# A conserved π-helix plays a key role in thermoadaptation of catalysis in the glycoside hydrolase family 4

Samar Bhallabha Mohapatra[1], Narayanan Manoj[1]*

[1]Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai 600036
*Correspondence: Narayanan Manoj, (nmanoj@iitm.ac.in)

**Abstract**

Here, we characterize the role of a π-helix in the molecular mechanisms underlying thermoadaptation in the glycoside hydrolase family 4 (GH4). The interspersed π-helix present in a subgroup is evolutionarily related to a conserved α-helix in other orthologs by a single residue insertion/deletion event. The insertional residue, Phe407, in a hyperthermophilic α-glucuronidase, makes specific interactions across the inter-subunit interface. In order to establish the sequence-structure-stability implications of the π-helix, the wild-type and the deletion variant (Δ407) were characterized. The variant showed a significant lowering of melting temperature and optimum temperature for the highest activity. Crystal structures of the proteins show a transformation of the π-helix to a continuous α-helix in the variant, identical to that in orthologs lacking this insertion. Thermodynamic parameters were determined from stability curves representing the temperature dependence of unfolding free energy. Though the proteins display maximum stabilities at similar temperatures, a higher melting temperature in the wild-type is achieved by a combination of higher enthalpy and lower heat capacity of unfolding. Comparisons of the structural changes, and the activity and thermodynamic profiles allow us to infer that specific non-covalent interactions, and the existence of residual structure in the unfolded state, are crucial determinants of its thermostability. These features permit the enzyme to balance the preservation of structure at a higher temperature with the thermodynamic stability required for optimum catalysis.

**1. Introduction**

Given that primordial life most likely existed in a hot environment, temperature adaptation is a key evolutionary driver of proteins in extant organisms. The primary mechanism driving molecular evolution is mutations and subsequent fixation of the new variants in the population. The likelihood of fixation is subject to the fitness of these variants under selective forces that may be neutral, deleterious or beneficial. A significant determinant of fitness is the selection at the level of protein structure and stability and the potential trade-off with the functional requirement [1,2]. In general, such evolutionary transitions are gradual, and small changes within a protein can account for functional divergence between orthologous members of a superfamily. Mutational events can either involve substitutions that may induce small changes in local structure and stability, or involve backbone mutations (insertion/deletion) (InDels) that introduce radical shifts in the structure-function properties (macromutations). While single residue InDels are the most frequently occurring backbone mutations among homologs, this

InDel occurring within secondary structure elements can constitute a macromutation that may be crucial to the emergence of novel properties, if the event passes the test of selection [2–5].

The favourable contribution of backbone hydrogen bonding (H-bond) patterns to protein folding and stability is well recognized, and the H-bonds between the carbonyl oxygen and amide nitrogen atoms of residues positioned at i and i+4 in the sequence define the dominant α-helical secondary structure [6–10]. An atypical helix that occurs infrequently is the π-helix, where the H-bonds form between residues positioned at i and i+5 in sequence. Although π-helices often tend to be misannotated or overlooked in structural descriptions of proteins, their prevalence is much greater than previously thought [11–13]. The lengths of π-helical regions identified in structures range from 5 to 18 residues, and these may occur at either terminus or appear interspersed between α-helices or be present independently. In several cases, the interspersed π-helices are restricted to regions of one turn within α-helices and have been designated as α-aneurisms or α-bulges. Studies have shown that such interspersed π-helices in protein structure families are conserved within subgroups of homologs, leading to the hypothesis that the π-helix has evolved by the insertion of a single residue into an existing α-helix [11,14–17]. In several instances, the presence of an interspersed π-helix was shown to be responsible for an adaptive gain-of-function event within the protein family [11,16,18–24].

The GH4 and GH109 families of glycoside hydrolases (GH) constitute an unusual NAD-dependent group among the 167 families that presently constitute the GH superfamily (CAZy database) [25]. The GH4 enzymes require reducing conditions, $NAD^+$ and a divalent metal ion for catalysis. Furthermore, these enzymes show diverse selectivity and specificity for carbohydrate substrates [26]. GH4 homologs are present only in Archaea and Bacteria and remain poorly characterized for their structure-function properties [27–30]. We have previously reported the crystal structure of the GH4 α-glucuronidase from the hyperthermophile *Thermotoga maritima*(TmAgu4B) in complex with $Co^{2+}$ and citrate, an inhibitor, at 1.95 Å resolution (PDB: 6KCX) [31]. A careful examination of the structure revealed a hitherto unrecognized single turn π-helix interspersed within an α-helix. Interestingly, this α-helix constitutes the inter-subunit interface of a conserved homodimeric assembly, with the inserted π-helix residue participating in significant non-bonded interactions across the interface.

In the present study, we carried out a comparative analysis of GH4 structures and noted that the π-helix is present at the same position in three out of eight homologs available in the PDB. We hypothesized that this interspersed π-helix residue (insertion of Phe407 in TmAgu4B), while typically destabilizing within an α-helix, is an evolutionarily conserved

3

protein backbone mutation that may provide a functional gain in the homologs that contain it. A phylogenetic analysis suggests that the π-helix InDel event defines a separate GH4 subfamily. A deletion variant lacking residue Phe407 (TmAgu4BΔ407) was created. The variant showed a significant decrease in the optimum temperature for glucuronidase activity. Crystal structures of the NAD complexes of the wild-type and the variant show that the π-helix transformed into a continuous α-helix while retaining the quaternary structure. By comparing the structural changes, the activity profiles and the thermodynamic profiles of the two proteins, we propose that specific interactions, and residual structure in the unfolded state, are crucial determinants of thermoadaptation of catalysis in the GH4 subfamily. To the best of our knowledge, the direct role of a π-helix as a structural determinant of protein thermostability has not been reported yet.

## 2. Material and methods

### 2.1 Phylogenetic analysis

A total of 15811 protein sequences available in GH4 family were extracted from CAZy database. Redundant sequences were removed with the CD-HIT program using a cut-off value of 98% [32]. The final dataset had 1005 sequences and was aligned using the MAFFT program [33]. Incomplete and fragmented sequences were manually removed. The DaliLite program was used to generate a structure-based multiple sequence alignment of eight available GH4 orthologs [34] and used as a constraint to realign the final dataset using MAFFT [33]. The BLOSUM62 scoring matrix and the option E-INS-I was enabled for better accuracy for a data set. The final alignment, after inspection was used to generate the WebLogo representation using WebLogo3 [35]. The sequence of the *T. maritima* GH4 α-glucuronidase TmAgu4B (UniProt ID: Q9WZL1, PDB: 6KCX), was taken as the reference sequence. The multiple sequence alignment was depicted using ESPript 3.0 [36]. Phylogenetic analysis was performed using Maximum likelihood approach implemented in RAxML8.2.9 [37]. ProtTest was used to determine the evolutionary model and parameters for phylogenetic reconstruction based on the Akaike Information Criterion (minAIC) [38]. Jones-Taylor-Thornton amino acid substitution matrix with gamma distribution (JTT+G) for rate change among sites was identified as the best model. The robustness of the tree topology was tested using bootstrapping with 500 iterations. Bootstrap values were mapped on the initial true tree, which were used to assess the tree topology. Trees were examined using FigTree software version v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

## 2.2 Cloning, protein expression, purification and oligomeric states

The plasmid containing *TM0752* gene, encoding protein TmAgu4B was obtained from the DNASU plasmid repository (Clone ID: TmCD00422573). The expression vector is pMH2T7 containing the arabinose-inducible araBAD promoter and included ampicillin resistance. The deletion mutant was generated by PCR using site-directed mutagenesis method using the primers (5'CGGATAAAGATCTATCTGTGGCCC-3') and (5'-GGGCCACAGATAGATCTTTATCCG-3'). TmAgu4B (UniProt ID: Q9WZL1) codes for 471 amino acids and the recombinant wild type protein contains an additional N-terminal 12 residue expression and purification tag. The sequences of *TmAgu4B* and single residue deletion construct (*TmAgu4BΔ407*) were confirmed by DNA sequencing. Recombinant wild type and mutant were expressed and purified as previously described [31]. Briefly, cells were sonicated in dialysis buffer (50 mM Tris-HClpH 7.5, 20 mM imidazole, 300 mM NaCl) and centrifuged. The supernatant was run on a Ni-Nitrilotriacetic acidsepharose column (GE Healthcare) and washed with buffer (50 mM Tris-HCl, pH 7.5, 300 mM NaCl, 50 mM imidazole). The elution buffer (50 mM Tris-HCl, pH 7.5,300 mM NaCl, 250 mM imidazole) was used to elute sample, followed by desalting in storage buffer (50 mM Tris-HCL, 100 mM NaCl, pH 7.5) using HiPrep$^{TM}$ 26/10 Desalting column. Oligomeric states were calculated using size exclusion chromatography (SEC) at 4 °C using a Superdex$^{TM}$ 200 (10/300 GL) analytical column (GE Healthcare). The column was pre-equilibrated with storage buffer. The loaded protein sample (5 mg ml$^{-1}$ in 200μl) was incubated at room temperature with 100 mM NAD$^+$, 300 mM DTT,100 mM MnCl$_2$and eluted at a flow rate of 0.5 ml min$^{-1}$. Chromatograms were obtained by measuring absorbance at 280 nm.

## 2.3 Steady-state enzyme assay

Synthetic substrate *p*-nitrophenyl-α-D-glucuronic acid (*p*NP-α-GUA), NAD$^+$, MnCl$_2$, and D-glucuronic acid were of analytical grade. Spectrophotometric data were measured in a Perkin Elmer Lambda 25 UV-visible spectrophotometer with a Peltier system. α-Glucuronidase activity was determined by monitoring the continuous release of product *p*-nitrophenol at 405 nm. The assay conditions are same as that previously reported for the TmAgu4B, with minor modification. TmAgu4B had an optimum pH and temperature of 8.0 and 90 $^{\circ}$C, respectively [31]. The molar absorption coefficient for *p*-nitrophenol under assay conditions used is 7200 mM$^{-1}$ cm$^{-1}$. The effect of temperature on α-glucuronidase activity of TmAgu4B and TmAgu4BΔ407 was determined by measuring the specific activity over the temperature range of 30 to 100 °C. The assay conditions include incubating 10 μg of enzyme, 50 mM of HEPES

pH 7.5, 0.2 mM $MnCl_2$, 0.5 mM of $NAD^+$, 30 mM DTT at 40 $^\circ$C for 2 min first. One unit of enzyme activity is defined as the amount of enzyme releasing 1 μmol of product per min. All measurements were performed in triplicates.

## 2.4 Protein crystallization

Initial crystallization trials were carried out by hanging drop vapour diffusion method using protein samples (10 mg $ml^{-1}$) using commercial crystallization screens. 1 μl of protein mixed with 1 μl of reservoir solution, was equilibrated with 0.5 ml of reservoir solution. Optimization was done by varying additives, pH, precipitants, protein concentration. Best crystals for NAD bound form of TmAgu4B wild-type were obtained by mixing 2 μ protein (40 mg/ml) with 7.27 mM of DTT, 7.27 mM of $NAD^+$ against 2 μl of reservoir solution composed of 12-18% PEG 3350, 0.1 M imidazole (pH 5.8-6.2), 2-propanol and 0.2 M trilithium citrate. NAD and $Mn^{2+}$ bound crystal forms of TmAgu4BΔ407 were obtained by mixing 2 μl protein (40 mg/ml) with 7.3 mM of DTT, 7.3 mM of $NAD^+$, 7.3 mM of $MnCl_2$ against 2 μl of reservoir solution containing 12-18% PEG 3350, 0.1 M imidazole (pH 5.8-.2), 2-propanol and 0.2 M trilithium citrate. Crystals of both proteins grew after 7 to 15 days at 20 °C.

## 2.5 Data collection, structure determination and refinement

X-ray diffraction data for TmAgu4B and TmAgu4BΔ407 crystals were collected at a home source on a BRUKER MICROSTAR rotating anode X-ray generator (Cu Kα = 1.5418 Å) with the MAR345 Image Plate. Crystals soaked in a cryoprotectant solution (39 % PEG 3350, 0.2 M trilithium citrate, 0.1 M imidazole) for ~5-10 s were mounted on cryo-loops and flash-frozen in a nitrogen gas stream at 100 K. MOSFLM and Aimless programs, from the CCP4 software package, were used for data processing and scaling [39–41]. Structure solution and refinement protocols were carried out using programs from the PHENIX program suite [42]. The structure was solved by molecular replacement using the PHASER program. One subunit of TmAgu4B (TM0752, PDB: 1VJT, Joint Centre for Structural Genomics, unpublished) was used as the search model. Iterative rounds of restrained maximum likelihood refinement were carried out. Model building was carried out using the Coot program [42–44]. The refinement also included simulated annealing to remove model bias. The stereochemical quality of the final models was evaluated using the MolProbity program [45]. Structure representations were made using Chimera  and PyMol programs [46,47]. All structural alignments were carried out using the

DALI program [34] and SuperPose Version 1.0 [48]. A summary of the data collection, refinement and validation statistics for the models are given in Table 1.

## 2.6 Circular dichroism spectroscopy

Circular dichroism (CD) experiments were carried out on samples in 20 mM sodium phosphate buffer (pH 7.4). Far-UV CD spectra were obtained on a Jasco J-815 spectropolarimeter equipped with a Jasco Peltier temperature controller. Spectra were recorded in the range 200-250 nm in a 1 mm path-length quartz cell containing protein sample (0.22 mg ml$^{-1}$) and a response time of 1s. The unfolding profile of protein was measured at wavelength 222 nm as a function of increasing temperature with a scan rate of 1 °C min$^{-1}$, until the protein was unfolded. The melting temperature $T_M$ (also referred as $T_G$, where the $\Delta G$ becomes zero) was calculated using the Boltzmann sigmoid equation using GraphPad Prism 5.0. Further the thermodynamics of the protein unfolding was investigated by monitoring the CD melting profile at 222 nm as a function of temperature, assuming a two-state unfolding i.e., molecule undergoes unfolding between two folded ($F$) and unfolded ($U$) states. The fraction folded at any temperature is $\alpha$

$$\alpha=(\theta_t-\theta_U)/(\theta_F-\theta_U)......1$$

where $\theta_t$ is the observed molar residue ellipticity at any temperature t, $\theta_F$ is the molar residue ellipticity of the native form. The equilibrium constant k between native and denatured states are calculated using the equation

$$K = \alpha/(1-\alpha)\ .......2$$

The Gibbs free-energy change between the denatured and native states of a protein at any temperature $T$ is given by

$$\Delta G° = -RT\ln K\ ......3$$

where $R$ is the gas constant (1.98 cal mol$^{-1}$), and $T$ is the absolute temperature in Kelvin.

The obtained values of $\Delta G°(T)$ was plotted against temperature (K) to obtain linear plot described by the modified Gibbs-Helmholtz equation

$$\Delta G°(T) = \Delta H°(1-T/T_G)-\Delta C_p°((T_G-T) + T\ln(T/T_G))\ ......4$$

where $T_G$ is the heat-denaturation (melting) temperature, $\Delta H°$ is the change in enthalpy, and $\Delta C_p°$ is the change in heat capacity between native and denatured states of the protein.

The $\Delta G$ vs. temperature (K) plot obtained above was fitted to equation 4 to generate the protein stability curve over range of temperature (30 – 100 °C), and the values of $\Delta H°$ and $\Delta C_p°$ were determined. The $T_M$ values earlier measured by thermal unfolding was included as a single reference point ($\Delta G = 0$). The values of $\Delta H(T)$ in the temperature range (30 – 100 °C) were calculated using the unfolding enthalpy function equation

$$\Delta H(T) = \Delta H° + \Delta C_p°(T\text{-}T_G) \ ......5$$

$\varDelta H(T)$ values were plotted for each temperature and $T_S$ is the temperature at which $\Delta S°$ is zero, which is also the temperature where the $\varDelta G°$ is maximum.

$\Delta S°$ value is obtained from

$$\Delta S° = \Delta H°/ T_G ......6$$

The temperature where enthalpy of unfolding is zero ($T_H$) is calculated from

$$T_{H =} T_S - \Delta G°(T_S)/\varDelta C_p° \ ......7$$

where $T_S$ is the temperature where $\varDelta G$ attains the maximum values ($\Delta G°(T_S)$) [49–51].Curve fitting and data analysis were performed using Microsoft Excel (2007).


## 3. Results

### 3.1 Prevalence of the π-helix in the GH4 family

The family 4 is unusual among the 167 families that make up the GH superfamily since these enzymes hydrolyze glycosidic linkages by a novel mechanism involving oxidation-elimination and addition reactions [52–56]. The Carbohydrate-Active EnZymes database (CAZy, http://www.cazy.org) lists a total of 15811 GH4 sequences, including 40 archaeal and 15770 bacterial enzymes. Although only 27 homologs have been characterized experimentally, these display a wide diversity of substrate specificity, and selectivity at the level of the α- or β-configuration in their substrates. Among the crystal structures of eight orthologs available to date, five have been reported in literature while three from structural genomics efforts remain unpublished (28, 35, 38, 39, 40–45). The previously unreported π-helix in the TmAgu4B α-glucuronidase (PDB: 6KCX, 1VJT) is sandwiched in a long α-helix (α14, residues 402-425) that constitutes the oligomerization domain of the GH4 fold. The secondary structure assignment algorithm DSSP (Definition of Secondary Structure in Proteins) identifies residues 404 to 409 present with an i, i+5 hydrogen-bonding pattern indicating an interspersed π-helix in an otherwise canonical α14-helix (Fig. 1, Fig. 2A, B, Fig. S1).
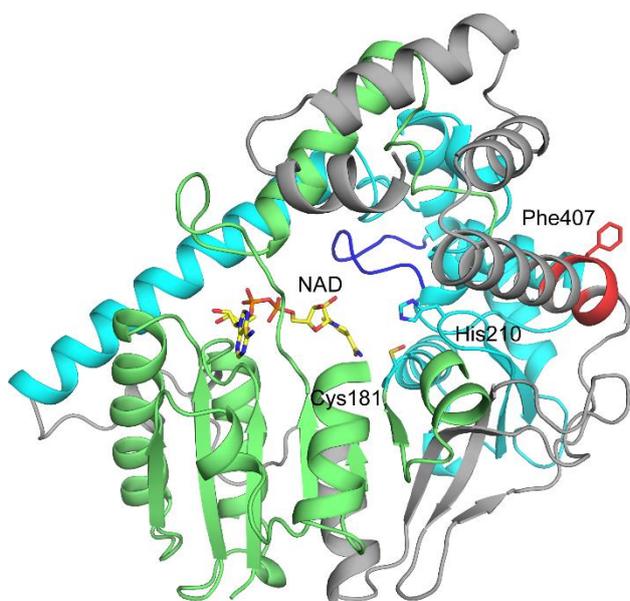
**Figure 1.** Tertiary structure of TmAgu4B (PDB ID: 7BRF). TmAgu4B is shown in ribbon representation showing the three domains, namely, the N-terminal Rossmann fold domain (lime), the central catalytic domain (cyan) and the C-terminal domain (grey). Active site loop comprising residues 238-242 is shown in blue and the catalytic residues, Cys181, His210 are shown as sticks. Bound NAD is shown in yellow stick representation. The interspersed π-helical region (red) with the insertional residue Phe407 (in stick) in the C-terminal oligomerization domain is shown.

A DALI search was carried out against the PDB using TmAgu4B as the query to identify structurally known homologs. The search identified all the expected GH4 homologs, including TmAglA (PDB: 1OBB), TnAgL (PDB: 3U95), GsLicH (PDB: 1S6Y), BsGlvA (PDB: 1U8X), TmBglT (PDB: 1UP6), KpAglB (PDB: 6DVV) and BsLplD (PDB: 3FEF). The homolog pairs had structural alignments with Z-scores in the range 31.2 to 61.8, root mean square deviations (rmsd) in the range 1.2 to 3.2 Å and sequence identities ranging from 18–89 %. The superpositions indicated a highly conserved tertiary structure, including that of the α14-helix containing the embedded π-helix (Fig. S2). Close inspection shows that TmAglA and TnAgL, which share sequence identities of 51 and 89 %, respectively, with TmAgu4B, contain a structurally equivalent π-helix, whereas the other five homologs contained a continuous canonical α-helix (Fig.2, Fig. S3, Fig. S4). Multiple structure alignment indicates that the insertion of a single residue (Phe407 in TmAgu4B) into an existing α-helix accounts for the π-helix geometry (Fig. S5).
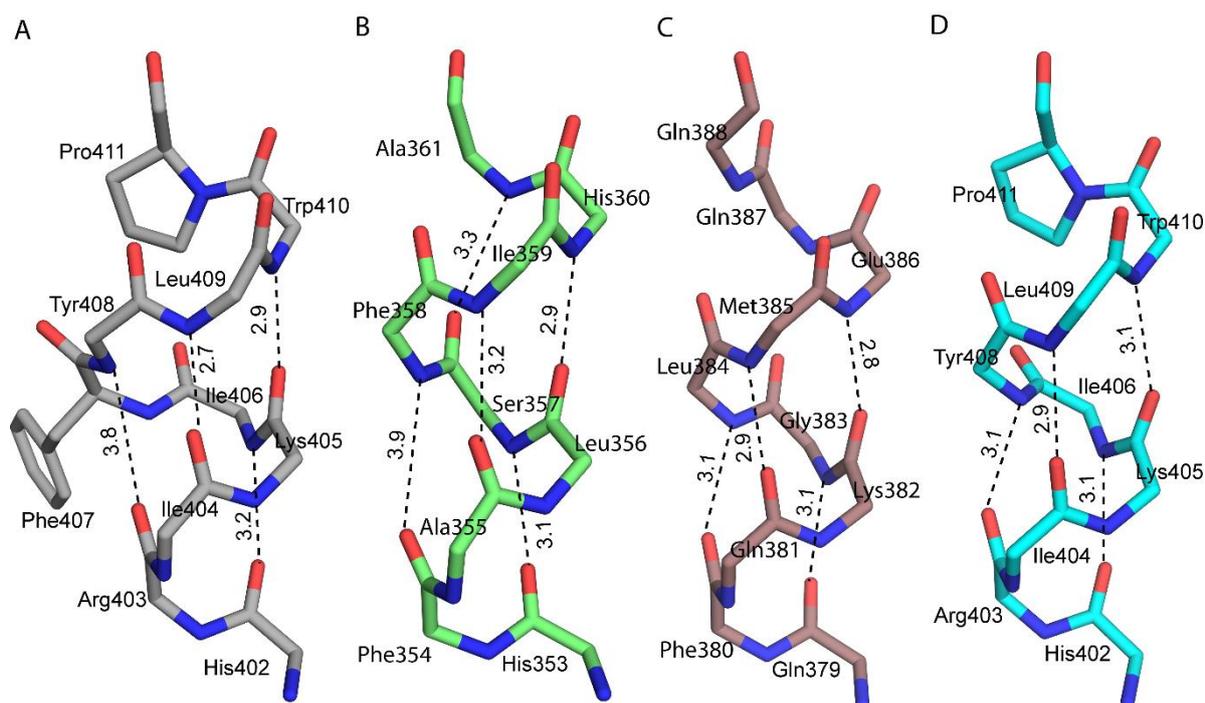
**Figure 2.** Comparison of backbone hydrogen bonding network in the α-14 helix.A) TmAgu4B (PDB ID: 7BRF). The π-type i, i+5 hydrogen bonding pattern is shown between residue pairs. Side chain of insertional residue Phe407 is shown in stick. B) *T. maritima* 6-phospho-β-glycosidase, TmBglT (PDB: 1UP6).C) *B. subtilis*6-phospho-α-glycosidase, BsGlvA (PDB: 1U8X). The α-type i, i+4 hydrogen bonding pattern is shown between equivalent residue pairs. D) Deletion variant, TmAgu4BΔ407 (PDB:7BR4). The α-type i, i+4 hydrogen bonding pattern is shown between equivalent residue pairs. Hydrogen bonds are shown as dashed lines and the values shown in Å.

To know the abundance of this feature within the GH4 family, a comprehensive survey of sequences available in the CAZy database was carried out. A set of 1005 non-redundant sequences were subjected to multiple sequence alignment (MSA) using the structurally-informed MSA as the seed. The alignment showed the expected conservation of all catalytic and cofactor binding residues. The region corresponding to the α14 helix is well aligned and conserved across the diverse set of sequences (Fig. 3, Fig. S5). A subset of 85 GH4 homologs contains the insertion, providing strong evidence that this naturally present π-helix feature is evolutionarily conserved within a subset. The alignment was next subject to phylogenetic analysis using the maximum likelihood method. The resulting unrooted tree resolved into well-supported clades that correspond to several subgroups. The subset of π-helix containing GH4 homologs cluster into a strongly supported clade (named GH4-π hereafter) (Fig. 4). This clade mainly includes sequences from thermophilic/hyperthermophilic organisms from both Archaea and Bacteria. Although the total number of archaeal sequences is fewer compared to the bacterial homologs, all except one archaeal sequence cluster into the GH4-π clade (Table S1).
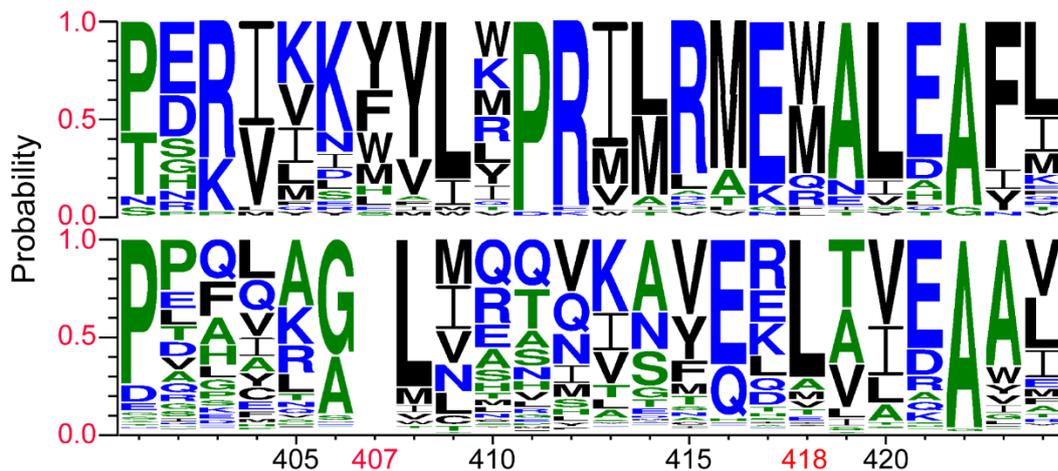
**Figure 3.** WebLogo plot of sequence conservation in α14-helix of the oligomerization domain of the GH4 family. The top and bottom panels represent the structure-based multiple sequence alignment of the π-helix containing orthologs and the α-helix containing orthologs, respectively. The residue numbering corresponds to the conserved α14-helix (residues 401-425) of TmAug4B. The insertional residue at position 407 in the π-helical region (residues 405-411) is mostly a large aromatic or aliphatic residue (Phe407 in TmAug4B). Pro411 marks the C-terminal position where the π-helix shifts back to an α-helix. The charged, polar and nonpolar residues are colored blue, green and black, respectively. Symbol heights within a stack represent relative frequency of each residue. Phe407 interacts with residues 416′-418′ from the two-fold related subunit in the dimeric interface.

The overall pattern of subgroup clades inferred here is largely consistent with a previous phylogenetic analysis of a smaller subset of 201 GH4 sequences [26]. That study was carried out to explore whether phylogenetic analysis and sequence motif annotation could be used to predict function within the GH4 family. A strong association between the substrate specificity and five major subgroups of GH4 homologs was proposed. Furthermore, they described a variable four-residue, active site Cys-containing motif that appeared correlated to the predicted catalytic function specific to each subgroup[26]. In our analysis, we adopt the same motif-based naming scheme with minor variations to label the seven significant subgroups identified (Fig. 4). Considering the strong evolutionary association between the sequences and the presence of the π-helix, it is evident that this feature is an adaptation defining a new GH4 subfamily.
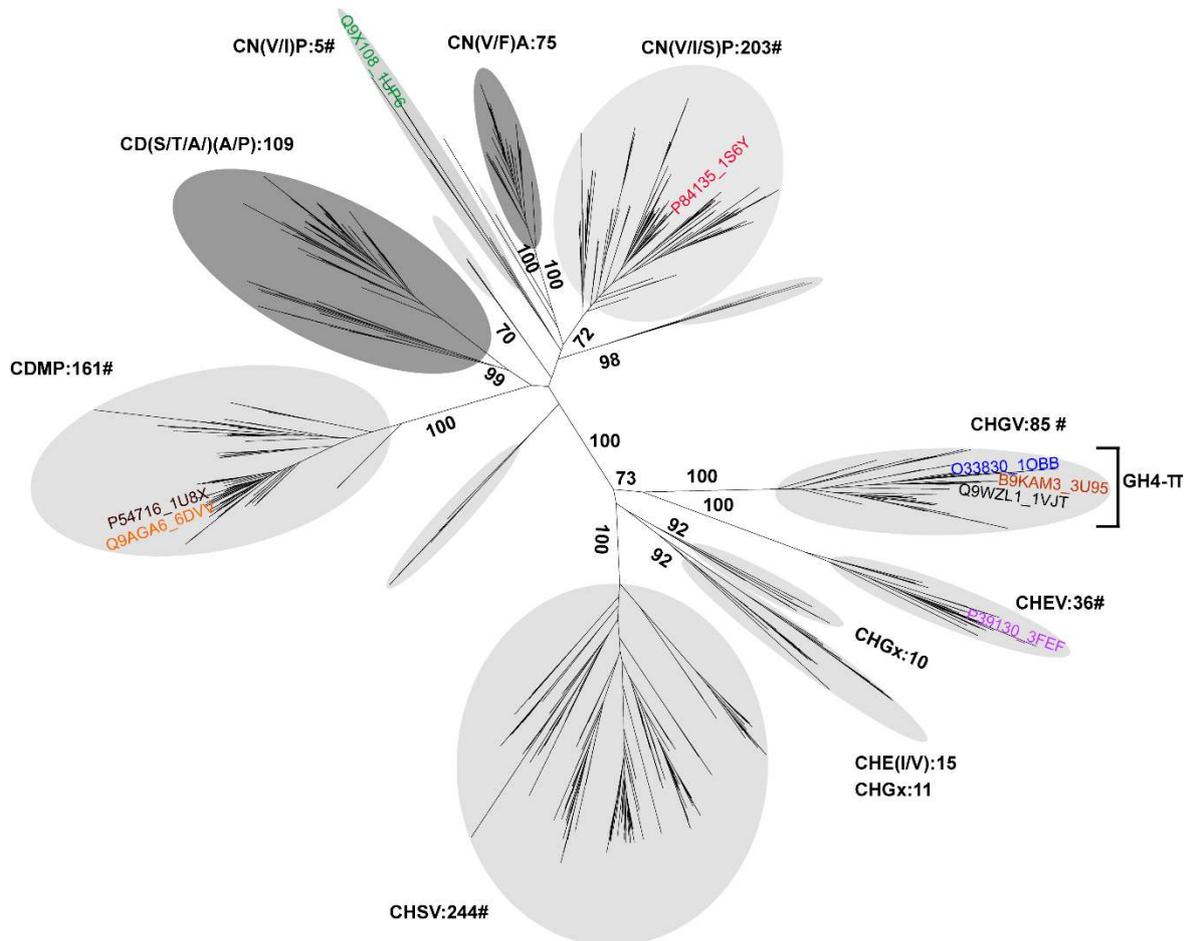
**Figure 4**. Phylogenetic analysis of the GH4 family. Unrooted maximum likelihood phylogenetic tree constructed with RAxML using a structurally-informed multiple sequence alignment of 1005 non-redundant GH4 homologs. The sequences were retrieved from the CAZy database. Major clades are shaded in grey background. The four-letter naming scheme for the major clades is as proposed by Hall et al., 2009, with minor variations. The name corresponds to a variable four-residue, active site Cys-containing motif, that appeared correlated to the predicted catalytic function specific to each clade. The number following the name indicates the total number of sequences within the clade. # symbol represents clades that are consistent with the subgroups defined by Hall et al, 2009.Two new major clades are shaded in darker grey background. Support values (% of bootstrap value) for the major nodes are shown. The GH4-π clade is marked separately. The structurally characterized homologs are labeled in color with Uniprot name and PDB ID.

## 3.2 Activity and thermal stability of TmAgu4B and TmAgu4BΔ407

Experimentally characterized GH4 family members exist in homo-dimeric or tetrameric forms. The tetrameric forms can be considered as formed by the association of two homodimers. The invariant dimeric interface across the GH4 family is dominated by interactions between the α14 helix and the corresponding two-fold symmetry related helix from the adjacent subunit packed in an antiparallel orientation (Fig. S6.A). The π-helix insertion residue common to TmAgu4B, TnAgl and, TmAglA, is completely buried within this interface in a pocket formed

12

by multiple residues from the other subunit, suggesting that this residue may influence the stability of the dimeric association (Fig. 5A, Fig. S7A, Fig. S8A, Fig. S9A, B). Interestingly the three GH4-π homologs are dimers, whereas the other five homologs are tetramers, although the dimer-dimer interface in the tetramers is distant from the intra-dimer interface.

In order to explore the structure-function implications of the π-helix in TmAgu4B, the deletion variant of residue Phe407 (TmAgu4BΔ407) was generated. Both wild-type and mutant proteins were expressed in *E. coli* and purified to yield ~35 mg of protein per lit of culture. The mutation did not affect the folding or the solubility. Both proteins showed the expected subunit molecular weights. Furthermore, the oligomeric state of the proteins examined by size exclusion chromatography (SEC) displayed peaks that correspond to dimeric species (Fig. S10A, B). Next, the optimum temperatures for catalytic activity were measured using α-glucuronidase assays. The temperature optimum for the wild-type is ~90 °C, with a steep drop at 100 °C. In contrast, the variant showed a near-linear increase in specific activity from 30 – 50 °C and the highest activity in a broad 50 – 60 °C range and an almost 'bell-shaped' behaviour of decreasing activity with higher temperature (Fig. 6A). A relative 50 % decrease in activity at the temperature optimum of the variant indicates a significantly deleterious effect on its stability. The secondary structure contents examined by far-UV CD spectra at 25 °C are nearly identical, indicating that the overall secondary structure of the protein was not affected (Fig. S10C).

The possibility of the loss of stability due to the mutation was evaluated from thermal denaturation profiles of the proteins monitored using the CD method. Since denaturation in both cases is irreversible, the analyses describe only the unfolding transition temperatures. The CD ellipticity of the wild-type at 222 nm shows a moderate linear decrease upon increase in temperature to 91 °C, signifying no changes in the secondary structure. The unfolding transition occurs at a temperature of 94 °C, consistent with that expected for a hyperthermophilic protein. However, there is a significant loss in ellipticity with temperature in the variant, with the transition at 78 °C (Fig. S10D). The thermal unfolding studies confirm that the deletion indeed has an adverse effect on the thermal stability.
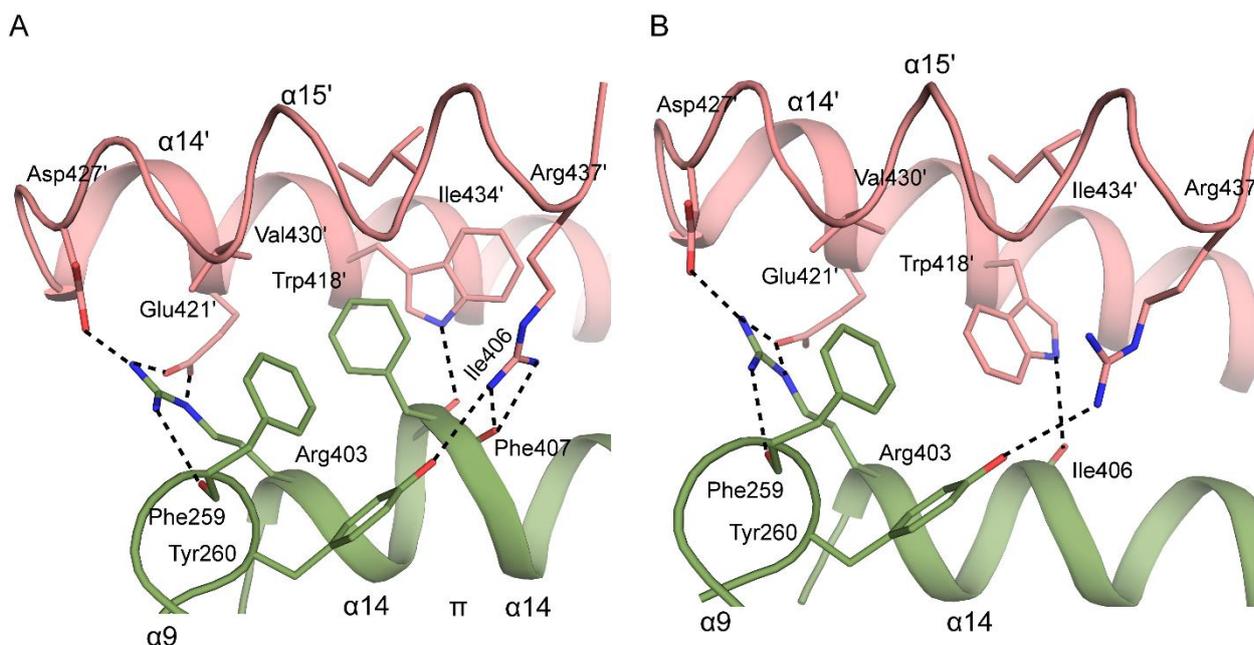
**Figure 5.** Comparison of geometry of interactions in the dimer interface A)Wild-type TmAgu4B (PDB ID: 7BRF) B) Variant TmAgu4BΔ407 (PDB ID:7BR4). Side-chains of residues involved in crucial non-covalent interactions (distances ≤ 4.0 Å) are shown. The hydrogen bonds and salt-bridges (distances ≤ 3.4 Å) are shown as dashed-lines. It is to be noted that these interactions constitute one-half of the interactions of the symmetric dimeric interface.
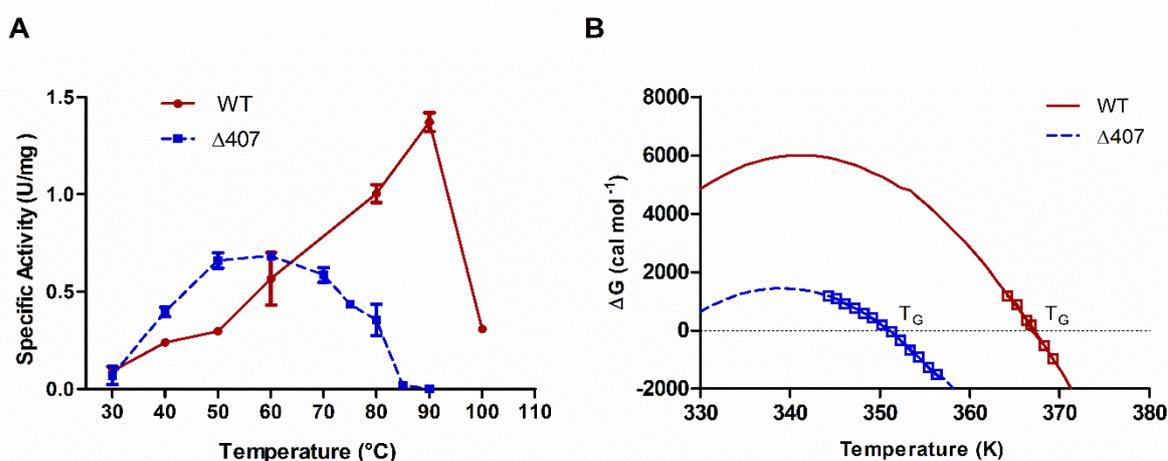


**Figure 6.** Temperature dependence of activity and free energy of unfolding. A) Effect of temperature on specific activity of the wild-type (red line) and the variant (blue line). B) Thermodynamic profiles depicting the free energy of thermal unfolding (*ΔG*) versus temperature (K), for wild-type (solid red line) and variant (dashed blue line). Square symbol (□) correspond to *ΔG* values obtained from temperature denaturation experiments. The temperature where *ΔG* equals zero is the $T_G$ (melting temperature of heat denaturation). The sold line and the dashed line plots represent the fits of protein stability data to the Gibbs-Helmholtz equation.

## 3.3 Overall structures of TmAgu4B and TmAgu4BΔ407

The crystal structures of the wild-type and the variant proteins in complex with NAD were

determined to evaluate how the deletion at the π-helix modifies the structure and affects the

functional and biophysical properties. Both proteins crystallized in the space group C121 under similar crystallization conditions. The wild-type and variant structures were solved by molecular replacement and refined to resolutions of 2.15 Å and 1.95 Å, respectively. The asymmetric unit contains one subunit, with the second subunit of the functional homodimer related by crystallographic two-fold symmetry. The data collection and refinement statistics are summarized in Table 1.

The overall fold of a mixed α/β protein class is unique to the GH4 family. The fold can be divided into three domains; N-terminal region (residues 1-180), central catalytic domain (residues 181-321) and the C-terminal (residues 322-468) (Fig. 1, Fig. S1). The N-terminal region forms the core Rossmann fold-like NAD-binding domain. The other two domains, each consisting of a mix of a twisted antiparallel β sheet and α helices, are positioned largely on top of the Rossmann domain [27,28,53]. The central catalytic region shows wide sequence and structural divergence across the GH4 family and is proposed to have a role in determining substrate specificity [57]. The conserved C-terminal region containing the oligomerization domain comprises of an antiparallel twisted three-stranded β sheet (β11-β13), and α helices, α13-α19.The NAD runs from the distal adenine moiety through to the nicotinamide ribose moiety spanning the surface to the catalytic site. Unlike that in the earlier reported structure of TmAgu4B with a partial model of $NAD^+$ (TM0752, PDB: 1VJT, Joint Centre for Structural Genomics, unpublished) at 2.5 Å resolution, the entire cofactor could be modelled in the wild-type, probably because of the better resolution of the data obtained here (Fig. S11A). Notably, none of the available GH4 structures reported so far represent the snapshot of the cofactor bound holo form. They are either apo forms, inhibitor bound forms or ternary complexes with substrate analogs/products. In the variant, the nicotinamide ring of the cofactor could not be modelled due to poor electron density in this region. However, $Mn^{2+}$ was bound to the conserved metal-binding site and formed the expected coordination network involving SH of Cys181, NE2 of His210and two water molecules (Fig. S11B) [31]. Overall, no significant structural changes occur between the wild-type and the mutant, and the subunits superpose with a rmsd of ~0.36 Å over 468 Cα atoms. Furthermore, the dimeric assembly of the two structures are primarily identical (Fig. S6).


### 3.4 Inter-subunit interfaces in TmAgu4B and TmAgu4BΔ407

The GH4 dimeric interface is entirely formed by symmetric interactions between the C-terminal helical regions, α14, and α15 from both subunits (Fig. 5A, Fig. S6A). In the wild-type, the dimer interface buries accessible surface area of ~1525 $Å^2$ per subunit. Of this, the α14

helix alone (402-425) contributes ~626 $\text{Å}^2$ per subunit of buried surface area, indicating a substantial involvement of this helix to the stability of the dimeric association. The $\pi$-helical region is identified by the $i$ to $i$+5 hydrogen bonding pattern between carbonyl groups of Ile404 and Lys405 to the amide groups of Leu409 and Trp410, respectively (2.7 and 2.9 Å). The extended helical regions on either end maintain the standard $\alpha$ helical $i$ to $i$+4 hydrogen bonding pattern (Fig. 2A). The $\pi$-helical insertion residue Phe407 faces the dimer interface and is nestled in a hydrophobic pocket (buried surface area of 106 $\text{Å}^2$ for Phe407), completely shielding the residue from the solvent (Fig. 5A). Significant interactions that stabilize the otherwise energetically unfavourable $\pi$-helix include an aromatic cluster where Phe407 is sandwiched between a T-$\pi$ (edge to face, C-H••$\pi$, 3.8 Å) stacking interaction with Phe259 of the same subunit, and a $\pi$-$\pi$ (face to face, 3.4 Å) stacking interaction with Trp418′ of the adjacent subunit (Fig. S12A). Other hydrophobic interactions involve sidechains of Val430′ and Ile434′ from the adjacent subunit. Another inter-subunit interaction involves a bifurcated hydrogen bond between the main chain carbonyl of Phe407 and the side chain of Arg437′ (2.8 Å). An almost identical hydrophobic cluster accommodates the equivalent Phe407 in TnAgL. In TmAglA, the equivalent Tyr417 is also completely buried and makes hydrophobic interactions with Met428′ and Ile440′. Moreover, the side chain OH is hydrogen-bonded to O$\epsilon$1 of Glu431′ (2.7 Å). It is striking that the inter-subunit hydrogen bond between the carbonyl group of the $\pi$-helix insertion residue and the Arg437′ sidechain is shared across all three GH4-$\pi$ homologs, TmAgu4B, TnAgl and TmAglA. Arginine at this position is also largely conserved across the GH4-$\pi$ group and is probably another crucial residue that stabilizes the $\pi$-helix geometry (Fig. S5, Fig. S9).

The TmAgu4B$\Delta$407 variant did not contain the $\pi$-helix, but instead, helix $\alpha$14 was a continuous canonical $\alpha$-helix, identical to that in orthologs lacking the $\pi$-helix (Fig. 2C, Fig. S4B, Fig. S6). The electron density maps demonstrate the variant model containing the canonical $\alpha$-helical i to i+4 main chain hydrogen bonding pattern between residues 404-405 and 409-410, respectively (Fig. S13). This is consistent with our original hypothesis that a single residue InDel event, evolutionarily relates the interspersed $\pi$-helix and the $\alpha$-helix in the GH4, with minimal conformational modifications required for the transition. Furthermore, the void at the core of the $\pi$-helix (C$\alpha$ of Tyr408 moves by ~1.5 Å) is now lost, resulting in the closer, more favourable cross-core van der Waals interactions of atoms on opposite sides of the helix. The overall dimeric interface in the variant is mostly identical to that in the wild type (surface area of ~1451 $\text{Å}^2$ per subunit) where the $\alpha$14 contributes ~552 $\text{Å}^2$. The deletion creates a void resulting in a relative loss of ~74 $\text{Å}^2$ of buried surface area per subunit in the variant.

The average B factor values of the equivalent residues for this region in the two structures are ~23 and ~19 Å$^2$, for the wild-type and mutant, respectively, indicating no major effect on the overall flexibility of the region. (Fig. 5A, B). Conformational changes are restricted to minor rearrangement of two side chains. Firstly, sidechain of Trp418′ flips ($\chi$2 rotation from -10˚ to 96˚) to fill the cavity created by the deletion. Despite this flip, the inter-subunit interaction between Ile406 O and Trp418′Nε is unchanged (2.9 – 3.1 Å). Secondly, a minor rearrangement of Arg437 sidechain occurs, while the salt-bridge between Arg403 and Asp427 Oδ2 (2.9 – 3.1 Å) is unchanged. The conserved main-chain–side-chain bifurcated hydrogen bond of the deleted residue Phe407 carbonyl oxygen with Arg437′ is lost. Together, it is clear that the most significant alteration at the interface is the relative loss of van der Waals and hydrophobic interactions in the variant. In particular, the aromatic-aromatic trimer cluster of Phe256, Phe407 and Trp418′ that stabilizes both the $\pi$-helix and the inter-subunit association in the wild-type, is lost in the variant (Fig. S12). Indeed, in large scale surveys of protein structures, aromatic trimers are energetically favourable and occur frequently, and the T-$\pi$ and $\pi$-$\pi$ type cluster is the most abundant type at 80 % or more (40–44).

An interesting feature is that a proline residue (Pro411 in TmAgu4B) is present at the C-terminal end where the $\pi$-helix shifts back to an $\alpha$-helix. This feature is strongly conserved across the GH4-$\pi$ subfamily (Fig. 2A, Fig. 3). The high propensity of Pro, a helix disrupter, at this position in interspersed $\pi$-helices or $\pi$-bulges has been previously reported in large scale comparative analyses [11,13]. This appears to be a defining characteristic of the $\pi$-helical feature in the GH4 subfamily as well. As a result, the main chain i to i+5 hydrogen bond between residues Ile406 and Pro411 in TmAgu4B and the equivalent i to i+4 hydrogen bond between these residues in the variant, are absent (Fig. 2A, C). However, the lack of this stabilizing hydrogen bond in the variant does not disrupt the standard $\alpha$-helical pattern in the residues preceding and succeeding Pro in $\alpha$14 (residues 402 to 424). The presence of the proline is another evolutionary marker that is robust to uncertainties in the MSA and tree topology employed to infer the reliability of the phylogenetically predicted GH4-$\pi$ clade.

## 3.5 Comparison of the active sites of TmAgu4B and TmAgu4BΔ407

The GH4 family active site is at the bottom of a cavity that is well conserved, particularly with respect to the NAD and metal-binding sites. The general catalytic mechanism and the role of the divalent metal, usually Mn$^{2+}$, is well described. Mn$^{2+}$ is present in a hexa-coordinated network of interactions involving SH of Cys181, NE2 of His210 and atoms O2 and O3 of the

sugar and three water molecules. The metal is expected to act to both stabilize the generated anions and polarize the carbonyl group generated in the intermediates during the catalytic cycle [28,31,53,55,60]. The catalytic residues His210 and Cys181 are absolutely conserved across the GH4 family.

Comparison of the active sites indicates negligible perturbations in the extended active site, including the catalytic residues, the substrate and metal-binding sites, and the NAD binding site. The mode of NAD binding is also identical between the two structures. Bound $Mn^{2+}$ coordinated to His210 and Cys181, and two water molecules is observed only in the mutant where the nicotinamide moiety is disordered. The metal ion is absent in the wild-type with Cys181 present in an oxidized cysteine sulfenic acid (Cys-SOH) state, likely preventing metal ion coordination (Fig. S11A). The ordered nicotinamide moiety of bound NADH wild-type shows that nicotinamide is oriented *anti* to the ribose with the N7 atom hydrogen-bonded to peptide carbonyl oxygens of residues Thr158 and Phe180.In contrast, the nicotinamide ring in the TmAglA-NAD ternary complex with substrate maltose is in *syn* orientation and stacks with maltose and likely represents a non-productive binding mode of the substrate and the nicotinamide moiety [27].

Given that the catalytic activity of the variant is disrupted only at high temperatures, the structures were examined to identify any direct structural contacts between the π-helical residues and the catalytic site. Interestingly, in the wild-type, residues Tyr408 and Leu409 makes multiple strong van der Waals interactions (< 4Å) with neighbouring residues in the region 208-213 that contains the catalytic His210. His210 is positioned at residue i+1 of a β-turn connecting strands β7 and β8 (Fig. 7). In the variant, however, residues 408 and 409 that have now transitioned to the more compact α-helical geometry show significantly altered patterns with a relatively weaker set of non-bonded interactions with the residues 208-213. For instance, the side chain of Tyr408 is shifted away from Gly211 by ~1.5 Å and the side chain of Trp213 by ~1.2 Å.
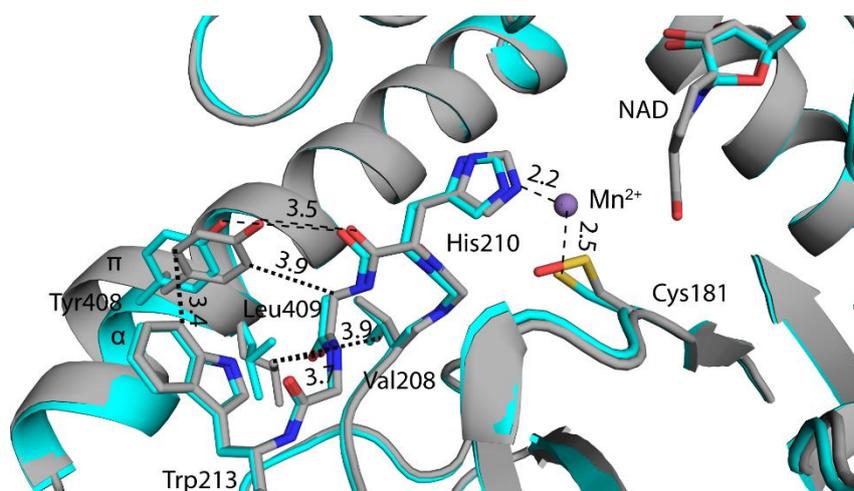
**Figure 7.** Dissection of interactions between the π helical region and the neighboring conserved catalytic loop. Superposition of the wild-type (grey) with π helix and the variant (cyan) with the equivalent α-helix is shown. Side chains of residues that contribute to non-covalent interactions (< 4.0 Å) between the two regions are shown in stick representation. The dashed lines indicate distances between corresponding atoms involving residues Tyr408 and Leu409, with the main chain and side chain atoms of residues in the conserved catalytic site loop (208-213). Stringently conserved catalytic residue His210 is involved in coordinating with $Mn^{2+}$ (purple sphere) along with another conserved residue Cys181, as observed in the variant. Cys181 in the wild-type is present in an oxidized cysteine sulfenic acid (Cys-SOH) state and lacks the bound metal ion. The local non-covalent interactions between the two regions are significantly relaxed in the variant compared to the wild-type. The observed structural plasticity at the π helical region may be a key feature responsible for the deleterious effect on the temperature optimum for highest activity (~60 °C) in the variant in comparison to the wild-type (~90 °C).

### 3.6 Thermodynamic characterization of TmAgu4B and TmAgu4BΔ407

The wild-type had an optimum temperature for the highest activity of 90 °C and the melting temperature of thermal denaturation was 94 °C. The corresponding values for the variant are 60 °C and 78 °C, indicating that the variant is structurally less stable. In order to establish the structure-stability implications of the π-helix, we characterized the thermodynamic parameters for both proteins using stability curves representing the temperature dependence of free energy of unfolding ($\Delta G$). The stability was evaluated by thermal denaturation, monitored by CD spectra. The simplest applicable two-state process of dissociation/unfolding of the native dimeric to the unfolded monomeric state was assumed here. For each protein, $\Delta G°(T)$ was plotted as a function of temperature. The two stability curves were fit to the Gibbs-Helmholtz equation to obtain the thermodynamic parameters, namely, the heat capacity change $\Delta C_p°$, the melting temperature $T_G$, and the enthalpy change at the melting temperature $\Delta H°$ (Table. 2). The CD data fitted well to the extrapolated thermodynamic parameters indicating that the two-state model reliably describes the thermal unfolding curves independent of the ellipticity parameter used for detecting the conformational changes.

As is expected, the differences in the thermodynamic stability of the two proteins, as estimated by the maximal protein stability $\Delta G°(T_S)$ values, are different (Fig. 6B, Table 2). The temperature at which the $\Delta G$ equals zero represents the thermal midpoint ($T_G$) of heat denaturation. The $\Delta H°$ determines the slope of the stability curve while heat capacity change upon unfolding ($\Delta C_p°$) gives the curvature. The $T_G$ values of heat denaturation are ~94 °C and 78 °C for the wild-type and mutant, respectively. The higher thermal stability of the wild-type is a result of both an upshift and a broadening of the stability curves as compared to the mutant, indicating that the wild-type is more stable at all temperatures. The larger $\Delta G(T_S)$ value for the wild-type appears to have been achieved by ~2-fold higher value of the calculated specific enthalpy of unfolding $\Delta H°$, and a lowering of the $\Delta C_p°$ value (6.2 kcal/mol/K for the wild-type vs. 6.7 kcal/mol/K for the variant). However, it must be noted that the $\Delta C_p°$ values here have not been determined experimentally, and we cannot rule out other paths to obtaining the higher $\Delta G°(T_S)$ in the wild-type. Although the maximum stabilities of the wild-type and the variant differ by ~12 kcal/mol, the respective temperatures of maximum stability ($T_S$) are very similar at 64~66 °C. Furthermore, the calculated values of $T_H$, which is the temperature at which the solubility of the protein is at a minimum, are close to the $T_S$ values as expected.

## 4. Discussion

Although an interspersed π-helix is often described as an α-helical distortion and considered an anomaly, the two secondary structures are related by a single residue InDel event constituting a structurally disruptive backbone mutation with evolutionary and functional implications. This energetically unfavourable insertion within an α-helix is subject to far stronger selection than residue substitutions. Previous studies have shown that many such naturally occurring π-helices are conserved and can affect function utilizing diverse mechanisms [18,61,62]. For instance, π-helices in the pore-lining helices of Transient Receptor Potential (TRP) channels have been implicated both in the assembly and channel opening dynamics involving an α-to-π secondary structure transition [63]. In contrast, in the lipooxygenases, a π-helix serves to position two histidine residues to coordinate the catalytic iron [64]. It has been proposed that the single residue insertional origin of some π-helices occurs within discrete subgroups of protein superfamilies as a mechanism to evolve divergent functionality [65]. However, testing this hypothesis is not always conclusive. For instance, enzymes belonging to the NAD(P)H:FMN reductase family are grouped into subgroups based on the absence or presence of a π-helix in the subunit interface. But, studies on variants of

homolog SsuE, indicate that a single residue insertion is not sufficient to generate a π-helix and that other alterations around the insertion site are also required [66].

The GH4 family is unique among the numerous glycosidase families, employing a novel reaction mechanism. Furthermore, the GH4 fold is equally unique, sharing a common ancestor with the Rossmann-fold containing NAD-dependent dehydrogenases. The oligomerization domain containing the α14-helix constitutes additional secondary structure elements that emerged in GH4 by divergent evolution [27,28,53]. The GH4-π group evolved a π-helix transition within the α14-helix, whereas the corresponding region in other orthologous groups is a canonical α-helix. Our phylogenetic analysis confirms that the π-helix is indeed an evolutionary marker that defines a new GH4-π subfamily (Fig. 4). Structural/sequence evaluation indicates that the insertion is usually a large aliphatic or aromatic residue. This residue, involved in multiple non-covalent interactions across the interface, was predicted to disturb the integrity/stability of the oligomerization and catalytic activity. Interestingly, the deletion of the specific residue Phe407 transforms the π-helix to a canonical α-helix, while the dimeric assembly in the variant remains intact. The largest conformational changes involve a sidechain rearrangement with concomitant loss of non-covalent interactions and hydrophobic buried surface area (Fig. 5).

The most dramatic change in properties of the variant is the loss of thermostability, with the apparent $T_M$ decreasing by ~15 °C, consistent with the poorer packing at the interface. The substantial decrease in the optimum temperature for the highest activity by ~30°C in the variant indicates the conversion from a hyperthermostable to a thermostable enzyme (Fig. 6A). However, the mutation is not deleterious for the activity of the variant at temperatures below 60 °C. This is in agreement with the absence of any discernible changes in the conformational geometry of the catalytic site and the extended active site residues between the two structures. In contrast, the effect on activity is significant at temperatures higher than 60 °C. The structures provide a plausible mechanism for how the compacting of the residues Tyr408 and Leu409 to the α-helical conformation in the mutant is transmitted to the active site. It appears that the relative loss of non-covalent interactions between the two residues and the neighboring region containing the catalytic His210, impairs catalytic activity by imposing a deleterious effect on the dynamics of the otherwise rigid metal-coordination geometry (Fig. 7). The effect is manifested at temperatures higher than 60 °C (Fig. 6A). This provides an example of how the intrinsic structural plasticity of the π-helical region that is not directly part of the active site can contribute to function.

Notably, the temperature dependence of the catalytic activity of the variant displayed an interesting trend. The mutant displays higher relative activity than the wild-type at temperatures in the range 30 – 50 °C. In fact, at 50 °C, the activity of the mutant is ~2.4-fold larger than that of the wild-type (Fig. 6A). This trend is typical of a moderately thermophilic homolog that has evolved an adaptation for catalysis at a lower temperature. Together with the thermostability data, it appears that the hyperthermostability of the wild-type is associated with greater structural rigidity and that the relatively restricted conformational flexibility in comparison to the variant, compromises its catalytic activity at temperatures lower than 50 °C.

Investigations of the molecular evolution of hyperthermophilic and thermophilic proteins have shown the lack of a universal structural mechanism to increase thermal resistance with respect to their mesophilic orthologs. Protein structure modifications stabilize structure through a combination of several factors, including non-covalent interactions and other global structural adaptations that comprise disulfide bonds, quaternary structure, content of positively charged residues, secondary structure propensity, protein compaction, among others [67–69]. Given that the crystal structures have captured the variations between the two proteins here, we carried out a thermodynamic comparison to provide a context for understanding the structural basis of thermoadaptation. Three models of modulating protein stability curves explain how thermophilic proteins achieve higher thermostability in comparison to the mesophilic counterpart [70,71]. In model I, altered interactions, for instance, by an increase in the number of enthalpic interactions, will shift the entire curve upwards to a higher $\Delta G°$ of unfolding. In model II, the stability curve of the thermostable protein is flattened to a higher melting temperature by a lowering of the $\Delta C_p°$ for the same $\Delta G°$ value of stability. Finally, in model III, the entire curve is horizontally shifted to the right towards higher temperatures, for the same $\Delta G°$ value, by lowering the change in entropy of folding. Different proteins have adopted these strategies independently or in combinations, to achieve thermostability [50,72].

A comparison of the free energy profiles indicates that both the increasing and flattening of the thermodynamic stability curve contribute to the higher melting temperature in the wild-type (Fig. 6B). The contribution from the larger value of enthalpy of unfolding $\Delta H°$ in the wild-type relative to the variant is consistent with the fact that specific non-covalent interactions that stabilize the wild-type are lost in the variant (Fig. 5). However, the relatively significant difference in $\Delta C_p°$ values between the two proteins is unexpected since the structural data sheds little light on the source of this variation. The $\Delta C_p$ of a globular protein, to a large extent, is accounted for by variations in water-accessible surface area (ASA) of non-polar

groups upon transition between the folded and the unfolded state (59–62). The observed difference between the values of $\Delta C_p°$ (~0.5 kcal/mol/K) for the two proteins would imply a relative change of ~2900 $\text{Å}^2$ and cannot be explained by the measured ASA change of ~148 $\text{Å}^2$, between the two folded states. The simplest explanation is that the nature of the residual structures in the unfolded states of these proteins may be different. Comparison of the stability curves together with the native structures and the activity data implies a role for the unfolded state in determining the thermostability. We propose that the π-helix insertion residue which is well packed within a largely hydrophobic pocket in the folded state contributes to a putative residual structure (perhaps, a hydrophobic cluster akin to the aromatic-aromatic trimer in the folded form), that persists in the unfolded state and reduces the total amount of surface area it must bury upon folding. The deletion of the insertional residue probably disturbs this residual structure in the unfolded state of the variant, increasing the $\Delta C_p°$, and making it a thermophile rather than a hyperthermophile (Fig. 5 and Fig. S12). Thermophilic proteins that use a combination of higher $\Delta H°$ and lower $\Delta C_p°$ to increase $T_M$, relative to mesophilic homologs, are the most commonly observed examples [76–79].

Regardless of the large differences in the heat transition temperatures and free energy values between the wild-type and the mutant, the temperatures of maximum stability ($T_S$) of both proteins are very similar (64~66 °C), and lower than the living temperature of *T. maritima* (80 °C) (Table 2, Fig. 6B). In general, maximum stability temperatures of thermophilic proteins are lower than the living temperatures of the source organisms and is closely correlated to the molecular flexibility that is required for optimal functionality at that temperature. Interestingly, the $T_S$ values of the two proteins are very close to the temperature at which their catalytic activities are nearly identical (Fig. 6A). It appears that this is a result of different manifestations of the complex interplay of molecular flexibility and rigidity within each protein. The lower $\Delta H°$ and $\Delta G°(T_S)$ values of the variant reflect a different state of optimum structural flexibility that contributes to its 2-fold higher catalytic activity compared to the wild-type, at 50 °C.

Our structure-function studies provide compelling evidence for the evolutionary relationship between the π-helix and the α-helix in the GH4. Many previous examples suggest that the insertional origin of π-helix is more prevalent in protein families [80]. Since these two features can interconvert both directions during protein evolution, we tried to rationalize their gain and loss in the GH4. The taxonomic distribution of the GH4 is restricted to the Archaea and Bacteria, suggesting a deep ancestry of this enzyme. Furthermore, organismal sources of the GH4-π subgroup proteins reveal a strong correlation between the presence of the π-helix

and their thermophilic/hyperthermophilic living temperatures. In contrast, taxonomic distribution of the α-helix containing homologs are largely mesophilic, with few thermophilic and hyperthermophilic homologs (Table. S1). It is generally accepted that the last universal common ancestors (LUCA) of Bacteria and Archaea were hyperthermophilic [81–83]. Experimental support for this elevated thermostability of LUCA comes from several studies that have characterized reconstructed ancestral proteins [84–86]. Considering that the π-helix plays a conserved role in the thermostability of the GH4-π subfamily, it is tempting to speculate that the hypothetical thermophilic ancestor of the GH4 family had a π-helix that was accepted since the conformational adjustment in the neighboring region was energetically and functionally favourable. Under this hot-start scenario, because reaction rates are temperature dependent, the evolution towards mesophilic enzymes must involve an adaptation to increase activity at lower temperatures, whereas the selection pressure on thermostability is relaxed. This activity-stability trade-off is exemplified in the properties of the moderately thermophilic variant, namely, a transition to an α-helix by deletion of a single residue, lower thermodynamic stability, but a higher catalytic rate at lower temperatures. The variant generated in our study can thus be considered as an equivalent of a reconstructed ancestral protein or a living protein fossil. Extant hyperthermophilic α-helix containing GH4 homologs could have re-evolved from this less thermostable α-helical ancestor by selection on the thermostability on return to a hotter environment, but with no pressure on the already optimized activity. An alternative path for the evolutionary origin of the π-helix in the GH-π subgroup envisages an independent evolution by insertion of one residue in a α-helix containing thermophilic/mesophilic ancestor and subsequent selection on stability and/or activity.

## 5. Conclusions

Our findings reveal a surprising use of an energetically unfavourable secondary structure backbone mutation as a structural mechanism for thermoadaptation. The presence of specific interactions and residual structure in the unfolded state, appear to be vital *in vivo* mechanisms to balance the need to preserve structure at higher temperature with the thermodynamic stability required for optimal catalysis.

## Acknowledgments

**Conflict of Interest:** The authors declare that they have no conflicts of interest with the contents of this article.

**References**
[1]     J.D. Bloom, F.H. Arnold, In the light of directed evolution: Pathways of adaptive protein evolution, Light Evol. 3 (2009) 149–163. doi:10.17226/12692.

[2]     Á. Tóth-Petróczy, D.S. Tawfik, Protein insertions and deletions enabled by neutral roaming in sequence space, Mol. Biol. Evol. 30 (2013) 761–771. doi:10.1093/molbev/mst003.

[3]     W.W. de Jong, L. Rydén, Causes of more frequent deletions than insertions in mutations and protein evolution, Nature. 290 (1981) 157–159. doi:10.1038/290157a0.

[4]     M.S. Taylor, C.P. Ponting, R.R. Copley, Occurrence and consequences of coding sequence insertions and deletions in mammalian genomes, Genome Res. 14 (2004) 555–566. doi:10.1101/gr.1977804.

[5]     J.A.J. Arpino, P.J. Rizkallah, D.D. Jones, Structural and dynamic changes associated with beneficial engineered single-amino-acid deletion mutations in enhanced green fluorescent protein, Acta Crystallogr. Sect. D Biol. Crystallogr. 70 (2014) 2152–2162. doi:10.1107/S139900471401267X.

[6]     J.K. Myers, C.N. Pace, Hydrogen bonding stabilizes globular proteins, Biophys. J. 71 (1996) 2033–2039. doi:10.1016/S0006-3495(96)79401-8.

[7]     G.D. Rose, P.J. Fleming, J.R. Banavar, A. Maritan, A backbone-based theory of protein folding, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 16623–16633. doi:10.1073/pnas.0606843103.

[8]     D.W. Bolen, G.D. Rose, Structure and Energetics of the Hydrogen-Bonded Backbone in Protein Folding, Annu. Rev. Biochem. 77 (2008) 339–362. doi:10.1146/annurev.biochem.77.061306.131357.

[9]     D.P. Teufel, C.M. Johnson, J.K. Lum, H. Neuweiler, Backbone-driven collapse in unfolded protein chains, J. Mol. Biol. 409 (2011) 250–262. doi:10.1016/j.jmb.2011.03.066.

[10]    C. Nick Pace, J. Martin Scholtz, G.R. Grimsley, Forces stabilizing proteins, FEBS Lett. 588 (2014) 2177–2184. doi:10.1016/j.febslet.2014.05.006.

[11]    J.P. Cartailler, H. Luecke, Structural and Functional Characterization of $\pi$ Bulges and Other Short Intrahelical Deformations, Structure. 12 (2004) 133–144. doi:10.1016/j.str.2003.12.001.

[12]    R.P. Riek, R.M. Graham, The elusive $\pi$-helix, J. Struct. Biol. 173 (2011) 153–160. doi:10.1016/j.jsb.2010.09.001.

[13]    P. Kumar, M. Bansal, Dissecting $\pi$-helices: Sequence, structure and function, FEBS J. 282 (2015) 4415–4432. doi:10.1111/febs.13507.

[14]    D.W. Heinz, W.A. Baase, F.W. Dahlquist, B.W. Matthews, How amino-acid insertions are allowed in an $\alpha$-helix of T4 lysozyme, Nature. 361 (1993) 561–564. doi:10.1038/361561a0.

[15]    L.J. Keefe, J. Sondek, D. Shortle, E.E. Lattman, The $\alpha$ aneurism: A structural motif revealed in an insertion mutant of staphylococcal nuclease, Proc. Natl. Acad. Sci. U. S. A. 90 (1993) 3275–3279. doi:10.1073/pnas.90.8.3275.

[16] J.A. Hardy, S.T. Walsh, H.C. Nelson, Role of an α-helical bulge in the yeast heat shock transcription factor 1 1Edited by F. E. Cohen, J. Mol. Biol. 295 (2002) 393–409. doi:10.1006/jmbi.1999.3357.

[17] R.B. Cooley, T.W. Rhoads, D.J. Arp, P.A. Karplus, A diiron protein autogenerates a valine-phenylalanine cross-link, Science 332. 6032 (2011) 929. doi:10.1126/science.1205687.

[18] T.M. Weaver, The π-helix translates structure into function, Protein Sci. 9 (2008) 201–206. doi:10.1110/ps.9.1.201.

[19] M.N. Fodje, S. Al-Karadaghi, Occurrence, conformational features and amino acid propensities for the π-helix, Protein Eng. Des. Sel. 15 (2002) 353–358. doi:10.1093/protein/15.5.353.

[20] A.E. Medlock, T.A. Dailey, T.A. Ross, H.A. Dailey, W.N. Lanzilotta, A π-Helix Switch Selective for Porphyrin Deprotonation and Product Release in Human Ferrochelatase, J. Mol. Biol. 373 (2007) 1006–1016. doi:10.1016/j.jmb.2007.08.040.

[21] J.A. Cotruvo, J. Stubbe, Escherichia coli Class Ib ribonucleotide reductase contains a dimanganese(III)-tyrosyl radical cofactor in vivo, Biochemistry. 50 (2011) 1672–1681. doi:10.1021/bi101881d.

[22] J.M. Musila, H.R. Ellis, Transformation of a Flavin-Free FMN Reductase to a Canonical Flavoprotein through Modification of the π-Helix, Biochemistry. 55 (2016) 6389–6394. doi:10.1021/acs.biochem.6b00452.

[23] A.R. Bernard, T.C. Jessop, P. Kumar, N.E. Dickenson, Deoxycholate-Enhanced Shigella Virulence Is Regulated by a Rare π-Helix in the Type Three Secretion System Tip Protein IpaD, Biochemistry. 56 (2017) 6503–6514. doi:10.1021/acs.biochem.7b00836.

[24] A.S. Chilton, A.L. Lamb, H.R. Ellis, D.L. Forbes, J.S. McFarlane, R.A. Hagen, Not as easy as π: An insertional residue does not explain the π-helix gain-of-function in two-component FMN reductases, Protein Sci. 28 (2018) 123–134. doi:10.1002/pro.3504.

[25] V. Lombard, H. Golaconda Ramulu, E. Drula, P.M. Coutinho, B. Henrissat, The carbohydrate-active enzymes database (CAZy) in 2013., Nucleic Acids Res. 42 (2014) D490-5. doi:10.1093/nar/gkt1178.

[26] B.G. Hall, A. Pikis, J. Thompson, Evolution and biochemistry of family 4 glycosidases: Implications for assigning enzyme function in sequence annotations, Mol. Biol. Evol. 26 (2009) 2487–2497. doi:10.1093/molbev/msp162.

[27] J.A. Lodge, T. Maier, W. Liebl, V. Hoffmann, N. Sträter, Crystal structure of Thermotoga maritima α-glucosidase AglA defines a new clan of $NAD^+$-dependent glycosidases, J. Biol. Chem. 278 (2003) 19151–19158. doi:10.1074/jbc.M211626200.

[28] S.S. Rajan, X. Yang, F. Collart, V.L.Y. Yip, S.G. Withers, A. Varrot, J. Thompson, G.J. Davies, W.F. Anderson, Novel catalytic mechanism of glycoside hydrolysis based on the structure of an $NAD^+/Mn^{2+}$-dependent phospho-α-glucosidase from Bacillus subtilis, Structure. 12 (2004) 1619–1629. doi:10.1016/j.str.2004.06.020.

[29] A. Varrot, V.L.Y. Yip, Y. Li, S.S. Rajan, X. Yang, W.F. Anderson, J. Thompson, S.G. Withers, G.J. Davies, $NAD^+$ and Metal-ion Dependent Hydrolysis by Family 4 Glycosidases: Structural Insight into Specificity for Phospho-β-D-glucosides, J. Mol. Biol. 346 (2005) 423–435. doi:10.1016/j.jmb.2004.11.058.

[30] B.Y. Yun, S.Y. Jun, N.A. Kim, B.Y. Yoon, S. Piao, S.H. Park, S.H. Jeong, H. Lee, N.C. Ha, Crystal structure and thermostability of a putative α-glucosidase from Thermotoga neapolitana, Biochem. Biophys. Res. Commun. 416 (2011) 92–98. doi:10.1016/j.bbrc.2011.11.002.

[31] S.B. Mohapatra, N. Manoj, Biochemical and Biophysical Research Communications Structure of an a -glucuronidase in complex with $Co^{2+}$ and citrate provides insights

into the mechanism and substrate recognition in the family 4 glycosyl hydrolases, Biochem. Biophys. Res. Commun. 518 (2019) 197–203. doi:10.1016/j.bbrc.2019.08.030.

[32]   W. Li, A. Godzik, Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences, Bioinformatics. 22 (2006) 1658–1659. doi:10.1093/bioinformatics/btl158.

[33]   K. Katoh, D.M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability, Mol. Biol. Evol. 30 (2013) 772–780. doi:10.1093/molbev/mst010.

[34]   L. Holm, Benchmarking Fold Detection by DaliLite v.5., Bioinformatics. (2019). doi:10.1093/bioinformatics/btz536.

[35]   G. Crooks, G. Hon, J. Chandonia, S. Brenner, WebLogo: a sequence logo generator, Genome Res. 14 (2004) 1188–1190. doi:10.1101/gr.849004.1.

[36]   X. Robert, P. Gouet, Deciphering key features in protein structures with the new ENDscript server, Nucleic Acids Res. 42 (2014) 320–324. doi:10.1093/nar/gku316.

[37]   A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies, Bioinformatics. 30 (2014) 1312–1313. doi:10.1093/bioinformatics/btu033.

[38]   D. Darriba, G.L. Taboada, R. Doallo, D. Posada, Europe PMC Funders Group ProtTest 3 : fast selection of best-fit models of protein evolution, Bioinformatics. 27 (2017) 1164–1165. doi:10.1093/bioinformatics/btr088.ProtTest.

[39]   M.D. Winn, C.C. Ballard, K.D. Cowtan, E.J. Dodson, P. Emsley, P.R. Evans, R.M. Keegan, E.B. Krissinel, A.G.W. Leslie, A. McCoy, S.J. McNicholas, G.N. Murshudov, N.S. Pannu, E.A. Potterton, H.R. Powell, R.J. Read, A. Vagin, K.S. Wilson, Overview of the CCP4 suite and current developments, Acta Crystallogr. Sect. D Biol. Crystallogr. 67 (2011) 235–242. doi:10.1107/S0907444910045749.

[40]   T.G.G. Battye, L. Kontogiannis, O. Johnson, H.R. Powell, A.G.W. Leslie, iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM., Acta Crystallogr. D. Biol. Crystallogr. 67 (2011) 271–281. doi:10.1107/S0907444910048675.

[41]   W. Kabsch, Automatic indexing of rotation diffraction patterns, J. Appl. Crystallogr. 21 (1988) 67–72. https://doi.org/10.1107/S0021889887009737.

[42]   P.D. Adams, P. V Afonine, G. Bunkoczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.-W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger, P.H. Zwart, PHENIX: a comprehensive Python-based system for macromolecular structure solution, Acta Crystallogr. Sect. D. 66 (2010) 213–221. https://doi.org/10.1107/S0907444909052925.

[43]   P. V Afonine, R.W. Grosse-Kunstleve, N. Echols, J.J. Headd, N.W. Moriarty, M. Mustyakimov, T.C. Terwilliger, A. Urzhumtsev, P.H. Zwart, P.D. Adams, Towards automated crystallographic structure refinement with phenix.refine., Acta Crystallogr. D. Biol. Crystallogr. 68 (2012) 352–367. doi:10.1107/S0907444912001308.

[44]   P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Features and development of Coot, Acta Crystallogr. Sect. D. 66 (2010) 486–501. https://doi.org/10.1107/S0907444910007493.

[45]   V.B. Chen, W.B. 3rd Arendall, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson, D.C. Richardson, MolProbity: all-atom structure validation for macromolecular crystallography., Acta Crystallogr. D. Biol. Crystallogr. 66 (2010) 12–21. doi:10.1107/S0907444909042073.

[46]   E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng,

T.E. Ferrin, UCSF Chimera--a visualization system for exploratory research and analysis., J. Comput. Chem. 25 (2004) 1605–1612. doi:10.1002/jcc.20084.

[47]    The PyMOL Molecular Graphics System, Version 2.0 Schrodinger, LLC.

[48]    R. Maiti, G.H. Van Domselaar, H. Zhang, D.S. Wishart, SuperPose: A simple server for sophisticated structural superposition, Nucleic Acids Res. 32 (2004) 590–594. doi:10.1093/nar/gkh477.

[49]    N.J. Greenfield, Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions, Nat. Protoc. 1 (2007) 2527–2535. doi:10.1038/nprot.2006.204.

[50]    S. Kumar, C.J. Tsai, R. Nussinov, Thermodynamic differences among homologous thermophilic and mesophilic proteins, Biochemistry. 40 (2001) 14152–14165. doi:10.1021/bi0106383.

[51]    A.A. Saboury, A.A. Moosavi-Movahedi, Derivation of the thermodynamic parameters involved in the elucidation of protein thermal profiles, Biochem. Educ. 23 (1995) 164–167. doi:10.1016/0307-4412(95)00049-9.

[52]    V.L.Y. Yip, S.G. Withers, Nature's many mechanisms for the degradation of oligosaccharides, Org. Biomol. Chem. 2 (2004) 2707–2713. doi:10.1039/b408880h.

[53]    A. Varrot, V.L.Y. Yip, Y. Li, S.S. Rajan, X. Yang, W.F. Anderson, J. Thompson, S.G. Withers, G.J. Davies, NAD$^+$ and metal-ion dependent hydrolysis by family 4 glycosidases: Structural insight into specificity for phospho-β-D- glucosides, J. Mol. Biol. 346 (2005) 423–435. doi:10.1016/j.jmb.2004.11.058.

[54]    V.L.Y. Yip, J. Thompson, S.G. Withers, Mechanism of GlvA from Bacillus subtilis: a detailed kinetic analysis of a 6-phospho-α-glucosidase from glycoside hydrolase family 4., Biochemistry. 46 (2007) 9840–52. doi:10.1021/bi700536p.

[55]    W. Huang, J. Llano, J.W. Gauld, Redox mechanism of glycosidic bond hydrolysis catalyzed by 6-phospho-α-glucosidase: A DFT study, J. Phys. Chem. B. 114 (2010) 11196–11206. doi:10.1021/jp102399h.

[56]    N. Sannikova, C.A.N. Gerak, F.S. Shidmoossavee, D.T. King, S. Shamsi, K. Abadi, A.R. Lewis, A.J. Bennet, Both Chemical and Non-Chemical Steps Limit the Catalytic Efficiency of Family 4 Glycoside Hydrolases, Biochemistry. 57 (2018) 3378–3386. doi:10.1021/acs.biochem.8b00117.

[57]    V.L.Y. Yip, A. Varrot, G.J. Davies, S.S. Rajan, X. Yang, J. Thompson, W.F. Anderson, S.G. Withers, An unusual mechanism of glycoside hydrolysis involving redox and elimination steps by a family 4 β-glycosidase from Thermotoga maritima, J. Am. Chem. Soc. 126 (2004) 8354–8355. doi:10.1021/ja047632w.

[58]    E. Lanzarotti, R.R. Biekofsky, D.A. Estrin, M.A. Marti, A.G. Turjanski, Aromatic-aromatic interactions in proteins: Beyond the dimer, J. Chem. Inf. Model. 51 (2011) 1623–1633. doi:10.1021/ci200062e.

[59]    M. Chourasia, G.M. Sastry, G.N. Sastry, Aromatic-Aromatic Interactions Database, A2ID: An analysis of aromatic π-networks in proteins, Int. J. Biol. Macromol. 48 (2011) 540–552. doi:10.1016/j.ijbiomac.2011.01.008.

[60]    S. Chakladar, L. Cheng, M. Choi, J. Liu, A.J. Bennet, Mechanistic evaluation of MelA α-galactosidase from citrobacter freundii: A family 4 glycosyl hydrolase in which oxidation is rate-limiting, Biochemistry. 50 (2011) 4298–4308. doi:10.1021/bi101808h.

[61]    R.P. Riek, A. a Finch, G.E. Begg, R.M. Graham, Wide turn diversity in protein transmembrane helices implications for G-protein-coupled receptor and other polytopic membrane protein structure and function., Mol. Pharmacol. 73 (2008) 1092–1104. doi:10.1124/mol.107.043042.

[62]    A. Gonzalez, A. Cordomí, G. Caltabiano, L. Pardo, Impact of Helix Irregularities on

Sequence Alignment and Homology Modeling of G Protein-Coupled Receptors, ChemBioChem. 13 (2012) 1393–1399. doi:10.1002/cbic.201200189.

[63] L. Zubcevic, S.Y. Lee, The role of $\pi$-helices in TRP channel gating, Curr. Opin. Struct. Biol. 58 (2019) 314–323. doi:10.1016/j.sbi.2019.06.011.

[64] W. Minor, J. Steczko, B. Stec, Z. Otwinowski, J.T. Bolin, R. Walter, B. Axelrod, Crystal structure of soybean lipoxygenase L-1 at 1.4 Å resolution, Biochemistry. 35 (1996) 10687–10701. doi:10.1021/bi960576u.

[65] R.B. Cooley, D.J. Arp, P.A. Karplus, Symerythrin structures at atomic resolution and the origins of rubrerythrins and the ferritin-like superfamily, J. Mol. Biol. 413 (2011) 177–194. doi:10.1016/j.jmb.2011.08.019.

[66] J.S. McFarlane, R.A. Hagen, A.S. Chilton, D.L. Forbes, A.L. Lamb, H.R. Ellis, Not as easy as $\pi$: An insertional residue does not explain the $\pi$-helix gain-of-function in two-component FMN reductases, Protein Sci. 28 (2019) 123–134. doi:10.1002/pro.3504.

[67] N.J. Fraser, J.W. Liu, P.D. Mabbitt, G.J. Correy, C.W. Coppin, M. Lethier, M.A. Perugini, J.M. Murphy, J.G. Oakeshott, M. Weik, C.J. Jackson, Evolution of Protein Quaternary Structure in Response to Selective Pressure for Increased Thermostability, J. Mol. Biol. 428 (2016) 2359–2371. doi:10.1016/j.jmb.2016.03.014.

[68] A. V. Glyakina, S.O. Garbuzynskiy, M.Y. Lobanov, O. V. Galzitskaya, Different packing of external residues can explain differences in the thermostability of proteins from thermophilic and mesophilic organisms, Bioinformatics. 23 (2007) 2231–2238. doi:10.1093/bioinformatics/btm345.

[69] A. Karshikoff, R. Ladenstein, Ion pairs and the thermotolerance of proteins from hyperthermophiles: A "traffic rule" for hot roads, Trends Biochem. Sci. 26 (2001) 550–557. doi:10.1016/S0968-0004(01)01918-1.

[70] H. Nojima, A. Ikai, T. Oshima, H. Noda, Reversible thermal unfolding of thermostable phosphoglycerate kinase. Thermostability associated with mean zero enthalpy change, J. Mol. Biol. 116 (1977) 429–442. doi:10.1016/0022-2836(77)90078-X.

[71] R. Jaenicke, G. Böhm, The stability of proteins in extreme environments, Curr. Opin. Struct. Biol. 8 (1998) 738–748. doi:10.1016/S0959-440X(98)80094-8.

[72] A. Razvi, J.M. Scholtz, Lessons in stability from thermophilic proteins, Protein Sci. 15 (2006) 1569–1578. doi:10.1110/ps.062130306.

[73] J.R. Livingstone, R.S. Spolar, M. Thomas Record, Contribution to the Thermodynamics of Protein Folding from the Reduction in Water-Accessible Nonpolar Surface Area, Biochemistry. 30 (1991) 4237–4244. doi:10.1021/bi00231a019.

[74] R.L. Baldwin, Energetics of Protein Folding, J. Mol. Biol. 371 (2007) 283–301. doi:10.1016/j.jmb.2007.05.078.

[75] G.I. Makhatadze, P.L. Privalov, Heat capacity of proteins. I. Partial molar heat capacity of individual amino acid residues in aqueous solution: hydration effect., J. Mol. Biol. 213 (1990) 375–384. doi:10.1016/S0022-2836(05)80197-4.

[76] S. Robic, M. Guzman-Casado, J.M. Sanchez-Ruiz, S. Marqusee, Role of residual structure in the unfolded state of a thermophilic protein, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 11345–11349. doi:10.1073/pnas.1635051100.

[77] W.T. Li, R.A. Grayling, K. Sandman, S. Edmondson, J.W. Shriver, J.N. Reeve, Thermodynamic stability of archaeal histones, Biochemistry. 37 (1998) 10563–10572. doi:10.1021/bi973006i.

[78] K.A. Luke, C.L. Higgins, P. Wittung-Stafshede, Thermodynamic stability and folding of proteins from hyperthermophilic organisms, FEBS J. 274 (2007) 4023–4033. doi:10.1111/j.1742-4658.2007.05955.x.

[79] S. Knapp, P.T. Mattson, P. Christova, K.D. Berndt, A. Karshikoff, M. Vihinen, C.I.

Smith, R. Ladenstein, Thermal unfolding of small proteins with SH3 domain folding pattern., Proteins. 31 (1998) 309–319. doi:10.1002/(sici)1097-0134(19980515)31:3<309::aid-prot7>3.0.co;2-d.

[80] R.B. Cooley, D.J. Arp, P.A. Karplus, Evolutionary Origin of a Secondary Structure: π-Helices as Cryptic but Widespread Insertional Variations of α-Helices That Enhance Protein Functionality, J. Mol. Biol. 404 (2010) 232–246. doi:10.1016/j.jmb.2010.09.034.

[81] K.O. Stetter, Hyperthermophiles in the history of life, Philos. Trans. R. Soc. B Biol. Sci. 361 (2006) 1837–1842. doi:10.1098/rstb.2006.1907.

[82] N.R. Pace, Origin of life-facing up to the physical setting., Cell. 65 (1991) 531–533. doi:10.1016/0092-8674(91)90082-a.

[83] D.W. Schwartzman, C.H. Lineweaver, The hyperthermophilic origin of life revisited., Biochem. Soc. Trans. 32 (2004) 168–171. doi:10.1042/bst0320168.

[84] S. Akanuma, Y. Nakajima, S.I. Yokobori, M. Kimura, N. Nemoto, T. Mase, K.I. Miyazono, M. Tanokura, A. Yamagishi, Experimental evidence for the thermophilicity of ancestral life, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) 11067–11072. doi:10.1073/pnas.1308215110.

[85] L.C. Wheeler, S.A. Lim, S. Marqusee, M.J. Harms, The thermostability and specificity of ancient proteins., Curr. Opin. Struct. Biol. 38 (2016) 37–43. doi:10.1016/j.sbi.2016.05.015.

[86] E.A. Gaucher, S. Govindarajan, O.K. Ganesh, Palaeotemperature trend for Precambrian life inferred from resurrected proteins, Nature. 451 (2008) 704–707. doi:10.1038/nature06510.