional allowed regions, residues that have shown 0.4% region in the generously allowed area of the graph, the residue that has shown 0.7% region in the dis-allowed area, amino acid residues that have shown 100% region are total no. of non-proline and non-glycine, 39 triangles shown as total no. of glycine amino acid residues, 23 triangles shown as no. of proline amino acid residues. So, 349 are the total no. of amino acid residues. In the most favored regions, the best model should have a percentage of over 90%.

Table 2. *Mutation analysis of substituted residues of S. acidoculdarius and A. fulgidus*

TOOL	MUTATION	SCORE	PREDICTION	TOOL	MUTATION	SCORE
Provean	H62A	-9.917	Deleterious	Provean	D204E	-3.978
	C38A	-8.925	Deleterious		H208R	-7.956
	C154A	-8.128	Deleterious		H285R	-7.843
	E63Q	-2.975	Deleterious		H271R	-7.870
MUpro	H62A	-0.7011	Decrease stability	MUpro	D204E	-0.2306
	C38A	-1.1164	Decrease stability		H208R	-0.3863
	C154A	-1.4602	Decrease stability		H285R	-0.2886
	E63Q	-0.6579	Decrease stability		H271R	-0.6351
IMutant	H62A	0.04	Increase	IMutant	D204E	0.47
	C38A	-0.88	Decrease		H208R	0.03
	C154A	-1.12	Decrease		H285R	-0.06
	E63Q	-0.49	Decrease		H271R	-0.15
SIFT	H62A	0.00	Intolerance	SIFT	D204E	0.00
	C38A	0.00	Intolerance		H208R	0.00
	C154A	0.00	Intolerance		H285R	0.00
	E63Q	0.00	Intolerance		H271R	0.00
DUET	H62A	-1.214	Destabilizing	DUET	D204E	O. O69
	C38A	-2.991	Stabilizing		H208R	-1.481
	C154A	-2.054	Destabilizing		H285R	-0.647
	E63Q	0.014	Destabilizing		H271R	-1.404
mCSM	H62A	-1.005	Destabilizing	mCSM	D204E	-0.298
	C38A	-2.795	Destabilizing		H208R	-1.501
	C154A	-1.942	Destabilizing		H285R	-0.785
	E63Q	-0.131	Destabilizing		H271R	-1.38
SDM	H62A	-1.58	Destabilizing	SDM	D204E	-0.17
	C38A	-1.15	Destabilizing		H208R	-1.26
	C154A	-1.02	Destabilizing		H285R	-0.74
	E63Q	-1.22	Destabilizing		H271R	-1.59

In *Archaeglobus fulgidus* mutated structure has amino acid residues that have shown 90.3% region in the most favored area, 8.5% amino acid residues are in the additional allowed regions, residues that have shown 1.2% region in the generously allowed area of the graph, the residue that has shown 0.0% region in the dis-allowed area, amino acid residues that have shown 100% region are total no. of non-proline and non-glycine, 25 triangles shown as total no. of glycine amino acid residue, 16 triangles shown as no. of proline amino acid residues and 374 are the total no. of amino acid residues (*Figure 10*).

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2102_925941

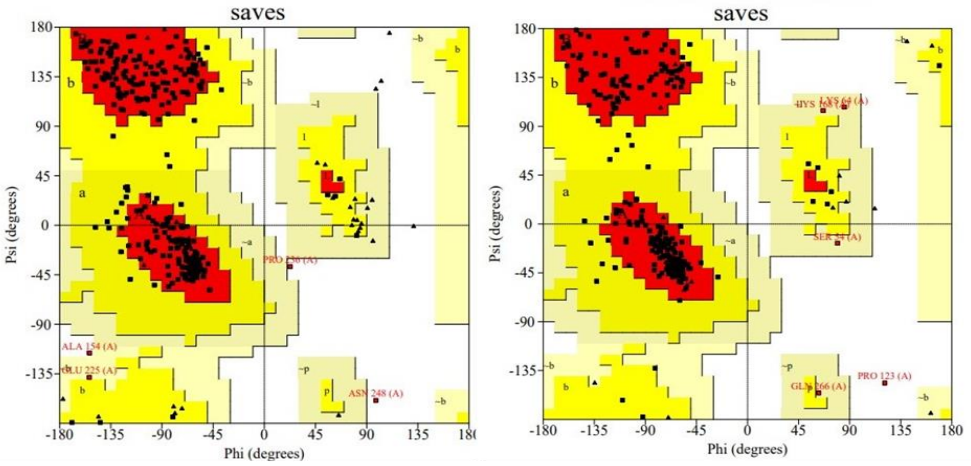

Figure 10. *Ramachandran graph of mutated S. acidoculdarius and mutated A. fulgidus. The graph was generated with the Procheck tool*

Detailed structure analyses of wild type and mutated structures

The influence of substituted residues on conformations of protein structure was studied using detailed structural analysis. The superimposition is done to compare wild and mutated structures with the structures, the distance between residues and metal ion values are given using Chimera. We concluded that His62, Glu63, Cys38, and Cys154 mutations cause a major disturbance in zinc binding and similarly His285, His271, Asp204, and His208 mutations cause a major disturbance in iron binding in the predicted model. The fluctuation resulted due to the point mutations at the active binding site. It completely disturbs the bonds between the residue and zinc and iron metal irons and which causes the prevention of metal ion binding. Both these mutated models were visualized in chimera to check the comparison of residues bind with zinc and iron with an original model, residues disturb after substitution. Therefore, in both mutated structures metal ions completely abolish from active-binding sites (*Figures 11-14, Table 3*) (Ahmad and Qazi, 2014; Damborský et al., 2019).

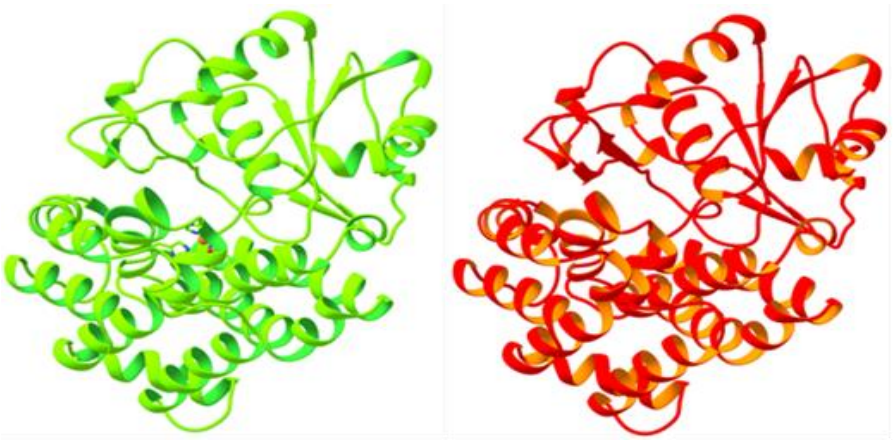

Figure 11. *Wild-type and Mutated-type structure of A. fulgidus. The first figure shows that wild-type protein, where iron metal ion is bind with the conserved residues and second figure shows Mutated protein, the iron metal ion is not bound in this structure after the substitution of binding residues which affect the binding pocket*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2102_925941

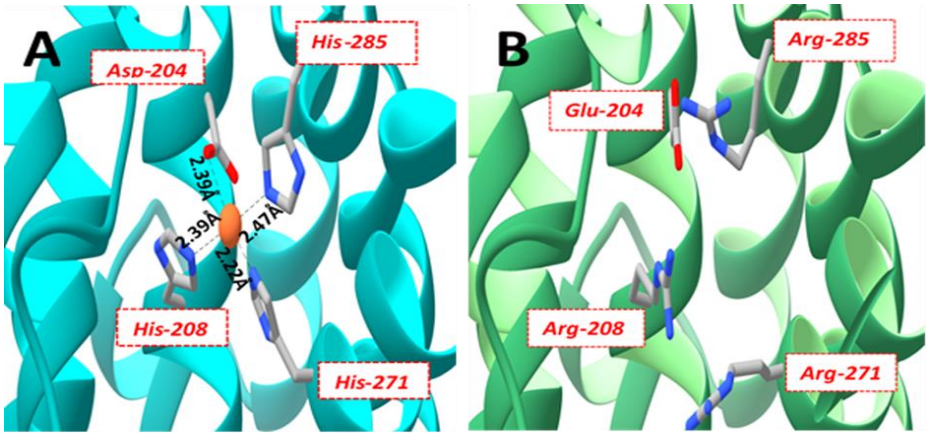

Figure 12. (A) *wild-type A. fulgidus protein, iron interaction with their binding residues and distance* ***(B)*** *Mutated A. fulgidus protein, substituting the residues which involve in iron-binding result shows that the binding conformation change so metal ion is abolish*

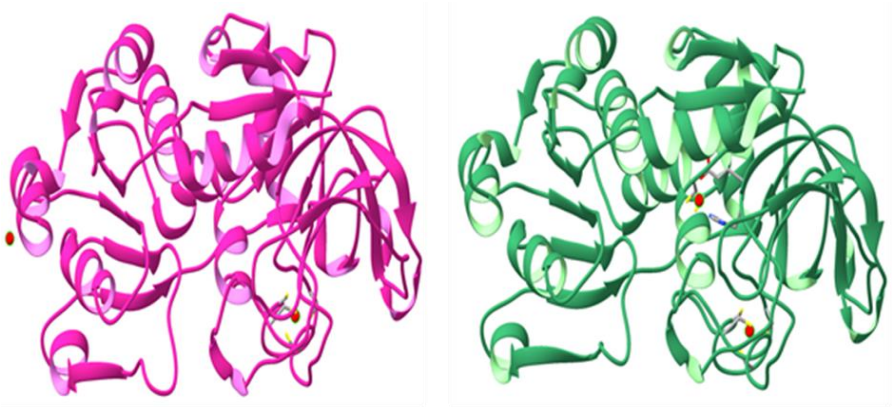

Figure 13. *Wild-type and Mutated structure of S. acidoculdarius. The first figure shows that Mutated protein, where zinc metal ion is not bound in our conserved residues after substitution and second figure shows Wild-type protein, zinc metal ion is present in conserved residues*

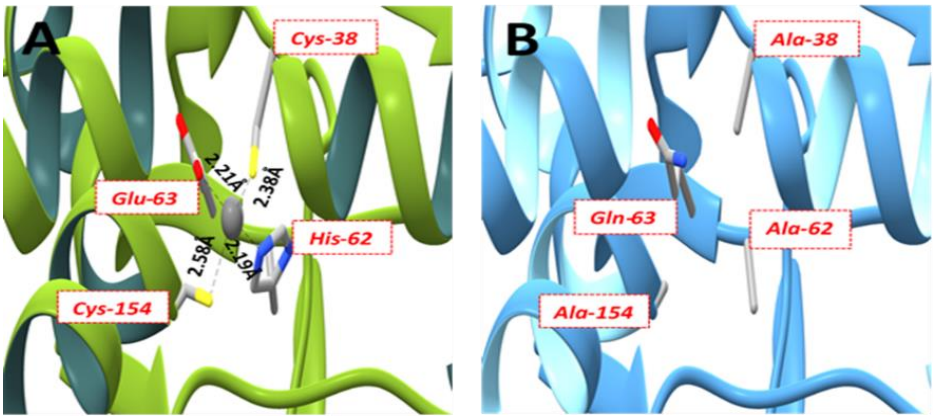

Figure 14. *(A) wild-type S. acidoculdarius protein, showing zinc interaction with their binding residues and distance (B) Mutated S. acidoculdarius protein, substituting the residues which involve in zinc binding result shows that the binding conformation changed and abolish the zinc ion*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2102_925941

Table 3. *Structural analysis of mutated and wild type strain*

Properties	**ADH, *A. fulgidus* (Wild Type)**	**ADH, *A. fulgidus* (Mutated)**	**ADH, *S. acidoculdarius* (Wild Type)**	**ADH, *S. acidocaldarius* (Mutated)**
No. of Amino acid	395	395	352	352
Molecular weight	44237.06	44308.22	37298.26	37167.09
Isoelectric point	5.11	5.14	6.71	7.05
(-) charged Residue	61	61	33	32
(+) charged Residue	43	46	32	32
Extinction coefficient	44140	44140	33265	33140
Instability index	36.66	37.82	22.54	19.66
Aliphatic index	90.15	90.15	99.66	100.51
GRAVY	-0.082	-0.091	0.107	0.118
Total no. atoms	6219	6240	5298	5290
ERRAT	94.2466	96.1219	87.5	85.3211
Poor rotamers	0.32%	0.64%	0.35%	0.00%
Favored rotamers	95.18%	96.46%	96.45%	96.77%
Ramachandran outliers	1.08%	1.34%	2.88%	3.17%
Ramachandran favored	94.09%	92.74%	92.80%	93.66%
Rama-distribution Z-score	-0.01 ± 0.44	-0.02 ± 0.43	-1.57 ± 0.45	-1.48 ± 0.45
Cβ deviation> 0.25Å	2.01%	1.15%	4.52%	4.19%
Bad bonds	0.00%	0.00%	0.00%	0.00%
Bad angles	0.82%	0.65%	1.67%	1.98%
Cis prolines	0.00%	0.00%	0.00%	0.00%
Tetrahedral geometry outliers	-	-	1	1

Mutation analysis

I-Mutant (Capriotti et al., 2005), Provean (Choi and Chan, 2015), mCSM (Pires et al., 2014), MUpro (Randall et al., 2006), SIFT (Ng et al., 1981), DUET (Pires et al., 2014) and SDM (Preissner et al., 2011) were used to perform the mutation Analysis. To predict whether the substitutions of residues affect the protein stability or not. These 7 tools were used to get mutation score and stability prediction (*Table 4*).

To inexact the protein's ddG esteem and expect the bearing of steadiness change, MUPRO and I-Mutant2.0 utilize an SVM model. Positive upsides of ddG in the I-Mutant 2.0 server show seriously balancing out information, while negative qualities demonstrate destabilizing data. On the off chance that the PROVEAN score is equivalent to or under a present edge, the protein change is probably going to have a "damaging/deleterious" impact (for example, - 2.28). If the PROVEAN score is over the edge, the variation is probably going to have a "tolerated/neutral" impact (Temelso et al., 2019).

Amino acid changes in protein-coding regions are recognized by single nucleotide polymorphism (SNP) exploration and irregular mutagenesis tests. The standardized probability that the amino acid change would be tolerated is showed by scores. Replacements with a score of under 0.05 are named malicious by SIFT. Some SIFT

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2102_925941

clients have found that replacements with a score of under 0.1 had a higher affectability for recognizing risky SNPs than replacements with a worth of more than 0.1 (Ng et al., 1981).

First, the I-Mutant score for both proteins predict that the substitutions decrease the stability of protein except residues H62A in Zn dependent protein and D204E in Fe dependent protein which increases the stability, Provean tool predicts that all the substituted residues for both proteins are deleterious, mCSM tool predicts that all the residues for both proteins destabilizing the protein, MUpro tool predict that all the residues substitution decreases the stability of both proteins, SIFT tool predict that all the residues substitution are intolerant, DUET tool predicts that all the substituted residues destabilizing the protein except residues C38A in Zn dependent protein and H208R in Fe dependent protein which stabilizing the protein and SDM tool predict that all the substituted residues for both proteins reduce the stability.

Table 4. *Applied tools*

Sr No.	Name of tool	purpose	reference
1	Uniprot Knowledgebase	Retrieval of protein sequences	https://www.uniprot.org/help/uniprotkb
2	Molprobity webserver	Protein structural analysis	(http://molprobity.biochem.duke.edu/).
3	T-Coffee	Alignment tool	(https://www.ebi.ac.uk/Tools/msa/tcoffee/
4	Residue depth	Protein active binding site prediction	(http://mspc.bii.astar.edu.sg/depth/)
5	ERRAT plot	Verification f protein structure	https://pubmed.ncbi.nlm.nih.gov/8401235/

Discussion

In a time when soaring petroleum prices and growing environmental problems are becoming major concerns. Countries that are facing this problem in its worst condition include Pakistan. Bioethanol is an engaging opportunity for meeting future energy needs. Biofuels produced using inexhaustible materials, for example, cereals and sugar stick for bioethanol and plant-inferred oils for biodiesel, are specifically noteworthy. These items are either joined with or supplant regular energizes like fuel and diesel, and they offer an approach to address critical natural difficulties, for example, unreasonable transportation-related ozone harming substance (GHG) discharges.

The alternative of petroleum-based transportation fuels is bio-diesel and bio-ethanol that are similar to them. It's anything but a reasonable wellspring of energy fuel that is delivered by changing over agricultural raw products and domestic wastes (fatty oils) into fluid energy fuel with insignificant or no natural effect (Khan et al., 2011).

This research is based on zinc and iron-dependent ADHs that are characterized as ADH 1 & 3 in literature. ADH studied in strains is majorly thermophilic can withstand high temperatures without their protein structure being altered. As it is known that ADH is responsible for the fermentation of sugars into ethanol, this enzyme is rendered useful in the production of bioethanol. Optimum conditions (temperature, pH, and anaerobic environment, medium) are always the prime requirements of any process, in this case where our potential feedstock primarily is lignocelluloses based; a lot of pre-treatments

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2102_925941

are required. Shortly use of thermophilic strains is expected to be the prime priority as all the parameters listed above can be easily maintained in these strains.

The advancement of systems dependent on the misuse of the carbohydrate part of lignocellulosic material has been practicing exclusively the focal point of procedures for expanding fermentative ethanol creation. The age of ethanol at higher temperatures has been prescribed as an approach to make the measure plan simpler (Morozkina et al., 2010).

During this research, the most imperative finding included the identification of conserved Histidine 62, Cystine 38, Cystine 154, and Glutamic acid 63 regions in *S. cerevisiae* and *S. acidoculdarius* ADHs that contain zinc, which additionally turns out to be the zinc active site residues. While in the Fe-based ADH site, the active binding site for Fe is present, where the metal ion is coordinated by an aspartic acid residue (Asp204) and three histidine residues (His208, His271, and His285, numbering in *A. Fulgidus* (Ying, 2007). Moreover, the underlying structural modelling finished with specific substitution of residues showed the most stable type of substitution as when the progressions were registered the outcomes were not positive.

The first mutation also shows the negative change in *A. fulgidus* (iron-based ADH). One aspartic acid and three histidine residues associating with metal in the catalytic site were replaced with glutamic acid and arginine at all four sites. The disruption of linkages between the metal and the residues was the result of mutations. The alterations entirely disrupted the pocket, preventing the formation of an enzyme-substrate complex. The research was performed that suggested that the mutation occurred close to the active site iron (Fodor et al., 2014).

The second mutation when done was based on zinc-based ADH *S. acidoculdarius*. In it, the zinc-binding residue site Histidine 62, Cystine 38, and Cystine 154 was replaced with alanine and Glutamic acid 63 was substituted with Glutamine. Not only did the modification damage the catalytic site, but it also disrupted all of the possible linkages that kept the active site work effectively. Due to the mutation, a negative change resulted. It eradicates the catalytic activity and enzyme's metal-binding capability as well as without changing the enzyme's entire secondary structure (Olson et al., 2015).

Studies on *S. cervisiae* reveal that this, too, led to a negative change. The modification was made to the zinc-binding residue Glutamic acid (Glu67), which was replaced with Glutamine (Gln). Even though both are acidic, the conformation of the binding site changed due to the varying characteristics of Glutamine, and interactions between metal and residue were broken to some extent, reducing the total efficiency of ethanol generation. The structural analysis validated the findings of previous research. Not only did the modification damage the catalytic site, but it also disrupted all of the possible linkages that kept the active site work effectively. Due to the mutation, a negative change resulted. It eradicates the catalytic activity and enzyme's metal-binding capability as well as without changing the enzyme's entire secondary structure (Radianingtyas and Wright, 2003).

In another work, the same results were found, showing that replacing Glutamic acid (Glu 63) with Glutamine (Gln63) reduces catalytic effectiveness by a hundredfold (Ahmad and Qazi, 2014). In a brief, the similarity between the strains indicates that they have a tremendous deal of potential for producing biofuels. Furthermore, the hostile and persistent habitat in which they may survive makes them ideal for practical investigations. Theoretical work on their mutations has resulted in a collection that may be used for exact practical demonstration.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2102_925941

Conclusion

In nutshell, the study of thermophilic bacterial strains theoretically proved to be of immense benefit. Interestingly, apart from withstanding high temperatures during pretreatment processes, the bacteria tended to maintain its original structure. The major goal behind targeting thermophiles is to isolate the cost-effective and environment-friendly strain that can produce Bioethanol. Based on their stability ADH from these strains can ensure a strong enzyme-substrate complex, promising a consistent and higher yield as compare to mesophilic bacterial strains. The prokaryotic nature of these strains makes them even more feasible for use as their structural mechanism is way simpler than eukaryotic organisms such as yeast. It is because genetic engineering in prokaryotes is easy and cheaper than in eukaryotes. Hence, this theoretically researched if taken practically can lead to positive and successful results rendering open the gates of effective commercialization for efficient bioethanol production.

Future prospective

According to current research, the remarkable heat stability qualities of hyper-thermostable proteins can help in different Biotechnological industries because of their novel proteins which are used during pre-treatments. Zn-dependent ADH and Fe-dependent ADH withstanding high temperatures during pre-treatments and that help in high yield of alcohol. In future research, we can use these thermophiles with genetically engineered enzymes to help in better ethanol yield rather than traditional mesophiles. Hyperthermophilic ethanol manufacturing research is still in its early stages. The genomes of several hyperthermophilic organisms have already been sequenced, and putative ADHs have been recognized, but none have been characterized. For characterization following tools have been used, such as quantitative reverse transcription PCR could be employed to evaluate the physiological role and expression level under specific condition.

REFERENCES

[1] Ahmad, Q.-u.-A., Qazi, J. I. (2014): Thermophilic fermentations of lignocellulosic substrates and economics of biofuels: prospects in Pakistan. – International Journal of Energy and Environmental Engineering 5(2): 94.

[2] Ammendola, S., Raia, C. A., Caruso, C., Camardella, L., D'Auria, S., De Rosa, M., Rossi, M. (2012): Thermostable NAD+-dependent alcohol dehydrogenase from Sulfolobussolfataricus: gene and protein sequence determination and relationship to other alcohol dehydrogenases. – Biochemistry 31(49): 12514-12523.

[3] Capriotti, E., Fariselli, P., Casadio, R. (2005): I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. – Nucleic Acids Research 33: W306-W310.

[4] Choi, Y., Chan, A. P. (2015): PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. – Bioinformatics 31(16): 2745-2747.

[5] Damborsky, Ș. A. (2019): Bioethanol from Cellulosic Materials: A Renewable Motor Fuel from Biomass. – Energy Sources 27(4): 327-337.

[6] Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X. Y., Murray, L. W., Arendall, W. B., Snoeyink, J., Richardson, J. S., Richardson, D. C. (2007): MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. – Nucleic Acids Research 35: W375-W383.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2102_925941

[7] Dehouck, M. S., Ma, K. (2009): Decarboxylation of pyruvate to acetaldehyde for ethanol production by thermophiles. – Biomolecules 3(3): 578-596.
[8] Deng, H., Jia, Y., Zhang, Y. (2018): Protein structure prediction. – International journal of modern physics B 32(18): 1840009.
[9] Guy, J. E., Isupov, M. N., Littlechild, J. A. (2003): The Structure of an Alcohol Dehydrogenase from the Hyperthermophilic Archaeon Aeropyrum pernix. – Journal of Molecular Biology 331(5): 1041-51.
[10] Hassan, M., Abbas, Q., Raza, H., Moustafa, A. A., Seo, S. Y. (2017): Computational analysis of histidine mutations on the structure stability of human tyrosinases leading to albinism insurgence. – Mol. BioSyst. 8.
[11] Khan, R. A., Khan, A. N., Ahmed, M., Khan, M. R., Shah, M. S., Azam, N., Sadullah, F., Dian, F., Ullah, S., Khan, N. (2011): Bioethanol sources in Pakistan: A renewable energy resource. – African Journal of Biotechnology 10(86): 19850-19854.
[12] Machielsen, R., Uria, A. R., Kengen, S. W. M., van der Oost, J. (2006): Production and characterization of a thermostable alcohol dehydrogenase that belongs to the aldo-keto reductase superfamily. – Applied and Environmental Microbiology 72(1): 233-238.
[13] Montella, C., Bellsolell, L., Pérez-Luque, R., Badía, J., Baldoma, L., Coll, M., Aguilar, J. (2005): Crystal structure of an iron-dependent group III dehydrogenase that interconverts L-lactaldehyde and L-1,2-propanediol in Escherichia coli. – Journal of Bacteriology 187(14): 4957-4966.
[14] Moon, J.-H., Lee, H.-J., Park, S.-Y., Song, J.-M., Park, M.-Y., Park, H.-M., Sun, J., Park, J.-H., Kim, B. Y., Kim, J. S. (2011): Structures of iron-dependent alcohol dehydrogenase 2 from Zymomonas mobilis ZM4 with and without NAD+ cofactor. – J Mol Biol 407(3): 413-24.
[15] Morozkina, E. V., Slutskaia, E. S., Fedorova, T. V., Tugaĭ, T. I., Golubeva, L. I., Koroleva, O. V. (2010): Extremophilic microorganisms: biochemical adaptation and biotechnological application. – Applied Biochemistry and Microbiology 46(1): 5-20.
[16] Ng, T. K., Ben-Bassat, A., Zeikus, J. G. (1981): Ethanol Production by Thermophilic Bacteria: Fermentation of Cellulosic Substrates by Cocultures of Clostridium thermocellum and Clostridium thermohydrosulfuricum. – Applied and Environmental Microbiology 41(6): 1337-1343.
[17] Olson, D. G., Sparling, L., Lynd, L. R. (2015): Ethanol production by engineered thermophiles. – Curr Opin Biotechnol 33: 130-41.
[18] Pires, D. E. V., Ascher, D. B., Blundell, T. L. (2014): mCSM: predicting the effects of mutations in proteins using graph-based signatures. – Bioinformatics 30(3): 335-342.
[19] Preissner, R., Blundell, T. L. (2011): SDM - a server for predicting effects of mutations on protein stability and malfunction. – Nucleic Acids Research 39(Web Server issue): W215-W222.
[20] Radianingtyas, H., Wright, P. C. (2003): Alcohol dehydrogenases from thermophilic and hyperthermophilic archaea and bacteria. – FEMS Microbiol Rev 27(5): 593-616.
[21] Ramaswamy, S., Plapp, B. V. (2014): Yeast Alcohol Dehydrogenase Structure and Catalysis. – Biochemistry 53(36): 5791-5803.
[22] Randall, A., Baldi, P. (2006): Prediction of protein stability changes for single-site mutations using support vector machines. – Proteins 62(4): 1125-32.
[23] Stank, A., Kokh, D. B., Fuller, J. C., Wade, R. C. (2015): Protein Binding Pocket Dynamics. – Accounts of Chemical Research 49(5): 809-815.
[24] Tan, K. P., Nguyen, T. B., Patel, S., Varadarajan, R., Madhusudhan, M. S. (2013): Depth: a web server to compute depth, cavity sizes, detect potential small-molecule ligand-binding cavities and predict the pKa of ionizable residues in proteins. – Nucleic Acids Research 41(W1): W314-W321.
[25] Temelso, B., Alser, K. A., Gauthier, A., Palmer, A. K., Shields, G. C. (2019): Structural Analysis of α-Fetoprotein (AFP)-like Peptides with Anti-Breast-Cancer Properties. – The Journal of Physical Chemistry B 118(17): 4514-4526.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2102_925941

[26] Ying, X. X. (2007): Characterization of iron- and zinc-containing alcohol dehydrogenases from anaerobic hyperthermophiles. – UWSpace. http://hdl.handle.net/10012/3446.
[27] Ying, X., Ma, K. (2011): Characterization of a zinc-containing alcohol dehydrogenase with stereoselectivity from the hyperthermophilic archaeon Thermococcus guaymasensis. – Journal of bacteriology 193(12): 3009-3019.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2102_925941