

ARTICLE

DOI: 10.1038/s42003-018-0167-7

OPEN

Loss of a conserved salt bridge in bacterial glycosyl hydrolase BglM-G1 improves substrate binding in temperate environments

Dipali Mhaindarkar¹, Raphael Gasper², Natalie Lupilov¹, Eckhard Hofmann² & Lars I. Leichert ¹

Salt bridges are the strongest electrostatic interactions in proteins. They substantially contribute to a protein's structural stability. Thus, mutations of salt bridges are typically selected against. Here, we report on the evolutionary loss of a highly conserved salt bridge in the GH1 family glycosyl hydrolase BglM-G1. BglM-G1's gene was found in the bacterial metagenome of a temperate, seasonally cold marine habitat. In BglM-G1, arginine 75 is replaced by a histidine. While fully retaining β -glucosidase activity, BglM-G1 is less heat stable than an H75R variant, in which the salt bridge was artificially re-introduced. However, the K_m toward its substrates was lower in wild type, leading to an overall higher catalytic efficiency. Our results indicate that this loss of the salt bridge leads to higher flexibility in BglM-G1's active site, trading structural stability at high temperatures, a trait not needed in a temperate, seasonally cold habitat, for a more effective catalytic activity.

¹Ruhr University Bochum, Fakultät für Medizin, Institute for Biochemistry and Pathobiochemistry, Microbial Biochemistry, Universitätsstr. 150, 44780 Bochum, Germany. ²Ruhr University Bochum, Faculty of Biology and Biotechnology, Department of Biophysics, Protein Crystallography, Universitätsstr. 150, 44780 Bochum, Germany. Correspondence and requests for materials should be addressed to L.I.L. (email: lars.leichert@ruhr-uni-bochum.de)

Structural stability is governed by hydrophilic and hydrophobic interactions between the backbone and side chains of individual amino acids in a protein. Hydrogen bonds and van der Waals interactions are often the largest contributors to overall protein stability, but individually, salt bridges can contribute substantial energy to protein folding and stability^{1–3}. However, our knowledge about mutations that disrupt or enhance these interactions, and their effect on the stability and function of proteins is derived primarily from lab studies, in which these salt bridges were introduced or removed on purpose^{4,5}. Indeed, very few studies have been conducted that look at the naturally occurring mutations in a protein, the β -lactamase TEM1 being a rare exception⁶. Since the typical thermodynamic stability of a protein fold, $\Delta G_{\text{folding}}$ is in the range of -20 to -65 kJ mol⁻¹¹⁷, the destruction of one particularly stabilizing interaction, such as a salt bridge or a hydrogen bond can lead to the complete unfolding of a protein. Thus, in general, most mutations are destabilizing and therefore are evolutionary selected against⁸.

Here we report, in contrast, on a naturally occurring mutation of an otherwise invariant residue in the β -glucosidase BglM-G1 that balances a trade-off between structural stability and enzyme activity. β -glucosidases are enzymes hydrolyzing β -glycosidic bonds in substrates ranging from glycosides, to glucosylated flavonoids, glucosylceramides, and oligosaccharides, releasing non-reducing terminal glucosyl residues (see Ketudat et al.⁹ for a comprehensive review). They are also one of the key enzymes in the bioconversion of cellulosic material^{10,11}. The important conserved amino acids in β -glucosidases have been well-defined through comparison of primary and 3D structures¹². In particular, 37 highly conserved amino acids have been identified that are crucial for the function of proteins from glycoside hydrolase family 1 (GH1)¹³. We studied the impact of the naturally occurring substitution of one of these amino acids. The otherwise invariant arginine at position 75 changed to a histidine in this protein, while the other 36 amino acids deemed necessary for the function and structural integrity of members of this protein family remained unchanged. This mutation decreased stability at high temperatures but increased catalytic efficiency toward its substrates. The molecular basis for these changes was investigated through X-ray crystallography of BglM-G1 and its H75R variant, which re-introduces the otherwise invariant arginine. Our data indicate that the natural mutation to histidine at position 75 leads to the loss of an extensive electrostatic network, including a salt bridge in the active site region and implies that evolution of BglM-G1 traded a decrease in thermal stability for improved catalytic efficiency.

Results

The metagenomic protein BglM-G1 is a bona fide β -glucosidase. β -glucosidases of the GH1 are one of the key enzymes in the conversion of cellulose into glucose (for a comprehensive resource for the classification of carbohydrate-active enzymes see the CAZy database¹⁴). As such, these biocatalysts are essential metabolic enzymes of cellulolytic organisms, as well as tools for the biotechnological conversion of cellulolytic biomass into biofuels and fine chemicals. To identify new β -glucosidases with favorable traits, we searched one of the largest metagenomic datasets, the global ocean sampling (GOS) dataset¹⁵, for GH1 family members with naturally occurring mutations in otherwise invariant residues. Among the 132 GH1 family members in this dataset, we found only two sequences which featured a substitution of the invariant arginine 75 to a histidine. Both sequences were found in DNA isolates from Chesapeake Bay, Maryland, USA, obtained from very cold water (1 °C) and had a

99% amino acid identity. One of the sequences was otherwise unchanged at the other 36 residues, which are considered highly conserved or invariant in this enzyme family¹³. We selected this protein for further studies and termed it BglM-G1 (GH1 family β -glucosidase of metagenomic origin). Bioinformatic analysis revealed that BglM-G1 exhibited 99% amino acid sequence identity with a β -glucosidase from *Actinobacteria bacterium* BACL2 MAG-120820-bin50 (GenBank accession KRO51423). This sequence was derived by genome assembly from DNA samples obtained from Baltic sea water surface and contains the same mutation of residue 75 to histidine. The closest well-characterized homolog of BglM-G1 was a β -glucosidase from the thermophilic, anaerobic eubacterium *Thermotoga neapolitana* with 50.1% amino acid identity (UniProt accession number P0C946, BGLA_TENN in Fig. 1a)¹⁶. A conservation analysis of the BglM-G1 sequence using ConSurf¹⁷ confirmed that arginine at position 75 is indeed invariant in all other known homologs of BglM-G1 (Fig. 1a, Supplementary Fig. 1). To test if BglM-G1 is, despite the mutation of an invariant amino acid, still a bona fide β -glucosidase, we created a synthetic gene *bglM-G1*, which we then could express from a plasmid in *Escherichia coli*. We tested this expression clone for β -glucosidase activity in an esculin-based plate assay. The presence of a black halo surrounding the clone confirmed its β -glucosidase activity (Fig. 1b).

Enzymatic characteristics of BglM-G1. To biochemically characterize BglM-G1, we expressed a recombinant, N-terminally His₆-tagged and TEV-cleavage site containing version in *E. coli* and purified it to homogeneity by Ni-NTA affinity chromatography. Activity of this protein was then tested using the chromogenic substrate *p*-nitrophenyl- β -D-glucopyranoside (pNPGlu). BglM-G1 showed optimal activity at pH 6.0 in citrate buffer and had more than 60 % activity in a pH range from 5.0 to 6.5 (Fig. 1c). BglM-G1 was stable in a range from pH 6.0 to 9.0 (Fig. 1d). Its temperature optimum was 45 °C (Fig. 1e), making it by most definitions a mesophilic enzyme. The specific activity of BglM-G1 at optimal pH and temperature was 56.3 ± 1.7 $\mu\text{mol min}^{-1} \text{mg}^{-1}$ (Table 1), comparable to the activity of other GH1 β -glucosidases^{18,19}. Other substrates, toward which BglM-G1 showed high activity were *p*-nitrophenyl- β -D-fucopyranoside (pNPFuc) and *p*-nitrophenyl- β -D-galactopyranoside. Further substrates containing β -glycosidic bonds such as *p*-nitrophenyl- β -D-cellobioside and *p*-nitrophenyl- β -D-xylopyranoside were also hydrolyzed, although the activity towards these substrates was only 4 and 14% of the activity towards BglM-G1's most preferred substrate, respectively (Table 1). No activity against any of the tested substrates that contained α -glycosidic bonds was detected.

Structure of BglM-G1 lacks electrostatic interaction network.

The fact that BglM-G1 evolved to be a fully active β -glucosidase with a naturally occurring substitution of the otherwise invariant arginine at position 75 provided us with the unique opportunity to study the structural implications of this amino acid in β -glucosidases. We, therefore, re-introduced arginine, replacing the histidine 75 in this enzyme, creating the BglM-G1 variant H75R. This variant showed enzymatic activity (Fig. 1b) and characteristics highly similar to the wild type enzyme with the same pH optimum and buffer preferences (Fig. 1f), optimal temperature (Fig. 1g), and comparable substrate specificities (Table 1). Circular dichroism (CD) spectroscopy demonstrated an overall structural similarity of both wild type (Fig. 2a) and the H75R variant of BglM-G1 (Fig. 2b). For a more detailed structural analysis, we then crystallized both wild type (PDB ID 5NS6) and the H75R variant (PDB ID 5NS7) of BglM-G1 and solved their X-ray structure using data to 1.50 and 1.54 Å, respectively

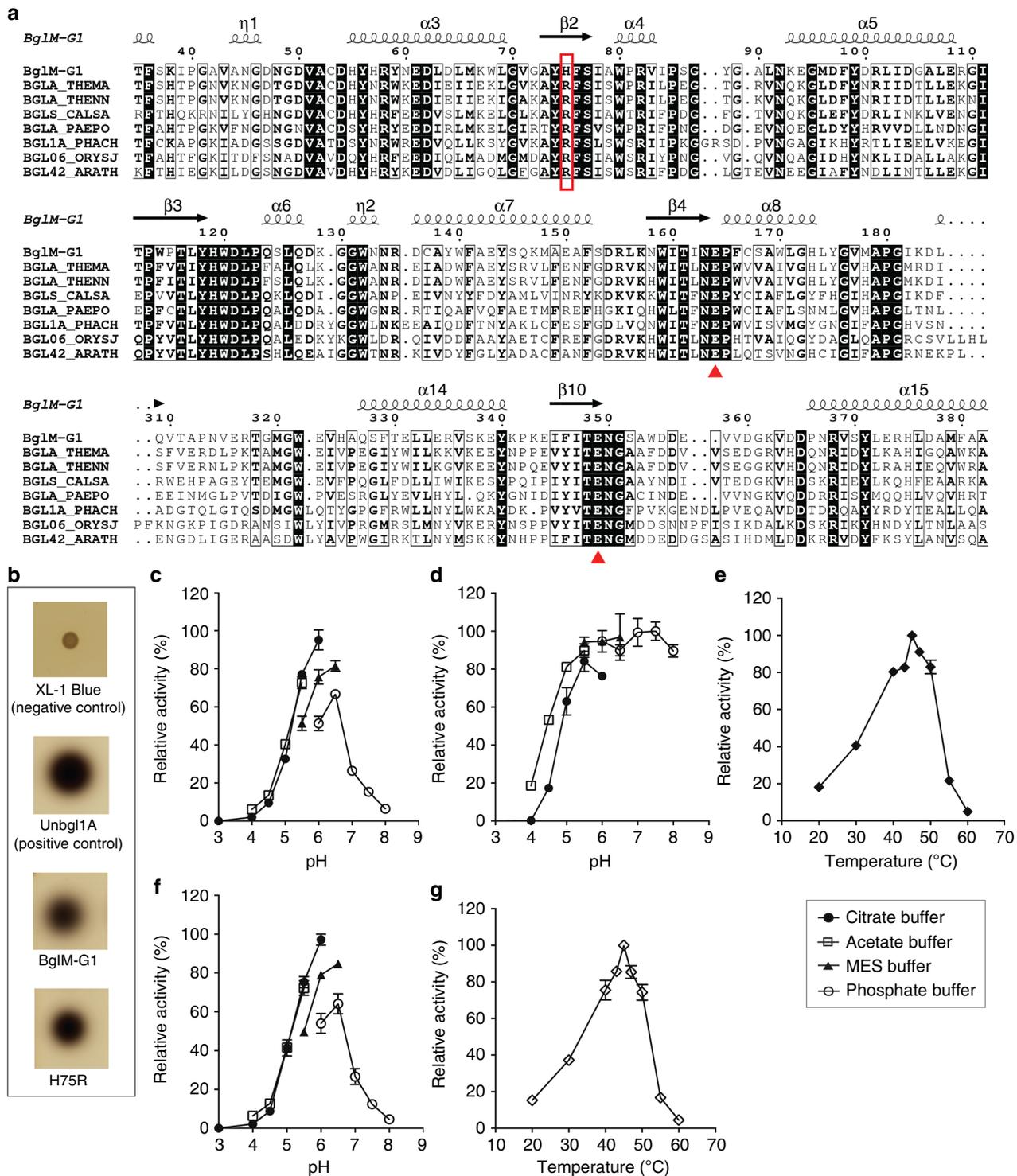


Fig. 1 Primary sequence alignment and enzymatic characteristics of BglM-G1 and its H75R variant. **a** Partial alignment of BglM-G1 with homologous GH1 β-glucosidases. BGLA_THEMEA: β-glucosidase A from *Thermotoga maritima* (UniProt ID Q08638); BGLA_THENN: 1,4-β-D-glucan glucohydrolase from *Thermotoga neapolitana* (UniProt ID B9K7M5); BGLS_CALSA: β-glucosidase A from *Caldicellulosiruptor saccharolyticus* (UniProt ID P10482); BGLA_PAEPO: β-glucosidase A from *Paenibacillus polymyxa* (UniProt ID P22073); BGLA1A_PHACH: β-glucosidase 1A from *Phanerochaete chrysosporium* (UniProt ID Q25BW5); BGL06_ORYSJ: β-glucosidase 6 from *Oryza sativa subsp. japonica* (UniProt ID Q8L7J2); BGL42_ARATH: β-glucosidase 42 from *Arabidopsis thaliana* (UniProt ID Q9FIW4). A box in red color indicates the position of conserved Arginine 75. Triangles in red color indicate the positions of the catalytic glutamate residues. **b** Esculin-based plate assay testing for β-glucosidase activity in *E. coli* expressing BglM-G1 and H75R. There was no background activity in the *E. coli* XL-1 Blue host. Positive control (Unbgl1A) was a β-glucosidase isolated from a soil metagenomic library (GenBank accession JX566949)⁵¹. **c** Relative enzyme activity of BglM-G1 over a range of different pHs. Optimum activity is achieved in citrate buffer at pH 6.0. **d** pH stability profile of BglM-G1. **e** Activity of BglM-G1 at different temperatures. Optimum activity is obtained at a temperature of 45 °C. **f** Relative enzyme activity of H75R over a range of different pHs. The optimum pH is virtually identical to wild type. **g** Additionally, the optimum temperature for enzyme activity of H75R was identical at 45 °C. Error bars represent standard deviation, where shown

(Table 2). Both BglM-G1 and BglM-G1's variant H75R crystallized in space group P43 with three monomers in the asymmetric unit (Fig. 2d). Based on size-exclusion chromatography, which

showed that both proteins exist in a monomeric form in solution (Fig. 2c), we assume that the trimeric assembly in the crystal is packing-generated. The structure of BglM-G1 was solved by molecular replacement using family 1 β -glucosidase from *Thermotoga maritima*, TmGH1 (PDB ID 1OD0, BGLA_THEME in Fig. 1a) as a search model. Comparison of the structures of BglM-G1 and TmGH1 monomers revealed a root mean square deviation (RMSD) value of 0.53 Å over 313 Ca atoms and the conformation of the active site glutamates were highly similar (Fig. 2e, f). Upon solving the structure, we observed both in wild type and H75R an unresolved electron density within the active site tunnel. This density does not seem to originate from a peptide chain and we were not able to fit an atomic structure to it. On a sequence level, BglM-G1 and TmGH1 show 49% amino acid sequence identity. Both BglM-G1 and its H75R variant showed a $(\beta/\alpha)_8$ TIM barrel topology, typical for the GH-A clan including GH1 members²⁰. Overall, the side chain and backbone topologies of the amino acid residues in BglM-G1 and its H75R variant were highly similar with an RMSD value of 0.069 Å over 1217 Ca atoms, confirming that the mutation in the active site did not lead to a substantial change in the overall fold. There was, however, a major difference in electrostatic interactions between the mutation site and the active site nucleophile. The naturally occurring

Table 1 Specific activity of BglM-G1 and H75R towards aryl-glycosides		
Substrate	Specific activity (U/mg)	
	BglM-G1	H75R
<i>p</i> -nitrophenyl- β -D-glucopyranoside	56.3 \pm 1.7	74.7 \pm 1.3
<i>p</i> -nitrophenyl- β -D-fucopyranoside	36.9 \pm 1.5	80.4 \pm 5.3
<i>p</i> -nitrophenyl- β -D-galactopyranoside	21.5 \pm 0.3	24.1 \pm 3.7
<i>o</i> -nitrophenyl- β -D-glucopyranoside	17.5 \pm 0.7	21.0 \pm 0.6
<i>o</i> -nitrophenyl- β -D-galactopyranoside	8.8 \pm 0.4	9.5 \pm 0.5
<i>p</i> -nitrophenyl- β -D-cellobioside	7.5 \pm 0.1	9.2 \pm 0.4
<i>p</i> -nitrophenyl- β -D-xylopyranoside	4.8 \pm 0.1	6.6 \pm 0.5
<i>p</i> -nitrophenyl- β -D-mannopyranoside	—	0.2 \pm 0.1
<i>p</i> -nitrophenyl- α -D-arabinofuranoside	—	—
<i>p</i> -nitrophenyl- α -D-rhamnopyranoside	—	—

The specific activities were determined at 45 °C in sodium citrate buffer at pH 6.0

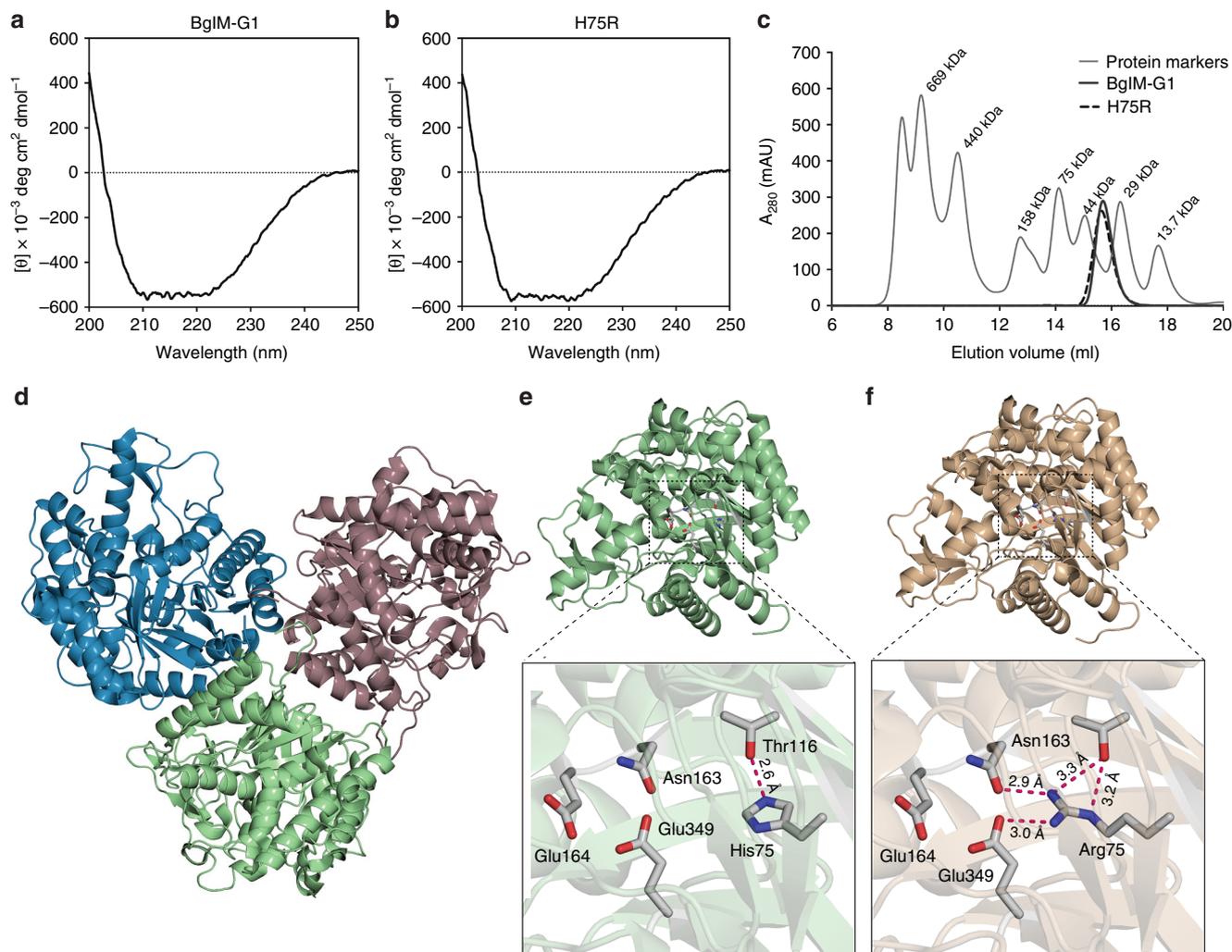


Fig. 2 CD spectroscopic analysis and X-ray crystal structure of BglM-G1 and its H75R variant. Far-UV CD spectra of BglM-G1 (a) and H75R (b) at 20 °C demonstrated overall similarity of the protein structures. Size-exclusion chromatography indicates a monomeric nature of BglM-G1 and H75R (c). However, the asymmetric unit of BglM-G1's crystal structure shows a trimeric assembly (d). Examination of the active site of BglM-G1 (e) and H75R (f) reveals a modified electrostatic interaction network around His75 and Arg75. Electrostatic interactions emanating from residue 75 below 3.5 Å were considered

Table 2 Data collection and refinement statistics for structural datasets

	BglM-G1 (PDB ID 5NS6)	H75R (PDB ID 5NS7)	H75R co-crystallized with 1-deoxynojirimycin (PDB ID 5NS8)
<i>Data collection</i>			
Space group	P43	P43	P43
<i>Cell dimensions</i>			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	138.02, 138.02, 97.91	137.92, 137.92, 98.04	137.75, 137.75, 97.80
α , β , γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution (Å)	48.80-1.50 (1.54-1.50) ^a	48.76-1.54 (1.58-1.54)	48.70-1.55 (1.59-1.55)
CC1/2	99.9 (29.9)	99.8 (31.5)	99.9 (39.4)
<i>R</i> _{sym} (%)	14.0 (238.2)	10.4 (182.1)	15.4 (233.4)
<i>I</i> / σ <i>I</i>	13.31 (1.05)	11.28 (0.89)	11.97 (0.80)
Completeness (%)	100.0 (100.0)	99.9 (99.7)	100.0 (100.0)
Redundancy	13.6 (13.7)	6.8 (6.2)	13.7 (13.0)
Wavelength (Å)	1.00	1.00	1.03
<i>Refinement</i>			
Resolution (Å)	48.80-1.50 (1.54-1.50) ^a	48.76-1.54 (1.58-1.54)	48.70-1.55 (1.59-1.55)
No. of reflections used	292,649	270,326	263,952
<i>R</i> _{work} / <i>R</i> _{free} (%)	14.8/17.2	15.5/17.2	16.1/17.2
<i>No. of atoms</i>			
Protein (non-H)	11845	11625	11612
Ligand/ion (non-H)	193	142	140
Water	859	756	747
<i>B-factors</i>			
Protein	22.7	26.25	24.14
Ligand/ion	56.16	62.41	43.66
Water	31.43	32.26	29.14
<i>Root mean square deviations</i>			
Bond lengths (Å)	0.006	0.009	0.005
Bond angles (°)	0.833	0.955	0.782

^aValues in parentheses represent the highest resolution shell

histidine at position 75 in BglM-G1 formed only a single hydrogen bond with the backbone oxygen of threonine 116 (Fig. 2e), while arginine at position 75 formed, in addition to hydrogen bond interactions with residues Gln163 and Thr116, a salt bridge with the catalytic nucleophile glutamate at position 349 (Fig. 2f)

BglM-G1 is less thermostable due to loss of salt bridge. We hypothesized that this loss of electrostatic interactions should destabilize the structure. Thus, we used far-ultraviolet (UV) CD spectroscopy to investigate temperature-induced structural transitions in both wild-type BglM-G1 and its H75R variant. Both proteins showed highly similar CD spectra with minima between 210 and 220 nm, characteristic for proteins with a significant amount of α -helices²¹. We observed an upward shift in these minima at 50 °C for wild type (Fig. 3a) and 55 °C in the H75R variant (Fig. 3b). The loss in secondary structure in wild type at temperatures lower than in the H75R variant pointed at a lower thermal stability of the former. The similarity in the far-UV CD spectra of the unfolded species also showed a match of the molten-globule state of both wild type and its variant. We then determined the melting temperature of both proteins by monitoring changes in ellipticity at 222 nm, while continuously increasing the temperature from 10 to 80 °C. For BglM-G1 the mid-point of transition was at 51.4 °C (324.6 K with a 95% confidence interval between 324.4 and 324.8 K), while the melting temperature of H75R was at 54.9 °C (328.1 K with a 95% confidence interval between 328.9 and 329.2 K) (Fig. 3c). This difference of almost 3.5 K corresponds to a difference in folding energy of ΔG 4.24 kJ mol⁻¹ at the melting point. This rather high increase in melting temperature in H75R reflects the structural

stability provided by the extensive electrostatic interaction network around residue 75 in β -glucosidases. The differences in melting temperature were also reflected in temperature stability of the enzyme. The variant H75R retained its activity at 45 °C over 40 min, while wild type already lost significant activity (*p* value = 3×10^{-5}). After 5 min at 50 °C, less than 10% activity was measured in the wild type, while H75R retained almost 50% of its initial activity under these conditions (*p* value = 4×10^{-6} , Fig. 3d, e).

Loss of the salt bridge decreases *K_m* and catalytic efficiency. Because the crystal structure showed that arginine at position 75 forms a strong electrostatic interaction with the active nucleophile glutamate 349, this should lead to a more rigid substrate binding site in the variant. To define the substrate binding site, we co-crystallized the H75R variant with the glucosidase inhibitor 1-deoxynojirimycin²², solved the structure with a resolution of 1.55 Å (Table 2), and identified residues that have polar contacts with this sugar derivative. Asparagine 163 and Glutamate 349 were the two residues that had electrostatic interactions with both the inhibitor and Arginine 75 (Supplementary Fig. 2). We thus inspected the B-factor of the atoms in the side chain of these two amino acids. A higher B-factor indicates higher flexibility in the crystal²³. Indeed, the H75R mutant showed lower normalized B-factor values in these residues, suggesting a more rigid substrate binding site in this mutant (Table 3).

We tested if such rigidity would lead to changes in the capability to bind substrate. We thus determined the kinetic parameters of both wild type and the H75R variant toward the two most preferred substrates at 45 °C. BglM-G1 had a *k_{cat}* value of 55.6 ± 1.5 s⁻¹ and a *K_m* of 135 ± 16 μ M toward the pNPGlu

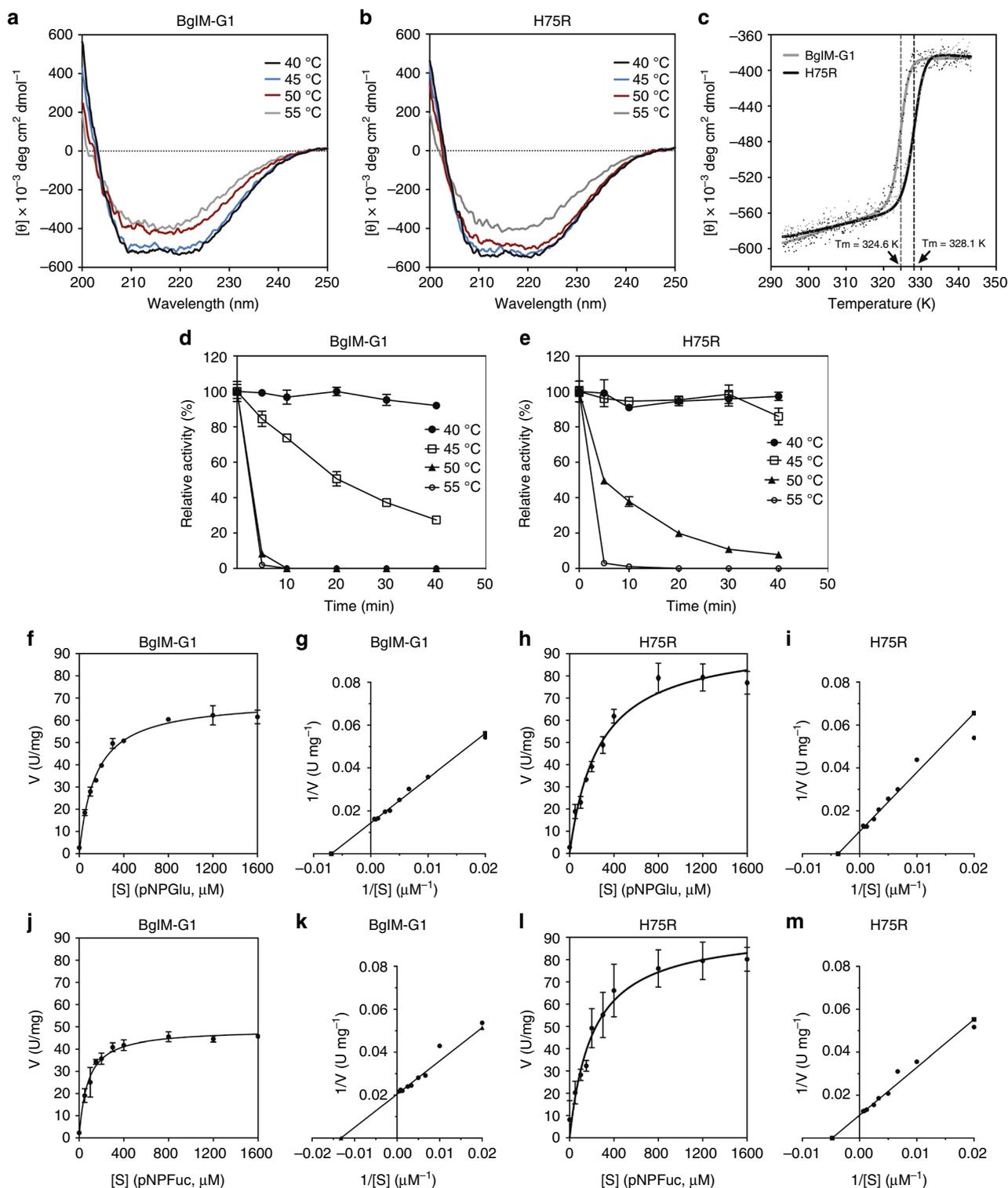


Fig. 3 Temperature stability and kinetic parameters of BglM-G1 and its H75R variant. Analysis of the secondary structure of BglM-G1 (**a**) and H75R (**b**) by CD spectroscopy at various temperatures. For BglM-G1 an upward shift in the spectrum, indicative of denaturation of protein, was observed at 50 °C while for H75R a similar spectral transition was observed at 55 °C. Melting curves of the proteins (**c**). The melting temperature (T_m) of H75R was almost 3.5 K higher than that of BglM-G1. Thermal stability of BglM-G1 (**d**) and H75R (**e**). Determination of kinetic parameters K_m and V_{max} of BglM-G1 and H75R at the optimum temperature of 45 °C toward their most preferred substrates and corresponding Lineweaver-Burk plots, pNPGlu (**f-i**) and pNPFuc (**j-m**). Data based on at least three technical replicates except (**a-c**) where duplicates were performed. Error bars represent standard deviation, where shown.

a, b Exemplary figures shown

Table 3 Average B-factor of the atoms in the side chains of residues Asparagine 163 and Glutamate 349

Protein	Average normalized B-factor Asn163	Average normalized B-factor Glu349	Average normalized B-factor Asn163 and Glu349 combined
BglM-G1	0.93	0.89	0.91
BglM-G1 H75R	0.86	0.85	0.85
BglM-G1 H75R with 1-deoxynorijimycin	0.79	0.82	0.81

The average B-factor of the atoms is normalized to the average B-factor of all atoms in the chain

Table 4 Kinetic parameters of BglM-G1 and H75R

Substrate	Enzyme	Temperature (°C)	k_{cat} (s ⁻¹)	K_m (μM)	k_{cat}/K_m (s ⁻¹ mM ⁻¹)
pNPGlu	BglM-G1	45	55.6 ± 1.5	135 ± 16	417 ± 44
		10	14.9 ± 0.2	72 ± 13	211 ± 38
	H75R	45	99.4 ± 16.3	414 ± 83	242 ± 19
pNPFuc	BglM-G1	10	14.2 ± 0.4	71 ± 3	199 ± 4
		45	41.4 ± 3.0	88 ± 38	519 ± 146
	10	14.4 ± 0.5	90 ± 11	160 ± 16	
	H75R	45	72.9 ± 8.1	222 ± 51	335 ± 38
		10	17.5 ± 0.6	154 ± 4	113 ± 2

Kinetic parameters were determined using *p*-nitrophenyl β-D-glucopyranoside (pNPGlu) and *p*-nitrophenyl β-D-fucopyranoside (pNPFuc) as substrates at 45 and 10 °C. Averages and standard deviations are based on at least three technical replicates

substrate. While k_{cat} of the H75R variant was higher at $99.4 \pm 16.3 \text{ s}^{-1}$, its K_m was $414 \pm 83 \text{ μM}$ (Fig. 3f–i, Table 4). These differences are significant with *p*-values of 0.0042 (k_{cat}) and 0.0014 (K_m). This decrease in the substrate affinity lead to an overall lower catalytic efficiency k_{cat}/K_m for the variant of $242 \pm 19 \text{ s}^{-1} \text{ mM}^{-1}$ versus $417 \pm 44 \text{ s}^{-1} \text{ mM}^{-1}$ for BglM-G1 (*p* value = 0.0002). A similar trend in the differences in the kinetic parameters between the enzymes was observed when pNPFuc was used as substrate (Fig. 3j–m, Table 4) (*p* values are of 4×10^{-5} (k_{cat}) and 0.0014 (K_m) and 0.0261 (catalytic efficiency). At 10 °C, the differences in k_{cat}/K_m became less pronounced, but were still significant in the case of pNPFuc as substrate (*p* value = 0.007) (Table 4). At 45 °C, the optimal temperature for the wild-type enzyme, the substitution of Arginine at position 75 in BglM-G1, while decreasing thermal stability, thus increased catalytic efficiency more than 70% toward pNPGlu and more than 50% toward pNPFuc, by enhancing substrate binding. Given the fact that BglM-G1 was isolated from a temperate, seasonally cold environment, the trade-off of lower thermal stability versus higher catalytic efficiency could probably prove overall beneficial.

Discussion

In this work, we examined the effects of a naturally occurring loss of a highly conserved network of electrostatic interactions close to the active site of the metagenomic β-glucosidase BglM-G1. This network is emanating from an Arginine, which was previously considered invariant. We identified the lack of the network by obtaining high-resolution X-ray structural data of wild type BglM-G1, in which this Arginine in question has been replaced through evolution with a Histidine. We then compared it to an H75R variant of BglM-G1, in which we re-introduced the Arginine. This revealed an extensive hydrogen bond network including a salt bridge between Arginine 75 and the active site Glutamate 349.

This hydrogen bond network, which is missing in BglM-G1 can be also identified in other structures of β-glucosidases. In fact, a comparison of the published structures of all homologs with at least 30% primary sequence identity to BglM-G1 (39 in total)

revealed the salt bridge between Arginine 75 and Glutamate 349 to be present in almost all of these proteins (Supplementary Data 1), except in Strictosidine-O-β-D-glucosidase from *Rauvolfia serpentina* (Serpentine wood) and the psychrophilic β-glucosidase BglU from *Micrococcus antarcticus*, a cold-adapted protein.

The formation of a salt bridge could potentially lead to changes in the ionization of the catalytic nucleophile and thus influence the overall pH optimum of the enzyme. However, a mutational study using the β-glucosidase from *Spodoptera frugiperda*, substituting the Arginine corresponding to Arginine 75 with a Methionine was shown to increase the pH optimum of the enzyme only slightly from 6.2 to 6.5²⁴, in line with our observations.

Nevertheless, as expected, this loss of electrostatic interactions led to a substantial loss of structural stability. This manifested itself in a decreased stability and melting temperature of the wild type when compared to the H75R variant. The observed change in melting temperature of 3.5 K is considerably higher than the increase of 1–2 K that is typically achieved by stabilizing mutations in proteins²⁵. Although substantially higher increases in the melting temperature have been achieved, these are often the direct effect of hydrophobic interactions or the removal of unfavorable hydrophobic residues on the solvent-exposed surface of proteins^{26–28}. While salt bridges certainly stabilize hyperthermophilic proteins²⁸, it has been suggested that the stability of mesophilic proteins often does not profit from salt bridges, as the energy required for desolvation is not larger than the energy provided by the electrostatic interactions²⁹. Consequently, the replacement of amino acids involved in salt bridges by amino acids that engage in hydrophobic interactions led to substantial stabilization of the Arc repressor³⁰. However, other theoretical and experimental work shows that while salt bridges can be both stabilizing and destabilizing, the majority seems to have a stabilizing effect^{5,31}, in line with our observations.

Importantly, this loss in thermal stability coincided with an increased catalytic efficiency of BglM-G1, especially at 45 °C, potentially providing a selective advantage. This is in contrast to what would typically be expected from the substitution of a highly conserved amino acid. For example, the aforementioned

substitution of the Arginine corresponding to Arginine 75 with a Methionine in the β -glucosidase from *S. frugiperda* resulted in a massive 500-fold decrease in the turnover of the substrate²⁴. Depending on the temperature of their environment, enzymes evolve to achieve the optimal balance between the structural rigidity necessary for stability and the flexibility important for substrate binding and enzyme activity^{32–34}. The increased ligand binding affinity in the wild-type protein would increase the catalytic efficiency by optimizing substrate binding in an environment where the amount of substrate is scarce. In some cold-adapted enzymes secreted from marine microorganisms, a low K_m has already been implied with the need to acquire more substrate in low substrate concentration environments (see Siddiqui and Cavicchioli³⁵ for a comprehensive review). Furthermore, it has been suggested that amino acid changes that destabilize protein structure are involved in the adaptation of enzymes to lower temperatures³⁶. The lower B-factors of residues Asparagine 163 and Glutamate 349 in the substrate binding site of BglM-G1's H75R variant indeed suggest that the higher K_m could be the result of a lower conformational flexibility caused by the stabilizing effect of the electrostatic network formed by Arginine 75 (Table 3). Although atoms in the active sites generally tend to be in low B-factor regions of a protein³⁷, it is noteworthy that the average normalized B-factor of these two residues in BglM-G1 is also lower in the majority of its homologs (Supplementary Data 2).

Although the temperature optimum of BglM-G1 classifies it as a mesophilic enzyme, the DNA encoding it was isolated from Chesapeake Bay, Maryland, USA, a seasonally cold marine environment, at only 1°C¹⁵. The trade-off in stability versus higher catalytic efficiency is thus probably advantageous to the organism carrying BglM-G1. It should be noted, however, that the gain in catalytic efficiency we observed at 45°C becomes less pronounced at lower temperatures, closer to the temperatures of the water where the gene encoding BglM-G1 was isolated. However, due to seasonal changes, environmental microbes are typically exposed to greatly varying temperatures. In fact, average maximum surface temperatures in shallow tributaries to Chesapeake Bay can be as high as 30°C and temperatures above 32°C have been measured in the vicinity of power plants³⁸. Thus, the fitness benefit provided by the substitution of Arginine 75 will be greater at higher temperatures.

Our study implies that while the naturally occurring mutation of the otherwise invariant Arginine 75 in this enzyme class is decreasing thermal stability, it can also lead to higher catalytic efficiency. This evolutionary trade-off is potentially beneficial in the temperate, seasonally cold marine environment where BglM-G1 was discovered.

Methods

Bioinformatic sequence analysis. The sequence of hypothetical protein GOS_608286, which we termed BglM-G1, was obtained from the GOS project^{15,39}. The J Craig Venter Institute derived the protein sequence (GenBank accession EDH42461.1) from the GOS marine metagenome scaffold JCVI_S-CAF_1096626901033 (GenBank accession EP597324.1). The DNA for this scaffold was isolated as part of the GOS project from a sample obtained on December 18, 2003 in Chesapeake Bay, MD, at a depth of 13.2 m and a water temperature of 1°C¹⁵.

BlastP analysis of the primary sequence of BglM-G1 was carried out against the NCBI nr and UniProtKB databases. We selected representative sequences with more than 30% amino acid sequence identity to BglM-G1 from UniProtKB from a diverse set of genera for further analysis. A multiple sequence alignment was created using MAFFT (version 7)⁴⁰ and the results visualized using the program ESPrpt 3.0⁴¹. Furthermore, the degree of conservation of each amino acid in BglM-G1 was evaluated using the ConSurf server (<http://consurf.tau.ac.il/2016/>)¹⁷. To analyze the structural homology of BglM-G1 and the electrostatic interaction network emanating from Arginine 75 in these proteins, BglM-G1 was blasted against the RCSB PDB database. We then selected one structure for any protein

with more than 30% amino acid sequence identity, preferably structures with high resolution and without ligand or inhibitor covalently bound to the active site.

Cloning of BglM-G1 into an expression vector. For the expression of BglM-G1 in *E. coli*, codon optimization of the gene sequence encoding BglM-G1 was carried out using the online tool JCat⁴². The codon-optimized gene sequence was synthesized (GeneArt[®], Regensburg, Germany) with additional 5' NdeI and 3' EcoRI restriction sites. Digested *bglM-G1* was then ligated into pOE, linearized with the same enzymes, for overexpression and purification via His-tag. pOE is a pET-11a (Merck Millipore, Billerica, MA, USA) derivative recently developed in our lab for this purpose⁴³. pOE contains an N-terminal 6× His-tag followed by a TEV-protease specific cleavage site. This added the additional N-terminal amino acid sequence MHHHHHHENLYFQGH to BglM-G1 during expression. After TEV cleavage, the scar peptide GH remained. The resultant plasmid pOEBglM-G1 was transformed into *E. coli* BL21 (DE3) for protein expression.

Site directed mutagenesis. To construct the H75R variant of BglM-G1, site directed mutagenesis of pOEBglM-G1 was carried out at position 224 to change the codon CAC to CGC, introducing the corresponding H75R substitution. The oligonucleotides used were: H75R_for: CCAAGCGATAGAGAAGCGGTAAGCACC AACACC and H75R_rev: GGTGTTGGTGCTTACCCTTCTCTATCGCTTGG. The mutagenic PCR reaction was carried out in a 50 μ l reaction mixture containing the following components: 400 μ M dNTPs, 1× *Pfu* polymerase buffer, 0.2 μ M of each primer (H75R_for and H75R_rev), 100 ng pOEBglM-G1, and 2.5 U *Pfu* polymerase. PCR conditions were: 1 min at 95°C to denature the template, followed by 12 cycles of 30 s at 94°C, 30 s at 55°C, and 7 min at 68°C. The PCR reaction products were treated with 0.1 U DpnI for 3 h at 37°C to digest methylated parental DNA, leaving nonmethylated, amplified DNA containing the desired mutation. Competent *E. coli* XL1-Blue cells were transformed with 1 μ l of the DpnI-treated PCR product and grown overnight at 37°C yielding pOEH75R. Successful mutagenesis was confirmed by DNA sequencing (Eurofins Genomics, Ebersberg, Germany).

Purification of recombinant proteins. For purification of recombinant proteins, *E. coli* BL21(DE3) harboring pOEBglM-G1 and pOEH75R were grown in 1 to 4.5 L of LB media supplemented with ampicillin (100 μ g/mL) at 37°C until cultures reached an optical density OD₆₀₀ of 0.6 \pm 0.2. Gene expression was induced at this point by the addition of 1 mM isopropyl-1-thio- β -D-galactopyranoside (IPTG). The culture was then incubated at 20°C for 16 h. Subsequently, cells were harvested by centrifugation at 4°C and 7800×g. The resulting cell pellet was resuspended in buffer A (50 mM sodium phosphate buffer, pH 8.0 with 300 mM NaCl), one dissolved pill of an EDTA-free protease inhibitor cocktail (Roche, Mannheim, Germany) and 1 mM phenylmethanesulfonyl fluoride. Cells were homogenized by high-pressure cell disruption (Constant systems, Daventry, UK) at 1.9 kbar. Cell debris was removed by centrifugation at 50,000×g for 1 h, at 4°C and the supernatant was filtered through a vacuum filter (Filtropur, Sarstedt, Germany). Protein purification was carried out using an ÄKTApurifier FPLC system (GE Healthcare): the filtrate was loaded on a 5 ml Ni-NTA agarose column (GE Healthcare) pre-equilibrated with buffer A. The column was washed with up to 30-column volumes of buffer A, followed by up to three column volumes with 2% buffer B (50 mM sodium phosphate buffer, pH 8.0 with 300 mM NaCl and 500 mM imidazole) until no further decrease in UV-absorption could be observed. The protein was eluted with a 30-column volume linear gradient running from 2 to 100% buffer B (10 mM to 500 mM imidazole) at a flow rate of 3.0 mL min⁻¹. Fractions were analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and those with highest target protein yield were pooled. To remove the N-terminal 6× His-tag, the pooled active fractions were incubated with 6× His-tagged TEV protease (30-fold molar excess of target protein to TEV protease), while dialyzing overnight at 4°C against buffer A with 1 mM EDTA. The protein solution was transferred to a Falcon tube, MgCl₂ was added (final concentration 2 mM) and the solution was centrifuged at 6000×g for 20 min at 4°C. To remove the cleaved 6× His-tag and the 6× His-tagged TEV protease, the protein solution was loaded on a Ni-NTA column. SDS-PAGE analysis of the flow-through revealed a single band with a molecular weight of ~50 kDa, which is consistent with BglM-G1's theoretical mass of 49,874 Da. Fractions of pure protein were collected as flow-through, the buffer exchanged to 50 mM sodium phosphate buffer, pH 7.0 and the protein was concentrated by ultrafiltration using Amicon[®] Ultra tubes (cutoff 10,000, MWCO, Millipore) and aliquots of concentrated pure protein were stored at -80°C. Over the course of our experiments we did not observe any loss of activity under these storage conditions. The purity of BglM-G1 and H75R was verified by SDS-PAGE and protein concentration was determined by UV spectroscopy determining the absorbance at 280 nm in 50 mM sodium phosphate buffer (pH 7.0). The extinction coefficient ($\epsilon_{280\text{ nm}}$) and molecular weight of the purified recombinant protein was calculated based on amino acid composition using the web-based tool ProtParam at ExPASy (<http://web.expasy.org/protparam/>)⁴⁴.

Size-exclusion chromatography. Analytical size exclusion chromatography was carried out with the ÄKTApurifier FPLC-system running UNICORN 5.31 software using a Superdex 200 Increase 10/300 GL column (GE Healthcare). The column

was pre-equilibrated with buffer solution (50 mM sodium phosphate, pH 7.0; 150 mM NaCl). The protein solution was loaded onto the column and elution was carried out at a flow rate of 0.4 mL min⁻¹. The eluate was analyzed for protein content by online measurement of the absorption at 280 nm. To determine the molecular weight of BglM-G1 and H75R, a mixture of standard calibration protein markers was prepared and 100 µL of the protein mix was loaded on the column. The calibration proteins used were as follows: Thyroglobulin (669.0 kDa), Ferritin (440.0 kDa), Aldolase (158.0 kDa), Conalbumin (75.0 kDa), Ovalbumin (44.0 kDa), Carbonic anhydrase (29.0 kDa), Ribonuclease A (13.7 kDa). The exclusion volume was determined by using Blue Dextran (200,000 kDa).

Plate-based analysis of β-glucosidase activity. Functional determination of β-glucosidase activity was performed using LB agar plates with 0.1 % esculin hydrate, 0.2 % ferric ammonium citrate, 100 µg mL⁻¹ ampicillin and 50 µM IPTG⁴⁵. Clones expressing β-glucosidase activity were identified by observing the plates for appearance of dark halos surrounding the colonies after 24 h of incubation at 37 °C.

Enzymatic characterization. β-glucosidase activity was determined using pNPGlu as substrate. At least three technical replicates were performed for these experiments. The assay mixture (1 mL) consisting of 2 mM pNPGlu in 50 mM sodium citrate buffer (pH 6.0) was incubated with the enzyme at 45 °C for 10 min. The reaction was stopped by adding 1 mL of 1 M Na₂CO₃ and the amount of *p*-nitrophenol released was determined by measuring the absorbance at 400 nm in a JASCO V-650 UV-VIS spectrophotometer running Spectra Manager Version 2 (Jasco, Tokyo, Japan). One unit of enzyme activity was defined as the amount of enzyme releasing 1 µmol of pNP per minute under the above-mentioned assay conditions. Specific activity was expressed in units of active enzyme per milligram of protein.

Optimum pH, optimum temperature, and temperature stability. The influence of the pH on the activity of BglM-G1 and its H75R variant was determined by measuring the enzyme activity in 100 mM buffers of various pH values (range: 3.0–10.0) with 1 mM pNPGlu as a substrate at 40 °C. At least three technical replicates were performed for each experiment. Buffers used were: 50 mM sodium citrate buffer (pH 3.0–6.0), 50 mM acetate buffer (pH 4.0–5.5), 50 mM MES buffer (pH 5.5–6.5), 50 mM sodium phosphate buffer (pH 6.0–8.0). The pH stability of BglM-G1 was determined by pre-incubation of the enzyme in buffers of varying pHs for 1 h and determining the residual enzyme activity using 100 mM sodium citrate buffer (pH 6.0) at 45 °C. The buffers used were: 50 mM sodium citrate buffer (pHs, 3.0, 4.0, 4.5, 5.0, 5.5, and 6.0), 50 mM acetate buffer (pHs, 4.0, 4.5, 5.0, and 5.5), 50 mM MES buffer (pHs, 5.5, 6.0, and 6.5), and 50 mM sodium phosphate buffer (pHs, 6.0, 6.5, 7.0, 7.5, and 8.0). The enzyme activity without pre-incubation in buffers was considered 100%. The optimum temperature for enzyme activity of BglM-G1 and H75R was determined by carrying out assays at temperatures over a range of 10–60 °C with 2.0 mM pNPGlu in 100 mM sodium citrate buffer (pH 6.0), as this was determined to be the optimal pH buffer. The thermostability of proteins was assessed by incubating enzymes without substrate at various temperatures (10–60 °C) for various time intervals in this buffer and the residual enzyme activity was then determined at 40 °C.

Substrate specificity. The substrate specificities of BglM-G1 and H75R with various aryl-glycosides were evaluated by incubating 2.0–207.0 pmol of enzymes in a 1 mL reaction mixture with 2 mM substrate in 50 mM sodium citrate, pH 6.0 at 45 °C. Substrate conversion was determined at various time intervals spectrophotometrically by measuring the release of *p*-nitrophenol. In all assays, an appropriate substrate blank was used to account for spontaneous hydrolysis of the substrate. At least three technical replicates were performed.

CD spectroscopy. Far-UV CD spectra of BglM-G1 and H75R were recorded in a JASCO J-815 (Jasco) coupled to a Peltier temperature control system (PTC 514, Jasco). Protein solutions were measured in quartz cuvettes (Hellma Analytics, Germany) with a path length of 1 mm at concentrations of 2.5–6.0 µM in 10 mM sodium phosphate buffer (pH 7.0). Spectra were continuously recorded between 185 and 250 nm in a 1 mm quartz cuvette at a bandwidth of 2 nm. Thermal unfolding was observed from 10 to 80 °C in 5 K intervals, using temperature interval measurement mode with a gradient of 2 °C/min. Measured ellipticity values, obtained in millidegree (θ), were converted to standard units of molar ellipticity [deg cm² dmol⁻¹].

Calculation of the melting point (T_m) and folding energy. For melting point determination, data points were obtained continuously at 222 nm from 10 to 80 °C with a ramp rate of 2 °C/min. Two datasets were measured, and the values obtained from both datasets were combined and analyzed by nonlinear regression (Curve fit) in GraphPad Prism (Version 5.0, GraphPad Software, La Jolla, California, USA) using Eq. (1). We assumed a two-state transition of a monomer from folded to unfolded form with an added data correction for pre- and post-transition linear

changes in ellipticity as a function of temperature⁴⁶.

$$y = a^*(\epsilon F + (CF^*T)) - (\epsilon U + (CU^*T)) + (\epsilon U(CU^*T)) \quad (1)$$

Here, ϵF and ϵU are the mean residual ellipticities of 100% folded and unfolded proteins, respectively, we used the ellipticities -580.7 and -348.0 for BglM-G1 and ellipticities -581.6 and -348.4 for H75R at 25 and 80 °C, respectively, through extrapolation. CF is the linear correction of folded protein as a function of temperature, CU is the linear correction of unfolded protein as a function of temperature and T is the absolute temperature (Kelvin). Eq. (1) depends on the fraction of the folded protein α , which is derived by using Eq. (2). The constant of folding (K) is dependent on free energy of unfolding (ΔG), gas constant (R), and temperature (T) (Eq. (3)). For each data point a value ΔG was derived using Eq. (4). ΔH is van't Hoff enthalpy, ΔCp is heat capacity change upon thermal denaturation, and T_m is the melting temperature.

$$a = K/(1 + K) \quad (2)$$

$$K = \exp^* / (-\Delta G / (RT)) \quad (3)$$

$$\Delta G = \Delta H^* (1 - T/T_m) - \Delta Cp^* ((T_m - T) + T^* \ln (T/T_m)) \quad (4)$$

Crystallization, data collection, and refinement. The proteins obtained after Ni-NTA affinity chromatography were further purified by size-exclusion using a HiLoad 26/60 Superdex 75 pg column (GE Healthcare), running isocratically with 40 mM sodium phosphate buffer (pH 7.0) and 200 mM NaCl. The eluted fractions from the representative protein peak were dialyzed against 20 mM MES, 50 mM NaCl (pH 7.0) and concentrated (final concentrations: 10.6 mg/mL, BglM-G1 and 8.2 mg/mL, H75R) using Vivaspin 20 10'000 MWCO centrifugal concentrators (Sartorius).

The crystals of BglM-G1 and H75R proteins were obtained by a sitting drop vapor diffusion technique in 2:1 mixture of protein solution:crystallization solution. The crystallization solution used was: 2% (v/v) polyethylene glycol 400, 0.1 M sodium HEPES (pH 7.5), 2.0 M ammonium sulfate. Crystals appeared after 72 h of incubation at 18 °C. The crystals formed were transferred to 1 µL of the reservoir solution with 25% glycerol and flash-cooled in liquid nitrogen. The complex of H75R with 1-deoxyojirimycin was obtained by soaking the inhibitor into pre-existing crystals of H75R. For this, crystals of H75R, obtained after 72 h, were picked and transferred to 10 µL of crystallization solution mixed with 25% glycerol and 6 mM 1-Deoxyojirimycin on a glass cover slide. The cover slide was kept inverted on a well of a 24 well plate containing the crystallization solution. The plates were incubated at 18 °C for 24 h and the crystals were picked and flash-cooled in the liquid nitrogen and stored. X-ray diffraction data of the proteins were collected at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) at beamline ID23-1 at a temperature of 100 K and a wavelength of 1.00 Å (5NS6 and 5NS7) and the DESY PETRA III facility (Deutsches Elektronen-Synchrotron, Hamburg, Germany) at beamline P11 at a temperature of 100 K and a wavelength of 1.03 Å (5NS8). The data were processed using XDS and XSCALE⁴⁷. The structure of BglM-G1 was solved by molecular replacement using PHASER⁴⁸, with coordinates of the closest homolog of BglM-G1, TmGH1 (PDB ID 1OD0)²². The structure was refined using phenix.refine⁴⁹ followed by several rounds of manual building using Coot⁵⁰. The structures of H75R and H75R co-crystallized with 1-deoxyojirimycin were solved using the coordinates of the final structure model of BglM-G1. For structures of both proteins, iterative cycles of refinement and manual rebuilding were carried out until the free *R*-factor values converged. Ramachandran statistics were calculated and are as follows (favored/allowed/outliers [%]): BglM-G1 (5NS6) 98.09/1.91/0; H75R (5NS7): 98.17/1.83/0; H75R with 1-deoxyojirimycin (5NS8): 97.79/2.21/0.

The data collection and refinement statistics of the structures are listed in Table 2. The RMSD of structural alignments were determined using the align command in PyMOL. Figures of structures were generated using PyMOL (Schrodinger LLC, New York, NY). Electrostatic interactions were calculated using the distance function of PyMOL with the parameters mode = 2, using a cutoff of 3.5 Å.

Kinetic analysis. The kinetic parameters of BglM-G1 and H75R were determined using *p*-nitrophenyl-β-D-glucopyranoside (pNPGlu) and *p*-nitrophenyl-β-D-fucopyranoside (pNPFuc) substrates. A substrate concentration in the range of 0.5–5.0 times K_m was used. The reaction velocity of enzymes towards substrates was determined in 100 mM sodium citrate, pH 6.0, at 45 °C and 10 °C. The kinetic constants K_m (mM) and k_{cat} (1/s) were determined with a nonlinear regression function assuming Michaelis–Menten kinetics for substrate concentration versus velocity in GraphPad Prism (Version 5.0, GraphPad Software).

Statistical analysis. *p* Values stated were calculated using the *t* test function (two-tailed, two-sample equal variance (homoscedastic)) of Excel for Mac (version 16.16.1 (180814), Microsoft, Redmond, WA). The 95% confidence intervals for the

melting points of BglM-G1 and its H75R variant were derived from the fitting data provided by GraphPad Prism.

Data availability

The protein sequence of BglM-G1 (hypothetical protein GOS_608286) can be found under GenBank accession EDH42461.1. Structural data described in this manuscript is deposited in the protein data bank under PDB ID 5NS6, 5NS7, and 5NS8. There are no restrictions on data availability.

Received: 2 May 2018 Accepted: 13 September 2018

Published online: 17 October 2018

References

- Robinson-Rechavi, M., Alibés, A. & Godzik, A. Contribution of electrostatic interactions, compactness and quaternary structure to protein thermostability: lessons from structural genomics of *Thermotoga maritima*. *J. Mol. Biol.* **356**, 547–557 (2006).
- Feller, G. Protein stability and enzyme activity at extreme biological temperatures. *J. Phys. Condens. Matter* **22**, 323101 (2010).
- Vogt, G. & Argos, P. Protein thermal stability: hydrogen bonds or internal packing? *Fold. Des.* **2**, S40–S46 (1997).
- Tamaki, F. K., Textor, L. C., Polikarpov, I. & Marana, S. R. Sets of covariant residues modulate the activity and thermal stability of GH1 β -glucosidases. *PLoS One* **9**, e96627 (2014).
- Lee, C. W., Wang, H.-J., Hwang, J.-K. & Tseng, C. P. Protein thermal stability enhancement by designing salt bridges: a combined computational and experimental study. *PLoS One* **9**, e112751 (2014).
- Salverda, M. L. M., de Visser, J. A. G. M. & Barlow, M. Natural evolution of TEM-1 β -lactamase: Experimental reconstruction and clinical relevance. *FEMS Microbiol. Rev.* **34**, 1015–1036 (2010).
- Branden, C. & Tooze, J. *Introduction to Protein Structure*. (Taylor & Francis Ltd., 1998).
- Tokuriki, N., Stricher, F., Serrano, L. & Tawfik, D. S. How protein stability and new functions trade off. *PLoS Comput. Biol.* **4**, 35–37 (2008).
- Ketudat Cairns, J. R. & Esen, A. β -Glucosidases. *Cell. Mol. Life Sci.* **67**, 3389–3405 (2010).
- Lynd, L. R., Weimer, P. J., van Zyl, W. H. & Pretorius, I. S. Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol. Mol. Biol. Rev.* **66**, 506–577 (2002).
- K.M., S. B., Bhat. Cellulose degrading enzymes and their potential industrial applications. *Biotechnol. Adv.* **15**, 583–620 (1997).
- Durand, P. et al. Active-site motifs of lysosomal acid hydrolases: Invariant features of clan GH-A glycosyl hydrolases deduced from hydrophobic cluster analysis. *Glycobiology* **7**, 277–284 (1997).
- Teze, D. et al. Semi-rational approach for converting a GH1 β -glucosidase into a β -transglycosidase. *Protein Eng. Des. Sel.* **27**, 13–19 (2014).
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, 490–495 (2014).
- Rusch, D. B. et al. The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* **5**, 0398–0431 (2007).
- Dakhova, O. N., Kurepina, N. E., Zverlov, V. V., Svetlichnyi, V. A. & Velikodvorskaya, G. A. Cloning and expression in *Escherichia coli* of *Thermotoga neapolitana* genes coding for enzymes of carbohydrate substrate degradation. *Biochem. Biophys. Res. Commun.* **194**, 1359–1364 (1993).
- Ashkenazy, H. et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350 (2016).
- Kulkarni, T. S. et al. Crystal structure of β -glucosidase 1A from *Thermotoga neapolitana* and comparison of active site mutants for hydrolysis of flavonoid glucosides. *Proteins Struct. Funct. Bioinforma.* **85**, 872–884 (2017).
- Uchiyama, T., Yaoi, K. & Miyazaki, K. Glucose-tolerant β -glucosidase retrieved from a Kusaya gravity metagenome. *Front. Microbiol.* **6**, 1–9 (2015).
- Henrissat, B. et al. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc. Natl Acad. Sci.* **92**, 7090–7094 (1995).
- Holzwarth, G. & Doty, P. The ultraviolet circular dichroism of polypeptides. *J. Am. Chem. Soc.* **87**, 218–227 (1965).
- Zechel, D. L. et al. Iminosugar glycosidase inhibitors: structural and thermodynamic dissection of the binding of Isofagomine and 1-Deoxyxojirimycin to β -Glucosidases. *J. Am. Chem. Soc.* **125**, 14313–14323 (2003).
- Yuan, Z., Zhao, J. & Wang, Z.-X. Flexibility analysis of enzyme active sites by crystallographic temperature factors. *Protein Eng. Des. Sel.* **16**, 109–114 (2003).
- Marana, S. R., Mendonça, L. M. F., Andrade, E. H. P., Terra, W. R. & Ferreira, C. The role of residues R97 and Y331 in modulating the pH optimum of an insect β -glycosidase of family 1. *Eur. J. Biochem.* **270**, 4866–4875 (2003).
- Kuchner, O. & Arnold, F. H. Directed evolution of enzyme catalysis. *Biofutur* **15**, B1–B11 (1999).
- Xiao, Z. et al. Improvement of the thermostability and activity of a pectate lyase by single amino acid substitutions, using a strategy based on melting temperature-guided sequence alignment. *Appl. Environ. Microbiol.* **74**, 1183–1189 (2008).
- Pakula, A. A. & Sauer, R. T. Reverse hydrophobic effects relieved by amino acid substitution at a protein surface. *Nature* **344**, 363–364 (1990).
- Hecht, M. H., Sturtevant, J. M. & Sauer, R. T. Effect of single amino acid replacements on the thermal stability of the NH₂-terminal domain of phage λ repressor (scanning calorimetry/circular dichroism/proteolysis/temperature-sensitive mutants). *Proc. Natl Acad. Sci. USA* **81**, 5685–5689 (1984).
- Hendsch, Z. S. & Tidor, B. Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci.* **3**, 211–226 (1994).
- Waldburger, C. D., Schildbach, J. F. & Sauer, R. T. Are buried salt bridges important for protein stability and conformational specificity? *Nat. Struct. Biol.* **2**, 122–128 (1995).
- Kumar, S. & Nussinov, R. Salt bridge stability in monomeric proteins 11 Edited by J. M. Thornton. *J. Mol. Biol.* **293**, 1241–1255 (1999).
- Beadle, B. M. & Shoichet, B. K. Structural bases of stability-function tradeoffs in enzymes. *J. Mol. Biol.* **321**, 285–296 (2002).
- Zavodszky, P., Kardos, J., Svingor, A. & Petsko, G. A. Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins. *Proc. Natl Acad. Sci.* **95**, 7406–7411 (1998).
- Fields, P. A. Review: protein function at thermal extremes: balancing stability and flexibility. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **129**, 417–431 (2001).
- Siddiqui, K. S. & Cavicchioli, R. Cold-adapted enzymes. *Annu. Rev. Biochem.* **75**, 403–433 (2006).
- Fields, P. A., Dong, Y., Meng, X. & Somero, G. N. Adaptations of protein structure and function to temperature: there is more than one way to “skin a cat”. *J. Exp. Biol.* **218**, 1801–1811 (2015).
- Zhang, H. et al. On the relation between residue flexibility and local solvent accessibility in proteins. *Proteins* **76**, 617–636 (2009).
- Ding, H. & Elmore, A. J. Spatio-temporal patterns in water surface temperature from Landsat time series data in the Chesapeake Bay, USA. *Remote Sens. Environ.* **168**, 335–348 (2015).
- Yooshep, S. et al. The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. *PLoS Biol.* **5**, e16 (2007).
- Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
- Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDScript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).
- Grote, A. et al. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res.* **33**, W526–W531 (2005).
- Masuch, T. et al. A combined bioinformatics and functional metagenomics approach to discovering lipolytic biocatalysts. *Front. Microbiol.* **6**, 1–11 (2015).
- Wilkins, M. R. et al. Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol.* **112**, 531–552 (1999).
- Eberhart, B., Cross, D. F. & Chase, L. R. β -Glucosidase system of *Neurospora crassa*. I. β -Glucosidase and cellulase activities of mutant and wild-type strains. *J. Bacteriol.* **87**, 761–770 (1964).
- Greenfield, N. J. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat. Protoc.* **1**, 2527–2535 (2006).
- Kabsch, W. Xds. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125–132 (2010).
- McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
- Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221 (2010).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
- Lu, J., Du, L., Wei, Y., Hu, Y. & Huang, R. Expression and characterization of a novel highly glucose-tolerant β -glucosidase from a soil metagenome. *Acta Biochim. Biophys. Sin. (Shanghai)* **45**, 664–673 (2013).

Acknowledgments

X-ray diffraction experiments for BglM-G1 and the H75R variant were performed on beamline ID23-1 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. We are grateful to Alexander Popov at the ESRF for providing assistance in using

the beamline. We also acknowledge DESY (Hamburg, Germany), a member of the Helmholtz Association HGF, for the provision of experimental facilities. Data collection for H75R co-crystallized with 1-Deoxynojirimycin was carried out at PETRA III (beamline P11). We thank Christian Herrmann and the Interfacial Systems Chemistry Department at the Ruhr University Bochum for the provision of the CD spectrophotometer. Principal funding for this work was provided by the European Research Council under the European Union's Seventh Framework Program (FP7/2007–2013)/ERC Grant agreement n° 281384–FuMe to L.I.L. This manuscript was written in part during a Writing Retreat funded by the Ruhr University Bochum Research School RURSplus.

Author contributions

D.M., E.H., and L.I.L. designed the experiments. D.M. performed the experiments. N.L. recombinantly expressed and purified the proteins and performed size exclusion chromatography. E.H. and R.G. collected the X-ray diffraction data. R.G., E.H. and D.M. analyzed the diffraction data and solved the structure. L.I.L. and D.M. performed bioinformatics analyses. L.I.L. and D.M. wrote the paper, E.H. and R.G. consulted on the manuscript.

Additional information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s42003-018-0167-7>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018