

Homology Modeling of Carbonic Anhydrase from *Dunaliella Viridis*

syed muhammad iqbal azimuddin (✉ iqbal.azimuddin@gmail.com)

Mohammad Ali Jinnah University <https://orcid.org/0000-0002-3376-3835>

Research Article

Keywords: Dunaliella Viridis, Carbonic Anhydrase, Halotolerant, homology modeling, Protein modelling

Posted Date: January 20th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-2459875/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Under salt and toxic metal stress condition, genetic make-up and proteins including some enzymes undergo specific changes to combat the posed harsh environment. Carbonic anhydrase is ubiquitous enzyme associated with respiratory disorder and cancerous condition. Strategies investigated and adopted by microorganisms to cope with hypersaline by bioinformatics analysis including homology modeling, Zinc metal interaction prediction, sequence analysis for Carbonic Anhydrase.

This study shows that carbonic anhydrase acquired more acidic residues on its surface as countermeasure and may form salt bridges with positive ions in response to hypersaline environment. It also reduced its surface area and had more polar residues. Zinc metal interaction with Nitrogen of Histidine residues were almost conserved in the studied and modelled structure with fewer changes in catalytic region of sequence pattern. Salt tolerance achieved by foster promising approaches at the genetic and proteomic level by modifying protein sequence. Surface residues are exposed to a hypersaline medium and are mainly composed of polar and acidic residues. The enzyme exists in its compact form and reduces its surface footprint exposed to outside environment. Beta pleated secondary structure configuration increases flexibility to counter, ions disruptive effect.

Introduction

Carbonic Anhydrase (CA) is a vital enzyme of ubiquitous nature has physiological and ecological role. This enzyme mediate mineralization process specifically deposition of CaCO_3 . (Kupriyanova et al., 2011) CA is involved in ion transport, photosynthesis, acid base balance, calcification, respiration etc. The enzyme is converting Carbon dioxide (CO_2) in HCO_3^- forms and termed to be one of the fastest enzymes in nature (Kanth, Lee, & Pack, 2013). There are five classes of CA α , β , γ , δ and ζ . CA α -class is monomer mostly, β -CA class is very diverse in nature from dimer to octamer, γ -class is trimer, δ -class is Cadmium containing and ζ -class is Iron or Cobalt containing enzymes recently discovered (Tripp, Smith, & Ferry, 2001) (Emameh, Barker, Tolvanen, Ortutay, & Parkkila, 2014) (So et al., 2004).

In salt stress condition *Dunaliella Salina* a green alga adapt expressed an internally duplicated CA with each repeat homologous to *Chlamydomonas reinhardtii* CA, CA activity observe at higher salt concentrations. It is because of fusion into single polypeptide chain of two repeats and it has excess of acidic over basic residues (Fisher, Gokhman, Pick, & Zamir, 1996). Cyanobacteria and salt tolerant algae overcome the limitation of decreased solubility of CO_2 by up regulation of plasma membrane CA which facilitates CO_2 acquisition (Fisher et al., 1996) (Liska, Shevchenko, Pick, & Katz, 2004). Arid conditions also stimulate bacteria to work at high salt concentration like terrestrial cyanobacterium *Nostoc Falgelliforme* withstands prolonged period of dehydration. External CA is activated by hydration with dew or rain wetting while internal CA is inactivated, in this way maintaining a nice balance of activity driven Carbon fixation by photosynthetic machinery (Ye, Gao, & Giordano, 2008). *Synechococcus* specie PCC 7942 a fresh water cyanobacterium show salt tolerance on transformation and expression of Na^+ / H^+ antiporter channel protein. (Waditee, Hibino, Nakamura, Incharoensakdi, & Takabe, 2002)

In this study CA of salt tolerant CA structural details studied and various bioinformatics tool were used to analyze *dunaliella viridis* CA. The eukaryote thrives in hypersaline ecological system.

Methodology

Homology Modeling

Sequence Hunt

Carbonic anhydrase sequence from hyper saline thriving *Dunaliella viridis* was chosen to construct homology model by Modeller version 9.20 (Andrej Šali & Blundell, 1993). The criteria to choose this sequence are

- (a) The sequence is a moderate length sequence
- (b) The enzyme sequence is expressed as functional enzyme (experimental evidence at transcript level)
- (c) It is an extracellular enzyme and a putative component of EPS
- (d) Enzyme is halotolerant enzyme as belong to the organism grows in salty in habitat (Brock, 1975).

The sequence accession number is "D2SP68" with entry name as "D2SP68_9CHLO" and gene name is "CA2". It has 325 amino acid residues and after removal of signal peptide predicted by SignalP 4.1 server (Nielsen, 2017) the sequence length truncated to 304 amino acids. The sequence is searched by key word "Dunaliella viridis", "halotolerant carbonic anhydrase" and later downloaded from UniprotKB (www.uniprot.org) ("UniProt: the universal protein knowledgebase," 2017).

Finding Template

A sequence database was created from 901 pdb files deposited in structural database of proteins www.rcsb.org as carbonic anhydrase under Enzyme Commission number 4.2.1.1 through modified python script align2d.py from modeller. The script reads downloaded pdb files without sequence alignment and extracts sequences from it and stored it in a text file in fasta format. The text file is further processed for removal of redundant entries and number of sequences decreased to 660 sequences with the help of another python script (seqclean.py). Position specific iterated BLAST (PSI BLAST) (Altschul et al., 1997) was performed on the customized database of carbonic anhydrase for "D2SP68" (truncated version, signal peptide removed) as query sequence. These sequences reduced to 427 sequences after PSI BLAST hit analysis and could be served as possible identified template.

Template search is also accomplished by build_profile.py modified version edited to generate alignment files in PIR (protein identification resource) format .pap alignment file in PAP format and .prf Profile file in text format. This python script first convert pdb_95.pir database file in binary format and process

sequences that has 30-4000 residues. Pdb_95.pir has pdb sequences for each representative structure that is clustered at 95% to remove redundancy. At the python terminal log file display information about the hits found with chances of error (0.0 is the best). The signal peptide removed query sequence "D2SP68" was provided in PIR format.

After analyzing output obtained from PSI BLAST and build_profile finally 11 pdb structures were selected and these were 4x5s, 5hpj, 4coq, 4twl, 1kop, 4xiw, 3fe4, 1y7w, 4cnx, 5a25 and 6eki. The parameter for the selection was higher identical residues number, lesser gaps in alignment and secondary structure assignment consistency among the template structures. Template with the highest resolution was preferred among the templates hit with similar sequences and fewer differences.

Aligning the sequences

Sequence alignment is the prerequisite for model building through any comparative modeling software. Modeller also needs aligned template sequences with query sequence, and this was achieved by edited salign.py script that catered the requirement of the specific case. The protocol adopted for sequence alignment can be divided in two parts:

(a) Multiple Sequence Alignment (MSA)

(b) Pairwise Sequence Alignment (PSA)

Above mentioned steps depicted the loop boundaries and gaps found in the query sequence conducted for structure prediction. The mutation trend showed by multiple sequence alignment helped in finding conserved / semi conserved locations and guided to do adjustments in the sequence alignment if required. The most problematic regions were loop regions those did not align with any of the template and the misaligned patches.

The comparison between MSA and PSA gave another dimension to figure out sequence pattern consistency and changes from MSA to PSA and vice versa. The information extracted helped in getting the most optimum alignment. Accurate alignment is prime requirement for getting accurate structure prediction by template base modeling (A. Šali, Overington, Johnson, & Blundell, 1990). The MSA and PSA also gave clue to understand signature patches relevant for mesophilic, thermophilic and halotolerant version of the enzyme.

Modeller is the package available for homology modelling by satisfying spatial restraints (Andrej Šali & Blundell, 1993). The package is available as python package and contains various modules, sub modules, methods and functions to perform many tasks required to accomplish homology modeling.

Various methods are available in alignment class object for sequence alignment job for example align, align2d, align3d, malign, malign3d and salign. Salign and align2d methods were used and selected for better results on the basis of alignments produced after manual inspection. Align2d was used for PSA

and salign was used for both PSA and MSA, while in the case of salign method for MSA progressive alignment was chosen with the generation of superposed structures in _fit file form considering 1y7w as reference structure.

All these methods were run in loop mode so that the job could be executed in a single stance and for this procedure python scripts were edited accordingly. Aln.write method created PAP, FASTA and PIR files, the PIR file served as sequence alignment file for modeling, the FASTA file used for generation of publishable quality of sequence alignment files by Aline a perl script written for this purpose (Bond & Schuttelkopf, 2009) and PAP file used to manually inspect alignment quality contain secondary structure assignments, solvent accessible surface area (SASA) etc.

The DSSP algorithm extract secondary structure information from pdb coordinates of protein in a standard format .dssp file. The file can be read by other sequence alignment, displaying soft wares and the information can be shown with sequence alignment. This information help in model building task and improve sequence alignment (Kabsch & Sander, 1983).

Automodel class was used from modeller package for model construction and optimized by variable target function method (VTFM) through conjugated gradient (CG). The quality of model was further improved by molecular dynamics (MD) with simulated annealing (SA) (Andrej Šali & Blundell, 1993).

Initially 10 models were constructed and subjected to further processing of evaluation to check the quality of model and refinement to improve the quality of the projected model. The process of energy minimization a due step to get quality models during model construction. The step is critically important otherwise any wrong selection of parameter could trap the whole process in the vicious cycle of model construction → model evaluation → again model construction. Energy minimization is the procedure to get a model with minimum or no bad contacts means atoms are not colliding to each other. The atom distance, angles and dihedral angles are within the geometrical limits for each amino acid. The bond angles and distances are not violating physical limits and the standards set by observed protein models determined by experimental techniques like X-ray diffraction and Nuclear Magnetic Resonance Spectrometer. The protocols used were VTFM with CG and MD with SA (Levitt, 1983) (Andrej Šali & Blundell, 1993).

Discrete Optimized Protein Energy (DOPE) (Shen & Sali, 2006) was employed at the beginning during the model construction and scores were written for every generated model in the log file. Model with the lowest score was the most suitable model and went through the next steps.

Model was checked for stereo chemical quality by Procheck (R. A. Laskowski, MacArthur, Moss, & Thornton, 1993) and generated Ramachandran plots (Ramachandran, Ramakrishnan, & Sasisekharan, 1963b) Ramachandran plot depicted phi and psi torsion angles for every amino acids and highlighted the problematic amino acid.

The energy profile of each amino acid was also checked and plotted through Prosa-web server and ProSa2003 a standalone python based application to find out residues with high energy (Wiederstein &

Sippl, 2007).

The process of refinement was conducted through loop modeling (Fiser & Sali, 2003) by modeller. The region with problems were marked and a loop modelling was performed through creating a class object MyLoop contain method select_loop_atoms for the selection of residue range subjected to loop modeling. The models were constructed by defining MD level refine.slow to get better quality model. The created models 10 in number were also assessed for DOPE scores and the best model were further processed if needed.

The model obtained from loop modeling were additionally refined via kinematic closure (KIC) routine of Pyrosetta (Schenkelberg & Bystroff, 2015) (Dasgupta, 2016) after manual inspection of Ramachandran plot (Ramachandran, Ramakrishnan, & Sasisekharan, 1963a) and out files for any errors. The KIC is based on the conformational sampling algorithm treated the first selected amino acid as hinge. The algorithm evaluate all possible torsion angle conformations for main chain and keep rejecting the low quality one until the best one is chosen for every amino acid in the loop (Mandell, Coutsiyas, & Kortemme, 2009).

The energy score was also speculated by gbsa script that calculated solvation energy for a model. Several other scripts with multiple options are also available and could be used for model improvement and checking the quality of model or restricting the region for the reconstruction. Picking of hot atoms and finding the low-quality region could also be determined through modeller scripts. Some of the scripts were tried but not mentioned in this work. The evaluation and refinement procedures were employed in cycle unless an acceptable quality model was obtained.

CAs are metallo-enzymes and most of them contain Zinc in their catalytic pocket (Lionetto, Caricato, Giordano, Erroi, & Schettino, 2012). The approach adopted to incorporate Zinc metal in the pdb coordinates of model was through FINDSITE-metal version 1.0 (Brylinski & Skolnick, 2011).

At first the mapping file was prepared by findsite_library command and the resultant mapping file contained pdb entries mapped to the template. This took several hours for a desktop computer depending on the computer configuration and input file provided. Then findsitemetal was executed with the input pdb structure and template.lst that hold list of matched pdb structures (Obtained through HHBlits)

The predicted CA model out of the amino acid sequence and template pdb models used for prediction were analyzed to determine Zinc nearby residues. Zinc interaction with these residues were calculated and projected in 2D and 3D space for the structures under consideration. All 2D projections were acquired and 3D were facilitated through LigPlot+ version 2.1 by Pymol extension (Roman A. Laskowski & Swindells, 2011). The HBPLUS calculated hydrogen bonds and hydrophobic interactions (McDonald & Thornton, 1994) as built in function.

Accessible surface area was calculated by freesasa version 1.1 with default parameters (Mitternacht, 2016). Pdbremix python module was used to calculate total volume of individual Cas.

Results

Multiple Sequence Alignment (MSA) exhibits some highly consistent residues including Histidine at catalytic pocket and cysteine that forms disulfide bridges in almost all sequences (mesophilic, halophilic and thermophilic carbonic anhydrases) Figure 3, Figure 76 and Table 1 (Jo, Seo, & Cha, 2014).

Table 1 Template selected for model building ID % with respect to D2SP68 Query Sequence

PDB ID	Status	Metal	Source	ID%	Length
1Y7W	Halotolerant	ZN+2	<i>Dunaliella salina</i>	45	291
4CNX	Halotolerant Engineered	ZN+2	<i>Bos Taurus</i>	32	262
5A25	Halotolerant Engineered	ZN+2	<i>Bos Taurus</i>	32	260
5HPJ	Psychrohalophilic	ZN+2	<i>Photobacterium profundum</i>	41	217
1KOP	Mesophilic	ZN+2	<i>Neisseria gonorrhoeae</i>	37	223
3FE4	Mesophilic	MG+2	<i>Homo sapiens</i>	35	278
4TWL	Mesophilic	None	<i>Dioscorea japonica</i>	40	246
4XIW	Mesophilic	ZN+2	<i>Chlamydomonas reinhardtii</i>	37	239
4COQ	Thermophilic	ZN+2	<i>Thermovibrio amonificans</i>	41	247
4X5S	Thermophilic	ZN+2	<i>Sulfurhydrogenibium azorense</i>	43	232
6EKI	Thermophilic	ZN+2	Hydrothermal chimney vent	41	226

Some sequence patterns emerges after CAs (Halotolerant, mesophilic and thermophilic) pairwise and multiple sequence alignments Figure 4 to Figure 14 and Figure 17 to Figure 19. and Table 2 .These patterns trend depict halotolerant CAs has conserved Y and almost conserved W except 1Y7W (pattern 1), while QsP is conserved except S replaced by T in 5HPJ (pattern 3). The catalytic pocket has consistent identical residues of QFHFH/EIHIV the region stretched in thermophilic CAs as LKQFHFHAPSEHT/EaHfVH (pattern 5 and 6), has similar polar region SEhtv (pattern 6) and shows

conservation of hydrophobic LLP (pattern 8) also present in thermophilic CAs and relatively polar GSLTTP (pattern 9) region that is extended as YSGSLTTPPC in thermophilic case.

Table 2 Sequences (Sequence count 12,04,03...) show consensus pattern: ! hydrophobic, # polar, % aromatic, upper case identical, lower case shows amino acid not conserved for example y (tyrosine) is not conserved and nX amino acid positions did not match (where n=2,3,4....)

Structure Count	pattern1	pattern2
12	hwXy	gpehW
04 Halo	hwgY	khngpehWh
04 Meso	y	gpeXWgnl
03 Thermo	HWsY	GpehWgdL
Structure Count	pattern3	pattern4
12	C2XGXnQSP!#i	v#nghXt2Xv
04 Halo	GeXQsP!DlXt	geatsXrmvn#ghsfnveydd
04 Meso	cXtGkrQSP!#	dvXNnGHtvqv3Xp
03 Thermo	mCkigKNQSP!Dis	!vNNGHTIkV
Structure Count	pattern5	pattern6
12	g2X%XlkQfHfH	pSEhtvXgkX%pXEaHIVh5Xg
04 Halo	ltXt%rlvQFHFHwg	gSEhtvdXkkY2XEIHIVhwntkygdfgta
04 Meso	lt!nXlgXeyXlkqfHfH	pSEhXmDgrrfXmEahlVHXn3Xk
03 Thermo	!dGkr%XLKQFHFH	APSEHT!ngkXYPfEaHfVH
Structure Count	pattern7	pattern8
12	IaVXg!2XkXg6XnpXl	idp3XLLP
04 Halo	GLAV!g!XlkvgdXanpaLqkXldaXd	pgXsLLP
04 Meso	laVlailXeXg4XnpXlS4XvXp	p2Xk2Xl2xgXdp\$lpk
03 Thermo	DKdGNiaVIGVfFKeGK	In3XLLP2Xk
Structure Count	pattern9	pattern10

12	%yX%XGSLTTPcXEgvXWXV	p2Xvs2XQplnXRX!
04 Halo	vldYwt%pGSLTTPp	pLkdRe!rgf
04 Meso	%yr%XGSLTtPPCtEg!sWfVl	nnXRplQplNXRvVey
03 Thermo	YYrYSGSLTTPCsEGVrWIV	dnNRPvQPINaRX!\$e

In the halotolerant category 1Y7W significant identical residues are QSP, HFHHFLS (catalytic pocket), QLAVIGI, GYDGSLTTP and CDERVKW. The secondary structures of catalytic pocket have no gaps in the catalytic region. The gaps are visible at the first beta pleated sheet region and some gaps are found throughout the aligned sequences in between and at the terminal residues. Most of the gaps are in the loop region with no secondary structure assignment.

The predicted model has one disulphide bridge (between Cys 41 and Cys 247), alpha helices, dominating beta pleated sheets (more pleated sheets compare to helices) and Zn +2 coordinated with Histidine 134,136 and 154 residues at catalytic pocket Figure 3 and Figure 20 . The catalytic triad is highly conserved except 4TWL that do not have metal ion Figure 3, Figure 20, Figure 25 and Table 2

Model quality is verified and acceptable by PROSA-web, DOPE scores and Procheck Figure 21 to Figure 24.

Zinc metal coordination geometry vary but D2SP68, 1Y7W and 5HPJ exhibit consensus in three Histidines' nitrogen interaction with Zinc Figure 26, Figure 27 and Figure 28.

The accessible surface area for 12 CAs and their respective polar residues percent at the surface are given in Table 3.

Table 3 Respective Solvent Accessible Surface Areas (SASA) in Angstrom square and Volume in Angstrom cube S/V is surface and volume ratio

PDB ID	Status	SASA Å ²	Polar %	Vol Å ³	S/V	Chains
1Y7W	Halotolerant	22559	51.5%	53242	0.424	2
4CNX	Halotolerant Engineered	11289	47.3%	28221	0.400	1
5A25	Halotolerant Engineered	22033	45.7%	50644	0.435	2
5HPJ	Psychrohalophilic	10311	49.1%	25455	0.405	1
1KOP	Mesophilic	20946	46.4%	41706	0.502	2
3FE4	Mesophilic	21688	43.7%	48050	0.451	2
4TWL	Mesophilic	22162	53.1%	43146	0.514	2
4XIW	Mesophilic	78026	40.8%	184390	0.423	8
4COQ	Thermophilic	21240	47.8%	52295	0.406	2
4X5S	Thermophilic	19199	49.0%	43945	0.437	2
6EKI	Thermophilic	57111	47.0%	131342	0.435	6
D2SP68	Halotolerant	19247	49.4%	25319	0.760	1

Discussion

In this study several input files were used in different format examples are PDB structures file for protein 3d structure, sequence files Fasta (various format), phylogenetic files etc. We generated output data by using algorithm and modeling software. The data generated was shown in different forms in the Result section and these are new information and were also evaluated for the accuracy for example homology model by PROSA, procheck.

The study mainly focuses on the new information obtained after running the mentioned software and these soft wares were used by many researchers for their work (Kang et al., 2014; Khoddami, Nadri, Moradi, & Sakhteman, 2015).

The significant amount of time was invested in getting right input files in right order so that the quality data generated, and the program could be executed without crashing or error. The obtained data files were later converted into tables or figures with the help of scripts and relevant soft wares. Sometime the data files were further processed to get information and could be presented in easier way.

Multiple and pair wise sequence alignment reveals mutation trend and favored mutation for halotolerant enzymes compare to mesophilic counterpart two halotolerant enzymes are wild type (1Y7W and 5HPJ) (Premkumar et al., 2005; Somalinga, Buhrman, Arun, Rose, & Grunden, 2016), while other two are engineered (mutated version of mesophilic enzyme, 4CNX and 5A25) (Warden et al., 2015). Cysteine residues engaged in forming disulfide bridges are intact in wild type corroborated the significance of bridges stabilize the overall structure in high salt conditions. Proline in coiled and beta pleated sheets are mostly conserved in salt tolerant CAs. Tyrosine is conserved and buried residue in all except in mesophilic (3FE4 and 4XIW) possibly due to missing residues, while in the same patch HW is conserved in halo and thermo CAs and not in mesophilic possibly because of misalignment (pattern 1, Table 2) so pattern 1 is almost same in all types. HWH is well conserved in all except in halo replaced by DWP in 1Y7W and DWR in D2SP68, while W is buried in almost all cases (pattern2 and Figure 76). QSP is also conserved in all cases except 5HPJ replaced by QTP indicates serine can be mutated by threonine for psychrohalophilic prokaryotic CA and the patch is well buried for all CAs. This patch is very significant and considers as signature pattern for the discussed CAs. Q(S/T)P!D is consistent in all halo and thermo CAs where ! is some hydrophobic amino acid, while in meso the signature is replaced by QSP!# where # is some polar amino acid. D is mutated by N or Q in the meso CAs compared to halo CAs. These small changes may bring subtle properties as mentioned before the QSP patch is buried well so the relevant amino acids in the patch should be buried as well but the D or N/Q residues more expose to surface compared to QSP, small polar side chain (D) may justify the balance between hydrophobic and polar environment in halotolerant enzyme without exposing QSP to the surface.

Catalytic triad is almost conserved in all 12 chosen structures (pattern 5 and pattern 6, Table 2) and shows meager differences among halo, meso and thermo CAs. The catalytic triad has typical signature of QfHfH and pSEhtv and the histidine triad is preserved for all 12 CAs. Q is replaced by R in 4TWL and F is replaced by L, I and M in D2SP68, 4TWL and 3FE4 respectively. In case of halotolerant (1Y7W and D2SP68) hydrophobic amino acid I and L appear as IQFHFH and LQLHFH respectively. This exhibits tendency of maintaining hydrophobic environment at the catalytic side relatively and it may help to cope at high salt concentration. In pSEhtv pattern proline is replaced by leucine in 1y7w and D2SP68 may be involved to maintain hydrophobic micro- environment. The halotolerant enzyme should not be exposed to too many ions and these amino acids could possibly work to keep ions away within their vicinity.

Another almost conserved pattern is GSLTTPPc a buried patch (pattern 9, Table 2) could serve as signature pattern for all 12 CAs. In this pattern PPc is substituted by PGC in 1Y7W giving tolerance of mutation of P to G in halotolerant CAs and the cysteine is involved in forming di sulfide bridge another feature of halotolerant enzyme.

At the catalytic pocket three histidines positioned around Zn atom, interact through Nitrogen atom with Zn and keep the entry of water molecule clear (Figure 26, Figure 27 and Figure 28). This suggests that the catalytic mechanism of the enzyme is same for halotolerant enzymes and that is also matched with mesophilic enzymes.

Halotolerant enzymes offered lesser SASA (Table 3) and these enzymes require limited surface exposure in high salt environment. The observation further corroborated as the conserved hydrophobic contact surfaces, hydrophobic interaction and aggregation of halo enzyme were reduced. Thus prevent the loss of activity by adopting the above mentioned changes (Siglioccolo, Paiardini, Piscitelli, & Pascarella, 2011). The Surface/Volume ratio is close to 0.4 in salt tolerant enzymes that is showing the compactness of the enzyme compared to their meso counterpart. This stabilize CA by alleviating exposure to high salt and heat stress conditions that leads to inactivation of CA. (Siddiqui et al., 2018) Polar residues contributed greater than 45% of the SASA in halo and thermo stable enzymes except plant origin meso 4TWL that can survive at extreme physical conditions (Okagami & Kawai, 1982) and prokaryotic 1KOP (Huang et al., 1998) from pathogenic bacteria.

In overall sense the halotolerant enzymes and thermos tolerant enzymes shares many features and acquired many structural changes compare to their mesophilic counterpart. In this way they keep their active site in right conformation and the active site is not flooded by the ions in the hyper saline conditions.

Salt Bridges

Salt bridges formed by D and E in halophilic enzymes resist the denaturing high salt condition (Kastiris, Papandreou, & Hamodrakas, 2007). Halophilic enzymes has more D and E amino acids at their surface, they surpass their mesophilic rivals by increase numbers of D and E and also the appearance of these residues on the surface (Figure 77).

The D2SP68 model built by comparative modeling has more beta pleated sheets (anti parallel) and fewer alpha helices and most of them stacked together (Figure 20). Other halotolerant enzymes also show the same trend that endorse beta pleated sheet is a favored secondary structure. The geometry of alpha helices are less flexible in contrast to beta pleated sheets and halotolerant enzymes undergo many conformational changes to cope the salty environment and keep them functional (Warden et al., 2015)

Conclusions

CA in hypersaline conditions has almost same type of catalytic pocket but the residues around are keeping salt ions out and avoiding the saturation of active site with the flood of salt ions and are mainly acidic residues. CA compact footprint minimizes the interaction of enzyme with ions. The buried residues also have occasionally acidic amino acid that keep the salty ions away from hydrophobic residues and maintain the hydrophobic-hydrophobic interaction of buried residues. This keeps the enzyme structure in native form and avoids enzyme structure disruption in high salt conditions.

References

- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17), 3389-3402. doi:10.1093/nar/25.17.3389
- Bond, C. S., & Schuttelkopf, A. W. (2009). ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Crystallogr D Biol Crystallogr*, 65(Pt 5), 510-512. doi:10.1107/s0907444909007835
- Brock, T. D. (1975). Salinity and the Ecology of *Dunaliella* from Great Salt Lake. *Journal of General Microbiology*, 89(2), 285-292.
- Brylinski, M., & Skolnick, J. (2011). FINDSITE-metal: Integrating evolutionary information and machine learning for structure-based metal-binding site prediction at the proteome level. *Proteins: Structure, Function, and Bioinformatics*, 79(3), 735-751.
- Dasgupta, A. (2016). *Sequence: Robust loop modeling using PyRosetta*. State University of New York at Buffalo.
- Enameh, R. Z., Barker, H., Tolvanen, M. E. E., Ortutay, C., & Parkkila, S. (2014). Bioinformatic analysis of beta carbonic anhydrase sequences from protozoans and metazoans. *Parasites & Vectors*, 7, 1-12.
- Fiser, A., & Sali, A. (2003). ModLoop: automated modeling of loops in protein structures. *Bioinformatics*, 19(18), 2500-2501. doi:10.1093/bioinformatics/btg362
- Fisher, M., Gokhman, I., Pick, U., & Zamir, A. (1996). A salt-resistant plasma membrane carbonic anhydrase is induced by salt in *Dunaliella salina*. *Journal of Biological Chemistry*, 271(30), 17718-17723.
- Huang, S., Xue, Y., Sauer-Eriksson, E., Chirica, L., Lindskog, S., & Jonsson, B.-H. (1998). Crystal structure of carbonic anhydrase from *Neisseria gonorrhoeae* and its complex with the inhibitor acetazolamide¹ Edited by R. Huber. *Journal of Molecular Biology*, 283(1), 301-310. doi:<https://doi.org/10.1006/jmbi.1998.2077>
- Jo, B. H., Seo, J. H., & Cha, H. J. (2014). Bacterial extremophilic carbonic anhydrases from deep-sea hydrothermal vents as potential biocatalysts for CO₂ sequestration. *Journal of Molecular Catalysis B: Enzymatic*, 109, 31-39. doi:<https://doi.org/10.1016/j.molcatb.2014.08.002>
- Kabsch, W., & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers: Original Research on Biomolecules*, 22(12), 2577-2637.
- Kang, Y., Gu, C., Yuan, L., Wang, Y., Zhu, Y., Li, X., . . . Yu, J. (2014). Flexibility and Symmetry of Prokaryotic Genome Rearrangement Reveal Lineage-Associated Core-Gene-Defined Genome Organizational Frameworks. *mBio*, 5(6), e01867-01814. doi:10.1128/mBio.01867-14

- Kanth, B. K., Lee, J., & Pack, S. P. (2013). Carbonic anhydrase: Its biocatalytic mechanisms and functional properties for efficient CO₂ capture process development. *Eng. Life Sci.*, 13, 422-431.
doi:10.1002/elsc.201200157
- Kastritis, P. L., Papandreou, N. C., & Hamodrakas, S. J. (2007). Haloadaptation: Insights from comparative modeling studies of halophilic archaeal DHFRs. *International Journal of Biological Macromolecules*, 41(4), 447-453. doi:<http://dx.doi.org/10.1016/j.ijbiomac.2007.06.005>
- Khoddami, M., Nadri, H., Moradi, A., & Sakhteman, A. (2015). Homology modeling, molecular dynamic simulation, and docking based binding site analysis of human dopamine (D₄) receptor. *J Mol Model*, 21(2), 1-10. doi:10.1007/s00894-015-2579-3
- Kupriyanova, E. V., Sinetova, M. A., Markelova, A. G., Allakhverdiev, S. I., Los, D. A., & Pronina, N. A. (2011). Extracellular β -class carbonic anhydrase of the alkaliphilic cyanobacterium *Microcoleus chthonoplastes*. *Journal of Photochemistry and Photobiology B: Biology*, 103(1), 78-86.
doi:<http://dx.doi.org/10.1016/j.jphotobiol.2011.01.021>
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26(2), 283-291.
doi:10.1107/S0021889892009944
- Laskowski, R. A., & Swindells, M. B. (2011). LigPlot+: Multiple Ligand-Protein Interaction Diagrams for Drug Discovery. *Journal of Chemical Information and Modeling*, 51(10), 2778-2786.
doi:10.1021/ci200227u
- Levitt, M. (1983). Protein folding by restrained energy minimization and molecular dynamics. *Journal of Molecular Biology*, 170(3), 723-764. doi:[https://doi.org/10.1016/S0022-2836\(83\)80129-6](https://doi.org/10.1016/S0022-2836(83)80129-6)
- Lionetto, M. G., Caricato, R., Giordano, M. E., Erroi, E., & Schettino, T. (2012). Carbonic Anhydrase and Heavy Metals. *Biochemistry Prof. Deniz Ekinci (Ed.)*, 205-225.
- Liska, A. J., Shevchenko, A., Pick, U., & Katz, A. (2004). Enhanced photosynthesis and redox energy production contribute to salinity tolerance in *Dunaliella* as revealed by homology-based proteomics. *Plant Physiology*, 136(1), 2806-2817.
- Mandell, D. J., Coutsiias, E. A., & Kortemme, T. (2009). Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nature methods*, 6(8), 551.
- McDonald, I. K., & Thornton, J. M. (1994). Satisfying hydrogen bonding potential in proteins. *Journal of Molecular Biology*, 238(5), 777-793.
- Mitternacht, S. (2016). FreeSASA: An open source C library for solvent accessible surface area calculations [version 1; referees: 2 approved]. *F1000Research*, 5(189).
doi:10.12688/f1000research.7931.1

- Nielsen, H. (2017). Predicting Secretory Proteins with SignalP. In D. Kihara (Ed.), *Protein Function Prediction: Methods and Protocols* (pp. 59-73). New York, NY: Springer New York.
- Okagami, N., & Kawai, M. (1982). Dormancy in *Dioscorea*: differences of temperature responses in seed germination among six Japanese species. *The botanical magazine =Shokubutsu-gaku-zasshi*, 95(2), 155-166.
- Premkumar, L., Greenblatt, H. M., Bageshwar, U. K., Savchenko, T., Gokhman, I., Sussman, J. L., & Zamir, A. (2005). Three-dimensional structure of a halotolerant algal carbonic anhydrase predicts halotolerance of a mammalian homolog. *Proc Natl Acad Sci U S A*, 102(21), 7493-7498. doi:[10.1073/pnas.0502829102](https://doi.org/10.1073/pnas.0502829102)
- Ramachandran, G. N., Ramakrishnan, C., & Sasisekharan, V. (1963a). Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, 7(1), 95-99. doi:[https://doi.org/10.1016/S0022-2836\(63\)80023-6](https://doi.org/10.1016/S0022-2836(63)80023-6)
- Ramachandran, G. N., Ramakrishnan, C., & Sasisekharan, V. (1963b). Stereochemistry of polypeptide chain configurations. *J Mol Biol*, 7, 95-99.
- Šali, A., & Blundell, T. L. (1993). Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Journal of Molecular Biology*, 234(3), 779-815. doi:<http://dx.doi.org/10.1006/jmbi.1993.1626>
- Šali, A., Overington, J. P., Johnson, M. S., & Blundell, T. L. (1990). FROM COMPARISONS OF PROTEIN SEQUENCES AND STRUCTURES TO PROTEIN MODELLING AND DESIGN. In R. A. Bradshaw & M. Purton (Eds.), *Proteins: Form and Function* (pp. 163-171). Oxford: Academic Press.
- Sayle, R. A., & Milner-White, E. J. (1995). RASMOL: biomolecular graphics for all. *Trends in Biochemical Sciences*, 20(9), 374-376. doi:[http://dx.doi.org/10.1016/S0968-0004\(00\)89080-5](http://dx.doi.org/10.1016/S0968-0004(00)89080-5)
- Schenkelberg, C. D., & Bystroff, C. (2015). InteractiveROSETTA: a graphical user interface for the PyRosetta protein modeling suite. *Bioinformatics*, 31(24), 4023-4025. doi:10.1093/bioinformatics/btv492
- Schrödinger, L. L. C. (2017). The PyMOL molecular graphics system, Version 1.8. 2015. *Google Scholar*.
- Shen, M.-y., & Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. *Protein Science*, 15(11), 2507-2524. doi:10.1110/ps.062416606
- Siddiqui, M. H., Alamri, S. A., Al-Khaishany, M. Y. Y., Al-Qutami, M. A., Ali, H. M., Al-Whaibi, M. H., . . . Alharby, H. F. (2018). Mitigation of adverse effects of heat stress on *Vicia faba* by exogenous application of magnesium. *Saudi Journal of Biological Sciences*, 25(7), 1393-1401. doi:<https://doi.org/10.1016/j.sjbs.2016.09.022>
- Siglioccolo, A., Paiardini, A., Piscitelli, M., & Pascarella, S. (2011). Structural adaptation of extreme halophilic proteins through decrease of conserved hydrophobic contact surface. *BMC Struct Biol*, 11(1), 50. doi:10.1186/1472-6807-11-50

- So, A. K.-C., Espie, G. S., Williams, E. B., Shively, J. M., Heinhorst, S., & Cannon, G. C. (2004). A novel evolutionary lineage of carbonic anhydrase (ϵ class) is a component of the carboxysome shell. *J Bacteriol*, 186(3), 623-630. doi:10.1128/JB.186.3.623-630.2004
- Somalinga, V., Buhrman, G., Arun, A., Rose, R. B., & Grunden, A. M. (2016). A High-Resolution Crystal Structure of a Psychrohalophilic α -Carbonic Anhydrase from *Photobacterium profundum* Reveals a Unique Dimer Interface. *PLoS ONE*, 11(12), e0168022. doi:10.1371/journal.pone.0168022
- Tripp, B. C., Smith, K., & Ferry, J. G. (2001). Carbonic Anhydrase: New Insights for an Ancient Enzyme. *Journal of Biological Chemistry*, 276, 48615-48618. doi:10.1074/jbc.R100045200
- UniProt: the universal protein knowledgebase. (2017). *Nucleic Acids Research*, 45(D1), D158-D169. doi:10.1093/nar/gkw1099
- Waditee, R., Hibino, T., Nakamura, T., Incharoensakdi, A., & Takabe, T. (2002). Overexpression of a Na⁺/H⁺ antiporter confers salt tolerance on a freshwater cyanobacterium, making it capable of growth in sea water. *Proceedings of the National Academy of Sciences*, 99(6), 4109-4114.
- Warden, A. C., Williams, M., Peat, T. S., Seabrook, S. A., Newman, J., Dojchinov, G., & Haritos, V. S. (2015). Rational engineering of a mesohalophilic carbonic anhydrase to an extreme halotolerant biocatalyst. *Nat Commun*, 6, 10278. doi:10.1038/ncomms10278
- Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, 35(suppl_2), W407-W410.
- Ye, C., Gao, K., & Giordano, M. (2008). The odd behaviour of carbonic anhydrase in the terrestrial cyanobacterium *Nostoc flagelliforme* during hydration–dehydration cycles. *Environmental Microbiology*, 10(4), 1018-1023.

Figures

```

1kopA   HTHWGYT..GHDSPE$WGNL  CSTGKNQSPVNIIT  ENNGH.TIQV
1y7wA   NDGYDYMQL...HGFDWPG  C.SGSNQSPIDINT  DLE..QGMKV
3fe4A   .....EAHWPHI  C.GGQRQSPINLQR  VNNGH.TVQI
4cnxA   ..HWGYD..KHNGPEHWHE  A.DGERQSPVDIDT  VNDGH.SFNV
4coqA   GAHWGYS..GSIGPEHWGD  CKIGKNQSPIDINS  VNNGH.TIKV
4twIA   .DEFSYIDGNPNGPENWGN  CGKGMESQSPIQLRD  RNSGH.DVLV
4x5sA   ..HWSYE..GENGPENWAK  CN.LKNQSPVDISD  VNNGH.TIQV
4xiwA   .....EVA...GPPTWK  CATGKRQSPINIP  LNTGHGTMQV
5a25A   .HHWGYG..KHNGPEHWHD  A.NGERQSPVDIDT  VNNGH.SFNV
5hpjA   ..EWSYT..GEHGEHWGDS  CAEGVNQITPIDIN  VNNGH.TIQA
6ekiA   ..HWSYH..GETGPQHWGD  CKIGKNQSPVDIS  TNNGH.TIKV
D2SP68A SDKYNIR.....ELGYDW  C.AGDMQSPVNIIR  EQ....DMQ
Consensus ..hw.y.....gpehW..l//c..G.nQSP!#i..//v#ngh.t..v//

1kopA   PSENIQIKGRTFPMEAHFVHLDE.NK  QPLVLAVLYEAG...K.TNGRL  LLP
1y7wA   LSEHTINGIHYPLELHIVMQEQDPA  QLAVIDIMYKYS..EN.GDAFL  LLP
3fe4A   GSEHTVDGIRHVIEIHIVHNSKYK  GLAVLAAFVEVK..NYPENTYY  MLP
4cnxA   GSEHTVDRKKYAAELHLVHNTKYG  GLAVVGVLKVG...DANPAL  LLP
4coqA   PSEHTVNGKHYPFEAHFVHLDK.NG  NITVLGVFFKVG...K.ENPEL  LLP
4twIA   PSEHEMNGERFDLEAQLVHESQ.DQ  KRAVVSILFRFG...R.ADPFL  LQI
4x5sA   PSEHTVNGKYPLEMHVHKDK.DG  NIAVIGVFFKEG...K.ANPEL  LLP
4xiwA   PSEHAMDGRRYAMEAHLVHKNSTG  NLAVLGIMLEPG..GLIKNPAL  LLP
5a25A   GSEHTVDRKKYAAELHLVHNTKYG  GLAVVGVLKVG...DANPAL  LLP
5hpjA   PSENYLKGKQYPLEAHFVHATD.KG  ELAVVAVMFDVG..PR.SNNEL  LLP
6ekiA   PSEHTIKGKSYPFEAHFVHADK.DG  NLAVIGVIFKEG...K.KNPIT  LLP
D2SP68A LSEHTVNGKHFPLEAHLVMKDESD  QLAVIDILYEYTEETQDDPFI  LLP
Consensus //pSEhtv.gk.%p.EaHlVh.....g//.laV.g!..k.g.....np.l//LLP

//

1kopA   YYRFAAGSLTTPPCTEGVSWLV  YDHIDQAQAEKF  NNRPVQPLNARVV
1y7wA   YAGYDGSLLTTPGCDERVKWV  PREVREQMKLF  NNRMIQDLGDREV
3fe4A   YYTYHGSLLTTPPCTENVHVF  FVKLSRTQWK  DYRRTOPLNHRVV
4cnxA   YWTYPGSLTTPPLLESVTWIV  PISVSSEQMLKF  NARPAQPLKDREV
4coqA   YYRYSGSLTTPPCSEGVRWIV  PVEMSREQLEKF  NNRPVQPLNARKV
4twIA   YYRYVGSSETAPPCTEGISWT  VATVSPRQVLL  NARPLQPTNFRSV
4x5sA   YYTYSGSLTTPPCTEGVWLV  PITASKQQIELF  NNRPTQPINSRYI
4xiwA   FVHYPGSLTTPPCSEGVDWV  PIKVPDSQILD  NTRPLQLLNSRLV
5a25A   YWTYPGSLTTPPLLESVTWIV  PISVSSQQMLKF  NARPAQPLKNRCV
5hpjA   YYRFNGSLTTPPCSEGVRWV  PQTSSKAQTEKL  NARPLQPLNARLI
6ekiA   YYRYSGSLTTPPCSEGVRWIV  EMELSKEQIEKF  TNRPVQPLNARMI
D2SP68A HYGYDGSLLTTPPCDERVKWV  TRKVSQAQMKVF  NNRVIOPLNCRQV
Consensus %y.%.GSLTTPPC.EgV.W.V//p..vs..Q..kf//nnRp.QpIn.R.!//

```

Figure 1

Figure 3 Multiple sequence alignment of 12 Carbonic anhydrase halophilic, thermophilic and mesophilic PDB structures by `salgn()`, Consensus regions and mutation trends are shown. Sequences edited by Aline (Bond & Schuttelkopf, 2009)

Halo MSA

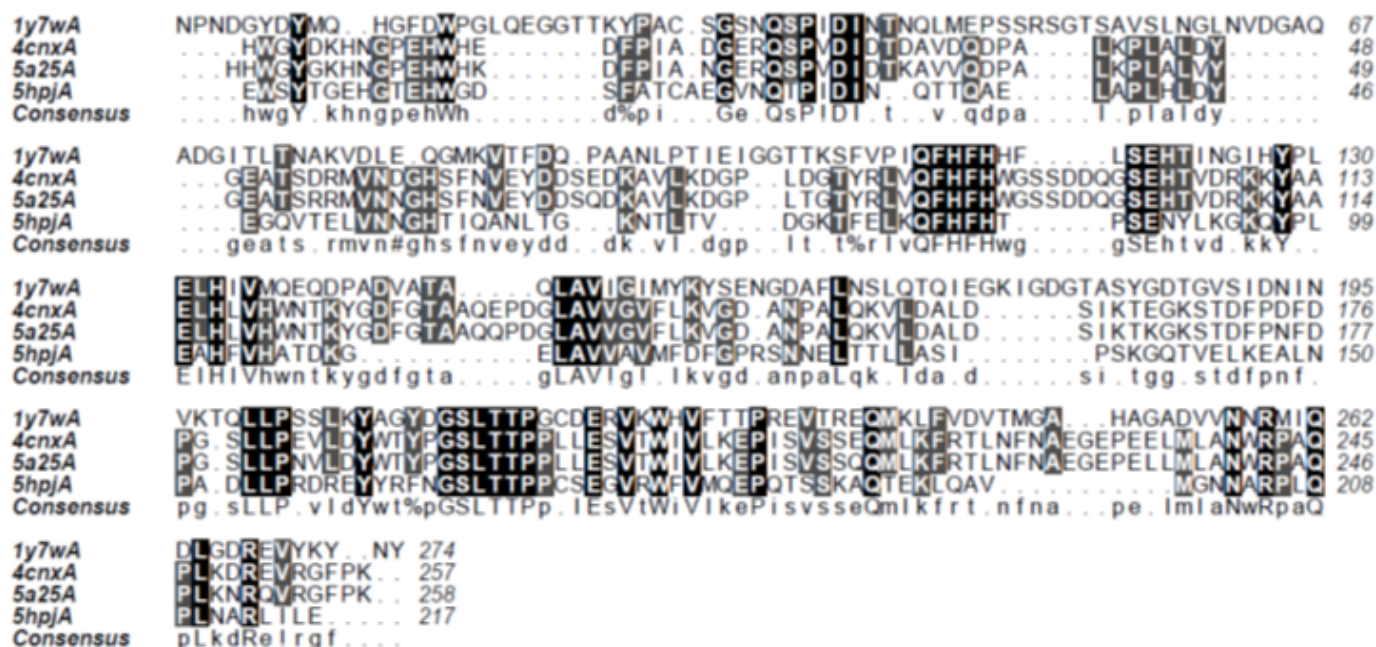


Figure 2

Figure 17 Multiple sequence alignment of halotolerant carbonic anhydrase (CA), shows mutation trend shaded in grey scale, consensus pattern: ! hydrophobic, # Polar, % Aromatic

Meso MSA

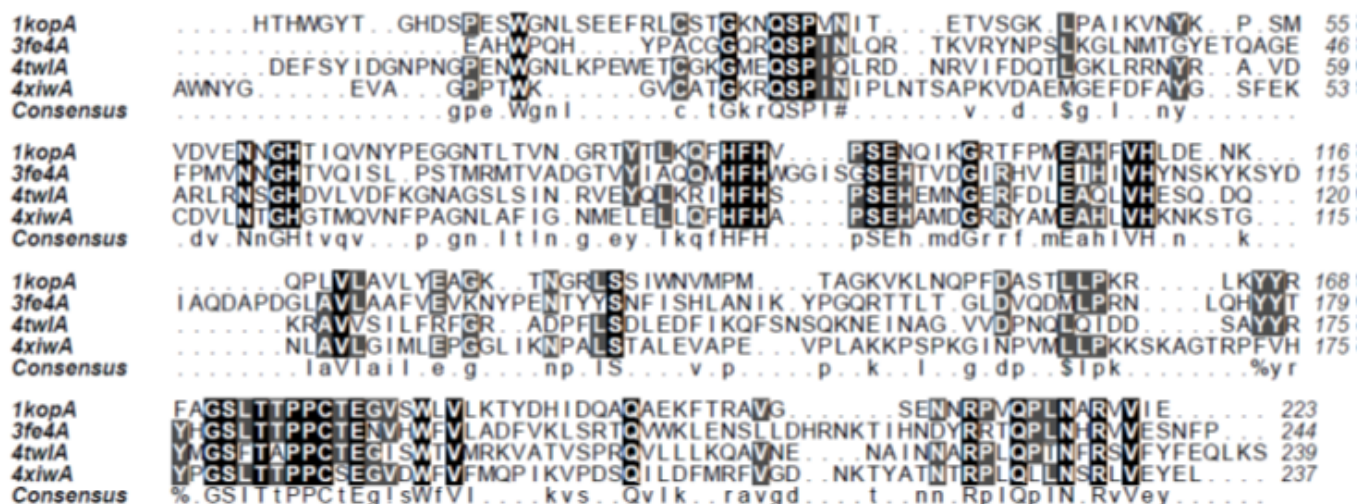


Figure 3

Figure 18 Multiple sequence alignment of mesophilic carbonic anhydrase, shows mutation trend shaded in grey scale, consensus pattern: ! hydrophobic, # Polar, % Aromatic

Thermo MSA



Figure 4

Figure 19 Multiple sequence alignment of thermophilic carbonic anhydrase, shows mutation trend shaded in grey scale, consensus pattern: ! hydrophobic, # Polar, % Aromatic

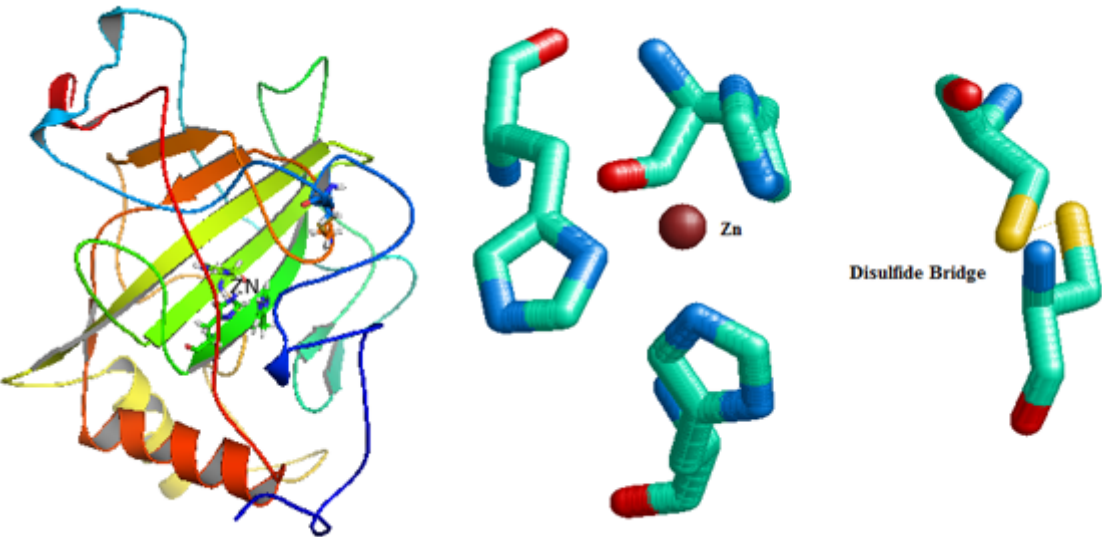


Figure 5

Figure 20 Catalytic triad of His 134,136 and 154 residues within Zn vicinity and Cys Bridge in D2SP68 model (Dunaliella viridis CA) projected by Pymol (Schrödinger, 2017)

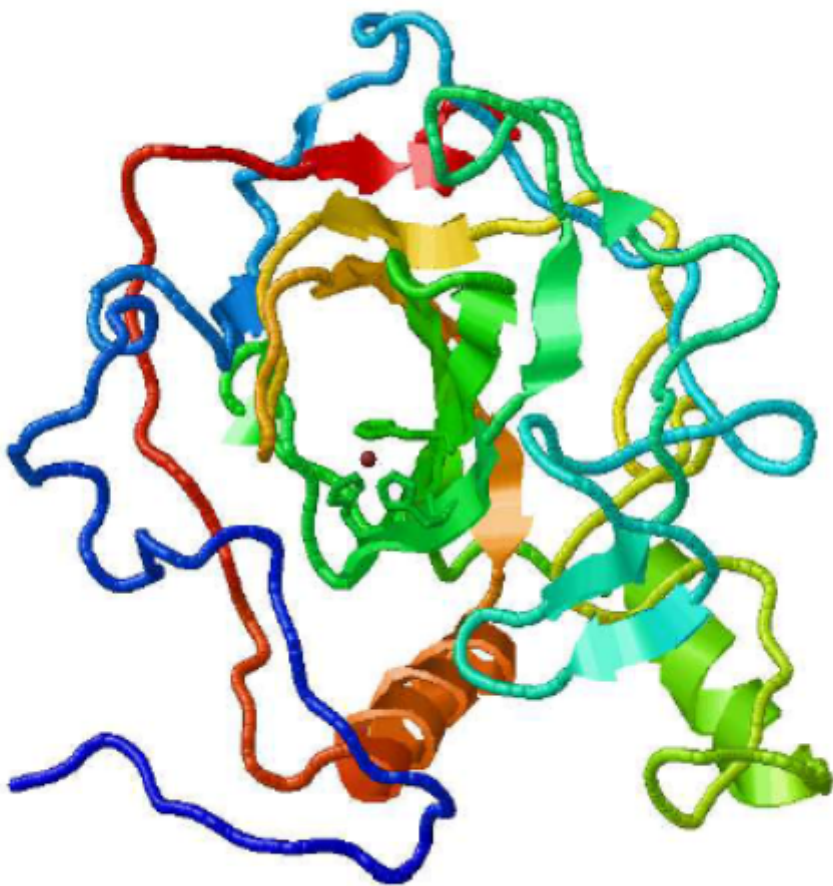


Figure 6

Figure 25 Histidine triad in the D2SP68 model beta pleated sheets shown by arrows and Zinc as sphere surrounded by histidine residues model projected by Rasmol (Sayle & Milner-White, 1995)

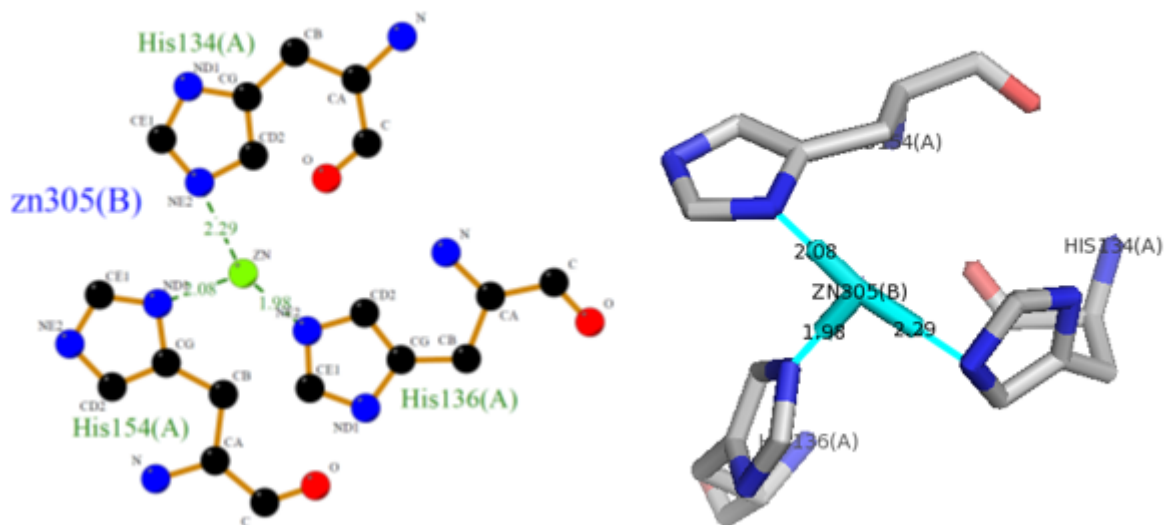


Figure 7

Figure 26 Possible Zn interaction with Nitrogen of Histidines' triad at catalytic site D2SP68 model (Atomic Distances mentioned are in Angstrom) by LigPlot (Roman A. Laskowski & Swindells, 2011)

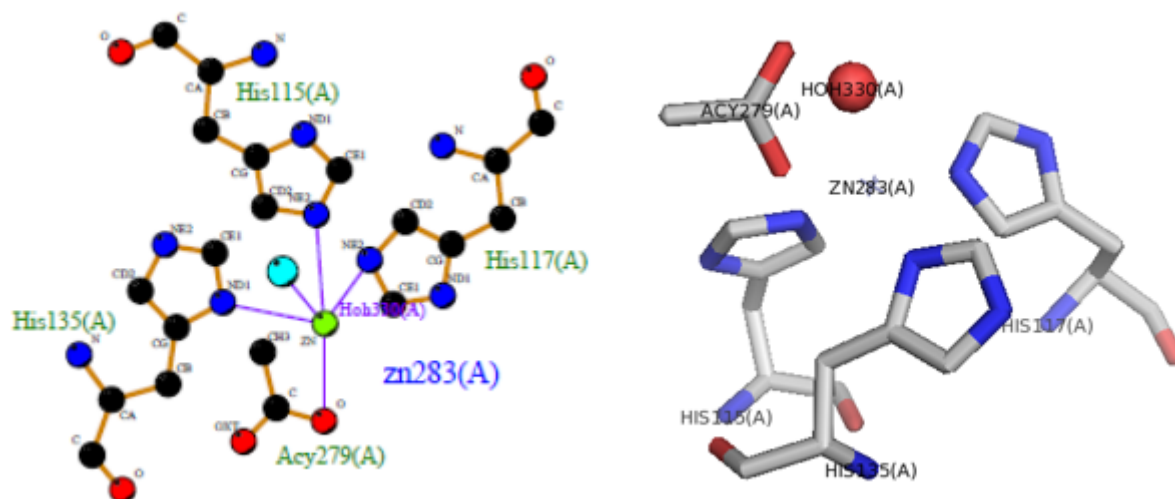


Figure 8

Figure 27 Zn interaction with Nitrogen of Histidines' triad at catalytic site 1Y7W (Dunaliella salina CA, ACY is acetic acid)

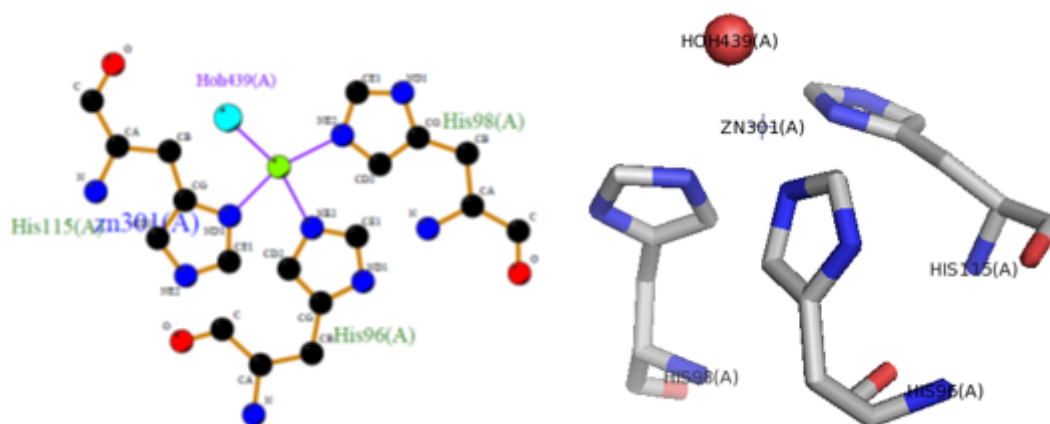


Figure 9

Figure 28 Zn interaction with Nitrogen of Histidines triad at catalytic site 5HPJ (Photobacterium profundum CA) by LigPlot

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupportingInformation.docx](#)