

Molecular Basis of the Thermostability and Thermophilicity of Laminarinases: X-ray Structure of the Hyperthermostable Laminarinase from *Rhodothermus marinus* and Molecular Dynamics Simulations

Lucas Bleicher,[†] Erica T. Prates,[‡] Thiago C. F. Gomes,[‡] Rodrigo L. Silveira,[‡] Alessandro S. Nascimento,[†] Adriana L. Rojas,^{†,§} Alexander Golubev,^{||} Leandro Martínez,^{*†} Munir S. Skaf,^{*‡} and Igor Polikarpov^{*†}

[†]Institute of Physics of São Carlos, University of São Paulo, Avenida Trabalhador São-Carlense, 400, CEP 13560-970 São Carlos, SP, Brazil

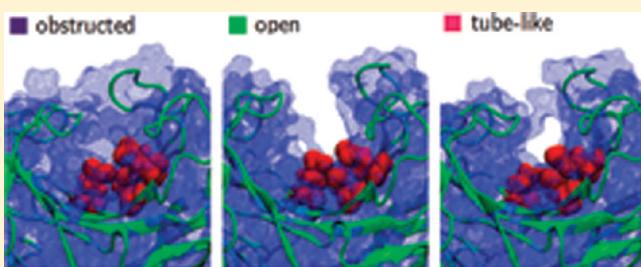
[‡]Institute of Chemistry, State University of Campinas — UNICAMP, Cx.P. 6154, Campinas, SP 13084-862, Brazil

[§]Structural Biology Unit, Center for Cooperative Research in Biosciences bioGUNE, Bizkaia Technology Park, 48160 Derio, Spain

^{||}Petersburg Nuclear Physics Institute, Gatchina, St. Petersburg, 188300, Russia

 Supporting Information

ABSTRACT: Glycosyl hydrolases are enzymes capable of breaking the glycosidic linkage of polysaccharides and have considerable industrial and biotechnological applications. Driven by the later applications, it is frequently desirable that glycosyl hydrolases display stability and activity under extreme environment conditions, such as high temperatures and extreme pHs. Here, we present X-ray structure of the hyperthermophilic laminarinase from *Rhodothermus marinus* (RmLamR) determined at 1.95 Å resolution and molecular dynamics simulation studies aimed to comprehend the molecular basis for the thermal stability of this class of enzymes. As most thermostable proteins, RmLamR contains a relatively large number of salt bridges, which are not randomly distributed on the structure. On the contrary, they form clusters interconnecting β -sheets of the catalytic domain. Not all salt bridges, however, are beneficial for the protein thermostability: the existence of charge–charge interactions permeating the hydrophobic core of the enzymes actually contributes to destabilize the structure by facilitating water penetration into hydrophobic cavities, as can be seen in the case of mesophilic enzymes. Furthermore, we demonstrate that the mobility of the side-chains is perturbed differently in each class of enzymes. The side-chains of loop residues surrounding the catalytic cleft in the mesophilic laminarinase gain mobility and obstruct the active site at high temperature. By contrast, thermophilic laminarinases preserve their active site flexibility, and the active-site cleft remains accessible for recognition of polysaccharide substrates even at high temperatures. The present results provide structural insights into the role played by salt-bridges and active site flexibility on protein thermal stability and may be relevant for other classes of proteins, particularly glycosyl hydrolases.



1. INTRODUCTION

Laminarin is an energy storage glucose polysaccharide found in brown algae. It is the nutrient rich part of algae that feeds hundreds of sea animals. Brown algae grow very fast, at a rate of up to 30 cm per day and synthesize laminarin directly by photosynthesis. As such, it is promising source of renewable biomass. Laminarin is a mostly $\beta(1 \rightarrow 3)$ linked β -D-glucan, such that glycosyl hydrolases (GHs) are able to depolymerize into fermentable sugars.¹ GHs are abundant proteins in several organisms for their essential role in the degradation of polysaccharides. As such, there are more than 100 families of GHs resulting from convergent evolution,^{2,3} which became an important biotechnological target for renewable energy exploration. However, usability

and cost-effectiveness of these enzymes in industrial applications require high stability and activity under extreme environments, particularly wide ranges of pH and elevated temperatures.

Thermophilic enzymes can be obtained from thermophilic microorganisms such as, for example, *Rhodothermus marinus*, a gram-negative thermophilic bacteria that grows between 54 and 77 °C, and optimally at 65 °C. Most of its characterized enzymes have optimum activity in temperatures ranging between 55 and 100 °C.⁴ In particular, the laminarinase from *R. marinus*

Received: January 12, 2011

Revised: April 14, 2011

Published: May 27, 2011

Table 1. Crystallographic Parameters

Data Processing Parameters ^a	
wavelength (Å)	1.42
resolution limits (Å)	30.2–1.95 (2.06–1.95)
space group	$P2_1$
lattice parameters (Å, degrees)	$a = 52.22, b = 108.29,$ $c = 64.59, \beta = 113.90$
completeness (%)	97.8 (94.9)
multiplicity	4.0 (4.1)
R_{merge} (%)	10.4 (37.2)
I/OI	5.9 (1.9)
total reflections	188058
unique reflections	46598
Refinement Parameters	
resolution limits (Å)	25.9–1.95 (1.98–1.95)
R_{factor} (%)	16.7 (23.9)
R_{free} (%)	19.8 (26.5)
reflections for R_{free}	4709 (187)
deviation from ideal values (bonds)	0.007
deviation from ideal values (angles)	1.130
twinning operator	$h, -k, -h-l$
twinning fraction	0.404

^aValues between parentheses correspond to the highest resolution shell.

(RmLamR) is a member of GH family 16, which was cloned, expressed, and characterized in 1998 by Krah et al.,⁵ displaying optimal activity in pH 5.5 and 88 °C. Therefore, RmLamR is a hyperthermophilic enzyme, and as such bears a considerable promise for industrial and biotechnological applications.

The optimal temperature for the stability and activity of each enzyme results from an intricate balance between charged, polar, nonpolar, and solvent interactions, which are affected differently by temperature. In general terms, it has been observed that thermostable proteins display enhanced polar interactions with the solvent and increased hydrophobic cores.^{1,2} However, the energetic balance is subtle, and some interactions may destabilize folded structures at different temperatures.⁶ For instance, surface salt bridges (SBs) do not contribute to protein stability at room temperature, but they do at high temperatures due to decreased desolvation penalties of charged groups.^{7–11} Therefore, understanding the interplay between different components of the protein–solvent and residue–residue interactions is a key factor for unveiling the molecular and structural reasons of a given enzyme's thermostability and thermophilicity.

Here, we discuss the molecular basis of the thermal stability and thermophilicity of family 16 laminarinases by solving the X-ray structure of the hyperthermstable RmLamR and by performing comparative molecular dynamics (MD) simulations of three structural homologues: the hyperthermophilic RmLamR from *R. marinus*, a thermostable laminarinase from alkaliphilic *Nocardiopsis*, and a mesophilic laminarinase from *Phanerochaete chrysosporium*. From the combination of structural and MD analysis we show that both the number and spatial distribution of SBs are determinant of protein thermal stability, and that distinct exposure of the active site occurs for laminarinases at different temperatures. The structural and dynamical features promoting thermal stability observed here can be useful for the rationalization of protein design, particularly for GHSs.

2. MATERIALS AND METHODS

2.1. Experimental Section. The laminarinase from *R. marinus* was cloned, expressed, purified, and crystallized as described.¹² The X-ray diffraction images from the crystal were collected at the MX-1 beamline at the Brazilian National Synchrotron Light Laboratory (LNLS), in Campinas, Brazil.¹³ The data set was indexed and integrated using MOSFLM,¹⁴ scaled in SCALA,¹⁵ and analyzed using phenix.xtriage.^{16,17} During this process it was detected that the crystal presented a rare case of pseudomeroedral twinning.¹² The structure of endo-β-1,3-glucanase from *Nocardiopsis sp.* (PDB code: 2HYK) was used as the search model for a molecular replacement in MOLREP.¹⁸ The structure was rebuilt with the correct sequence for RmLamR (GenBank: AAC69707.1; GI: 2896144) using RESOLVE^{19,20} as implemented in phenix.autobuild.²¹ After real-space model manipulation using COOT²² alternated with least-squares twin refinement as implemented in phenix.refine,²³ a final model consisting of two chains with 251 residues each, a glycerol molecule and two calcium ions converged at $R_{\text{factor}} = 16.7\%$ and $R_{\text{free}} = 19.8\%$. The full set of crystallographic parameters is listed in Table 1. The final crystallographic model and structure factors were deposited in the Protein Data Bank under PDB code 3ILN.

Secondary structural elements were assessed using Stride.²⁴ Analysis of GHF16 amino acid distributions was evaluated using version 23.0 of the PFAM database,²⁵ family PF00722. SSM SUPERPOSE²⁶ was used for three-dimensional superposition of protein structures.

2.2. Molecular Dynamics. The simulation boxes of the three homologous laminarinases were built using their crystallographic structures. The structure of laminarinase from *R. marinus* was obtained as described above. Structural models 2HYK and 2CL2 were obtained from the protein data bank for the alcaliphilic *Nocardiopsis sp.* strain F96 (aNLam)²⁷ and for laminarinase Lam 16A from *Phanerochaete chrysosporium* (PcLam),²⁸ respectively. Initially, the tridimensional aNLam structure had to be completed for missing residues at the extremities (TESDMR sequence peptide at the N-terminus and dipeptide LG at the C-terminus), according to its primary sequence. The coordinates of these residues were modeled by alignment of 2HYK to the RmLamR α-carbons. The positions of the other missing atoms were modeled according to the CHARMM27 topology file.²⁹ The enzymes were hydrated by 15 000 water molecules in a cubic box with sides of approximately 80 Å using Packmol,^{30,31} such that the solvent layer around the protein surface is at least 12 Å thick.

To set the ionization states of the ionizable residues (K, R, H, D, and E), the correspondent pK_a values were determined according to the specified pH 7 and the molecular environment (high dielectric constant at the protein surface and low dielectric constant in its interior) using the H++ server.^{32–34} Special attention was paid to the choice of the ionization states of the residues in the catalytic site. The residues D131 and E134 of RmLamR were considered protonated, so that they may interact with each other and with the substrate by means of hydrogen bonding. The nucleophile residue E129 was kept in its charged form. For all the simulated systems, there were 50 chloride and 50 sodium basal ions, in addition to the crystallographic calcium ion. To keep the system's electroneutrality, an excess of 1, 13, and 3 sodium ions were added in the RmLamR, 2HYK (aNLam), and 2CL2 (PcLam) systems, respectively. The resulting salt concentration is approximately 0.16 M.

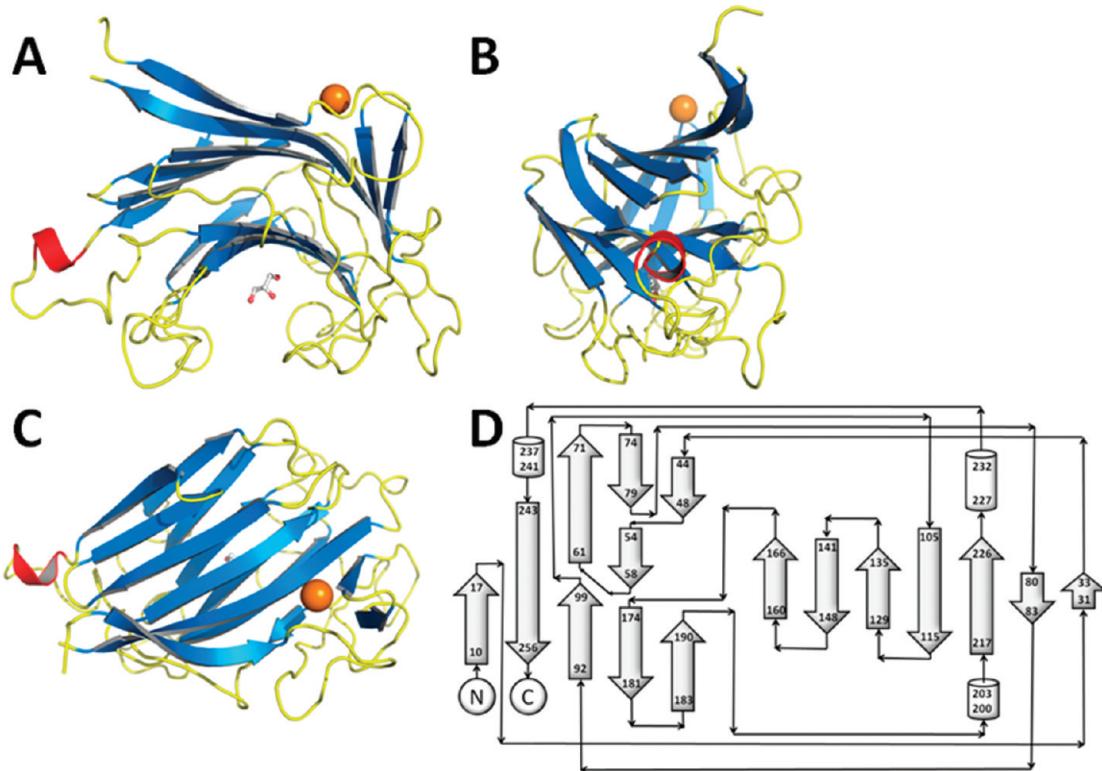


Figure 1. (A) Overall structure of the laminarinase from *R. marinus*, showing the calcium ion, which is related to stability (orange) and a glycerol molecule in the active pocket. (B,C) Same as A, after a 90° rotation upon the y and x axis. (D) Topology of RmLamR.

The energy of the system was initially minimized by 500 steps of the conjugate gradient (CG) method^{35,36} as implemented in NAMD³⁷ to eliminate bad contacts. After minimization, we performed equilibration runs consisting of three phases: In the first 10 ps, the positions of all residues, except the modeled ones, were kept fixed; in the next 100 ps, the lateral chains were released; and in the last 890 ps, the entire system was allowed to move. To run simulations at higher temperature (363 K), the equilibration period were extended in two other steps: 500 ps of equilibration at an intermediate temperature (330.5 K) and 1 ns of equilibration at 363 K. From the equilibrated systems, we carried out 10 independent 12 ns simulations for RmLam, eight 12 ns simulations for aNLam, and eight 9 ns runs for Pclam at 25 °C. At 90 °C, we performed five independent 9 ns simulations for each protein.

All simulations were performed with NAMD³⁷ applying periodic boundary conditions and CHARMM parameters. The TIP3P model was used for water.³⁸ Langevin dynamics and Nosé-Hoover Langevin piston methods^{39,40} were used to keep the temperature and pressure constant. The RESPA multiple-time step algorithm⁴¹ was used with the shortest time step of 2 fs. All hydrogen-to-heavy-atom bonds were kept rigid using SHAKE.⁴² A 12 Å cutoff with smooth switching function starting at 10 Å was used for the van der Waals interactions, whereas electrostatic forces were treated via the particle mesh Ewald method.⁴³

3. RESULTS AND DISCUSSION

3.1. The Structure of *R. marinus* Laminarinase: A Member of GH Family 16.

The laminarinase from *R. marinus* (RmLamR) was crystallized in space group P2₁, with a dimer in the asymmetric

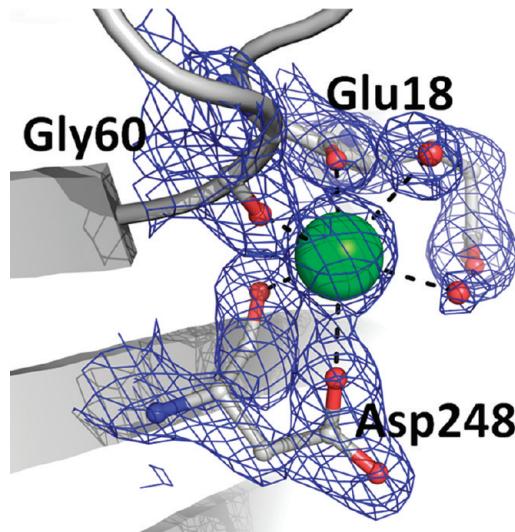


Figure 2. Coordination of the calcium ion bound to both chains in the RmLamR structure. The 6-fold coordination involves Asp240 (main chain and side chain), Gly60 (main chain), Glu18 (main chain), and two water molecules.

unit.¹² This dimer and all other possible quaternary assemblies formed by contacts in the crystal are unlikely to be found in solution, as indicated by theoretical estimates of the solvation free energy gain upon interface formation for each assembly obtained using the PISA server.⁴⁴ As previously observed by Krah et al.,⁵ the RmLamR structure is composed of two leaflets of antiparallel β -sheets in a complex jelly roll topology, which includes lengthy

Table 2. Amino Acid Content of RmLamR for Thermo-stability-Related Residues¹ Compared to Averages of Family 16 GHs^a

Residue type	GHF16	RmLamR
Glu	5.08 ± 0.19	6.5
Lys	4.29 ± 0.21	1.1
Arg	4.23 ± 0.19	8.7
Tyr	4.88 ± 0.18	5.4
Phe	4.84 ± 0.18	4.0
Asp	6.72 ± 0.17	6.5
His	2.42 ± 0.11	3.3

^a Although not originally included in Jaenicke and Böhm's classification,¹ results for the SB forming residues Asp and His are also shown for comparison.

loop connections, segmented β -strands, and three small helix segments, as shown in Figure 1. Its secondary structure content is much closer to that of *B. macerans* endo-1,3-1,4- β -glucanase than to that predicted by circular dichroism spectra:⁵ instead of 16% α -helices and 37% β -strands, our reported structure contains 1% α -helices (resulting solely from the small 3₁₀ helix segment) and 48% β -strands, similar to the 3% and 47%, respectively, in the *B. macerans* homologue.

Calcium ions bound to each monomer were clearly identified (Figure 2), in equivalent sites of each polypeptide chain in the asymmetric unit cell (ASU), and displaying the same coordination patterns. The calcium ions are known to be important for protein stability of the GH 16 family members.^{45,46} Coordination of the calcium ion is 6-fold, and involves Gly60 (main chain), Asp248 (main chain and side chain), Glu18 (main chain), and two water molecules.

A search using the DALI server⁴⁷ shows that the endo- β -1,3-glucanase from alkaliphilic *Nocardiopsis* sp. strain F96 (PDB code 2HYK), used in molecular replacement as a search model, is indeed the most similar structure in the PDB, but in addition to the expected similarity to other GH 16 structures, RmLamR also shares three-dimensional similarity to lectins and agglutinins (e.g., PDB codes 1FAY and 2E53) and also the *Botulinum neurotoxin* type B (PDB code 2NP0).

3.2. The RmLamR Active Site. Sequence analysis of members of the GH family 16 and site-directed mutagenesis have shown that catalytic activity of RmLamR is dependent on two glutamate and one aspartate residues located in a highly conserved motif (WX_{1–4}E[LIV]D[LIVF]X_{0–1}EX_{1–3}[GQ]).⁵ Moreover, the GH family 16 can be divided in two groups, which differ by the presence of an additional catalytic residue, usually a methionine preceding the second glutamate. The Met residue is responsible for a β -strand distortion, resulting in a β -bulge.^{48,49} The laminarinase from *R. marinus* has this additional methionine. Its presence results in one of six β -bulges in RmLamR, as deduced from the hydrogen bonding pattern between β -sheet containing Met133 and the neighboring β -sheet. As for other members of GH family 16,^{49,50} this β -bulge does not imply conformational changes in active site residues. The RmLamR structure also reveals a glycerol molecule fortuitously bound to the active site of the enzyme molecule. Anchoring of glycerol molecules by several hydrogen bonds to active site residues has been previously observed in the X-ray structure of the endo- β -1,3-glucanase from *Nocardiopsis* sp. strain F96.²⁷

Table 3. Number of SBs on RmLamR and Other GHF16 Structures on the PDB

structure (organism)	number of SBs
RmLamR (<i>R. marinus</i>)	18
2HYK (<i>Nocardiopsis</i> sp.)	7
1DYP (<i>P. carriageonovora</i>)	11
1GBG (<i>B. licheniformis</i>)	5
1MVE (<i>F. succinogenes</i>)	8
1UPS (<i>C. perfringens</i>)	17 (22) ^a
1U0A (<i>B. macerans</i>)	5
1UN1 (<i>P. tremula</i>)	11
1O4Y (<i>Z. galactanivorans</i>)	11
2CL2 (<i>P. chrysosporium</i>)	11

^a Structure 1UPS has an additional domain, residues 17–286, which correspond to the GHF 16 catalytic domain, have 17 SBs, while the full molecule has a total of 22 SBs.

3.3. Structure and Thermostability. One common feature of proteins bearing high thermal stability is the low content of loop regions.⁵¹ RmLamR, on the contrary, displays a large number of loops of considerable lengths (Figure 1). Remarkably, all loops and turns can be readily modeled from the electron density maps, even in regions where side-chains are barely seen. Thus, number and lengths of loops do not imply an abundance of disordered regions, and loops do not adopt multiple conformations in the crystal.

The sequence of RmLamR contains an increased number of arginines and glutamate residues relative to other GH family 16 members, as shown in Table 2. The abundance of charged and aromatic residues can be correlated to thermophilicity, as proposed by Jaenicke and Böhm.¹ This correlation was further refined to the abundance of Glu, Lys, and Arg (but not Asp), and to Phe and Tyr (but not Trp) by Goldstein.² The full amino acid content of RmLamR compared to averages of family 16 GHs is provided in the Supporting Information. The abundance of charged residues is correlated with the abundance of SBs, a factor widely regarded as crucial for the stability of thermophilic proteins.^{52,53} We evaluated the abundance of SBs in a nonredundant set of structures from RmLamR homologues, as shown in Table 3. RmLamR has the highest number of SBs, thus reinforcing their possible role in thermal stability (see below). The only comparable number of SBs comes from the endo- β -glucanase from *Clostridium perfringens*. This is a ubiquitous Gram-positive anaerobic bacterium, which can be found in diverse environments, from soil to intestinal tract of animals.⁵⁴ The high abundance of SBs may also stabilize this protein in the wide range of physiological conditions the organism exists. However, comparison of laminarinases from, say, *Nocardiopsis* sp. and *Phanerochaete chrysosporium*, shows nearly equal numbers of SBs, despite the fact that the former is a thermophilic protein with optimal activity at 77 °C,²⁷ whereas the latter is mesophile.²⁸ These differences motivated us to compare the dynamic behavior of these enzymes using MD simulations, with the aim of obtaining further insights into the structural nature of the thermal stability and thermophilicity of RmLamR.

Three structural models: of RmLamR (reported here) and of homologues endo- β -1,3-glucanase from alkaliphilic *Nocardiopsis* sp. strain F96 (aNLam, PDB entry 2HYK) and mesophilic *Phanerochaete chrysosporium* (PcLam, PDB entry 2CL2), respectively, were subjected to MD simulations, comprising structures

RmLamR	7	- - RLPHWELVWSDEFD - - YNGLPDPAKWDYDVGGHGWNQELQYYTRARI
aNLam	1	TESDMRATLVWSDEFDGPAGSAPPANHETGDHGWNNEQNNTDSRA
PcLam	1	- - - - ATYHLEDNW - - VGSALSTFTTHEAIADPTHGRVNVDQATALA
RmLamR	53	ENARVGGVLIIIEARRESGREYTSARLVTRGKASWTYGRFEIRAL - -
aNLam	51	NSALDGNLNLTARQEADGG - - YTSLRLLTQNKVQPQYGRVEASIQI - -
PcLam	42	KNLTYASGDTLILRADHTTTLSPSGPRNSVRIRSIKTTTHAVFDVRH
RmLamR	101	- PSGRTWPWMLPDRQTYGSAYWPDNGEIDIMEHVGFPNDPVHGTWHT
aNLam	97	- PRQGWIWPAFWML - - GADFPNTPWPDSGEIDIMENIGREPHLVHGSLG
PcLam	92	MPQCGTWPAAW - - - ETDEGDPWPNNGEVDIIEGVNDQSPNAMTLHTG
RmLamR	150	KAYNHLLGTQRGG - - - - - SIRVPTA - - RTDFHVYIAE
aNLam	144	PGY - FGGEPPLTG - - - - - SYMHPQQWSFADTFHTFADV
PcLam	136	ANCAMPASRTMTGHATNNNCVDNTDNTCCGVQAPTANSYGPSFNANGGG
RmLamR	180	WTPEEIR - - - - - WFVDDSLYYRFPNERLTNPSEADWRHWPFQPFH - - -
aNLam	175	WRPGSIT - - - - - WSDGVAYQTY - - - TSADTRGNPWFQPFH - - -
PcLam	168	WYAMERTNSFIKWFFPRNAGNVNPNDIASGPATINTDWGTPAFFPNTN
RmLamR	220	----- LIMNIAVGGTWGGQQGVDPFAFPAAQLVVVDYRVRYRWE - - -
aNLam	210	----- MILNVAVGGDWPGYPDGSTO - FPQEMRVDVYRVYELG - - -
PcLam	236	CDIGSHFDANNIIINLTFCGDWAGQASIFNGAGCPGSCVDYVNNNPSAFA
RmLamR		-----
aNLam		
PcLam	286	NAYWDIASVRVYQ

Figure 3. Structure-based sequence alignment of RmLamR, aNLam, and PcLam. Conserved catalytic (Glu, Asp, Glu) triad is highlighted in bold. The sequence identity between RmLamR and aNLam, and between RmLamR and PcLam are 44% and 13%, respectively.

with different thermal properties: the hyperthermophilic RmLamR, the thermophile aNLam, and the mesophile PcLam.

Despite the restricted primary sequence identity (Figure 3), the tertiary structures of these enzymes are very similar, and the catalytic residues harbored at the S8 β -sheet are highly preserved. The overall fold of the β -sheet leaflets is essentially identical for all three structures (Figure 4). Loop segments are fairly similar between RmLamR and aNLam, but differ substantially for PcLam, which contains numerous α -helices in peripheral regions. Because of the fairly low sequence identity and size differences between RmLam and PcLam, one should be cautious when comparing the behavior of these two proteins. Nevertheless, among the mesophilic members of this family of proteins with known crystal structures, PcLam is the closest to RmLam.

We performed simulations at low (25 °C) and high (90 °C) temperatures, to address the stability and mobility of the three structures. The global folds, as measured by structural parameters within MD simulations (Table 4), are mostly preserved in all three structures in the time-scale of our simulations, at both temperatures. For instance, the radii of gyration of the macromolecules are similar and invariant as the temperature increases from 25 to 90 °C. R_g of the heavier PcLam is slightly larger, but is also temperature independent. Departures from the corresponding crystal structures, measured by the root mean square displacements (RMSDs) of the protein backbone atoms, are close to 1 Å for all structures at all temperatures. The overall structural stability of the RmLamR and aNLam is consistent with their thermal properties since these enzymes preserve their activity at high temperature and, therefore, their structures should be preserved as well. Judging from the computed RMSD and R_g values alone, the PcLam structure seems to be reasonably stable at 90 °C, even though this enzyme is actually less thermostable. Further inspection of the time history of these structural parameters (see Supporting Information), suggests that thermal denaturation of these proteins may be under way, but the time scale of the simulations seems too short to reveal any major loss of structure. Increased stabilization effects may also stem from shortcomings

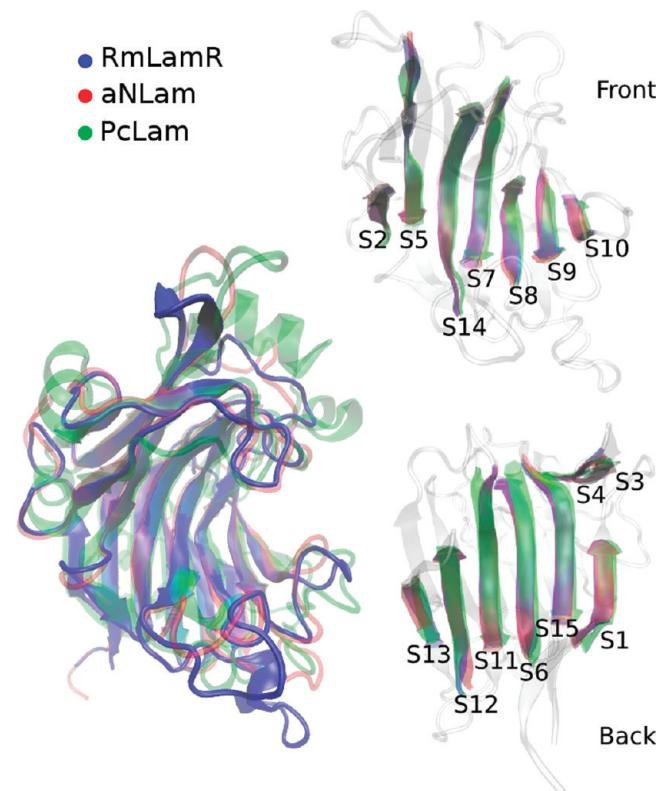


Figure 4. Structural alignment of laminarinases from *R. marinus* (RmLamR - blue), *Nocardiopsis* sp. (aNLam - red), and *P. chrysosporium* (PcLam - green). In detail, the β -sheets, numbered 1 through 15, are the major matching portions of these three structures.

of the CHARMM force field, which has been found to overestimate the secondary structure of proteins.⁵⁵

At the side-chain level, however, stability differences between these enzymes become noticeable. The average number of hydrogen bonds (HB) decreases from 77 at 25 °C to 65 at 90 °C for RmLamR, from 59 to 53 for aNLam, and from 77 to 62 for PcLam (a sizable proportion of PcLam H-bonds is present in peripheral regions). At the same time, the SBs of all structures remain nearly intact when the temperature increases from 25 to 90 °C. Importantly, the average number of hydrophobic contacts (HC) is approximately 10% larger in RmLamR than in the other two proteins. In all three homologues, the hydrophobic core of the protein is formed by the contacts between residues located in the opposite leaflets of antiparallel β -sheets. Nevertheless, in the RmLamR structure the hydrophobic core is more closely packed.

In summary, these results mainly reveal that, in comparison to its structural homologues, the thermophile RmLamR exhibits a substantially larger number of SBs and higher degrees of both intraprotein hydrogen bonding and hydrophobic contacts, thus possibly contributing to thermal stability. In terms of these parameters, aNLam and PcLam are quite similar. However, taking into consideration that PcLam is significantly larger than its homologues (RmLamR, aNLam, and PcLam have 251, 245, and 298 residues, respectively), the content of hydrogen bonds and hydrophobic contacts per residue is lower in PcLam.

How these features influence laminarinase thermostability? As RmLamR, aNLam, and PcLam are formed mainly by β -sheets that engulf the catalytic site and form the enzymes' scaffold, the

Table 4. Data Computed from MD Simulations of the Homologous Laminarinases (RmLamR, aNLam, and PcLam) at 25 and 90 °C^a

	RmLamR		aNLam		PcLam	
	25 °C	90 °C	25 °C	90 °C	25 °C	90 °C
R _g (Å)	17.7 ± 0.1	17.7 ± 0.1	17.3 ± 0.1	17.4 ± 0.1	18.3 ± 0.1	18.4 ± 0.1
RMSD (Å·ps ⁻¹)	0.82	0.94	0.92	1.04	0.72	0.94
Sec. Struc.	49.0%	48.8%	45.6%	44.2%	50.5%	49.8%
HB	77 ± 7	65 ± 7	59 ± 6	53 ± 6	77 ± 8	62 ± 6
HB per residue ^b	0.31	0.26	0.24	0.22	0.26	0.21
SB	24 ± 3	24 ± 2	9 ± 1	10 ± 2	11 ± 1	11 ± 1
HC	166 ± 2	165 ± 3	150 ± 2	151 ± 3	153 ± 3	152 ± 4

^a Structural parameters are radius of gyration (R_g), content of secondary structure (Sec. Struc.), number of hydrophobic contacts (HC), number of hydrogen bonds (HB), average number of salt bridges (SB), and backbone root mean square displacement (RMSD). Error bars were computed from the standard deviations obtained from the set of independent simulations for each system. ^b RmLamR, aNLam, and PcLam have 251, 245, and 298 residues, respectively.

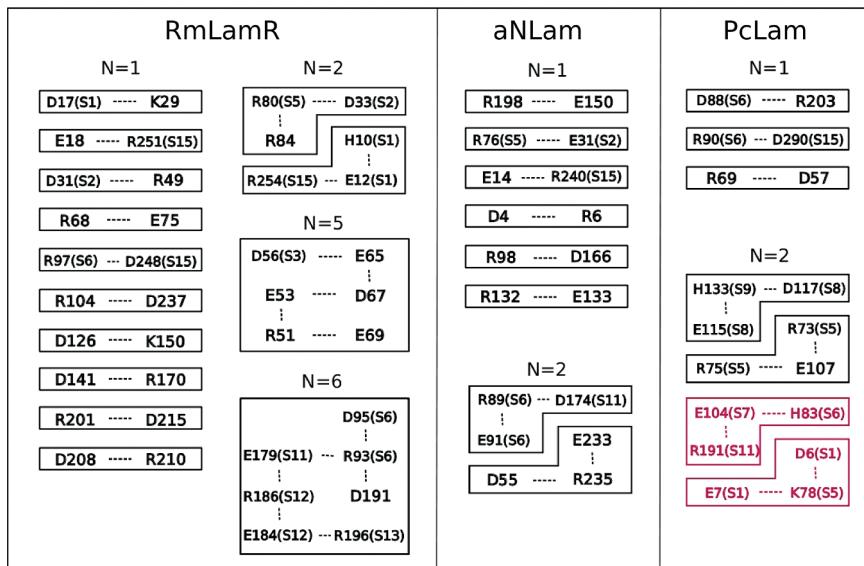


Figure 5. SB topological features of simulated laminarinases. Each dashed line represents an SB that remains formed during at least 50% of the simulation time. In parentheses are β -sheet labels. In RmLamR, there are clusters containing up to $N = 6$ interconnected SBs. In PcLam, there are SBs connecting residues located in the inner and outer β -sheet leaflets (marked red).

β -sheets also harbor most of the SBs (or salt clusters) present in these enzymes. Interestingly, the spatial distribution of SBs forms belt- or rim-like patterns connecting adjacent β -sheets on each leaflet. This is particularly striking for RmLamR: the connectivity maps of SBs, considering only those that persist for at least 50% of the total simulation time, show that in all three structures there are several isolated SBs (that is, SB clusters of size $N = 1$), as well as clusters containing $N = 2$ or more adjacent SBs (Figures 5 and 6). Remarkably, only RmLamR exhibits clusters of $N = 5$ or 6 interconnected SBs. The clusters of interconnected SBs link several sheets, as shown in Figure 6 (top panel), and confer an additional stabilization to the β -sheet scaffold for RmLamR.

Further inspection of the SB topology reveals that the three homologues differ in another important aspect: the SBs in the thermophilic structures do not connect residues belonging to the inner and outer leaflets of β -sheets, as opposed to the mesophile PcLam, in which four SBs connect the two leaflets. In PcLam (Figure 6, bottom panels), the Glu104 of S7 is simultaneously

bound to the His83 of S6 and Arg191 of S11; and the Lys78 of S8 forms ion bonds with the Asp6 and Glu7 positioned on S1. This is a very revealing feature: it indicates that in PcLam there are hydrophilic interactions bridging inner and outer β -sheet leaflets, whereas in more thermostable RmLamR and aNLam the hydrophobic core is comprised exclusively by the hydrophobic contacts between residues.

The presence of SBs in the hydrophobic core of the proteins has direct effects on their thermal resilience. Snapshots taken from simulations of the three proteins at 25 and 90 °C (Figure 7) show that RmLamR and aNLam strongly coordinate, respectively, one and two crystallographic water molecules in their hydrophobic cores at all temperatures, whereas in the less thermostable PcLam there is a marked increase in the number of water molecules penetrating the protein hydrophobic core, from an average of three water molecules at 25 °C up to six water molecules at 90 °C (see Supporting Information for further details). This suggests the onset of protein denaturation, as discussed above.

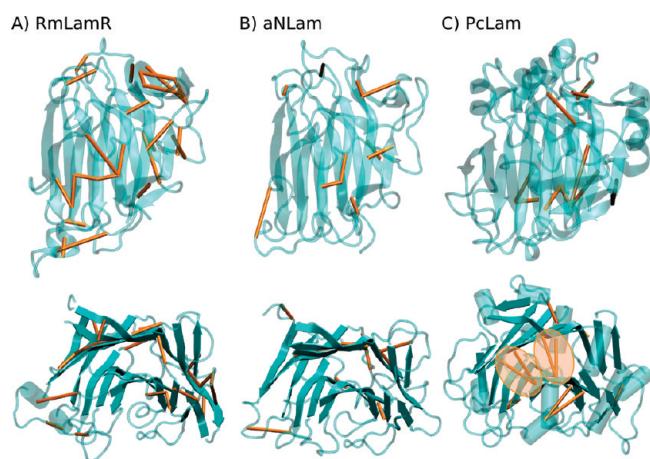


Figure 6. Topological organization of SBs of laminarinases. (a) Hyperthermophilic RmLamR. (b) Thermophilic aNLam. (c) Mesophilic PCLam. The connectivity of the salt-bridges, described in Figure 5 is represented in the top panels by connecting the Ca atoms of corresponding residues. The bottom panel shows that only in the mesophilic enzyme SBs are formed between the two leaflets of β -sheets. These SBs are observed to destabilize the hydrophobic core of the structure by facilitating water penetration.

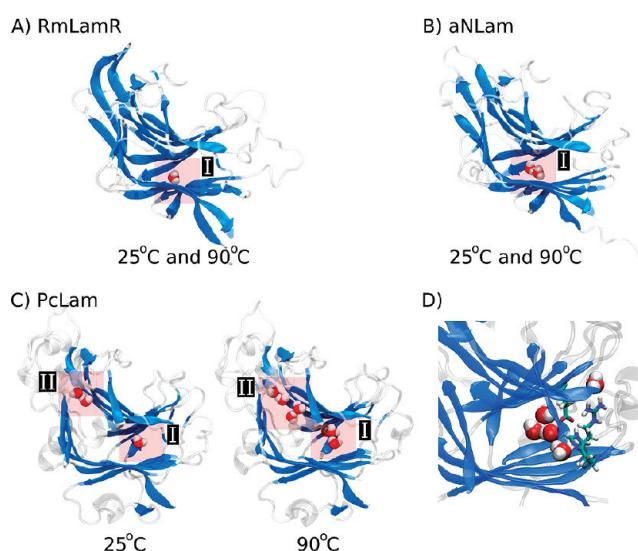


Figure 7. Details of water penetration between the inner and outer β -sheet leaflets. (A) One water molecule is present in the hydrophobic core of RmLamR at 25 and 90 °C (region I). The snapshot is representative of either set of simulations. (B) Two water molecules penetrate the hydrophobic core of aNLam at 25 and 90 °C (region I). (C) The hydrophobic core of PCLam is more water accessible, as three water molecules penetrate the core at 25 °C and up to six water molecules are found in that region of the protein at 90 °C (regions I and II). (D) Water penetration into the hydrophobic core of PCLam is facilitated by SBs that link the two β -sheet leaflets (see Figure 6).

The simulations also reveal that the SBs between the opposite β -sheet leaflets present in PCLam serve as a doorway for the entrance of water molecules into its hydrophobic core (Figure 7D), thus contributing to the thermal destabilization of PCLam compared to the other two more thermostable counterparts. Therefore, the disposition of the SBs is a crucial factor for

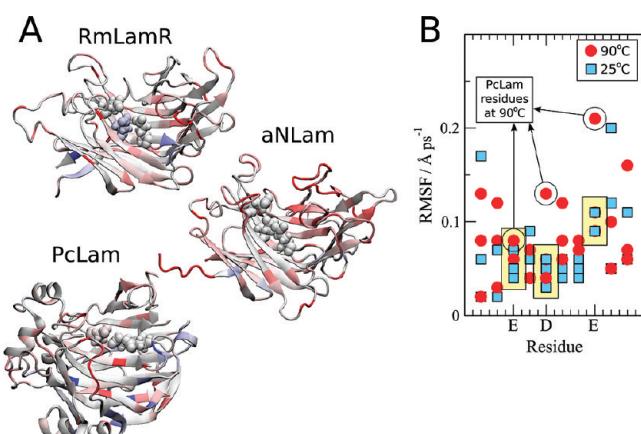


Figure 8. Mobilities of side chains of amino acid residues computed by their RMSF. (A) Differences in mobilities observed for structures at 25 and 90 °C. From blue to red, the colors indicate an increase in mobility at higher relative to lower temperatures. White indicates no change. (B) Mobilities of side chains of residues adjacent to the catalytic triad (E126, D128, and E131 for RmLamR); each symbol represents the mobility of one residue at one temperature. The mobility of the catalytic triad increases for the mesophilic PCLam, but is not sensitive to temperature variations for the thermophiles aNLam and RmLamR.

the thermal stability: SBs improperly oriented in the structure may play a destabilizing role.

3.4. Binding Cleft Dynamics and Accessibility. The properties of the side chains of the residues that delineate the substrate binding cleft must be preserved to enable enzymatic activity at high temperatures. Figure 8A depicts the relative mobility of the side chains displayed over the protein structures, with red and blue indicating higher and lower mobilities at 90 °C, respectively, relative to the mobilities at 25 °C. All structures exhibit regions in which the mobility of side chains have increased with increasing temperature, particularly the loops, and regions in which the side chains have been stabilized at 90 °C. Interestingly, the three side chains of the catalytic triad (Glu126, Asp128, and Glu131; RmLamR numbering) are among regions that have been stabilized or barely perturbed in the thermostable RmLamR and aNLam. At the same time, a noticeable increase in the mobility of the side chains can be observed for the active site residues in PCLam.

The mobility of the catalytic triad side-chains is not only marginally affected by the increased temperature in RmLamR and aNLam, but is also quantitatively similar to the mobilities of corresponding PCLam residues at 25 °C, as shown in Figure 8B. Thus, both the conformation and mobilities of the active site residues remain unperturbed for the thermophilic enzymes at high temperature. However, considerable alterations in the mobility are observed for residues Asp117 and Glu120 of the active site of PCLam at high temperature.

These results indicate that the thermophilic enzymes manage to preserve the active site flexibility and conformation that characterize the functional state, while experiencing other structural perturbations, particularly in loop regions. The compromise between structural stability and flexibility of proteins in different regions is a key aspect of biomolecular function in extreme environments.^{56,57} The essential adaptive alterations tend to bias common characteristics of these proteins toward the respective optimal environmental conditions in such a way as to maintain these proteins in the so-called “corresponding states” regarding

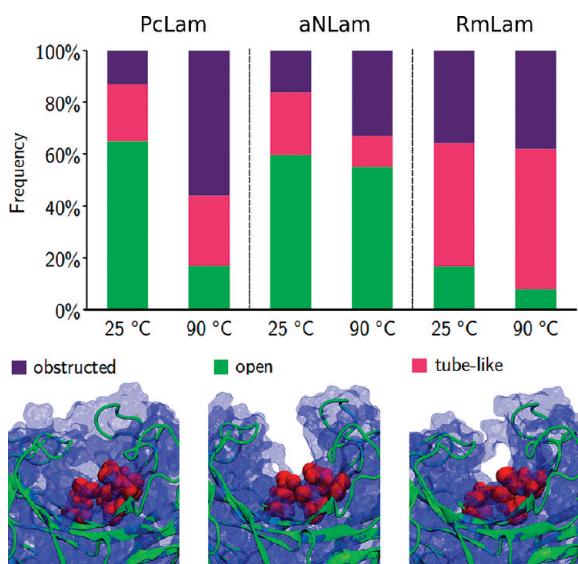


Figure 9. Frequency distribution of the conformational states of the substrate binding cleft. The obstructed conformation is likely to strongly impair enzymatic activity, whereas open and tube-like conformations are expected to be active. Increase in the relative population of closed (inactive) conformations agrees qualitatively with the thermophilicity of the proteins. In addition, relative populations are quite insensitive to variations in temperature for hyperthermophile RmLamR. Representative snapshots of simulations illustrate the structural nature of each configuration, with catalytic residues depicted by red spheres.

folding topology, structural flexibility, and solvent exposure.⁵⁸ As discussed above, the RmLamR, aNLam, and PcLam display active-site flexibilities that are similar at temperatures of corresponding optimal activity.

The side-chain mobility of residues in the vicinity of the active site exposes yet another key differential aspect of the thermophilic vs mesophilic enzymes at high temperature. As shown in Figure 9, the exposure of the active sites of the three enzymes is markedly affected in different ways by the motions of nearby side chains. In all structures, the substrate binding cleft can be cast in three major characteristic conformations: (1) A fully *closed* or *obstructed* state, in which the binding cleft is blocked by the side chains of nearby residues: Trp220, Trp230, and Trp257, located right above the catalytic triad for RmLam, aNLam, and PcLam, respectively; Trp37 and Phe113 located in the opposing loops of aNLam; and Pro26 and Trp110, located in opposite loops of PcLam. (2) A fully *open* or *productive* state, in which the active site channel is free and fully accessible to either solvent or substrate, and (3) a *tube-like* state, in which some side chains partially cover the binding cleft forming a tube-like structure in one end of the cleft without blocking the channel along its full extension. We classified all conformations of the binding cleft generated during the entire course of the simulations into these three main states. Detailed quantitative information regarding the structural criteria we use to classify the conformational states of the binding cleft and calculate the corresponding frequencies is provided in the Supporting Information.

The state density or population is similar for PcLam and aNLam at 25 °C, with a large fraction of the population being the open channel states. However, their response to the temperature increase differs: The conformational population of PcLam at 90 °C is dominated by *obstructed* states, whereas open cleft

conformations prevail for aNLam even at high temperature. Thus, given that obstructing the binding channel is likely to yield lower levels of enzymatic activity, the simulations provide a plausible rationale, at the molecular level, for the observed higher thermophilicity of aNLam compared to PcLam.

The distribution of the substrate binding cleft conformations of RmLamR, on the other hand, is remarkably less sensitive to the temperature increase. The conformational population is dominated by the formation of *tube-like* structures in one end of the catalytic cleft at 25 °C, and this population remains predominant at 90 °C. From 25 to 90 °C the population of open (productive) cleft configuration reduces toward both closed and tube-like conformations. The low population of obstructed channel conformations of RmLamR at both 25 and 90 °C relative to PcLam and aNLam is consistent with its higher activity at elevated temperatures.

We conjecture that the tubular conformation, which is highly populated at both temperatures for RmLamR is, in fact, the most active. It could allow for the permeation of the polysaccharide substrate into the active site channel and restrain its release to solvent, thus stabilizing the activated complex by providing more effective enzyme–substrate interactions. The formation of this tunnel is dependent on the mobility of the side chain of the Trp230 residue located in one of the loops that embrace the substrate binding cleft. The simulations suggest that the side chain mobilities of these residues seem to be crucial for the enzyme's activity.

4. CONCLUSIONS

In this work we studied the structural basis for the thermostability and thermophilicity of laminarinases by solving the X-ray structure of hyperthermophilic laminarinase from *Rhodothermus marinus* at 1.95 Å and by performing MD simulations on structurally homologous laminarinases with different thermophilicities. RmLamR is formed mostly by β -sheets in a complex jelly roll topology that is conserved among members of glycosid hydrolase family 16. The enzyme contains a high number of SBs, only comparable with other GHs that must be resistant to multiple extreme environmental conditions.

We show that the thermostability of the structures is dependent on the number of SBs, but that these cannot be randomly distributed. Thermal stability is achieved by clustering SBs and by avoiding their interaction with the hydrophobic core of the proteins. SBs across the hydrophobic core of one structure actually introduce a destabilizing factor by facilitating water permeation. Furthermore, we show that mobility and substrate accessibility to the active site do not depend on temperature in a trivial fashion. Yet, MD simulations provided valuable insights into structural alterations which are consistent with the observed thermophilicity. In particular, we describe two alternative conformations of the active site channel which are not observed in the crystallographic models, one of which putatively disrupts activity, whereas the other may enhance enzymatic activity by entrapping the polysaccharide chain within the binding site. The populations of these structures seem to reflect the relative activity of the laminarinases at low and high temperatures.

We advocate that the X-ray structure of RmLamR presented here and its structural and MD analyses might provide useful insights for the engineering of the enzymes of GHF16 and other GHs with enhanced properties for industrial applications. This, without a doubt, can be of considerable importance in the present

era of protein engineering research applied to biotechnology and to second-generation biofuel technologies.

■ ASSOCIATED CONTENT

Supporting Information. Further simulation details are provided in Table S1. The full amino acid content of RmLamR is provided in Table S2. Structural parameters reported in Table 4 as functions of time for each independent simulation are shown in Figures S1–S7. Details on the calculation of the hydration levels of the hydrophobic core are provided, and the results as functions of time are shown in Figure S8. A geometric criterion for the conformational states of the binding cleft is introduced and illustrated by Figures S9–S11. This information is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail addresses: ipolikarpov@ifsc.usp.br (I.P.); skaf@iqm.unicamp.br (M.S.S.); leandro@ifsc.usp.br (L.M.).

Author Contributions

L. Bleicher, E. T. Prates, T. C. F. Gomes, and R. L. Silveira contributed equally to this work.

■ ACKNOWLEDGMENT

We thank the Brazilian funding agencies CNPq and FAPESP (Grant Nos. 08/56225-9, 10/16947-9, 10/18849-4, 09/14107-6, 10/08680-2, 10/16947-9) and the MCT/CNPq/FAPESP EU-Brazil Collaboration program in Second Generation Biofuels (Grant No. 490022/2009-0) for financial support. We also thank the Brazilian National Synchrotron Laboratory for the utilization of the MX1 beamline. L.B. also thanks Ralf Grosse-Kunstleve and Peter Swart.

■ REFERENCES

- (1) Jaenicke, R.; Bohm, G. *Curr. Opin. Struct. Biol.* **1998**, *8*, 738–748.
- (2) Goldstein, R. A. *Protein Sci.* **2007**, *16*, 1887–1895.
- (3) Davies, G.; Henrissat, B. *Structure* **1995**, *3*, 853–859.
- (4) Bjornsdottir, S. H.; Blöndal, T.; Hreggvidsson, G. O.; Eggertsson, G.; Petursdottir, S.; Hjorleifsdottir, S.; Thorbjarnardottir, S. H.; Kristjansson, J. K. *Extremophiles* **2006**, *10*, 1–16.
- (5) Krah, M.; Misselwitz, R.; Politz, O.; Thomsen, K. K.; Welfle, H.; Borriß, R. *Eur. J. Biochem.* **1998**, *257*, 101–111.
- (6) Argos, P.; Rossmann, G. M.; Grau, U. M.; Zuber, H.; Frank, G.; Tratschin, J. D. *Biochemistry* **1979**, *18*, 5698–5703.
- (7) Vogt, G.; Woell, S.; Argos, P. *J. Mol. Biol.* **1997**, *269*, 631–643.
- (8) Strop, P.; Mayo, S. L. *Biochemistry* **2000**, *39*, 1251–1255.
- (9) Elcock, A. H. *J. Mol. Biol.* **1998**, *284*, 489–502.
- (10) Priyakumar, U. D.; Ramakrishna, S.; Nagarjuna, K. R.; Reddy, K. S. *J. Phys. Chem. B* **2010**, *114*, 1707–1718.
- (11) Xiao, L.; Honig, B. *J. Mol. Biol.* **1999**, *289*, 1435–1444.
- (12) Golubev, A. M.; Rojas, A. L.; Nascimento, A. S.; Bleicher, L.; Kulminskaya, A. A.; Eneyskaya, E. V.; Polikarpov, I. *Protein Pept. Lett.* **2008**, *15*, 1142–1144.
- (13) Polikarpov, I.; Perles, L. A.; Oliveira, R. T.; Oliva, G.; Castellano, E. E.; Garratt, R. C.; Craievich, A. J. *Synchrotron Radiat.* **1998**, *5*, 72–76.
- (14) Leslie, A. G. *Acta Crystallogr. D* **1999**, *55*, 1696–1702.
- (15) Evans, P. *Acta Crystallogr. D* **2006**, *62*, 72–82.
- (16) Zwart, P. H.; Grosse-Kunstleve, R. W.; Adams, P. D. *CCP4 Newsletter* **2005**, *42*, 58–67.
- (17) Zwart, P. H.; Grosse-Kunstleve, R. W.; Adams, P. D. *CCP4 Newsletter* **2005**, *43*, 27–35.
- (18) Vagin, A.; Teplyakov, A. *J. Appl. Crystallogr.* **1997**, *30*, 1022–1025.
- (19) Terwilliger, T. C. *Acta Crystallogr. D* **2003**, *59*, 38–44.
- (20) Terwilliger, T. C. *Acta Crystallogr. D* **2003**, *59*, 45–49.
- (21) Terwilliger, T. C.; Grosse-Kunstleve, R. W.; Afonine, P. V.; Moriarty, N. W.; Zwart, P. H.; Hung, L. W.; Read, R. J.; Adams, P. D. *Acta Crystallogr. D* **2008**, *64*, 61–69.
- (22) Emsley, P.; Cowtan, K. *Acta Crystallogr. D* **2004**, *60*, 2126–2132.
- (23) Afonine, P. V.; Grosse-Kunstleve, R. W.; Adams, P. D. *CCP4 Newsletter* **2005**, *42*, 43–49.
- (24) Frishman, D.; Argos, P. *Proteins* **1995**, *23*, 566–579.
- (25) Finn, R. D.; Tate, J.; Mistry, J.; Coggill, P. C.; Sammut, S. J.; Hotz, H. R.; Ceric, G.; Forslund, K.; Eddy, S. R.; Sonnhammer, E. L. L.; Bateman, A. *Nucleic Acids Res.* **2008**, *36*, D281–D288.
- (26) Krissinel, E.; Henrick, K. *Acta Crystallogr. D* **2004**, *60*, 2256–2268.
- (27) Fibriansah, G.; Masuda, S.; Koizumi, N.; Nakamura, S.; Kumazaka, T. *Proteins* **2007**, *69*, 683–690.
- (28) Vasur, J.; Kawai, R.; Larsson, A. M.; Igarashi, K.; Sandgren, M.; Samejima, M.; Stahlberg, J. *Acta Crystallogr. D* **2006**, *62*, 1422–1429.
- (29) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Wantanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (30) Martínez, J. M.; Martínez, L. *J. Comput. Chem.* **2003**, *24*, 819–825.
- (31) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. *J. Comput. Chem.* **2009**, *30*, 2157–2164.
- (32) H++ Server. <http://biophysics.cs.vt.edu/H++>. Accessed November 26, 2010.
- (33) Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A. *Nucleic Acids Res.* **2005**, *33*, W368–W371.
- (34) Anandakrishnan, R.; Onufriev, A. *J. Comput. Biol.* **2008**, *15*, 165–184.
- (35) Fletcher, R. *Practical Methods of Optimization*, 2nd ed.; Wiley: Padstow, 2000.
- (36) Fletcher, R.; Reeves, C. M. *Comput. J.* **1964**, *7*, 149–154.
- (37) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- (38) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (39) Schneider, T.; Stoll, E. *Phys. Rev. B* **1978**, *17*, 1302–1322.
- (40) Martyna, G. J.; Tobias, D. J.; Klein, M. L. *J. Chem. Phys.* **1994**, *101*, 4177–4189.
- (41) Tuckerman, M. E.; Berne, B. J.; Martyna, G. J. *J. Chem. Phys.* **1992**, *97*, 1990–2001.
- (42) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (43) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (44) Krissinel, E.; Henrick, K. *J. Mol. Biol.* **2007**, *372*, 774–797.
- (45) Keitel, T.; Meldgaard, M.; Heinemann, U. *Eur. J. Biochem.* **1994**, *222*, 203–214.
- (46) Tsai, L. C.; Shyur, L. F.; Lee, S. H.; Lin, S. S.; Yuan, H. S. *J. Mol. Biol.* **2003**, *330*, 607–620.
- (47) Holm, L.; Kaariainen, S.; Rosenstrom, P.; Schenkel, A. *Bioinformatics* **2008**, *24*, 2780–2781.
- (48) Neustroev, K. N.; Golubev, A. M.; Sinnott, M. L.; Borriß, R.; Krah, M.; Brumer, H., III; Eneyskaya, E. V.; Shishlyannikov, S.; Shabalin, K. A.; Peshechonov, V. T.; Korolev, V. G.; Kulminskaya, A. A. *Glycoconjugate J.* **2006**, *23*, 501–511.

- (49) Allouch, J.; Jam, M.; Helbert, W.; Barbeyron, T.; Kloareg, B.; Henrissat, B.; Czjzek, M. *J. Biol. Chem.* **2003**, *278*, 47171–47180.
- (50) Vasur, J.; Kawai, R.; Hanna, K.; Jonsson, M.; Widmalm, G.; Engström, A.; Frank, M.; Andersson, E.; Hansson, H.; Forsberg, Z.; Igarashi, K.; Samejima, M.; Sandgren, M.; Ståhlberg, J. *J. Am. Chem. Soc.* **2010**, *132*, 1724–1730.
- (51) Thompson, M. J.; Eisenberg, D. *J. Mol. Biol.* **1999**, *290*, 595–604.
- (52) Kumar, S.; Nussinov, R. *ChemBioChem* **2002**, *3*, 604–617.
- (53) Dominy, B. N.; Minoux, H.; Brooks, C. L., III. *Proteins* **2004**, *57*, 128–141.
- (54) Petit, L.; Gibert, M.; Popoff, M. R. *Trends Microbiol.* **1999**, *7*, 104–110.
- (55) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III. *J. Comput. Chem.* **2004**, *25*, 1400–1415.
- (56) Tsou, C.-L. *Biochemistry* **1988**, *27*, 1809–1812.
- (57) Tsou, C.-L. *Biochemistry* **1998**, *63*, 253–258.
- (58) Jaenicke, R. *Eur. J. Biochem.* **1991**, *202*, 715–728.