FEBS 21562 FEBS Letters 445 (1999) 375–383

Crystal structure of the β-glycosidase from the hyperthermophile *Thermosphaera aggregans*: insights into its activity and thermostability

Young-In Chi[1,a], Luis A. Martinez-Cruz[a], Jarmalia Jancarik[a], Ronald V. Swanson[b], Dan E. Robertson[b], Sung-Hou Kim[a,*]

[a] *Department of Chemistry and Lawrence Berkeley National Laboratory, University of California, Berkeley, CA 94720, USA*
[b] *Diversa Corp., San Diego, CA 92121, USA*

Received 3 November 1998; received in revised form 1 January 1999

Abstract The glycosyl hydrolases are an important group of enzymes that are responsible for cleaving a range of biologically significant carbohydrate compounds. Structural information on these enzymes has provided useful information on their molecular basis for the functional variations, while the characterization of the structural features that account for the high thermostability of proteins is of great scientific and biotechnological interest. To these ends we have determined the crystal structure of the β-glycosidase from a hyperthermophilic archeon *Thermosphaera aggregans*. The structure is a $(\beta/\alpha)_8$ barrel (TIM-barrel), as seen in other glycosyl hydrolase family 1 members, and forms a tetramer. Inspection of the active site and the surrounding area reveals two catalytic glutamate residues consistent with the retaining mechanism and the surrounding polar and aromatic residues consistent with a monosaccharide binding site. Comparison of this structure with its mesophilic counterparts implicates a variety of structural features that could contribute to the thermostability. These include an increased number of surface ion pairs, an increased number of internal water molecules and a decreased surface area upon forming an oligomeric quaternary structure.

Key words: β-Glycosidase; Crystal structure; Glycosyl hydrolase; Hyperthermophile; Thermostability

1. Introduction

Glycosyl hydrolases are a widespread group of enzymes that hydrolyze the various stereochemical conformations of the carbohydrate glycosidic bond. These enzymes are found in almost all organisms and have diverse biological functions, ranging from the essential processing of glucosides in the mammalian brain to the cleavage of cellobiose, a biomass degradation product. Glycosyl hydrolases and related enzymes have been classified into more than 60 families on the basis of amino acid sequence homology and structural similarities rather than substrate selectivities [1]. The β-glycosidase from the hyperthermophilic archeon *Thermosphaera aggregans* (Gly-*Thermosphaera*) [2,3], isolated from the Obsidian hot pool at Yellowstone National Park (Wyoming, MI, USA), belongs to the glycosyl hydrolase family 1 which includes β-glucosidases, 6-phospho-β-galactosidases, 6-phospho-β-glucosidases and some myrosinases. Family 1 belongs to a larger superfamily called 'clan GH-A' which contains at least 11 families and now comprises more than 250 members representing 18 different substrate specificities [4]. Crystal structures of some of the members of this superfamily revealed that these enzymes adopt a $(\beta/\alpha)_8$ barrel fold which is not shared by members of different families such as the lysozymes (families 22–25) or glucoamylases (family 15).

Enzymatic hydrolysis of the glycosidic bond takes place via general acid catalysis that requires two critical residues: a proton donor and a nucleophile/base [5]. The hydrolysis occurs via two major mechanisms giving rise to either an overall retention, or an inversion of anomeric configuration [6]. In both mechanisms, the position of the proton donor is identical, within hydrogen bonding distance of the glycosidic oxygen. However, in retaining enzymes, the nucleophilic catalytic base is in close proximity to the sugar anomeric carbon, whereas in inverting enzymes, this base is more distant in order to accommodate a nucleophilic water molecule between the base and the sugar [7].

Gly-*Thermosphaera* has a broad β-glycosidase activity. Analysis of the substrate specificity of the Gly-*Thermosphaera* with a variety of methylumbelliferone linked sugar substrates revealed the following preference: β-D-fucose > β-D-glucose > β-D-galactose (data not shown). This is the same substrate preference displayed by a homologous glycosidase from *Sulfolobus sulfataricus* (Gly-*Sulfolobus*) [8].

One of the potential industrial applications for thermostable glycosidases is the degradation of cellulose to glucose. This reaction requires a number of enzymatic activities including endo and exoglycosidases. Because heat is an important component of the reaction to help to breakdown higher order cellulose structures, the resistance to heat denaturation and consequently high efficiency and low cost during high temperature processes would be very beneficial to biotechnological applications such as pulp and paper production. The second important industrial application for these enzymes is the synthesis of various compounds via transglycosidation or 'reverse hydrolysis'. Synthesis of novel polysaccharides and glycoconjugates may be accomplished more effectively in high concentrations of organic solvents or in solvent aqueous systems rather than in pure aqueous systems. Thermostable proteins are usually more resistant to organic solvents than are corresponding mesophile derived proteins. Gly-*Thermosphaera* is extremely thermostable and retains 95% of its activity after incubation at 80°C for 130 h (unpublished observation). *Thermosphaera aggregans* can grow between 65 and 90°C and its optimum growth temperature is 85°C [3].

*Corresponding author. Fax: (1) (510) 486-5272.
E-mail: shkim@lbl.gov

[1] Current address: Joslin Diabetes Center, Harvard Medical School, Boston, MA 02215

PII: S0014-5793(99)00090-3

Many observations and theories on the structural basis of the thermostability of proteins have accumulated over the past years [9] and have guided our analysis of Gly-*Thermosphaera*. The comparison of this structure with its mesophilic counterparts provides considerable insight into the underlying principles of protein architecture which is of great interest for designing a protein with a proper folding and stabilization and for improving the efficiency of many industrial processes.

2. Materials and methods

2.1. Protein preparation and crystallization

The Gly-*Thermosphaera* gene was originally cloned using the methods previously described [10]. After DNA sequencing (Genbank accession number AF053078), PCR was used to amplify the active gene. The resulting fragment was subcloned into an *Escherichia coli* expression vector using standard techniques. Following expression of the protein in *E. coli*, the cells were harvested and lysed using a microfluidizer. After centrifugation the lysate was heat-treated (80°C, 10 min) to denature most of the native *E. coli* proteins. The recombinant Gly-*Thermosphaera* product which remained in solution after centrifugation was lyophilized.

The lyophilized protein samples were further purified on the Bio-Cad system using the Hi-trap Q column (Pharmacia) pre-equilibrated with buffer A (20 mM HEPES, 1 mM EDTA, pH 7.4). The column was thoroughly washed with buffer A after the protein samples had been loaded. Then the protein was eluted at 0.35 M NaCl over a 75 ml linear gradient of 0.0–1.0 M NaCl in buffer A. Fractions (1 ml) were collected and those containing the protein of interest were pooled and dialyzed against buffer A overnight, and concentrated by ultracentrifugation (Amicon YM3).

The crystallization was carried out by using the hanging drop vapor diffusion method with 4 μl drops (containing equal volumes of protein and reservoir solution) equilibrated against 500 μl of reservoir solution at room temperature. The bipyramidal-shaped crystals were obtained with 3 M sodium formate in 20 mM HEPES, pH 7.5. These crystals appeared in 3 days and continued to grow for about a week.

2.2. Data collection

The native data set from a single crystal were collected at 100 K on an R-AXISIIC imaging plate detector coupled with a Rigaku Rotaflex X-ray generator running at 50 kV and 100 mA. The crystals were transferred stepwise to mother liquor containing 30% (v/v) glycerol, acting as a cryo protectant, before being flash frozen by liquid nitrogen stream at 100 K for data collection. The data set was processed with DENZO [11] and scaled with Scalepack [11]. The space group is $P2_12_12$ with two molecules (1/2 tetramer) in the asymmetric unit. The unit cell parameters are $a = 117.51$ Å, $b = 102.21$ Å and $c = 95.78$ Å, which indicates the solvent content of 52.2% based on a protein density of 1.35 g/cc. In all, 256 310 observations of 45 287 unique structure amplitudes to 2.4 Å resolution were obtained and the data statistics are shown in Table 1.

2.3. Structure determination

The structure was determined by the molecular replacement method using the AMORE software package [12]. The coordinates of β-glycosidase from *Solfolobus solfataricus* (Gly-*Sulfolobus*) were obtained from the Protein Data Bank (access code 1GOW) and used as the search model. Solutions were easily obtained by using the entire model without any side chain substitutions or deletions/insertions. A search carried out with the data between 15.0 and 4.0 Å and center of mass cut off at 20 Å, produced a solution with a correlation coefficient of 0.406 and an R factor of 42.7%. The following rigid body refinement, using 10.0 and 2.4 Å data with X-PLOR [13], resulted in the R_{free} and R values of 46.1% and 45.9%, respectively. The proper side chains for the mutated residues were built into the σ_A-weighted 2Fo-Fc map and one round of positional refinement was carried out. The subsequent Fo-Fc maps clearly showed electron densities for the inserted and deleted regions between two species.

2.4. Model refinement

The refinement of the corrected model was then continued with the group B factor refinement followed by the simulated annealing protocols implemented in X-PLOR 3.1. Restraints were placed on bond lengths, bond angles, non-bonded contacts and temperature factors of neighboring atoms. The non-crystallographic symmetry restriction was imposed and the bulk solvent correction was also applied. The σ_A-weighted 2Fo-Fc maps as well as omit maps were calculated at regular intervals to allow manual rebuilding of insertion/deletion loops and side chains with a different rotamer. Model building was done using O [14] and the backbone geometry was regularly checked against a structural database with the pep flip option. Solvent molecules (all regarded as water) based on higher than 3σ peaks in the Fo-Fc σ_A-weighted maps were added gradually and conservatively with regard for their environment including potential interactions with hydrogen bond partners. This solvent model was further comprehensively checked several times during the refinement by omitting all water molecules that had high B values (> 60 Å^2) or made either too close contacts with each other or with protein atoms or a too sharp angle with potential hydrogen bonding partners. Inclusion of individual atomic temperature factors and the removal of NCS restriction during the final stages was validated by a substantial decrease in the value of R_{free}. The program Procheck [15] was used to check the quality of the structure. At the end of the refinement, the crystallographic R factor was 20.1% with the R_{free} value of 24.9%. The detailed refinement statistics are shown in Table 1.

Table 1
Data processing and refinement statistics

Data statistics					
Resolution	Reflections	Redundancy	Data coverage (%)	I/σ	R_{sym}(%)
2.4 Å	45 287	5.6 (4.0)	98.9 (95.3)	12.4 (4.5)	6.1 (19.8)

Refinement Statistics	
Resolution range	20.0–2.4 Å
Number of non-H protein atoms	7858
Number of solvent molecules	439
R factor for 42922 reflections ($F > 2\sigma$)	21.0%
R_{free} value for 2268 reflections ($F > 2\sigma$, 5%)	24.9%
rms deviation from ideal geometry	
Bond lengths	0.007 Å
Bond Angles	1.473°
Dihedral angles	22.193°
Improper angles	0.738°
Average temperature factors	
Main chain atoms only	28.828 Å^2
All protein atoms	30.355 Å^2
Waters	5.288 Å^2

Values in parentheses correspond to the highest resolution shell (2.49–2.4).

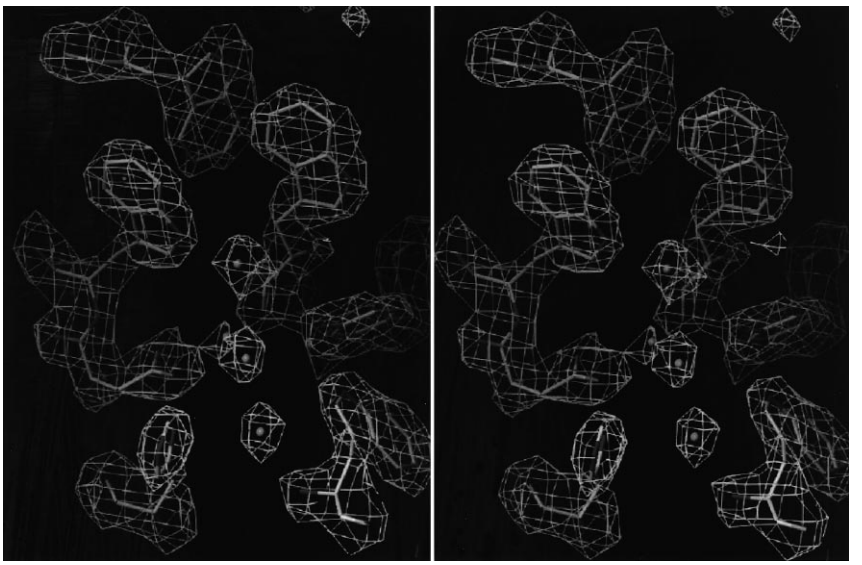


Fig. 1. Representative portion of the σ_A-weighted 2Fo-Fc map with the final model showing a group of aromatic residues along with water molecules around the active site. The map is contoured at 1.5σ value.

3. Results and discussion

3.1. Description of the structure

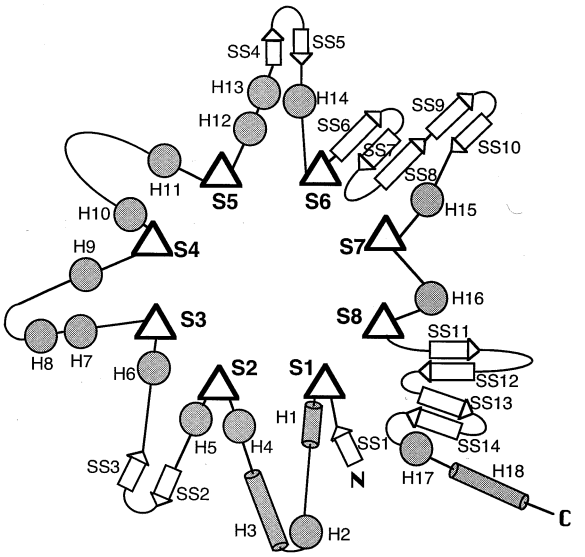
The current refined model at 2.4 Å resolution (Table 1) contains all protein atoms and 439 water molecules. A representative portion of the σ_A weighted 2Fo-Fc map with the final model is shown in Fig. 1. The overall fold is very similar to that of Gly-*Sulfolobus* [8], which shares 53% sequence identity and was used as a search model for molecular replacement. When the two structures are superimposed using only Cα atoms, the equivalent Cα positions (449 residues out of 481 residues) are in very good agreement (rms value of 0.84 Å, Fig. 2). The major differences only occur at three long loops containing short β-strands: between H5 and H6, between H13 and H14 and between S6 and H15 (Figs. 2 and 3), where insertions and/or deletions have occurred. The schematic representation of the secondary structure of a monomer is shown in Fig. 3. This classic $(\beta/\alpha)_8$ barrel (TIM-barrel) fold has been observed in all known structures of family 1 glycosyl hydrolases. The comparison with structures of other family 1 members is summarized in Table 2.

The protein subunits are organized to form a tetramer (Fig. 4), with a 222 point symmetry, two dyads of which are crystallographic, with dimensions of roughly $120\times100\times70$ Å along the respective two-fold axes. The constituent monomers each have identical subunit interactions. This tetrameric arrangement is very similar to the one observed for the Gly-*Sulfolobus* structure [8].

3.2. Catalysis and the putative active center

Catalysis by β-glycosyl hydrolases involves two essential carboxylates, one acting as the proton donor and the other as the nucleophile. Two conserved glutamates near the carboxy-terminal end of the $(\beta/\alpha)_8$ barrel have been implicated to

Fig. 2. Wire stereo representation of the two superimposed hyperthermophilic β-glycosidases. The main chains of the Gly-*Thermosphaera* and the Gly-*Sulfolobus* structures are shown as a thick wire and a thin wire, respectively. The overall folding of the two structures are very similar except the three loops shown on the top portion of the molecules in this figure.



SS1: 1-3
S1: 7-14
H1: 15-19
H2: 31-38
H3: 39-45
H4: 62-72
S2: 75-81
H5: 82-86
SS2: 96-100
SS3: 107-111
H6: 124-142
S3: 143-151
H7: 155-158
H8: 160-167
H9: 178-197
S4: 201-208
H10: 208-218
H11: 230-255
S5: 260-272
H12: 277-289
H14: 290-296
SS4: 297-298
SS5: 304-306
H15: 306-310
S6: 315-321
SS6:322-329
SS7: 332-335
SS8: 343-349
SS9: 351-356
SS10: 358-361
H15: 364-378
S7: 381-388
H16: 396-414
S8: 417-424
SS11: 427-429
SS12: 438-441
SS13: 443-447
SS14: 450-454
H17: 455-468
H18: 471-478

Fig. 3. Schematic drawing of the secondary structure arrangement in the Gly-*Thermosphaera* structure. The core β-strands forming the $(\beta/\alpha)_8$ barrel, approximately perpendicular to the plane of the paper, are depicted as triangles. The β-strands surrounding the barrel, almost along the plane of the paper are depicted as arrows. The α-helices approximately perpendicular to the plane of the paper and almost along the plane of the paper are depicted as shadowed circles and shadowed rods, respectively. The first turn of the H9 and H15, and the last turn of the H18 form 3_{10} helices. All secondary structure elements are denoted by letters followed by numbers and the first and the last residue numbers of each secondary structure element are indicated on the right panel of the figure.

serve these roles from previously determined glycosyl hydrolase family 1 crystal structures [10]. Indeed, in the Gly-*Thermosphaera* structure as well as in the Gly-*Sulfolobus* structure, the highly conserved Glu-208 and Glu-386 are found at the proposed site in β-strands four and seven, respectively (Figs. 3 and 5) [16]. Glu-208 is preceded by Asn-207 which is also typical for members of the '4/7 superfamily' enzymes [16]. The exact role of this invariant asparagine residue is not clear.

Glu-386 has been identified in family 1 as the catalytic nucleophile by covalent modification with a mechanism-based inhibitor of the equivalent residue in the *Agrobaterium* enzyme [17] and later confirmed by mutational studies [18]. In the Gly-*Thermosphaera* structure, Glu-208, which acts as a proton donor, is 5.05 Å away from Glu-386 (between C_δ atoms from both residues, the closest distance is 3.88 Å between $O_{\varepsilon 2}$ atom of Glu-208 and $O_{\varepsilon 1}$ atom of Glu-386). This is consistent with the retaining mechanism [19]. In the inverting mechanism, the distance would be larger (around 10 Å) to accommodate a water molecule between the nucleophilic base and the substrate [4]. The covalent intermediate in the retaining mechanism has recently been trapped using a mechanism-based inhibitor, 2-fluoro-glycoside, and crystallographically observed in a related β-glycosidase from *Cellulomonas fimi* [20] and myrosinase from *Sinapis alba* [21]. The active site architecture of the covalent intermediate clearly showed the substrate covalently bound to the catalytic nucleophile and the proton donor near enough to protonate the bound glycoside substrate and deprotonate an incoming water molecule during hydrolysis.

The surrounding residues, which would stabilize the substrate, determine the substrate specificity. The accumulated findings demonstrate the remarkable versatility of the TIM-barrel fold, which evolution has finely tuned to generate many different substrate specificities around the same catalytic reaction with a similar disposition of identical catalytic residues on the same ancestral structure [19]. The active center in this type of glycosyl hydrolase is typically surrounded by aromatic groups which, along with polar residues, constitute the glycoside binding site [8,22,23]. In the Gly-*Thermosphaera* structure, these include Tyr-321, Trp-424, Phe-440, Phe-341, Trp-360 and Phe-358 on one face and Phe-219, Phe-224, Trp-35, Trp-32, Trp-432 and Trp-152 on the other face (Fig. 5). In

Table 2
Structural comparison of the glycosyl hydrolase family 1 members

	Gly-*Sulfolobus*	6-P-β-galactosidase[a]	β-glucosidase[b]	Myrosinase[c]	β-glucosidase A[d]
Gly-*Thermosphaera*	0.84Å (449)	1.56 Å (342)	1.36 Å (357)	1.41 Å (356)	1.49 Å (363)
Gly-*Sulfolobus*	–	1.46 Å (371)	2.05 Å (364)	1.16 Å (308)	1.56 Å (368)
6-P-β-galactosidase[a]	–	–	1.79 Å (358)	1.24 Å (340)	1.77 Å (416)
β-glucosidase[b]	–	–	–	1.40 Å (438)	2.04 Å (424)
Myrosinase[c]	–	–	–	–	2.02 Å (398)

The root mean square values between two structures were calculated by a least square fit after superposing the structurally equivalent C_α atoms of the two molecules in comparison (the number of equivalent C_α atoms are indicated in parenthesis). The hyperthermophilic proteins are indicated as bold letters.

[a]6-phospho-β-galactosidase structure from a mesophilic bacterium *Lactococcus lactis*.
[b]β-glucosidase structure from white clover *Trifolium repens*.
[c]Myrosinase structure from white mustard seed *Sinapis alba*.
[d]β-glucosidase A structure from *Bacillus polymyxa*.

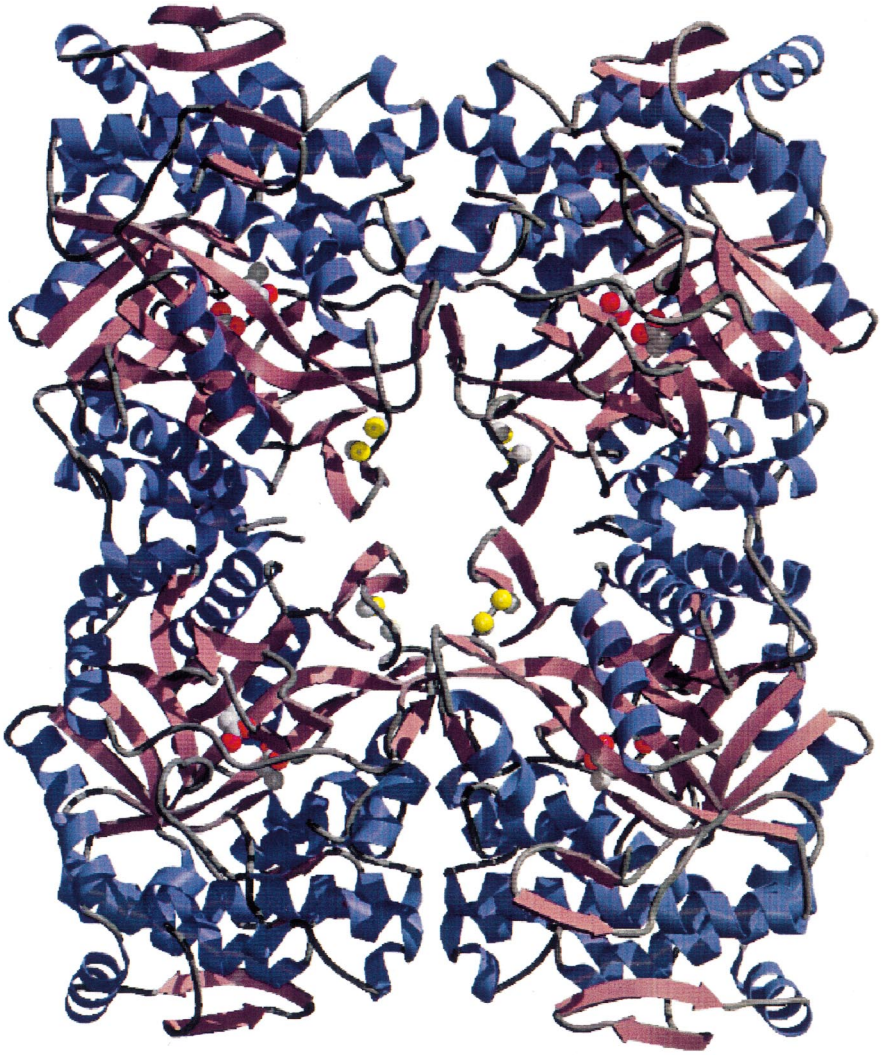


Fig. 4. Raster 3D [43] rendered a ribbon diagram showing the tetrameric arrangement of the subunits. The catalytic glutamates and the cystein residues forming a disulfide bond are indicated as ball and stick representations. The oxygen atom and sulfur atoms are represented in red and yellow, respectively. Each monomer has identical subunit interactions.

addition, an extensive hydrogen bond network (Fig. 6) stabilizes the ionization states of key residues around the active site and may play a role in modulating the ionization state of the catalytic residues and in controlling the local charge balance during the reaction [20]. This pocket-shaped active site topology with essential catalytic residues and supporting neighboring residues (Fig. 5) is optimal for the recognition of a monosaccharide as a substrate and has been observed in β-galactosidase, β-glucosidase, sialidase, neuraminidase and β-amylase structures [1]. In contrast, for the recognition of a polysaccharide, cleft-shaped or tunnel-shaped active site topologies are typically observed as seen in endoglucanase from *Thermonospora fusca* [24] and cellobiohydrolase II from *Trichoderma reesei* [25].

3.3. Thermostability

Many different crystal structures of proteins from thermophilic organisms have provided valuable information on the molecular basis of thermostability. The comparison of these structures with their mesophilic counterparts suggests that thermostability is achieved by subtle structural differences residing over the entire molecule without large changes in the overall polypeptide chain folding. These factors include increases in the number of hydrogen bonds, the number of surface ion pairs, subunit interactions and hydrophobic residue substitutions as well as a reduction of the surface area by shortening the connecting loops and/or a decrease in the number of cavities [9].

Among the members of the glycosyl hydrolase family 1, five more structures, other than the Gly-Thermosphaera structure, have been previously determined by crystallographic methods. They are the structures of Gly-*Sulfolobus* [8], 6-phospho-β-galactosidase from a mesophilic bacterium *Lactococcus lactis* [22], β-glucosidase from white clover *Trifolium repens* [23], myrosinase from white mustard seed *Sinapis alba* [21] and β-glucosidase A from *Bacillus polymyxa* [26]. When these structures are compared, there is an increase in the number of ion pairs in the hyperthermophilic enzymes (Table 3). For example, of the 24 arginine residues in Gly-*Thermosphaera*, only five do not participate in salt bridges with acidic residues. For the remaining charged residues, more than half of them are involved in either surface ion pairs or internal ion pairs.

Fig. 5. Ball and stick representation of the aromatic residues surrounding the active site at the C-terminal end of the TIM-barrel-folded molecule shown as a wire representation. The catalytic residues are also shown as a ball and stick model and labelled.

The number of ion pairs found in the Gly-*Thermosphaera* structure is almost twice as high as the number found in the myrosinase despite the shorter polypeptide chain length (Table 3). The list of ion pairs in the Gly-*Thermosphaera* monomer is shown in Table 4. These ion pairs exert cross-linking effects within a folded protein structure to prevent it from unfolding at high temperature. This dramatic increase in the number of ion pairs has been observed previously in many different proteins from hyperthermophiles [27–29]. One of the most remarkable ion pair contributions to protein stability in these examples is seen in the structure of glutamate dehydrogenase from *Pyrococcus furiosus* where a striking series of ion pair networks are found on the surface of the protein subunits and at both interdomain and intrasubunit interfaces [29]. Another striking feature, potentially linked to the thermostability of Gly-*Thermosphaera*, is the increase in the number of internal waters. Water molecules are rarely found within hydrophobic cavities except when they have specific functions such as proton relay [30] or substrate binding [31]. Even though buried waters are typically defined as ones that cannot be connected by a continuous series of water-water hydrogen bonds to bulk water molecules [32], we included those waters residing in the cavities, connected by a chain of waters to the surface waters (thus called 'cleft waters' [32]). The total number of water molecules reported in a presenting crystal structure depends on certain factors such as resolution and crystallization condition. Therefore, we used the ratio of internal waters to the total number of water molecules in the PDB data bank files for the comparison, instead of the total absolute number of internal waters. A total of 96 internal water molecules were found for the Gly-*Thermosphaera* dimer in the asymmetric unit, which accounts for 22% of the total number of water molecules (Table 3, also see Section 2 for the criteria used for inclusion of water molecules). These internal water molecules are found mostly in the form of chains, rather than as isolated molecules, to facilitate the formation of hydrogen bond networks. These buried water molecules may stabilize a protein structure by providing the otherwise missing van der Waals interactions for those atoms bordering a cavity and hydrogen bonding to otherwise unsatisfied protein hydrogen bonding groups [32]. This type of increase in the internal hydrogen bonding network associated with water molecules

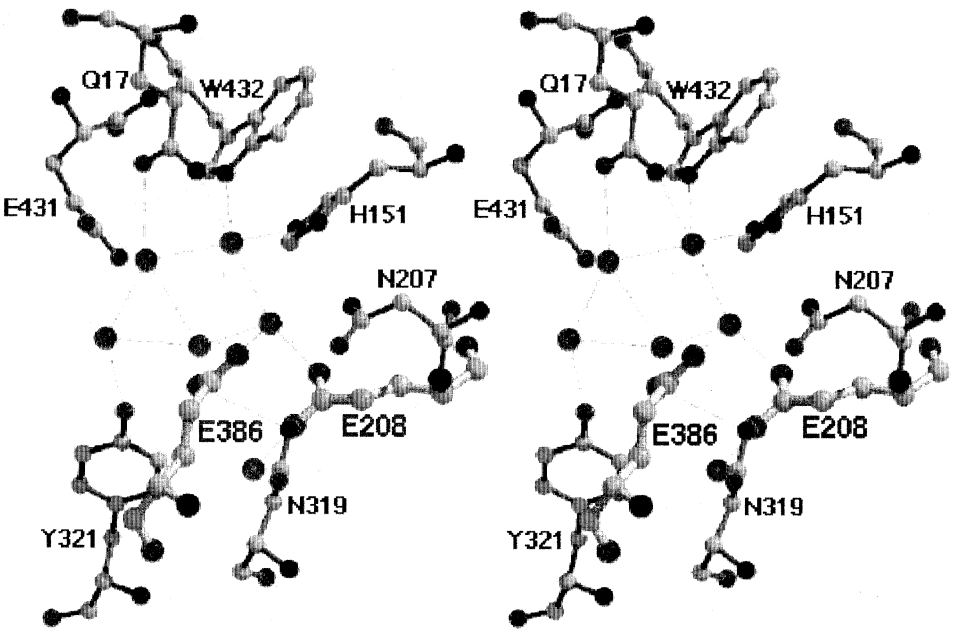


Fig. 6. Ball and stick stereo representation of the hydrogen bond network associated with water molecules at the core of the active site. The water molecules are shown in pink.

Table 3
Comparison of some of the features between hyperthermophilic and mesophilic glycosyl hydrolase family 1 members that are generally known to affect the thermostability

	Number of residues	Number of ion pairs	Internal water ratio*	Functional unit	Number of proline	Number of disulfide
Gly-*Thermosphaera*	482	**40**	**22%** (96/439)	**tetramer**	30	1
Gly-*Sulfolobus*	489	**37**	**25%** (72/286)	**tetramer**	26	0
6-P-β-galactosidase[a]	468	32	16% (67/417)	dimer	20	0
β-glucosidase[b]	490	22	12% (53/436)	dimer	26	1
Myrosinase[c]	501	25	9% (62/672)	dimer	26	3
β-glucosidase A[d]	448	28	10% (160/1541)	dimer**	18	0

The same proteins as shown in Table 2 were used in these comparisons and the symbol denotations for each protein are identical to those in Table 2. The hyperthermophilic proteins and the features affecting their thermostability are also indicated as bold letters.
*internal water ratio = (number of internal waters)/(total number of waters) × 100.
**The β-glucosidase A structure from *Bacillus polymyxa* was reported as an octamer which is in fact made of a crystallographic tetramer of the dimers containing similar interacting surfaces to other related family 1 glycosidase dimers.

has also been observed in other thermophilic protein structures [33,34].

Reduction of the surface area is another factor implicated in thermostability seen in many thermostable proteins [35,36]. In the Gly-*Thermosphaera* structure, shorter connecting loops and a decrease in the number of cavities do not seem to be contributing factors for thermostability because some loops (e.g. the loop between S2 and S3 strands) are even longer than their mesophilic counterparts. However, it is noteworthy to point out that the active form of this enzyme is a tetramer, whereas the mesophilic counterparts are active as dimers (Table 3). The surface area buried at the tetramer interface for each monomer is 2.231 Å^2 (using a probe radius of 1.4 Å) which accounts for a 14% reduction of the solvent-accessible surface area upon tetramer formation. This type of oligomerization through 'sticky patches' via intersubunit hydrophobic contacts as a means of achieving thermostability has also been observed in a number of enzymes [35,37,38].

The presence or introduction of disulfide bonds to a protein has been shown to enhance its thermostability [39,40]. The disulfide bond (Cys-343/Cys-355) in the Gly-*Thermosphaera* structure may contribute to its slightly higher thermostability than the Gly-*Sulfolobus* structure which has none (data not shown). This feature was also seen in the myrosinase from white mustard seed *Sinapis alba* where the presence of three disulfide bonds, in the context of otherwise similar structural features to other mesophilic proteins, is thought to be responsible for its stability against denaturation [41] and its maximal activity at temperatures between 55°C and 65°C [42].

An increase in proline residues has been suggested to be a factor in the thermostability of some other enzymes by reducing the flexibility and providing more hydrophobicity [35]. However, the number of proline residues in this group of enzymes does not seem to be correlated with thermostability (Table 3).

Table 4
List of ion pairs found in the crystallographic structure of the β-glycosidase from *Thermosphaera aggregans*

Internal ion pairs	
Arg-78/Glu-386	Arg-85/Glu-19
Arg-397/Asp-394	His-151/Glu-386
His-159/Asp-31	
Intersubunit ion pairs	
Arg-171/Asp-170	Arg-462/Glu-463
Lys-71/Glu-473	
Surface ion pairs	
Arg-100/Glu-179	Arg-100/Glu-180
Arg-118/Glu-121	Arg-141/Asp-67
Arg-141/Glu-70	Arg-141/Asp-137
Arg-143/Glu-70	Arg-165/Glu-40
Arg-166/Glu-40	Arg-247/Asp-250
Arg-247/Asp-308	Arg-254/Asp-110
Arg-306/Asp-308	Arg-312/Asp-250
Arg-323/Glu-361	Arg-377/Asp-292
Arg-452/Asp-65	Arg-455/Asp-428
Lys-125/Asp-120	Lys-136/Glu-140
Lys-144/Asp-6	Lys-187/Glu-184
Lys-195/Asp-120	Lys-253/Asp-314
Lys-283/Glu-373	Lys-286/Glu-214
Lys-296/Asp-292	Lys-410/Glu-365
Lys-419/Asp-381	His-37/Glu-53
His-66/Asp-137	His-130/Glu-133

3.4. Sequence comparison with other hyperthermophilic β-glycosidases

When we compare the structures of Gly-*Thermosphaera* and Gly-*Sulfolobus*, the differences are only prominent at three solvent exposed loop regions where insertions and/or deletions have occurred (Fig. 2). Moreover, the majority of internal waters (52 out of 72 reported internal waters for the Gly-*Sulfolobus* structure) are found at the same locations in the two known structures of the hyperthermophilic β-glycosidases.

In a sequence comparison of β-glycosidases from hyperthermophilic archea whose sequences are known, 28% of the residues are strictly conserved (Fig. 7). Among these, there appears to be a greater degree of conservation of residues which participate in the formation of ion pairs, in particular those which form buried ion pairs. All nine residues forming internal ion pairs are strictly conserved and three out of six residues forming intersubunit ion pairs are conserved (Table 4Fig. 7). However, only 13 out of 53 residues (26%) forming surface ion pairs are conserved, consistent with the average sequence conservation among these sequences. From these findings, mutations which affect the buried ion pairs are likely to be more detrimental to the molecule's integrity than those mutations in surface ion pairs.

In summary, the thermostability of the Gly-*Thermosphaera* structure appears to be mostly attributed to the increased ionic interactions, the increased number of internal water molecules to enhance the internal packing by hydrogen bond formations and the oligomerization to reduce the surface area and to enhance the intermolecular packing interactions.

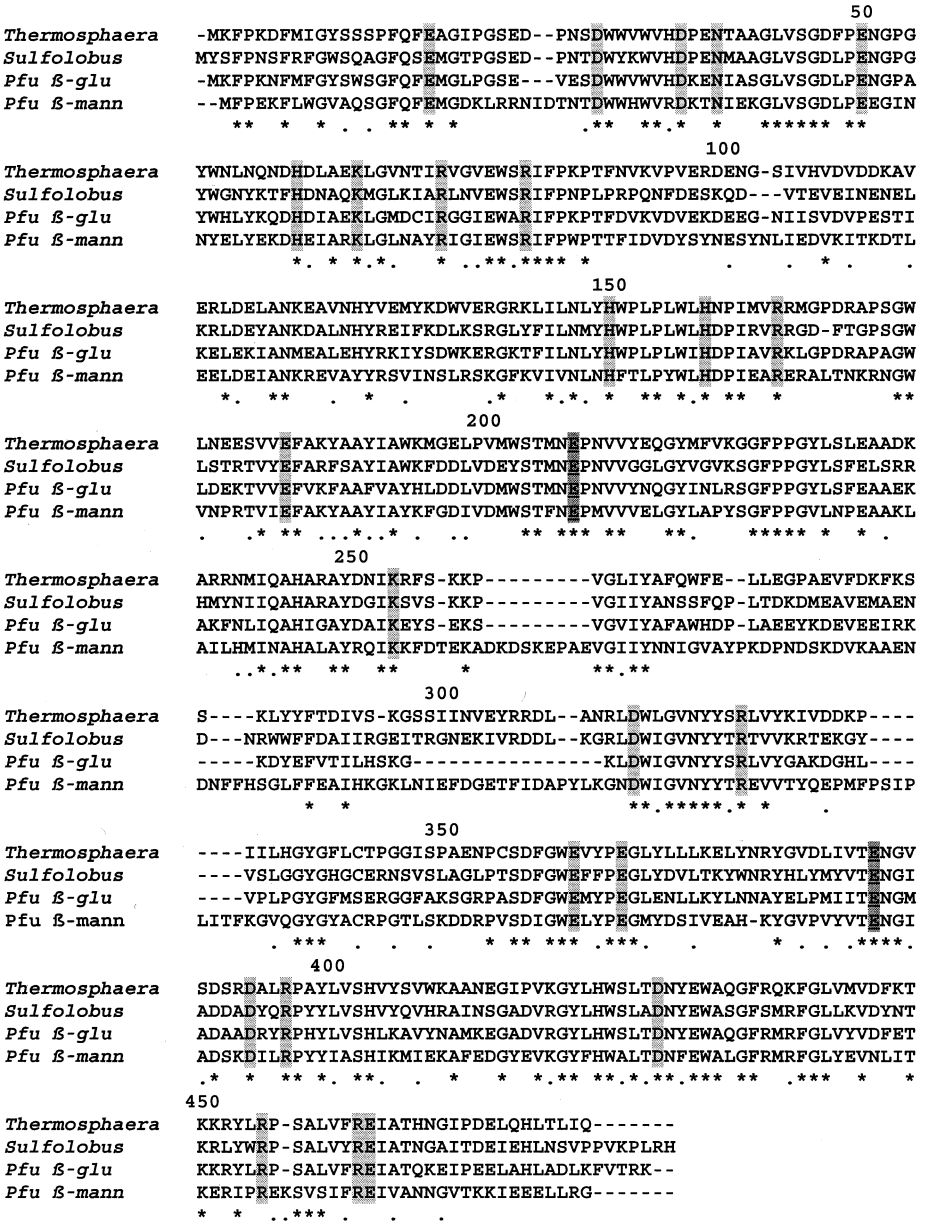


Fig. 7. Sequence alignment of four β-glycosidases or its family members from three hyperthermophilic archea. The overall sequence identity within the alignment is 28%. The red color denotes those residues involved in ion pair formations which are strictly conserved among these species. The two catalytic glutamate residues are underlined. The positions marked by an asterisk or a dot contain identical residues or similar residues, respectively. The sequence numbers are based on Gly-*Thermosphaera*. The four different sequences in the alignment are from *Thermosphaera aggregans*, *Sulfolobus solfataricus, Pyrococcus furiosus* (β-glucosidase and β-mannosidase).

Furthermore, the residues participating in the internal ion pair formation are strictly conserved among this type of enzymes in hyperthermophilic archea.

Acknowledgements: We would like to thank the many individuals and groups at Diversa Corp. who contributed to the cloning, sequencing, expression and characterization of Gly-*Thermosphaera*. In particular we would like to acknowledge the contributions of the following people: M. Snead, E. Bylina, A. Lenox, S. Peralta, D. Murphy and J. Reid. We are also grateful to Dr K.O. Stetter for generously supplying the original *Thermosphaera aggregans* biomass from which DNA libraries were constructed. Our gratitude also goes to Thomas Livingston for his assistance in the preparation of the manuscript. This work was supported by the Grant from Office of Biosciences and Environmental Research, Office of Energy Research, US Department of Energy (DE-AC03-76SF00098).

References

[1] Henrissat, B. and Bairoch, A. (1993) Biochem. J. 293, 781–788.
[2] Huber, R., Burggraf, S., Mayer, T., Barns, S.M., Rossnagel, P. and Stetter, O.K. (1995) Nature 376, 57–58.
[3] Huber, R., Dyba, D., Huber, H., Burggraf, S. and Rachel, R. (1998) Int. J. Syst. Bacteriol. 48, 31–38.
[4] Henrissat, B. and Bairoch, A. (1996) Biochem. J. 316, 695–696.
[5] Davies, G. and Henrissat, B. (1995) Structure 3, 853–859.
[6] Sinnott, M.L. (1990) Chem. Rev. 90, 1171–1202.
[7] MaCarter, J.D. and Withers, S.G. (1994) Curr. Opin. Struct. Biol. 4, 885–892.
[8] Aguilar, C.F., Sanderson, I., Moracci, M., Ciaramella, M., Nucci, R., Rossi, M. and Pearl, L.H. (1997) J. Mol. Biol. 271, 789–802.

[9] Querol, E., Perez-Pons, J.A. and Mozo-Villarias, A. (1996) Protein Eng. 9, 265–271.
[10] Bauer, M., Bylina, E.J., Swanson, R.V. and Kelly, R.M. (1996) J. Biol. Chem. 271, 23749–23755.
[11] Ozwinowsk, Z. (1993) in: Data Collection and Processing (Sawyer, L., Issacs, N. and Bailey, S., Eds.), pp. 56–62, SERC Darsberry Laboratory, Warrington, UK.
[12] Navaza, J. (1994) Acta Cryst. A50, 157–163.
[13] Brünger, A.T. (1992) X-PLOR manual, version 3.1., Yale University, New Haven, CT.
[14] Jones, T.A., Zhou, J.-Y., Cowan, S.W. and Kjeldgaard, M. (1991) Acta Cryst. A47, 110–119.
[15] Laskowski, R.A., Macarthur, M.W., Moss, D.S. and Thornton, J.M. (1993) J. Appl. Crystallogr. 26, 283–291.
[16] Jenkins, J., Leggio, L.L., Harris, G. and Pickersgill, R. (1995) FEBS Lett. 362, 281–285.
[17] Withers, S.G., Warren, R.A.J., Street, I.P., Rupitz, K., Kempton, J.B. and Aebersold, R. (1990) J. Am. Chem. Soc. 122, 5887–5889.
[18] Withers, S.G., Rupitz, K., Trimbur, D. and Warren, R.A.J. (1992) Biochemistry 31, 9979–9985.
[19] Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.P. and Davies, G. (1995) Proc. Natl. Acad. Sci. USA 92, 7090–7094.
[20] White, A., Tull, D., Johns, K., Withers, S.G. and Rose, D.R. (1996) Nat. Struct. Biol. 3, 149–154.
[21] Burmeister, W.P., Cottaz, S., Driguez, H., Iori, R., Palmieri, S. and Henrissat, B. (1997) Structure 5, 663–675.
[22] Wiesmann, C., Beste, G., Hengstenberg, W. and Schultz, G.E. (1995) Structure 3, 961–968.
[23] Barrett, T., Suresh, C.G., Tolley, S.P., Dodson, E.J. and Hughes, M.A. (1995) Structure 3, 951–960.
[24] Spezio, M., Wilson, D.B. and Karplus, P.A. (1993) Biochemistry 32, 9906–9916.
[25] Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J.K.C. and Jones, T.A. (1990) Science 249, 380–386.
[26] Sanz-Aparicio, J., Hermoso, J.A., Martinez-Ripoll, M., Lequerica, J.L. and Polaina, J. (1998) J. Mol. Biol. 275, 491–502.
[27] Hennig, M., Darimont, B., Sterner, R., Kirschner, K. and Jansonius, J.N. (1995) Structure 3, 1295–1306.
[28] Korndorfer, I., Steipe, B., Huber, R., Tomschy, A. and Jaenicke, R. (1995) J. Mol. Biol. 246, 511–521.
[29] Yip, K.S., Stillmas, T.J., Britton, K.L., Artymiuk, P.J., Baker, P.J., Sedelnikova, S.E., Engel, P.C., Pasquo, A., Chiaraluce, R. and Consalvi, V. (1995) Structure 3, 1147–1158.
[30] Martinez, S.E., Huang, D., Ponomarev, M., Cramer, W.A. and Smith, J.L. (1996) Protein Sci. 5, 1081–1092.
[31] Meyer, E., Cole, G., Radhakrishnan, R. and Epp, O. (1988) Acta Cryst. B44, 26–55.
[32] Williams, M.A., Goodfellow, J.M. and Thornton, J.M. (1994) Protein Sci. 3, 1224–1235.
[33] Tanner, J.J., Hecht, R.M. and Krause, K.L. (1996) Biochemistry 35, 2597–2609.
[34] Macedo-Ribeiro, S., Darimont, B., Sterner, R. and Huber, R. (1996) Structure 4, 1291–1301.
[35] Delboni, L.F., Mande, S.C., Rentier-Delrue, F., Mainfroid, V., Turley, S., Vellieux, F.M., Martial, J.A. and Hol, W.G. (1995) Protein Sci. 4, 2594–2604.
[36] Hwang, K.Y., Song, H.K., Chang, C., Lee, J., Lee, S.Y., Kim, K.K., Choe, S. and Suh, S.W. (1997) Mol. Cell. 7, 251–258.
[37] Harris, G.W., Pickersgill, R.W., Connerton, I., Debeire, P., Touzel, J.P., Brenton, C. and Perez, S. (1997) Proteins Struct. Funct. Genet. 29, 77–86.
[38] Ermler, U., Merckel, M., Thauer, R. and Shima, S. (1997) Structure 15, 635–646.
[39] DeDecker, B.S., O'Brien, R., Fleming, P.J., Geiger, J.H., Jackson, S.P. and Sigler, P.B. (1996) J. Mol. Biol. 264, 1072–1084.
[40] Ko, J.H., Jang, W.H., Kim, E.K., Lee, H.B., Park, K.D., Chung, J.H. and Yoo, O.J. (1996) Biochem. Biophys. Res. Commun. 221, 631–635.
[41] Pessina, A., Thomas, R.M., Palmieri, S. and Luisi, P.L. (1990) Arch. Biochem. Biophys. 280, 383–389.
[42] Hochkoeppler, A. and Palmieri, S. (1992) Biotechnol. Prog. 8, 91–98.
[43] Merrit, E.A. and Murphy, M.E.P. (1994) Acta Cryst. D50, 869–873.