J Comput Aided Mol Des (2014) 28:721–734
DOI 10.1007/s10822-014-9754-y

New insight in the structural features of haloadaptation in α-amylases from halophilic *Archaea* following homology modeling strategy: folded and stable conformation maintained through low hydrophobicity and highly negative charged surface

Mohamed Amine Zorgani · Kevin Patron · Mickaël Desvaux

Received: 13 February 2014 / Accepted: 16 May 2014 / Published online: 28 May 2014

Abstract Proteins from halophilic archaea, which live in extreme saline conditions, have evolved to remain folded, active and stable at very high ionic strengths. Understanding the mechanism of haloadaptation is the first step toward engineering of halostable biomolecules. Amylases are one of the main enzymes used in industry. Yet, no three-dimensional structure has been experimentally resolved for α-amylases from halophilic archaea. In this study, homology structure modeling of α-amylases from the halophilic archaea *Haloarcula marismortui*, *Haloarcula hispanica*, and *Halalkalicoccus jeotgali* were performed. The resulting models were subjected to energy minimization, evaluation, and structural analysis. Calculations of the amino acid composition, salt bridges and hydrophobic interactions were also performed and compared to a set of non-halophilic counterparts. It clearly appeared that haloarchaeal α-amylases exhibited lower propensities for helix formation and higher propensities for coil-forming regions. Furthermore, they could maintain a folded and stable conformation in high salt concentration through highly negative charged surface with over representation of acidic residues, especially Asp, and low hydrophobicity with increase of salt bridges and decrease in hydrophobic interactions on the protein surface. This study sheds some light on the stability of α-amylases from halophilic archaea and provides strong basis not only to understand haloadaptation mechanisms of proteins in microorganisms from hypersalines environments but also for biotechnological applications.

Keywords α-Amylase · Archaea · Halophile · Haloadaptation · Stability · Homology modeling

Introduction

Strategies adopted by some microorganisms to colonize extreme environments have attracted a lot of attention [1]. These extremophiles are adapted to thrive and survive in extreme environments characterized by the presence of hostile conditions, such as high or low temperatures, extreme values of pH, high pressure or high salt concentrations. Considering the latter, there is a wide range of halophilic microorganisms belonging to the domains of *Archaea* and *Bacteria* [2, 3]. To be able to survive, actively grow and propagate at high salinity these microorganisms must keep their cytoplasm at least isosmotic with their external environment, in which salt concentration, i.e. NaCl, in the cytoplasm can reach up to 4 M [4].

The stable and unique native structure of a protein is a basic requirement for its proper catalytic activity [5]. Understanding the haloadaptation mechanisms of proteins in microorganisms from hypersalines environments, i.e. salt lakes, is not only important to address fundamental problems related to protein folding, stability and solubility but also of particular interest given their great potential

Electronic supplementary material The online version of this article (doi:10.1007/s10822-014-9754-y) contains supplementary material, which is available to authorized users.

M. A. Zorgani · M. Desvaux
UR454 Microbiologie, INRA, 63122 Saint-Genès Champanelle, France

Present Address:
M. A. Zorgani (✉) · K. Patron
UMR1282 Bactérie et risques materno-fœtal, Faculté de médecine, INRA, Université François Rabelais de Tours, 10 Boulevard Tonnellé, 37032 Tours, France
e-mail: mohamedamine.zorgani@etu.univ-tours.fr

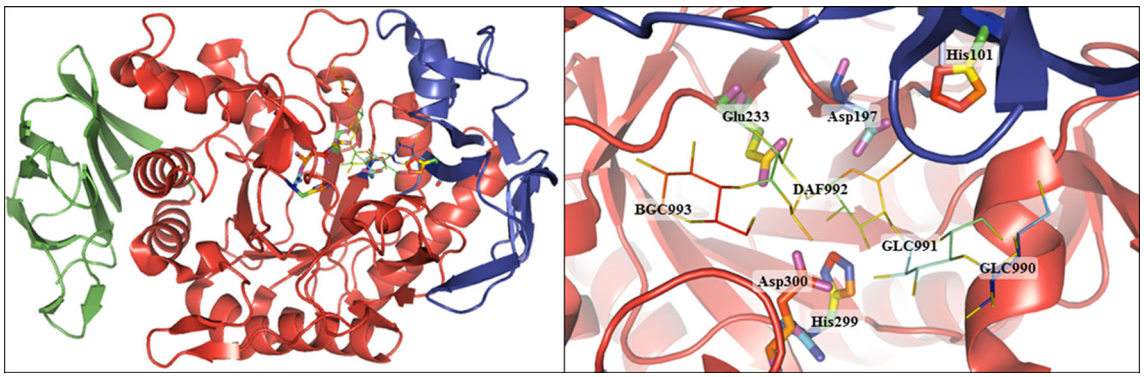


Fig. 1 Structure of a mammalian α-amylase (PDB ID: 1PPI) in a complex with a carbohydrate inhibitor [11]. Left panel shows the domain organization of the α-amylase. Domains A, B and C are shown in red, blue and green respectively. Right panel shows the active center highlighting the amino acids involved in the catalytic site (His101, Asp197, Glu233, His299 and Asp300) in complex with acarbose (a pseudotetrasaccharide α-amylase inhibitor composed of GLC990, GLC991, DAF992 and BGC993). The carboxylic oxygens of the catalytic residues Glu233 and Asp300 form hydrogen bonds with the glycosidic component of the acarviosine group (DAF992 and BGC993). The third residue of the catalytic triad Asp197 is located on the opposite side of the inhibitor binding cleft. GLC: α-D-Glucose, BGC: β-D-glucose, DAF: 4,6-dideoxy-4-{[(1S,5R,6S)-3-formyl-5,6-dihydroxy- 4-oxocyclohex-2-en-1-yl]amino}-α-D-xylo-hex-5-enopyranose

application in the biotechnology industry. For instance, haloarchaea and their enzymes can be used as biocatalysts in application requiring low water activity (A_w), e.g. biochemical reactions in high salt or organic solvent concentrations [6]. Furthermore, these enzymes from halophilic microorganisms could be used in many harsh industrial processes where the concentrated salt solutions used would otherwise inhibit many enzymatic conversions. For instance, halophilic α-amylases have valuable applications in many industries such as pharmaceutical, textile, sugar, brewing, food, and detergents [7, 8]. Even that, halophilic proteins have not been studied as widely as thermophilic proteins due to the difficulty of crystallizing them at high ionic strengths; most halophilic α-amylases are significantly thermotolerant and remain stable at room temperature over long periods [9].

α-Amylases (α-1,4-glucan-4-glucanohydrolase, EC 3.2.1.1) are endoenzymes that act on internal α-glycosidic bonds of polysaccharides, e.g. starch, glycogen, and hydrolyze these bonds to produce α-anomeric mono- or oligosaccharides [10]. On structural level, α-amylases present a single polypeptide chain folded into three domains named A, B, and C (Fig. 1) [11]. The N-terminal of the polypeptide is constituted of the catalytic domain A, which is made up of eight parallel β-strands arranged in a barrel encircled by eight α-helices called TIM barrel (Triose Phosphate Isomerase barrel). Domain B connects the third β-strand and the third α-helix of the TIM barrel presented as extension of domain A. Domain C constitutes the C-terminal part of the polypeptide and it consists of eight anti-parallel β-strands [10]. The TIM barrel is made up of four regions involved in the active site and highly conserved among α-amylase family [11, 12]: (1) the His positioned in C-terminal of the third β-strand interacts with the glucose residues of the substrate; (2) the Asp of the fourth β-strand plays the role of nucleophile during catalysis; (3) the fifth β-strand with Glu acting as proton donor/acceptor; and (4) the fourth region at the seventh β-strand with His and Asp residues that may form hydrogen bonds with glucose residues of the substrate.

The most intensively studied α-amylase for which a three-dimensional (3D) structure has been solved are isolated from *Alteromonas haloplanctis* (AHA) [13], pig pancreatic (PPA) [14], *Aspergillus oryzae* (TAKA) [15], *Bacillus subtilis* (BSUA) [16], and *Pyrococcus woesei/furiosus* (PWA/pfa) [17]. However, no 3D structure has been resolved experimentally for any of α-amylases isolated from the halophilic archaea, namely *Haloarcula marismortui*, *Haloarcula hispanica*, and *Halalkalicoccus jeotgali*. Therefore, there is not any available information about the structural features allowing these haloarchaeal α-amylases to fold in active and stable conformation in very high salt concentrations. However, it has been noted in several studies that the high frequency of negatively charged residues on protein surface is one of the most prominent features in proteins from halophilic micro-organisms [3, 18, 19]. Relatively low hydrophobicity has been pointed as another explanation for the adaptation to high salt environments [19]. Overall, structural insights gathered from all known halophilic protein structures suggest that they exhibit a notable instability in law salt concentration, and rather than being unfolded in high salt concentration they appear to be dependent on the presence of salts [20]. The current study aimed at characterizing the haloadaptation phenomenon and molecular stability in salty environments of some haloarchaeal α-amylases through 3D structure homology modeling strategy using different bioinformatics approaches.

Methods

Sequence retrieval and similarity search

Three sequences of α-amylase isolated from the halophilic archaea *H. marismortui* (amyHm), *H. hispanica* (amyHh), and *H. jeotgali* (amyHj) were retrieved from UniProtKB (http://www.uniprot.org/) under accession numbers: Q5UZY3, Q4A3E0, and D8J7H2, respectively. No 3D structure has been experimentally resolved for any of the three haloarchaeal α-amylases. In order to find the most similar sequence with known 3D structure (template) to each of the target sequences, similarity search were performed using BLASTp [21]. The 3D structures of the templates were extracted from the Protein Data Bank (PDB) (http://www.rcsb.org/). The multiple alignments with the targets and the templates were performed using ESPript v3.0 [22].

In silico structure modeling

With a view to study the 3D structures of the selected haloarchaeal α-amylases, comparative modeling was performed using the program MODELLER 9v8 [23]. The encoded 3D structures extracted from the PDB database 1WZA; 1MXG; and 1UOK, were used as templates for the construction of a set of 20 models for the target sequences amyHm; amyHh; and amyHj, respectively and separately. The template 1WZA is an α-amylase isolated from the species *Halothermothrix orenii*, and exhibits halophilic and thermophilic characteristics at the same time with no acidic-surface [24]. The template with the PDB ID 1MXG is an α-amylase produced by the hyperthermophilic *Pyrococcus furiosus,* and has the ability to bind both metals zinc and calcium [17]. The last template with the PDB ID 1UOK isolated from *Bacillus cereus*, is an oligo-1,6-glycosidase (dextrin 6-alpha-glucanohydrolase, EC 3.2.1.10) very close to α-amylase family [25].

Based on the templates chosen for each haloarchaeal α-amylase, the resulting models generated by MODELLER were conducted into steepest descent energy minimization using Gromacs 4.0.5 with the GROMOS96 45a3 force field [26].

The quality of our models was assessed by two types of energy functions: a semi-empirical force field GROMOS and a knowledge-based energy function Anolea [27]. The GROMOS and Anolea scores were calculated and normalized into *Z score* as described by [28]. We have selected for each target, the best model presenting the lowest average *Z score* for further structural analysis.

Structural analysis and electrostatic potential calculations

Pairwise structural alignments for each of our selected haloarchaeal models with a set of 19 non-halophilic homologues α-amylases with known 3D structure (Table 1) were performed to identify structural similarities and differences using Dalilite [29]. The secondary structure was predicted using DSSP program implemented in Dalilite [30]. A visual inspection of the constructed models regarding the amino acid composition and localization was also performed with Pymol v1.1 [31]. The surface electrostatic potential of the α-amylases was calculated using the program ABPS (Adaptive Poisson-Boltzmann Solver) implemented in Pymol [32].

Table 1 PDB IDs of the non-halophilic α-amylases used in the structural comparison with the haloarchaeal models

α-Amylase origin	PDB ID
Pyrococcus woesei	1MW0
Sulfolobus solfataricus	1EH9
Bacillus subtilis	1UA7
Bacillus amyloliquefaciens	3BH4
Bacillus halmapalus	2JGP
Bacillus sp KR 8104	3DCO
Bacillus stearothermophilus	1HVX
Bacillus licheniformis	1OB0
Bacillus sp KSM-K38	1UD2
Deinococcus radiodurans	2BHU
Lactobacillus plantarum	3DHU
Thermoactinomyces vulgaris	2D0F
Thermotoga maritima	2B5D
Aspergillus niger	2AAA
Aspergillus oryzae	3KWX
Homo sapiens	3OLD
Hordeum vulgar	3BSG
Sus scrofa	3L2L
Tenebrio molitor	1CLV

Surface amino acids, hydrophobic interactions, and salt bridges calculations

The average number of charged amino acids (Asp, Glu, His, Lys, and Arg) that lies on the surface of the haloarchaeal models, was calculated and compared to the average number of charged amino acids that lies on the surface of the above mentioned set of non-halophilic structures, using a cut-off of 7 as relative solvent accessibility value (*rsav*). An amino acid that has all atoms *rsav* less than or equal to 7 was considered as buried, whereas all other resides with >7 *rsav* were considered to lie on the surface of the structures. This procedure was carried out with PIC program using NACCESS as calculator [33].

The potential number of intraprotein hydrophobic interactions, as well as the number of salt bridges on the constructed models of haloarchaeal α-amylases was calculated

utilizing PIC program and compared to that of non-halophilic homologues. The parameters of calculations for both hydrophobic interactions and salt bridges were set to default [33].

Amino acid composition of halophilic and non-halophilic α-amylases

To find the differences in amino acid composition (ΔAAC) between haloarchaeal and non-halophilic α-amylases, the average value of AAC from the haloarchaeal α-amylases sequences was calculated and compared to the average value of AAC from 5 sets of non-halophilic α-amylases sequences deposited in UniProtKB under EC number 3.2.1.1. This procedure was carried out using MEGA v5.05 [34]. After reducing sequence redundancy using UniRef100 from UniProt, these 5 sets includes 806 α-amylase sequences: 58 sequences from *Archaea*, 525 sequences from *Bacteria* species, 34 sequences from fungi, 139 sequences from metazoan group, and 50 α-amylase sequences belong to plants. The differences in the relative hydrophobic values (ΔRHV) of polar (D, E, H, R, K, Q, N, T, S) and apolar amino acids (V, A, I, L, P, C, M, F, W, Y, G) that composed α-amylase sequences were calculated as described by [35].

Statistical analysis

Differences between the structural properties, AAC and RHV regarding haloarchaeal α-amylases and non-halophilic homologues were calculated whenever was possible. For instance, ΔAAC indicates haloarchaeal minus non-halophilic amino acid composition. The differences significance was evaluated using the *t* test where the null hypothesis is that the average Δ was 0 at 0.05 *p* value. A *p* value <0.05 is considered as significant. A *p* value <0.005 is considered as very significant. A *p* value <0.0005 is considered as very high significant.

Results and discussion

Sequence similarity analysis

Similarity search were performed in order to find the most similar sequence with known structure to generate 3D structure models for the sequences of the haloarchaeal α-amylases isolated from *H. marismortui* (amyHm), *H. hispanica* (amyHh), and *H. jeotgali* (amyHj). The two sequences amyHm and amyHh shares similar sequence identity with the sequences of 1WZA and 1MXG, respectively, other homologues with experimentally determined structure. Whereas, the haloarchaeal α-amylase secreted by *H. jeotgali* shares high sequence identity and similarity with an oligo-1,6-glycosidase secreted by *Bacillus cereus* (Table 2). In general, the overall sequence similarity of α-amylases is low and can fall below 10 % with some members [17].

The multiple sequence alignment of the three haloarchaeal α-amylase sequences (amyHm, amyHh, and amyHj) with their respective template sequences shows that they present conserved active site residues known among α-amylase family (Fig. 2). The overall conservation of the catalytic residues among the haloarchaeal α-amylases suggests that their enzymatic mechanism is very similar to the double-displacement catalytic mechanism of other known α-amylases [11, 12]. Of note those α-amylases are most certainly secreted via the Sec export pathway as indicated by the presence of N-terminal signal peptide in amyHh for instance [36–38]. In general, aspects related to protein secretion in *Archaea* would require further in-depth investigations. Furthermore, the secondary structure alignment shows that all the 3D structures used as templates (1WZA, 1MXG and 1UOK) to build the haloarchaeal α-amylase models conserve the same global structural organization known among α-amylase family, domains A (TIM barrel), B, and C (Fig. 2). Except for the template 1UOK, where in addition to the TIM barrel, it presents five more α-helix positioned as it follows: α-helix (Nalpha6′) between the sixth β-strand (Nbeta6) and the sixth α-helix (Nalpha6); Nalpha7′ between Nbeta7 and Nalpha7; and three α-helix (Nalpha8′, Nalpha8″, and Nalpha8‴) between Nbeta8 and Nalpha8 [25]. In order to get further insight into the sequence similarity with regards to structural resemblance and as non-structural information is available in the literature; the structure of haloarchaeal α-amylases was modeled using homology modeling and the three selected templates were used to generate the haloarchaeal models.

Table 2 Sequence identity and similarity between the sequences of haloarchaeal α-amylase and their respective template sequences. The *rmsd* values (atoms superimposed/atoms aligned) are calculated after superposition with Pymol [31]

Target ID	Template PDB ID	Sequence identity (%)	Sequence similarity (%)	E value[a]	*rmsd*
amyHm	1WZA	30.0	46.0	$2.0e^{-35}$	0.359 (241/343)
amyHh	1MXG	27.0	45.0	$2.0e^{-33}$	0.273 (226/346)
amyHj	1UOK	46.0	64.0	$1.0e^{-150}$	0.173 (442/545)

[a] The E value indicates the reliability of the alignment, as match is lower as the alignment is significant

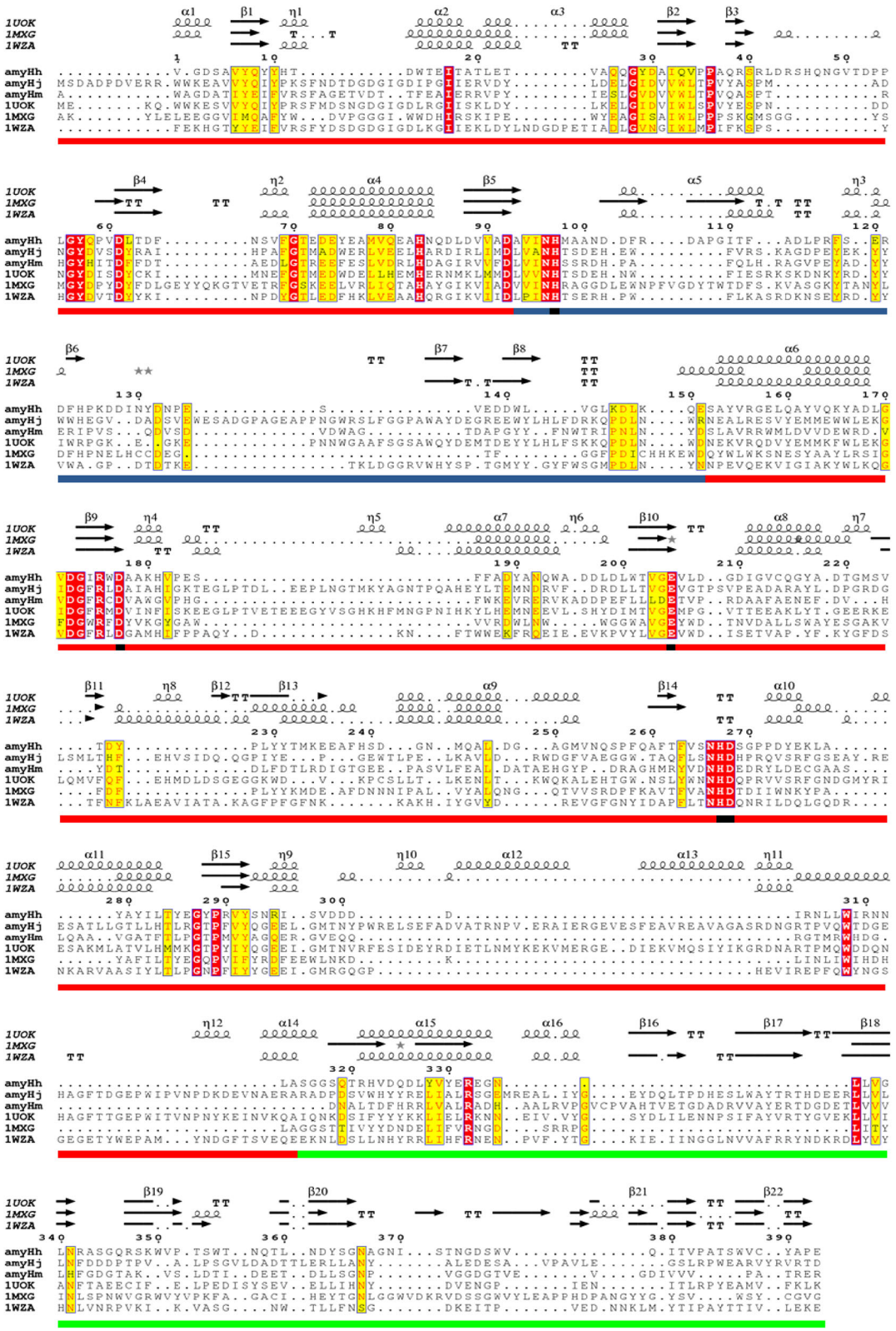


Fig. 2 Sequence similarity analysis of α-amylases. The multiple sequence alignment of the haloarchaeal α-amylases and their respective templates was performed by ESPript v3.0 [22]. On the top of the sequence alignment is represented the secondary structure alignment of the templates. Conserved residues are *white* colored in *red squares*. Similar residues are *red* colored in *yellow squares*. Domains A, B and C are delimited by the underscore *squares* colored in *red*, *blue* and *green* respectively. Residues involved in the active site are indicated with *black underscore squares*. α: α-helix, β: β-strand, η: 3_{10}-helices, T: turn, TT: strict β-turns

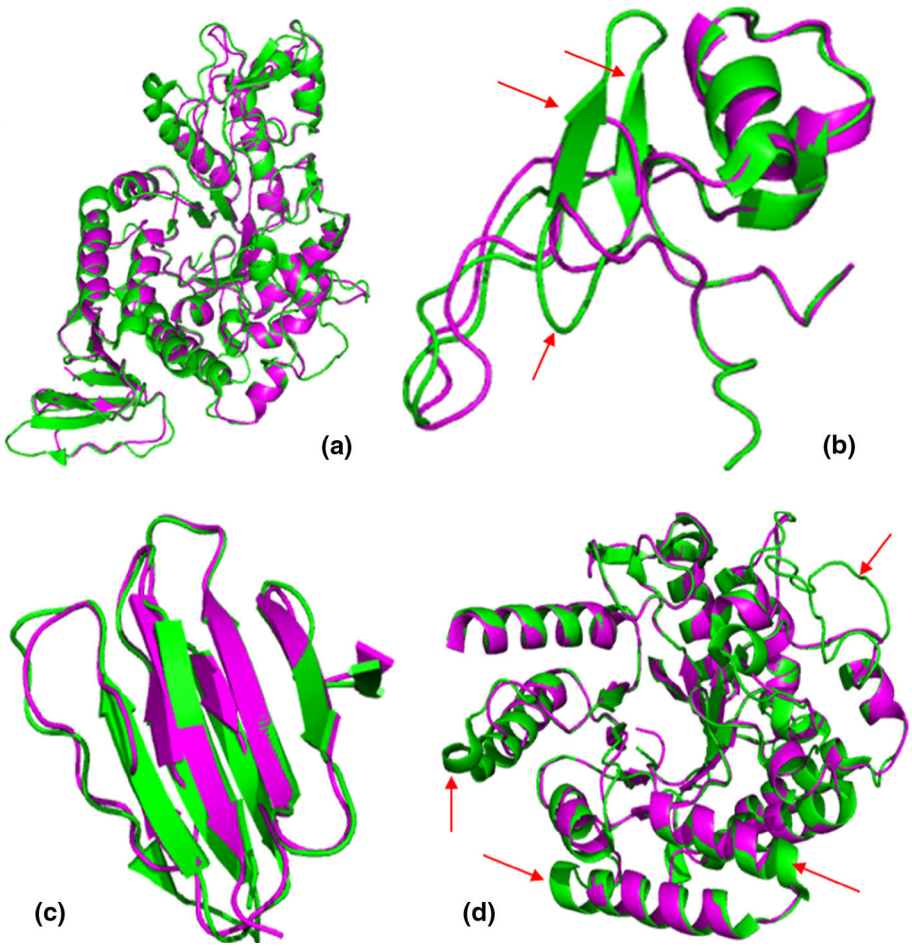


Fig. 3 Structural analysis of the amyHm haloarchaeal model. **a** Superposition of the selected haloarchaeal model of α-amylase from *H. marismortui* (in *magenta*, amyHm) and its respective template (in *green*, PDB ID: 1WZA), superimposed and visualized with Pymol [31]. **b–d** Zoom on domain B, C, and A respectively. The *red arrows* are pointed to the structural differences between the target and the template

Building of α-amylase models

The haloarchaeal α-amylase models were generated by homology modeling. The model with the lowest *Z score* was selected for further structural analysis. Model number 19 (amyHh_m19) was selected for amyHh, amyHj_m17 for amyHj, and amyHm_m12 for amyHm (see supplementary material, Tables S1, S2, and S3).

Structural superposition of the three selected models with their respective templates, shows structural resemblance highlighted with very low *rmsd* values (Table 2). The overall structural resemblance (target/template) covers domains A, B, and C. This result should enhance the reliability of our selected models, on structural level, to be used for further structural analysis. The Fig. 3 shows the structural resemblance between the selected α-amylase model for amyHm and its template (PDB ID: 1WZA). Interestingly, structural analysis of domain A, B, and C separately show while there is a marked conservation of the rigid structures (α-helix and β-strand), there is structural differences located especially on the coil-forming regions (Fig. 3b–d). In domain B of the template (Fig. 3b), two β-strands were replaced by coils in the same domain of the target. Whereas, in domain A (Fig. 3d), parts of the α-helixes were also replaced by coil-forming regions. The same structural trends where observed when the haloarchaeal models of amyHh and amyHj where analyzed and compared to their respective templates 1MXG and 1UOK (Figs. 4, 5). This unique structural organization of the haloarchaeal α-amylase models could highlight the fact that these coils regions could be specific characteristics of haloarchaeal α-amylase.

Comparison with known protein structures

In order to deeply analyze the previous structural differences between the haloarchaeal α-amylases models and their non-halophilic counterparts, we have performed structural comparisons taking into account the various forms of protein folding (α-helix, β-strand and coil-forming regions). Compared to non-halophilic structures, the haloarchaeal models present an increase in coil-forming regions (52.7 % for amyHh, 55.1 % for amyHm, and 52.3 % for amyHj) and a decrease in α-helix and β-strands-forming regions (Table 3). Note that, the halophilic enzymes must assume a balance between being

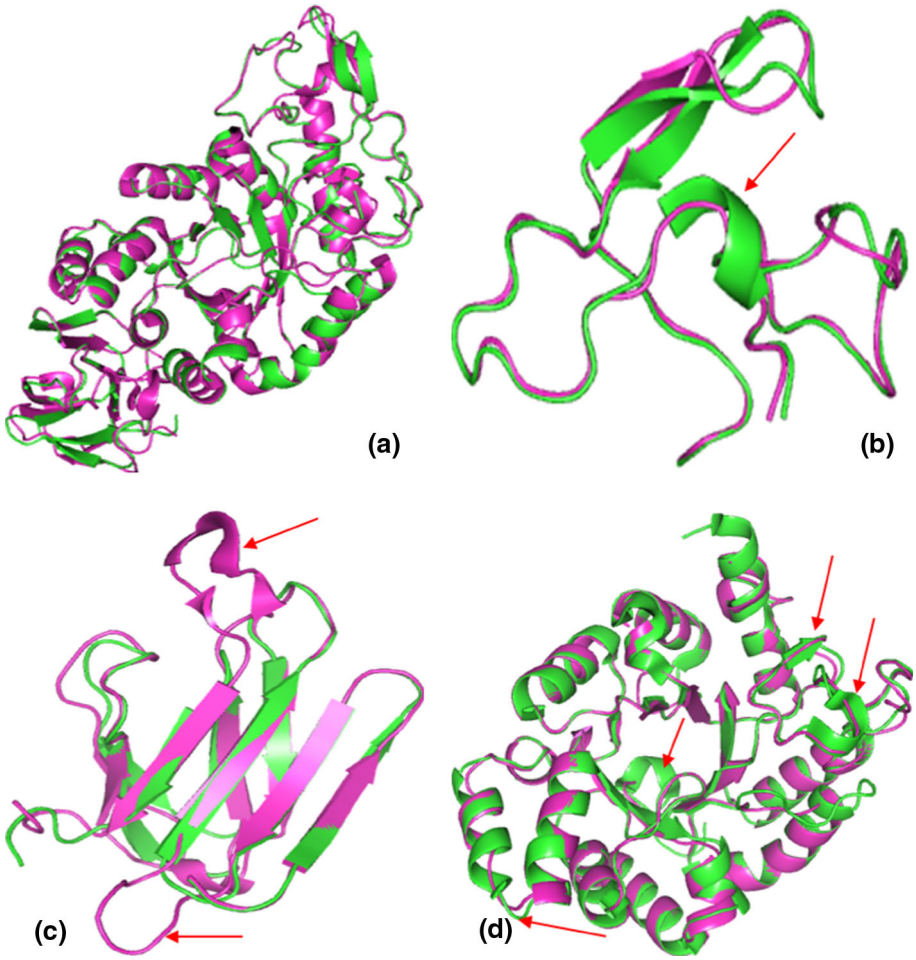


Fig. 4 Structural analysis of the amyHh haloarchaeal model. **a** Superposition of the selected haloarchaeal model of α-amylase from *H. hispanica* (in *magenta*, amyHh) and its respective template (in *green*, PDB ID: 1MXG), superimposed and visualized with Pymol [31]. **b–d** Zoom on domain B, C, and A respectively. The *red arrows* are pointed to the structural differences between the target and the template

flexible enough to enable catalytic activities and rigid enough to avoid unfolding. Presumably, the coil regions provide a flexible conformation to the haloarchaeal α-amylases, which could enhance their stability and activity in environments with high ionic strength. Non-halophilic α-amylases, in other hand, present an increase in α-helix and β-strand-forming regions (Table 3). In this case, these rigid conformations could be one of the reasons causing lowering in dynamic motions of the enzyme in high salt concentration, known to be required for efficient catalysis and activity in salty environments [39].

Moreover, the structural inspection of the haloarchaeal models shows that the amino acid Asp is mostly in coil conformation, whereas the amino acid Glu is mostly in α-helix or β-strand conformations (Fig. 6b). From a structural viewpoint, Asp is recognized as an α-helix breaker, whereas Glu is favorable for α-helix formation [40]. In general, the coil regions of proteins are known to prefer Asp over Glu [41]. This could be probably the reason why halophilic proteins use more prevalently Asp than Glu residues.

As it will be discussed further below, residues having lower propensities for α-helical and β-strand regions and higher propensities for forming coil regions, such as Asp and Val, are preferred more in haloarchaeal α-amylases than in non-halophilic counterparts. This particular flexible structural organization among the haloarchaeal α-amylases, could be one of features conferring stability of the 3D structure, to help the enzyme in avoiding the loss of function in extreme salt concentrations by staying folded in the right conformation.

Table 3 Comparison of the average number of secondary structure between the α-amylase models of haloarchaeal species (in bold), the templates (underlined), and the structures of non-halophilic species

α-amylase origin	α-Helix (%)	β-Strand (%)	Coil (%)
Haloarcula hispanica	**27.48**	**19.84**	**52.67**
Haloarcula marismortui	**28.14**	**16.78**	**55.08**
Halalkalicoccus jeotgali	**31.11**	**16.58**	**52.30**
Halothermothrix orenii	31.35	18.03	50.61
Bacillus cereus	33.51	17.38	49.10
Pyrococcus woesei	30.66	21.88	47.46
Bacillus stearothermophilus	26.51	25.25	48.24
Lactobacillus plantarum	35.80	19.66	44.54

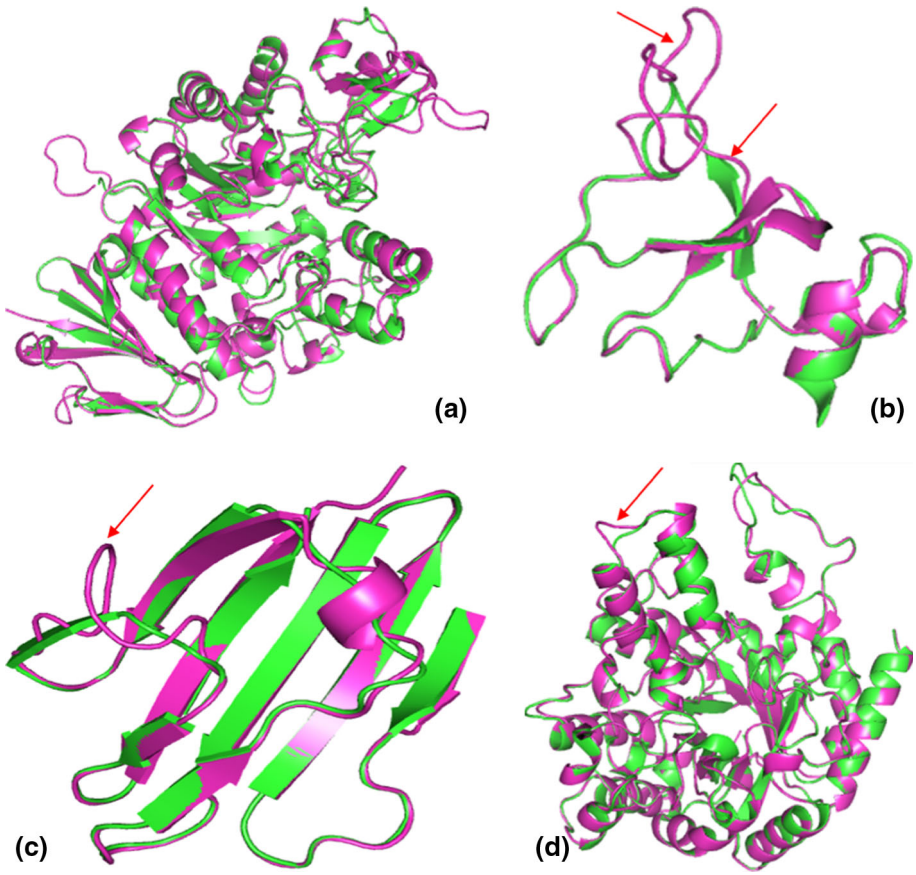


Fig. 5 Structural analysis of the amyHj haloarchaeal model. **a** Superposition of the selected haloarchaeal model of α-amylase from *H. jeotgali* (in *magenta*, amyHj) and its respective template (in *green*, PDB ID: 1UOK), superimposed and visualized with Pymol [31]. **b–d** Zoom on domain B, C, and A respectively. The *red arrows* are pointed to the structural differences between the target and the template

Surface electrostatic potential and intraprotein interactions

The stability of a protein is limited by the nature of the surrounding environment. Halophilic archaea lives in a continuous contact with high salt concentrations. In such harsh conditions, the solvent exposed regions of proteins play an important role for their stability. The calculation of surface electrostatic potential reveals a highly acidic surface for the three haloarchaeal models (Fig. 6a). Structural insights gathered from all known halophilic proteins structures suggest that the conserved acidic surface enhances the stability of the protein through an increased water binding capacity [42], for instance the esterase protein from *H. marismortui* [43]. Unlike all the other haloarchaeal models, the template used for the homology modeling of amyHm (1WZA in Fig. 6a) lacks the conserved acidic surface and both positively and negatively charged residues are distributed equally [24]. It should be mentioned that the α-amylase from the thermo-halophilic *Halothermothrix orenii* with the structure 1WZA exhibit a different haloadaptation mechanism. While 1WZA does not have an acidic surface it does have other structural characteristics, such as, the presence of an equal number of positively charge residues, binding of two calcium ions and surface exposed salt bridges that form ions network known to stabilize halophilic and thermophilic proteins [24]. Further inspection of the surface-amino acid composition reveals significant differences between the haloarchaeal α-amylase and the 19 non-halophilic counterparts structures (Table 4). The haloarchaeal models present significantly more Asp residues on the surface of the protein than the non-halophilic homologues, and significantly less Lys residues. The higher usage of negatively charged amino acid residues, leads to the organization of a hydrated salt ion network at the surface of the protein and formation of salt bridges with strategically positioned basic residues. Otherwise, by accumulating a large number of negatively charged residues in the surface, mainly Asp residues, the tendency to aggregation will be reduced and the solubility at physiological pH will be increased due to the lowering in the isoelectric point.

Interestingly, the calculation of the number of salt bridges on the haloarchaeal models did not show any significant trends, compared to non-halophilic homologues (Table 5). Nevertheless, the balance between the coil-forming regions and the salt bridges on haloarchaeal α-amylases may confer the necessary rigidity to the enzyme

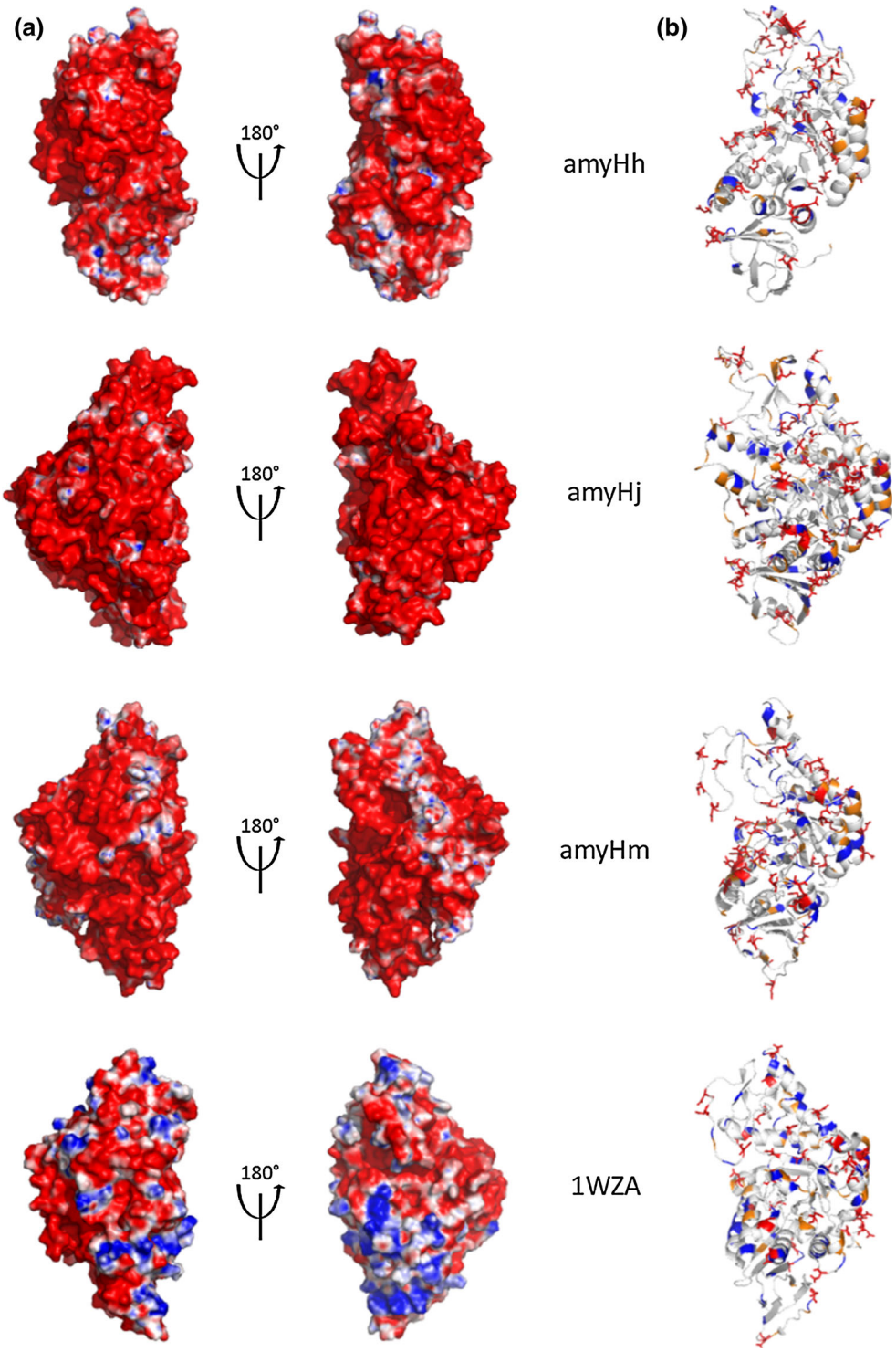


Fig. 6 Electrostatic properties and surface charge of the of haloarchaeal α-amylase models. **a** The electrostatic surface potential. Electrostatic drawings were produced using the program ABPS (Adaptive Poisson-Boltzmann Solver) implemented in Pymol [32]. Surface colors represent the potential from $-10k_BT^{-1}$ (*red*) to $+10k_BT^{-1}$ (*blue*). **b** Structural representation of the 3D structures showing the amino acids proportion and localization. Negatively charged residues Asp and Glu are shown in *red sticks* and *orange*, respectively. Positively charged residues His, Arg, and Lys are shown in *blue* and neutral residues are in *white*

to avoid unfolding in high ionic strength. At high salt concentrations, the negative free energy of ion binding to the sites formed by the salt bridge networks will counterbalance screening of protein-charged groups by solvent ions [44]. With a stabilizing effect and taking in account the balance between the flexibility provided by the coil-forming regions and the rigidity provided by the salt bridges, this features constitutes an important structural signature of haloarchaeal α-amylase that could contribute to their stability.

The number of predicted intraprotein hydrophobic interactions shows a significant decrease in the average

Table 4 Average number of charged residues lying on the surface of the haloarchaeal α-amylase models and the non-halophilic homologues

α-Amylase origin	Negatively charged residues		Positively charged residues		
	ΔAsp[a]	ΔGlu	ΔHis	ΔLys	ΔArg
Halophilic archaea					
Non-halophilic archaea	13.04	−5.55	1.12	−6.10	4.38
Bacteria	13.74	7.83	0.06	−4.61	5.29
Fungi	3.15	13.67	1.81	−5.63	8.81
Metazoan group	8.41	11.84	1.81	−7.19	7.98
Total[b]	38.35	27.79	4.80	−23.52	26.46
Average[c]	9.59	6.95	1.20	−5.88	6.61
t test[d]	**0.001**	0.110	0.240	**0.022**	**0.003**

[a] Number of Asp residues lying on the surface of the haloarchaeal α-amylase models minus the number of Asp lying on the surface of the non-halophilic homologues

[b] Total Δ residue (Asp, Glu, His, Lys or Arg)

[c] Average Δ residue (Asp, Glu, His, Lys or Arg)

[d] Boldfaced numbers indicate significant *p* values

number of hydrophobic interactions in haloarchaeal models, highlighted by an increased number of hydrophobic interactions in case of non-halophilic structures (Table 5). Indeed, maintaining stability and native structure through a correct protein folding in hypersaline environment may require less hydrophobic interaction to avoid collapse. As it was mentioned in a previous study, halophilic protein present 2-fold reduction in the proportion of lysine residues in the sequence leading to a 4-fold reduction in the exposed hydrophobic accessible surface compared to non-halophilic homologues [20]. For halophilic proteins, the lowering in the strength of hydrophobic interactions is commonly viewed as the main reason for preventing aggregation of proteins at high salt concentration [19]. Indeed, all structural analyses for the haloarchaeal models, mainly intraprotein interactions, were conducted on structures obtained from homology modeling. However, it is worth mentioning that before calculating these intraprotein interactions, the models were conducted to energy minimization and evaluation which makes the resulting models more reliable for such structural analysis.

Table 5 Intraprotein salt bridges and hydrophobic interactions

α-Amylase PDB ID	Sequence length	Salt bridges (SB)			Hydrophobic interactions (HI)		
		Number of SB	SB (%)[a]	ΔSB[b]	Number of HI	HI (%)[c]	ΔHI[d]
Haloarchaeal α-amylases							
1WM0	435	55	6.32	−0.85	487	55.98	−20.32
1EH9	558	95	8.51	−3.04	546	48.92	−13.26
1UA7	425	53	6.24	−0.77	362	42.59	−6.93
3BH4	484	68	7.02	−1.55	400	41.32	−5.66
2JGP	485	52	5.36	0.11	502	51.75	−16.09
3DCO	425	73	8.59	−3.12	597	70.24	−34.58
1HVX	483	58	6.00	−0.53	440	45.55	−9.89
1OB0	483	57	5.90	−0.43	409	42.34	−6.68
1UD2	480	57	5.94	−0.47	414	43.13	−7.47
3DHU	427	45	5.27	0.20	436	51..05	−15.39
2BHU	602	90	7.48	−2.01	532	44.19	−8.53
2D0F	637	46	3.61	1.86	565	44.35	−8.69
2B5D	528	95	9.00	−3.53	524	49.62	−13.96
2AAA	476	29	3.05	2.42	457	48.00	−12.34
3KWX	476	37	3.89	1.58	467	49.05	−13.39
3OLD	496	59	5.95	−0.48	424	42.74	−7.08
3BSG	404	44	5.45	0.02	395	48.89	−13.23
3L2L	496	60	6.05	−0.58	423	42.64	−6.98
1CLV	471	43	4.56	0.91	403	42.78	−7.12
Total[e]				−10.25			−227.59
Average[f]				−0.54			−11.98
t test[g]				0.30			**0.004**

[a] Percentage of salt bridges predicted on the structures of non-halophilic α-amylases

[b] Percentage of salt bridges predicted on the structures of the haloarchaeal α-amylases minus the percentage of salt bridges predicted on the structures of non-halophilic homologues

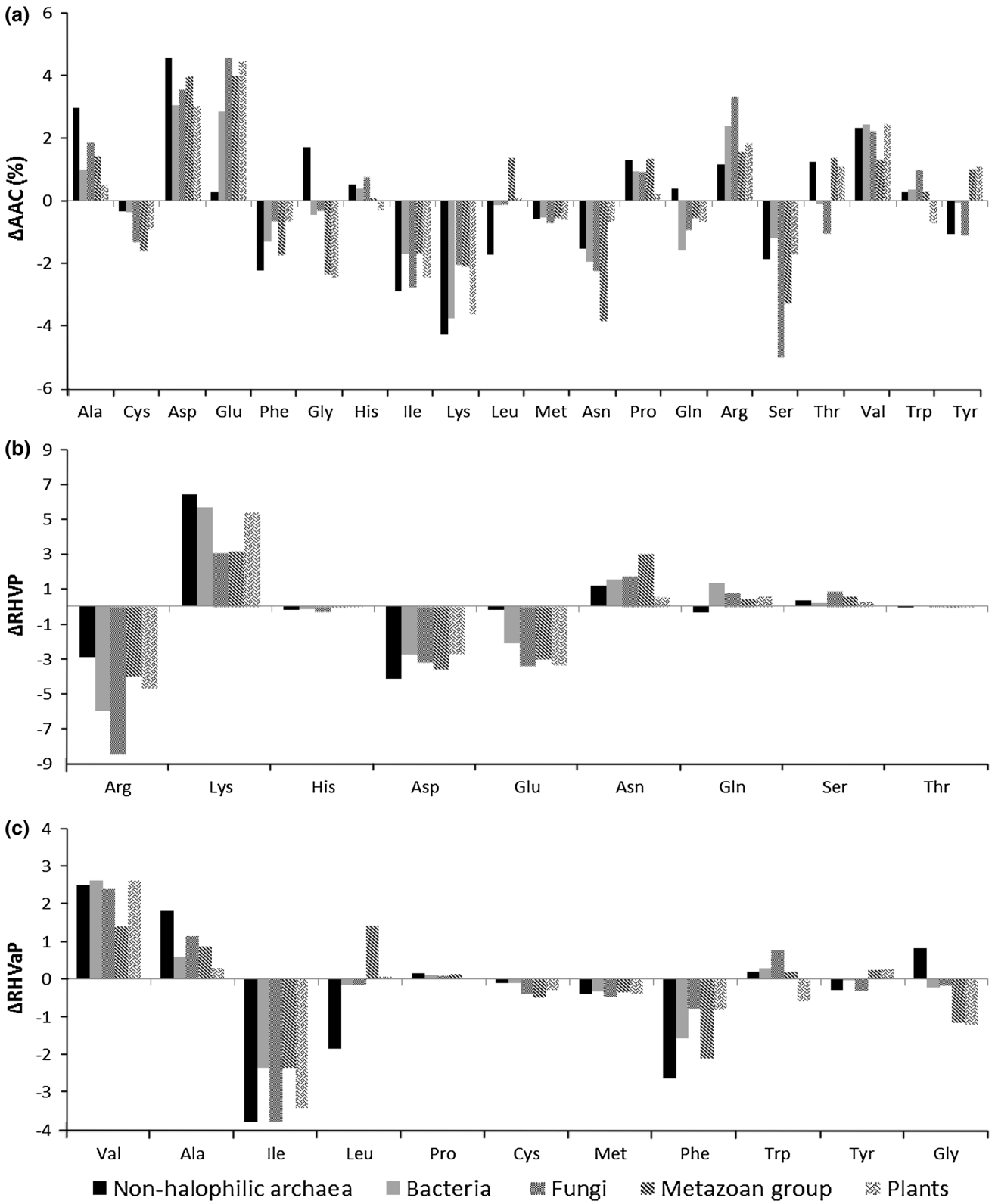
[c] Percentage of hydrophobic interactions predicted on the structures of non-halophilic α-amylases

[d] Percentage of hydrophobic interactions predicted on the structures of the haloarchaeal α-amylases minus the percentage of hydrophobic interactions predicted on the structures of non-halophilic homologues

[e] Total ΔSB and ΔHI

[f] Average ΔSB and ΔHI

[g] Boldfaced numbers indicate significant *p* values


(a)
ΔAAC (%)
6
4
2
0
-2
-4
-6
Ala Cys Asp Glu Phe Gly His Ile Lys Leu Met Asn Pro Gln Arg Ser Thr Val Trp Tyr
(b)
ΔRHVP
9
7
5
3
1
-1
-3
-5
-7
-9
Arg Lys His Asp Glu Asn Gln Ser Thr
(c)
ΔRHVaP
4
3
2
1
0
-1
-2
-3
-4
Val Ala Ile Leu Pro Cys Met Phe Trp Tyr Gly
Non-halophilic archaea
Bacteria
Fungi
Metazoan group
Plants

◀**Fig. 7** Amino acid signature of α-amylases. **a** Column diagram reporting the ΔAAC of α-amylases, calculated by MEGA v5.05. ΔAAC report the differences between the amino acid compositions calculated within haloarchaeal α-amylases and non-halophilic homologues (see methods for calculation). Table under the column diagram report the *p* values of the AAC. *Underlined* digits report significant *p* value <0.05. Boldfaced digits report very significant *p* value <0.005. *Gray square* digits report very high significant *p* value <0.0005. **b**, **c** Column diagrams reporting the ΔRHV of polar and apolar residues of α-amylases, respectively. ΔRHV report the differences between the relative hydrophobic values of polar or apolar residues calculated within haloarchaeal α-amylases and non-halophilic homologues (see "Methods" for calculation). The relative hydrophobic values were calculated according to the normalized consensus hydrophobic scale of Eisenberg [49]. Labels of the amino acids are reported on the horizontal axes. AAC: amino acid composition. RHV: relative hydrophobic values. RHVP: relative hydrophobic values of polar residues. RHVaP: relative hydrophobic values of apolar residues

Amino acid composition of haloarchaeal α-amylases: A signature of haloadaptation

Genome sequence analyses of halophilic proteins suggest that they have a unique amino acid composition [45, 46]. To investigate the later characteristic, AAC of the haloarchaeal α-amylase was calculated and compared to a set of 806 non-halophilic homologues. The results report a very significant increase in Asp and Glu residues with an average ΔAAC 3.63 % and 3.23 %, respectively (Fig. 7a). Whereas, non-halophilic homologues present an overall decrease in the average number of acidic residues for all phyla included in this study, except from non-halophilic archaeal α-amylases, where it is clearly observed that there is an increased usage of Glu (8.08 %). It should be noted, that Asp residue is the most dominant negatively charged residue in haloarchaeal α-amylases compared to his homologues Glu. As pointed by [47], flexible but ordered proteins are characterized by higher average hydrophilicity and higher occurrence of negatively charged residues, especially Asp.

With highly hydrophobic side chain, Lys residue was significantly less present in haloarchaeal α-amylases then non-halophilic homologues (Fig. 7a). It was previously pointed that low occurrence of Lys combined with an increase in Arg content is another feature of amino acid composition of halophilic proteins [35]. In succeeding in reducing the interaction surface between the protein and the solvent, which is beneficial in an environment where water availability is limited (hydration of ion salts at high concentration) [48], amino acids with short side chains are more prevalent in halophilic proteins when compared to amino acids such as Lys [20].

According to the Eisenberg normalized consensus hydrophobicity scale [49], except the low significant frequency of Lys residues, hydrophilic residues are dominant in the sequences of haloarchaeal α-amylase in comparison to non-halophilic homologues from other phyla (Fig. 7b). Moreover, there is a significant decreased percentage in large hydrophobic residues such as Ile and Phe with slight increase in small hydrophobic residues such as Val and Ala, in the sequences of haloarchaeal α-amylases compared to non-halophilic counterpart (Fig. 7c). Increased usage of Val in haloarchaeal α-amylases may make them more flexible, because the strong hydrophobic residue Val has a lower helix formation propensity than other strong hydrophobic residues such as Leu, Ile and Met, or positively charged such as Lys [41]. From a structural view point, it has been demonstrated that halophilic proteins present a significant decrease of apolar surface and an increase in polar surface area [20]. These observations are correlated to the increased frequency of Asp and Glu combined to decreased frequency of Lys on the proteins surface.

Conclusion

In order to investigate the haloadaptation mechanism, the present study constitutes the first report focused on structural features of haloarchaeal α-amylases, as well as comparison of the amino acid composition between a large set of homologues protein. Compared to non-halophilic ones, haloarchaeal α-amylases present a different 3D structure organization, to maintain their stability and activity in high salt concentrations. It here appeared that their structure is characterized by (1) an increase in strategically positioned coil regions, (2) an under-representation of α-helix and β-strand-forming regions, (3) highly acidic and negatively charged surface, and (4) the presence of intraprotein interactions such as salt bridges and significantly lower proportion of hydrophobic interactions. In a salty environment, the hydrophobic residues of newly synthesized proteins are exposed to high salt concentrations, leading to non-specific inter- or intramolecular interactions of their side chains, which may compete with proper intramolecular burial within the correct conformation. Those features might contribute to avoid aggregation by enabling flexible but stable conformation changes at high salt concentrations. While performing identical enzymatic functions as their non-halophilic counterparts, the use of haloarchaeal α-amylases could be advantageous as biocatalysts in application requiring low water activity. Eventually, the results of this study could lead to the development of halostable biomolecules.

References

1. Madigan MT, Marrs BL (1997) Extremophiles. Sci Am 276(4):82–87

2. Paul S, Bag SK, Das S, Harvill ET, Dutta C (2008) Molecular signature of hypersaline adaptation: insights from genome and proteome composition of halophilic prokaryotes. Genome Biol 9 (4):R70.
3. Eisenberg H (1995) Life in unusual environments: progress in understanding the structure and function of enzymes from extreme halophilic bacteria. Arch Biochem Biophys 318(1):1–5
4. Pieper U, Kapadia G, Mevarech M, Herzberg O (1998) Structural features of halophilicity derived from the crystal structure of dihydrofolate reductase from the Dead Sea halophilic archaeon, *Haloferax volcanii*. Structure 6(1):75–88
5. Jaenicke R, Bohm G (1998) The stability of proteins in extreme environments. Curr Opin Struct Biol 8(6):738–748
6. Joo WA, Kim CW (2005) Proteomics of halophilic archaea. J Chromatogr B Anal Technol Biomed Life Sci 815(1–2):237–250
7. Rasiah IA, Rehm BH (2009) One-step production of immobilized alpha-amylase in recombinant *Escherichia coli*. Appl Environ Microbiol 75(7):2012–2016
8. Ghollasi M, Khajeh K, Naderi-Manesh H, Ghasemi A (2010) Engineering of a *Bacillus* alpha-amylase with improved thermostability and calcium independency. Appl Biochem Biotechnol 162(2):444–459
9. Hutcheon GW, Vasisht N, Bolhuis A (2005) Characterisation of a highly stable alpha-amylase from the halophilic archaeon *Haloarcula hispanica*. Extremophiles 9(6):487–495
10. Prakash O, Jaiswal N (2010) Alpha-amylase: an ideal representative of thermostable enzymes. Appl Biochem Biotechnol 160(8):2401–2414
11. Qian M, Haser R, Buisson G, Duee E, Payan F (1994) The active center of a mammalian alpha-amylase. Structure of the complex of a pancreatic alpha-amylase with a carbohydrate inhibitor refined to 2.2-A resolution. Biochemistry 33(20):6284–6294
12. van der Maarel M, van der Veen B, Uitdehaag J, Leemhuis H, Dijkhuizen L (2002) Properties and applications of starch-converting enzymes of the alpha-amylase family. J Biotechnol 94(2):137–155
13. Aghajari N, Feller G, Gerday C, Haser R (1998) Structures of the psychrophilic *Alteromonas haloplanctis* alpha-amylase give insights into cold adaptation at a molecular level. Structure 6(12): 1503–1516
14. Gilles C, Astier JP, Marchis-Mouren G, Cambillau C, Payan F (1996) Crystal structure of pig pancreatic alpha-amylase isoenzyme II, in complex with the carbohydrate inhibitor acarbose. Eur J Biochem 238(2):561–569
15. Swift HJ, Brady L, Derewenda ZS, Dodson EJ, Dodson GG, Turkenburg JP, Wilkinson AJ (1991) Structure and molecular model refinement of *Aspergillus oryzae* (TAKA) alpha-amylase: an application of the simulated-annealing method. Acta Crystallogr B 47(Pt 4):535–544
16. Kagawa M, Fujimoto Z, Momma M, Takase K, Mizuno H (2003) Crystal structure of *Bacillus subtilis* alpha-amylase in complex with acarbose. J Bacteriol 185(23):6981–6984
17. Linden A, Mayans O, Meyer-Klaucke W, Antranikian G, Wilmanns M (2003) Differential regulation of a hyperthermophilic alpha-amylase with a novel (Ca, Zn) two-metal center by zinc. J Biol Chem 278(11):9875–9884
18. Elcock AH, McCammon JA (1998) Electrostatic contributions to the stability of halophilic proteins. J Mol Biol 280(4):731–748
19. Mevarech M, Frolow F, Gloss LM (2000) Halophilic enzymes: proteins with a grain of salt. Biophys Chem 86(2–3):155–164
20. 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 11:50
21. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215(3):403–410
22. Gouet P, Robert X, Courcelle E (2003) ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Res 31(13):3320–3323
23. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234(3):779–815
24. Sivakumar N, Li N, Tang JW, Patel BK, Swaminathan K (2006) Crystal structure of AmyA lacks acidic surface and provide insights into protein stability at poly-extreme condition. FEBS Lett 580(11):2646–2652
25. Watanabe K, Hata Y, Kizaki H, Katsube Y, Suzuki Y (1997) The refined crystal structure of *Bacillus cereus* oligo-1,6-glucosidase at 2.0 A resolution: structural characterization of proline-substitution sites for protein thermostabilization. J Mol Biol 269(1):142–153
26. Bekker H, Berndsen HJC, Dijkstra EJ, Achterop S, Van Drunen R, Van Der Spoel D, Sijbers A, Keegstra H, Reitsma B, Renardus MKR (1993) Gromacs: a parallel computer for molecular dynamics simulations. Phys Comput 92:252–256
27. Melo F, Devos D, Depiereux E, Feytmans E (1997) ANOLEA: a www server to assess protein structures. Proc Int Conf Intell Syst Mol Biol 5:187–190
28. Khemili S, Kwasigroch JM, Hamadouche T, Gilis D (2012) Modelling and bioinformatics analysis of the dimeric structure of house dust mite allergens from families 5 and 21: Der f 5 could dimerize as Der p 5. J Biomol Struct Dyn 29(4):663–675
29. Dietmann S, Holm L (2001) Identification of homology in protein structure classification. Nat Struct Biol 8(11):953–957
30. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22(12):2577–2637
31. Delano WL (2002) The pymol molecular graphics system. DeLano Scientific
32. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci USA 98(18):10037–10041
33. Tina KG, Bhadra R, Srinivasan N (2007) PIC: protein interactions calculator. Nucleic Acids Res 35(Web Server issue):W473–W476
34. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24(8):1596–1599
35. Kastritis PL, Papandreou NC, Hamodrakas SJ (2007) Haloadaptation: insights from comparative modeling studies of halophilic archaeal DHFRs. Int J Biol Macromol 41(4):447–453
36. Chagnot C, Zorgani MA, Astruc T, Desvaux M (2013) Proteinaceous determinants of surface colonization in bacteria: bacterial adhesion and biofilm formation from a protein secretion perspective. Front Microbiol 4:303
37. Desvaux M, Hebraud M, Talon R, Henderson IR (2009) Secretion and subcellular localizations of bacterial proteins: a semantic awareness issue. Trends Microbiol 17(4):139–145
38. Desvaux M, Parham NJ, Scott-Tucker A, Henderson IR (2004) The general secretory pathway: a general misnomer? Trends Microbiol 12(7):306–309
39. Sawaya MR, Kraut J (1997) Loop and subdomain movements in the mechanism of *Escherichia coli* dihydrofolate reductase: crystallographic evidence. Biochemistry 36(3):586–603
40. Karlin S, Brocchieri L, Bergman A, Mrazek J, Gentles AJ (2002) Amino acid runs in eukaryotic proteomes and disease associations. Proc Natl Acad Sci USA 99(1):333–338
41. Costantini S, Colonna G, Facchiano AM (2006) Amino acid propensities for secondary structures are influenced by the protein structural class. Biochem Biophys Res Commun 342(2):441–451
42. Bieger B, Essen LO, Oesterhelt D (2003) Crystal structure of halophilic dodecin: a novel, dodecameric flavin binding protein from *Halobacterium salinarum*. Structure 11(4):375–385
43. Muller-Santos M, de Souza EM, Pedrosa Fde O, Mitchell DA, Longhi S, Carriere F, Canaan S, Krieger N (2009) First evidence

for the salt-dependent folding and activity of an esterase from the halophilic archaea *Haloarcula marismortui*. Biochim Biophys Acta 1791(8):719–729
44. Ebrahimie E, Ebrahimi M, Sarvestani NR (2011) Protein attributes contribute to halo-stability, bioinformatics approach. Saline Syst 7(1):1
45. Dym O, Mevarech M, Sussman JL (1995) Structural features that stabilize halophilic malate dehydrogenase from an archaebacterium. Science 267(5202):1344–1346
46. Fukuchi S, Yoshimune K, Wakayama M, Moriguchi M, Nishikawa K (2003) Unique amino acid composition of proteins in halophilic bacteria. J Mol Biol 327(2):347–357
47. Radivojac P, Obradovic Z, Smith DK, Zhu G, Vucetic S, Brown CJ, Lawson JD, Dunker AK (2004) Protein flexibility and intrinsic disorder. Protein Sci 13(1):71–80
48. Tadeo X, Lopez-Mendez B, Trigueros T, Lain A, Castano D, Millet O (2009) Structural basis for the aminoacid composition of proteins from halophilic archea. PLoS Biol 7(12):e1000257. doi:10.1371/journal.pbio.1000257
49. Eisenberg D, Schwarz E, Komaromy M, Wall R (1984) Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J Mol Biol 179(1):125–142