Structural characterization and directed evolution of a novel acetyl xylan esterase reveals thermostability determinants of the Carbohydrate Esterase 7 family

Running title: Structure of a novel carbohydrate esterase 7 enzyme

Fiyinfoluwa A. Adesioye,[a] Thulani P. Makhalanyane,[a] Surendra Vikram,[a] Bryan T. Sewell,[b] Wolf-Dieter Schubert,[c] Don A. Cowan [a] #

Centre for Microbial Ecology and Genomics, Department of Genetics, University of Pretoria, Pretoria, South Africa[a]; Institute of Infectious Disease and Molecular Medicine, University of Cape Town, South Africa[b]; Department of Biochemistry, University of Pretoria, Pretoria, South Africa[c]

#Address correspondence to Don A. Cowan, don.cowan@up.ac.za

ABSTRACT

A hot desert hypolith metagenomic DNA sequence dataset was screened *in-silico* for genes annotated as acetyl xylan esterases (AcXEs). One of the genes identified encoded a ~36 kDa protein (NaM1). The synthesised gene was cloned, expressed and the resulting protein, purified. NaM1 was optimally active at pH 8.5 and 30°C, and functionally stable at salt concentrations up to 5 M. The specific activity and catalytic efficiency were 488.9 Umg^{-1} and $3.26x10^{6}$ $M^{-1}s^{-1}$, respectively. The crystal structure of wild type NaM1 was solved at a resolution of 2.03 Å and a comparison with the structures and models of more thermostable carbohydrate esterase (CE) 7 family enzymes and variants of NaM1 from a directed evolution experiment, respectively, suggest that reduced side chain volume of protein core residues are relevant to the thermal stability of NaM1. Surprisingly, a single point mutation (N96S), not only

resulted in a simultaneous improvement in thermal stability and catalytic efficiency, but also increased the acyl moiety substrate range of NaM1.

IMPORTANCE

Acetyl xylan esterases (AcXEs) belong to nine carbohydrate esterase families (CE 1-7, 12 and 16), of which CE7 enzymes possess a unique and narrow specificity for acetylated substrates. All structurally characterised members of this family are moderately to highly thermostable. The crystal structure of a novel, mesophilic CE7 AcXE (NaM1), from a soil metagenome, provides a basis for comparisons with thermostable CE7 enzymes. Using error-prone polymerase chain reaction (PCR) and site-directed mutagenesis, we enhanced both the stability and activity of the mesophilic AcXE. With comparative structural analyses, we have also been able to identify possible thermal stability determinants. These are valuable for understanding the thermal stability of enzymes within this family and as a guide for future protein engineering of CE7 and other α/β hydrolase enzymes.

KEYWORDS: acetyl xylan esterase, carbohydrate esterase 7, x-ray crystallography, sequence-based metagenomics, directed evolution, thermal stability.

INTRODUCTION

Acetyl xylan esterases (AcXEs) are carbohydrate-active enzymes (CAZymes) that hydrolyse ester bonds to liberate acetic acid from acetylated hemicellulose - typically polymeric xylan and xylooligosaccharides. Although several AcXEs have been identified from culturable organisms, the discovery and characterization of new AcXEs using metagenomic methods avoids the limitations of culture-based techniques (1) and may aid the identification of enzymes with improved stability and activity. Several studies have reported the identification and expression of full length CAZyme-encoding open reading frames (ORFs) in metagenomic datasets (2, 3). The Namib Desert hypolith

metagenome, designated as a ‘hypolithome’ in this study, has also been shown to possess a large number of genes encoding cell wall–degrading enzymes (4-6). Members of the Carbohydrate esterase 7 enzymes are intracellular enzymes with a narrow acyl moiety specificity for acetates. Some members, however, are also active on butyrates. CE7 enzymes preferably deacetylate cephalosporin C (CPC), 7-aminocephalosporanic acid (7-ACA), simple acetylated sugars, aryl alcohols and acetylated xylooligosaccharides (AcXOS), rather than polymeric xylan (AX) (7).

Crystal structures for nearly a third of biochemically characterised AcXEs are available (8) including those from *Bacillus subtilis* (*Bs*CAH) (9-11), *B. pumilus* (*Bp*AXE CECT5027 and *Bp*AXE PS213) (12-15), *Thermotoga maritima* TM007 (*Tm*AcE) (16-19) and *Thermoanaerobacterium saccharolyticum*. JW/SL YS485 (*Ts*AcE) (20, 21), all of which are moderately (T_{opt} 45°C) to highly (T_{opt} 90°C) thermostable. High thermal stability is important for xylan deacetylation during industrial hemicellulose bioconversion processes, but not required for CPC or 7-ACA deacetylation processes (22). All characterised CE7 enzymes deacetylate CPC and/or 7-ACA, but not all deacetylate AX. *Bp*AXE PS213 (T_{opt} 45°C) deacetylates xylan while highly thermostable homologs, *Tm*AcE and *Ts*AcE do not. Hence, the natural substrates of *Tm*AcE and *Ts*AcE are not known (17, 21). A number of thermolabile CAZymes (23), including AcXEs (24), have been modified to create more thermostable variants for industrial lignocellulose deconstruction. Thermal stability (25, 26), substrate specificity (18, 27) and activity (17, 28, 29) determinants of CE7 AcXEs have been investigated. In *Tm*AcE, a β-loop interface (25) and an N-terminal extension (26) were identified as modulators of thermal stability and activity. Both elements are, however, also present in less thermostable CE7 enzymes, making it difficult to identify specific factors responsible for CE7 enzyme stability.

Directed evolution (DE), in particular error-prone PCR (EpPCR), is a powerful tool for optimizing enzyme traits (30). EpPCR (31) and site-directed mutagenesis (SDM) (32) have been used to improve CAZyme activity. DE remains the best method for improving thermal stability and k_{cat} of an enzyme despite structural and functional data being available for homologs (33).

Here, we describe the *in-silico* screening of a Namib Desert hypolith metagenomic dataset for novel AcXE-encoding genes by combining sequence-based and synthetic metagenomic approaches, together with the structural and functional characterization of a novel CE7 AcXE.

RESULTS

In-silico biomining

A Namib desert hypolith metagenomic dataset (>600 million bp) was screened for AcXEs of CE families 1 to 7 with 71 AcXE homologs from the CAZy database (34). Primary sequence hits were filtered for specific AcXE domain ORFs and complete N and/or C termini. This analysis identified three putative AcXE sequences, with AcXE1 or XynB-like AcXE conserved domains of CAZy families CE3 or CE7. The three putative AcXEs, designated *NaMet*1 (CE7), *NaMet*2 (CE7) and *NaMet*3 (CE3), had 64, 69 and 59% sequence identities, respectively, to known AcXE sequences from actinobacteria (Table **S1**) and distinct domain arrangements. AXE1 domains of NaM1 and NaM2 were encoded by nucleotides 19-966 and 19-972 of *NaMet1* and *NaMet2*, respectively, and shared 46.5% sequence identity. *NaMet1* and *NaMet2* were located on separate contigs. The NaM1, NaM2 and NaM3 protein sequences all exhibited a Ser-His-Asp(Glu) catalytic triad and a GXSXG or GDS(L) motifs typical of CE7 or CE3 families, respectively. The NaM3 sequence contained catalytic residues typical of SGNH hydrolases: Ser49, Gly106, Asn163 and His225 (Figure **S2a**). The proteins, NaM1, NaM2 and NaM3, had isoelectric point (pI) values of 5.1, 5.5 and 5.8 and molecular weight (MW) values of 35.6, 35.9 and 26.0 kDa, respectively. Residues 1-32 of NaM3 encoded a signal peptide, while no signal peptide sequences were identified for NaM1 and NaM2 implying intracellular localization. Catalytic residues of NaM1/NaM2 corresponded to Ser185/187, His304/307 and Asp275/273 (Figure **S2b**).

Cloning, expression and purification

Synthetic genes were amplified (primers in Table **S3**) and gene sizes of 966, 990 and 735 bp for *NaMet*1 to 3 confirmed by agarose gel electrophoresis. Expression in *Escherichia coli* BL21 DE cells produced proteins of ~35.6 kDa (NaM1), 35.9 kDa (NaM2) and 26 kDa (NaM3)(Figure **S4**) with optimal production at 25°C for NaM1 and NaM2, and 18°C for NaM3. Clear zones in tributyrin agar confirmed that the expressed NaM1 protein was active while NaM2 and NaM3 showed no tributyrin-hydrolysing activity. Consequently, NaM1 was selected for further biochemical and structural characterization. IMAC purification of NaM1 culture extract yielded >95% pure protein (Figure **S4**, inset).

Functional Characterisation

***pH optimum and stability*:** NaM1 was most stable at pH 8 and showed optimal activity at pH 8.5 (Figure 1a).

***Temperature optima and stability*:** NaM1 was optimally active at 30°C without NaCl (Figure 1b), but at 35°C with 1 M NaCl (Figure **S5a**). It retained 10% of initial activity after incubation at 40°C for 1h, but 33% activity at 50°C with 1M NaCl over a similar period. All residual activity was lost after incubation at 55°C for 1h (Figures 1c and **S5b**).

Effects of solutes: Thermal stability and activity of NaM1 was improved with increasing NaCl concentration peaking at 1 M NaCl and retaining ~50% activity in 5 M NaCl at 40°C after 1 h (Figure **S6a**). NaM1 activity and thermal stability were not affected by BSA or other stabilising solutes (results not shown) – other than NaCl and trehalose (Figure **S6b**).

Error-prone PCR, cloning, library analysis and screening

DNA sequencing after EpPCR revealed one or two mutations per clone, without any obvious bias of transitions over transversions (30) being observed. Of 3100 variant enzymes screened for thermal stability, 11 were putative thermostable mutants (4% of tested and 5% of active mutants).

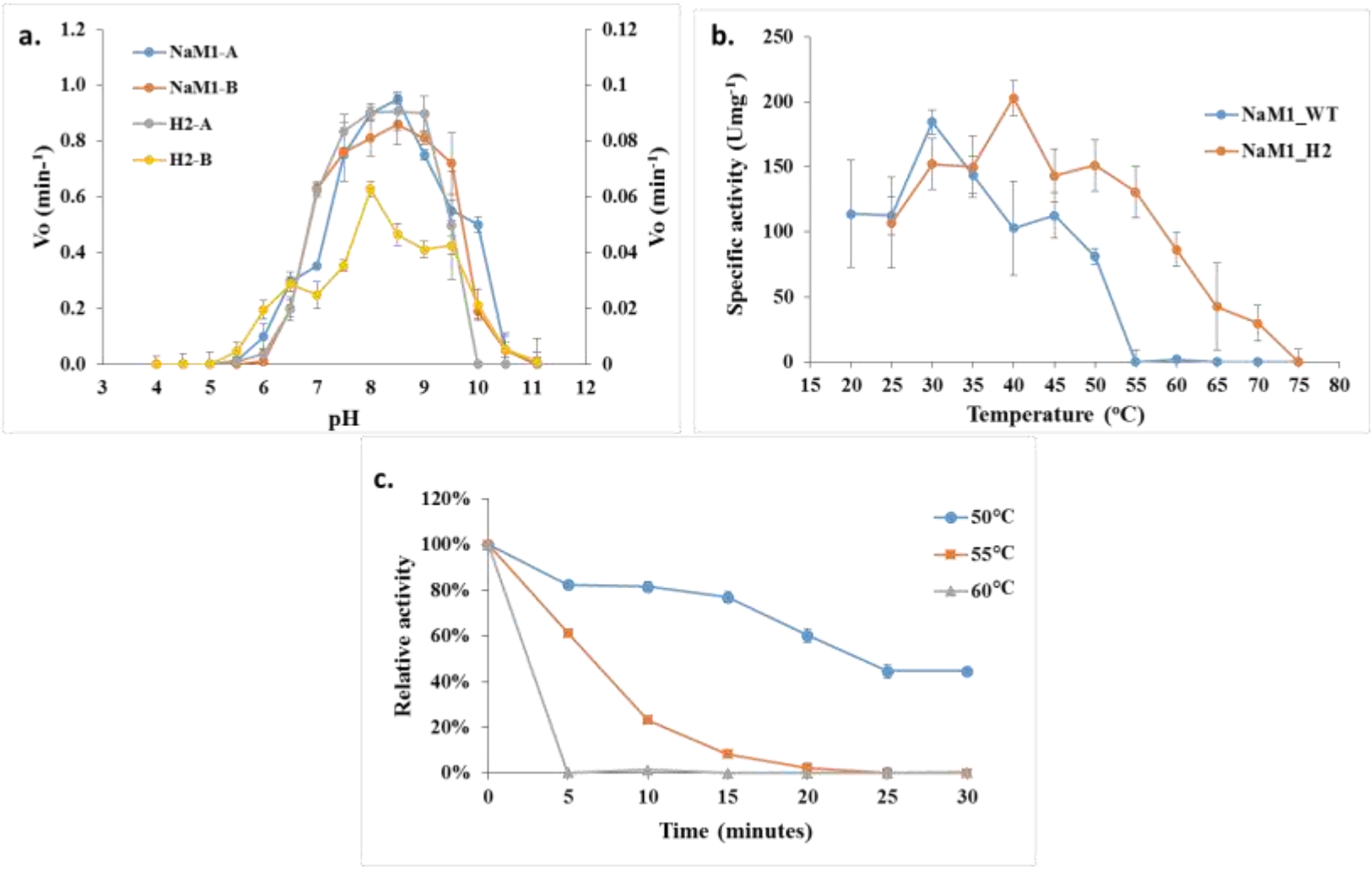

Figure 1: a. pH activity (A) and stability (B) profiles of $NaM1_{WT}$ (primary y-axis) and $NaM1_{H2}$ (secondary y-axis). **b.** Thermal optima for $NaM1_{WT}$ and $NaM1_{H2}$ shown at 30 and 40°C, respectively. **c.** Thermal stability profile of $NaM1_{WT}$.

Sequence and structure to function analyses of mutants

NaM1 variants with increased thermostability and reduced k_{cat} were all derived from transitions (Table **S7**). Enzyme activity and thermostability are often inversely related (35, 36), but mutations improving both are possible (33). The variants, $NaM1_{H2}$, $NaM1_{D8}$ and $NaM1_{B4}$, each contained two point mutations, one of which was silent in both $NaM1_{D8}$ and $NaM1_{B4}$, leaving only $NaM1_{H2}$ with two amino acid substitutions: N96S and F210L. Figure **S8** shows the location of substitutions.

Thermal stability and inactivation assays

Six NaM1 putative thermostable variants were identified as false positives by thermal stability assays. Of the remaining five variants, $NaM1_{E1}$ and $NaM1_{F9}$ retained wild-type thermal stabilities but had higher catalytic activity, while $NaM1_{B4}$, $NaM1_{D8}$ and $NaM1_{H2}$ were more thermostable but less active than $NaM1_{WT}$ (Figures 2 and 3a). Thermal inactivation assays on cell-free extracts of the five mutant clones revealed that $NaM1_{H2}$ retained 100% activity after 20 min at 50 and 55°C (Figures 2a and b). The half-life of purified $NaM1_{H2}$ at 65°C was 23 min (Figure 3b), equivalent to that of $NaM1_{WT}$ at 50°C (Figure 1c). As the most thermally stable mutant generated, $NaM1_{H2}$ was used in all further analyses.

Site-Directed Mutagenesis

Amino acid substitutions yielding the enhanced stability of $NaM1_{H2}$ were investigated by SDM. Substitutions N96S and F210L of $NaM1_{H2}$ resulted from point mutations A293G and T634C in *NaMet1*. The mutations were separately introduced into *NaMet1* by SDM and the variant proteins designated $NaM1_{N96S}$ and $NaM1_{F210L}$. Incubating purified $NaM1_{N96S}$ and $NaM1_{F210L}$ at 50°C for 20 min gave 50% and 20% residual activity, respectively, compared to 100% for $NaM1_{H2}$ (Figure 2a). Both variants were, thus, substantially less thermostable than $NaM1_{H2}$, but more thermostable than the WT enzyme, implying that both substitutions contributed to the increased thermal stability of $NaM1_{H2}$.

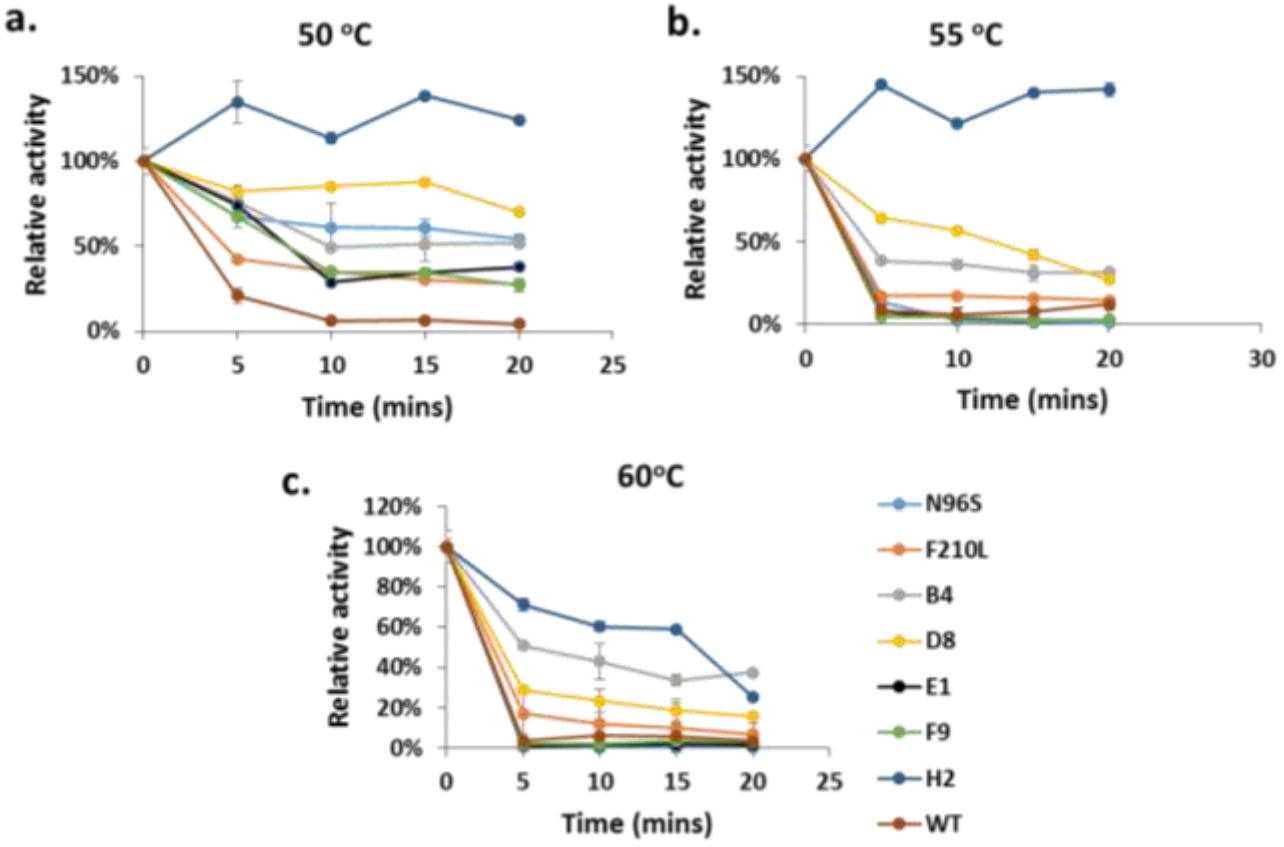

Figure 2: Thermal stability profiles of NaM1 and its variants using their cell free extracts.

Substrate specificity and enzyme kinetics

NaM1 activity, specific activity and kinetic constants were determined for substrates *p*-NPA, *p*-NPB, 4-MUA, 2-NA, 7-ACA, *p*-NPO and *p*-NPP (Table 1). No detectable activity was observed for *p*-NPO and *p*-NPP, while p-NPA yielded the highest specific activity (488.9 Umg^{-1}), lowest K_M (0.1 mM), and highest catalytic efficiency ($3.3x10^6$ M^{-1} s^{-1}). The lowest activity was observed for AX.

The temperature and pH optima for $NaM1_{H2}$ were 40°C (Figure 1b) and 7.5-8.5 (Figure 1a), respectively. $NaM1_{F210L}$ and $NaM1_{H2}$ had the lowest catalytic efficiencies (k_{cat}/K_M) on *p*-NPA and retained only ~20 and 30% activity of $NaM1_{WT}$, respectively, while the activity of $NaM1_{N96S}$ was >30% higher than that of $NaM1_{WT}$ (Table 2, Figure 3c). The catalytic efficiency of $NaM1_{WT}$ was lower at 40°C than at 25°C, whereas $NaM1_{N96S}$, at 40°C, had approximately the same efficiency as NaM1 at 25°C (Tables 1 and 2).

Table 1: Substrate specificity and enzyme kinetics of $NaM1_{WT}$ and $_{H2}$

Substrate	NaM1	V_{max} (Uml^{-1})	Specific activity (Umg^{-1})	K_M (mM)	k_{cat} (s^{-1})	Cat. Eff. ($M^{-1}s^{-1}$)
***p*-NPA**	WT	0.88 ± 0.016	488.9 ± 16.71	0.1 ± 0.015	293.3 ± 10.03	3.26×10^6
	H2	1.15 ± 0.11	348.25 ± 34.17	0.76 ± 0.22	205.96 ± 20.21	2.72×10^5
***p*-NPB**	WT	0.05 ± 0.005	12.96 ± 1.41	0.7 ± 0.14	7.67 ± 0.83	1.10×10^4
	H2	0.1 ± 0.01	20.46 ± 2.23	1.34 ± 0.27	12.18 ± 1.33	9.09×10^3
***p*-NPO**	WT	-	-	-	-	-
	H2	0.01 ± 0.003	1.86 ± 0.56	0.19 ± 0.25	1.11 ± 0.33	5.83×10^3
4-MUA	WT	0.5 ± 0.01	277.8 ± 1.89	0.13 ± 0.02	166.7± 1.13	1.28×10^6
	H2	1.06 ± 0.06	211.67 ± 12.65	0.47 ± 0.08	125.99 ± 7.53	2.68×10^5
2-NA	WT	0.4 ± 0.001	222.2 ± 12.72	0.2 ± 0.03	133.3 ± 7.63	6.67×10^5
	H2	0.39 ± 0.05	78.51 ± 9.30	0.41 ± 0.14	46.73 ± 5.54	1.14×10^5
7-ACA	WT	0.8 ± 0.02	200 ± 0.038	0.46 ± 0.05	120 ± 12.53	2.60×10^5
	H2	5.84 ± 1.84	106.25 ± 33.40	0.51 ± 0.39	62.83 ± 19.75	1.22×10^2
0.5% AX	WT	0.24 ± 0.08	6.05 ± 1.99	-	-	-
	H2	ND	-	-	-	-

*One enzyme unit (U) is the amount of enzyme that releases 1 µmol of product from substrate per minute under standard assay conditions. ND – not detected.

Table 2: Comparison of the kinetics of $NaM1_{H2}$ with $NaM1_{N96S}$, $NaM1_{F210L}$ and $NaM1_{WT}$ at 40°C on *p*-NPA.

NaM1 Variant	V_{max} (Uml^{-1})*	Specific activity (Umg^{-1})*	K_M (mM)	k_{cat} (s^{-1})	Catalytic Efficiency ($M^{-1}s^{-1}$)
‡**H2**	1.15 ± 0.11	348.25 ± 34.17	0.76 ± 0.22	205.96 ± 20.21	2.72×10^5
N96S	3.15 ± 0.19	955.39 ± 59.02	0.42 ± 0.10	565.02 ± 34.90	1.34×10^6
F210L	0.54 ± 0.05	162.17 ± 15.53	0.63 ± 0.17	95.91 ± 9.18	1.52×10^5
WT	2.75 ± 0.34	832.36 ± 102.50	0.58 ± 0.21	492.26 ± 60.62	8.43×10^5

*One enzyme unit (U) is the amount of enzyme that releases 1 µmol of product from substrate per minute under standard assay conditions.

‡$NaM1_{H2}$ kinetic data on *p*-NPA is same as was described in Table 1.

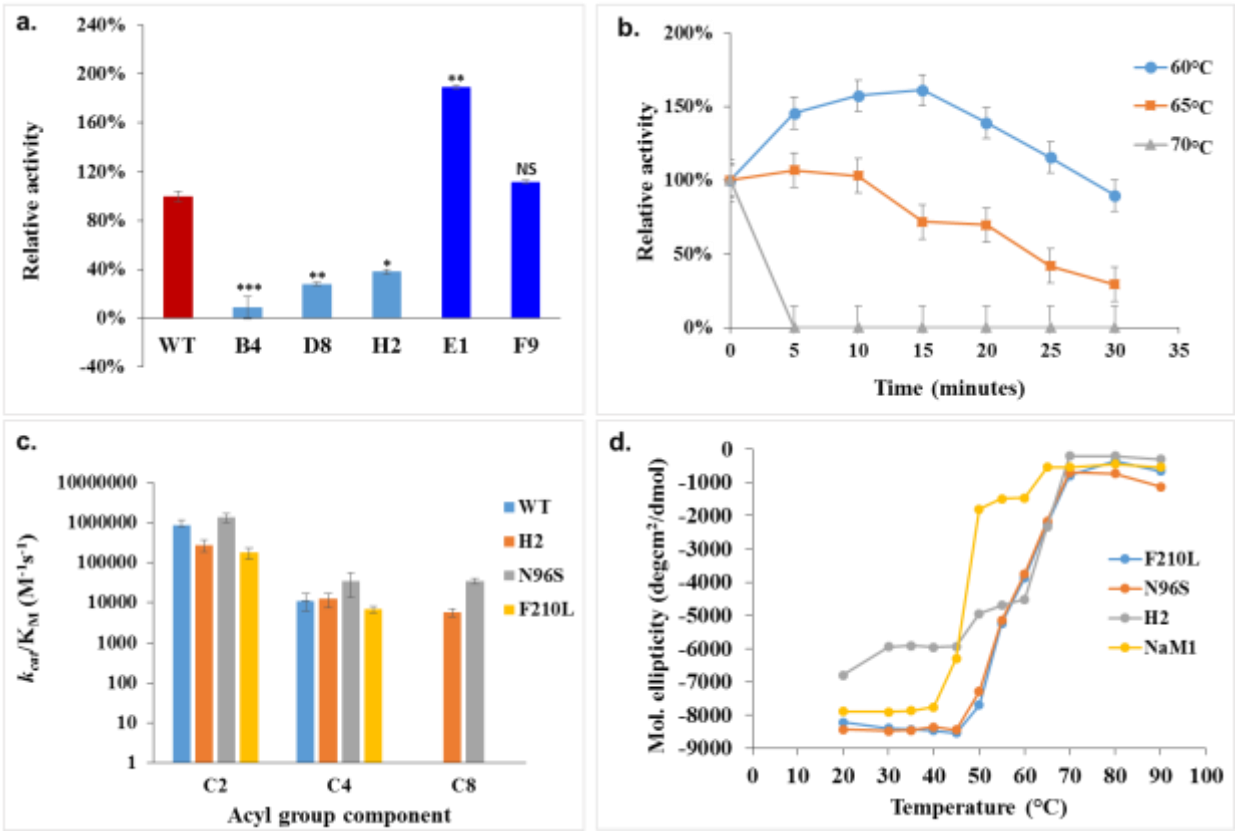

Figure 3: a. Deacetylase activity of selected NaM1 variants relative to the wild type (*P<0.01, **P<0.001, ***P<0.0001, NS – not statistically significant). **b.** Thermal stability profile of $NaM1_{H2}$. **c.** Catalytic efficiencies of $NaM1_{WT}$, $NaM1_{H2}$, $NaM1_{N96S}$ and $NaM1_{F210L}$ on substrates with acyl moiety length as indicated. **d.** Thermal unfolding profiles (CD spectroscopy at 210 nm) of $NaM1_{WT}$, $NaM1_{H2}$, $NaM1_{N96S}$ and $NaM1_{F210L}$. Inferred melting temperatures are around 47, 63, 52 and 52°C, respectively.

Thermal unfolding of NaM1 and variant proteins

CD spectra of $NaM1_{WT}$ and $NaM1_{H2}$ at 25°C were indistinguishable (Figure **S9**), implying that the amino acid substitution in $NaM1_{H2}$ had minimal effect on the overall fold. Plotting ellipticity at 210 nm against temperature indicated a melting temperature (T_m) of 47°C and an unfolding transition at 45°C for $NaM1_{WT}$. These were shifted to 63 and 60°C, respectively, for $NaM1_{H2}$ (Figure 3d). $NaM1_{N96S}$ and $NaM1_{F210L}$ initiated thermal unfolding at 50°C, had a melting temperature of ~52°C and lost all secondary structures at 70°C (Figure 3d). These data suggest that both substitutions contributed to the thermal stability of $NaM1_{H2}$.

Crystallization, data collection and structure determination

Single crystals of NaM1 were obtained in 0.1 M 2-(N-morpholino) ethanesulfonic acid (MES) buffers pH 8.5 and 25% PEG 8000 (w/v). X-ray diffraction data to 2.0 Å resolution were collected and the crystal structure of NaM1 was solved by molecular replacement, locating one NaM1 hexamer per asymmetric unit of the orthorhombic space group with a solvent content of 47% (Table **3**). For NaM1 components, see Table **S10**. A single oligomeric species of NaM1 was observed by size exclusion chromatography and the hexameric nature of NaM1 was confirmed by non-denaturing and SDS-PAGE (Figure **S11**). These showed the MW of the oligomer (native NaM1) to be ~216 kDA, six times the MW of NaM1.

Residues Arg56 to Gly320 of NaM1 adopted a typical α/β hydrolase fold, dominated by a central nine-stranded β-sheet arranged in sequence with strands 4 and 5 exchanged, and β-strands 1235 running antiparallel, but strands 546789, parallel (Figures **S12a** and **c**). NaM1 exhibited the α/β-hydrolase motif GXSXG (GYSQG) in loop β6-α5. Like other CE7 enzymes, NaM1 deviated from the canonical α/β hydrolase fold by an N-terminal extension (α-helices α1 and 2, Phe9 to Ala38 and β-strand β1, Val44 to Pro47), a three-helix insertion (α6, α7 and α8, Phe214 to Val254), and an additional interface loop region (Gly120 to Leu140) (25).

Table 3: Summary of data collection and structure solution parameters

Data collection	
*Resolution range (Å):	89.23 – 2.03 (2.10 - 2.03)
Space Group:	$P2_12_12_1$
Unit cell parameters: a, b, c (Å), α, β, γ (°)	107.7, 116.8, 159.4, 90, 90, 90
*Completeness (%)	99.94 (99.95)
*I/σI	10.02 (2.24)
*R_{merge}	0.07 (0.35)
Mosaicity (°)	0.02
Refinement	
Solution method	Molecular replacement
*Refinement resolution range	89.32 - 2.03 (2.05-2.03)
*R-free / R-work (%)	22.48 (31.97) / 17.22 (26.84)
RMSD: angles (°) and bonds (Å)	0.89 and 0.007
Average and Wilson B-factors ($Å^2$)	20.90 and 18.29
Number of chains, residues, ligands, atoms, waters	6, 1932, 27, 17600, 2238
Ramachandran favoured, allowed, outliers (%)	96.67, 2.92, 0.42

*shell of highest resolution in brackets

The NaM1 hexamer has a D3 point group symmetry and consists of two interdigitating trimers, involving monomers A, B, C and D, E, F, respectively (Figure **S12b**). The hexamer encloses a central void accessed through pores around the three-fold rotational axis on either side. The combined surface area of the six NaM1 monomers is 64900 $Å^2$ of which 21430 $Å^2$ is buried in the hexamer. Each NaM1 monomer directly interacts with the two monomers from the same trimer and with two monomers from the opposite trimer. Thus, monomer A interacts with B and C, as well as D and E (Table **S13**). The A-D interface (equivalent to B-F and C-E), covering a surface area of about 837 $Å^2$, represents the largest monomer-monomer interface and involves two salt bridges, 12 hydrogen bonds and ~140 van der

Waals interactions. Surprisingly, a deletion mutant *Tm*AcEΔ26 of *Tm*AcE identified these interactions as non-essential for oligomer formation (25). The A-B subunit interface (repeated in B-C, C-A, D-E, E-F and F-D) involves six salt bridges, six H-bonds and ~50 van der Waals interactions (Table **S13** and Figure 4a). The A-E interface (equal to B-D and C-F) covers 753 $Å^2$ and involves 12 hydrogen bonds, no salt bridge and >130 van-der-Waals interactions. MES atoms were also found to be involved in inter-subunit interactions (Figure 4b)

Active site and oxyanion hole

The active sites of all NaM1 monomer face the central space of the NaM1 hexamer (Figure **S12b**). The catalytic triad consists of Ser185, Asp275 and His304. Ser185 is located towards the end of a concave substrate binding pocket that extends to the S2 pocket (Figure 4c) accommodating the substrate acyl moiety (17, 18). His304 bridges Ser185 and Asp275 (Figure 4d). In the presence of substrate, Asp275 polarizes His304, to deprotonate Ser185 allowing the latter attack the carbonyl carbon of the substrate ester to initiate bond cleavage (17, 37). Residues Phe210, Pro226, Ile277 and Cys278 provide a hydrophobic environment around the active site (15). Active site residues indirectly interact with adjacent monomers, implying that the quaternary structure of NaM1 is critical for substrate binding and catalysis. His304 of subunit A (His304A) hydrogen-bonds Glu305A, which in turn forms a salt bridge with Arg235D in α-helix α7 (Table **S13** and Figure 4a) . Helix α7 is part of the triple helix insertion of

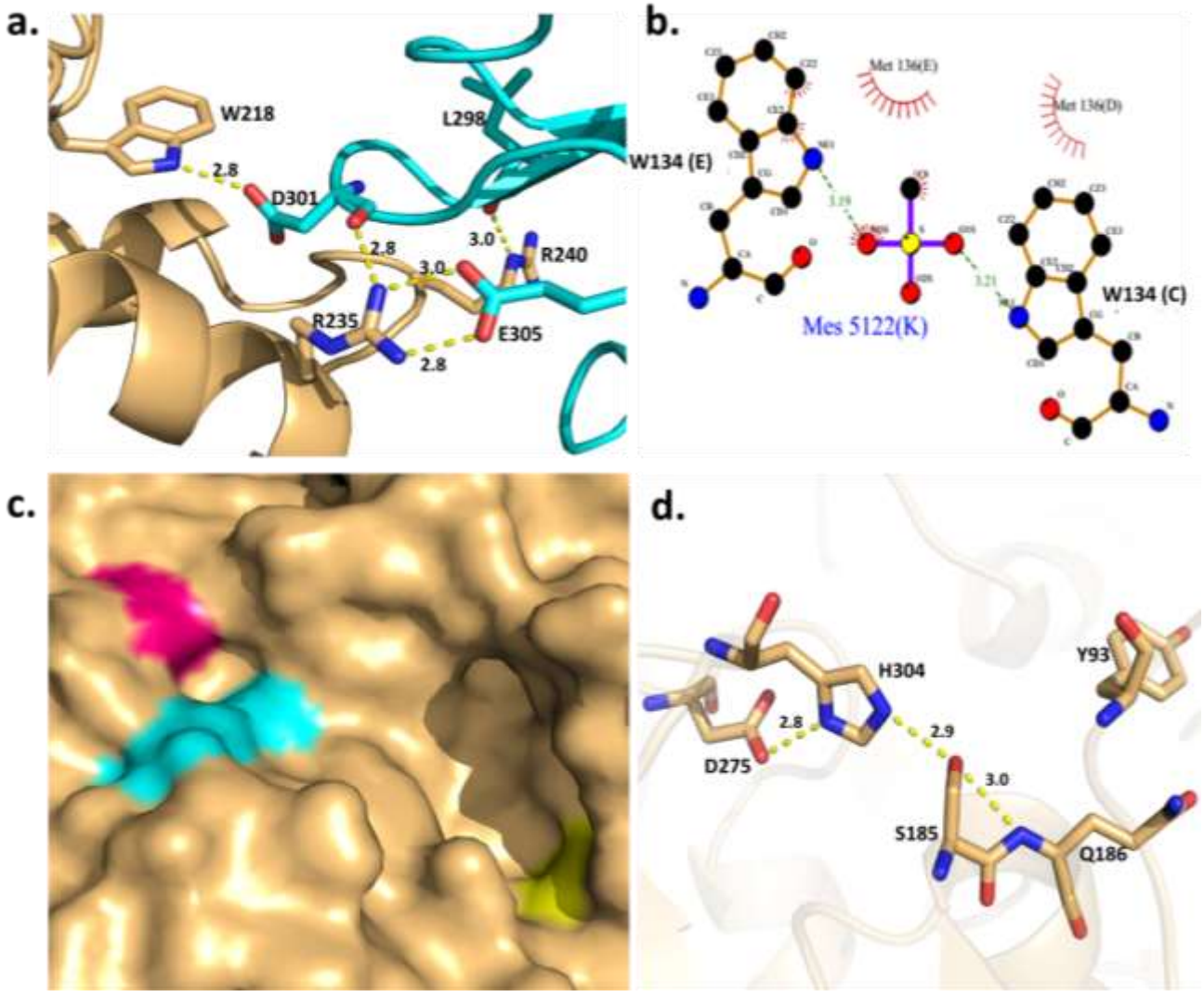

Figure 4: a. Interactions between subunits A (brown) and B (cyan). Interacting residues are represented as sticks. **b.** **T**rimer-trimer interactions between subunits D and E facilitated by MES. Non-bonded interactions are represented by brushed lines (red). Figure was generated using Ligplot (61). **c.** NaM1 substrate-binding cavity: Tyr106 (yellow) and Pro226 (pink) mark the proximal and distal ends. Catalytic Ser185 is shown in cyan. **d.** Directed interactions for NaM1 catalytic and oxyanion hole residues. All distances are measured in Å.

NaM1 confirming its likely importance in oligomerization and catalysis (15, 17). The NaM1 insertion residue (Pro226) nearest the active site is 7.3 Å from His304 and delimits the substrate-binding site at one end, while the tryptophan, typically delimiting the other end in CE7 enzymes, is replaced by a tyrosine in NaM1 (Figure 4c) and *Ts*ACE. Laterally, the binding site is delimited by α-helix α3 and loop α3-β4 on the one side and the loop β9-α11 on the other separated by 17.1 Å. The "floor" of the

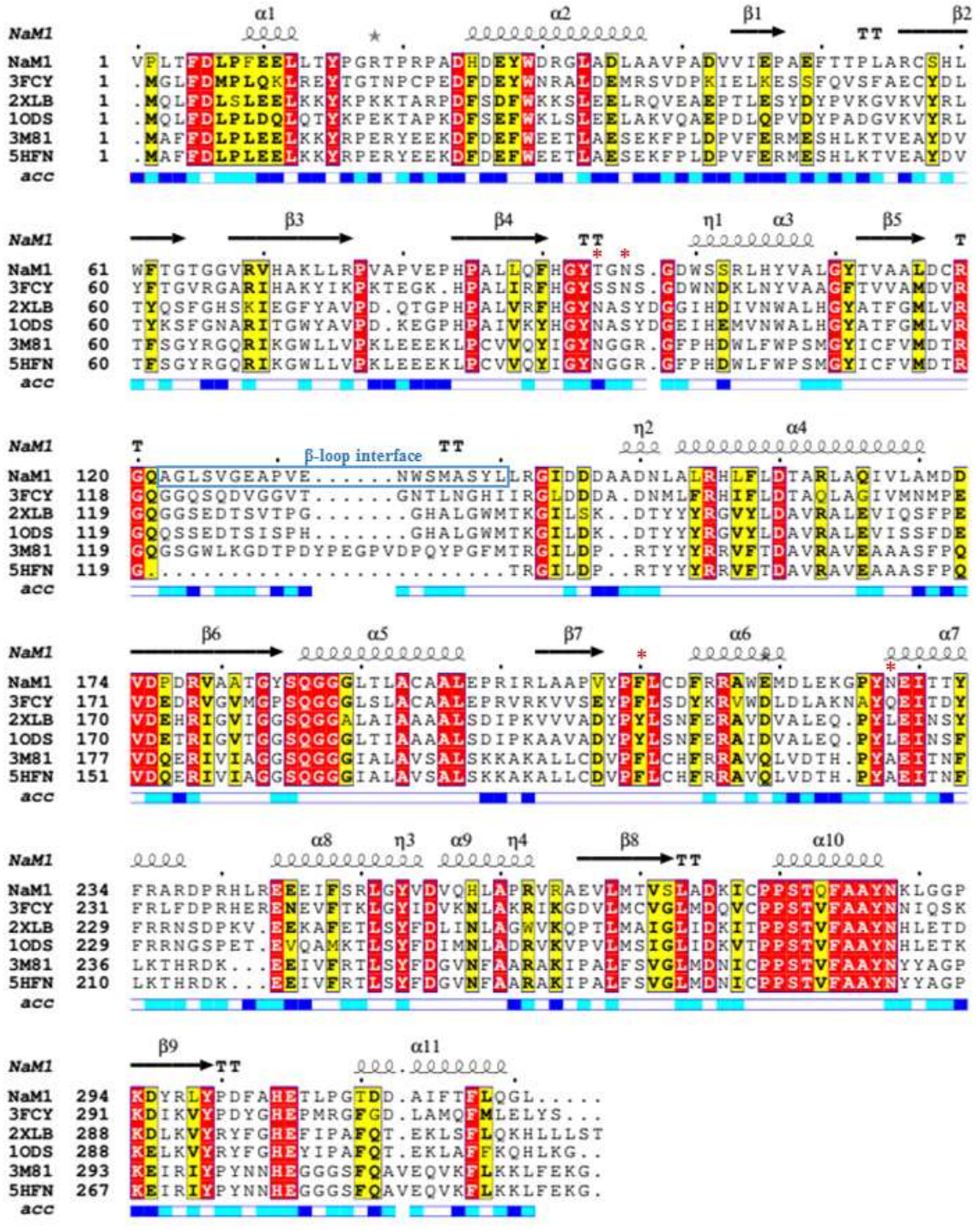

Figure 5: Structure-based alignment of NaM1 with structural homologs from the CE7 family generated using Multalin (66) and EsPript 3.0 (67) software. Black dots and red asterisks above sequences denote multiples of 10 and the locations of thermostabilizing substitutions respectively. Strictly and moderately conserved residues are highlighted in red and yellow, α-Helices and β-strands as coils and arrows; and Strict α and β-turns as TTT and TT, respectively. Relative accessibilities of residues are indicated as accessible (blue), intermediate (cyan) or buried (white).

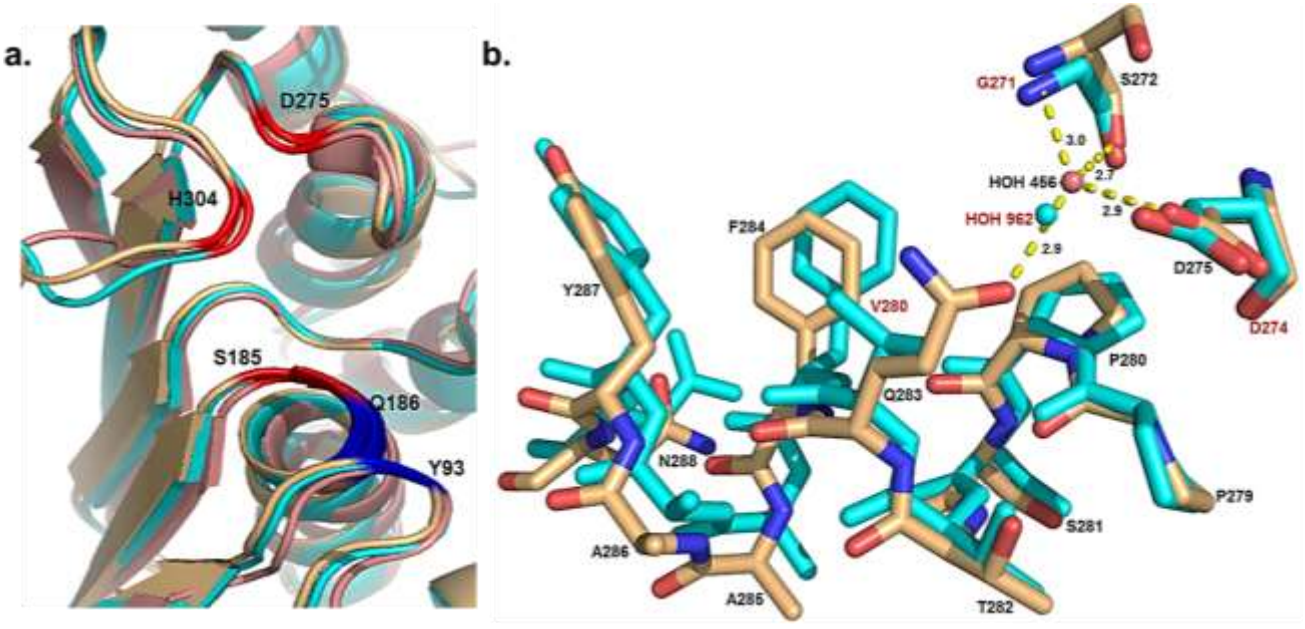

Figure 6: a. Superposition of NaM1 (light orange) with thermostable CE7 enzymes from *Thermoanaerobacterium* sp. (RMSD 1.2, PDB: 3FCY, cyan) and *Thermotoga maritima* (RMSD 2.1, PDB: 3M81, salmon) revealing conserved folds. Active site and oxyanion hole residues are coloured red and blue, respectively. **b.** Hydrophobic valine in the strictly conserved PPST**V**FAAYN motif of *Tm*AcE (cyan) is replaced by polar Gln283 in NaM1 (light orange) creating water-mediated hydrogen bonds to catalytic Asp275. Residues labelled in red are unique to *Tm*AcE. Interactions shown are between NaM1 residues and are measured in Å.

binding site is defined by the N-terminus of α5 and β-strands, β6 and β5. The oxyanion hole is located between the β6-α5 turn and the adjacent loop, involving the backbone amide groups of conserved Tyr93 and Gln186.

Comparison with other CE7 esterase structures

*Ts*AcE (PDB code - 3FCY), *Bp*AXE (3FVT and 2XLB), *Tm*AcE (3M81) and *Bs*CAH (1ODS) are structural homologs to NaM1 (Table **S14**). Alignment of NaM1 with CE7 homolog structures revealed low sequence (Figure 5), but high structural (Figures 6a and **S15**) conservation within the family.

Removing the β-interface loop in *Tm*AcE (residues 120 to 145) negatively affected thermal and conformational stability, as well as activity (25). In NaM1, the loop (residues 121 to 140) is six residues shorter (Figure 5) possibly affecting stability and activity. Some residues of this loop,

conserved in thermostable enzymes, are replaced in NaM1: two glycines are replaced by Ala122 and Ser138 in NaM1 and asparagine, lysine and valine are replaced by His258, Arg264 and Gln283, respectively (Figure 5).

DISCUSSION

In-silico mining

In sequence-based metagenomic mining for functional genes, ambiguous hits may be excluded through strict screening selections. Despite stringent protocols being employed in this study, a large number of initial hits for putative AcXEs proved to be false positives. The three putative AcXEs selected showed a reasonable degree of sequence novelty, with sequence identities of below 70%. The similarity of the selected AcXE enzymes to actinobacterial homologs (Table **S1**) is not surprising as this phylum is common to arid soils and hypolithons (5, 6).

Functional characterization

Of three isolated enzymes only NaM1 was active, mirroring similar observations of low folding success in published metagenomic screening experiments (2, 3) . Most characterised CE7 enzymes have moderate to high temperature 'optima' (45 to 90°C) while NaM1, surprisingly, had a temperature 'optimum' of only 30°C, despite the hot desert soil origins of the sample. This may be attributed to an adaptation of the enzyme to the refuge niche of the hypolithon (38).

NaM1 showed a preference for short chain fatty acid esters with the highest affinity for C2 and C4 substrates and no detectable activity on long chain (≥C8) fatty acid substrates. NaM1 displayed 7-ACA deacetylase activity which was comparable or higher than that for other CE7 enzymes (Table **S16**). NaM1 was least active on AX, in line with the low xylan deacetylation rates of other CE7 AcXEs (7).

Thermostability of CE7 family enzymes

The structure of NaM1 was compared with more thermostable CE7 enzymes to identify possible determinants of thermal stability. Most notably, a hydrophobic valine in the strictly conserved

PPST**V**FAAYN motif is replaced, in NaM1, with a polar glutamine residue (Figure 5). This Gln283 (α-helix α10) interacts with active-site residue Asp275 via a water-mediated H-bond (Figure 6b), an interaction, is impossible with a valine and possibly impacts NaM1 active site hydrophobicity and enzyme stability. Hydrophobic cores contribute to protein thermal stability such that meso- to thermophilic proteins often replace uncharged polar residues in less stable proteins by non-polar residues (39). The larger size of glutamine may result in less compact packing in the core of NaM1, potentially impairing thermal stability (40), although, the increased entropy may also be critical. Furthermore, NaM1 has fewer residues than the more thermostable *Tm*AcE, but a 3% larger surface area (17, 25). This observation is consistent with the protein compactness which is strongly positively correlated with thermostability (40).

Improving thermostability through point mutations often reduces enzyme activity (35). Correspondingly, all, but one ($NaM1_{N96S}$), of the thermostabilised NaM1 variants had reduced specific activity. For variant $NaM1_{H2}$, the optimal temperature (40°C) was increased by 10°C relative to the $NaM1_{WT}$, but with a 70% reduction in specific activity. The highest substrate affinity, K_M 0.42mM for *p*-NPA was observed for $NaM1_{N96S}$ at 40°C. The catalytic efficiency of $NaM1_{F210L}$ was half that of $NaM1_{H2}$, while that of $NaM1_{N96S}$ was 5-fold higher than $NaM1_{H2}$. Thus, the two substitutions had opposing effects on $NaM1_{H2}$ activity, with F210L decreasing and N96S increasing catalytic efficiencies. The improved stability with reduced activity observed in $NaM1_{H2}$ supports the notion that improving both thermostability and activity may require multiple substitutions throughout the protein (33).

Variant F210L

Phe210 of NaM1 is conserved as either tyrosine or phenylalanine in all CE7 esterases. In $NaM1_{WT}$, Phe210-N forms a main chain hydrogen bond to catalytic Ser185-O and this interaction is retained by Leu210 in $NaM1_{H2}$. However, while the Phe210 side chain faces away from Ser185 in $NaM1_{WT}$, modelling of $NaM1_{H2}$ indicated that the Leu210 side-chain faces Ser185, creating an unfavourable

interaction that interferes with substrate positioning and may prevent the Ser185 O^{γ} nucleophilic attack on the substrate. F210L also appears to increase the hydrogen bond length between catalytic-triad residues His304-N^{δ} and Asp275-O^{δ} from 3.1 to 3.4 Å, affecting deprotonation of the nucleophile Ser185 (Figure 7a) and reducing catalytic efficiency of both $NaM1_{H2}$ and $NaM1_{F210L}$.

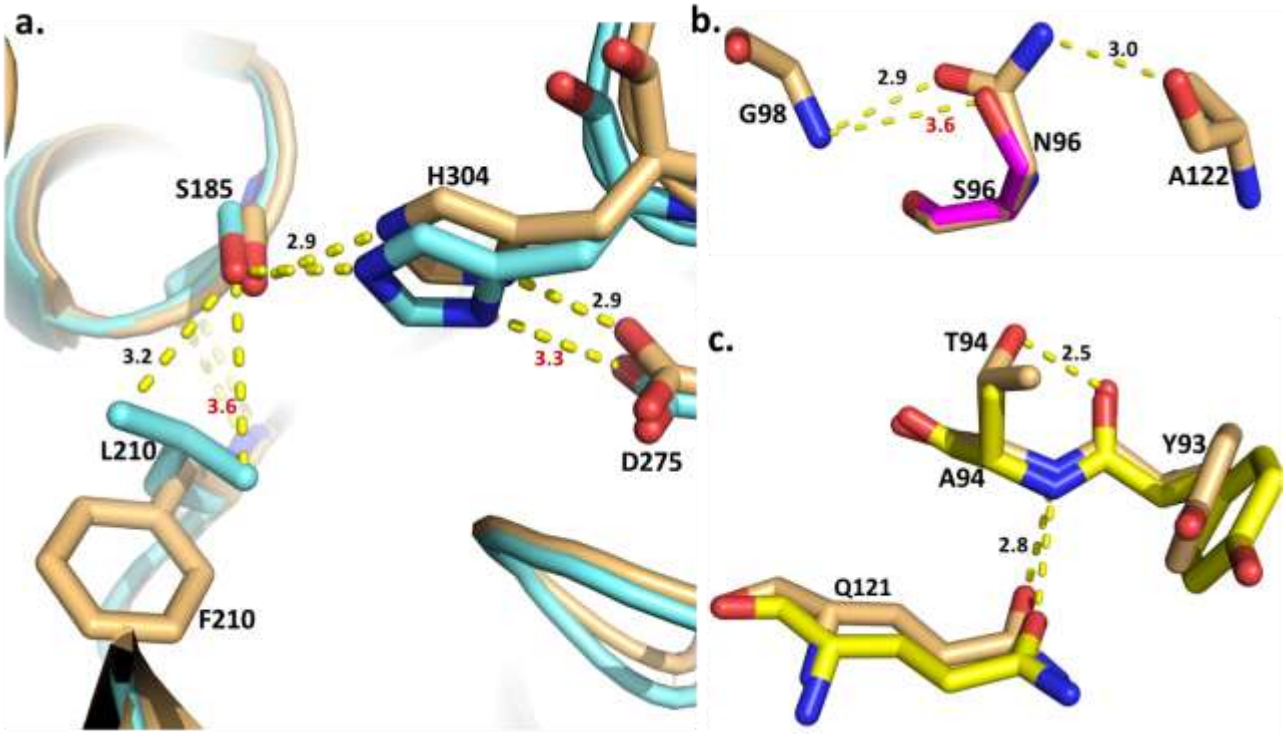

Figure 7: Possible structural effects of substitutions in thermostable NaM1 variants. **a.** F210L: Creation of unfavourable interactions between Leu210-$C^{\delta 1/2}$ and catalytic Ser185-$O^{\gamma 1}$, and increased distance between catalytic His304 and Asp275. **b.** N96S: Loss of H-bond with G98 and A122. **c.** T94A: Loss of H-bond, Thr94-$O_{\gamma 1}$ to Tyr93-O. NaM1 is shown in light brown colour while each mutant is represented in other colour. Extended H-bonds are labelled in red. All distances are measured in Å.

However, the F210L mutation may have improved protein compactness by promoting hydrophobic packing in the core, which might be expected to increase thermal stability in $NaM1_{F210L}$.

In analogy to *Tm*AcE, residues Phe210, Pro226 and Ile277 of NaM1 are likely to constrain acyl group size and orientation in the S2 binding site (19, 27). *Bp*AXE CECT5072 with a tyrosine in place of

Phe210 does not metabolise *p*-NP butyrate (15) while NaM1, with less bulky Phe210, accommodates substrates with up to C4 acyl groups (17, 21).

$NaM1_{WT}$ was less active on *p*-NPB than $NaM1_{H2}$, but was inactive on >C4 substrates (Figure 3d) while $NaM1_{H2}$ was active on the C8 substrate, *p*-NPO. Since Pro226 and Ile277 were unchanged in $NaM1_{H2}$, differences in substrate preferences were thought to be due to the F210L substitution. The smaller volume of leucine could potentially accommodate longer acyl chains in the active site.

Variant N96S

Asn96 is located on the same loop as oxyanion hole-forming residue, Tyr93, and forms stabilising H-bonds with Ala 122. Ala122 is located in the β-interface loop (Gly120 to Leu 140), which is important for thermal stability and activity in the *T. maritima* enzyme (25). Although an N96S substitution deletes the stabilising H-bonds Asn96-O^{δ} to Gly98-N and Asn96-$N^{\delta 2}$ to Ala122-O (Figure 7b), thermal stability was improved. This may have resulted from improved NaM1 packing, possibly due to the replacement with the smaller serine side-chain, as observed for an Asn to Ser substitution in a thermostable variant of a thermolabile subtilisin E (35). It is not clear why this mutation resulted in improved activity.

Surprisingly, $NaM1_{N96S}$ cleaved *p*-NPO in contrast to $NaM1_{WT}$ and $NaM1_{F210L}$, confirming that the N96S substitution, rather than F210L, facilitated the change in substrate specificity. Therefore, the N96S substitution not only improved thermal stability and catalytic efficiency, but also expanded the substrate specificity of NaM1.

Variant T94A

Position 94 in NaM1 is invariably occupied by hydrophilic serine, threonine or asparagine in CE7 enzymes (Figure 5). Replacing polar Thr94 with non-polar and smaller alanine (T94A) in $NaM1_{D8}$, presumably improved side-chain packing and hydrophobicity, increasing thermostability. The increased hydrophobicity may have resulted in reduced affinity for the intermediates of hydrophilic substrates, such as *p*-NPA with a consequential reduction in enzyme activity. Alternatively, reduced activity may

have resulted from an increase in flexibility of the substrate intermediate, since the T94A substitution eliminated an oxyanion hole H-bond, Tyr93-O and Thr94- $O^{\gamma 1}$ (Figure 7c) (41).

Variant N228D

Asn228 is located close to Pro226, the "barrier gate" to S2 binding site (19, 27). Replacing Asn228 by aspartate (N228D) in $NaM1_{B4}$ can improve protein thermal stability (35), although, the same is true for the reverse (42). The reduced activity of $NaM1_{B4}$ may be due to the introduction of a negative charge near the active site, potentially affecting charge distribution during catalysis and reducing enzyme activity (36).

Conclusion

This study demonstrates the value of metagenomics as an alternative screening method to clone libraries and functional metagenomics. Presently, NaM1 is the only CE of metagenomic origin to be structurally and functionally resolved to date. NaM1 is a moderately alkalophilic, halophilic and mesophilic AcXE, the least thermostable of the characterized CE7 esterases. Through our comparison of NaM1 with thermostable CE7 enzymes and our study of single thermostabilising substitutions, we have identified residues that influence thermal stability. We conclude that, in the CE7 enzyme family, reduced residue side chain volume and increased enzyme compactness enhance thermal stability. The N96S substitution in NaM1 further showed how a single mutation influences the narrow acyl moiety specificity unique to this CE family.

MATERIALS AND METHODS

In-silico mining and screening

A Namib hypolith metagenomic dataset (5) was searched using known AcXE homolog sequences with the hmmsearch tool of the HMMER 3.0 (43). Hits were filtered by e-values, sequence length and the presence of open reading frames (ORFs) or contiguous sequences (contig). ORFs coding for less than 100 aa were excluded. The NCBI conserved domain database (CDD) was used to identify AcXE

domains (44). Complete AcXE domains were included in multiple sequence alignments (MSA) to check for conserved motifs from the NCBI-CDD domain models and Carbohydrate Active EnZyme (CAZy) database (34) using MAFFT-Align (45), Clustal-Omega (46) and MEGA 6.0 softwares (47). ExPASy was used to compute the isoelectric point (pI) and molecular weight (MW) (48) and SignalP 4.1 (49) was used to identify signal peptides.

Gene synthesis, sub-cloning and expression

Three putative AcXE-encoding genes were codon-optimized, synthesized and cloned into the EcoRV site of a pUC57 (2710 bp) vector (GenScript, Piscataway, NJ, USA). Putative AcXE-encoding genes were PCR amplified using specific primers (Table **S3**) and Dream Taq DNA polymerase (Thermoscientific, MA, USA). *NaMet1*, *NaMet2* and *NaMet3* amplicon sizes were confirmed by agarose gel electrophoresis and synthesised genes were sub-cloned into a pET28a (5369 bp) expression vector via *XhoI* and *EcoRI* restriction sites. The gene construct coding for an N-terminal His_6-tag was transformed into *Escherichia coli* BL21 DE(3) for isopropyl β-D-1-thiogalactopyranoside (IPTG)-induced gene expression.

Error-prone and site-directed mutagenesis

The epPCR protocol was adapted from published protocols (50). Taq polymerase was induced to generate errors by an imbalance in deoxynucleotide (dNTP) concentrations in the PCR mix. The epPCR cycling conditions are described in Table **S3**. The pET vector system T7 forward and reverse primers were used for direct amplification from the kanamycin-resistant pET28a vector used in this study. PrimerX tool (http://www.bioinformatics.org/primerx/documentation.html) was used to design mutagenic primers. SDM PCR conditions are described in Table **S3**. The annealing temperature was determined by gradient PCR as outlined by the Stratagene QuikChange™ SDM protocol (Agilent technologies, CA, USA). The PCR product was treated with DpnI (New England Biolabs, MA, USA) for 1 h at 37°C to digest template DNA. Cloning and transformation of EpPCR and SDM-PCR

products were carried out as described above. A published protocol was adapted for mutant library construction (51).

Protein production and purification

Proteins were produced for 6 to 8 h in *E. coli* in LB medium after 0.1 mM IPTG induction during the mid-log growth phase at 25°C. Cells were harvested at 5000 x g for 15 min at 4°C using a Sorvall Lynx 6000 centrifuge (Thermoscientific, USA). Harvested cells were re-suspended in 50 mM NaH_2PO_4, 300 mM (pH 7) NaCl (lysis buffer) and lysed on ice with a Q-Sonica Q500 sonicator (Newtown, Connecticut, USA). Soluble protein was separated from cell debris by centrifugation at 17000 x g for 1 h and purified by immobilised metal (Cobalt) affinity chromatography (IMAC). For wash and elution steps, the lysis buffer was supplemented with 10 and 250 mM imidazole, respectively. Target protein fractions were pooled and stored at 4°C in 25 mM Tris-HCl, 25 mM NaCl (pH 8) following buffer exchange using Amicon filters (MWCO 10 kDa) (Clontech, CA, USA). Cell free extracts were used to test for tributyrin hydrolysis. The His_6-tagged gene product was confirmed using KPL His_6-Detector Western Blot AP Colorimetric Kit after sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

Mutant library cultivation and screening

Mutant library cultivation, expression and thermostability screening were adapted from established protocols (51, 52). Cells harvested at 4000 x g for 10 min at 4°C were re-suspended in 20 µL BPER reagent (Thermoscientific, USA) incubated for 15 min at room temperature and diluted 1:1 with 25 mM Tris-HCl, 25 mM NaCl, pH 8. Duplicate plates containing 10 µL aliquots of cell free extracts were incubated at 25 and 55°C for 15 min and assayed spectrophotometrically for *para*-nitrophenol acetate *(p*-NPA) hydrolysis activity (a rapid amber to yellow colour change). Cell free extracts of selected mutants were checked to determine the thermal stability of esterase activities. The thermostable mutant $NaM1_{H2}$ and its SDM-derived variants, $NaM1_{N96S}$ and $NaM1_{F210L}$, were purified from large scale (1L) fermentations.

Enzyme functional characterization

Acetyl esterase activity was determined spectrophotometrically by quantifying the release of *para*-nitrophenol (*p*-NP) from *para*-nitrophenol acetate at $OD_{405\,nm}$ and 25°C. Each assay mix contained 0.5 mM *p*-NPA (Sigma, Switzerland), 50 mM Tris-HCl pH 8 and 0.2 µg NaM1 except where otherwise indicated. One enzyme unit is defined as the amount of enzyme that releases 1 µmol of *p*-NP min^{-1} at pH 8 and 25°C.

Temperature and pH optima were respectively determined by performing acetyl esterase assays over a temperature range of 20 to 75°C and between pH 4 and 11 at 25°C, in 50 mM buffers: phosphate citrate pH 4 to 7, Tris-HCl pH 7 to 9 and 3-cyclohexylamino-1-propanesulphonic acid pH 9.5 to 11.1.

Thermal stability was monitored at 5°C intervals between 4 and 65°C by incubating the enzyme in assay buffer at specified temperatures for 1 h. Thermal inactivation was quantified by determining residual activity at 5 min intervals for 30 min at 50 to 65°C. pH stability between pH 5 and 11 was determined by measuring residual activity after $NaM1_{WT}$ and $NaM1_{H2}$ incubation for 1 h in buffers at the appropriate pH at 35 and 40°C, respectively.

Effects of stabilizing solutes: The effects of stabilizing salts (NaCl, KCl), sugars (0.25 to 1.5 M trehalose, 0.25 M sucrose, 0.25 M glucose), sugar alcohol (0.25 M mannitol) and non-catalytic protein (0.1 to 10 mg ml^{-1} bovine serum albumin -BSA) on the thermal stability of NaM1 were evaluated. The residual activity of the enzyme was determined after incubation in assay buffer supplemented with a specified solute at 30 to 65°C for 15min (1 h for trehalose. NaM1 activity for increasing NaCl concentrations was tested by supplementing assay buffer with NaCl. The stability of NaM1 in NaCl at 40°C was determined by measuring residual activity after incubation in varying NaCl concentrations for 1 h.

Substrate specificity

NaM1 activity on 0.05 to 2 mM *p*-NP butyrate, *p*-NP octanoate and *p*-NP palmitate was assayed as described for *p*-NPA. The release of 4-methylumbelliferone (4-MU) from 4-methylumbelliferone

acetate (4-MUA) (0.01 to 1 mM) and 2-naphthol from 2-NA (0.02 to 1 mM) were monitored spectrophotometrically at 354 nm and 330 nm, respectively. Molar absorption coefficients for *p*-NP (18.3 $mM^{-1}cm^{-1}$), 4-MU (10.47 $mM^{-1}cm^{-1}$) and 2-naphthol (1.5 $mM^{-1}cm^{-1}$) were determined experimentally (53). Deacetylase activity on 0.1 to 2.0 mM 7-ACA and 0.5% (w/v) AX were determined in 50 mM phosphate buffer, pH 8, in 20 min assays with enzyme concentrations of 0.04 and 0.2 mg ml^{-1} respectively. Both deacetylation reactions (1 ml) were terminated with 1 N H_2SO_4. Assay solutions were placed on ice immediately after adding the stop solution. The release of acetic acid was determined at 210 nm using reverse-phase high performance liquid chromatography (RP-HPLC) on a Dionex Ultimate 3000 HPLC unit with a Phenomenex Luna C18 (2) column. The mobile phase was a 93:7 (v/v) mix of 25 mM NaH_2PO_4, pH 2.5: methanol at 1 ml/min flow rate, a 20 µl injection volume and a 25°C column temperature. Elution was achieved by increasing the acetonitrile concentration to 60% (v/v) (54). Acetic acid standards (0.1-10 mM) were prepared using HPLC grade acetic acid. $NaM1_{H2}$ activities on AX, 7-ACA and substrates with acyl moieties of more than two carbon atoms (*p*-NPB, *p*-NPO and *p*-NPP), were also determined. The Michaelis-Menten constant (K_M), maximum velocity (V_{max}), catalytic constant (k_{cat}) and catalytic efficiency were derived from a non-linear regression curve using GraphPad Prism 5 (San Diego, CA, USA). Kinetics of the $NaM1_{WT}$, $NaM1_{H2}$, $NaM1_{N96S}$ and $NaM1_{F210L}$ proteins (0.4 µg) on *p*-NPA were determined at 40°C. All error bars indicate means ± standard deviation.

Circular Dichroism Spectroscopy

Protein thermal stabilities were investigated by circular dichroism (CD) spectroscopy (Applied Photophysics Chirascan, USA) at 5°C intervals between 20 to 65°C. Proteins (4 µM in 25 mM Tris HCl pH 7.0, 25 mM NaCl) in 2 mm path length quartz cuvettes were incubated for 5 min at each temperature and ellipticities (θ) recorded from 195 to 250 nm. The thermal unfolding profile at 210 nm was extracted and fitted to a Boltzmann sigmoidal curve. Blanks confirmed the non-interference of buffer components.

Size exclusion chromatography and Native PAGE

The oligomeric state of NaM1 was studied by size exclusion chromatography using a Superdex 200 10/300 GL column (GE Healthcare Life Sciences, Buckinghamshire, UK) in an AKTA 900 fast pressure liquid chromatography system (Amersham Biosciences, New Hampshire, USA) and 500 µl of 11 mg ml^{-1} NaM1 solution in crystallization buffer at a flowrate of 0.3 ml min^{-1}. Protein eluents were collected in 1 ml fractions and their native molecular weights were further studied using native PAGE.

Crystallization, data collection, structure solution and analysis

A 96-well, sitting-drop crystallization experiment was set up using Qiagen PEG Suite (Hilden, Germany) with a 1:1 ratio between reservoir and 8 mg ml^{-1} NaM1 protein in crystallization buffer (25 mM Tris HCl pH 7, 25 mM NaCl), incubated at 18°C. Initial crystals were iteratively optimized and cryo-protected by 20 to 25% (v/v) PEG 400 in reservoir fluid. X-ray diffraction data were collected at 100 K on beamline ID23-1 of the European Synchrotron Radiation Facility (ESRF), Grenoble, France and processed by the Grenoble automatic data processing system (GrenADeS) (55).

The structure was solved by molecular replacement (MR) using the structure of AcXE from *Ts*ACE (Protein Data Bank [PDB] code 3FCY) as a model in Phaser, followed by Phenix AutoBuild (56), manual correction in COOT (57) and refinement in Phenix Refine (58). Protein geometry was analysed using COOT and MolProbity (59). Ligand interactions were analysed using PDBsum (60) and Ligplot (61), and inter-subunit interactions were determined using the Protein Interfaces, Surfaces and Assemblies service (PISA) (62). To identify structural parameters which affect NaM1 (variant) thermostability, the sequence and structure was aligned with CE7 and related α/β-hydrolases. Proteins were modelled using SwissModel (63) and compared against PDB structures using the DALI server (64). Pymol (65) was used for structural superpositions, analyses and molecular images.

Accession numbers:

NaMet1, *NaMet2* and *NaMet3* have been deposited in the NCBI database under the following nucleotide accession numbers: KX818842, KX818843 and KX818844. Protein identifiers for NaM1,

NaM2 and NaM3 are ATB18054, ATB18055 and ATB18056. Raw reads of the metagenomic dataset were deposited under the SRA accession number SRR2124832.

Acknowledgments

The authors thank Mr Kgama Mathiba, Centre for Scientific and Industrial Research, Pretoria, South Africa for technical assistance during HPLC analyses, and the ESRF for beamline facilities. Funding from the South Africa Bio-catalysis Initiative, Department of Science and Technology, the University of Pretoria Genomics Research Institute (DAC, TPM), the National Research Foundation (WDS, BTS), the Research Development Program (TPM, SV) and the Organisation for Women in Science in the Developing World (OWSD) (FAA) is gratefully acknowledged.

Conflict of interest

The authors declare no conflict of interest.

REFERENCES

1. Adesioye FA, Makhalanyane TP, Biely P, Cowan DA. 2016. Phylogeny, classification and metagenomic bioprospecting of microbial acetyl xylan esterases. Enzyme and Microbial Technology 93:79-91.
2. Allgaier M, Reddy A, Park JI, Ivanova N, D'haeseleer P, Lowry S, Sapra R, Hazen TC, Simmons BA, VanderGheynst JS. 2010. Targeted discovery of glycoside hydrolases from a switchgrass-adapted compost community. Plos one 5:e8812.
3. Dougherty MJ, D'haeseleer P, Hazen TC, Simmons BA, Adams PD, Hadi MZ. 2012. Glycoside hydrolases from a targeted compost metagenome, activity-screening and functional characterization. BMC Biotechnology 12:38.

4. Adriaenssens EM, Van Zyl L, De Maayer P, Rubagotti E, Rybicki E, Tuffin M, Cowan DA. 2015. Metagenomic analysis of the viral community in Namib Desert hypoliths. Environmental Microbiology 17:480-95.
5. Vikram S, Guerrero LD, Makhalanyane TP, Le PT, Seely M, Cowan DA. 2016. Metagenomic analysis provides insights into functional capacity in a hyperarid desert soil niche community. Environmental Microbiology 18:1875-1888.
6. Le PT, Makhalanyane TP, Guerrero L, Vikram S, Van de Peer Y, Cowan DA. 2016. Comparative metagenomic analysis reveals mechanisms for stress response in hypoliths from extreme hyperarid deserts. Genome Biology and Evolution 8:2737-2747.
7. Biely P. 2012. Microbial carbohydrate esterases deacetylating plant polysaccharides. Biotechnology Advances 30:1575-88.
8. CAZy. 2017. Carbohydrate Active enZyme Database (CAZy).
9. Tian Q, Song P, Jiang L, Li S, Huang H. 2014. A novel cephalosporin deacetylating acetyl xylan esterase from *Bacillus subtilis* with high activity toward cephalosporin C and 7-aminocephalosporanic acid. Applied Microbiology and Biotechnology 98:2081-9.
10. Vincent F, Charnock SJ, Verschueren KH, Turkenburg JP, Scott DJ, Offen WA, Roberts S, Pell G, Gilbert HJ, Davies GJ. 2003. Multifunctional xylooligosaccharide/cephalosporin C deacetylase revealed by the hexameric structure of the *Bacillus subtilis* enzyme at 1.9 Å resolution. Journal of Molecular Biology 330:593-606.
11. Takimoto A, Mitsushima K, Yagi S, Sonoyama T. 1994. Purification, characterization and partial amino acid sequences of a novel cephalosporin-C deacetylase from *Bacillus subtilis*. Journal of Fermentation and Bioengineering 77:17-22.
12. Degrassi G, Kojic M, Ljubijankic G, Venturi V. 2000. The acetyl xylan esterase of *Bacillus pumilus* belongs to a family of esterases with broad substrate specificity. Microbiology 146:1585-1591.
13. Degrassi G, Okeke BC, Bruschi CV, Venturi V. 1998. Purification and characterization of an acetyl xylan esterase from *Bacillus pumilus.* Applied and Environmental Microbiology 64:789-792.

14. Martinez-Martinez I, Montoro-Garcia S, Lozada-Ramirez JD, Sanchez-Ferrer A, Garcia-Carmona F. 2007. A colorimetric assay for the determination of acetyl xylan esterase or cephalosporin C acetyl esterase activities using 7-amino cephalosporanic acid, cephalosporin C, or acetylated xylan as substrate. Analytical Biochemistry 369:210-217.

15. Montoro-García S, Gil-Ortiz F, García-Carmona F, Polo LM, Rubio V, Sánchez-Ferrer Á. 2011. The crystal structure of the cephalosporin deacetylating enzyme acetyl xylan esterase bound to paraoxon explains the low sensitivity of this serine hydrolase to organophosphate inactivation. Biochemical Journal 436:321-330.

16. Drzewiecki K, Angelov A, Ballschmiter M, Tiefenbach KJ, Sterner R, Liebl W. 2010. Hyperthermostable acetyl xylan esterase. Microbial Biotechnology 3:84-92.

17. Levisson M, Han GW, Deller MC, Xu Q, Biely P, Hendriks S, Ten Eyck LF, Flensburg C, Roversi P, Miller MD, McMullan D, von Delft F, Kreusch A, Deacon AM, van der Oost J, Lesley SA, Elsliger MA, Kengen SW, Wilson IA. 2012. Functional and structural characterization of a thermostable acetyl esterase from *Thermotoga maritima*. Proteins 80:1545-59.

18. Singh MK, Manoj N. 2016. Crystal structure of *Thermotoga maritima* acetyl esterase complex with a substrate analog: Insights into the distinctive substrate specificity in the CE7 carbohydrate esterase family. Biochemical and Biophysical Research Communications 476:63-68.

19. Singh MK, Manoj N. 2017. Structural role of a conserved active site cis proline in the *Thermotoga maritima* acetyl esterase from the carbohydrate esterase family 7. Proteins: Structure, Function, and Bioinformatics 85:694-708.

20. Lorenz WW, Wiegel J. 1997. Isolation, analysis, and expression of two genes from *Thermoanaerobacterium* sp. strain JW/SL YS485: a beta-xylosidase and a novel acetyl xylan esterase with cephalosporin C deacetylase activity. Journal of Bacteriology 179:5436-5441.

21. Shao W, Wiegel J. 1995. Purification and characterization of two thermostable acetyl xylan esterases from *Thermoanaerobacterium* sp. strain JW/SL-YS485. Applied and Environmental Microbiology 61:729-733.

22. Tan Q, Song Q, Wei D. 2006. Single-pot conversion of cephalosporin C to 7-aminocephalosporanic acid using cell-bound and support-bound enzymes. Enzyme and microbial technology 39:1166-1172.

23. Moraïs S, Stern J, Kahn A, Galanopoulou AP, Yoav S, Shamshoum M, Smith MA, Hatzinikolaou DG, Arnold FH, Bayer EA. 2016. Enhancement of cellulosome-mediated deconstruction of cellulose by improving enzyme thermostability. Biotechnology for Biofuels 9:164.

24. Koseki T, Miwa Y, Fushinobu S, Hashizume K. 2005. Biochemical characterization of recombinant acetyl xylan esterase from Aspergillus awamori expressed in Pichia pastoris: mutational analysis of catalytic residues. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics 1749:7-13.

25. Singh MK, Manoj N. 2016. An extended loop in CE7 carbohydrate esterase family is dispensable for oligomerization but required for activity and thermostability. Journal of Structural Biology 194:434-445.

26. Singh MK, Shivakumaraswamy S, Gummadi SN, Manoj N. 2017. Role of an N-terminal extension in stability and catalytic activity of a hyperthermostable α/β hydrolase fold esterase. Protein Engineering, Design and Selection:1-12.

27. Hedge MK, Gehring AM, Adkins CT, Weston LA, Lavis LD, Johnson RJ. 2012. The structural basis for the narrow substrate specificity of an acetyl esterase from *Thermotoga maritima*. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics 1824:1024-1030.

28. Singh MK, Manoj N. 2017. Structural role of a conserved active site cis proline in the Thermotoga maritima acetyl esterase from the carbohydrate esterase family 7. Proteins: Structure, Function, and Bioinformatics.

29. Krastanova I, Guarnaccia C, Zahariev S, Degrassi G, Lamba D. 2005. Heterologous expression, purification, crystallization, X-ray analysis and phasing of the acetyl xylan esterase from Bacillus pumilus. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics 1748:222-230.

30. Kumar A, Singh S. 2013. Directed evolution: tailoring biocatalysts for industrial applications. Critical Reviews in Biotechnology 33:365-378.

31. Lin L, Fu C, Huang W. 2016. Improving the activity of the endoglucanase, Cel8M from Escherichia coli by error-prone PCR. Enzyme and Microbial Technology 86:52-58.

32. DiCosimo R, Payne MS, Gavagan JE. 2013. Perhydrolase variant providing improved specific activity. Google Patents.

33. Currin A, Swainston N, Day PJ, Kell DB. 2015. Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. Chemical Society Reviews 44:1172-1239.

34. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490-5.

35. Wintrode PL, Arnold FH. 2001. Temperature adaptation of enzymes: lessons from laboratory evolution. Advances in Protein Chemistry 55:161-225.

36. Daniel RM, Danson MJ. 2013. Temperature and the catalytic activity of enzymes: A fresh understanding. FEBS Letters 587:2738-2743.

37. Hakulinen N, Tenkanen M, Rouvinen J. 2000. Three-dimensional structure of the catalytic core of acetylxylan esterase from Trichoderma reesei: insights into the deacetylation mechanism. J Struct Biol 132:180-90.

38. Pollo SM, Zhaxybayeva O, Nesbø CL. 2015. Insights into thermoadaptation and the evolution of mesophily from the bacterial phylum Thermotogae. Canadian journal of microbiology 61:655-670.

39. Pucci F, Rooman M. 2017. Physical and molecular bases of protein thermal stability and cold adaptation. Current Opinion in Structural Biology 42:117-128.

40. Tompa DR, Gromiha MM, Saraboji K. 2016. Contribution of main chain and side chain atoms and their locations to the stability of thermophilic proteins. Journal of Molecular Graphics and Modelling 64:85-93.

41. Lee L-C, Lee Y-L, Leu R-J, Shaw J-F. 2006. Functional role of catalytic triad and oxyanion hole-forming residues on enzyme activity of Escherichia coli thioesterase I/protease I/phospholipase L1. Biochemical Journal 397:69-76.

42. Williams JC, Zeelen JP, Neubauer G, Vriend G, Backmann J, Michels PA, Lambeir A-M, Wierenga RK. 1999. Structural and mutagenesis studies of leishmania triosephosphate isomerase: a point mutation can convert a mesophilic enzyme into a superstable enzyme without losing catalytic power. Protein Engineering 12:243-250.

43. Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Research 39:W29-37.

44. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI. 2014. CDD: NCBI's conserved domain database. Nucleic Acids Research:gku1221.

45. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution 30:772-780.

46. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology 7:539.

47. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution:mst197.

48. Gasteiger E, Hoogland C, Gattiker A, Duvaud Se, Wilkins MR, Appel RD, Bairoch A. 2005. Protein identification and analysis tools on the ExPASy server. Springer.

49. Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature methods 8:785-786.

50. Copp JN, Hanson-Manful P, Ackerley DF, Patrick WM. 2014. Error-prone PCR and effective generation of gene variant libraries for directed evolution. Directed Evolution Library Creation: Methods and Protocols:3-22.

51. Sambrook J, Russell D. 2001. Molecular Cloning: A Laboratory Manual,(pp3. 17-3.32). Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

52. Reetz MT, Carballeira JD. 2007. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. Nature Protocols 2:891-903.

53. Kelly SJ, Butler LG. 1977. Enzymic hydrolysis of phosphonate esters. Reaction mechanism of intestinal 5'-nucleotide phosphodiesterase. Biochemistry 16:1102-1104.

54. Cawthray GR. 2003. An improved reversed-phase liquid chromatographic method for the analysis of low-molecular mass organic acids in plant root exudates. Journal of Chromatography A 1011:233-240.

55. Delagenière S, Brenchereau P, Launer L, Ashton AW, Leal R, Veyrier S, Gabadinho J, Gordon EJ, Jones SD, Levik KE. 2011. ISPyB: an information management system for synchrotron macromolecular crystallography. Bioinformatics 27:3186-3192.

56. McCoy AJ. 2007. Solving structures of protein complexes by molecular replacement with Phaser. Acta Crystallographica Section D: Biological Crystallography 63:32-41.

57. Emsley P, Cowtan K. 2004. Coot: model-building tools for molecular graphics. Acta Crystallographica Section D: Biological Crystallography 60:2126-2132.

58. Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH, Adams PD. 2012. Towards automated crystallographic structure refinement with phenix. refine. Acta Crystallographica Section D: Biological Crystallography 68:352-367.

59. Deis L, Verma V, Videau L, Prisant M, Moriarty N, Headd J, Chen V, Adams P, Snoeyink J, Richardson J. 2013. Phenix/MolProbity hydrogen parameter update. Computational Crystallographic Newsletter 4:9-10.

60. de Beer TA, Berka K, Thornton JM, Laskowski RA. 2014. PDBsum additions. Nucleic Acids Research 42:D292-D296.

61. Wallace AC, Laskowski RA, Thornton JM. 1995. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Engineering 8:127-134.

62. Krissinel E, Henrick K. 2007. Protein interfaces, surfaces and assemblies service PISA at European Bioinformatics Institute. J Mol Biol 372:774-797.

63. Schwede T, Kopp J, Guex N, Peitsch MC. 2003. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Research 31:3381-3385.

64. Holm L. 2010. Dali server: conservation mapping in 3D. Nucleic Acids Research 38:W545-W549.

65. DeLano WL. 2002. The PyMOL molecular graphics system.

66. Corpet F. 1988. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Research 16:10881-10890.

67. Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Research 42:W320-W324.