Structural and functional characterization of a novel acetyl xylan esterase from a desert soil metagenome

by

Fiyinfoluwa Adenike ADESIOYE

Submitted in partial fulfilment of the requirements for the degree

Doctor of Philosophy (PhD) Genetics

In the Faculty of Natural & Agricultural Sciences

University of Pretoria

Pretoria

October, 2017

Supervisor: Professor Don A. Cowan

Co-supervisors: Dr. Thulani Makhalanyane

Professor Wolf-Dieter Schubert

“I do not feel obliged to believe that the same God who has endowed us with sense, reason, and intellect has intended us to forgo their use."

Galileo Galilei (1564-1642)

Declaration

I Fiyinfoluwa Adenike ADESIOYE declare that this thesis, which I hereby submit for the degree of *Doctor of Philosophy* (PhD) in Genetics at the University of Pretoria, is my own original work and has not been previously submitted by me for a degree at this or any other tertiary institution.

SIGNATURE: ……

DATE: ……

Acknowledgements

My profound gratitude goes to Professor Don Cowan for availing me the opportunity of conducting my PhD studies in his renowned research unit. I am grateful for your excellent guidance, support and fatherly understanding during my PhD studies. I particularly appreciate your clear explanations of enzymology and training in scientific writing. I am also grateful to my co-supervisor, Dr. Thulani Makhalanyane, for his valuable suggestions and training during the course of my studies, as well as for consistently lending a much-needed listening ear to my lab woes and victories, so many times. I immensely appreciate my co-supervisor, Professor Wolf-Dieter Schubert for patiently teaching me crystallography. Thank you for encouraging me and inciting my interest in the discipline beyond my pre-conceived misgivings.

I would like to thank all the members of my thesis jury for accepting to be part of my final defense committee and their useful suggestions. I am grateful to Professor Paulette Bloomer for having that introductory and nerve-calming meeting with me on my first arrival at the Department of Genetics for PhD studies. I cannot but thank Professor Trevor Sewell for his mentorship and contributions to the success of my structural data collection.

Special thanks to my colleagues at the Centre for Microbial Ecology and Genomics and the Structural Biology Group particularly Eloy, Surendra, Leandro, Clifford, Valentine, Leticia and Habibu for their assistance. I also thank Ms. Alacia Armstrong and Ms. Yashini Naidoo for their supportive secretarial and administrative work.

I am thankful to the Department of Science and Technology (DST), the Council for Scientific and Industrial Research (CSIR), the Organisation for Women in Science in the Developing World (OWSD) and Adekunle Ajasin University Akungba-Akoko (AAUA), Nigeria, my home institution, for financial and career support during my studies.

I express my sincere thanks to the pastorate and members of the Redeemed Christian Church of God, Jesus House for All Nations, Pretoria. You became the other reason for my being in Pretoria. Thanks for making my stay a joyous and memorable one. I am particularly grateful to all my housemates at Tuksdorp residence and friends, particularly to Dr. and Mrs Adeola, for always welcoming me into their home.

I am immensely grateful to my family especially my siblings and parents, Pastors Henry and Olutoyin

Adesioye and Mrs Comfort Olaitan. You have contributed so much to helping me get here in life. I'm sorry I missed so many of your birthdays, weddings and infant-naming ceremonies during my studies. Thank you all for your love, encouragement, prayers and support.

I would like to thank my husband, Dr. Abiola Olumuyiwa Olaitan. Biola, words cannot describe my deep appreciation of your especially tender love, patience, prayers, intellectual support and commitment. I have great respect for your solidarity with me the whole time I, your brand-new bride, was mostly away from you studying. You are the best!

Above all, I am supremely grateful to God Almighty, for giving me life and the aptitude for learning the science and literature of our time. My ever-faithful Father and my Rock. Thank you.

ABSTRACT

A Namib Desert soil hypolith metagenomic dataset was screened, *in silico*, for novel acetyl xylan esterase (AcXE) - encoding genes. AcXEs hydrolyse ester bonds to liberate acetic acid in acetylated polymeric xylan and xylooligosaccharides during bioconversion of lignocellulosic biomass for sustainable biofuel production. One of the identified genes (*NaMet1*) was synthesized, cloned and expressed to produce a ~36 kDa protein. This protein, NaM1, was confirmed to be functional and was purified and characterized. NaM1, a carbohydrate esterase (CE) 7 enzyme, was optimally active on *p*ara-nitrophenol acetate at pH 8.5 and 30 ^{o}C, and remained active in up to 5 M NaCl and 65% DMSO. The specific activity and catalytic efficiency were 488.9 Umg^{-1} and $3.26x10^{6}$ $M^{-1}s^{-1}$, respectively. NaM1 deacetylated *para*-nitrophenol acetate and butyrate, 7-aminocephalosporanic acid and acetylated xylan.

Most investigations of CE7 esterases have been carried out using structural information from thermostable members of this family and little is known about thermolabile members. A 2.03 Å crystal structure of native NaM1, the first CE structure of metagenomic origin to be submitted to the Protein data bank, was solved. The structure was compared with those of thermostable CE7 enzymes and used to study the thermal stability determinants of this enzyme family. This comparison showed strong structural conservation between both enzyme types and suggested that differences in several key residues, as well as, packing within the core, were responsible for thermal stability.

Directed evolution (DE) of NaM1 yielded thermostable variants, including a variant with 10^{o}C improved stability. Analyses of the kinetic and putative structural characteristics of selected variants in comparison with those of the wild-type provided insights to the role of residues influencing the thermal stability, substrate specificity and activity of NaM1. A single substitution was found to expand acyl moiety specificity and improve both thermal stability and activity of NaM1. Knowledge of key residues identified during NaM1 DE is useful for the future engineering of CE7 and α/β hydrolase enzymes in order to improve catalytic turnover, substrate specificity and thermal stability.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

CHAPTER ONE: LITERATURE REVIEW

METAGENOMIC BIOPROSPECTING OF MICROBIAL ACETYL XYLAN ESTERASES

1.0 INTRODUCTION

The use of renewable sources of energy as alternatives to fossil fuels is the focus of world efforts at sustainable energy development. Consequently, lignocellulosic biomass has been harnessed as a renewable feedstock in the production of biofuels. However, lignocellulosic biomass is recalcitrant to complete bioconversion due to certain structural complexities such as acetylation. Acetyl xylan esterases (AcXEs) are Carbohydrate-Active Enzymes (CAZy) that hydrolyse ester bonds to liberate acetic acid in acetylated polymeric xylan and xylooligosaccharides. They belong to Carbohydrate Esterase 1-7 and 16 families and are known to act on a variety of substrates. Although, several AcXEs have been identified from a range of lignocellulose-degrading microorganisms, saccharification remains a major bottle-neck during industrial biomass-to-biofuel conversions. Metagenomic screening methods allow access to novel metabolites/gene products of the ≥90% uncultured microorganisms within any given environmental sample. This review provides comprehensive information on the phylogeny and ambiguity of bioinformatic annotations of AcXEs. Furthermore, it synthesizes available information on the classification, metagenomic bioprospecting of unique extremophilic environments, large-scale screening limitations and industrial applications of AcXEs. This will facilitate knowledge about this class of accessory enzymes for potential improvements in current and developing AcXE bioprospecting and/or engineering technologies.

1.1 Article: Phylogeny, classification and metagenomic bioprospecting of microbial acetyl xylan esterases.

Fiyinfoluwa A. Adesioye[a], Thulani P. Makhalanyane[a], Peter Biely[b], Don A. Cowan[a].

[a] Centre for Microbial Ecology and Genomics, Genomics Research Institute, Natural Sciences 2 Building, University of Pretoria, Hatfield 0028, Pretoria, South Africa.

[b] Institute of Chemistry, Slovak Academy of Sciences, Dúbravská cesta 9, 845 38, Bratislava, Slovak Republic.

Published in Enzyme and Microbial Technology 93–94 (2016) 79–91

Enzyme and Microbial Technology 93–94 (2016) 79–91

Contents lists available at ScienceDirect

Enzyme and Microbial Technology

journal homepage: www.elsevier.com/locate/emt

Review

Phylogeny, classification and metagenomic bioprospecting of microbial acetyl xylan esterases

Fiyinfoluwa A. Adesioye[a], Thulani P. Makhalanyane[a], Peter Biely[b], Don A. Cowan[a,*]

[a] Centre for Microbial Ecology and Genomics, Genomics Research Institute, Natural Sciences 2 Building, University of Pretoria, Hatfield 0028, Pretoria, South Africa
[b] Institute of Chemistry, Slovak Academy of Sciences, Dúbravská cesta 9, 845 38 Bratislava, Slovak Republic

article info

Article history:
Received 16 February 2016
Received in revised form 18 June 2016
Accepted 1 July 2016
Available online 29 July 2016

Keywords:
Acetyl xylan esterase
Carbohydrate active enzymes
Carbohydrate esterases
Metagenomics

abstract

Acetyl xylan esterases (AcXEs), also termed xylan deacetylases, are broad specificity Carbohydrate-Active Enzymes (CAZymes) that hydrolyse ester bonds to liberate acetic acid from acetylated hemicellulose (typically polymeric xylan and xylooligosaccharides). They belong to eight families within the Carbohydrate Esterase (CE) class of the CAZy database. AcXE classification is largely based on sequence-dependent phylogenetic relationships, supported in some instances with substrate specificity data. However, some sequence-based predictions of AcXE-encoding gene identity have proved to be functionally incorrect. Such ambiguities can lead to mis-assignment of genes and enzymes during sequence data-mining, reinforcing the necessity for the experimental confirmation of the functional properties of putative AcXE-encoding gene products.

Although one-third of all characterized CEs within CAZy families 1–7 and 16 are AcXEs, there is a need to expand the sequence database in order to strengthen the link between AcXE gene sequence and specificity. Currently, most AcXEs are derived from a limited range of (mostly microbial) sources and have been identified via culture-based bioprospecting methods, restricting current knowledge of AcXEs to data from relatively few microbial species. More recently, the successful identification of AcXEs via genome and metagenome mining has emphasised the huge potential of culture-independent bioprospecting strategies. We note, however, that the functional metagenomics approach is still hampered by screening bottlenecks.

The most relevant recent reviews of AcXEs have focused primarily on the biochemical and functional properties of these enzymes. In this review, we focus on AcXE phylogeny, classification and the future of metagenomic bioprospecting for novel AcXEs.

Contents

Abbreviations: AcXE, acetyl xylan esterase; AcE, acetyl esterase; CE, carbohydrate esterase; CAZy, carbohydrate-active enzyme database; CAZymes, carbohydrate-active enzymes; HTP, high through-put; XOS, xylooligosaccharides; T_{opt}, optimum temperature.

* Corresponding author.
E-mail address: don.cowan@up.ac.za (D.A. Cowan).

http://dx.doi.org/10.1016/j.enzmictec.2016.07.001
0141-0229/

1. Introduction

Acetyl xylan esterases (AcXEs) (EC 3.1.1.72) are a class of hydrolytic enzymes primarily involved in the hydrolysis of acetyl xylan esters. They were first described by Biely, Puls [1,2] and have since attracted steadily increasing attention from researchers (Fig. 1A), with a number of recent reviews focusing principally on the chemistry and specificity of AcXE catalysis [3–6]. A number of AcXE-producing microorganisms have been isolated from various environments (further discussed below) and AcXE activity was recently reported for esterases from some plants; e.g., *Populus*-secreted pectin acetyl esterase [7] and *Arabidopsis thaliana* [8]. However, the classification of AcXEs remains unclear, with the CAZy (Carbohydrate Active Enzyme) database classification placing AcXEs within the very broad category of carbohydrate esterases (CEs). Such assignments introduce problems during sequence-based screening and genome annotation and it is therefore important that the linkage between sequence-based phylogenies and functional characteristics of AcXEs (particularly substrate specificity) are properly understood.

The importance of AcXEs in lignocellulose degradation [9] supports the need for continuous bioprospecting for novel AcXEs with improved functional properties [10] using both culture-dependent and genomic bioprospecting approaches. Although, several AcXEs have been identified, the discovery and characterization of new AcXEs of microbial origin with improved catalytic activities via functional metagenomics will result in the availability of a huge reservoir of xylan deacetylases for maximising xylan hydrolysis given different processing conditions [11]. A limited number of xylan deacetylating gene products, identified principally by functional screening of metagenomic libraries, have been reported [12]. While the convention is that novel genes are best identified by activity-based screening of metagenomic libraries [13], identification of putative AcXE genes by annotation of metagenomic sequence datasets is technically feasible, albeit confounded by the miss-annotation of some non-specific esterases as AcXEs. A careful analysis of conserved AcXE phylogenetic data should therefore assist in the discovery of novel, but distantly-related AcXEs.

We note, from a review of literature, that AcXEs are variously termed acetyl xylan esterases, acetylxylan esterases or xylan deacetylases. While all three terms are broadly appropriate, the use of multiple name variants impacts significantly on the search outputs from sequence databases, purely because each is perceived as a different input by search engines and therefore yields different search results (Fig. 1B). Although the term *acetylxylan esterase* is the more correct chemical nomenclature, the term *acetyl xylan esterase* is in more common usage and is consistent with the widely accepted acronym *AcXE*, and we therefore argue strongly for the general acceptance and consistent usage of this term.

2. Acetylated xylan

Many plants have evolved mechanisms for protection of their tissues from physical and biochemical attack (e.g., from insects, microorganisms and enzymes) [14,15]. Such mechanisms include acetylation and 'feruloyation' of plant cell polymers (xylan, mannan, pectin, peptidoglycan and chitin), thereby reducing the efficiency of polysaccharide hydrolases [4,5]. Acetylation occurs mainly in hardwood (acetyl glucuronoxylan), typically at the C2 or C3 or both positions, depending on whether it is a mono- or di-*O*-acetylated xylopyranosyl unit (Fig. 3). On non-reducing xylopyranosyl residues of oligosaccharides or xylopyranosides, acetyl groups can migrate to position C4 [5,16]. The migration is accelerated by increased pH and temperature [17]. The migration between position 2 and 3 on internal xylopyranosyl residues in the polysaccharide is also anticipated, however it has not been proven and demonstrated experimentally. Other substitutions, such as C5 acetylation of the L-arabinofuranosyl side chain residues of xyloglucan, have also been reported [18].

It has been established [3,5,19] that acetylation of xylans generally increases the recalcitrance of plant polymers to depolymerisation by interfering with access of endoxylanases to glycosidic linkages on the xylopyranose main chain and resulting in incomplete hydrolysis of polymeric xylan (hemicellulose). This is a significant factor in reducing the efficiency of saccharification of lignocelluosic biomass for biofuel production [3,5,19].

3. The carbohydrate-active enZyme (CAZy) database

3.1. Carbohydrate esterases (CEs)

Carbohydrate esterases are a large group of carbohydrate-active enzymes that catalyse the removal of ester substituents from the glycan chains of polysaccharides. The catalysis of the de-*O*- or de-*N*-acylation of sugar substituents may be via a Ser-His-Asp catalytic triad, a Ser-His diad or a Zn^{2+} catalytic pathway (or other mechanism) depending on the CE family [20]. With exception of CE15, which act on substrates in which the sugar acts as the acid, all other CE families hydrolyze esters in which sugars act as alcohols. Currently, there are 16 recognized CE families and a CE17 family has been proposed [4,21]. CEs act on soluble substrates and de-esterify saccharide residues esterified with short carboxylic acids such as acetic, propionic, butyric and phenolic acids (ferulic or *p*-coumaric acid) including peptidoglycan (CE4, 9, 11), glucuronic acid (CE14 and 15), chitin (CE4, 14), rhizobial Nod factors (CE4), xylan (CE1-7 and 16), pectin (CE8, 12, 13) and rhamnogalacturonan (CE12) [20,22]. These activities may facilitate the action of other CAZymes associated with plant cell wall degradation. Interestingly, CEs (CE2

and CE16) have also been implicated in transesterification reactions [23,24].

3.2. Acetyl xylan esterases (AcXEs)

Acetyl xylan esterases (EC 3.1z.1.72) belong to CE families 1–7 and 16 [4,20]. Family CE12 includes two enzymes designated as AcXEs, but their ability to deacetylate xylan has not been confirmed. The classification into these separate enzyme families is primar- ily based on similarities in sequence, structure and activity on xylan and per-*O*-acetylated saccharides or methylated saccharides [25,26]. While some acetyl esterases (AcEs) have been wrongly designated as AcXEs [27] due to the similarity of their sequences and inadequate information on their substrate specificities, clear distinctions between these enzyme classes have been made with the creation of the AcXE sub-class (EC 3.1.1.72) in 1999. In some cases, esterases exhibiting xylan deacetylating activity as a side reaction only were assigned as acetyl xylan esterases, such as the CE2 esterases (see below). General esterase activity assays, using α-naphtyl acetate [28] and 4-nitrophenyl acetate as substrates, have frequently been used to determine AcXE activity [19], but cannot distinguish AcEs from AcXEs. AcEs differ from AcXEs in that AcEs are generally not capable of deacetylating polymeric or oligomeric xylan. Although some AcEs may act non-specifically on short acetylated xylooligosaccharides [5], here we define AcXEs as enzymes that may deacetylate polymeric xylan; e.g., AcXE6A from *Fibrobacter succinogenes* S85 [29], and/or xylanase-cleaved XOS; e.g., AcXE from *Trichoderma reesei* [30], regardless of their activity on other substrates. The term 'AcE' is also generally used to describe enzymes which deacetylate substrates other than carbohydrates [31,32]. Simply stated, all AcXEs are AcEs in that they are deacetylating esterases, but not all AcEs are AcXEs.

The closest to the natural substrate of AcXEs is partially depolymerized *O*-acetyl-4-*O*-methyl-d-glucurono-d-xylan (Fig. 3), released from hardwood by steam explosion [33] or extracted from delignified hardwood pulp by DMSO [34]. Chemically acetylated alkali-extracted glucuronoxylan is an alternative substrate used in functional assays to confirm AcXE activity [35]. While the substrate specificity of AcXEs has been associated with whether xylan was naturally or chemically acetylated [36], it remains unclear whether

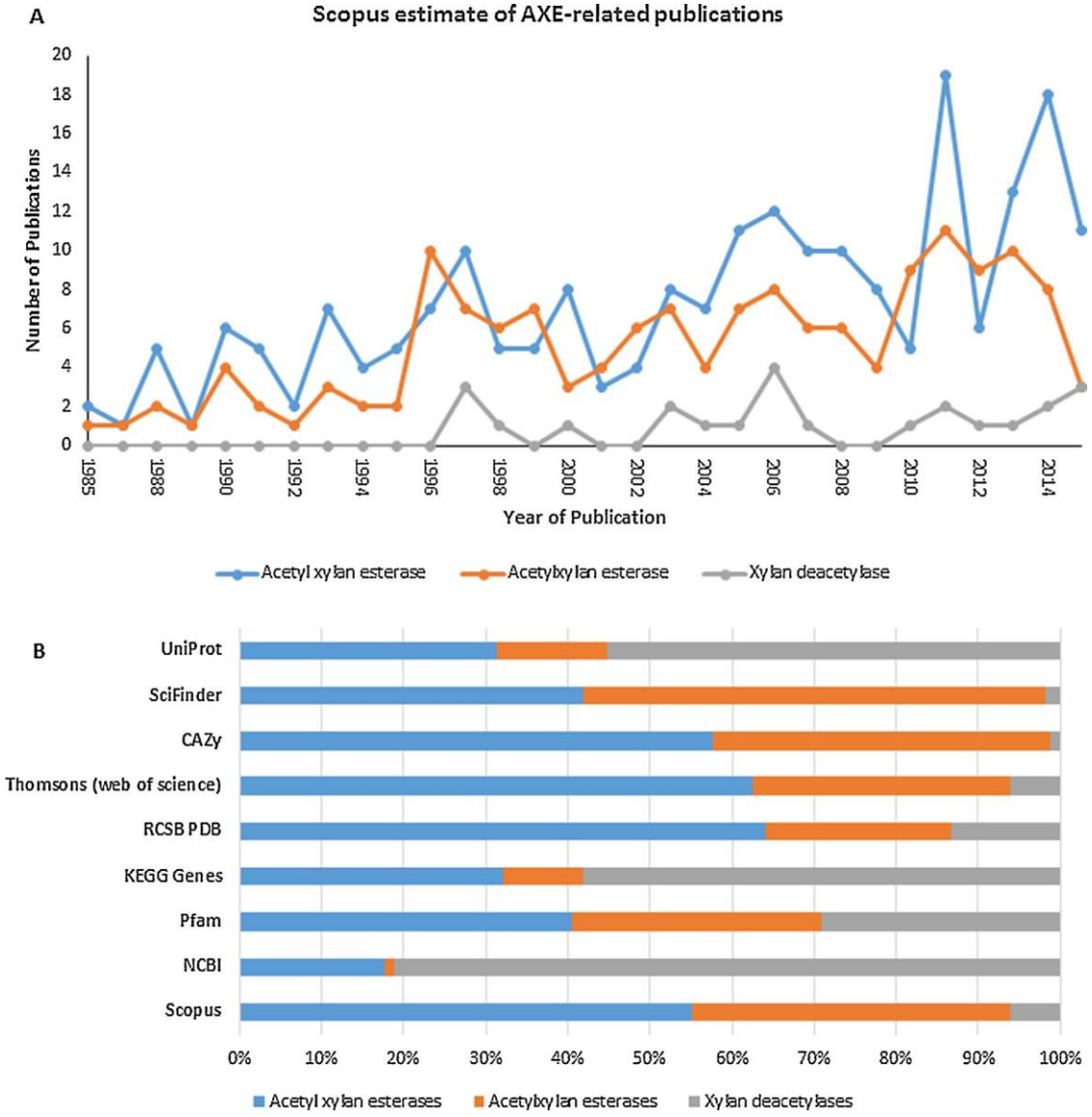


Fig. 1. (A) SAcopus estimate of AcXE–related publications since 1985; (B) Search hits from various databases using 'acetyl xylan esterases', 'acetylxylan esterases' and 'xylan deacetylases[a]' as key words.
[a] Some database search results include non-xylan deacetylases

the acetyl groups in chemically acetylated xylan are located on xylopyranosyl residues only.

Some AcXEs are produced as bimodular or multimodular enzymes [37,38]. Frequently their catalytic domain is linked to a different catalytic module or to a carbohydrate binding module (CBM), emphasizing their role in plant cell wall degradation. The CBM may not be a xylan binding module but a cellulose binding module [39]. Certain AcXEs were reported to bind to other plant cell wall polymers such as cellulose and arabinoxylan while independently deacetylating xylan, thus increasing accessibility of other lignocellulases to decaying plant material [29,40]. Other substrates on which AcXEs have been reported to be active include acetylated monosaccharides [23,41], triacetin [30], α-napthyl acetate [42], tri-*O*-acetyl-d-galactal [43], *p*-nitrophenyl acetate [1], short chain alkyl acetates and 4-methylumbelliferyl acetate [44]. Certain AcXEs have also been shown to have a higher rate of deacetylation when in synergistic action with xylanases [43,45].

3.3. AcXE phylogeny

The phylogenetic relationships between AcXEs are important in the classification of these enzymes into appropriate CE families. A maximum-likelihood phylogenetic tree (Fig. 2), using only characterized AcXE sequences from the CAZy database, shows a clear correlation between sequence homology and predicted activity. This topology also provides strong evidence for inter- and intra-specific horizontal AcXE gene transfers, both within and between bacteria and fungi [46–48]. Non-AcXEs and AcXEs within the same family are shown to originate from the same ancestor (e.g., members of the CE4 family).

Phylogenetic classification is also relevant for higher order classifications; e.g., of CAZyme classes into sub-families [48]. Consequently, certain CAZymes have been grouped into sub-families to serve as a platform through which CAZyme substrate specificities can be directly related to their structures [26], but no CE families have been sub-classified. So far, only the sub-classification of the CE16 family has been strongly supported [21,49].

In the sequence homology-based annotation of AcXEs, it has been demonstrated [50,51] that the use of conserved catalytic residues and block patterns in multiple sequence alignments for constructing phylogenetic trees is a more reliable guide to predicting enzyme specificity [52]. However, while phylogenetic relationships provide clear guidance to assigning AcXE sequences to most CE families, certain AcXEs belonging to families 2, 3 and 16 (all SGNH hydrolases) show substantial substrate promiscuity.

Fig. 2 shows up to 80% branch support for a cluster of CE2 and CE3 AcXEs together with non-AcXEs from the CE8 family. These observations highlight the risk to either functional annotation or prediction of enzyme substrate specificity based on phylogenetic relationships alone; and the importance of functional characterization data [3,53]. Hence, the general consensus that there is still inadequate information on substrate specificity in relation to CAZyme sequence information for automated prediction of CAZyme substrates [26]. It is therefore clear that more phylogenetic data derived from characterized functional AcXEs are required in

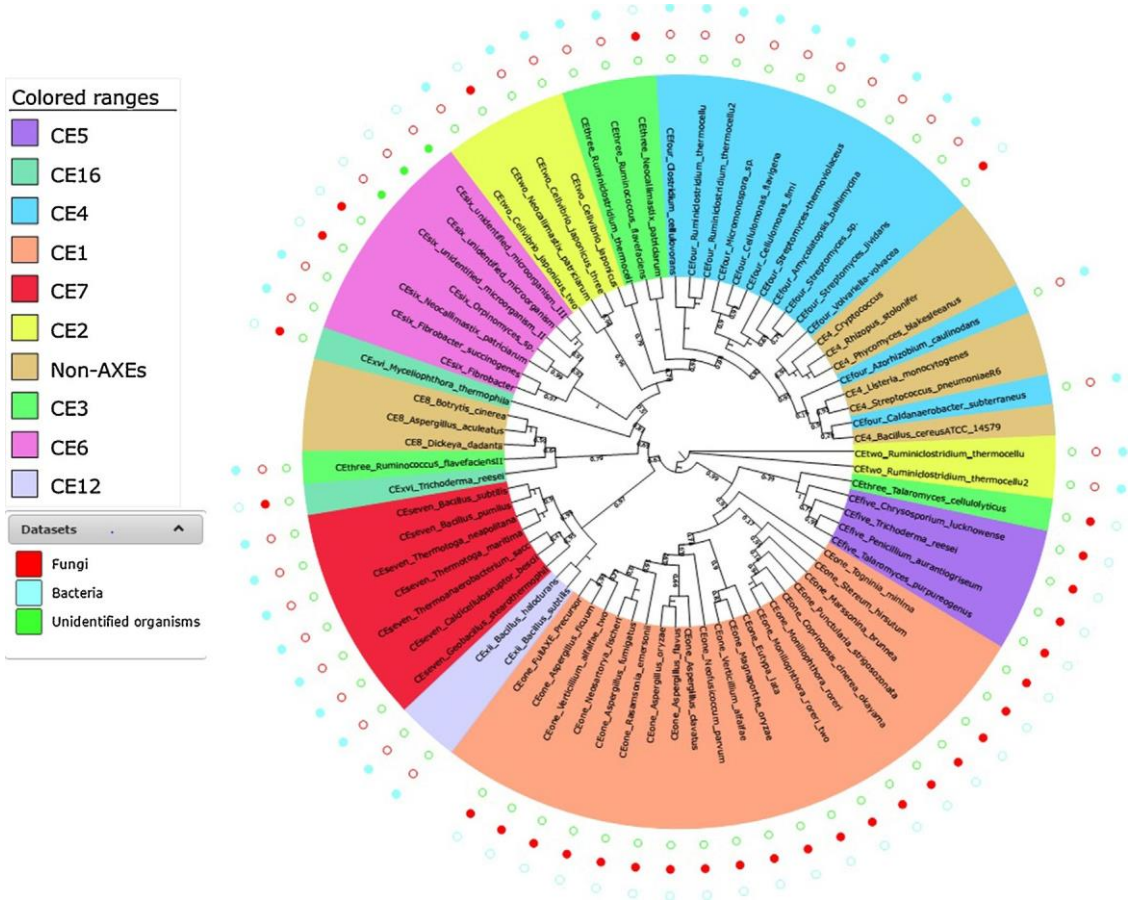


Fig. 2. Unrooted maximum likelihood (1000 bootstrap) phylogenetic tree of characterized AcXEs within CE 1–7 and CE16 families, including some non-AcXEs, constructed with their protein sequence alignments (MAFFT align) using the protein phylogeny methods comparison tool of Mobyle 1.5.4 version of the Pasteur Institute programme [170].

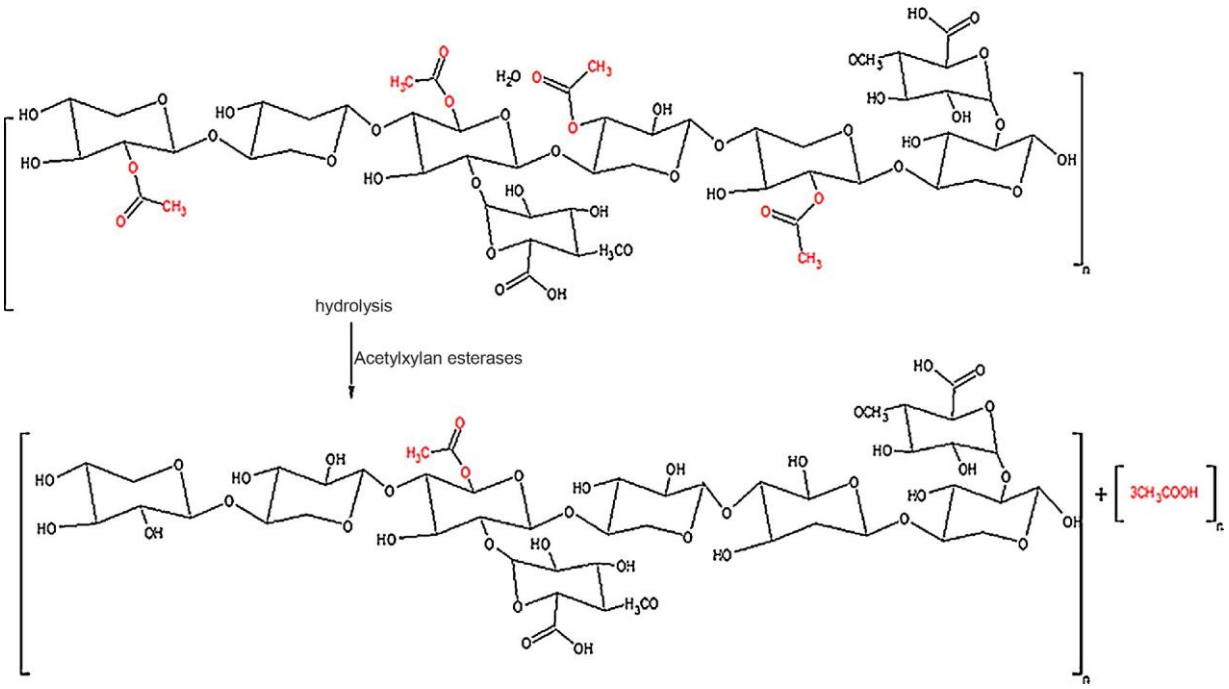


Fig. 3. Schematic representation of xylan deacetylation by acetyl xylan esterases (ACD/Chemsketch v.11.0 −Freeware).

order to establish reliable predictive relationships between gene sequences and substrate specificities.

3.4. AcXE classification—CAZy

AcXE phylogeny largely determines the classification of AcXEs. While biochemical properties and substrate specificities of AcXEs have been described in detail in previous reviews [3,4], we have summarised the most important characteristics of these enzymes, including source organism, substrate preference, conserved sequences (Table 3) and the catalytic residues that are unique to AcXEs within each family. These data clearly show the independence of AcXE classification on any individual characteristic, but strongly confirm the linkage betweensequence homology and functional relationships with other CEs.

3.4.1. CE1

The CE1 family consists of over 3700 CEs of archaeal, bacterial and eukaryotic (fungal and protozoan) origin including AcXEs, feruloyl esterases, carboxylesterases, cinnamoyl esterases, *S*-formylglutathione hydrolases, trehalose 6-*O*-mycolyltransferases, PHB depolymerases and diacylglycerol *O*-acyltransferases [20]. All members of this family possess an *α-J3-α* sandwich three-dimensional structure and are grouped as serine-type esterases possessing the GXSXG conserved sequence with the Ser-His-Asp catalytic triad as well as lengthy, highly conserved sequence blocks. They are active on acetyl xylan as well as acetylated glucosides, galactosides, galactoglucomannan and cellulose [4,54]. AcXEs belonging to this family of enzymes have been isolated largely from fungi such as *Schizophyllum commune* [44,55], *Penicillium purpurogenum* (also *Talaromyces purpureogenus*) [56,57], *Aspergillus ficuum* [58], *Aspergillus awamori* [59], *A. niger* [60], *Volvariella volvacea* [61] and *Chrysosporium lucknowense* (*Myceliophtora thermophile)* [62]. However, while CE1 AcXEs share sufficient homology with certain bacterial CEs to warrant classification into the same family, they do not possess strong phylogenetic relationships with any known AcXEs of bacterial origin. Evidently, classification of AcXEs into this family is not dependent on their possession of sequence homology unique to fungi or fungal AcXEs (which exist in other CE families).

3.4.2. CE2

This family consists of over 175 acetyl esterases (ACEs) and AcXEs of bacterial and fungal origins, only 6 of which have been characterized in detail [20]. There are currently no reported CE2 AcXE sequences from archaeal or protozoan sources. CE2 AcXEs, which were first described by Dalrymple, Cybinski [63] in the fungus *Neocallimastix patriciarum*, are active on acetylated birchwood xylan and *α*-naphthyl acetate. The CE2 AcXEs from *Cellvibrio japonicus* and *Clostridium thermocellum* have been best characterized and are reported to be generally active on aryl esters and selectively active on acetylated xylan, showing higher activities on acetylated konjac glucomannan than on acetylated birchwood xylan and diacetylated xylopyranosides [23,64]. The CE2 AcXEs are serine-type esterases with a Ser-His catalytic diad and an *α-J3* hydrolase protein fold [64]. They possess the GDS(L) conserved motif and exhibit deacetylase specificity for the C6 position of acetylated gluco- and manno-pyranosyl residues of hemicellulose substrates, and for the C4 and C3 positions on mono-acetylated xylopyranosyl residues [23]. 6-*O*-deacetylase activity has not been reported for any member of other CE families. Hence, it is unique to the CE2 family. The 4- and 6-*O*-deacetylation specificities of this family are unlike the regioselectivity for position 2 observed in AcXEs belonging to other CE families. AcXEs from this family catalyse transesterification reactions from vinyl acetate to the O-6 position of hexopyranosides and hexooligosaccharides [4,23].

3.4.3. CE3

Members of the CE3 family belong to the SGNH hydrolase superfamily [21] and are included in the serine-type AcEs which possess the Ser-His-Asp (SHD) catalytic triad and *α-J3-α* sandwich protein structure [39]. The CE3 family has over 190 members of archaeal, bacterial and eukaryotic (fungal) origin. This family also consists of the GDSL lipases and esterases which possess the GDSL motif rather than the GXSXG motif synonymous with lipases and esterases [65]. They are often multi-functional, including AcXE activity. However,

Table 1
Characteristics of CE families to which AcXEs belong. (For interpretation of the references to colour in this Table, the reader is referred to the web version of this article.)

CE Family	Member Enzymes	3D Structure	Enzyme source					Characterised				Total
			Archaea	*Bacteria*	*Fungi*	*Other Eukaryotes*	*Unclassified*	CEs	AcXEs	CEs	AcXEs	
1	AcXEs, cinnamoyl esterases, feruloyl esterases, carboxylesterases,S-formylglutathione hydrolase diacylglycerol O-acyltransferases, trehalose 6-O-mycolyltransferases	α/β/α sandwich	3	3561	99	3	54	8	-	41	12	3720
2	6-O deacetylases with some activity on xylan		-	171	8	-	1	4	4	6	5	180
3	AcXEs	α/β/α sandwich	8	131	54	-	-	1	2	6	5	193
4	AcXEs, chitin deacetylases, chitooligosaccharide deacetylases, peptidoglycan GlcNAc deacetylases, peptidoglycan N-acetylmuramic acid deacetylases	$(\beta/\alpha)_7$ barrel	16	7329	237	13	16	19	2	58	6	7611
5	AcXEs and Cutinases	α/β/α sandwich	-	795	149	2	8	8	2	18	4	954
6	AcXEs	α/β/α sandwich	1	135	3	22	4	2	2	7	7	165
7	AcXEs and cephalosporin C deacetylases	α/β/α sandwich	-	442	-	-	5	4	4	7	7	447
16	Acetyl esterase (multiple substrates)	-	1	29	61	-	-	-	-	2	2	91
	Totals		**29**	**12593**	**611**	**40**	**90**	**46**	16	**145**	48	**13361**

Structurally characterized **Functionally characterized** Culled from www.cazy.org 2015

studies on some of these enzymes to confirm their true substrate specificities are still required [4]. In the biochemical characterization of the catalytic module (CtCes3-1) of CtCes3 (*Clostridium thermocellum)* CE3 AcXE, it was observed that substrate specificity was "restricted to xylans", typical of true AcXEs rather than AcEs [39]. Conversely, in a recent study [66], only very low or no activity of CE3 AcXE (CtCE3-ABN52033) on 2-*O*, 3-*O*, and 2,3-di-*O*-acetylated xylopyranosyl and 3-*O*-acetylated xylopyranosyl 2-*O*-substituted with methylglucuronic acid from *Eucalyptus* was detectable. Other characterized CE3 AcXEs have been obtained from animal rumen bacteria; such as *Ruminococcus flavefaciens* 17 [37], and the fungus *Neocallimastix patriciarum* [45,63].

3.4.4. CE4

The CE4 family is the largest CE category with over 7600 sequenced enzyme members [20]. All enzymes in this CE family are referred to as 'aspartate metalloenzymes' [67]. CE4 family enzymes are characterized by their dependence, for catalytic activity (deacetylation), on coordination of either a Co^{2+} or Zn^{2+} cation [68], and by the use of the His-His-Asp metal binding triad[38]. CE4 AcXEs operate an acid (aspartic acid) − base (histidine) catalytic mechanism and are also known as 'NodB homologs' [69–71]. They possess the highly conserved catalytic NodB domain [72] characteristic of other rhizobial NodB enzymes in the CE4 family such as the chitin deacetylases, chitooligosaccharide deacetylases, rhizobial nod factor deacetylases as well as peptidoglycan N-acetylglucosamine deacetylases and N-acetylmuramic deacetylases which de-esterify *N*- or *O*-acetyl bonds of plant cell wall polymers or oligomers [73–75]. Family CE4 AcXEs possess an eight-stranded $(J3\text{-}\alpha)_8$ barrel protein fold which has an irregular structure [38]. CE4 AcXEs prefer longer chain oligosaccharide substrates and exhibit positional deacetylation specificity [67,73]. They are active on various acetylated xylan residues and are reported to show preference for methyl per-*O*-acetyl-*J*3-d-xylopyranoside [25], but unlike most AcXEs in other families, do not show activity on para-nitrophenyl acetate (pNPA) or 4-methylumbelliferyl acetate [67,73]. On the contrary, an enzyme recently reported as a CE4 AcXE from *Anoxybacillus flavithermus* showed activity on pNPA, but its activity on acetylated xylan was not tested to verify its identity as a 'true' AcXE [76]. On the polymeric substrate, CE4 AcXEs efficiently deacetylate only singly acetylated xylopyranosyl residues. Doubly acetylated xylopyranosides do not serve as substrates because these esterases require a free vicinal OH-group for deacetylation of position 2 or position 3 [77–79]. Characterized CE4 AcXEs include enzymes from *Streptomyces lividans* [80] and *Clostridium thermocellum* [38].

3.4.5. CE5

The CE5 family consists of over 900 protein sequences of AcXEs and cutinases. Both are serine-type esterases with $\alpha\text{-}J3\text{-}\alpha$ sandwich structures and the Ser-His-Asp catalytic triad. CE5 AcXE sequences possess the GXSXG conserved motif as well as highly conserved blocks found within the cutinase domain [81] and belong to the cutinase superfamily [82]. These enzymes preferentially deacetylate xylopyranosyl residues of acetylated glucuronoxylan at the C2 position [4,21] as well as at the C3 position of acetylated methyl xylopyranosides [67]. The first AcXE assigned to the CE5 family [81] was isolated from *Trichoderma reesei* and showed similarly as CE4 esterases. some requirement for a free unsubstituted hydroxyl group for deacetylation of adjacent positions [16]. Other characterized CE5 AcXEs have been isolated from *Chrysosporium lucknowense C1* [62], *Penicillium aurantiogriseum* [83] and *P. purpurogenum* MYA 38 [84]. There are currently no records of characterized bacterial AcXEs in this CE family.

3.4.6. CE6

Over 160 enzymes designated as AcXEs belong to the CE6 family [20]. They are serine esterases with conserved regions similar to the GDS(L) family of esterases and belong to the SGNH hydrolase superfamily [85]. They possess an $\alpha\text{-}J3\text{-}\alpha$ sandwich formation and the Ser-His-Asp catalytic triad typical of esterases as well as a unique, conserved HQGE catalytic site motif positioned between the N-terminus and center of their sequences [86]. CE6 esterases possess a broad substrate specificity [4]. Unlike CE4 AcXEs, they do not require a free vicinal hydroxyl group for deacetylation activity. Characterized AcXEs belonging to this family include Axe6A and

Axe6B from the bacterium *Fibrobacter succinogenes* [29,87], BnaA from *Neocallimastix patriciarum* [45,63], AxeA from *Orpinomyces* sp. [43] and AxeA from a bovine metagenome [12], all of which are of rumen origin. These AcXEs have very low sequence homology to other CE6 enzymes and it is not certain if this is as a result of functionally-convergent or sequence-divergent evolution [29]. AxeA from *Orpinomyces* [43], which is the only commercially available CE6 esterase, deacetylates positions 2 and 3 on both mono- and di-*O*-acetylated xylopyranosyl residues of polymeric xylan [78].

3.4.7. CE7

The CE7 family is made up of over 450 AcXEs and cephalosporin C deacetylases, all of bacterial origin. All possess the α/β/α sandwich conformation common to other serine esterases, the Ser-His-Asp catalytic triad and the GXSXG conserved motif common to esterases, lipases and thioesterases [88]. They are intracellular enzymes which exhibit unique substrate and positional specificities, making the classification of some members of this family as AcXEs uncertain (Levisson et al., 2012). Most AcXEs in this family deacetylate 7-aminocephalosporanic acid and cephalosporin C but possess high sequence homology with non-AcXE cephalosporin deacetylases. Prior to the establishment of the CAZy database and the CE7 family, these enzymes were all classified as cephalosporin deacetylases [4]. AcXE from *Bacillus pumilus* [88] has been well characterized and is often considered as a model organism for this family. Some CE7 AcXEs possess a substrate preference for monoacetylated short chain XOS rather than polymeric xylan [88,89], and act on acetates of simple sugars and various aryl alcohols [88,90]. Other characterized CE7 AcXEs are derived from *Bacillus subtilis* [42] and the thermophiles *Thermotoga maritima* [90] and *Thermoanaerobacterium saccharolyticum* [89,91]. It is also worth noting that all characterized AcXEs within this family are moderately (T_{opt} − 50 °C) to highly (T_{opt} − 90 °C) thermostable.

3.4.8. CE16

The CE16 family is the most recently established [92]. It consists of approximately 100 AcEs, only some of which exhibit AcXE activity. CE16 enzymes possess the GDS(L) catalytic motif and the Serine, Glycine, Asparagine and Histidine (SGNH) catalytic residues. They are mainly of bacterial and fungal origin and the most well-characterized member, a fungal AcE (Aes1) from *Hypocrea jecorina*, an anamorph of *Trichoderma reesei* [92,93] is regarded as the model enzyme for this family [4]. Aes1 does not deacetylate polymeric xylan, but shows a preference for deacetylating positions 3 and 4 on non-reducing end xylopyranosyl residues in XOS. It behaves as an exo-acting deacetylase. It also deacetylates monoacetylated 4-nitrophenyl β-d-xylopyranoside residues and naphtyl acetate [92,93] and can transacetylate saccharides in water or aqueous organic solvents [24,67]. AcEs showing AcXE activity in this family have been referred to generally as AcXEs or specifically as 'AcE-type AcXEs' (Acetyl esterase-AcXEs) because of their inability to act on polymeric xylan [94]. However, this situation has recently changed as reported from two independent studies of CE16 members, one from *Podospora anserina* (*Pa*CE16) [49] and one from *Aspergillus niger* (*An*CE16) [66]. These enzymes showed some catalytic properties compatible with exo-deacetylase activity of Aes1 from *T. reesei*, but also exhibited some activity on polymeric xylan. Similarly, as typical AcXEs they did not attack the 3-*O*-acetyl group on methylglucuronic acid substituted xylopyranosyl residues. Other 'AE-type AcXEs' in this family include AcXEs from *Myceliophthora thermophila* and *Trichoderma harzanium* [94]. Analysis of the xylan-deacetylation activities of these enzymes on acetylated XOS and methylglucuronic acid substituted acetylated XOS showed low individual deacetylation efficiencies but synergistic catalysis when used with the CE1 AcXE from *M. thermophila* and the CE5 AcXE from *T. reesei* [94]. Little is known about the mechanistic properties of AcXEs in this family and no CE16 enzyme structural data is currently available.

3.5. Positional specificity of AcXEs

In addition to substrate specificity, AcXEs exhibit positional specificity (regioselectivity) where a specific carbon position is preferably deacetylated over others [95]. The positional specificities of AcXEs and other CEs have been investigated using techniques capable of monitoring the release of unique acetate groups from acetyl xylan: there include matrix assisted laser desorption/ionisation time of flight (MALDI-TOF) mass spectrometry [62], enzyme coupled assays [96], proton-NMR (^{1}H NMR) [66,67,78] and capillary electrophoresis with laser-induced fluorescence (CE-LIF) [62].

A recent evaluation of the positional specificity of eight fungal AcXEs from seven CE families (CE1-6 and CE16) on mono- and di-*O* acetylated XOS and on 3-*O* acetylated XOS 2-*O* substituted with methylglucuronic acid [66], based on the rate of deacetylation at particular positions, and other recent studies [67,78] show that CE4 AcXEs deacetylate only singly acetylated xylopyranosyl residues, while CE1, CE5 and CE6 AcXEs are capable of also deacetylating doubly acetylated xylopyranosyl residues. None of the typical AcXEs recognized the 3-*O*-acetyl groups on methylglucuronic acid-substituted xylopyranosyl residues as a substrate. 3-*O*-acetyl groups were also resistant to CE16 esterase, but became a substrate when located on the non-reducing end xylopyranosyl residue. An example of such a substrate is the acetylated aldotetrauronic acid ($Ac^3MeGlcA^3Xyl_3$) generated by GH10 xylanase [3,94]. *Trichoderma reesei* CE16 (TrCE16) deacetylated in this aldouronic acid both position 3 and 4 (Puchart et al., unpublished data). In a comparative study of *Pa*CE16 and *An*CE16, their positional specificities were observed to differ from TrCE16 in two ways (Puchart et al., unpublished data). Firstly, they deacetylated the aldouronic acid only when the acetyl group migrated to position 4. This position was originally involved in glycosidic linkage. Secondly, they showed ability to deacetylate polymeric substrate at positions similar to AcXEs of CE families 1, 4, 5 and 6. Based on these studies, it is viewed that the positional specificity of AcXEs varies considerably with the enzyme and the assigned CE family, implying that the positional preference of any AcXE within a CE family cannot be confidently assumed. However, it has also been suggested that AcXE positional specificity may vary with natural and artificial substrates [4,66].

4. Bioprospecting for microbial AcXEs

Bioprospecting of microbial cultures and genomes for lignocellulose-degrading enzymes has increased dramatically over the past decade [97] and a substantial number of novel AcXEs have been identified, though not all have been functionally characterized. A range of environments containing a diversity of lignocellulose-degrading organisms, such as forest soils, termite guts, animal rumens and compost have yielded a range of AcXE-expressing organisms, including the fungal species *Aspergillus* [58,59], *Anoxybacillus* [76], *Trichoderma* [98], *Orpinomyces* [43], *Penicillium* [57], *Schizophyllum* [98], *Fusarium* [99], *Coprinopsis* [100], *Rhodotorula* [41] and *Neocallimastix* [63] and bacteria such as *Bacillus* [101], *Butyrivibrio* [102], *Caldicellulosiruptor* [103], *Caldocellum* [104], *Clostridium* [72], *Fibrobacter* [87], *Geobacillus* [21], *Pseudomonas* [105], *Streptomyces* [35], *Thermoanaerobacterium* [106] and *Thermobifida* [107]. Notably, genomic mining for AcXEs has, until recently, been limited to cultured species. With the rapidly increasing number of available fungal and bacterial genomes, sequence-based genomic mining for AcXEs would be a

useful tool in AcXE bioprospecting. However, the use of culture-independent metagenomic screening methods to access novel AcXE genes is an obvious objective [13,97,108] and is the focus of this review.

4.1. Mining metagenomes for novel AcXEs

The techniques of metagenomics can be used to either identify putative enzyme genes in large environmental metagenome data sets using sequence-based homology searches or by identification of functional expression of genes without prior sequence knowledge: both have potential for the discovery of a wide range of novel biomolecules [109,110].

Sequence homology-based screening involves nucleotide sequence-based screening of metagenomic datasets for genes of interest, where the selection of 'hits' is based on the identification of consensus sequences or motifs that are unique to specific gene product classes or functions. For identification of AcXE-encoding genes from metagenomic sequence data, it is generally recommended to conduct searches based on conserved AcXE domains [26,38,86,111]. However, the prediction of AcXE genes based on CE family sequence homologies from the CAZy database is complex and unreliable without confirmatory functional studies, since individual CE families include non-AcXE members and members of different CAZy classes may be grouped into a similar Protein Family (PFAM) class. For confirmatory studies, 'synthetic metagenomics' [112] can be employed to synthesize predicted genes or the genes can be mined from the actual metagenomic libraries via PCR analysis using specific oligonucleotide probes [113]. Sequence homology-based metagenomic screening has been employed to identify CAZymes [50,51,114], but only one of such experiments have resulted in the functional characterization of AcXEs [51]. However, considering their sequence similarities with other non-AcXE CEs, sequence homology-predicted AcXEs may not provide accurate information on functional properties. A recent study [51] showed that only one of two predicted and synthesized AcXE-encoding genes was expressed in the functional screen.

Activity-based screening, also referred to as functional metagenomics, involves functional detection, typically by enzyme assay, of genes expressing from metagenomic DNA fragments cloned into a suitable expression vector system [109,115]. AcXEs have been identified using functional screening of metagenomic libraries from a bovine metagenome [12]. The screening technologies specifically used in identifying AcXEs are summarised in Table 2. Functional metagenomic mining can be targeted to highly specific functional properties, while genome or sequence-based metagenome mining only accesses enzymes by class, without explicit specificity of properties. Few reported metagenomic screening studies have specifically targeted AcXEs, with most focusing on other CAZy classes or failing to classify identified CEs into families [116–119]. The limited focus on AcXEs is due, in part, to a lack of high throughput (HTP) metagenomic screening methods for direct detection of AcXE activity. In consequence, the general approach is to use a non-specific HTP screen for lipases/esterases, followed by secondary screening for esterases and third-level assays for AcXEs using acetylated xylan [12]. The polymeric substrate would enable the monitoring of AcXE production on the basis of precipitation of soluble polysaccharide in the screening medium due to deacetylation [98]. Acidification of the screening medium would be a less reliable alternative. The fact that acetylated xylan is not currently commercially available also represents a bottleneck in the AcXE screening technologies. Several researchers [29,36,90,101,120] report the in-house (and rather time-consuming) synthesis of this substrate using the method of [35], while others extract acetylated xylan from natural substrates using DMSO [34,66,78,121] or steam [33,58,62].

Table 2
Acetyl xylan esterases identified via metagenomic screening.

Predicted AcXE/CE Family	Source	library/cloning vector	Host	Library/dataset size	No. of clones screened	Hit rate/No. of hits	Screening substrate	Assay substrate	Xylan deacetylation	Reference
CE12	Compost	PDonR221/pET57&60 DEST	*E. coli* BL21 DE	NS	NS	2	NA	pNP-acetate	non-specific	[51]
AcXE1 (geneB); CE1, CE4	Tamar Wallaby	Fosmid (pCC1Fos)	*E. coli*	2.3Mbp	2.3Mbp	27	NA	Not examined	Not tested	[50]
AcXEs (NS)	Termite gut	Fosmids (pCC1Fos); pBK-CMV	*E. coli*	71Gbp	71Gbp	4–34	NA	Not examined	Not tested	[114]
AcXEs (CE1, CE6)	Bovine[a] Rumen	λ Phage; pBKr	*E. coli* XLOLR	200,000 clones	14,000 clones	1:24	α-naphthyl acetate	Acetylated birchwood xylan, glucose pentaacetate, tri-*O*-acetyl-d-galactal, xylose tetraacetate.	Specific	[12]

NA: Not applicable; NS: Not specified.
[a] Function-based metagenomic screen.

Table 3
Whole conserved sequence regions derived from multiple sequence alignments consisting of at least four[a] characterized functional AcXEs within each CE family and their encoded function depicted by listed substrate preference.

CE family	AcXE source organism (GenBank accession number)	Substrate preference	Conserved sequences	References
1	*Aspergillus oryzae* RIB40 (XP_001826329.1)	AX	LEQ**V**TDFG**D**NP**S**NV**K**MY**I**YVP; VA**I**HYC**T**GTA; G**S**PYA**Q**LA; GF**I**VIYP**E**SP; CWDVSS–LTH**N**GGGNSN**S**IANMV;VFVTGT**S**SGAMMTNVMAATYPN-**L**FAAG; YA**G**VP**A**GCF; WNSTCA**Q**GQ; MYP**D**YSGSRP**K**MQI**Y**HG**N**VDTTLYPQNY**E**E; KQW**A**GVFGY; PNLQGI**L**AGGVGH**N**I**Q**I; DM**K**WFG	[58]
2	*Neocallimastix patriciarum* (AAB69091.1)	α-Naphthyl acetate, AX	KIEFIGDSITC**A**YG; **A**S**K**Q**L**NA; SG**F**GI; PD**L**VVINLNGTND; G**K**G**I**DWHP; AE**E**L**V**A**E**I	[45,63]
3	*Ruminococcus flavefaciens* 17 (CAB55348.1)	J3-Naphthyl acetate, AX	IKIMPLGDSIT; D**E**GGYRKYL; VD**L**VDPEG; YDDNHAGYSGYTI; **S**PDIILL**Q**IGTNDV**S**N**G**H; YNE**L**IKKVA; NV**I**YADIH; DGVHPN**A**G**G**Y**E**KMG	[37]
4	*Streptomyces lividans* 1326 (AAC06115.2)	AX	GYV**G**LTFDDGP; ALNR**Q**N**G**LRAT**M**F**N**QGQ; RAQVDAG; V**A**NH**S**Y**T**HP**H**MTQ; SRTQQAI; GLCSGMISPQTGRAVAIP**D**G**S**G**G**G**G**D**G**G; FRPPYGET-NAT**L**R**S**VEAKYGLT**E**VIWDVDS**Q**DWNNAS**T**DAIVQA**V**S**R**L**G**NGQ; VILMHDWPANTLAAIPRIQTLA	[80]
5	*Chrysosporium lucknowense* (ADZ98863.1)	AX	SCPEVHVFGARETTAPPGYGTSQGLVNMVVQAYPGATSEAINYPA-CGGQASCGGIDYNTSANQGTQAVVSAVTSFNQRCPDTKIVLI-GYSQGGQIMDNAYCGGA; SA**L**NAVKATV**W**F**G**NP; Y**R**VGTC**Q**AGGF**A**ARP**P**GF; IKSYCDA**E**DPYCCNGNDAN; YGQQALAFIKSKL	[62]
6	*Orpinomyces* sp. PC-2 (AAC14690.1)	AX	PDPNFHIYLA**L**GQSNMEGQG; GEWYPA**L**PP; L**G**P**V**DYFGRTL; AKKAQK**A**GVIKGILLHQGETN; L**N**L**K**AEEVPLLAGEVV; LPE**V**IPTAHVISA**E**G; D**D**LHF; YR**I**LG**E**RYA	[43]
7	*Bacillus pumilus* (CAB76451.2)	short chain AcXOS	FDLSLE**E**LKKY; DF**S**DFW; M**L**VRGQGG; DTYYYR**G**VYLDAVRA; RIGV**I**GGSQGG**A**LA**I**A**A**AALS; PPSTVFAAYN; Y**F**GHE	[88,101]
16	[a] *Trichoderma reesei* (ABI34466.1)	Non-reducing end AcXOS	KYLITFGDSY; TASGGL**Q**W; WIG**T**ND; AG**G**R**R**FV**I**L; S**Y**LWYDELHP	[4,92]

AX—acetylated xylan, AcXOS—Acetylated xylooligosaccharides; Bold letters are amino acids that are unique to the representative AcXE listed.
[a] Pairwise alignment with *Myceliophthora thermophile* [94].

It is widely acknowledged that metagenomic screening approaches result in low hit rates [122,123], the reasons for which (low gene frequency, promoter incompatibility, rare codon usage, nascent protein folding limitations etc.) have been extensively reviewed [113,124,125]. Attempts to address the various limitations of functional metagenomics include selective enrichment of the uncultured microbial communities prior to construction of metagenomic libraries [126] and the use of heterologous hosts other than *Escherichia coli*; e.g., *Streptomyces lividans*, *Pichia pastoris*, *Saccharomyces cerevisae*, *Pseudomonas putida*, *Rhizobium leguminosarum* and *Bacillus subtilis* [59,109,127–131] to improve functional expression of heterologous genes. Recently, metagenomic libraries from thermophilic habitats were successfully screened for novel esterases using the bacterium *Thermus thermophilus* as host [132].

The consequence of limitations in functional metagenomic studies, however, is a relatively slow rate of discovery of novel AcXEs and this is evident in CAZy statistics. Only five (Table 2) of all functionally-characterized AcXEs catalogued in the CAZy database (Table1) were discovered via metagenomic analysis and all derived from one metagenomic screening study [12].

4.2. AcXEs from extremophiles

Extremophiles, organisms which typically inhabit environments which are considered to offer one or more 'extreme' characteristics (high or low temperatures, pH, salinity, pressures, high radiation; severe desiccation; etc.) [133,134], are popular targets for bioprospecting on the basis that their gene products are very likely to be adapted to function optimally under the specific extreme conditions [135–137]. A number of AcXEs have been identified from thermophiles and hyperthermophiles, including *Caldicellulosiruptor owensensis* [103], *Thermomonospora fusca* [138], *Thermoanaerobacterium* sp. [106], *Thermobifida fusca* [107,120], *Thermotoga maritima* [90], *Talaromyces emersonii* [139] and *Geobacillus stearothermophilus* [21]. AcXE1 from *T. maritima*, with optimal activity at 90 °C, is the most thermostable AcXE reported [90]. AcXEs have also been identified from alkaliphiles such as *Acremonium alcalophilum* [140]. To our knowledge, no characterized 'extremozymes' capable of deacetylating xylans have been reported from metagenomic screening projects.

There is clearly considerable scope to expand the bioprospecting of other extreme environments with the capacity for lignocellulosic degradation for AcXEs. For example, the seasonal or intermittent plant productivity of hot desert soils [141] offers an alternative, and largely unexplored, catalogue of thermophilic and/or halophilic and/or alkaliphilic lignocellulose-degrading organisms [132,142–145]. Cold-active AcXEs might also become a focus for future biomining projects, given the recent development of strategies for bioconversion of lignocellulosic substrates using psychrophilic CAZymes [146,147].

5. Applications of AcXEs

The aims of most enzyme bioprospecting projects are largely application-based. The most prominent potential application of AcXEs is specifically in the saccharification of acetylated hemicellulose extracted from plant biomass under non-alkaline conditions [148–150]. The potential role of AcXEs in enhanced lignocellulosic biomass degradation processes has been extensively reviewed [3,5,6,9,19,151–155]. While alkaline and certain non-alkaline pre-treatment methods in current use remove most acetyl groups from lignocellulosic biomass, they each have their drawbacks based on factors such as type of lignocellulosic feedstock, temperature requirements, intensity of pre-treatment, detoxification requirements and economic viability [154–156]. No single pre-treatment

method can be considered to be ideal for deacetylation of xylans. A recent analysis of the possible role of AcXEs in the bioconversion of lignocellulosic substrates without pre-treatment steps suggests an increased future focus on this enzyme class [103].

AcXEs are also potentially usable in other, less prominent, processes such as in the paper and pulp industry for treating complex carbohydrate waste to enhance its accessibility to xylanases during bleaching of sulfite pulp prepared under mild acidic conditions [157]. Deacetylated xylan is a substrate for the chemical production of hydrogels, which have potential as drug delivery agents [158]. Deacetylated xylooligosaccharides (XOS) are used as food additives and in the synthesis of stereoisomers of monosaccharide sugars [159]. XOS are classified as prebiotics and recommended as high roughage-diets for enhanced gastrointestinal activity in humans and animals [107,159]. Some AcXEs were found to catalyse perhydrolysis to produce peroxycarboxylic acids, that is, transacylation to peroxide instead of water [160–162]. A CE7 AcXE gene has recently been modified by site-directed mutagenesis to create a perhydrolytic enzyme with improved activity for the production of perhydrolytic acids applicable in laundry care and disinfectant formulations [163]. AcXEs have also been employed in deacetylation of other acetylated polysaccharides [40] as well as in the post-synthetic (*in planta*) deacetylation of xylan to enhance lignocellulose saccharification [7,8,164,165]. In synthetic carbohydrate chemistry, AcXEs can potentially be harnessed in catalyzing regioselective transacetylation reactions [23,24,166].

6. Future prospects and conclusion

It is argued that the need for improvements in the efficiency of industrial scale biodegradation of plant biomass necessitates continuous bioprospecting for novel lignocellulose-hydrolysing enzymes [10]. One of the areas where significant advances are possible is considered to be the exploitation of synergies derived from use of accessory enzymes (such as AcXEs) in conjunction with endo-acting cellulases and hemicellulases [2,29,45,107,167–169]. Options for identification of novel (and possibly superior) variants of such enzymes include the use of new screening methodologies, and the targeting of under-explored biological communities. Microbial communities within unique (and extremophilic) ecological niches, such as haloalkaline lacustrine habitats, acid-mine drainage, hot and cold deserts soils, etc., are valid targets for metagenomic bioprospecting for novel AcXEs.

The current level of research interest in AcXEs, shown by their proportion of characterized CEs (>30%: Table 1) and publications surveys (Fig. 1A), is moderately high. Despite this fact, only four functional metagenomic surveys have targeted AcXE-encoding genes (Table 2), suggesting that the potential for identifying novel genetic, functional and structural AcXE variants remains high. However, we argue that the effective and efficient exploitation of this potential is dependent on the future development of HTP functional screening assays, for example using acetylated xylan or XOS substrates, which are specific to this CAZy group.

Conflict of interest

The authors declare no financial or commercial conflict of interest.

Acknowledgements

This research was funded by the South African Department of Science and Technology Bio-catalysis Initiative and National Research Foundation (DAC, TPM) and the University of Pretoria's Genomics Research Institute (DAC) and Research Development Program (TPM). FAA was supported by funds from the Organisation for Women in Science in the Developing World (OWSD).

References

[1] P. Biely, J. Puls, H. Schneider, Acetyl xylan esterases in fungal cellulolytic systems, FEBS Lett. 186 (1985) 80–84.
[2] P. Biely, C. MacKenzie, J. Puls, H. Schneider, Cooperativity of esterases and xylanases in the enzymatic degradation of acetyl xylan, Nat. Biotechnol. 4 (1986) 731–733.
[3] P. Biely, B. Westereng, V. Puchart, P. de Maayer, D.A. Cowan, Recent progress in understanding the mode of action of acetylxylan esterases, J. Appl. Glycosci. 61 (2014) 35–44.
[4] P. Biely, Microbial carbohydrate esterases deacetylating plant polysaccharides, Biotechnol. Adv. 30 (2012) 1575–1588.
[5] P.M. Pawar, S. Koutaniemi, M. Tenkanen, E.J. Mellerowicz, Acetylation of woody lignocellulose: significance and regulation, Front. Plant Sci. 4 (2013) 118.
[6] V. Juturu, J.C. Wu, Insight into microbial hemicellulases other than xylanases: a review, J. Chem. Technol. Biotechnol. 88 (2) (2013) 353–363.
[7] J.Y. Gou, L.M. Miller, G. Hou, X.H. Yu, X.Y. Chen, C.J. Liu, Acetylesterase-mediated deacetylation of pectin impairs cell elongation, pollen germination, and plant reproduction, Plant Cell 24 (2012) 50–65.
[8] P.M.A. Pawar, M. Derba-Maceluch, S.L. Chong, L.D. Gómez, E. Miedes, A. Banasiak, C. Ratke, C. Gaertner, G. Mouille, S.J. McQueen-Mason, Expression of fungal acetyl xylan esterase in Arabidopsis thaliana improves saccharification of stem lignocellulose, Plant Biotechnol. J. 14 (2016) 387–397.
[9] M.D. Sweeney, F. Xu, Biomass converting enzymes as industrial biocatalysts for fuels and chemicals: recent developments, Catalysts 2 (2012) 244–263.
[10] S.K. Khare, A. Pandey, C. Larroche, Current perspectives in enzymatic saccharification of lignocellulosic biomass, Biochem. Eng. J. 102 (2015) 38–44.
[11] R.P. de Vries, M. Nadal, J. van den Brink, D.A. Vivas-Duarte, H. Stalbrand, Fungal Degradation of Plant Oligo and Polysaccharides, Pan Stanford Publishing Pte Limited, Singapore, 2012.
[12] M. Ferrer, O.V. Golyshina, T.N. Chernikova, A.N. Khachane, D. Reyes-Duarte, V.A. Santos, C. Strompl, K. Elborough, G. Jarvis, A. Neef, M.M. Yakimov, K.N. Timmis, P.N. Golyshin, Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora, Environ. Microbiol. 7 (2005) 1996–2010.
[13] D. Cowan, Q. Meyer, W. Stafford, S. Muyanga, R. Cameron, P. Wittwer, Metagenomic gene discovery: past, present and future, Trends Biotechnol. 23 (2005) 321–329.
[14] J. Pérez, J. Munoz-Dorado, T. de la Rubia, J. Martinez, Biodegradation and biological treatments of cellulose, hemicellulose and lignin: an overview, Int. Microbiol. 5 (2002) 53–63.
[15] E.A. Rennie, H.V. Scheller, Xylan biosynthesis, Curr. Opin. Biotechnol. 26 (2014) 100–107.
[16] M. Mastihubová, P. Biely, Lipase-catalysed preparation of acetates of 4-nitrophenyl *J3-d*-xylopyranoside and their use in kinetic studies of acetyl migration, Carbohydr. Res. 339 (2004) 1353–1360.
[17] V. Puchart, P. Biely, Redistribution of acetyl groups on the non-reducing end xylopyranosyl residues and their removal by xylan deacetylases, Appl. Microbiol. Biotechnol. 99 (2015) 3865–3873.
[18] M. Hoffman, Z. Jia, M.J. Peña, M. Cash, A. Harper, A.R. Blackburn, A. Darvill, W.S. York, Structural analysis of xyloglucans in the primary cell walls of plants in the subclass Asteridae, Carbohydr. Res. 340 (2005) 1826–1840.
[19] L.P. Christov, B.A. Prior, Esterases of xylan-degrading microorganisms: production, properties, and significance. Enzyme and microbial technology, 1 5 (1993) 460–475.
[20] CAZy Carbohydrate Active enZyme Database (CAZy) (2015).
[21] O. Alalouf, Y. Balazs, M. Volkinshtein, Y. Grimpel, G. Shoham, Y. Shoham, A new family of carbohydrate esterases is represented by a GDSL hydrolase/acetylxylan esterase from Geobacillus stearothermophilus, J. Biol. Chem. 286 (2011) 41993–42001.
[22] F. Vincent, S.J. Charnock, K.H. Verschueren, J.P. Turkenburg, D.J. Scott, W.A. Offen, S. Roberts, G. Pell, H.J. Gilbert, G.J. Davies, Multifunctional xylooligosaccharide/cephalosporin C deacetylase revealed by the hexameric structure of the Bacillus subtilis enzyme at 1.9 Å resolution, J. Mol. Biol. 330 (2003) 593–606.
[23] E. Topakas, S. Kyriakopoulos, P. Biely, J. Hirsch, C. Vafiadi, P. Christakopoulos, Carbohydrate esterases of family 2 are 6-*O*-deacetylases, FEBS Lett. 584 (2010) 543–548.
[24] L.r. Kremnicky, V.r. Mastihuba, G.L. Côté, Trichoderma reesei acetyl esterase catalyzes transesterification in water, J. Mol. Catal. B: Enzyme 30 (2004) 229–239.
[25] M. Tenkanen, J. Eyzaguirre, R. Isoniemi, C.B. Faulds, P. Biely, Comparison of catalytic properties of acetyl xylan esterases from three carbohydrate esterase families, Appl. Enzymes Lignocellulosics 855 (2003) 211–229.
[26] 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.
[27] M. Levisson, G.W. Han, M.C. Deller, Q. Xu, P. Biely, S. Hendriks, L.F. Ten Eyck, C. Flensburg, P. Roversi, M.D. Miller, D. McMullan, F. von Delft, A. Kreusch,

A.M. Deacon, J. van der Oost, S.A. Lesley, M.A. Elsliger, S.W. Kengen, I.A. Wilson, Functional and structural characterization of a thermostable acetyl esterase from Thermotoga maritima, Proteins 80 (2012) 1545–1559.
[28] M. Rosenberg, V. Roegner, F.F. Becker, The quantitation of rat serum esterases by densitometry of acrylamide gels stained for enzyme activity, Anal. Biochem. 66 (1975) 206–212.
[29] D.K. Kam, H.S. Jun, J.K. Ha, G.D. Inglis, C.W. Forsberg, Characteristics of adjacent family 6 acetylxylan esterases from Fibrobacter succinogenes and the interaction with the Xyn10E xylanase in hydrolysis of acetylated xylan, Can. J. Microbiol. 51 (2005) 821–832.
[30] K. Poutanen, M. Sundberg, An acetyl esterase of Trichoderma reesei and its role in the hydrolysis of acetyl xylans, Appl. Microbiol. Biotechnol. 28 (1988) 419–424.
[31] G.M. Kasundra, A.N. Bhargava, B. Bhushan, K. Shubhakaran, I. Sood, Polyneuritis cranialis with generalized hyperreflexia as a presenting manifestation of thyrotoxicosis, Annal. Indian Acad. Neurol. 18 (2015) 240.
[32] Z. Alrefaie, Vitamin D 3 improves decline in cognitive function and cholinergic transmission in prefrontal cortex of streptozotocin-induced diabetic rats, Behav. Brain Res. 287 (2015) 156–162.
[33] F. Kormelink, B. Lefebvre, F. Strozyk, A. Voragen, Purification and characterization of an acetyl xylan esterase from Aspergillus niger, J. Biotechnol. 27 (1993) 267–282.
[34] E. Hägglund, B. Lindberg, J. McPherson, Dimethylsulphoxide, a solvent for hemicelluloses, Acta Chem. Scand. 10 (1956) 1160–1164.
[35] K. Johnson, J. Fontana, C. MacKenzie, Measurement of acetylxylan esterase in Streptomyces spp, Methods Enzymol. 160 (1988) 551–560.
[36] A. Khan, K. Lamb, R. Overend, Comparison of natural hemicellulose and chemically acetylated xylan as substrates for the determination of acetyl-xylan esterase activity in Aspergilli, Enzyme Microb. Technol. 12 (1990) 127–131.
[37] V. Aurilia, J.C. Martin, S.I. McCrae, K.P. Scott, M.T. Rincon, H.J. Flint, Three multidomain esterases from the cellulolytic rumen anaerobe Ruminococcus flavefaciens 17 that carry divergent dockerin sequences, Microbiology 146 (2000) 1391–1397.
[38] E.J. Taylor, T.M. Gloster, J.P. Turkenburg, F. Vincent, A.M. Brzozowski, C. Dupont, F. Shareck, M.S. Centeno, J.A. Prates, V. Puchart, L.M. Ferreira, C.M. Fontes, P. Biely, G.J. Davies, Structure and activity of two metal ion-dependent acetylxylan esterases involved in plant cell wall degradation reveals a close similarity to peptidoglycan deacetylases, J. Biol. Chem. 281 (2006) 10968–10975.
[39] M.A. Correia, J.A. Prates, J. Brás, C.M. Fontes, J.A. Newman, R.J. Lewis, H.J. Gilbert, J.E. Flint, Crystal structure of a cellulosomal family 3 carbohydrate esterase from Clostridium thermocellum provides insights into the mechanism of substrate recognition, J. Mol. Biol. 379 (2008) 64–72.
[40] G. Mai-Gisondi, O. Turunen, O. Pastinen, N. Pahimanolis, E.R. Master, Enhancement of acetyl xylan esterase activity on cellulose acetate through fusion to a family 3 cellulose binding module, Enzyme Microb. Technol. 79 (2015) 27–33.
[41] H. Lee, R.J. To, R.K. Latta, P. Biely, H. Schneider, Some properties of extracellular acetylxylan esterase produced by the yeast Rhodotorula mucilaginosa, Appl. Environ. Microbiol. 53 (1987) 2831–2834.
[42] Q. Tian, P. Song, L. Jiang, S. Li, H. Huang, A novel cephalosporin deacetylating acetyl xylan esterase from Bacillus subtilis with high activity toward cephalosporin C and 7-aminocephalosporanic acid, Appl. Microbiol. Biotechnol. 98 (2014) 2081–2089.
[43] D.L. Blum, X.-L. Li, H. Chen, L.G. Ljungdahl, Characterization of an acetyl xylan esterase from the anaerobic fungus Orpinomyces sp. strain PC-2, Appl. Environ. Microbiol. 65 (1999) 3990–3995.
[44] P. Biely, K.K.Y. Wong, I.D. Suckling, S. Špániková, Transacetylations to carbohydrates catalyzed by acetylxylan esterase in the presence of organic solvent, BBA Gen. Subj. 1623 (2003) 62.
[45] D. Cybinski, I. Layton, J. Lowry, B. Dalrymple, An acetylxylan esterase and a xylanase expressed from genes cloned from the ruminal fungus Neocallimastix patriciarum act synergistically to degrade acetylated xylans, Appl. Microbiol. Biotechnol. 52 (1999) 221–225.
[46] M. Bruto, C. Prigent-Combaret, P. Luis, Y. Moënne-Loccoz, D. Muller, Frequent, independent transfers of a catabolic gene from bacteria to contrasted filamentous eukaryotes, Proc. R. Soc. B: Biol. Sci. 281 (1789) (2014) 20140848.
[47] L. Boto, Horizontal gene transfer in evolution: facts and challenges, Proc. R. Soc. B: Biol. Sci. 277 (1683) (2010) 819–827.
[48] M. Chaib De Mares, J. Hess, D. Floudas, A. Lipzen, C. Choi, M. Kennedy, I.V. Grigoriev, A. Pringle, Horizontal transfer of carbohydrate metabolism genes into ectomycorrhizal Amanita, New Phytol. 205 (2015) 1552–1564.
[49] V. Puchart, J.-G. Berrin, M. Haon, P. Biely, A unique CE16 acetyl esterase from Podospora anserina active on polymeric xylan, Appl. Microbiol. Biotechnol. (2015) 1–12.
[50] P.B. Pope, S.E. Denman, M. Jones, S.G. Tringe, K. Barry, S.A. Malfatti, A.C. McHardy, J.F. Cheng, P. Hugenholtz, C.S. McSweeney, M. Morrison, Adaptation to herbivory by the Tammar wallaby includes bacterial and glycoside hydrolase profiles different from other herbivores, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 14793–14798.
[51] M.J. Dougherty, P. D'haeseleer, T.C. Hazen, B.A. Simmons, P.D. Adams, M.Z. Hadi, Glycoside hydrolases from a targeted compost metagenome, activity-screening and functional characterization, BMC Biotechnol. 12 (2012) 38.
[52] G.A. Petsko, D. Ringe, Protein Structure and Function, 2nd ed., New Science Press, 2008.
[53] U.T. Bornscheuer, Microbial carboxyl esterases: classification, properties and application in biocatalysis, FEMS Microbiol. Rev. 26 (2002) 73–81.
[54] C. Altaner, B. Saake, M. Tenkanen, J. Eyzaguirre, C.B. Faulds, P. Biely, L. Viikari, M. Siika-aho, J. Puls, Regioselective deacetylation of cellulose acetates by acetyl xylan esterases of different CE-families, J. Biotechnol. 105 (2003) 95–104.
[55] N. Halgasova, E. Kutejova, J. Timko, Purification and some characteristics of the acetylxylan esterase from Schizophyllum commune, Biochem. J. 298 (1994) 751–755.
[56] F. Gordillo, V. Caputo, A. Peirano, R. Chavez, J. Van Beeumen, I. Vandenberghe, M. Claeyssens, P. Bull, M.C. Ravanal, J. Eyzaguirre, Penicillium purpurogenum produces a family 1 acetyl xylan esterase containing a carbohydrate-binding module: characterization of the protein and its gene, Mycol. Res. 110 (2006) 1129–1139.
[57] L. Egana, R. Gutierrez, V. Caputo, A. Peirano, J. Steiner, J. Eyzaguirre, Purification and characterization of two acetyl xylan esterases from Penicillium purpurogenum, Biotechnol. Appl. Biochem. 24 (1996) 33–99.
[58] H.-J. Chung, S.-M. Park, H.-R. Kim, M.-S. Yang, D.-H. Kim, Cloning the gene encoding acetyl xylan esterase from Aspergillus ficuum and its expression in Pichia pastoris, Enzyme Microb. Technol. 31 (2002) 384–391.
[59] T. Koseki, Y. Miwa, T. Akao, O. Akita, K. Hashizume, An Aspergillus oryzae acetyl xylan esterase: molecular cloning and characteristics of recombinant enzyme expressed in Pichia pastoris, J. Biotechnol. 121 (2006) 381–389.
[60] L. De Graaff, J. Visser, H. Van den Broeck, F. Strozyk, F. Kormelink, J. Boonman, Cloning, expression and use of acetylxylan esterases from fungal origin., Eur Patent Application (1992).
[61] S. Ding, J. Cao, R. Zhou, F. Zheng, Molecular cloning, and characterization of a modular acetyl xylan esterase from the edible straw mushroom Volvariella volvacea, FEMS Microbiol. Lett. 274 (2007) 304–310.
[62] L. Pouvreau, M.C. Jonathan, M.A. Kabel, S.W. Hinz, H. Gruppen, H.A. Schols, Characterization and mode of action of two acetyl xylan esterases from Chrysosporium lucknowense C1 active towards acetylated xylans, Enzyme Microb. Technol. 49 (2011) 312–320.
[63] B.P. Dalrymple, D.H. Cybinski, I. Layton, C.S. McSweeney, G.-P. Xue, Y.J. Swadling, J.B. Lowry, Three Neocallimastix patriciarum esterases associated with the degradation of complex polysaccharides are members of a new family of hydrolases, Microbiology 143 (1997) 2605–2614.
[64] C. Montanier, V.A. Money, V.M. Pires, J.E. Flint, B.A. Pinheiro, A. Goyal, J.A. Prates, A. Izumi, H. Stalbrand, C. Morland, A. Cartmell, K. Kolenova, E. Topakas, E.J. Dodson, D.N. Bolam, G.J. Davies, C.M. Fontes, H.J. Gilbert, The active site of a carbohydrate esterase displays divergent catalytic and noncatalytic binding functions, PLoS Biol. 7 (2009) e71.
[65] C.C. Akoh, G.C. Lee, Y.C. Liaw, T.H. Huang, J.F. Shaw, GDSL family of serine esterases/lipases, Prog. Lipid Res. 43 (2004) 534–552.
[66] K.G. Neumuller, A.C. de Souza, J.H. van Rijn, H. Streekstra, H. Gruppen, H.A. Schols, Positional preferences of acetyl esterases from different CE families towards acetylated 4-O-methyl glucuronic acid-substituted xylo-oligosaccharides, Biotechnol. Biofuels 8 (2015) 7.
[67] P. Biely, M. Cziszarova, J.W. Agger, X.L. Li, V. Puchart, M. Vrsanska, V.G. Eijsink, B. Westereng, Trichoderma reesei CE16 acetyl esterase and its role in enzymatic degradation of acetylated hemicellulose, Biochim. Biophys. Acta 1840 (2014) 516–525.
[68] F. Caufrier, A. Martinou, C. Dupont, V. Bouriotis, Carbohydrate esterase family 4 enzymes: substrate specificity, Carbohydr. Res. 338 (2003) 687–692.
[69] J.E. Urch, R. Hurtado-Guerrero, D. Brosson, Z. Liu, V.G. Eijsink, C. Texier, D.M. Van Aalten, Structural and functional characterization of a putative polysaccharide deacetylase of the human parasite Encephalitozoon cuniculi, Protein Sci. 18 (2009) 1197–1209.
[70] J.I. Laurie, J.H. Clarke, A. Ciruela, C.B. Faulds, G. Williamson, H.J. Gilbert, J.E. Rixon, J. Millward-Sadler, G.P. Hazlewood, The NodB domain of a multidomain xylanase from Cellulomonas fimi deacetylates acetylxylan, FEMS Microbiol. Lett. 148 (1997) 261–264.
[71] D. Kafetzopoulos, G. Thireos, J.N. Vournakis, V. Bouriotis, The primary structure of a fungal chitin deacetylase reveals the function for two bacterial gene products, Proc. Natl. Acad. Sci. 90 (1993) 8005–8008.
[72] A. Kosugi, K. Murashima, R.H. Doi, Xylanase and acetyl xylan esterase activities of XynA, a key subunit of the Clostridium cellulovorans cellulosome for xylan degradation, Appl. Environ. Microbiol. 68 (2002) 6399–6402.
[73] D.E. Blair, A.W. Schuttelkopf, J.I. MacRae, D.M. van Aalten, Structure and metal-dependent mechanism of peptidoglycan deacetylase, a streptococcal virulence factor, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 15429–15434.
[74] A. Marchler-Bauer, C. Zheng, F. Chitsaz, M.K. Derbyshire, L.Y. Geer, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, C.J. Lanczycki, F. Lu, S. Lu, G.H. Marchler, J.S. Song, N. Thanki, R.A. Yamashita, D. Zhang, S.H. Bryant, CDD: conserved domains and protein three-dimensional structure, Nucleic Acids Res. 41 (2013) D348–52.
[75] O. Hekmat, K. Tokuyasu, S. Withers, Subsite structure of the endo-type chitin deacetylase from a Deuteromycete, *Colletotrichum lindemuthianum*: an investigation using steady-state kinetic analysis and MS, Biochem. J. 374 (2003) 369–380.

[76] A. Eminoğlu, S. Ülker, C. Sandallı, Cloning, purification and characterization of acetyl xylane esterase from *Anoxybacillus flavithermus* DSM 2641T with activity on low molecular-Weight acetates, Protein J. 34 (2015) 237–242.
[77] V. Puchart, M.-C. Gariépy, F. Shareck, C. Dupont, Identification of catalytically important amino acid residues of *Streptomyces lividans* acetylxylan esterase A from carbohydrate esterase family 4, BBA Proteins Proteomics 1764 (2006) 263–274.
[78] I. Uhliariková, M. Vršanská, B.V. McCleary, P. Biely, Positional specifity of acetylxylan esterases on natural polysaccharide: an NMR study, BBA Gen. Subj. 1830 (2013) 3365–3372.
[79] P. Biely, M. Mastihubová, V. Puchart, The vicinal hydroxyl group is prerequisite for metal activation of *Clostridium thermocellum* acetylxylan esterase, Biochim. Biophys. Acta (BBA) Gen. Subj. 1770 (2007) 565–570.
[80] F. Shareck, P. Biely, R. Morosoli, D. Kluepfel, Analysis of DNA flanking the xlnB locus of *Streptomyces lividans* reveals genes encoding acetyl xylan esterase and the RNA component of ribonuclease P, Gene 153 (1995) 105–109.
[81] E. Margolles-Clark, M. Tenkanen, H. Söderlund, M. Penttilä, Acetyl xylan esterase from *Trichoderma reesei* contains an active-site serine residue and a cellulose-binding domain, Eur. J. Biochem. 237 (1996) 553–560.
[82] A. Marchler-Bauer, M.K. Derbyshire, N.R. Gonzales, S. Lu, F. Chitsaz, L.Y. Geer, R.C. Geer, J. He, M. Gwadz, D.I. Hurwitz, CDD: NCBI's conserved domain database, Nucleic Acids Res. (2014) gku1221.
[83] E.H. Hansen D.A. Skovlund H.R. Soerensen, Polypeptides having acetylxylan esterase activity and polynucleotides encoding same Google Patents (2009).
[84] R. Gutiérrez, E. Cederlund, L. Hjelmqvist, A. Peirano, F. Herrera, D. Ghosh, W. Duax, H. Jörnvall, J. Eyzaguirre, Acetyl xylan esterase II from *Penicillium purpurogenum* is similar to an esterase from *Trichoderma reesei* but lacks a cellulose binding domain, FEBS Lett. 423 (1998) 35–38.
[85] E. Bitto, C.A. Bingman, J.G. McCoy, S.T. Allard, G.E. Wesenberg, G.N. Phillips, Jr.: The structure at 1.6 Angstroms resolution of the protein product of the At4g34215 gene from *Arabidopsis thaliana*, Acta Crystallogr. D Biol. Crystallogr. 61 (2005) 1655–1661.
[86] N. Lopez-Cortes, D. Reyes-Duarte, A. Beloqui, J. Polaina, I. Ghazi, O.V. Golyshina, A. Ballesteros, P.N. Golyshin, M. Ferrer, Catalytic role of conserved HQGE motif in the CE6 carbohydrate esterase family, FEBS Lett. 581 (2007) 4657–4662.
[87] K.P. McDermid, C. Forsberg, C. MacKenzie, Purification and properties of an acetylxylan esterase from *Fibrobacter succinogenes* S85, Appl. Environ. Microbiol. 56 (1990) 3805–3810.
[88] G. Degrassi, M. Kojic, G. Ljubijankic, V. Venturi, The acetyl xylan esterase of Bacillus pumilus belongs to a family of esterases with broad substrate specificity, Microbiology 146 (2000) 1585–1591.
[89] W.W. Lorenz, J. Wiegel, 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, J. Bacteriol. 179 (1997) 5436–5441.
[90] K. Drzewiecki, A. Angelov, M. Ballschmiter, K.J. Tiefenbach, R. Sterner, W. Liebl, Hyperthermostable acetyl xylan esterase, Microbiol. Biotechnol. 3 (2010) 84–92.
[91] D. Currie, A.M. Guss, C. Herring, R.J. Giannone, C.M. Johnson, P.K. Lankford, S.D. Brown, R. Hettich, L.R. Lynd, Profile of secreted hydrolases, associated proteins, and SlpA in Thermoanaerobacterium saccharolyticum during the degradation of hemicellulose, Appl. Environ. Microbiol. 80 (2014) 5001–5011.
[92] X.L. Li, C.D. Skory, M.A. Cotta, V. Puchart, P. Biely, Novel family of carbohydrate esterases, based on identification of the Hypocrea jecorina acetyl esterase gene, Appl. Environ. Microbiol. 74 (2008) 7482–7489.
[93] K. Poutanen, M. Sundberg, H. Korte, J. Puls, Deacetylation of xylans by acetyl esterases of Trichoderma reesei, Appl. Microbiol. Biotechnol. 33 (1990) 506–510.
[94] S. Koutaniemi, M.P. van Gool, M. Juvonen, J. Jokela, S.W. Hinz, H.A. Schols, M. Tenkanen, Distinct roles of carbohydrate esterase family CE16 acetyl esterases and polymer-acting acetyl xylan esterases in xylan deacetylation, J. Biotechnol. 168 (2013) 684–692.
[95] P. Biely, V. Puchart, Recent progress in the assays of xylanolytic enzymes, J. Sci. Food Agric. 86 (2006) 1636–1647.
[96] P. Biely, M. Mastihubová, D.C. la Grange, W.H. van Zyl, B.A. Prior, Enzyme-coupled assay of acetylxylan esterases on monoacetylated 4-nitrophenyl *J*3-d-xylopyranosides, Anal. Biochem. 332 (2004) 109–115.
[97] M.N. Xing, X.Z. Zhang, H. Huang, Application of metagenomic techniques in mining enzymes from microbial communities for biofuel synthesis, Biotechnol. Adv. 30 (2012) 920–929.
[98] P. Biely, C. MacKenzie, H. Schneider, Production of acetyl xylan esterase by Trichoderma reesei and Schizophyllum commune, Can. J. Microbiol. 34 (1988) 767–772.
[99] P. Christakopoulos, D. Mamma, D. Kekos, B. Macris, Enhanced acetyl esterase production by Fusarium oxysporum, World J. Microbiol. Biotechnol. 15 (1999) 443–446.
[100] V. Juturu, C. Aust, J.C. Wu, Heterologous expression and biochemical characterization of acetyl xylan esterase from Coprinopsis cinerea, World J. Microbiol. Biotechnol. 29 (2013) 597–605.
[101] G. Degrassi, B.C. Okeke, C.V. Bruschi, V. Venturi, Purification and characterization of an acetyl xylan esterase from Bacillus pumilus, Appl. Environ. Microbiol. 64 (1998) 789–792.
[102] M. Till, D.C. Goldstone, G.T. Attwood, C.D. Moon, W.J. Kelly, V.L. Arcus, Structure and function of an acetyl xylan esterase (Est2A) from the rumen bacterium Butyrivibrio proteoclasticus, Proteins: Struct. Funct. Bioinf. 81 (2013) 911–917.
[103] X. Peng, W. Qiao, S. Mi, X. Jia, H. Su, Y. Han, Characterization of hemicellulase and cellulase from the extremely thermophilic bacterium Caldicellulosiruptor owensensis and their potential application for bioconversion of lignocellulosic biomass without pretreatment, Biotechnol. Biofuels 8 (2015) 1.
[104] E. Lüthi, D. Love, J. McAnulty, C. Wallace, P. Caughey, D. Saul, P. Bergquist, Cloning, sequence analysis, and expression of genes encoding xylan-degrading enzymes from the thermophile Caldocellum saccharolyticum, Appl. Environ. Microbiol. 56 (1990) 1017–1024.
[105] L. Ferreira, T.M. Wood, G. Williamson, C. Faulds, G.P. Hazlewood, G.W. Black, H.J. Gilbert, A modular esterase from Pseudomonas fluorescens subsp cellulosa contains a non-catalytic cellulose-binding domain, Biochem. J. 294 (1993) 349–355.
[106] W. Shao, J. Wiegel, Purification and characterization of two thermostable acetyl xylan esterases from Thermoanaerobacterium sp. strain JW/SL-YS485, Appl. Environ. Microbiol. 61 (1995) 729–733.
[107] Y.-C. Huang, G.-H. Chen, Y.-F. Chen, W.-L. Chen, C.-H. Yang, Heterlogous expression of thermostable acetylxylan esterase gene from Thermobifida fusca and its synergistic action with xylanase for the production of xylooligosaccharides, Biochem. Biophys. Res. Commun. 400 (2010) 718–723.
[108] J. Handelsman, M.R. Rondon, S.F. Brady, J. Clardy, R.M. Goodman, Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products, Chem. Biol. 5 (1998) R245–R249.
[109] P. Lorenz, J. Eck, Metagenomics and industrial applications, Nat. Rev. Microbiol. 3 (2005) 510–516.
[110] P.K. Pindi, N. Kishore, S. Reddy, Metagenomics: principles, methodology and biotechnological potentials, Textbook Mol. Biotechnol. (2009) 343.
[111] R. Daniel, The metagenomics of soil, Nat. Rev. Microbiol. 3 (2005) 470–478.
[112] T.S. Bayer, D.M. Widmaier, K. Temme, E.A. Mirsky, D.V. Santi, C.A. Voigt, Synthesis of methyl halides from biomass using engineered microbes, J. Am. Chem. Soc. 131 (2009) 6508–6515.
[113] L. Chistoserdova, Recent progress and new challenges in metagenomics for biotechnology, Biotechnol. Lett. 32 (2010) 1351–1359.
[114] F. Warnecke, P. Luginbühl, N. Ivanova, M. Ghassemian, T.H. Richardson, J.T. Stege, M. Cayouette, A.C. McHardy, G. Djordjevic, N. Aboushadi, Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite, Nature 450 (2007) 560–565.
[115] P. Lorenz, K. Liebeton, F. Niehaus, J. Eck, Screening for novel enzymes for biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space, Curr. Opin. Biotechnol. 13 (2002) 572–577.
[116] M. Nyyssönen, H.M. Tran, U. Karaoz, C. Weihe, M.Z. Hadi, J.B. Martiny, A.C. Martiny, E.L. Brodie, Coupled high-throughput functional screening and next generation sequencing for identification of plant polymer decomposing enzymes in metagenomic libraries, Front. Microbiol. 4 (2013).
[117] L. Tasse, J. Bercovici, S. Pizzut-Serin, P. Robe, J. Tap, C. Klopp, B.L. Cantarel, P.M. Coutinho, B. Henrissat, M. Leclerc, Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes, Genome Res. 20 (2010) 1605–1612.
[118] G. Bastien, G. Arnal, S. Bozonnet, S. Laguerre, F. Ferreira, R. Fauré, B. Henrissat, F. Lefèvre, P. Robe, O. Bouchez, Mining for hemicellulases in the fungus-growing termite Pseudacanthotermes militaris using functional metagenomics, Biotechnol. Biofuels 6 (2013) 78.
[119] U. Rabausch, N. Ilmberger, W.R. Streit, The metagenome-derived enzyme RhaB opens a new subclass of bacterial B type alpha-L-rhamnosidases, J. Biotechnol. 191 (2014) 38–45.
[120] C.H. Yang, W.H. Liu, Purification and properties of an acetylxylan esterase from Thermobifida fusca, Enzyme Microb. Technol. 42 (2008) 181–186.
[121] P. Biely, M. Mastihubova, M. Tenkanen, J. Eyzaguirre, X.L. Li, M. Vrsanska, Action of xylan deacetylating enzymes on monoacetyl derivatives of 4-nitrophenyl glycosides of beta-d-xylopyranose and alpha-l-arabinofuranose, J. Biotechnol. 151 (2011) 137–142.
[122] T. Uchiyama, K. Miyazaki, Functional metagenomics for enzyme discovery: challenges to efficient screening, Curr. Opin. Biotechnol. 20 (2009) 616–622.
[123] J. Raes, E.D. Harrington, A.H. Singh, P. Bork, Protein function space: viewing the limits or limited by our view, Curr. Opin. Struct. Biol. 17 (2007) 362–369.
[124] A. Felczykowska, A. Krajewska, S. Zielińska, J.M. Łoś, S.K. Bloch, B. Nejman-Faleńczyk, The most widespread problems in the function-based microbial metagenomics, Acta Biochim. Pol. 62 (2015) 161–166.
[125] L. Ufarté, G. Potocki-Véronèse, E. Laville, Discovery of new protein families and functions: new challenges in functional metagenomics for biotechnologies and microbial ecology, Front. Microbiol. 6 (2015) 563.
[126] J. Yun, S. Ryu, Screening for novel enzymes from metagenome and SIGEX, as a way to improve it, Microb. Cell Fact. 4 (2005) 8.
[127] L. Wang, V. Mavisakalyan, E.R. Tillier, G.W. Clark, A.V. Savchenko, A.F. Yakunin, E.R. Master, Mining bacterial genomes for novel arylesterase activity, Microbiol. Biotechnol. 3 (2010) 677–690.
[128] A. Martinez, S.J. Kolvek, C.L.T. Yip, J. Hopke, K.A. Brown, I.A. MacNeil, M.S. Osburne, Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts, Appl. Environ. Microbiol. 70 (2004) 2452–2463.

[129] K.M. DeAngelis, J.M. Gladden, M. Allgaier, P. D'haeseleer, J.L. Fortney, A. Reddy, P. Hugenholtz, S.W. Singer, J.S. Vander Gheynst, W.L. Silver, B.A. Simmons, T.C. Hazen, Strategies for enhancing the effectiveness of metagenomic-based enzyme discovery in lignocellulolytic microbial communities, BioEnergy Res. 3 (2010) 146–158.
[130] J.L. Adrio, A.L. Demain, Microbial enzymes: tools for biotechnological processes, Biomolecules 4 (2014) 117–139.
[131] T. Koseki, Y. Miwa, S. Fushinobu, K. Hashizume, Biochemical characterization of recombinant acetyl xylan esterase from Aspergillus awamori expressed in Pichia pastoris: mutational analysis of catalytic residues, Biochim. Biophys. Acta (BBA)—Proteins Proteomics 1749 (2005) 7–13.
[132] B. Leis, A. Angelov, M. Mientus, H. Li, V.T. Pham, B. Lauinger, P. Bongen, J. Pietruszka, L.G. Gonçalves, H. Santos, Identification of novel esterase-active enzymes from hot environments by use of the host bacterium Thermus thermophilus, Front. Microbiol. 6 (2015).
[133] K. Horikoshi, A.T. Bull, Prologue: definition, categories, distribution, origin and evolution, pioneering studies, and emerging fields of extremophiles, in: Extremophiles Handbook, Springer, 2011, pp. 3–15.
[134] L.J. Rothschild, R.L. Mancinelli, Life in extreme environments, Nature 409 (2001) 1092–1101.
[135] D.A. Cowan, A. Arslanoglu, S. Burton, R.A. Cameron, G. Baker, J.J. Smith, Q. Meyer, Metagenomics, gene discovery and the ideal biocatalyst, Biochem. Soc. Trans. 32 (2) (2004) 298–302.
[136] D. Cowan, J. Ramond, T. Makhalanyane, P. De Maayer, Metagenomics of extreme environments, Curr. Opin. Microbiol. 25 (2015) 97–102.
[137] M. Ferrer, O. Golyshina, A. Beloqui, P.N. Golyshin, Mining enzymes from extreme environments, Curr. Opin. Microbiol. 10 (2007) 207–214.
[138] S.L. Bachmann, A.J. McCarthy, Purification and cooperative activity of enzymes constituting the xylan-degrading system of Thermomonospora fusca, Appl. Environ. Microbiol. 57 (1991) 2121–2130.
[139] D.M. Waters, P.G. Murray, Y. Miki, A.T. Martinez, M.G. Tuohy, C.B. Faulds, Cloning, overexpression in Escherichia coli, and characterization of a thermostable fungal acetylxylan esterase from Talaromyces emersonii, Appl. Environ. Microbiol. 78 (2012) 3759–3762.
[140] E.O. Pereira, A. Tsang, T.A. McAllister, R. Menassa, The production and characterization of a new active lipase from Acremonium alcalophilum using a plant bioreactor, Biotechnol. Biofuels 6 (2013) 111.
[141] V. Saul-Tcherkas, A. Unc, Y. Steinberger, Soil microbial diversity in the vicinity of desert shrubs, Microb. Ecol. 65 (2013) 689–699.
[142] T.P. Makhalanyane, A. Valverde, E. Gunnigle, A. Frossard, J.B. Ramond, D.A. Cowan, Microbial ecology of hot desert edaphic systems, FEMS Microbiol. Rev. 39 (2015) 203–221.
[143] F. Canganella, J. Wiegel, Anaeroic thermophiles, Life (Basel) 4 (2014) 77–104.
[144] M. Neifar, S. Maktouf, R.E. Ghorbel, A. Jaouani, A. Cherif, Extremophiles as source of novel bioactive compounds with industrial potential, Biotechnol. Bioact. Compd. Sources Appl. (2015) 245.
[145] C. Heath, X.P. Hu, S.C. Cary, D. Cowan, Identification of a novel alkaliphilic esterase active at low temperatures by screening a metagenomic library from antarctic desert soil, Appl. Environ. Microbiol. 75 (2009) 4657–4659.
[146] R. Tiwari, S. Singh, P. Shukla, L. Nain, Novel cold temperature active *J3*-glucosidase from Pseudomonas lutea BG8 suitable for simultaneous saccharification and fermentation, RSC Adv. 4 (2014) 58108–58115.
[147] H.-X. Fan, L.-L. Miao, Y. Liu, H.-C. Liu, Z.-P. Liu, Gene cloning and characterization of a cold-adapted *J3*-glucosidase belonging to glycosyl hydrolase family 1 from a psychrotolerant bacterium Micrococcus antarcticus, Enzyme Microb. Technol. 49 (2011) 94–99.
[148] C.J. Barr, J.A. Mertens, C.A. Schall, Critical cellulase and hemicellulase activities for hydrolysis of ionic liquid pretreated biomass, Bioresour. Technol. 104 (2012) 480–485.
[149] M.J. Selig, E.P. Knoshaug, W.S. Adney, M.E. Himmel, S.R. Decker, Synergistic enhancement of cellobiohydrolase performance on pretreated corn stover by addition of xylanase and esterase activities, Bioresour. Technol. 99 (2008) 4997–5005.
[150] M.J. Selig, W.S. Adney, M.E. Himmel, S.R. Decker, The impact of cell wall acetylation on corn stover hydrolysis by cellulolytic and xylanolytic enzymes, Cellulose 16 (2009) 711–722.
[151] D. Dodd, I.K. Cann, Enzymatic deconstruction of xylan for biofuel production, Glob. Change Biol. Bioenergy 1 (2009) 2–17.
[152] L.L. Li, S.R. McCorkle, S. Monchy, S. Taghavi, D. van der Lelie, Bioprospecting metagenomes: glycosyl hydrolases for converting biomass, Biotechnol. Biofuels 2 (2009) 10.
[153] P. Alvira, E. Tomas-Pejo, M. Ballesteros, M.J. Negro, Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: a review, Bioresour. Technol. 101 (2010) 4851–4861.
[154] S.H. Mood, A.H. Golfeshan, M. Tabatabaei, G.S. Jouzani, G.H. Najafi, M. Gholami, M. Ardjmand, Lignocellulosic biomass to bioethanol, a comprehensive review with a focus on pretreatment, Renew. Sustain. Energy Rev. 27 (2013) 77–93.
[155] X. Chen, J. Shekiro, M.A. Franden, W. Wang, M. Zhang, E. Kuhn, D.K. Johnson, M.P. Tucker, The impacts of deacetylation prior to dilute acid pretreatment on the bioethanol process, Biotechnol. Biofuels 5 (2012) 1.
[156] P. Alvira, E. Tomás-Pejó, M. Ballesteros, M. Negro, Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: a review, Bioresour. Technol. 101 (2010) 4851–4861.
[157] L. Christov, P. Biely, E. Kalogeris, P. Christakopoulos, B. Prior, M. Bhat, Effects of purified endo-*J3*-1, 4-xylanases of family 10 and 11 and acetyl xylan esterases on eucalypt sulfite dissolving pulp, J. Biotechnol. 83 (2000) 231–244.
[158] Z.W.H. Van A.F.A. Chimphango J.F. Görgens, An enzymatic method of producing a hydrogel from xylan Google Patents (2014).
[159] F. Motta C. Andrade M. Santana A review of xylanase production by the fermentation of xylan: classification, characterization and applications Sustainable Degradation of Lignocellulosic Biomass-Techniques, Appl. Commercialization (2013).
[160] S.-M. Park, Acetyl xylan esterase of Aspergillus ficcum catalyzed the synthesis of peracetic acid from ethyl acetate and hydrogen peroxide, J. Biosci. Bioeng. 112 (2011) 473–475.
[161] H.Y. Yoo, J.H. Lee, Y.J. Suh, S.B. Kim, S.-M. Park, S.W. Kim, Immobilization of acetyl xylan esterase on modified graphite oxide and utilization to peracetic acid production, Biotechnol. Bioprocess E 19 (2014) 1042–1047.
[162] W. Tao, Q. Xu, H. Huang, S. Li, Efficient production of peracetic acid in aqueous solution with cephalosporin-deacetylating acetyl xylan esterase from Bacillus subtilis, Process Biochem. (2015).
[163] R. DiCosimo, M.S. Payne, J.E. Gavagan, Perhydrolase variant providing improved specific activity Google Patents (2015).
[164] G. Pogorelko, O. Fursova, M. Lin, E. Pyle, J. Jass, O.A. Zabotina, Post-synthetic modification of plant cell walls by expression of microbial hydrolases in the apoplast, Plant Mol. Biol. 77 (2011) 433–445.
[165] G. Pogorelko, V. Lionetti, O. Fursova, R.M. Sundaram, M. Qi, S.A. Whitham, A.J. Bogdanove, D. Bellincampi, O.A. Zabotina, Arabidopsis and Brachypodium distachyon transgenic plants expressing Aspergillus nidulans acetylesterases have decreased degree of polysaccharide acetylation and increased resistance to pathogens, Plant Physiol. 162 (2013) 9–23.
[166] P. Biely, G. Côté, Microbial hemicellulolytic carbohydrate esterases. in: C. Hou ed. Handbook of industrial biocatalysis (2005).
[167] A. Badhan, Y. Wang, R. Gruninger, D. Patton, J. Powlowski, A. Tsang, T. McAllister, Formulation of enzyme blends to maximize the hydrolysis of alkaline peroxide pretreated alfalfa hay and barley straw by rumen enzymes and commercial cellulases, BMC Biotechnol. 14 (2014) 31.
[168] A. Várnai, T.H. Costa, C.B. Faulds, A.M. Milagres, M. Siika-aho, A. Ferraz, Effects of enzymatic removal of plant cell wall acylation (acetylation, p-coumaroylation, and feruloylation) on accessibility of cellulose and xylan in natural (non-pretreated) sugar cane fractions, Biotechnol. Biofuels 7 (2014) 1–11.
[169] P. Biely, M. Cziszárová, I. Uhliariková, J.W. Agger, X.-L. Li, V.G.H. Eijsink, B. Westereng, Mode of action of acetylxylan esterases on acetyl glucuronoxylan and acetylated oligosaccharides generated by a GH10 endoxylanase, BBA Gen. Subj. 1830 (2013) 5075–5086.
[170] B. Néron, H. Ménager, C. Maufrais, N. Joly, J. Maupetit, S. Letort, S. Carrere, P. Tuffery, C. Letondal, Mobyle: a new full web bioinformatics framework, Bioinformatics 25 (2009) 3005–3011.

1.2 SUMMARY AND CONCLUSIONS

In this chapter, the phylogeny of all known CE families to which microbial AcXEs belong was reviewed and analysed using characterized AcXEs as models per family. To our knowledge, this is the only comprehensive phylogenetic analyses of these class of enzymes to be published and is a reference point for similar analyses of other industrially important enzymes towards enhancing future metagenomic bioprospecting studies. The importance of correct annotation, nomenclature and classification of AcXEs was emphasised in relation to the overall effects on metagenomic bioprospecting of AcXEs. The majority of AcXE-encoding genes in a variety of environments have yet to be explored and the full capacity of AcXEs identified within metagenomes are yet to be harnessed. Currently, there is only one commercially available AcXE and most characterized AcXEs have low activity and varying specificities for a wide range of substrates. To attain industrial relevance, many characterized AcXEs require considerable custom engineering to achieve optimal activity, thermal or chemical stability and/or specificity towards each type of lignocellulosic biomass or substrate. Otherwise, novel AcXEs with improved traits must be identified. According to a recent Scopus estimate, the number of AcXE-related publications have consistently increased since 1985. There was a 119% increase between 2010 and 2016 alone. Hence, it seems there's increasing research aimed at bioprospecting and characterising new AcXEs from microbial genomes and metagenomes, as well as, at protein engineering to fully understand and exploit available AcXEs. This understanding is important towards overcoming acetylation barriers during bioconversion of lignocellulose to biofuels and other industrial applications of AcXEs.

Author contributions

FAA performed the bioinformatic and phylogenetic analyses and wrote the first draft of the manuscript. All other authors edited subsequent versions of the article.

CHAPTER TWO

IN SILICO MINING OF THE NAMIB DESERT HYPOLITH METAGENOMIC DATASET FOR NOVEL AcXE-ENCODING GENES

2.0 INTRODUCTION

Current recommendations for bioprospecting of lignocellulose-decomposing enzymes emphasize the need for identifying novel bio-products from unique, non-conventional environments (Mirete et al., 2016, Adesioye et al., 2016). The focus of this chapter, mining AcXEs which are accessory enzymes involved in plant cell wall (hemicellulose) degradation from the Namib Desert soil environment, is in agreement with these concepts. The Namib Desert soil hypolith metagenome, designated as a hypolithome in this study, has been reported to possess a large retinue of genes encoding cell wall–degrading enzymes via *in silico* studies (Le et al., 2016, Vikram et al., 2016, Adriaenssens et al., 2015). Peak soil temperatures (66°C) recorded in the Namib Desert, coupled with the arid conditions, make it a suitable environment for identifying thermostable enzymes possibly capable of activity under low hydration conditions. Enzymes possessing such characteristics are desirable for industrial processes (Klibanov, 2001, Kumar et al., 2016). Due to the advent of better sequencing technologies and bioinformatic tools, sequence-based metagenomics is increasingly used for the identification of genes encoding bioactive compounds. The availability of improved high-throughput assay systems for validating predicted activity from *in silico* screening outcomes have also encouraged the use of bioinformatic tools for the design and execution of bioprospecting studies (Davids et al., 2013). *In silico* analyses of metagenomes have proven useful for understanding microbial community structure, metabolism and function. Novel bioactive compounds (Reen et al., 2016, Rocha-Martin et al., 2014, Lorenz et al., 2002, Hailes et al., 2017, Ferrer et al., 2016), including AcXEs (Pope et al., 2010, Warnecke et al., 2007, Dougherty et al., 2012) have been identified from various environmental samples and investigated using sequence-based metagenomics. The aim of the work reported in this chapter is to mine the Namib hypolith metagenomic dataset for novel functional AcXEs. The *in silico* screening regimen, as well as, selection strategies used in identifying potentially functional AcXEs are detailed and analyzed here.

2.1 MATERIALS AND METHODS

Sequence-based screening of the Namib metagenome for functional genes encoding AcXE activity was carried out in six distinct stages using methods similar to those reported by Pope et al. (2010) and Dougherty et al. (2012).

2.1.1 Retrieval of AcXE protein sequence homologs

Protein sequences of characterized AcXEs in the CAZy database were retrieved in FASTA format per carbohydrate esterase (CE) family from the NCBI database via their accession numbers. Each group of sequences per CE family was aligned using the ClustalW (Larkin et al., 2007) alignment tool to confirm their homology. Phylogenetic trees were also constructed using the Mega6 software (Tamura et al., 2013) to ascertain their relatedness (Figure 2, Chapter 1).

2.1.2 Interrogation of the Namib hypolith metagenomic dataset for AcXE-encoding genes

The Namib hypolith metagenomic dataset (>600 million bp), previously described by (Vikram et al., 2016), was screened for AcXEs using the HMMER '*hmmsearch'* tool (Finn et al., 2011). Multiple sequence alignments (MSAs) of characterized AcXEs from each CE family (1-7) were used to build profile hmms (hidden Markov models) which served as the *hmmsearch* query against the Namib metagenomics dataset. The trusted cut offs (cut_tc) per-sequence, per-domain reporting and inclusion score threshold format was selected and each hit was presented with its E-value, score and bias. All the hits were then filtered based on various parameters (below) to select the most probable AcXE hits.

2.1.2.1 E-values and sequence length

The HMMER output domain table operates such that, if a significant independent expected (E) value less than one (<1) is displayed for a domain, the domain in that hit is likely to be homologous to the query sequence. Since the average length of an AcXE domain is 275.5 amino acid residues (275.5aa) (Bateman et al., 2004), Namib metagenome *hmmsearch* hit sequences with ≥100aa and with e^- values not greater than 10^{-3} ($\leq 10^{-3}$) were selected from the HMMER top hits list of sequences.

2.1.2.2 Contig analysis for protein coding regions

The ORFs of the genes coding for each of the selected hits described in section 2.1.3.1 were manually identified on the contiguous sequences (contigs) using the general feature format (Gff. file) index of proteins from contigs. Only hit sequences with full reading frames within the contigs and with at least 5-10 flanking nucleotide bases at both the 5' and 3' ends were selected for further analysis.

2.1.3 Conserved domain search and filter

All the selected protein sequences with full ORFs were subjected to a conserved domain (CD) search using the NCBI Batch CD-search tool for single submission of multiple query proteins. The search was conducted against the Conserved Domain Database (CDD) (Marchler-Bauer et al., 2014, Boratyn et al., 2012) with a default E-value threshold of 0.01 (10^{-2}) and the full result mode was selected. The output was filtered and only proteins with full specific domain hits or AcXE multi-domain hits were selected. Partial hits, i.e. proteins with partial CD sequences (>20% missing amino acids) either at the N and/or C termini, were subsequently excluded. Simultaneously, conserved motifs unique to each CE family were identified from multiple sequence alignments (MSAs) of characterized AcXEs within each family (Table 3, Chapter 1).

2.1.4 Confirmation of selected domain function with NCBI curated models

A *blastx* search against non-redundant proteins in the NCBI database was run for each of the selected sequences to determine the percentage similarity and identity of each selected Namib hit sequence to known protein sequences. MSAs of NCBI-curated conserved domain models for each prospective AcXE protein sequence were investigated. Details of the function and characteristics of each protein sequence used to curate the models were retrieved via their accession numbers and reviewed. This was to ascertain from published reports of characterized proteins that the catalytic function of the domain is likely to be xylan deacetylation. Next, the selected sequences were correlated with the MSA models by including each query sequence in the alignment model. Then, the actual amino acids that represent the catalytic and active sites for AcXE activity were identified. Further nucleotide and protein sequence analyses involved MSAs alongside characterized AcXEs

retrieved from the CAZy database (Cantarel et al., 2009) using MAFFT-align (Katoh and Standley, 2013) and Clustal omega (Sievers et al., 2011) software packages. Protein physicochemical parameters were computed using ExPASy - Compute pI/Mw (Gasteiger et al., 2005) and SignalP 4.1 (Petersen et al., 2011) tools.

2.1.5 PCR mining of the Namib Desert soil metagenomic libraries and sequencing

Mining of metagenomes via polymerase chain reaction (PCR) analysis has been successfully applied as a sequence-based screening method (Leis et al., 2013, Itoh et al., 2016). Since the target sequences in this study were now known, this approach was employed. Specific primers were designed for selected ORFs using the Primer3 software (Untergasser et al., 2012). These were used to mine the Namib Desert soil metagenomic clone libraries for the purpose of amplifying and isolating the selected ORFs. The PCR conditions used are described in Table 2.3. Following confirmation of the amplicon sizes by agarose gel electrophoresis, the PCR products were purified and the DNA sequences were confirmed by capillary sequencing. All PCR product purifications and gel DNA extractions in this study were carried out using the NucleoSpin Gel and PCR Clean-up kit by Macherey-Nagel GmbH & Co, Dùren, Germany.

2.1.6 Gene synthesis and primer design

Following consistent non-specific PCR amplification, due to the complexity of the Namib metagenomic library, the selected putative AcXE-encoding genes were codon-optimized for *Escherichia coli* expression and chemically synthesized *de novo* by GenScript, Piscataway, NJ, USA. The Primer3 software (Untergasser et al., 2012) was employed in designing appropriate primers for PCR amplification of the putative AcXE-encoding gene-inserts and these were purified and sequenced as described in section 2.1.7.

2.2 RESULTS AND DISCUSSION

2.2.1 AcXE protein sequence homologs per CE family

AcXEs belonging to the same CE family were more homologous with each other than with those from other CE families. After the exclusion of non-homologous sequences, CE1 family had the largest number of input homologs (twenty) for the *hmmsearch* query while the least number of input homologs (five) was obtained from the CE5 family. This is attributable to the CE1:CE5 available sequence ratio in favor of CE1 AcXEs (Table 1, Chapter 1). MSAs were not established for characterized AcXEs in CE families 12 and 16, largely because of the unconfirmed or limited number of enzymes displaying AcXE activity within these families at the time of study (Biely, 2012, Adesioye et al., 2016, Biely et al., 2016).

2.2.2 AcXE-encoding gene hits within the Namib hypolith metagenomic dataset

A total of 737 hits were recovered from the AcXE homolog HMMER search against the Namib metagenomics dataset. The CE7 family had the largest number of hits (346) as well as the hit with the highest score (382.4), lowest E-value (3.1 e^{-113}) and best alignment (Appendix 1A). The fewest hits (4) were recorded for the CE5 family, three of which had high E-values (0.22, 3.2 and 7.8) and showed poor alignments. Of all the top E-values recorded per family, CE6 family had the poorest and consequently, the least maximum score (27.5). It was observed that some hits, such as gene_ID 20333, from the CE2 *hmmsearch* against the Namib metagenome, were also repeated in the CE3 *hmmsearch* output. This might be as a result of the similarity in the CE2 and CE3 sequence signatures as both families consist of GDSL esterases belonging to the SGNH hydrolase superfamily (Akoh et al., 2004).

AcXEs in the CE5 family are also known as cutinases (cuticle-degrading esterases) (Margolles-Clark et al., 1996). Although, bacterial cutinase sequences have been deposited in the CAZy database, characterized AcXEs belonging to the CE5 family were all observed to be of fungal origin (Adesioye et al., 2016). Cutin is a major structural component of desert plant surfaces that helps reduce water loss. Consequently, microbial cutinases were expected to be present in the Namib Desert soil environment, induced by decaying plants, but the Namib Desert hypolithome is mainly dominated by

bacteria while fungal members were estimated to constitute only about 3% of the community (Le et al., 2016). In another study, secretome analysis of the fungus, *Aureobasidium pullulans* (*A. namibiae*) isolated beneath a dolomite rock in the Namib Desert (Zalar et al., 2008), revealed a low number of cutinase-encoding genes (Gostinčar et al., 2014). These, put together may explain the limited number of CE5 hits identified in the Namib Desert hypolithome.

2.2.3 E-value and sequence length filter outputs

From the *hmmsearch* output (Appendix 1A), a total of 209 hits of the initial 709 Namib metagenome AcXE hits had E-values below 10^{-3} (0.001) and were longer than 100aa residues (Table 2.1). The CE7 family maintained the largest number (83) of selected hits while family CE6 had the fewest (3) hits selected (Table 2.1). No hits from the CE5 family hmmsearch output met the above-stated criteria. However, the only hit from this family which had an E-value less than zero (7.2 e^{-10}) and 79aa residues, was selected for the next screening stage.

Table 2.1: Summary of Acetyl xylan esterase (AcXE) *in silico* screening results from the Namib Desert metagenome

CE Family	No of homologs in search query	Total No of hits	No of selected hits[a]	Full length ORFs	*NCBI Batch Conserved Domains Database (CDD) search using full length ORFs as query*					
					Total CD hits	AcXE CD hits[b]	Complete AcXE CDs[c]	Conserved Domains (CDs)	Hit type	Description
CE1	20	146	40	6	73	-	-	-	-	-
CE2	12	46	11	2	54	-	-	-	-	-
CE3	9	138	55	6	111	1	**1**	**Xyn B**	**Specific**	**SGNH_hydrolase subfamily, similar to AcXE from *Ruminococcus flavefaciens* XynB (Aurilia et al., 2000).**
CE4	14	48	16	3	132	6†	3	CE4_SlAcXE_like;	Non-specific	Catalytic NodB domain of *Streptomyces lividans* AcXE and its bacterial homologs (Puchart et al., 2006)
							3	CE4_CtAcXE_like	Non-specific	Catalytic NodB domain of *Clostridium thermocellum* AcXE and its bacterial homologs (Taylor et al., 2006)
CE5	4	4	1*	-	-	-	-	-	-	-
CE6	5	8	3	-	-	-	-	-	-	-
CE7	7	346	83	5	75	5	**2**	**Axe1**	**Multi-domain**	**Esterase-Lipase super-family bacterial AcXE (Degrassi et al., 2000, Degrassi et al., 1998)**
Total	71	737	209	22	*427*	*12*	*9*			

*Sequence length <100, but the only sequence with the CE5 superfamily conserved domain and AcXE identity.
[a] Selection of probable AcXEs was achieved using the following filters: Sequence length ≥ 100 and Expected value (e-value) ≤10^{-3}.
[b] Selected after excluding all non-specific hits and specific hits whose CDs were not curated with characterized AcXEs or AcXEs at all.
[c] ORFs with complete CD hits at the N and C terminuses. ORFs: Open reading frames
† Three peculiar CE4 ORFs each with 2 Non-specific AcXE domain hits.
In bold: Three putative Namib metagenome AcXEs selected for primer design and PCR analysis.

2.2.4 Protein coding regions hits with full open reading frames (ORFs)

Most of the 209 selected hits were found to be partial gene sequences with incomplete reading frames; i.e. with either stop or start codons absent from the ORFs within the contigs. Full reading frames with at least 5-10 nucleotide bases flanking both the 5' and 3' ends of the protein coding regions within the contigs were found for only 22 of the Namib metagenome hits. Several hits were incomplete ORFs within contigs which is attributable to sequence trimming during contig assembly. The 22 hits included hits from CE families 1-4 and 7. There were 6 hits (the most per family) each from CE1 and 3 families and 2 hits (the least per family) from CE2 family (Table 2.1).

2.2.5 AcXE conserved domain hits

The NCBI Batch-CD search output (full mode) provides a full list of all the conserved domain hits (specific hits, superfamilies or multi-domains) found on each of the 22 protein sequences searched against the Conserved Domain Database (CDD). This comprehensive list (Appendix 1B) states the CD hit type, CD location, E^{-}value, bit score, accession number, CD name and superfamily details for each of several (10-35) domain hits found within each of the 22 protein sequences. A total of 427 CD hits were found and from these 17 CD hits (9 specific hits, 6 non-specific AcXE domain hits and 2 AcXE multi-domain hits) were selected after excluding all non-AcXE and non-specific domain hits as well as incomplete (N and/or C termini) AcXE domain hits (Appendix 1C).

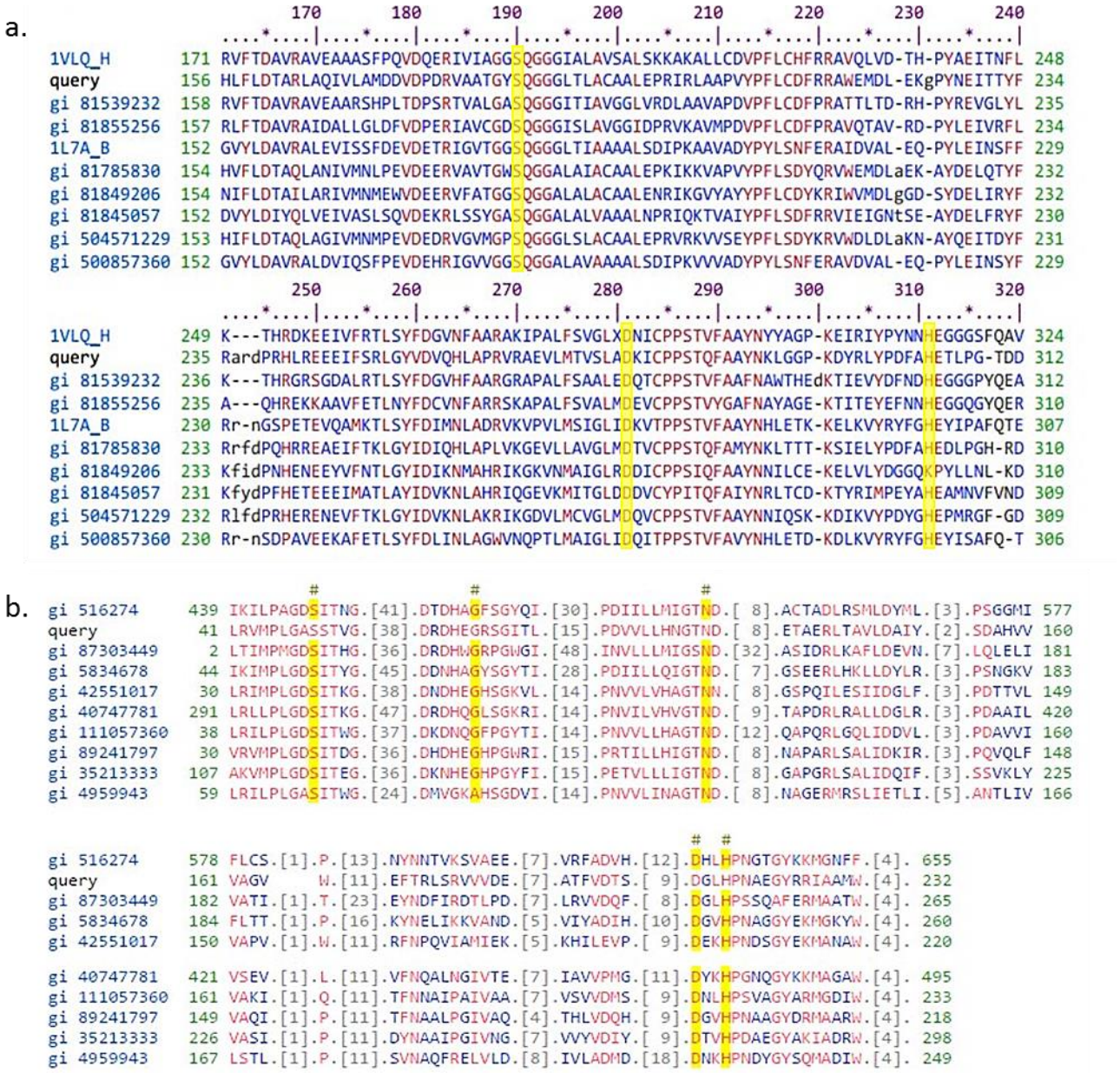


Figure 2.1: Sequence alignment of NCBI curated AcXE domains.

a. pfam05448: Axe1 with user query (gene ID – 39518) included. b. cd01833: XynB_like, with user query (gene ID – 173248) included. SHD and SGNHD catalytic residues are highlighted in yellow respectively. Conserved residues are displayed in red letters (Marchler-Bauer et al., 2014). Axe1 domain was curated with esterases from *Thermotoga maritima* (IVLQ_H), *Streptomyces coelicolor* (gi 81539232), *Mesorhizobium loti* (gi 81855256), *Bacillus subtilis* (IL7A_B), *Bacillus halodurans* (gi 81785830), *Clostridium perfringens* (gi 81849206), *Streptococcus pneumoniae* R6 (gi 81845057), *Thermoanaerobacterium aotearoense* (gi 504571229) and *Bacillus pumilus* (gi 500857360). XynB_like domain was curated with esterases from *Ruminococcuc flavefaciens* (gi 516274), *Synechococcus* sp. WH 5701 (gi 87303449), *Ruminococcus flavefaciens* 17 (gi 5834678),

Gibberella zeae PH-1 (gi 42551017), *Aspergillus nidulans* FGSC A4 (gi 40747781), *Phaeosphaeria nodorum* SN15 (gi 111057360), *Actinoplanes* sp. SE50/110 (gi 89241797), *Gloeobacter violaceus* PCC 7421 (gi 35213333) and Aspergillus terreus (gi 4959943)

2.2.6 NCBI curated models and confirmation of AcXE domain function

Evaluation of each sequence that made up the NCBI curated domain models of each domain hit revealed that only 9 domain hits (1 Xyn_B specific domain, 2 Axe1 multi-domain and 6 other non-specific domain hits), had their domain models curated by well characterized proteins duly reported to express AcXE activity. Each ORF possessing these domains was embedded as a query within the MSA of the domain models (Figure 2.1) and was found to possess the accurate motifs that depict the AcXE active sites (Table 3, Chapter 1). The protein sequences used to curate the 8 other domain hits were mostly not AcXEs; some were acetyl esterases (ACEs) with unconfirmed functions.

Blastx sequence homology data showed that Namib ORFs with gene IDs 39518, 27345 and 173248 were all most closely related to AcXEs from actinobacteria, but shared not more than 64%, 69% and 59% identity, respectively, with these AcXE sequences (Table 2.2). The other ORFs had non-specific AcXE hits and showed more similarity with characterized chitin deacetylases in the CE4 family than with AcXEs in that family.

Hence, three ORFs, with gene IDs 27345, 39518 and 173248, were selected for further studies and named *NaMet1*, *NaMet2* and *NaMet3*, respectively (Appendix 1D). The three putative AcXEs were <70% identical to any known enzyme and can be defined as ‘novel’ (Tindall et al., 2010). The relatedness of these novel AcXE-encoding ORFs to genes from organisms in the phylum Actinobacteria may be indicative of their origin. This is expected as members of this phylum are commonly found in arid soils and hypolithons (Wong et al., 2010, Makhalanyane et al., 2015, Makhalanyane et al., 2013, Vikram et al., 2016). Furthermore, this suggests that actinobacteria which are one of the dominant species in hypolithons, are chiefly responsible for lignocellulose metabolism. Cyanobacteria, on the other hand, are known to be responsible for most of the functional processes in these niches (Cowan et al., 2011). Considering the cyanobacteria to actinobacteria ratio (85:3) reported in the Namib desert hypolithons (Makhalanyane et al., 2013), it can be inferred that lignocellulose deconstruction is a secondary activity in hypolithons.

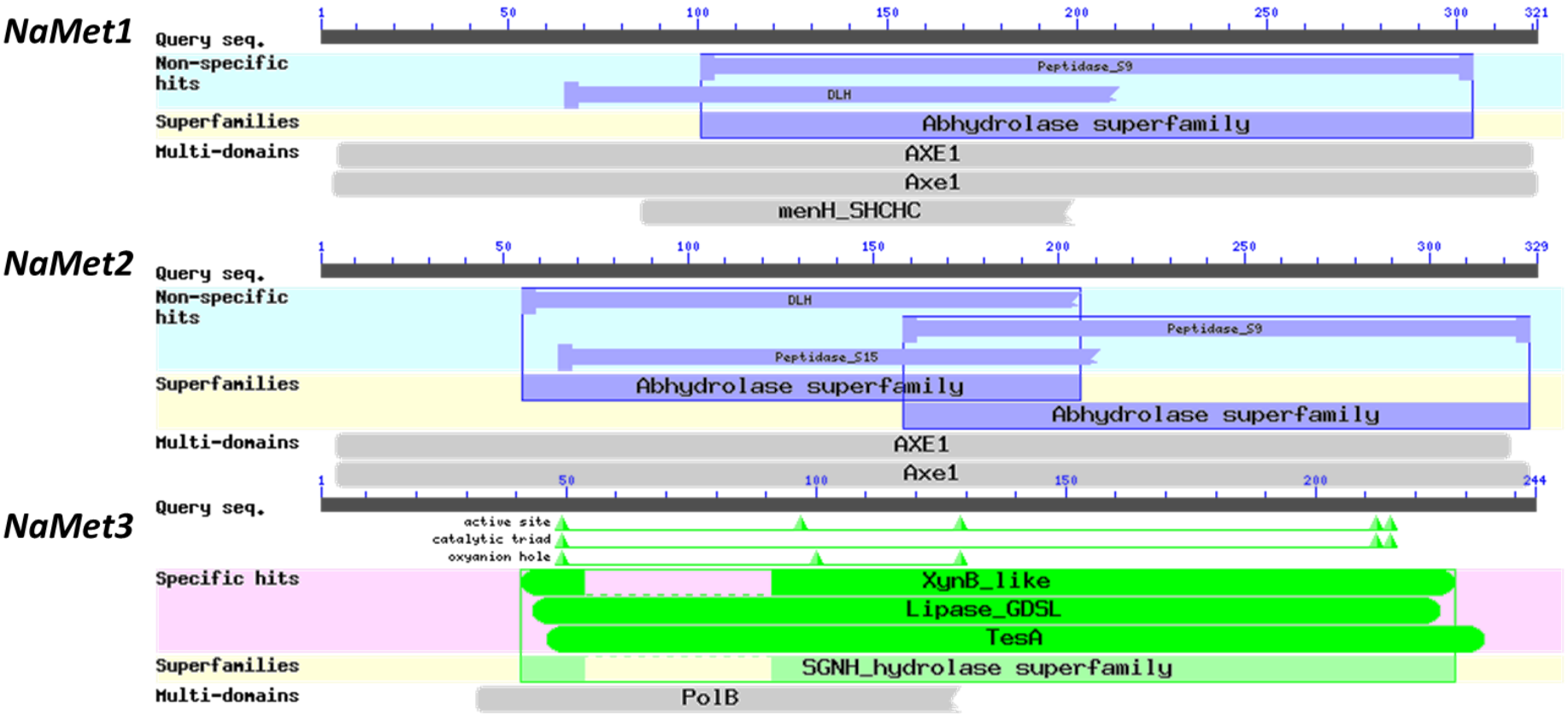


Figure 2.2: Domain architecture of *NaMet1-3*.

Positions occupied by each domain on the nucleotide sequence from a NCBI BlastX-CDD search output are shown (Marchler-Bauer et al., 2014).

2.2.7 Nucleotide and protein sequence analyses

The sizes of *NaMet1-3* were estimated to be 966bp, 990bp and 735bp, respectively. All three genes possessed high G+C content (>70%) in agreement with their predicted actinobacterial origin. Axe1 domains in *NaMet1* and *NaMet2* spanned positions 19-966 bp and 19-972 bp, respectively (Figure 2.2). Pairwise alignment of *NaMet1* and *NaMet2* showed that they shared ~35% and 47% identity and similarity, respectively. However, previous studies have shown that more than one AcXE, sometimes belonging to the same CE family, may be encoded within the same organism (Blum et al., 1999, Shao and Wiegel, 1995, Montanier et al., 2009).

In silico characterization of the translated proteins, NaM1-3 (Appendix 1E), indicated that they had the following isoelectric points and molecular weights: 5.11/35.6 kDa, 5.47/35.9 kDa and 5.78/26 kDa, respectively. A signal peptide sequence was predicted between positions 1-32 of NaM3. No signal peptide sequence was found in NaM1 and NaM2 suggesting that both proteins were intracellular. Signal peptides were also absent in previously characterized CE7 enzymes (Degrassi et al., 2000, Tian et al., 2014, Levisson et al., 2012, Drzewiecki et al., 2010). The GxSQG and GDS(L) conserved motifs, characteristic of the CE7 (NaM1 and NaM2) and CE3 (NaM3) CAZy families, respectively, were identified. MSAs showed that NaM1 and NaM2 possessed the Ser-His-Asp catalytic triad, typical of CE7 esterases (Biely, 2012). Active site residues were located at Ser185&187, His304&307 and Asp275&273, respectively. Residues synonymous with SGNH hydrolase active sites (Akoh et al., 2004) were located at Ser49, Gly106, Asn163 and His225 of NaM3.

Table 2.2: Most closely related proteins to the putative novel AcXE-encoding genes (NaMet1, 2 and 3) and their source organisms.

	Protein	Organism	Phylum/Family	Query cover (%)	E-value	Identity (%)	%GC content
NaMet1	Hypothetical	*Actinopolymorpha alba*	*Actinobacteria* / *Norcardiodiacea*	98	5e-135	64	
	AcXE	*Saccharothrix espanaensis*	*Actinobacteria* / *Pseudonorcardiacea*	98	8e-131	61	71.4
	ACE	*Paenibacillus ehimensis*	*Firmicutes* / *Paenibacillaceae*	96	4e-130	56	
NaMet2	AcXE	*Arthrobacter* sp. 35W	*Actinobacteria* / *Micrococcacea*	96	2e-151	69	
	AcXE	*Microbacterium mangrove*		95	9e-129	65	72.5
	AcXE	*Georgenia* sp. SUBG003	*Actinobacteria* / *Bogoriellacea*	97	3e-125	62	
NaMet3	Lipolytic	*Pseudonocardia dioxanivorans* CB1190		86	6e-67	58	
	Hypothetical	*Pseudonocardia dioxanivorans*	*Actinobacteria* / *Pseudonorcardiacea*	86	6e-67	58	74.8
	Hypothetical	*Pseudonocardia asaccharolytica*		81	8e-64	59	

Data was retrieved from the non-redundant database using BLASTX 2.4.0+ (Altschul et al., 1997).

2.2.8 PCR mining and sequencing of putative novel AcXE-encoding genes

PCR amplification using specific primers resulted in non-specific amplification (Figure 2.3a). Amplicons within the range of the expected gene sizes (~966bp, 990bp and 735bp) were gel-purified and confirmed as single bands on agarose gels (Figure 2.3b). However, DNA sequencing repeatedly yielded poor results, synonymous with mixed DNA templates (Figure 2.3c). Repeated PCR amplification, agarose gel electrophoresis and then gel or direct PCR product purification gave no improvement. The metagenomic database was interrogated using the specific primer sets, as well as ~50 other randomly designed primer sets specific to the three ORFs. This enquiry revealed multiple possible amplicons per set including the selected ORFs. Some of these potential amplicons were observed to be of similar sizes with the target ORFs making it difficult for the amplification of single genes. Thus, following PCR amplification, each single band observed on agarose gels consisted of more than one amplicon set resulting in poor sequencing data. This amplification of mixed DNA has been reported (Polz and Cavanaugh, 1998) and PCR methods are known to be detrimental to efficient evaluation of microbial diversity within communities (Gonzalez et al., 2012). Consequently, the option of gene synthesis was explored.

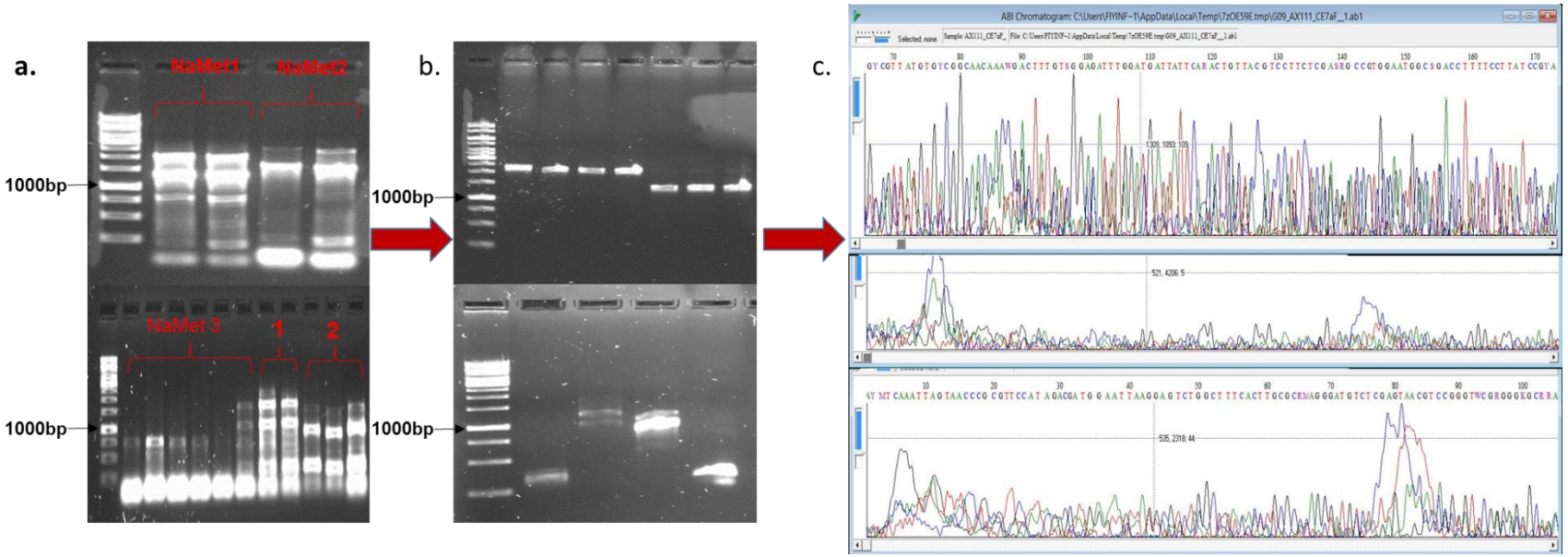


Figure 2.3: Agarose gel images of PCR screening of Namib metagenomic library

a. Non-specific-amplification using primers specific for *NaMet1-3* **b.** Gel purified putative AcXE bands from Figure 2.3a; c. Sequence electrophoresis chromatogram showing mixed DNA.

Table 2.3: Primers used in this study

Primer Name	Primer sequence (5'-3')	PCR conditions	Target region	References
CE7aF	AAAACATATGGTGCCGCTGACGTTC	95°C, 8min; (95°C, 45s; 55°C, 25 s; 72°C, 1min) 30X; 72°C, 10min; 25µL reaction vol.	966bp*	This study
CE7aR	GAAGAATTCCTACAACCCTTGCAGG			
CE7bF	AAAACATATGGTGGCCCGCGCCGA		990bp*	
CE7bR	AAAAGAATTCTCAGGATCGGCCCAGC			
CE3F	GAAACATATGGTGGGAACGATCCGG		735bp*	
CE3R	AAAAGAATTCTCACCGCGCTGCTGG			
NM1F	AAAACTCGAGGTCCCGCTGACCTTTGATC		978 bp	This study
NM1R	ACCTCATATGTTACAGACCTTGCAGGAAC			
NM2F	ATAACTCGAGGTGGCGCGTGCCGATC		1002 bp	
NM2R	ATTTCATATGTTACGAGCGGCCCAGAAC			
NM3F	AAAACTCGAGGTTGGCACCATTCGCTTTGG		747 bp insert	
NM3R	AATACATATGTTAACGAGCTGCCGGACG			
A293G.for	CACGGTTATACCGGCAGCAGCGGTGATTGGAG	95°C, 5min; (98°C, 20s; 67°C, 15 s; 72°C, 3min 30s) 25X; 72°C, 7min; 100µL reaction vol.	978 bp SDM insert	This study
A293G.rev	CTCCAATCACCGCTGCTGCCGGTATAACCGTG			
T634C.for	CAGCTCCGGTGTACCCGCTTCTGTGTGACTTCCGTC			
T634C.rev	GACGGAAGTCACACAGAAGCGGGTACACCGGAGCTG			
[a]T7F	TAATACGACTCACTATAGGG	95°C, 3min; (95°C, 30s; 47°C, 30s; 72°C, 1min) 25X; 72°C, 10min; 100µL reaction vol.	T7-region *NaMet*1-3	Novagen pET28 Primers
[a]T7R	GCTAGTTATTGCTCAGCGG			

* For amplification from metagenome [a] For EpPCR (section 5.1.2) and sequencing

2.2.9 Synthesis and amplification of putative AcXE-encoding genes

De novo DNA synthesis is commonly used as it excludes the PCR mining process. Also, with cheaper and more advanced technologies for gene synthesis, it is relatively affordable (Kosuri and Church, 2014). Synthesized codon-optimized AcXE-encoding genes were received in a 2.71kb pUC57 construct. Each gene possessed an *Eco*RI N-terminus and *Nde*I C-terminus increasing the insert sizes to 978bp, 1002bp and 747bp for *NaMet1-3* respectively. The inserts were flanked by *Xho*I and *Bam*HI restriction sites from the plasmid. The genes were amplified using gene-specific primers (Table 2.3). Agarose gel electrophoresis (Figure 2.4) and sequencing confirmed that the amplicons were of the expected sizes and correct nucleotide sequence (Appendix 1F).

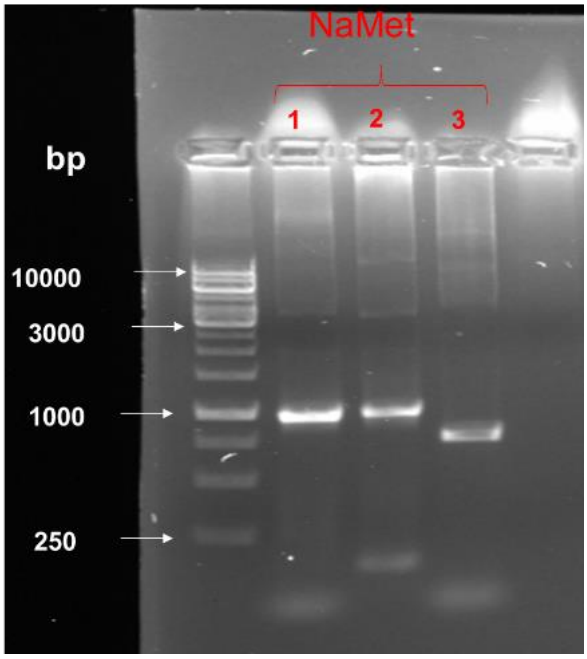


Figure 2.4: PCR amplification of synthesized putative AcXE-encoding genes.

2.3 CONCLUSION

This chapter describes the successful screening and isolation of AcXEs, from a soil metagenome using stringent screening protocols. Three putative AcXEs belonging to the CE3 and CE7 CAZy families were successfully identified using sequence-based, culture-independent methods. These results demonstrate the usefulness of sequence-based metagenomics as an alternative to function-based metagenomics. The diversity and richness of hydrolases within this metagenome (Le et al.,

2016, Makhalanyane et al., 2015, Vikram et al., 2016), was evident from the initial AcXE screening output and the subsequent inability to successfully isolate selected ORFs from the metagenomic library via PCR methods, without amplifying non-specific sequences. It is also worthy of note that none of the three putative AcXEs selected are known AcXEs. This indicates the uniqueness of the enzymes within the Namib Desert hypolithic community and suggests that there are more novel and potentially useful bioactive compound-encoding genes within uncultured microbial communities that are yet to be harnessed (Júnior et al., 2017, Wang et al., 2016). In view of this, it can be concluded with caution that *in silico* screening and gene synthesis is a valid alternative to the preparation of clone libraries and the use of functional screening strategies for gene-mining metagenomic projects.

2.4 REFERENCES

ADESIOYE, F. A., MAKHALANYANE, T. P., BIELY, P. & COWAN, D. A. 2016. Phylogeny, classification and metagenomic bioprospecting of microbial acetyl xylan esterases. *Enzyme and Microbial Technology,* 93**,** 79-91.

ADRIAENSSENS, E. M., VAN ZYL, L., DE MAAYER, P., RUBAGOTTI, E., RYBICKI, E., TUFFIN, M. & COWAN, D. A. 2015. Metagenomic analysis of the viral community in Namib Desert hypoliths. *Environmental Microbiology,* 17**,** 480-95.

AKOH, C. C., LEE, G. C., LIAW, Y. C., HUANG, T. H. & SHAW, J. F. 2004. GDSL family of serine esterases/lipases. *Progress in Lipid Research,* 43**,** 534-52.

AURILIA, V., MARTIN, J. C., MCCRAE, S. I., SCOTT, K. P., RINCON, M. T. & FLINT, H. J. 2000. Three multidomain esterases from the cellulolytic rumen anaerobe *Ruminococcus flavefaciens* 17 that carry divergent dockerin sequences. *Microbiology,* 146**,** 1391-1397.

BATEMAN, A., COIN, L., DURBIN, R., FINN, R. D., HOLLICH, V., GRIFFITHS-JONES, S., KHANNA, A., MARSHALL, M., MOXON, S. & SONNHAMMER, E. L. 2004. The Pfam protein families database. *Nucleic Acids Research,* 32**,** D138-D141.

BIELY, P. 2012. Microbial carbohydrate esterases deacetylating plant polysaccharides. *Biotechnology Advances,* 30**,** 1575-88.

BIELY, P., SINGH, S. & PUCHART, V. 2016. Towards enzymatic breakdown of complex plant xylan structures: State of the art. *Biotechnology Advances,* 34**,** 1260-1274.

BLUM, D. L., LI, X.-L., CHEN, H. & LJUNGDAHL, L. G. 1999. Characterization of an acetyl xylan esterase from the anaerobic fungus *Orpinomyces* sp. strain PC-2. *Applied and Environmental Microbiology,* 65**,** 3990-3995.

BORATYN, G. M., SCHAFFER, A. A., AGARWALA, R., ALTSCHUL, S. F., LIPMAN, D. J. & MADDEN, T. L. 2012. Domain enhanced lookup time accelerated BLAST. *Biology Direct,* 7**,** 12.

CANTAREL, B. L., COUTINHO, P. M., RANCUREL, C., BERNARD, T., LOMBARD, V. & HENRISSAT, B. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Research,* 37**,** D233-D238.

COWAN, D. A., SOHM, J. A., MAKHALANYANE, T. P., CAPONE, D. G., GREEN, T. G., CARY, S. C. & TUFFIN, I. M. 2011. Hypolithic communities: important nitrogen sources in Antarctic desert soils. *Environmental Microbiology Reports,* 3**,** 581-6.

DAVIDS, T., SCHMIDT, M., BÖTTCHER, D. & BORNSCHEUER, U. T. 2013. Strategies for the discovery and engineering of enzymes for biocatalysis. *Current Opinion in Chemical Biology,* 17**,** 215-220.

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.

DEGRASSI, G., OKEKE, B. C., BRUSCHI, C. V. & VENTURI, V. 1998. Purification and characterization of an acetyl xylan esterase from *Bacillus pumilus*. *Applied and Environmental Microbiology,* 64**,** 789-792.

DOUGHERTY, M. J., D'HAESELEER, P., HAZEN, T. C., SIMMONS, B. A., ADAMS, P. D. & HADI, M. Z. 2012. Glycoside hydrolases from a targeted compost metagenome, activity-screening and functional characterization. *BMC Biotechnology,* 12**,** 38.

DRZEWIECKI, K., ANGELOV, A., BALLSCHMITER, M., TIEFENBACH, K. J., STERNER, R. & LIEBL, W. 2010. Hyperthermostable acetyl xylan esterase. *Microbial Biotechnology,* 3**,** 84-92.

FERRER, M., MARTÍNEZ-MARTÍNEZ, M., BARGIELA, R., STREIT, W. R., GOLYSHINA, O. V. & GOLYSHIN, P. N. 2016. Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. *Microbial Biotechnology,* 9**,** 22-34.

FINN, R. D., CLEMENTS, J. & EDDY, S. R. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research,* 39**,** W29-37.

GONZALEZ, J. M., PORTILLO, M. C., BELDA-FERRE, P. & MIRA, A. 2012. Amplification by PCR artificially reduces the proportion of the rare biosphere in microbial communities. *PLoS One,* 7**,** e29973.

HAILES, H., BAUD, D., JEFFRIES, J., MOODY, T. & WARD, J. M. 2017. A metagenomics approach for new biocatalyst discovery: Application to transaminases and the synthesis of allylic amines. *Green Chemistry,* 19**,** 1134-1143.

ITOH, N., KAZAMA, M., TAKEUCHI, N., ISOTANI, K. & KUROKAWA, J. 2016. Gene-specific amplicons from metagenomes as an alternative to directed evolution for enzyme screening: a case study using phenylacetaldehyde reductases. *FEBS Open Bio,* 6**,** 566-575.

JÚNIOR, A. D. N. F., DAMÁSIO, A. R. L., PAIXÃO, D. A. A., ALVAREZ, T. M. & SQUINA, F. M. 2017. Applied Metagenomics for Biofuel Development and Environmental Sustainability. *Advances of Basic Science for Second Generation Bioethanol from Sugarcane.* London: Springer International Publishing.107-129

KATOH, K. & STANDLEY, D. M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution,* 30**,** 772-780.

KLIBANOV, A. M. 2001. Improving enzymes by using them in organic solvents. *Nature,* 409**,** 241-246.

KOSURI, S. & CHURCH, G. M. 2014. Large-scale de novo DNA synthesis: technologies and applications. *Nature Methods,* 11**,** 499-507.

KUMAR, A., DHAR, K., KANWAR, S. S. & ARORA, P. K. 2016. Lipase catalysis in organic solvents: advantages and applications. *Biological Procedures Online,* 18**,** 2.

LARKIN, M. A., BLACKSHIELDS, G., BROWN, N., CHENNA, R., MCGETTIGAN, P. A., MCWILLIAM, H., VALENTIN, F., WALLACE, I. M., WILM, A. & LOPEZ, R. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics,* 23**,** 2947-2948.

LE, P. T., MAKHALANYANE, T. P., GUERRERO, L., VIKRAM, S., VAN DE PEER, Y. & COWAN, D. A. 2016. Comparative metagenomic analysis reveals mechanisms for stress response in hypoliths from extreme hyperarid deserts. *Genome Biology and Evolution,* 8**,** 2737-2747.

LEIS, B., ANGELOV, A. & LIEBL, W. 2013. Screening and expression of genes from metagenomes. *Advances in Applied Microbiology,* 83**,** 1-68.

LEVISSON, M., HAN, G. W., DELLER, M. C., XU, Q., BIELY, P., HENDRIKS, S., TEN EYCK, L. F., FLENSBURG, C., ROVERSI, P., MILLER, M. D., MCMULLAN, D., VON DELFT, F., KREUSCH, A., DEACON, A. M., VAN DER OOST, J., LESLEY, S. A., ELSLIGER, M. A., KENGEN, S. W. & WILSON, I. A. 2012. Functional and structural characterization of a thermostable acetyl esterase from *Thermotoga maritima*. *Proteins,* 80**,** 1545-59.

LORENZ, P., LIEBETON, K., NIEHAUS, F. & ECK, J. 2002. Screening for novel enzymes for biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space. *Current Opinion in Biotechnology,* 13**,** 572-577.

MAKHALANYANE, T. P., VALVERDE, A., GUNNIGLE, E., FROSSARD, A., RAMOND, J. B. & COWAN, D. A. 2015. Microbial ecology of hot desert edaphic systems. *FEMS Microbiology Reviews,* 39**,** 203-21.

MAKHALANYANE, T. P., VALVERDE, A., LACAP, D. C., POINTING, S. B., TUFFIN, M. I. & COWAN, D. A. 2013. Evidence of species recruitment and development of hot desert hypolithic communities. *Environmental Microbiology Reports,* 5**,** 219-24.

MARCHLER-BAUER, A., DERBYSHIRE, M. K., GONZALES, N. R., LU, S., CHITSAZ, F., GEER, L. Y., GEER, R. C., HE, J., GWADZ, M. & HURWITZ, D. I. 2014. CDD: NCBI's conserved domain database. *Nucleic Acids Research,* 43**,** D222-D22.

MARGOLLES-CLARK, E., TENKANEN, M., SÖDERLUND, H. & PENTTILÄ, M. 1996. Acetyl Xylan Esterase from *Trichoderma reesei* Contains an Active-Site Serine Residue and a Cellulose-Binding Domain. *European Journal of Biochemistry,* 237**,** 553-560.

MIRETE, S., MORGANTE, V. & GONZÁLEZ-PASTOR, J. E. 2016. Functional metagenomics of extreme environments. *Current Opinion in Biotechnology,* 38**,** 143-149.

MONTANIER, C., MONEY, V. A., PIRES, V. M., FLINT, J. E., PINHEIRO, B. A., GOYAL, A., PRATES, J. A., IZUMI, A., STALBRAND, H., MORLAND, C., CARTMELL, A., KOLENOVA, K., TOPAKAS, E., DODSON, E. J., BOLAM, D. N., DAVIES, G. J., FONTES, C. M. & GILBERT, H. J. 2009. The active site of a carbohydrate esterase displays divergent catalytic and noncatalytic binding functions. *PLoS Biol,* 7**,** e71.

POLZ, M. F. & CAVANAUGH, C. M. 1998. Bias in template-to-product ratios in multitemplate PCR. *Applied and Environmental Microbiology,* 64**,** 3724-3730.

POPE, P. B., DENMAN, S. E., JONES, M., TRINGE, S. G., BARRY, K., MALFATTI, S. A., MCHARDY, A. C., CHENG, J. F., HUGENHOLTZ, P., MCSWEENEY, C. S. & MORRISON, M. 2010. Adaptation to herbivory by the Tammar wallaby includes bacterial and glycoside hydrolase profiles different from other herbivores. *Proceedings of the National Academy of Sciences USA,* 107**,** 14793-8.

PUCHART, V., GARIEPY, M. C., SHARECK, F. & DUPONT, C. 2006. Identification of catalytically important amino acid residues of Streptomyces lividans acetylxylan esterase A from carbohydrate esterase family 4. *Biochimica et Biophysica Acta,* 1764**,** 263-74.

REEN, F. J., DOBSON, A. D. & O'GARA, F. 2016. Metagenomics as a Tool for Biodiscovery and Enhanced Production of Marine Bioactives. *The Marine Microbiome.* Springer.377-400

ROCHA-MARTIN, J., HARRINGTON, C., DOBSON, A. D. & O'GARA, F. 2014. Emerging strategies and integrated systems microbiology technologies for biodiscovery of marine bioactive compounds. *Marine Drugs,* 12**,** 3516-3559.

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.

SIEVERS, F., WILM, A., DINEEN, D., GIBSON, T. J., 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.

TAMURA, K., STECHER, G., PETERSON, D., FILIPSKI, A. & KUMAR, S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution,* 30**,** 2725-2729.

TAYLOR, E. J., GLOSTER, T. M., TURKENBURG, J. P., VINCENT, F., BRZOZOWSKI, A. M., DUPONT, C., SHARECK, F., CENTENO, M. S., PRATES, J. A., PUCHART, V., FERREIRA, L. M., FONTES, C. M., BIELY, P. & DAVIES, G. J. 2006. Structure and activity of two metal ion-dependent acetylxylan esterases involved in plant cell wall degradation reveals a close similarity to peptidoglycan deacetylases. *Journal of Biological Chemistry,* 281**,** 10968-75.

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.

TINDALL, B. J., ROSSELLÓ-MORA, R., BUSSE, H.-J., LUDWIG, W. & KÄMPFER, P. 2010. Notes on the characterization of prokaryote strains for taxonomic purposes. *International Journal of Systematic and Evolutionary Microbiology,* 60**,** 249-266.

UNTERGASSER, A., CUTCUTACHE, I., KORESSAAR, T., YE, J., FAIRCLOTH, B. C., REMM, M. & ROZEN, S. G. 2012. Primer3—new capabilities and interfaces. *Nucleic Acids Research,* 40**,** e115-e115.

VIKRAM, S., GUERRERO, L. D., MAKHALANYANE, T. P., LE, P. T., SEELY, M. & COWAN, D. A. 2016. Metagenomic analysis provides insights into functional capacity in a hyperarid desert soil niche community. *Environmental microbiology,* 18**,** 1875-1888.

WANG, C., DONG, D., WANG, H., MÜLLER, K., QIN, Y., WANG, H. & WU, W. 2016. Metagenomic analysis of microbial consortia enriched from compost: new insights into the role of Actinobacteria in lignocellulose decomposition. *Biotechnology for biofuels,* 9**,** 22.

WARNECKE, F., LUGINBÜHL, P., IVANOVA, N., GHASSEMIAN, M., RICHARDSON, T. H., STEGE, J. T., CAYOUETTE, M., MCHARDY, A. C., DJORDJEVIC, G. & ABOUSHADI, N. 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature,* 450**,** 560-565.

WONG, F. K., LACAP, D. C., LAU, M. C., AITCHISON, J. C., COWAN, D. A. & POINTING, S. B. 2010. Hypolithic microbial community of quartz pavement in the high-altitude tundra of central Tibet. *Microbial Ecology,* 60**,** 730-739.

CHAPTER THREE

EXPRESSION, PURIFICATION AND FUNCTIONAL CHARACTERISATION OF NOVEL AcXEs

3.0 INTRODUCTION

Since the first description of AcXEs (Biely et al., 1986, Biely et al., 1985), substantial amount of research has been aimed at understanding AcXEs (see Figure 1, Chapter 1). This process starts with identification and isolation of potential AcXE-encoding genes. Functional capacity can then only be confirmed after cloning and expression of such genes as well as purification and characterization of the resultant proteins using suitable assays. Several genes encoding AcXEs (Waters et al., 2012, Eminoğlu et al., 2015, Vincent et al., 2003) and other esterases (Rashamuse et al., 2014, Soliman et al., 2014, Zhu et al., 2013) have been successfully expressed in the pET vector and *E.coli* BL21 host systems as well as other bacterial and fungal host systems (Adesioye et al., 2016, Leis et al., 2013). Considering the codon preference of host organisms, codon-optimization of the target gene is advised where possible (Felczykowska et al., 2015) often resulting in higher gene expression and protein production levels (Leis et al., 2013, Elena et al., 2014). Different substrates, as described by Adesioye et al. (2016), have been used to evaluate AcXE function and biochemical characteristics. As accessory enzymes during bioconversion of lignocellulosic substrates to ethanol, AcXEs are typically introduced following a pretreatment step to remove residual acetyl groups from the substrate. Pretreatment steps include chemical (acidic or alkaline) or physical (steam explosion) processes. Alkaline treatment removes 57-90% of acetyl side groups from the polymeric xylan (Tang et al., 2018) and accessory enzymes have been found to be sensitive to the background sugar concentration in bioreactor slurries (Chen et al., 2012). Therefore, depending on the pretreatment method used, AcXEs to be introduced should be stable under the prevalent pH, temperature, and solvent conditions of the medium (Sharma et al., 2017). The average optimal temperature of enzymes used in the bioconversion of lignocellulose to biofuel is ~50°C (Oliva et al., 2017, Liguori et al., 2016).

In this chapter, the sub-cloning and expression of *NaMet1-3* as well as the procedures involved in the production, purification and characterization of recombinant NaM1, the only functional protein of the three, are described and discussed. Considering the extreme environmental conditions in the

Namib Desert, the robustness of the enzyme for activity under various conditions, including high solute concentrations and organic solvents, was investigated.

3.1 MATERIALS AND METHODS

3.1.1 Materials

Para-nitrophenol acetate (*p*-NPA), butyrate (*p*-NPB), octanoate (*p*-NPO) and palmitate (*p*-NPP), 4-methylumbelliferyl acetate (4-MUA), 2-naphtyl acetate (2-NA), birchwood xylan, 7-aminocephalosporanic acid (7-ACA) and dialysis tubing (MWCO 14 kDa, flat width 43 mm) were purchased from Sigma (Switzerland, Germany and China). Dream Taq DNA polymerase, Fast Digest restriction enzymes and T4 DNA ligase as well as PAGE stained and unstained molecular weight markers were purchased from ThermoFisher Scientific, Massachusetts, USA. Bovine serum albumin (BSA) and AcXE from Orpinomyces (AxeA) (Blum et al., 1999) were purchased from Melford Biolaboratories, Ipswich, UK and Megazyme, Wicklow, Ireland, respectively. All primers were synthesized by Inqaba Biotech, Pretoria, South Africa. Immobilised metal (cobalt) affinity Talon resins and Amicon ultracentrifugal filters (MWCO 10 kDa) were purchased from Takara, Clontech, CA, USA and Merck Millipore, South Africa respectively. Microwell (96-well) plates manufactured by Greiner Bio One, Frickenhausen, Germany were used. All competent cells used in this study were made calcium-competent (Sambrook and Russell, 2006). All solutions were prepared using distilled water. Other reagents and solvents used were of analytical grade.

3.1.2 Sub-cloning and transformation of putative novel AcXEs

Synthesized putative AcXE-encoding genes were sub-cloned into a pET28a (5369 bp) expression vector via an XhoI site on the pUC57 plasmid and EcoRI site on the insert. Attempts at PCR cloning into the NdeI – EcoRI sites on the inserts were unsuccessful. Cloning was done using a modified protocol (Sambrook et al., 1989). Extracted DNA of pET28a expression vector and each pUC57 construct containing *NaMet*1, 2 or 3 genes was double-digested with FastDigest restriction enzymes, XhoI and EcoRI, at 37°C for 10 min. The enzymes were subsequently inactivated by incubation at 80°C for 5 min and the digest mix was purified using the DNA clean-up kit described in section 2.1.7. Ligation of *NaMet1-3* into pET28a was achieved with a 3:1 (insert:vector) DNA ratio using T4 DNA ligase in a ligation mix prepared according to a modified manufacturer's instructions. The mix was

incubated at 16ºC overnight and the ligase was inactivated at 65°C for 10 min. Following this, 4µL of the ligation mix was transformed into *E.coli* JM109 and BL21 DE competent cells, according to the method of Sambrook et al. (1989), for storage and cloning, respectively. The transformation mix was plated on kanamycin-supplemented Luria Bertani agar (LBA-kan) plates and incubated overnight (12-18 hr) at 37°C. Transformants were selected for overnight cultivation and plasmid extraction using the Qiagen Plasmid Miniprep kit. To confirm successful cloning and transformation, restriction digestion (XhoI and EcoRI), PCR amplification and sequencing of the extracted plasmid DNA were carried out. The pUC57-*NaMet1-3* constructs were each directly transformed into *E.coli* JM109 cells for long-term storage. Duplicate glycerol (25%) stocks of all constructs were stored at -80°C.

3.1.3 Gene expression and protein production

The pET28a gene constructs (Section 3.1.2) each possessed an N-terminal 6X His-tag and were expressed in *E. coli* BL21 DE(3) host cells suitable for isopropyl β-D-1-thiogalactopyranoside (IPTG)-induced gene expression. Glycerol stocks of each construct was plated out on LBA-Kan. Starter cultures (10 mL LBB-kan) for gene expression were inoculated with one clone from each plate and cultivated overnight with shaking (150 rpm). In small scale protein production experiments, flasks containing 100 mL LBB-kan were inoculated with 5 mL starter culture. These were incubated at 37°C with shaking at 180 rpm until an optical density $\geq 0.7_{600\ nm}$ was attained. The degree of expression of each gene was evaluated over a range of temperatures (18-30°C), growth periods and IPTG concentrations (0.1-0.5 mM) by subsequently estimating the amount of protein produced using SDS-PAGE analysis. The lac operon of *E. coli* BL21 host cells, cultivated in LBB, was induced for expression of *NaMet*1, 2 and 3 genes with IPTG during the mid-log growth phase (0.6 - 0.7 $OD_{600\ nm}$). The cultures were then incubated at the most appropriate temperature for expression of each gene for 6 – 8 hrs. Subsequently, protein production was scaled-up to 1 L culture volumes in 2 L flasks.

3.1.4 Protein isolation and purification

Cells were harvested using a Thermoscientific Sorvall Lynx 6000 centrifuge (USA) at 5000 x g for 15 min at 4°C. Harvested cells were re-suspended in lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl,

pH 7) and lysed on ice using a Q-Sonica Q500 sonicator (Newtown, Connecticut, USA). The crude protein extract was separated from the cell debris by centrifugation at 17000 x g for 60 min at 4°C. The supernatant (crude protein) was purified by immobilized metal (Cobalt) affinity chromatography (IMAC). Lysis buffer was supplemented with 10 mM and 250 mM imidazole for wash and elution steps, respectively. The bound protein was eluted in five 2 mL fractions and analyzed by SDS-PAGE alongside the unbound and wash fractions. Pure protein fractions were pooled and the high salt buffer contents were exchanged with 25 mM Tris-HCl, 25 mM NaCl (pH 8) using Amicon filters (MWCO 10 kDa). The purified protein was stored at 4°C.

3.1.5 Polyacrylamide gel electrophoresis and Western blotting

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and native PAGE were performed to estimate the molecular weight of the crude and purified protein extracts. Polyacrylamide gels (10-12%) were cast as described (Laemmli, 1979). For SDS-PAGE, protein samples to be analyzed were incubated at 95°C for 5 min in 5X SDS sample buffer composed of 0.225 M Tris-HCl (pH 6.8), 5% (w/v) SDS, 30% (v/v) glycerol, 0.05% (w/v) bromophenol blue and 0.01 M DTT. In the case of native PAGE, the protein samples were not heated, and all buffers and polyacrylamide gels used were prepared without SDS or any denaturants. An unstained molecular weight marker (SDS-PAGE) or protein of known native molecular weight (native PAGE) and protein samples were loaded unto the gel. Gels were electrophoresed for 30 min at 40 A. Following this, the gels were stained for 20 minutes in staining solution: 0.1% (w/v) Coomassie Blue, 10% (v/v) glacial acetic acid and 50% (v/v) ethanol. The gels were de-stained in similar solution, without Coomassie Blue, with gentle rocking (60 rpm) until the background was clear. An immunoassay, using a KPL His-Detector Western Blot AP Colorimetric Kit, was carried out to confirm the presence of the His-tagged gene product. In this case, a pre-stained PAGE molecular weight marker was used.

3.1.6 Preparation of acetylated xylan

Birch wood xylan was acetylated by a modified method of Johnson et al. (1988). Dry xylan (10 g) was slowly added to 250 mL dimethyl sulfoxide (DMSO) with gentle stirring at room temperature until an even suspension of xylan was achieved. The suspension was heated to 55°C for 15-20 min, and

2 g of potassium borate was added with continued stirring at 55°C for an additional 10 min. Pre-heated (60°C) acetic anhydride (200 mL) was carefully added to the mix while stirring for 10 min until the xylan showed a slightly darker hue. Simultaneously, two sections of dialysis tubing were pre-treated according to the manufacturer's instructions and cut to suitable lengths. The reaction mixture (~500 mL) was placed in the dialysis tubing, sealed at the ends with pegs and dialyzed against tap water in 5 L tanks at 4°C for 4-5 days. The water was periodically changed during this period until the odor of either DMSO or acetic anhydride was completely absent. Distilled water was used for dialysis during the final 24 hours. After dialysis, the acetylated xylan (AX) mixture was distributed into 6-8 round bottom flasks and frozen at -20°C overnight, then lyophilized in an Alpha 2-4 LD Plus Christ freeze dryer at -10°C and 2.5 mbar for 48-72 hrs. Lyophilized AX was stored at -20°C.

3.1.7 Functional characterization of NaM1

Acetyl esterase activity was determined spectrophotometrically (Thermoscientific Multiskan GO spectrophotometer) by monitoring the release of *para*-nitrophenol (*p*-NP) from 100 µl substrate at $OD_{405\ nm}$ and 25°C using the molar absorption coefficient of *p*-NP. The assay mix contained 0.5 mM *p*-NPA, 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. The standard curves for products detected in assays are collated in Appendix 2A. All experiments were carried out in triplicates and error bars shown in figures reflect the standard deviation of data points from the mean.

3.1.7.1 Determination of molar absorption coefficients (ε)

The molar absorptivity (molar extinction coefficients) of *p*-NP and 4-methylumbelliferrone (4-MU) at pH 8 and at 405 and 354 nm respectively was determined according to the Beer-Lambert law, using the formula:

$$A = \varepsilon \, l \, c \qquad \textbf{Equation 1}$$

$$\varepsilon = \frac{A}{l \, c}$$

Where A is the absorbance of light by sample at a given wavelength, ε is the molar absorption

coefficient (L $mol^{-1}cm^{-1}$), l is the path length of light (cm) and c is the concentration of the light absorbing sample (mol L^{-1}).

Slopes from plots of absorbance vs concentration (0.01 - 0.5 mM) were used to determine ε for *p*-NP and 4MUA. ε for 2-naphthol at 330 nm and pH 8 was previously described by (Kelly and Butler, 1977). However, since all assays were carried out in microplates with a path length of 0.32 cm for ~100 µl assay mix, the slope was divided by the path length to determine the actual extinction coefficient.

Protein concentrations were determined using equation 1. ε for NaM1 was calculated using the formula (Pace et al., 1995, Edelhoch, 1967):

$$\varepsilon_{280}\ (M\text{-}1\ cm\text{-}1) = \Sigma W\ (5500) + \Sigma Y\ (1490) + \Sigma C\ (125) \qquad \textbf{Equation 2}$$

Where ΣW, ΣY and ΣC are the total number of tryptophan, tyrosine and cysteine residues, respectively, in the protein and the values are constants corresponding to the extinction coefficients of each residue.

3.1.7.2 Effects of temperature and pH

The 'optimum temperature (T_{opt})' for NaM1 activity was determined using acetyl esterase assays at a range of temperatures (20-60°C). The 'pH optimum' was determined using assays in 50 mM phosphate citrate (pH 4-7), Tris-HCl (pH 7-9) and CAPS (3-cyclohexylamino-1-propanesulphonic acid) (pH 9.51-11.1) buffers.

The thermal stability of the enzyme was monitored at 5°C intervals between 30 and 65°C by incubating NaM1 at the specified temperature for 1hr. pH stability was determined between pH 5 and 11 by incubating NaM1 in buffers at the desired pH at 35°C for 1 hr. Residual activity was also recorded after 1, 3 and 24 hrs.

Following thermal stability studies, thermal inactivation of NaM1 was observed by incubating NaM1 at 50, 55 and 60°C in assay buffer supplemented with 1M NaCl and measuring residual activity at 5-min intervals over a for 30-min period.

3.1.7.3 Effects of stabilizing solutes

Using methods modified from previous studies (Kavruk et al., 2013, Kaushik and Bhat, 2003,

Charoenkitpaiboon, 2014, Anderson, 2007), the effects of stabilizing solutes including salts (NaCl, KCl), sugars (0.25-1.5 M trehalose, 0.25 M sucrose, 0.25 M glucose), sugar alcohol (0.25 M mannitol) and non-catalytic protein (0.1-10 mg mL^{-1} BSA) on NaM1 thermal stability were evaluated. Residual activity of the enzyme was determined after incubation at temperatures between 30 and 55°C for 15 min (1 hr for trehalose) in assay buffer supplemented with a specified solute. The activity of NaM1 in the presence of increasing NaCl concentrations was tested by supplementing assay buffer with NaCl. Stability in NaCl at 40°C was determined by measuring residual activity after incubating NaM1 in varying NaCl concentrations for 1 hr.

3.1.7.4 Effects of metal ions and chemical agents

Relative activity was determined after incubating NaM1 at 40°C for 1 hr in assay buffer supplemented with potential activators or inhibitors including 10 mM divalent cations (Ca^{2+}, Fe^{2+}, Co^{2+}, Mn^{2+}, Cs^{2+}, Mg^{2+}, Ni^{2+} and Cu^{2+}), organic solvents (1% di-methyl sulphoxide, 1% toluene, 1% ethanol, 1% methanol), non-ionic detergents (1% Triton X-100, 1% Tween 20), ionic detergent (2 mM sodium dodecyl sulphate), 2 mM ethylenediaminetetraacetic acid, 2 mM β-mercaptoethanol and 0.1 – 10 mM phenylmethylsulfonyl fluoride. Following this, NaM1 activity with decreasing water content in assay buffer was further evaluated. Assay buffer was supplemented with increasing concentrations (1-100% v/v) of the organic solvents: DMSO, methanol and toluene.

3.1.7.5 Substrate specificity studies

Activity on *p*-NPB (0.05-2 mM), *p*-NPO and *p*-NPP was assayed as described for *p*-NPA (above). The release of 4-methylumbelliferone (4-MU) from 4-MUA (0.01-1 mM) and 2-naphthol from 2-NA (0.02-1 mM) in 50 mM Tris HCl, 250 mM NaCl (pH 8) were monitored spectrophotometrically at 354 nm and 330 nm, respectively. Known molar absorption coefficients 18.3 $mM^{-1}cm^{-1}$ (*p*-NP), 10.47 $mM^{-1}cm^{-1}$ (4-MU) and 1.5 $mM^{-1}cm^{-1}$ (2-naphthol) (Kelly and Butler, 1977) were used. Deacetylase activity on 7-ACA (0.1-2 mM) and 0.5% AX was determined in 1 ml assays as described by Johnson et al. (1988) and Kormelink et al. (1993). The 10-minute 7-ACA and AX assays were composed of 50 mM potassium phosphate buffer (pH 8) and final enzyme concentrations of 0.004 and 0.08 mg mL^{-1}, respectively. Both deacetylation reactions were terminated with 1 N H_2SO_4 and assay solutions

were placed on ice immediately after adding the stop solution. AcXE from *Orpinomyces* sp. (Blum et al., 1999) was used as a positive control to ascertain successful chemical acetylation of xylan, while birchwood xylan was used as a negative control. The release of acetic acid was determined by reverse-phase high performance liquid chromatography (RP-HPLC) using a Dionex Ultimate 3000 HPLC unit with a Phenomenex Luna C18 (2), 150 x 4.6 mm column and a Dionex photodiode array UV detector (210 nm). A modified method of Cawthray (2003) was employed using a 25 mM NaH_2PO_4 (pH 2.5): methanol (93:7 v/v) mobile phase in a gradient elution with the following conditions; mobile phase: acetonitrile (40:60) at 1 mL min^{-1} flow rate, 20 µL injection volume, 25°C column temperature. Acetic acid standards (0.1-10 mM) were prepared using HPLC grade acetic acid and a calibration curve was drawn using the Agilent 1100/1200 HPLC ChemStation software. The enzyme kinetic constants per substrate were determined by fitting NaM1 activity at varying substrate concentrations to an Eadie-Hofstee plot (Hofstee, 1959, Eadie, 1942).

3.2 RESULTS AND DISCUSSION

Three genes *NaMet*1, 2 and 3 were codon optimized, synthesized and cloned into a pUC57 cloning vector. These genes were further sub-cloned for protein production and functional characterization.

3.2.1 Cloning, expression and purification

Following transformation into *E. coli* JM109 cells (section 3.1.2), synthesized genes were successfully sub-cloned into pET28a expression vectors (Figure 3.1a) and transformed into *E.coli* BL21 cells yielding several positive clones. Agarose gel electrophoretic analysis of the digested recombinant genes confirmed the sizes as ~ 966, 990 and 735 bp for *NaMet*1, 2 and 3 respectively (Figure 3.1b).

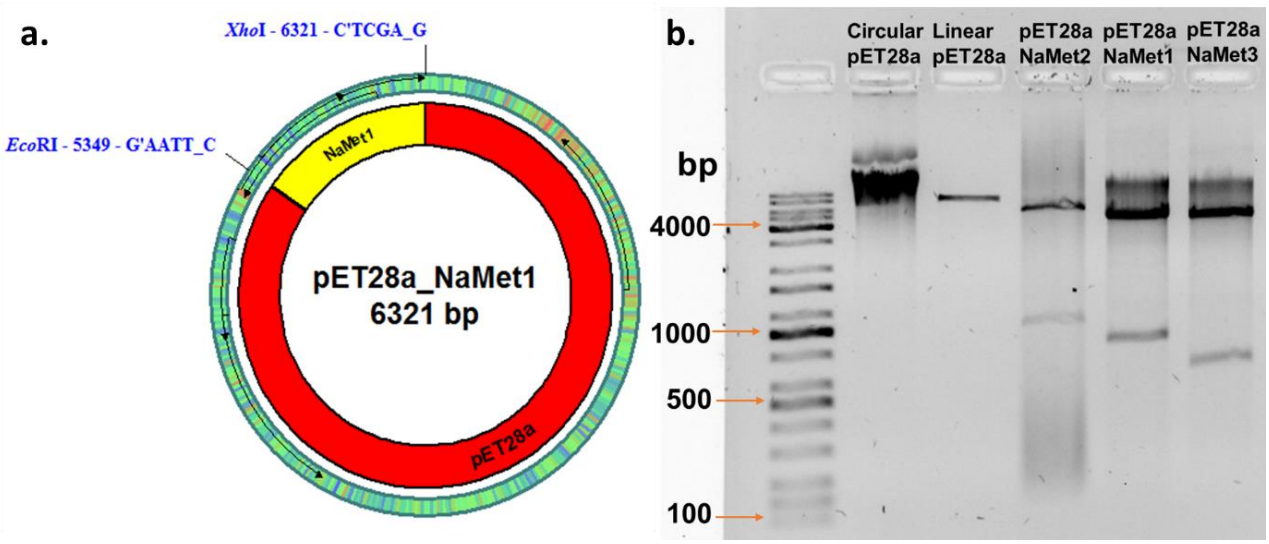


Figure 3.1: Recombinant *NaMet1* and Restriction digest

a: Recombinant *NaMet1* in pET28a. **b.** Restriction digest (lanes 4-6) showing empty pET28a (5369 bp) vector and *NaMet1-3* (735-990bp) indicating successful cloning. Recombinant *NaMet1* map was drawn using pDRAW32 software.

The most favorable expression conditions for *NaMet*1, 2 and 3 was observed at 25, 18 and 25 ^{o}C, respectively, while variations in IPTG concentrations had no detectable effects on protein production. After expression in *E.coli* BL21 DE cells, proteins of estimated molecular weights (~ 35.6 (NaM1), 35.9 (NaM2) and 26 (NaM3) kDa) were produced in the soluble fraction of the cell lysate. NaM1 was produced with a minimum yield of 26 mg per liter of *E.coli* BL21 cell culture (Figure 3.2a) visible via SDS-PAGE analysis of the soluble fraction, indicating that NaM1 was intracellularly produced, in agreement with the absence of a signal peptide (section 2.2.7). The presence of 6X His-tag NaM1 protein was confirmed with the appearance of a ~36 kDa band following Western blotting (Figure 3.2b).

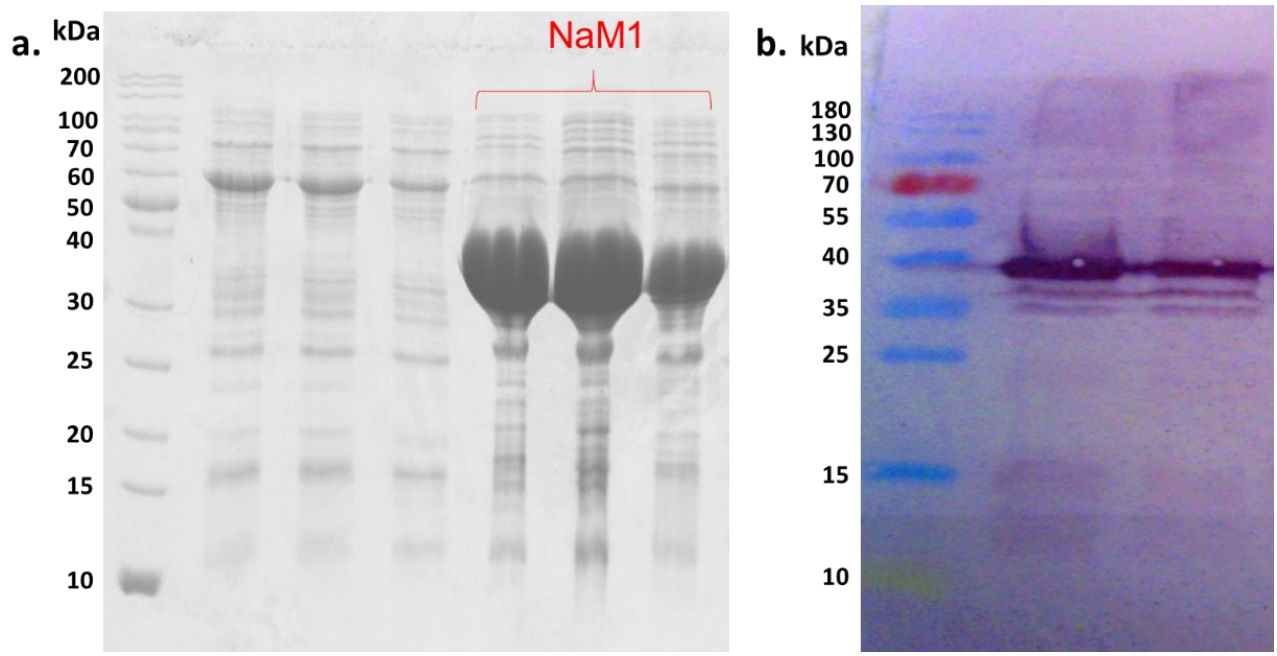


Figure 3.2: Overexpression and Western blot of NaM1

a. SDS-PAGE gel showing expression of empty pET28a vector (lanes 2-4) over-expression of NaM1 under optimum conditions (lanes 5-7). **b.** Western blot of 6X His-tagged NaM1.

Preliminary esterase activity tests using tributyrin agar well diffusion assays confirmed that NaM1 was an active enzyme as shown by the presence of clear zones (halos) around wells impregnated with crude extracts of NaM1. NaM1 was shown to be functional and was, thus, selected for further characterization. NaM2 and 3 were poorly expressed compared to NaM1 and did not show deacetylase activity. This is in agreement with a similar study (Dougherty et al., 2012) where only 7 of 22 predicted CAZy ORFs (including 1 of 2 predicted AcXEs) from sequence-based metagenomic screens showed functionality after expression. In another *in silico* CAZy bioprospecting study, only 1 of 2 predicted CAZyme ORFs was active (Allgaier et al., 2010). The low expression levels and inactivity on tributyrin of NaM2 and 3 might be due to factors such as the choice of expression vector and/or host system or substrate preference (Adesioye et al., 2016) which were not further explored in this study.

IMAC His-tag purification of NaM1 culture extract yielded a >95% pure protein and further processing using ion-exchange fast-pressure liquid chromatography (FPLC) methods yielded no significant differences in purity (Figure 3.3).

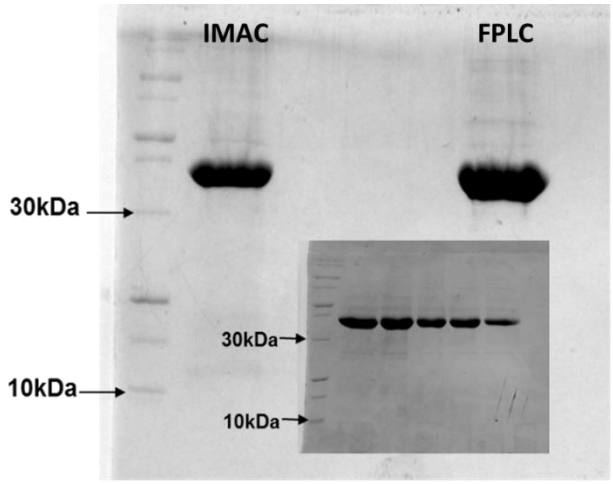


Figure 3.3: SDS-PAGE gels of IMAC cobalt-affinity (inset) and anion exchange FPLC purified fractions of ~36 kDa NaM1

3.2.2 Functional Characterization of NaM1

Functional characterization of NaM1 involving spectrophotometric activity was carried out using experimentally determined extinction coefficients of NaM1 (Section 3.1.7.1). The ε_{280} for *p-NP*A, 4MUA and NaM1 calculated using equations 1 and 2 were estimated to be 18.3, 10.47 and 48,985 $M^{-1}cm^{-1}$ respectively.

3.2.2.1 pH 'optimum' and stability

NaM1 was found to be stable between pH 6.5 and pH 10 and retained up to 40% activity after 72 hrs of incubation at 35°C at pH 8. 'Optimal' activity was observed at pH 8.5 (Figure 3.4).

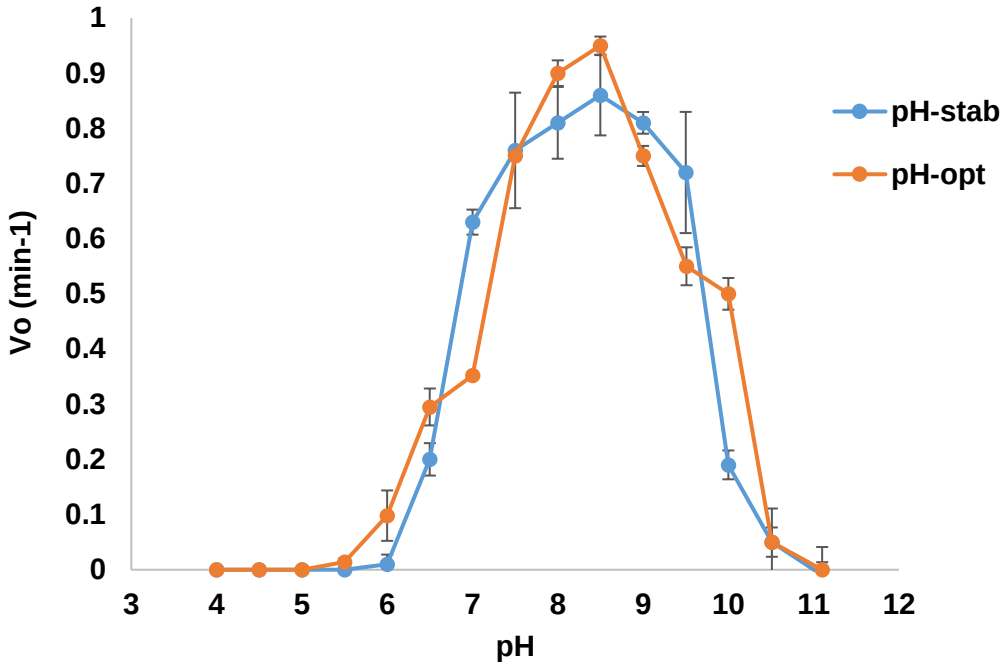


Figure 3.4: pH 'optimum' and stability profile of NaM1

3.2.2.2 Temperature 'optimum' and stability

Although, NaM1 shared high homology with thermostable CE7 AcXEs which possess moderate (T_{opt} – 45°C) (Degrassi et al., 2000) to high (T_{opt} - 90°C) (Levisson et al., 2012) temperature activity profiles, initial assays showed that NaM1 was 'optimally' active at 30°C (Figure 3.5a). This is the lowest T_{opt} reported for characterized CE7 enzymes. Less than 10% activity was retained after incubation of NaM1 at 40°C for 1 hr. However, in the presence of 1 M NaCl, the 'temperature optimum' increased to 35°C. Also, NaM1 retained 33% activity after incubation at 50°C for 1 hr in 1M NaCl, but no significant residual activity was detected with similar treatment at 55°C (Figure 3.5b). Hence, NaCl is clearly a stabilizing solute for NaM1 up to a restricted temperature limit, beyond which temperature-induced unfolding occurs.

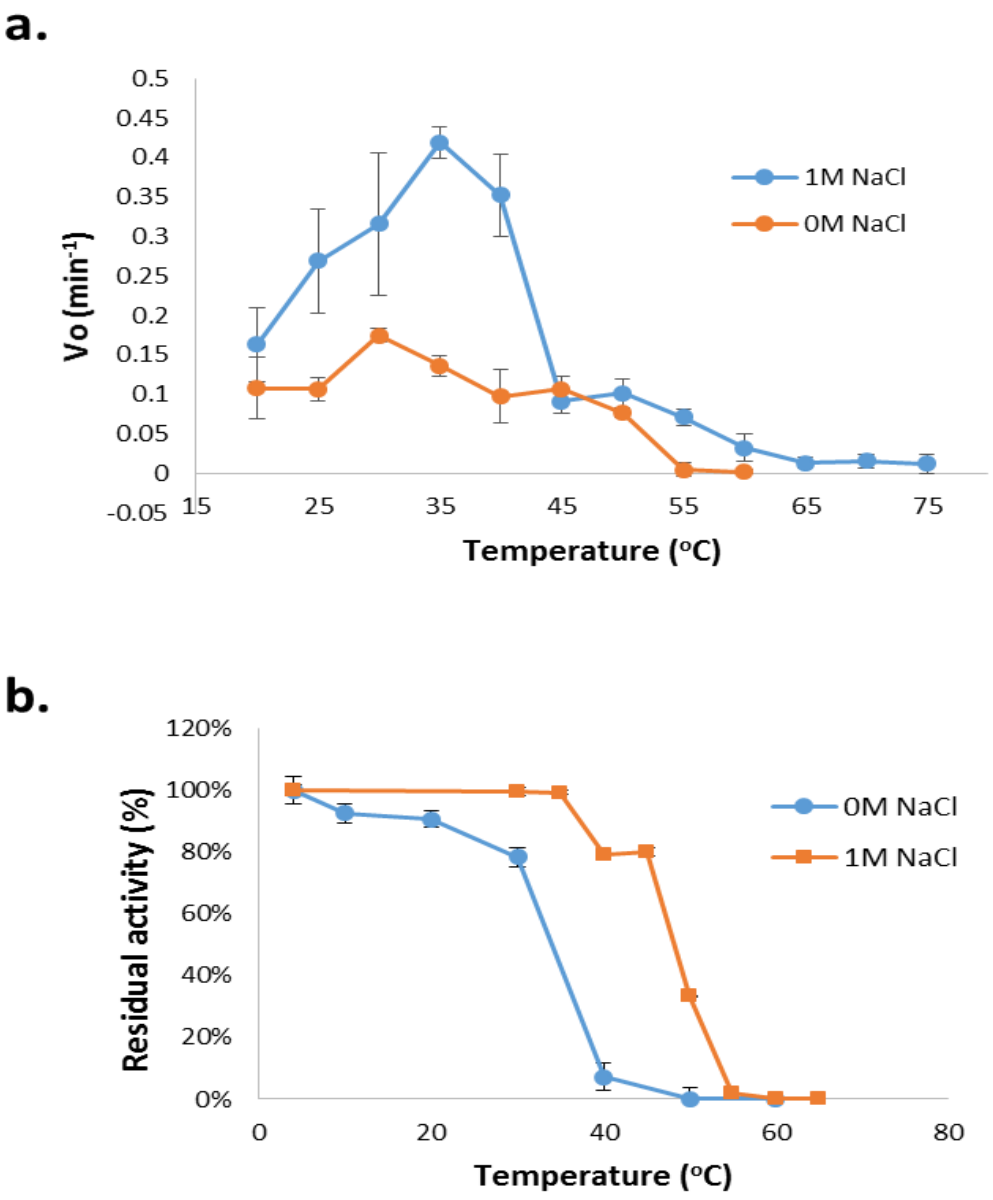


Figure 3.5: Effects of temperature on NaM1

a. Effects of temperature (20-75°C) on NaM1 activity with and without the addition of NaCl in assay buffer. **b.** Residual NaM1 acetyl esterase activity after incubation at various temperatures (4-60°C) for 1 hr.

Daniel and Danson (2013) showed in their studies that thermal inactivation occurs as a result of the relationship between the reaction temperature and the equilibrium temperature (T_{eq}) of the enzyme. Based on an equilibrium theory proposed by Daniel et al. (2001), the reduction in initial rates of NaM1 at assay temperatures higher than 30°C suggest that the reaction temperatures are above the T_{eq} of NaM1 (Daniel and Danson, 2013). Thus, the temperature at which the catalytically active and

inactive (but not denatured) species of NaM1 are at an equilibrium is shifted from ≤ 30°C (in the absence of NaCl) to ≤ 35°C (with NaCl). This suggests that increased ionic strength of the assay medium plays a role in delaying thermal unfolding and inactivation of NaM1.

3.2.2.3 Thermal inactivation profile

The thermal inactivation profile of NaM1 in 1M NaCl showed a sharp difference between thermal stability at 50°C and 55°C. After 15 minutes incubation at 55°C, only ~10% residual activity was recorded compared with ~80% residual activity after similar incubation period at 50°C (Figure 3.6). At 60°C, no activity was recorded for NaM1 after 5 minutes of incubation. These suggested that complete thermal denaturation of NaM1 occurred within 5 minutes of incubation at 60°C. However, this assay is more an indication of 'thermoactivity' (Daniel et al., 2001) rather than thermal stability. Hence, thermal stability was later confirmed by CD spectroscopy studies (section 5.2.6).

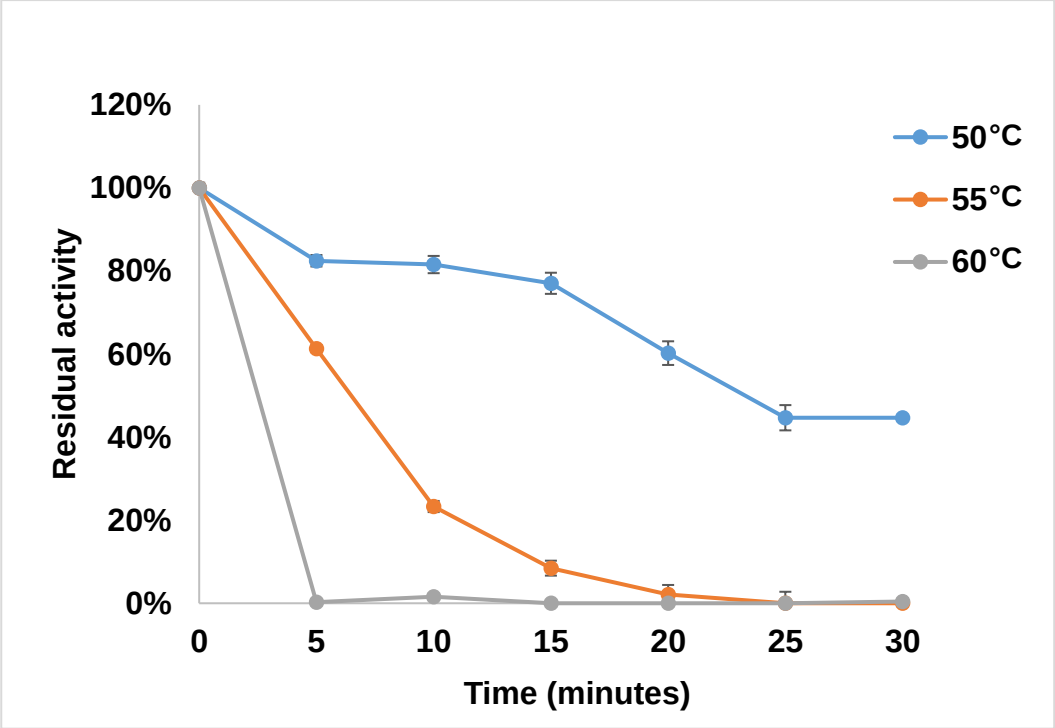


Figure 3.6: Thermal inactivation profile of NaM1 at 50, 55 and 60°C.

3.2.2.4 Effects of stabilizing solutes on activity and thermal stability

Solutes including polyols, sugars, salts, polyamine, amino acids and proteins are known to accumulate intracellularly during desiccation. These solutes act as osmo-protectants capable of

stabilizing native protein structures (Sleator and Hill, 2002, Giri, 2011). The stability and activity of NaM1 improved with increasing NaCl concentration. More than 50% residual NaM1 activity was detected after incubation in 5 M NaCl at 40°C for 1 hr (Figure 3.7) and optimal stability at 40°C was recorded in 1 M NaCl. Hence, further thermal stability studies were carried out in 1 M NaCl. The ability to remain active at high salt concentrations, conditions which may have caused some proteins to aggregate, suggests that NaM1 is a moderately halotolerant enzyme. The capacity to withstand extremely dry conditions, xerophily, is thought to be linked to halophily because, desiccation may occur as a result of high salt concentrations in an organism's immediate environment (Capece et al., 2013). These results suggest that NaM1 is potentially suitable for hydrolysis under conditions of low hydration or high solute contents.

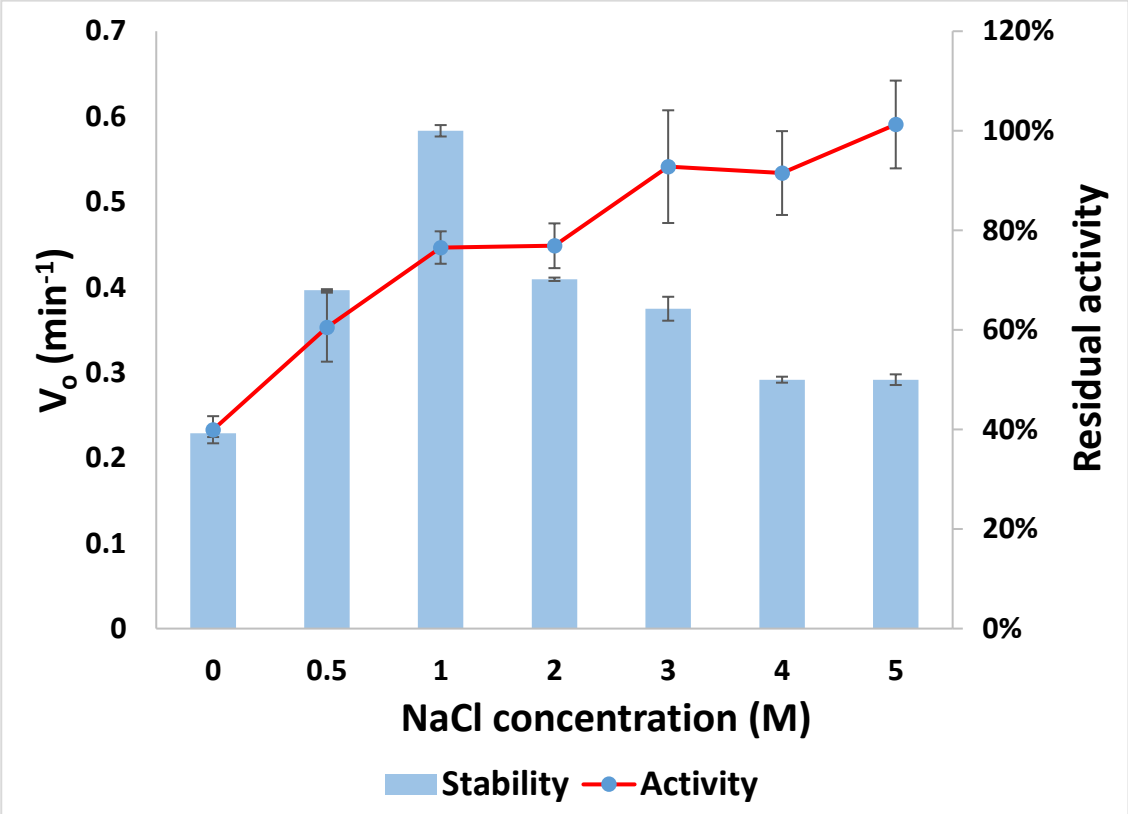


Figure 3.7: Effects of increasing NaCl concentration on NaM1 activity (primary y-axis) and stability at 40°C (secondary y-axis).

Trehalose is known as a compatible solute in halophilic organisms and has been reported to be produced in some bacteria in response to high salt concentrations (Kempf and Bremer, 1998). It is also a known protein stabilizing agent (Kaushik and Bhat, 2003, Jain and Roy, 2009) and has been

reported to stabilize a CE7 AcXE (Dicosimo et al., 2011). NaM1 was stabilized by trehalose when co-incubated for 1hr at 45°C. This was observed only when trehalose was added at concentrations ≥1 M resulting in 11% and 40% increased stability at 1 M and 1.5 M, respectively. However, it was observed that, the increased stabilizing effect of 1.5 M trehalose at 45°C was accompanied with reduced enzyme activity at temperatures ≤35°C (Figure 3.8). Increased thermal stability with reduced enzyme activity has been reported for enzymes incubated with trehalose (Sebollela et al., 2004, Luo et al., 2008, Carninci et al., 1998).

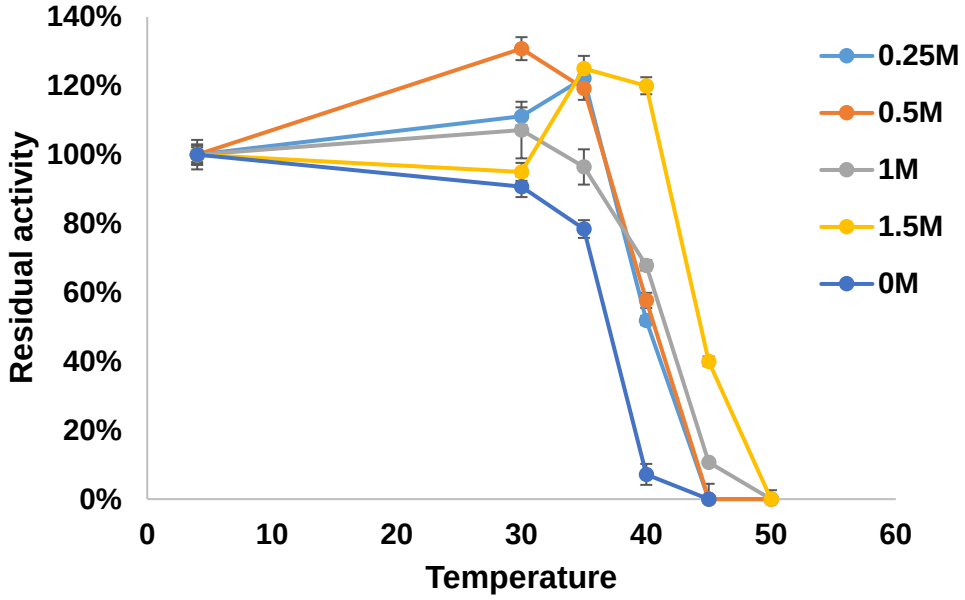


Figure 3.8: Effects of various trehalose concentrations on NaM1 thermal stability after incubation for 1 hr at various temperatures (4-50°C)

Trehalose and NaCl improved NaM1 thermal stability suggesting that they may be essential solutes in the original intracellular environment. From these results, it is inferred that NaCl and trehalose, to varying degrees, do not only act as osmoprotectants, but also slow down irreversible thermal denaturation of NaM1 in solution. There was no significant effect on NaM1's activity or thermal stability after 15 minutes of incubation with increasing concentrations of BSA. No increase in thermal stability beyond 45°C was observed for NaM1 when incubated in all stabilizing solutes tested, except NaCl, irrespective of further increases in solute concentration. This temperature profile further implies that NaM1 is well-adapted to the hypolith-niche-specific mesophilic conditions.

3.2.2.5 Effects of di-valent metal ions and chemical agents

NaM1 was stable in up to 10 mM of most divalent cations tested as illustrated in Figure 3.9a, but was completely inhibited by Cu^{2+} in agreement with a previous report of strong inhibition of the AcXE from *Bacillus subtilis* at 1 mM Cu^{2+} (Tian et al., 2014). NaM1 had lost 90% activity when incubated in 1mM Cu^{2+}. When assay solutions were supplemented with 1 – 10 mM Cu^{2+}, Mn^{2+}, Mg^{2+} or Co^{2+}, NaM1 was increasingly inhibited with increasing metal ion concentration. Slight enzyme activation by 10mM Ca^{2+} and Fe^{2+} contrasted observations with a 7-ACA deacetylase, EstD1, which was strongly inhibited by 1 mM Ca^{2+} (Ding et al., 2016), and an AcXE from *B. subtilis* which was slightly inhibited by 1 mM Ca^{2+} and Fe^{2+} (Tian et al., 2014). The stability observed for NaM1 in 10mM of several metal ions indicates that NaM1 may be useful for deacetylation of substrates from complex environmental sources containing high concentrations of such metal ions (Mirete et al., 2016).

Organic solvents are important for improving the solubility of several enzyme substrates during industrial processes. They also allow for reverse reactions and reduce unwanted water-dependent side reactions (Sharma and Kanwar, 2014). However, not all enzymes are stable in organic solvents. This has resulted in considerable industrial and research efforts at studying enzyme stability in organic solvents and engineering of enzymes with the aim of achieving stability in desired solvents (Stepankova et al., 2013, Kumar et al., 2016). Hence, the effects of various solvents, detergents and potential inhibitors were tested on NaM1 activity.

The most significant effects were observed with Triton-X, ethanol and tween-20 which severely inhibited (59-96%) hydrolysis by NaM1 (Appendix 2B). EDTA had a very minimal inhibitory effect on the enzyme activity contrasting a report by Ding et al. (2016) of strong inhibitory effects of EDTA, minimal influence of triton-X and ethanol, on a 7-ACA deacetylase (EstD1) from *Alicyclobacillus tengchongensis*. PMSF, a known serine protease inhibitor, had minimal effect on NaM1 activity. NaM1 maintained ~50% activity when assayed in a medium supplemented with 10 mM PMSF (Appendix 2C). This supports previous reports that CE7 enzymes possess low sensitivity to serine protease inhibitors (Montoro-García et al., 2011, Tian et al., 2014). The effect of increasing DMSO, toluene and methanol concentrations on NaM1 was further evaluated (Figure 3.9b). Most significantly, NaM1 retained 30% activity in the presence of 60% DMSO, but after 24-hr incubation in same, only ~2% activity was recorded. Methanol and DMSO which had the least inhibitory effects

on NaM1 may be recommended as solvents for its substrates in future applications. NaM1 may be considered for studies requiring enzyme activity at low hydration levels termed 'non-aqueous enzymology' (Carrea and Riva, 2000) such as those involving organic solvents or vapour phase substrates (Lind et al., 2004).

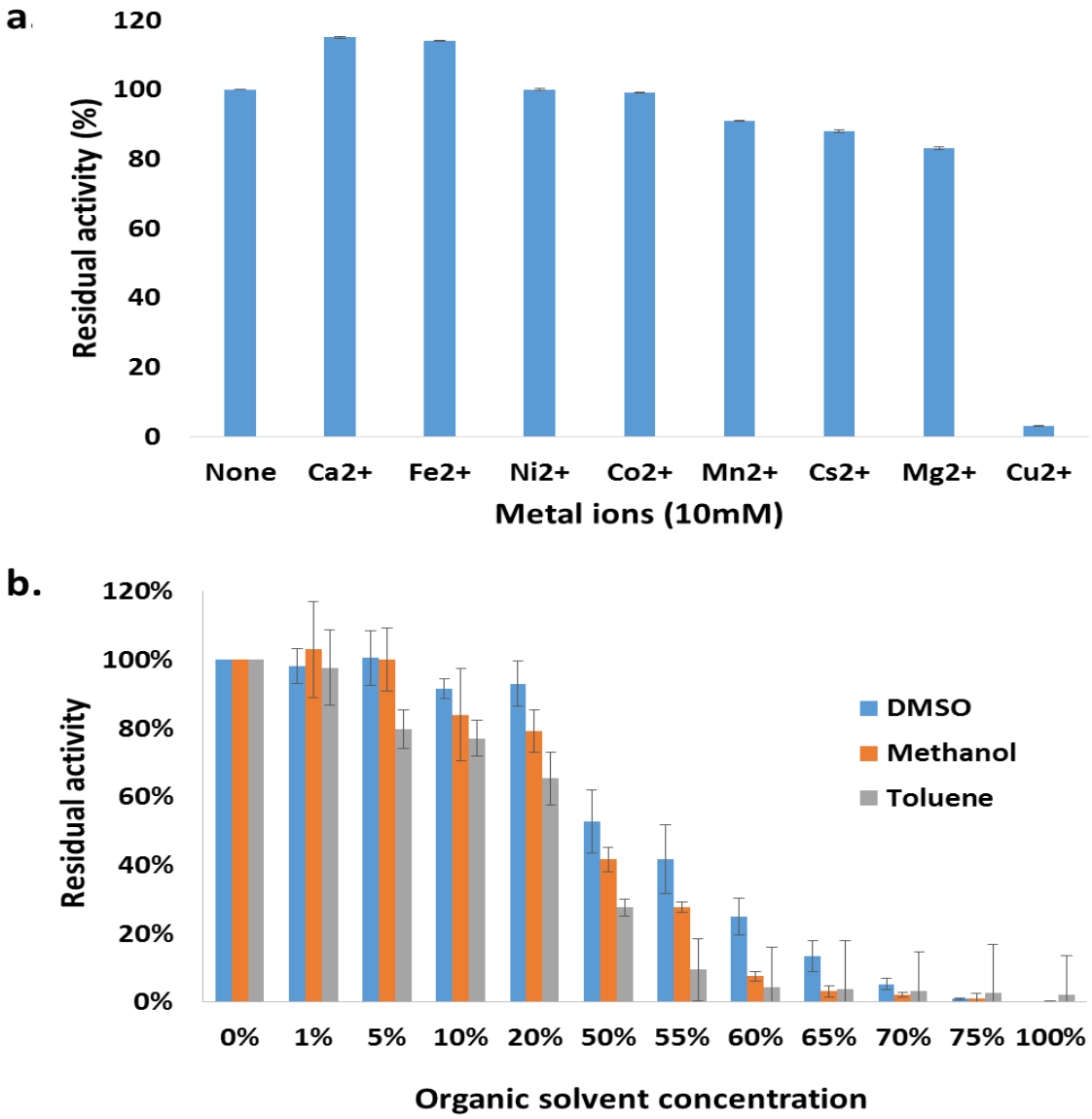


Figure 3.9: Effects of metal ions and organic solvents on NaM1 activity

a. Effects of 10 mM divalent metal ions on NaM1 activity after 1hr incubation at 40°C. **b.** Effects of increasing concentrations of solvents (DMSO, methanol and toluene) on NaM1 activity at 40°C.

3.2.2.6 Substrate specificity and enzyme kinetics

CE7 enzymes are known to have narrow substrate specificities. They are selectively active on short chain acetyl groups (Hedge et al., 2012) including substrates with C4 acyl moieties. However, *B. pumilus* CECT5072 CE7 esterase was reported to be inactive on butyrate (Montoro-García et al.,

2011) indicating that activity on butyrate was not uniform across all members. NaM1 showed a preference for short chain fatty acid esters by displaying strongest affinity for C2 and C4 substrates and no activity on long chain fatty acid substrates (*p*-NPO and *p*-NPP). Enzyme activity, specific activity and kinetic constants were determined for six substrates; *p*-NPA, *p*-NPB, 4-MUA, 2-NA, 7-ACA and acetylated xylan (AX) (Table 3.1). The lowest K_M, 0.1 mM, and highest catalytic efficiency, $3.26x10^6$ M^{-1} s^{-1}, were recorded for *p*-NPA.

Cephalosporin-C deacetylase (CCD) and 7-ACA deacetylase activities are unique to CE7 AcXEs. NaM1 also displayed 7-ACA deacetylase activity, at specific activity levels comparable to or higher than those observed for other CE7 7-ACA deacetylases (Table 3.2). NaM1 could, therefore, be a potential candidate for production of deacetoxy-7-aminocephalosporanic acid, an important precursor for the synthesis of β-lactamases in the pharmaceutical industry.

NaM1 showed lowest activity on acetylated xylan (AX) compared with other tested substrates (Table 3.1). This is in agreement with the low xylan deacetylating activity generally displayed by CE7 AcXEs (Biely, 2012). The activity of NaM1 on AX is similar to that reported for Axe1 from *Thermoanaerobacterium* sp. strain JW/SL YS485 (5.2 Umg^{-1}) (Shao and Wiegel, 1995), but lower than that of *B. pumilus* PS213 Axe1 (41 Umg^{-1}) (Degrassi et al., 2000) (Table 3.2). It is worth noting that apart from NaM1, only recombinant *B. pumilus* CE7 AcXEs have been reported to deacetylate AX (Degrassi et al., 2000, Martinez-Martinez et al., 2007, Montoro-García et al., 2011). CE7 esterases from *T. maritima* and *Thermonaerobacterium* which were previously annotated as AcXEs (Drzewiecki et al., 2010, Shao and Wiegel, 1995) have been re-annotated as non-AcXEs (Lorenz and Wiegel, 1997, Levisson et al., 2012), because deacetylation of xylan observed in the native enzyme was not observed in the recombinant enzyme. This was attributed to the cooperativity of enzymes in crude extracts (Biely et al., 1986) which allows short chain acetylated xylooligosaccharides, as opposed to bulky polymeric xylan, to be transported into the cell for deacetylation by intracellularly produced 'AcXEs' in whole cell culture extracts (Lorenz and Wiegel, 1997). Repeated assays on NaM1 showed consistently that NaM1 deacetylated chemically acetylated xylan (CAX), but with >60-fold lower specific activity than for *p*-NPA. Many AcXEs including CE7 AcXEs have shown increased activity on AX in the presence of xylanases indicating a preference for acetylated xylooligosaccharides. (Vincent et al., 2003, Blum et al., 1999). Since

NaM1 is produced intracellularly and is poorly active on AX, there is the possibility that NaM1 does prefer acetylated xylooligosaccharides. Alternatively, the bacterial species bearing NaM1 may have been exposed to a substrate similar to AX within the cell and, as a result, evolved to accommodate large acetylated substrates. This AcXE activity can be improved by substrate specificity-focused enzyme engineering.

Table 3.1: Substrate specificity and enzyme kinetics of NaM1

Substrate	V_{max} (Uml^{-1})	Specific activity (Umg^{-1})	K_M (mM)	k_{cat} (s^{-1})	Cat. Eff. ($M^{-1}s^{-1}$)
p-NPA	0.88 ± 0.02	488.90 ± 16.71	0.10 ± 0.015	293.30 ± 10.03	3.26×10^6
p-NPB	0.05 ± 0.005	12.96 ± 1.41	0.70 ± 0.14	7.67 ± 0.83	1.1×10^4
4-MUA	0.50 ± 0.01	277.8 ± 1.89	0.13 ± 0.02	166.70 ± 1.13	1.28×10^6
2-NA	0.40 ± 0.001	222.20 ± 12.72	0.20 ± 0.03	133.30 ± 7.63	6.67×10^5
7-ACA	0.80 ± 0.02	200 ± 0.038	0.46 ± 0.05	120 ± 12.53	2.6×10^5
0.5% AX	0.24 ± 0.08	6.05 ± 1.99	-	-	-

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

Table 3.2: Specific activities (Umg^{-1}) of functionally characterized 7-ACA deacetylases

Specific Activity (U/mg)	AX	7-ACA	CPC	*p*-NPA	Butyrate	α-NA	4-MUA	pH	T_{opt} (°C)	References
B. pumilus PS213 (rAXE)	13	179	26	88	NS	144	67	8.5	45	(Degrassi et al., 2000, Degrassi et al., 1998)
B. subtilis CICC 20034	-	888	484	2949	2.72	741	1086	7.0	50	(Tian et al., 2014)
B. subtilis SHS 0133	NS	12.4*	7.0*	201*	NS	NS	NS	7.0	55	(Takimoto et al., 1994)
B. subtilis 168‡	-	NS	1.18*	266.7*	NS	NS	NS	6.5	50	(Vincent et al., 2003)
B. pumilus CECT5072‡	NS	23.35	81	16.5	-	NS	NS	7.0	45	(Martinez-Martinez et al., 2007, Montoro-García et al., 2011)
A. tengchongensis (EstD1)	-	2.71	NS	NS	7.4*	79.6*	NS	8.5	65	(Ding et al., 2016)
T. maritima TM0077‡	-	1140[a]	376[a]	310.8*	NS	NS	NS	7.5	90	(Levisson et al., 2012)
Thermoanaerobacterium sp. JW/SL YS485 (rAXE) ‡	-	41	91	NS	NS	NS	47	7.0	75	(Lorenz and Wiegel, 1997)
T. maritima (EstA)	-	80	80	1095*	239*	NS	NS	8.5	95	(Levisson et al., 2009)
B. subtilis (YesT)	0.02	1235	356	1580	NS	217	NS	8.5	35	(Martínez-Martínez et al., 2008)
NaM1	6	200	NS	488.9	12.96	222	277.8	8.5	30	This study

+ active, - inactive, *p-NP*B: para-nitrophenol butyrate, AX: polymeric acetylated xylan, 7-ACA: 7-aminocephalosporanic acid, CPC: cephalosporin C deacetylase, α-NA: α-naphthyl acetate, 4-MUA: 4-methylumbelliferryl acetate, *catalytic efficiency k_{cat}/K_M ($s^{-1}mM^{-1}$), [a]turnover rate (s^{-1}), NS: not stated, ‡structurally characterized.

3.2.3 Database accession numbers

The nucleotide sequences of *NaMet*1, 2 and 3 have been deposited in the NCBI database under the following nucleotide accession numbers: KX818842, KX818843 and KX818844, respectively. 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.

3.3 CONCLUSION

In this chapter, predicted enzyme function based on sequence information was demonstrated and validated. The results obtained from functional characterization of NaM1 indicate that the Namib Desert soil hypolithic communities may produce highly adapted functional carbohydrate-active enzymes. The process of bioconversion of lignocellulose to biofuel is complex and involves various substrates, pretreatment steps, enzymes and conditions (Menon and Rao, 2012). Hence, each bioreactor protocol is custom-tailored per substrate. The suitability of NaM1 for any bioconversion process is determined by the substrate, pretreatment method and requirements of other hydrolyzing enzymes involved in the specific process (Menon and Rao, 2012, Sharma et al., 2017). In conclusion, NaM1 is a moderately alkalophilic, halophilic and mesophilic enzyme that is capable of xylan deacetylation. It is, however, unable to deacetylate xylan at the high temperatures typical of industrial lignocellulose-to-biofuel conversion processes (Hu and Saddler, 2018). NaM1 is, however, suitable for 7-ACA deacetylation industrial processes at mesophilic temperatures. It is a suitable candidate for further structural studies and enzyme engineering to improve its thermal stability and possibly AcXE activity.

3.4 REFERENCES

ADESIOYE, F. A., MAKHALANYANE, T. P., BIELY, P. & COWAN, D. A. 2016. Phylogeny, classification and metagenomic bioprospecting of microbial acetyl xylan esterases. *Enzyme and Microbial Technology,* 93**,** 79-91.

ANDERSON, J. A. 2007. Additive effects of alcohols and polyols on thermostability of pepper leaf extracts. *Journal of the American Society for Horticultural Science,* 132**,** 67-72.

BIELY, P., MACKENZIE, C., PULS, J. & SCHNEIDER, H. 1986. Cooperativity of esterases and xylanases in the enzymatic degradation of acetyl xylan. *Nature Biotechnology,* 4**,** 731-733.

BIELY, P., PULS, J. & SCHNEIDER, H. 1985. Acetyl xylan esterases in fungal cellulolytic systems. *FEBS Letters,* 186**,** 80-84.

BLUM, D. L., LI, X.-L., CHEN, H. & LJUNGDAHL, L. G. 1999. Characterization of an acetyl xylan esterase from the anaerobic fungus *Orpinomyces* sp. strain PC-2. *Applied and Environmental Microbiology,* 65**,** 3990-3995.

CAPECE, M. C., CLARK, E., SALEH, J. K., HALFORD, D., HEINL, N., HOSKINS, S. & ROTHSCHILD, L. J. 2013. Polyextremophiles and the constraints for terrestrial habitability. *Polyextremophiles.* Springer.3-59

CARNINCI, P., NISHIYAMA, Y., WESTOVER, A., ITOH, M., NAGAOKA, S., SASAKI, N., OKAZAKI, Y., MURAMATSU, M. & HAYASHIZAKI, Y. 1998. Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. *Proceedings of the National Academy of Sciences,* 95**,** 520-524.

CARREA, G. & RIVA, S. 2000. Properties and synthetic applications of enzymes in organic solvents. *Angewandte Chemie International Edition,* 39**,** 2226-2254.

CAWTHRAY, G. R. 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.

CHAROENKITPAIBOON, C. 2014. *Effects of osmolytes on the conformational stability and hydrodynamic radii of immunity protein 9 and human serum albumin.* University of East Anglia.

CHEN, X., SHEKIRO, J., FRANDEN, M. A., WANG, W., ZHANG, M., KUHN, E., JOHNSON, D. K. & TUCKER, M. P. 2012. The impacts of deacetylation prior to dilute acid pretreatment on the bioethanol process. *Biotechnology for Biofuels,* 5**,** 8.

DANIEL, R. M. & DANSON, M. J. 2013. Temperature and the catalytic activity of enzymes: A fresh understanding. *FEBS Letters,* 587**,** 2738-2743.

DANIEL, R. M., DANSON, M. J. & EISENTHAL, R. 2001. The temperature optima of enzymes: a new perspective on an old phenomenon. *Trends in Biochemical Sciences,* 26**,** 223-225.

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.

DEGRASSI, G., OKEKE, B. C., BRUSCHI, C. V. & VENTURI, V. 1998. Purification and characterization of an acetyl xylan esterase from *Bacillus pumilus*. *Applied and Environmental Microbiology,* 64**,** 789-792.

DING, J.-M., YU, T.-T., HAN, N.-Y., YU, J.-L., LI, J.-J., YANG, Y.-J., TANG, X.-H., XU, B., ZHOU, J.-P. & TANG, H.-Z. 2016. Identification and Characterization of a New 7-Aminocephalosporanic Acid Deacetylase from Thermophilic Bacterium *Alicyclobacillus tengchongensis*. *Journal of Bacteriology,* 198**,** 311-320.

DRZEWIECKI, K., ANGELOV, A., BALLSCHMITER, M., TIEFENBACH, K. J., STERNER, R. & LIEBL, W. 2010. Hyperthermostable acetyl xylan esterase. *Microbial Biotechnology,* 3**,** 84-92.

EADIE, G. 1942. The inhibition of cholinesterase by physostigmine and prostigmine. *Journal of Biological Chemistry,* 146**,** 85-93.

EDELHOCH, H. 1967. Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry,* 6**,** 1948-1954.

ELENA, C., RAVASI, P., CASTELLI, M. E., PEIRÚ, S. & MENZELLA, H. G. 2014. Expression of codon optimized genes in microbial systems: current industrial applications and perspectives. *Recombinant Protein Expression in Microbial Systems,* 5**,** 59-65.

EMINOĞLU, A., ÜLKER, S. & SANDALLı, C. 2015. Cloning, Purification and Characterization of Acetyl Xylane Esterase from *Anoxybacillus flavithermus* DSM 2641T with Activity on Low Molecular-Weight Acetates. *The Protein Journal,* 34**,** 237-242.

FELCZYKOWSKA, A., KRAJEWSKA, A., ZIELIŃSKA, S., ŁOŚ, J. M., BLOCH, S. K. & NEJMAN-FALEŃCZYK, B. 2015. The most widespread problems in the function-based microbial metagenomics. *Acta Biochimica Polonica,* 62**,** 161-166.

GIRI, J. 2011. Glycinebetaine and abiotic stress tolerance in plants. *Plant Signaling and Behavior,* 6**,** 1746-1751.

HEDGE, M. K., GEHRING, A. M., ADKINS, C. T., WESTON, L. A., LAVIS, L. D. & JOHNSON, R. J. 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.

HOFSTEE, B. 1959. Non-inverted versus inverted plots in enzyme kinetics. *Nature,* 184**,** 1296-1298.

HU, J. & SADDLER, J. N. 2018. Why does GH10 xylanase have better performance than GH11 xylanase for the deconstruction of pretreated biomass? *Biomass and Bioenergy,* 110**,** 13-16.

JAIN, N. K. & ROY, I. 2009. Effect of trehalose on protein structure. *Protein Science,* 18**,** 24-36.

JOHNSON, K., FONTANA, J. & MACKENZIE, C. 1988. Measurement of acetylxylan esterase in *Streptomyces* spp. *Methods in Enzymology,* 160**,** 551-560.

KAUSHIK, J. K. & BHAT, R. 2003. Why is trehalose an exceptional protein stabilizer? An analysis of the thermal stability of proteins in the presence of the compatible osmolyte trehalose. *Journal of Biological Chemistry,* 278**,** 26458-26465.

KAVRUK, M., ÖZALP, V. C. & ÖKTEM, H. A. 2013. Portable bioactive paper-based sensor for quantification of pesticides. *Journal of Analytical Methods in Chemistry,* 2013.

KELLY, S. J. & BUTLER, L. G. 1977. Enzymic hydrolysis of phosphonate esters. Reaction mechanism of intestinal 5'-nucleotide phosphodiesterase. *Biochemistry,* 16**,** 1102-1104.

KEMPF, B. & BREMER, E. 1998. Uptake and synthesis of compatible solutes as microbial stress responses to high-osmolality environments. *Archives of Microbiology,* 170**,** 319-330.

KORMELINK, F., LEFEBVRE, B., STROZYK, F. & VORAGEN, A. 1993. Purification and characterization of an acetyl xylan esterase from *Aspergillus niger*. *Journal of Biotechnology,* 27**,** 267-282.

KUMAR, A., DHAR, K., KANWAR, S. S. & ARORA, P. K. 2016. Lipase catalysis in organic solvents: advantages and applications. *Biological Procedures Online,* 18**,** 2.

LAEMMLI, U. 1979. Slab gel electrophoresis: SDS–PAGE with discontinuous buffers. *Nature,* 227**,** 680-5.

LEIS, B., ANGELOV, A. & LIEBL, W. 2013. Screening and expression of genes from metagenomes. *Advances in Applied Microbiology,* 83**,** 1-68.

LEVISSON, M., HAN, G. W., DELLER, M. C., XU, Q., BIELY, P., HENDRIKS, S., TEN EYCK, L. F., FLENSBURG, C., ROVERSI, P., MILLER, M. D., MCMULLAN, D., VON DELFT, F., KREUSCH, A., DEACON, A. M., VAN DER OOST, J., LESLEY, S. A., ELSLIGER, M. A., KENGEN, S. W. & WILSON, I. A. 2012. Functional and structural characterization of a thermostable acetyl esterase from *Thermotoga maritima*. *Proteins,* 80**,** 1545-59.

LEVISSON, M., SUN, L., HENDRIKS, S., SWINKELS, P., AKVELD, T., BULTEMA, J. B., BARENDREGT, A., VAN DEN HEUVEL, R. H., DIJKSTRA, B. W. & VAN DER OOST, J. 2009. Crystal structure and biochemical properties of a novel thermostable esterase containing an immunoglobulin-like domain. *Journal of Molecular Biology,* 385**,** 949-962.

LIGUORI, R., VENTORINO, V., PEPE, O. & FARACO, V. 2016. Bioreactors for lignocellulose conversion into fermentable sugars for production of high added value products. *Applied microbiology and biotechnology,* 100**,** 597-611.

LORENZ, W. W. & 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.

LUO, Y., LI, W.-M. & WANG, W. 2008. Trehalose: protector of antioxidant enzymes or reactive oxygen species scavenger under heat stress? *Environmental and Experimental Botany,* 63**,** 378-384.

MARTINEZ-MARTINEZ, I., MONTORO-GARCIA, S., LOZADA-RAMIREZ, J. D., 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.

MARTÍNEZ-MARTÍNEZ, I., NAVARRO-FERNÁNDEZ, J., DANIEL LOZADA-RAMÍREZ, J., GARCÍA-CARMONA, F. & SÁNCHEZ-FERRER, Á. 2008. YesT: a new rhamnogalacturonan acetyl esterase from Bacillus subtilis. *Proteins: Structure, Function, and Bioinformatics,* 71**,** 379-388.

MENON, V. & RAO, M. 2012. Trends in bioconversion of lignocellulose: biofuels, platform chemicals & biorefinery concept. *Progress in Energy and Combustion Science,* 38**,** 522-550.

MIRETE, S., MORGANTE, V. & GONZÁLEZ-PASTOR, J. E. 2016. Functional metagenomics of extreme environments. *Current Opinion in Biotechnology,* 38**,** 143-149.

MONTORO-GARCÍA, S., GIL-ORTIZ, F., GARCÍA-CARMONA, F., POLO, L. M., 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.

OLIVA, J. M., NEGRO, M. J., MANZANARES, P., BALLESTEROS, I., CHAMORRO, M. Á., SÁEZ, F., BALLESTEROS, M. & MORENO, A. D. 2017. A Sequential Steam Explosion and Reactive Extrusion Pretreatment for Lignocellulosic Biomass Conversion within a Fermentation-Based Biorefinery Perspective. *Fermentation,* 3**,** 15.

PACE, C. N., VAJDOS, F., FEE, L., GRIMSLEY, G. & GRAY, T. 1995. How to measure and predict the molar absorption coefficient of a protein. *Protein Science,* 4**,** 2411-2423.

RASHAMUSE, K., RONNEBURG, T., SANYIKA, W., MATHIBA, K., MMUTLANE, E. & BRADY, D. 2014. Metagenomic mining of feruloyl esterases from termite enteric flora. *Applied Microbiology and Biotechnology,* 98**,** 727-37.

SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. 1989. *Molecular cloning: a laboratory manual*, Cold spring harbor laboratory press.1546.

SAMBROOK, J. & RUSSELL, D. W. 2006. Preparation and transformation of competent *E. coli* using calcium chloride. *Cold Spring Harbor Protocols,* 1**,** 3932.

SEBOLLELA, A., LOUZADA, P. R., SOLA-PENNA, M., SARONE-WILLIAMS, V., COELHO-SAMPAIO, T. & FERREIRA, S. T. 2004. Inhibition of yeast glutathione reductase by trehalose: possible implications in yeast survival and recovery from stress. *The International Journal of Biochemistry and Cell Biology,* 36**,** 900-908.

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.

SHARMA, H. K., XU, C. & QIN, W. 2017. Biological Pretreatment of Lignocellulosic Biomass for Biofuels and Bioproducts: An Overview. *Waste and Biomass Valorization***,** 1-17.

SHARMA, S. & KANWAR, S. S. 2014. Organic solvent tolerant lipases and applications. *The Scientific World Journal,* 2014.

SLEATOR, R. D. & HILL, C. 2002. Bacterial osmoadaptation: the role of osmolytes in bacterial stress and virulence. *FEMS Microbiology Reviews,* 26**,** 49-71.

SOLIMAN, N. A., ABDEL-FATTAH, Y. R., MOSTAFA, H. E. & GABALLA, A. 2014. Heterologous expression of thermostable esterase gene from *Geobacillus thermoleovorans* YN under different expression promoters. *International Journal of Environmental Science and Technology,* 11**,** 119-126.

STEPANKOVA, V., BIDMANOVA, S., KOUDELAKOVA, T., PROKOP, Z., CHALOUPKOVA, R. & DAMBORSKY, J. 2013. Strategies for stabilization of enzymes in organic solvents. *ACS Catalysis,* 3**,** 2823-2836.

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.

TANG, Y., DOU, X., HU, J., JIANG, J. & SADDLER, J. N. 2018. Lignin Sulfonation and SO2 Addition Enhance the Hydrolyzability of Deacetylated and Then Steam-Pretreated Poplar with Reduced Inhibitor Formation. *Applied biochemistry and biotechnology,* 184**,** 264-277.

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.

VINCENT, F., CHARNOCK, S. J., VERSCHUEREN, K. H., TURKENBURG, J. P., SCOTT, D. J., OFFEN, W. A., ROBERTS, S., PELL, G., GILBERT, H. J. & DAVIES, G. J. 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.

WATERS, D. M., MURRAY, P. G., MIKI, Y., MARTINEZ, A. T., TUOHY, M. G. & FAULDS, C. B. 2012. Cloning, overexpression in Escherichia coli, and characterization of a thermostable fungal acetylxylan esterase from *Talaromyces emersonii*. *Applied Environmental Microbiology,* 78**,** 3759-62.

ZHU, Y., LI, J., CAI, H., NI, H., XIAO, A. & HOU, L. 2013. Characterization of a new and thermostable esterase from a metagenomic library. *Microbiological Research,* 168**,** 589-597.

CHAPTER FOUR

STRUCTURAL CHARACTERISATION OF NaM1

4.0 INTRODUCTION

Understanding the three-dimensional structure of a protein may often be helpful in comprehensively understanding its function. X-ray crystallography is one method used to elucidate the atomic structures of macromolecules. X-ray diffraction data from suitable protein crystals reveals the distribution of electrons in the crystal allowing the structure to be inferred. This technique is often used in protein engineering for deciphering and/or improving molecular interactions that determine stability, activity or substrate specificity of a particular protein. Crystal structures are available for about a third of biochemically characterized AcXEs (CAZy, 2017) including five CE7 family enzymes with 'temperature optima' between 45 and 90°C. X-ray crystallography has supported the elucidation of CE7 AcXE substrate specificities including their preference for acetyl groups and short chain xylooligosaccharides (Biely et al., 2014). Highly thermostable CE7 enzymes tend not to degrade polymeric xylan (Tables 3.2). The moderately thermophilic enzyme from *Bacillus pumilus* is, by contrast, an active AcXE. Additional structural information for both groups of CE7 esterases would allow for a better understanding of the broad substrate specificity in this enzyme family. While structural modeling predicts the structure of a protein fairly well, actual x-ray diffraction data from a protein gives a more accurate estimate of the coordinates of the atoms and molecules within the structure. Given the structural conservation observed from the superposition of a NaM1 model and other thermostable CE7 enzymes, molecular factors that determined the functional properties of these family of enzymes were more likely to be encoded in interactions between individual amino acids rather than between secondary elements of each protein. Hence, it would be more informative to study the functional properties of thermolabile NaM1 using a thermolabile CE7 AcXE model with similar structural conservation rather than a thermostable CE7 AcXE model. Thus, it became necessary to determine the crystal structure of NaM1.

This chapter describes the crystallization, X-ray crystallographic structure determination and structural characterization of the CE7 AcXE, NaM1.

4.1 MATERIALS AND METHODS

4.1.1 Crystallization and data collection

A 96-well, sitting-drop crystallization screen was set up for NaM1 using the Qiagen (USA) PEGs Suite and incubated at 18^{o}C. The reservoir solution and 7 to 10 mg/ml NaM1 solution were mixed in a 1:1 or 1:2 ratio for crystallization drops, where NaM1 was in 25 mM Tris HCl pH 7 and 25 mM NaCl. Initial crystal hits were improved by seeding as well as varying the crystallization conditions including pH, polyethylene glycol (PEG) concentration, temperature, buffer components and drop to buffer ratios to obtain optimal, single crystals for X-ray diffraction. To cryo-protect the crystals, the reservoir fluid was supplemented with 20 to 25% PEG 400 and the solution was cryo-cooled in the liquid nitrogen flow of an X-ray diffractometer to check for glass-like conditions. Crystals were harvested, flash-cooled in liquid nitrogen and stored or transported in an appropriate Dewar. X-ray diffraction data was collected on beamline ID23-1 of the European Synchrotron Radiation Facility, Grenoble, France. Diffraction data were processed automatically by the Grenoble automatic data processing system (GrenADeS) in fast processing mode (EDNA_proc, grenades_fastproc, grenades_parallelproc and autoPROC) (Delagenière et al., 2011)

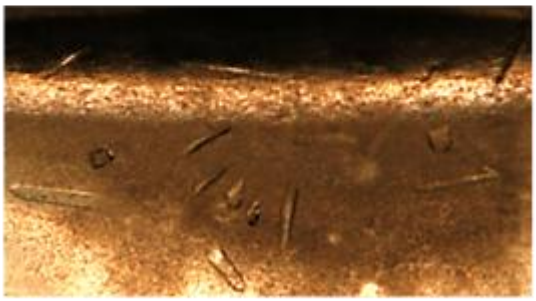
4.2 RESULTS AND DISCUSSION

4.2.1 Crystallization

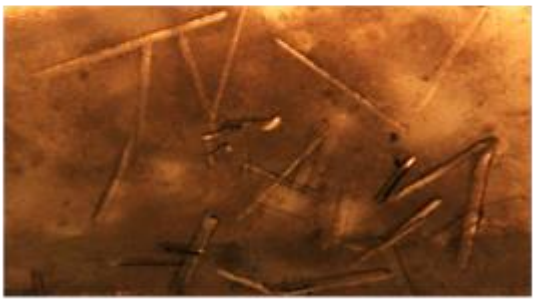
Hexagonal crystals were obtained in 0.1 M Tris HCl or 2-(N-morpholino) ethanesulfonic acid (MES) buffers pH 6.5 to 8.5 and 20 to 25% PEG 6000 to 10000 (w/v) (Figure 4.1).

0.1 M Tris HCl, pH 8.5, 20 % PEG 10000 (w/v)

0.1 M Tris HCl, pH 8.5, 25 % PEG 6000 (w/v)*



0.1 M MES, pH 6.5, 25 % PEG 8000 (w/v)



0.1 M MES, pH 6.5, 20 % PEG 10000 (w/v)

* obtained by seeding

Figure 4.1: NaM1 crystals obtained from sitting drop crystallization trials

4.2.2 Data collection and processing

Eight diffraction data sets were collected at 100 K from 1.7 to 2.7 Å resolution. Diffraction data were downloaded from the ESRF “information system for protein crystallography beamlines” (ISPyB) database. The space group was orthorhombic $P2_12_12_1$ with six monomers per asymmetric unit (Table 4.1) resulting in a solvent content of ~47 %.

Table 4.1: 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 (%)	22.03 (32.81) / 16.92 (27.34)
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, 1931, 27, 17600, 2264
Ramachandran favoured, allowed, outliers (%)	96.82, 2.76, 0.42
PDB code	6FKX

*shell of highest resolution in brackets

4.2.3 Structure determination

Starting phases for the structure factors, derived from the diffraction experiment, were determined by molecular replacement (MR) using hexameric AcXE from *Thermoanaerobacterium saccharolyticum* JW/SL YS485 (Protein Data Bank [PDB] code 3FCY) as a model. Phaser (McCoy, 2007), incorporated into PHENIX (Adams et al., 2010), located one hexamer per asymmetric unit using a maximum-likelihood algorithm and its AutoBuild module (Terwilliger et al., 2008) generated a first model for NaM1 based on the amino acid sequence of NaM1. The structure was cyclically improved by manual correction in COOT (Emsley and Cowtan, 2004) and refinement in PHENIX Refine (Afonine et al., 2012). Coordinates for MES and formate ligands for incorporation into the model were downloaded from the Hetero-compound Information Centre, Uppsala (HIC-Up 12.1) server (Kleywegt et al., 2003). The final model was structurally aligned to homologs using the COOT SSM superpose function.

The protein geometry was analyzed using COOT validate geometry analysis. Data collection and structure solution details are summarized in Table 4.1. The quality of the structure solution was evaluated using PROCHECK (Laskowski et al., 1993) and PHENIX MolProbity (Chen et al., 2010, Deis et al., 2013) validation tools. The overall NaM1 model was deemed to be of good quality on the basis of >90% of residues being located in the most favored regions of the Ramachandran plot (Ramachandran et al., 1963). Ligands including water molecules were located based on isolated peaks of electron density within reasonable distance of protein side chains. Ligand and inter-subunit interactions were analyzed using the Protein Interfaces, Surfaces and Assemblies' service (PISA) (Krissinel and Henrick, 2007), PDBsum (de Beer et al., 2014) and Ligplot (Wallace et al., 1995). Parameters describing the components of the NaM1 structure including ligands and water molecules are provided in Appendix 2E. All molecular images were generated using Pymol 1.6.0.0 (DeLano, 2002) except otherwise stated.

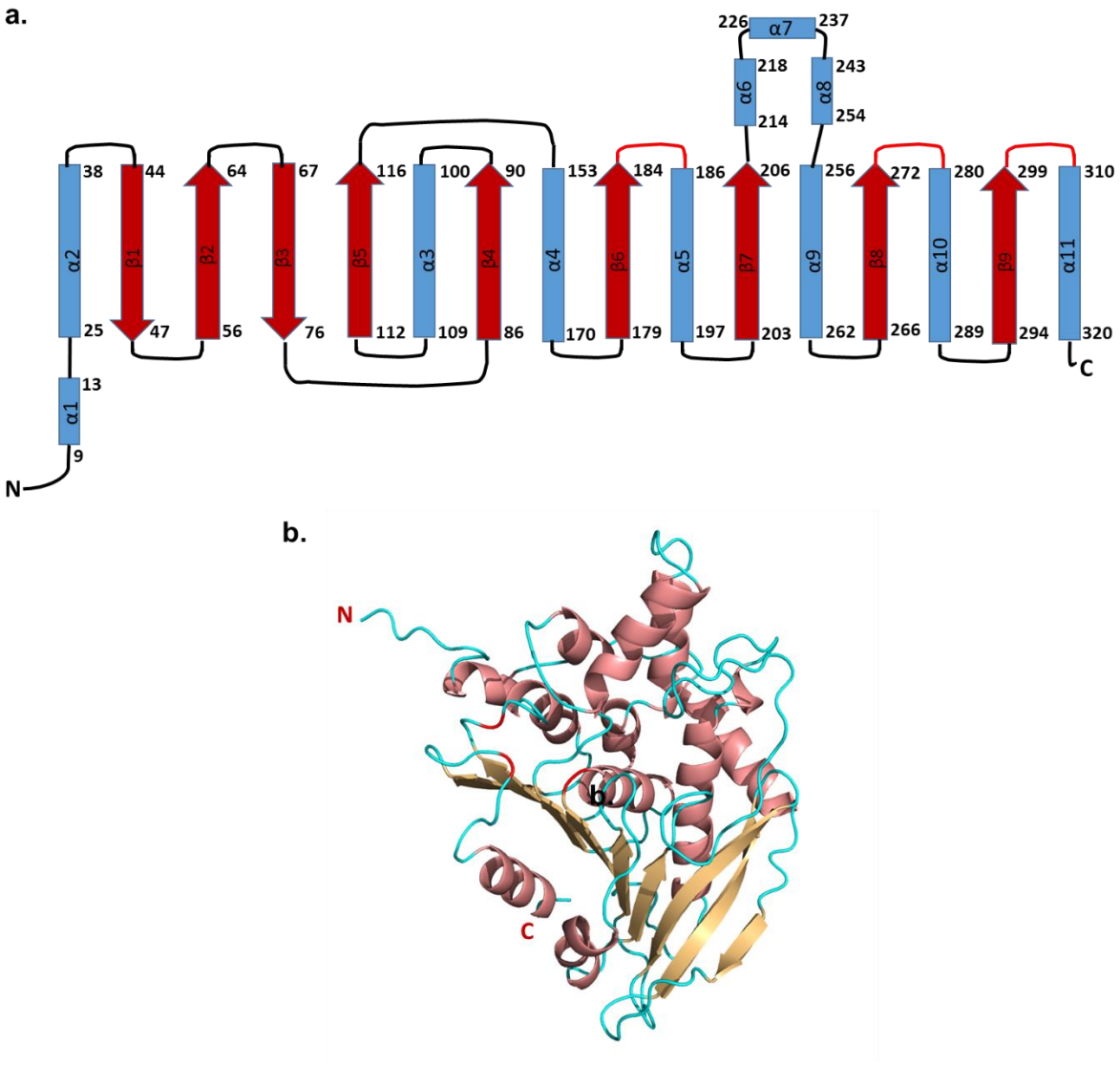


Figure 4.2: Tertiary structure of NaM1

a. NaM1 Topology: α-Helices (α1-11) are represented by blue cylinders, β-strands (β1-9) by red arrows with relative orientation, and connecting loops by black lines. Loops carrying active site residues (Ser185, His304, Asp275) are colored red. The typical α/β hydrolase fold begins with a β-hairpin (Arg56 – Arg76) and ends at the C-terminus, excluding the triple α-helix extension (Phe214 - Val254). **b.** NaM1 tertiary structure; α-helices are dark salmon in color, β-strands, tan and loops, cyan. Catalytic residues are in red.

4.2.4 Structure of NaM1

Each NaM1 monomer contains a central nine-stranded β-sheet of the sequence order 123546789 with an antiparallel arrangement for the first four β-strands (1235) and parallel for the last six (546789) strands. Two of 11 α-helices (3 and 11) are located on the one side of the β-sheet while the remaining nine are associated with the other side. The first five β-strands of the mixed β-sheet curve around α-helix 4, while the second half is largely flat (Figures 4.2a and b).

Residues Arg56 to Gly320 of NaM1 create a typical α/β hydrolase fold. NaM1 also shares a GXSXG (GYSQG) motif with other α/β-hydrolases located in the loop connecting β-strand 6 and α-helix 5 (β6-α5). NaM1, like other CE7 enzymes, differs from the canonical α/β hydrolase fold by the following additional structural elements: a three-helix insertion (α6, α7 and α8, Phe214 to Val254), an interface loop region (Gly120 to Leu140), and an N-terminal extension consisting of two α-helices (α1 and 2, Phe9 to Ala38) and a β-strand (β1, Val44 - Pro47) (Singh and Manoj, 2016b).

The NaM1 hexamer has a D3 point group symmetry, and may be thought of as a dimer of trimers, with one trimer consisting of chains A, B, C and the other of chains D, E, F (Figure 4.3a). The two trimers are arranged in an offset, interdigitating face-to-face. Each trimer has C3 symmetry with monomers rotated by 120^{o} around a central axis relative to each other. The N-terminus of each monomer is located near the three-fold symmetry axis, limiting the size of the central opening (Figure 4.3a and b). The NaM1 hexamer encloses a significant void between the two trimers with small pores around the three-fold rotational axis on either side. The total surface and buried area of the NaM1 hexamer are 64900 Å and 21430 Å, respectively.

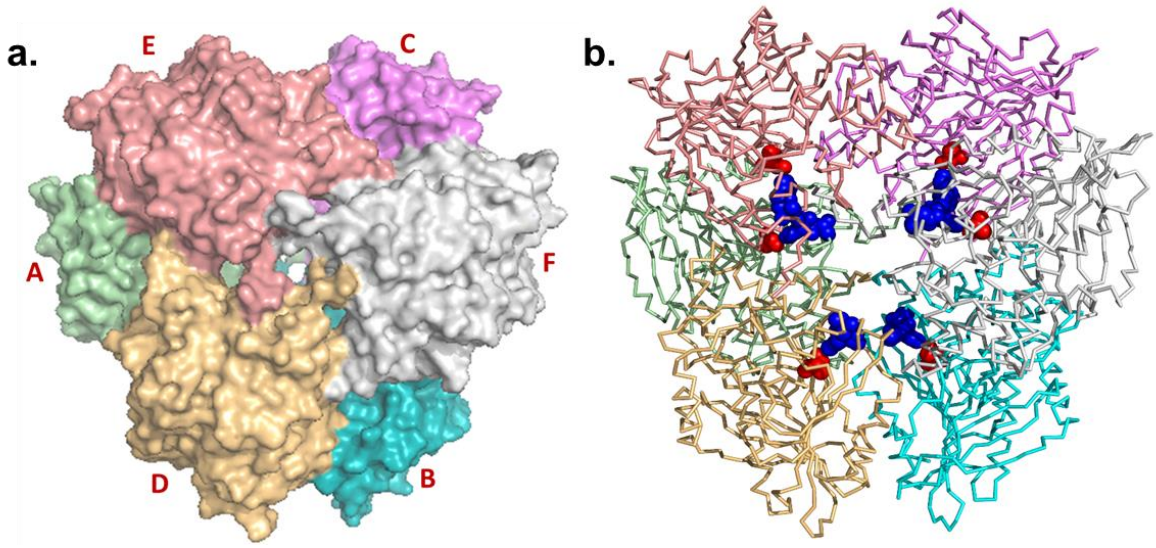


Figure 4.3: NaM1 quaternary structure

a. Surface representation of the flat spherical hexamer (top-view) showing subunit labels (A-F). **b.** Ribbon representation with active site residues shown as spheres located around the three-fold rotational axes. Catalytic Asp275 and His304 are in blue, while Ser185 is in red.

4.2.4.1 Active site and oxyanion hole

The NaM1 hexamer accommodates six active sites facing its inner void (Figure 4.3b). The catalytic triad consists of Ser185, Asp275 and His304. The nucleophilic Ser185 is adjacent to the S2 pocket that accommodates the substrate acyl moiety within the extended substrate binding cavity (Figure 4.4a) (Levisson et al., 2012, Singh and Manoj, 2016a). His304 forms a hydrogen bond to Ser185 and a salt bridge to Asp275 to stabilize the catalytic triad and activate Ser185 (Figure 4.4b). The residues, Phe 210, Pro 226, Ile 277 and Cys278, surround and embed the active site within the hydrophobic core of the domain – a typical feature of CE7 esterases (Montoro-García et al., 2011). Pro226 is part of a three-helix extension of CE7 enzymes that caps the substrate binding pocket at the active site end.

The oxyanion hole is formed by the backbone amide groups of conserved residues, Tyr93 and Gln186, located in the β6-α5 turn and the β4-α3 loop, respectively.

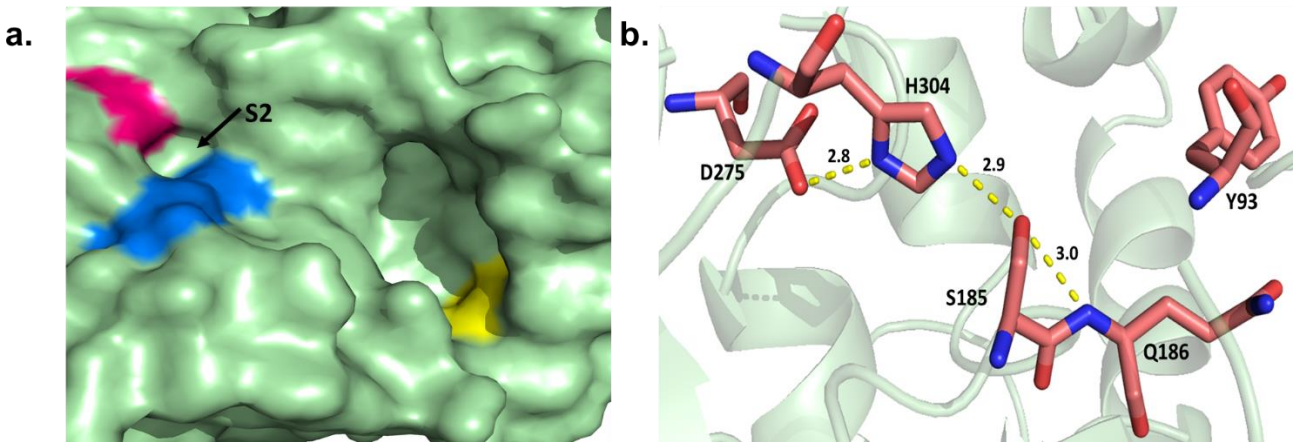


Figure 4.4a: NaM1 substrate-binding cavity

Pro226 (pink), located directly opposite the catalytic Ser185, marks the distal end and caps the active centre (blue). Tyr106 (yellow) marks the proximal end. **b.** NaM1 catalytic and oxyanion hole residues showing directed interactions as yellow dashed lines. All interactions are measured in Å

4.2.4.2 Inter-subunit interactions

Each NaM1 monomer interacts with four other monomers within the hexamer - two monomers from the same trimer and two from the opposite trimer. Thus, Monomer A interacts with subunits B and C, as well as D and E (Appendix 2F) resulting in a total buried surface area of 13650 Å due to interactions involving subunit A. The A-B subunit interface is repeated for B-C, C-A, D-E, E-F and F-D, and includes six salt bridges, and six hydrogen bonds and around 50 van der Waals interactions (Table 4.2 and Figure 4.5a). The average area per A-B – type monomeric interaction amounts to 547 $Å^2$. The A-D monomer-monomer interface (equivalent to B-F and C-E), connecting the two trimers involves two salt bridges, 12 hydrogen bonds and around 140 van der Waals interactions. These interactions form the largest monomer-monomer interfaces each with an average area of 837 Å. For *Thermotoga maritima* CE7 AcE mutant (TmAcEΔ26), the A-D interfaces (plus symmetry related) were found to be non-essential for oligomer formation (Singh and Manoj, 2016b). The A-E interface (equivalent to B-D and C-F), another monomer-monomer contact between trimers, also involves 12 hydrogen bonds and >130 van-der-Waals interactions, but no salt bridges. Each A-E – type interface has an average area of 753 Å.

Table 4.2: NaM1 inter-monomer interactions involving subunit A (PDBsum analysis)

Subunit	Residues from respective subunits	Respective atom names	Interaction type	Bond length (Å)*
A – B	Arg 18 – His 0	NH2 – ND1	H-bond	2.93
	Trp 218 – Asp 301	NE1 – OD1	H-bond	2.83
	Arg 235 – Glu 305	NH2 – OE1	Salt bridge Salt	2.80
		NH1 – OE2	bridge	3.00
	Arg 235 – Asp 301	NH1- O	Salt bridge	2.77
	Arg 240 – Leu 298	NE – O	H-bond	2.96
A – D	Gly 120 – Met 136	O – N	H-bond	2.79
	Gln 121 – Arg 237	O – NE	H-bond	3.09
		O – NH2	H-bond	2.84
	Val 126 – Trp 134	O – N	H-bond	2.87
	Glu 128 – Arg 142	OE1 –NH2	Salt bridge	2.98
	Glu 128 – Glu 128	OE2 - OE2	H-bond	3.07
	Trp 134 – Val 126	N – O	H-bond	2.89
	Met 136 – Gly 120	N – O	H-bond	2.90
	Arg 142 – Glu 128	NH2 – OE1	Salt bridge	2.94
	Arg 237 – Gln 121	NH2 – O	H-bond	2.87
A – C	Leu 298 – Arg 240	O – NE	H-bond	3.02
	Asp 301 – Trp 218	OD1 – NE1	H-bond	2.98
	Asp 301 – Arg 235	O –NH1	Salt bridge	2.97
	Glu 305 – Arg 235	OE2 – NH1	Salt bridge Salt	3.09
		OE1 – NH2	bridge	2.94
A – E	Asp 99 – Asp 312	OD2 – N	H-bond	2.95
	Trp 100 – Asp 312	N – OD2	H-bond	2.88
	Ser 101 – Asp 312	OG – OD1	H-bond	2.74
	Ser 101 – Asp 312	N – OD1	H-bond	2.77
	Asp 312 – Asp 99	N – OD2	H-bond	2.82
	Asp 312 – Ser 101	OD1 –N	H-bond	2.82
		OD1 – OG	H-bond	2.69
	Asp 312 – Trp 100	OD2 – N	H-bond	2.88

***Only interactions with inter-atomic distances between 2.6-3.1Å are listed.**

4.2.4.3 Non-physiological interactions

MES molecules from the crystallisation buffer were identified near tryptophan residues 61, 100 and 134

on the outer surface of each NaM1 subunit. Two MES molecules contribute to trimer-trimer interactions, A-D, B-F and C-E. In each case, the MES O3S and O2S atoms form hydrogen bonds with Trp134 N^{ε} atoms of both subunits (Figure 4.5b). MES molecules were also located between adjacent hexamers possibly supporting the crystal packing.

Additional electron density at the N-terminus of each monomer, except chain C, was modelled as one to three histidine residues of the His_6-tag. Due to the location of the N-terminus near the threefold axis at the trimer centre, the His_6-tags reduce the size of the central pore. The fact that these histidine residues are visible in the electron density implies that they are stabilized by interactions with the NaM1 protein. Only a single hydrogen bond is observed between His0 and Arg18 at subunit interfaces B-A, D-F, E-D and F-E respectively. Other interactions merely involve van der Waals interactions, indicating that the His_6-tag does not appear to contribute significantly to the stability of the NaM1 monomer or the hexamer.

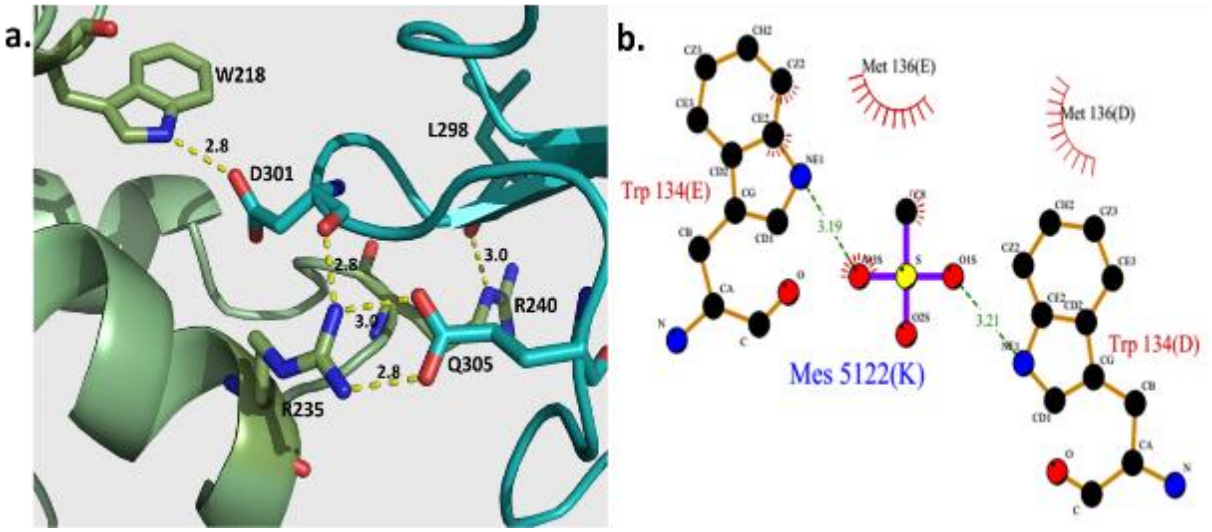


Figure 4.5: Inter-subunit interactions in NaM1

a. Interactions between subunits A (pale green) and B (cyan). Active residues are represented as sticks, showing salt bridges (Arg235A and Glu 305B) and H-bonds (Arg 235A – Asp 301B, Trp 218A – Asp 301B, Arg 240A – Leu 298B). **b.** MES trimer-trimer interactions between subunits D and E. Non-bonded

contacts are represented by red brushes directed towards interacting atoms. Figure 4.5b was generated using Ligplot (Wallace et al., 1995). The distance between atoms are measured in Å.

4.2.4.4 Comparison with other CE7 esterase structures

Using the Dali server (Holm, 2010) the closest structural homologs to NaM1 were identified as AcXE from *Thermoanareobacterium saccharolyticum* (PDB code 3FCY, RMSD = 1.2 Å), cephalosporin C deacetylase (CAH) from *B. subtilis* 168 (1ODS, 1.7 Å), AcXE from *Bacillus pumilus* (3FVT, 1.8 Å and 2XLB, 1.8 Å) and AcE from *Thermotoga maritima* (3M81, 2.1 Å). Table 4.3 shows details on the structural alignments. All of these proteins form hexamers but required unique conditions for crystallization. The pH varied from 6.0 to 7.5 and the resolution of the refined structures ranged from 1.8 to 2.5 Å. The parameters for NaM1 crystallization mirrors those of the related structures.

Table 4.3: Structural homologs of NaM1

Organism	**Protein**	**PDB ID**	**Number of residues**	**Identity to NaM1 (%)**	**Z-score**	**RMSD (Å)**
T. saccharolyticum JW/SL-YS485	AcXE	3FCY	317	50	49.6	1.2
B. subtilis 168	CAH	1ODS	316	33	42.8	1.7
B. pumilus PS213	AcXE	3FVT	317	33	42.3	1.8
B. pumilus CECT5072	AcXE	2XLB	317	32	42.3	1.8
T. maritima TM007	AcE/CAH	3M81	322	35	40.6	2.1

Superimposing NaM1 and all CE7 homologs indicates strong structural conservation within this family of enzymes (Appendix 2G). One difference in topology between NaM1 and other CE7 enzymes involves residues 174-178 of NaM1 which form an α4-β6 loop while forming the corresponding β-strand in all other AcXE structures. Multiple sequence alignment (MSA) nevertheless indicates that three constituent residues in this region, Val174, Asp175 and Arg178, are conserved in this family of enzymes (Figure

4.6). The difference in secondary elements may be caused by the replacement of polar Glu or Gln, found in thermostable 2XLB, 1ODS, 3FCY and 3M81, with rigid and hydrophobic Pro176 (NaM1).

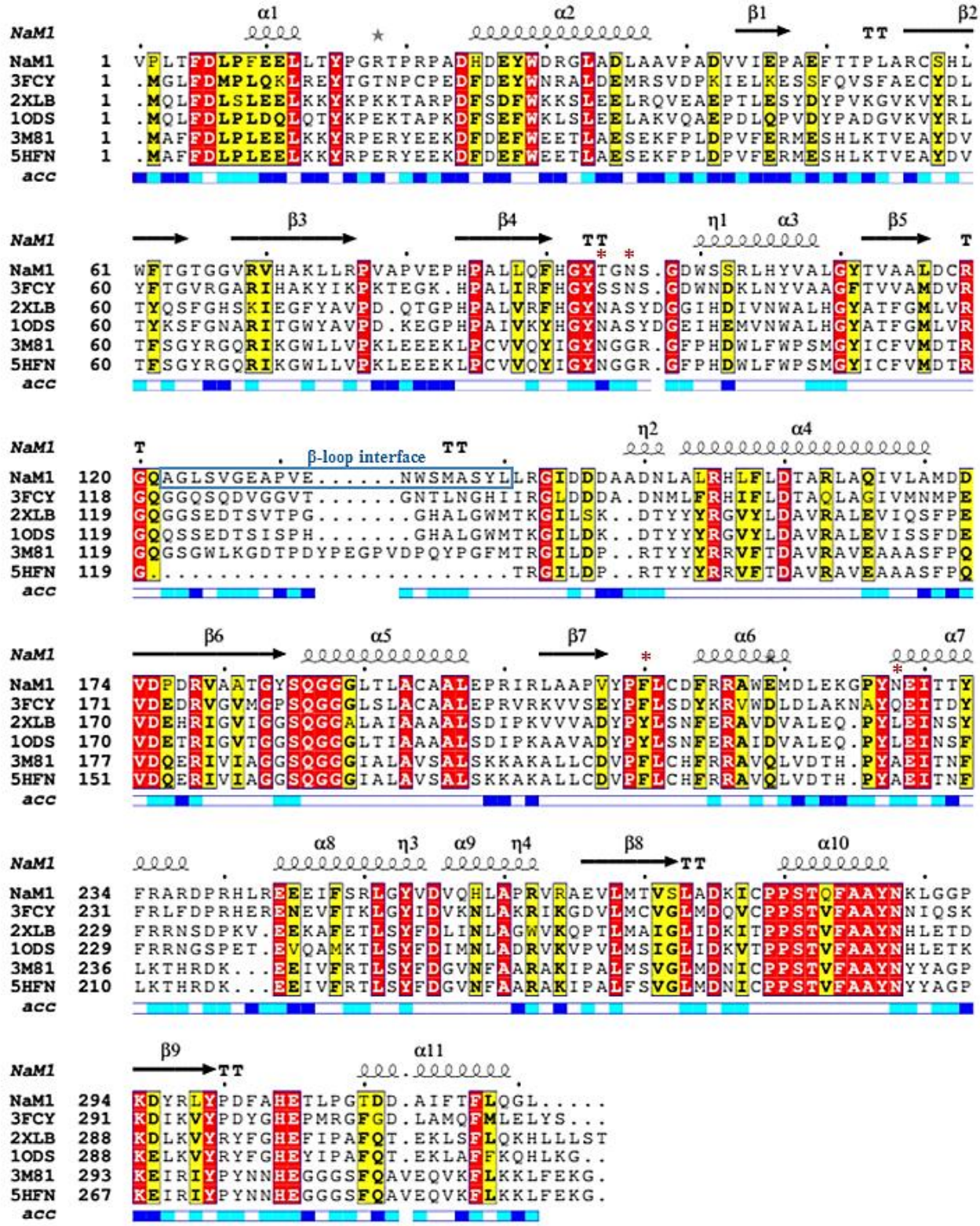


Figure 4.6: Structure-based sequence alignment of NaM1 and characterized CE7 family homologs.

This alignment was generated with Multalin (Corpet, 1988) and EsPript 3.0 (Robert and Gouet, 2014). Black dots and red asterisks above above sequences denote multiples of 10 and the locations of thermostabilizing substitutions in NaM1, respectively. Strictly and moderately conserved residues are coloured 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 accessibility band at the base of sequences indicates residues as accessible (blue), intermediate (cyan) or buried (white).

4.2.4.5 Substrate binding site

Unexpectedly, the substrate-binding cavity of NaM1, of around 420 $Å^2$, is structurally most similar to those of 3FVT and 2XLB (Figure 4.7) rather than to its closest homolog, 3FCY. This may indicate a larger overlap in substrate specificity than with other CE7 enzymes.

In NaM1, active site residues interact indirectly with residues from adjacent monomers implying the quaternary structure of NaM1 to be critical for substrate binding and catalysis. His304 of subunit A (His304A) is hydrogen-bonded to Glu305A, which in turn forms a salt bridge with Arg235D of α-helix α7. The corresponding Arg235A interacts with Glu305B (Table 4.2 and Figure 4.5a). Helix α7 is part of the triple helix insertion of NaM1 confirming its importance in oligomerization and catalysis (Montoro-García et al., 2011, Levisson et al., 2012). Such insertions or ‘caps’ are common in α/β hydrolases and appear to restrict the size of the substrate (Nardini and Dijkstra, 1999). In NaM1, Pro226 is the insertion residue nearest the active site. It is 7.3 Å from His304 and delimits the substrate-binding site on one side, with Tyr106 marking the other end (Figure 4.4a). In most CE7 enzymes, a conserved tryptophan residue, replaced by Tyr104 in 3FCY as in NaM1, delimits the substrate binding cavity at one end. 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 binding cleft is provided by the N-terminus of α5 and β-strands, β6 and β5.

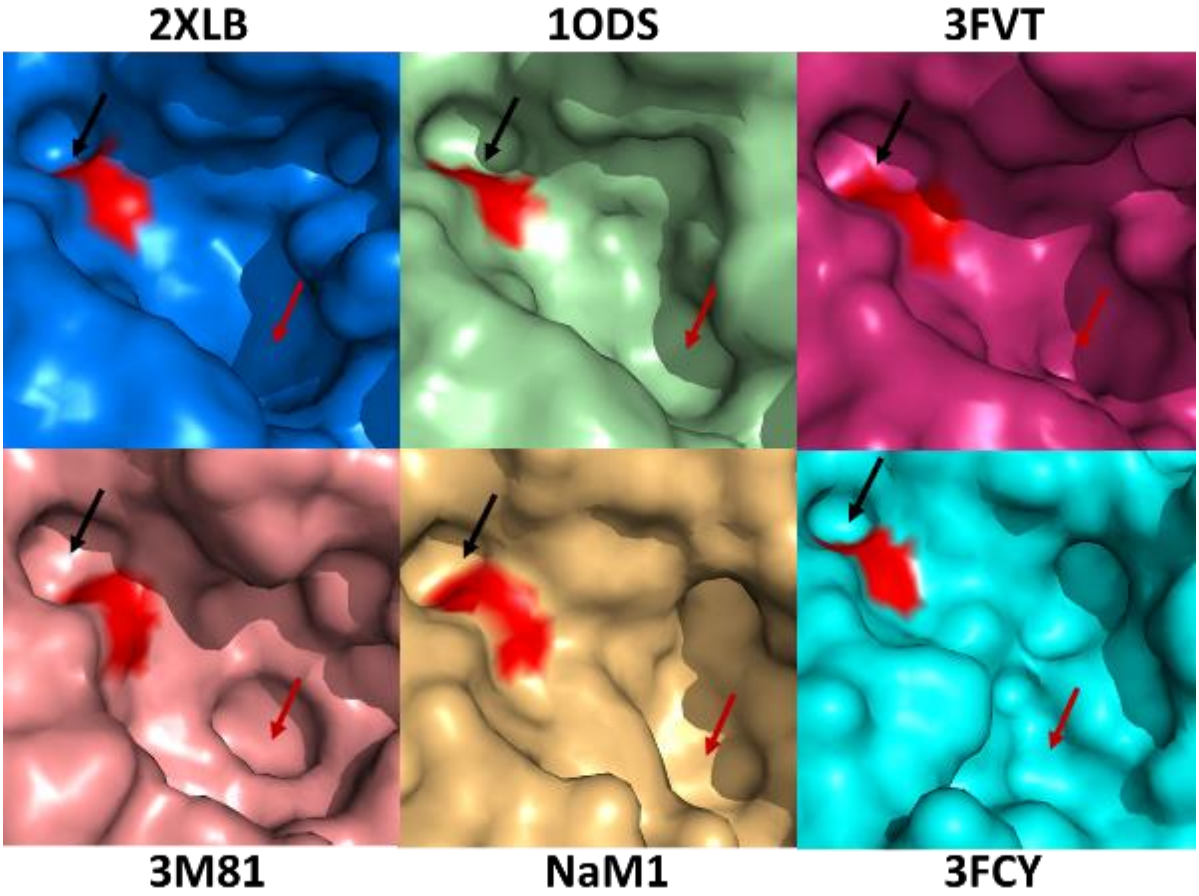


Figure 4.7: Surface representations of substrate binding cavities of CE7 esterases.

Catalytic serine is shown in red and binding pockets S1 and S2 (Levisson et al., 2012) are marked by red and black arrows, respectively.

4.2.5 Structural basis for substrate specificity

The location of acetate, the product of CE7 AcXEs, has been observed and described previously (Vincent et al., 2003, Montoro-García et al., 2011, Levisson et al., 2012). Many CE7 AcXEs accommodate substrates with up to C4 acyl groups (Shao and Wiegel, 1995, Levisson et al., 2012). AcXE of *B. pumilus* does not metabolise *p*-NP butyrate and this was attributed to Tyr206 restricting access to the active centre of groups larger than acetate (Montoro-García et al., 2011). In 2XLB, the Tyr206 interrupts the hydrophobic environment around the catalytic core; its hydroxyl group forms van der Waals interactions with active site-capping residue, Pro221 (see above), stabilizing the cap domain

and enhancing substrate selectivity. It also hydrogen-bonds Met138 (S^{δ}) and Tyr222 (N) of the hydrophobic core which are typically located near the active site of CE7 esterases. In NaM1, the S2 binding site-restricting Tyrosine is replaced by non-polar Phe210, creating a slightly larger binding site which can accommodate larger structures such as a butyryl group (Figure 4.8). This explains its high K_M and low catalytic efficiency for *p*-NP butyrate and lack of activity for *p*-NP octanoate and palmitate (Table 3.1; section 3.2.2.6). Other CE7 AcXEs with substrate delimiting phenylalanine rather than tyrosine were also active on C4 acyl groups (Levisson et al., 2012, Shao and Wiegel, 1995) implying that phenylalanine accommodates slightly longer acyl group chains (Montoro-García et al., 2011).

In 3M81, hydrophobic Phe210, Pro228 and Ile276 (corresponding to Phe210, Pro226 and Ile277 in NaM1) limit acetate specificity of this *T. maritima* enzyme (Hedge et al., 2012).

Replacing cis Pro228 in 3M81 with alanine switched the 227-228 peptide bond to trans and reduced catalytic activity to less than 10%, but without accommodating larger acyl moieties (Singh and Manoj, 2017). Hence, it was not clear if Pro228, of all the residues that make up the S2 binding site, influenced acyl component size restrictions the most significantly or not. Phe210 of NaM1 would seem to be more significant in this respect (Montoro-García et al., 2011).

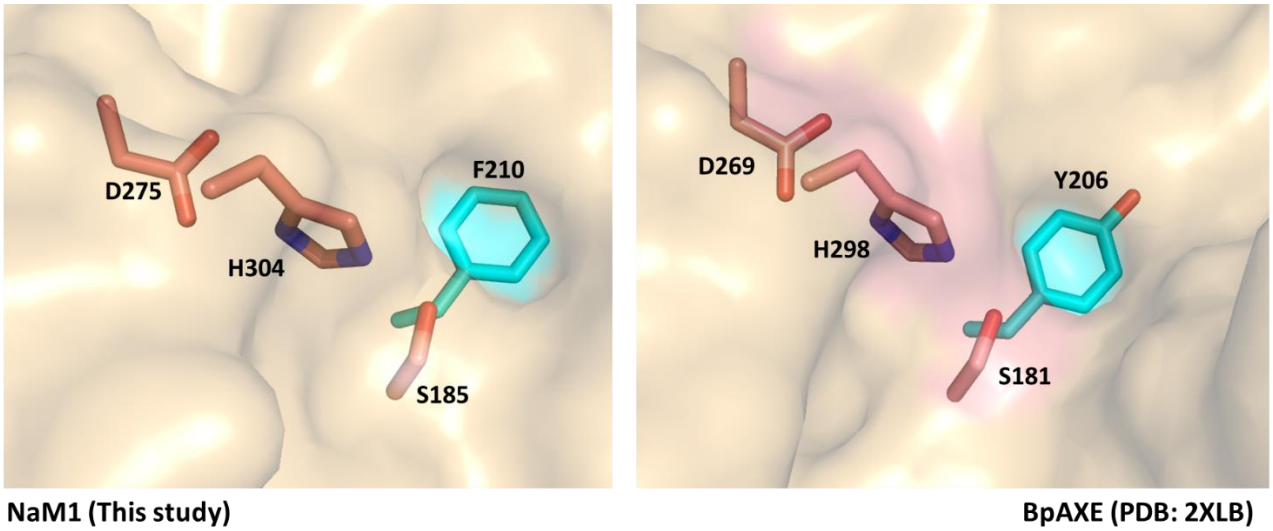


Figure 4.8: Residues proposed to control acyl length specificity.

a. In Nam1, Phe210 accommodates substrates with C4 acyl groups. **b.** In *Bp*AXE, Tyr206 excludes substrates with ≥C4 acyl groups.

NaM1 deacetylates acetylated xylan with low activity as do the *B. pumilus* PS213 (Degrassi et al., 2000) and CECT5072 (Martinez-Martinez et al., 2007) AcXEs. The similar substrate-binding cavity of these three enzymes (Figure 4.7) appears to accommodate this substrate while those of other CE7 enzymes do not. Substrate binding by CE7 esterases is inferred from bound inhibitors such as paraoxon and phenylmethylsulfonyl fluoride (PMSF) (Levisson et al., 2012, Montoro-García et al., 2011). To describe CE7 esterase specificities more fully, additional crystal structures with bound substrate, similar to that of non-cognate substrate 2-(2-oxo-1,3-dihydroindol-3-yl) acetate TmAcE (Singh and Manoj, 2016a), will be required.

4.2.6 Structural basis for thermostability

At 30°C, NaM1 has the lowest 'temperature optimum' of all CE7 esterases analysed to date. 'Thermal optima' normally lie between 45 and 90°C. Yet, there is high structural conservation among enzymes within this family (Figure 4.9a, Appendix 2G). In TmAcE, removal of the extended β-interface loop (residues 120 - 145) impacted thermal and conformational stability as well as enzyme activity (Singh and Manoj, 2016b). A sequence alignment of NaM1 with CE7 homologs (Figure 4.6) indicates that NaM1 lacks six residues of this loop (residues 121-140) (Figure 4.2a). Also within this loop, some conserved residues in thermostable enzymes are replaced in NaM1: For example, two glycines of thermostable homologs are replaced by alanine (Ala122 adjacent to a conserved RGQ motif) and serine (Ser138) in NaM1 (Figure 4.6). Other conserved residues replaced in NaM1 include His258, Arg264 and Gln283 which correspond with asparagine, lysine and valine, respectively, in more thermostable CE7 homologs. The substitution of hydrophobic valine, in a strictly conserved PPST**V**FAAYN motif, with polar glutamine in NaM1 appears to be the most dramatic change (Figure 4.9b). Gln283 in α-helix, α10, interacts with the active-site residue Asp275 through a water-mediated H-bond (Figure 4.9b). This polar interaction, impossible with valine, potentially disrupts the hydrophobic core and the hydrophobic active site environment in NaM1.

A decrease in inter-domain hydrophobic interactions was reported to be responsible for the

thermolability of a hormone-sensitive lipase family of esterases (Li et al., 2015) and hydrophobic interactions in the core of proteins are known to contribute to thermal stability. The replacement of uncharged polar residues with non-polar residues, capable of partaking in hydrophobic interactions, has been observed in the evolution of mesophilic to thermophilic proteins (Fields, 2001, Pucci and Rooman, 2017). A polar glutamine at this position may thus, contribute to NaM1 being more thermolabile. Interestingly, Gln283 of NaM1 is conserved as glutamine in several uncharacterized CE7 AcXE homologs from mesophilic microorganisms (Appendix 2E). Glutamine, as a larger residue than valine, leads to less compact packing near the active site potentially affecting thermal stability. In comparison to TmAcE, NaM1 has fewer residues, but a 3% more accessible surface area than TmAcE (Levisson et al., 2012), as calculated by PISA. This suggests more compact packing in TmAcE which may have enhanced its stability. As illustrated by (Singh and Manoj, 2016b), the structures of moderately thermostable CE7 enzymes were characterized with larger accessible surface areas than highly thermostable ones. AcXE from *T. saccharolyticum*, however, had a lower accessible surface area, but was less thermostable than it's highly thermostable counterpart, TmAcE.

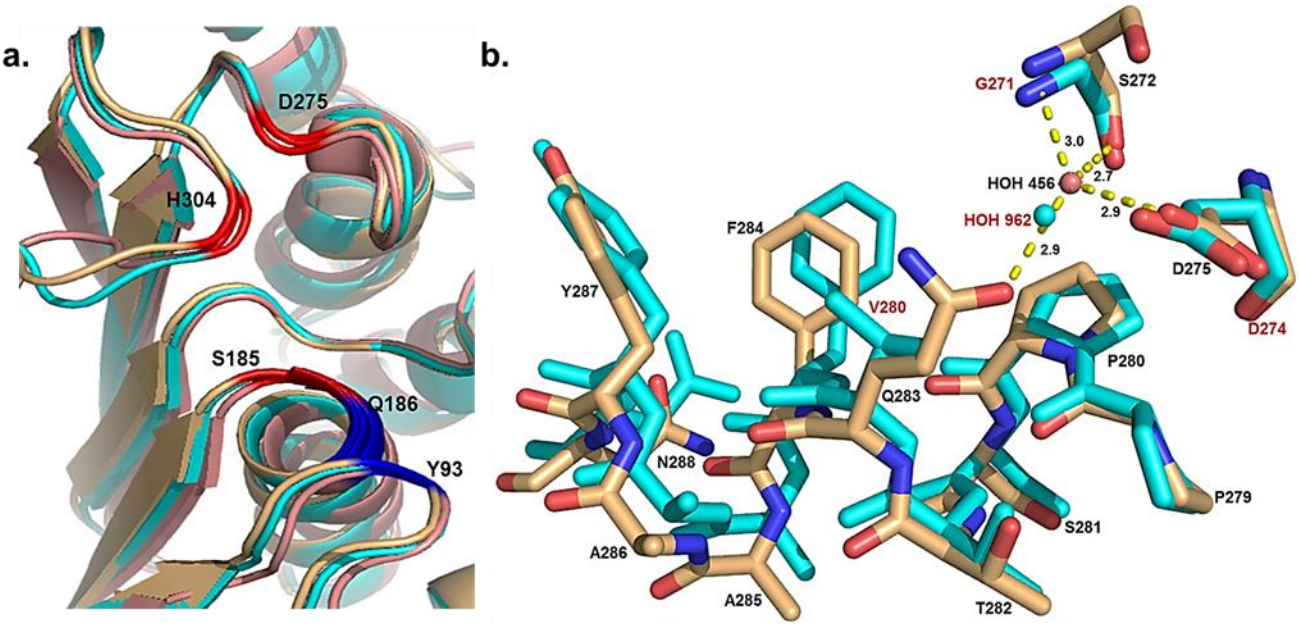


Figure 4.9: Structural conservation in NaM1

a. Superposition of NaM1 (tan) with thermostable CE7 enzymes from *Thermoanaerobacterium* sp.

(RMSD 1.2, PDB: 3FCY, cyan) and *Thermotoga maritima* (RMSD 2.1, PDB: 3M81, salmon) showing high structural conservation. Active site and oxyanion hole residues are colored red and blue, respectively. **b.** A hydrophobic valine in the strictly conserved PPST**V**FAAYN motif of *TmAcE* (cyan) is replaced by a polar Gln283 in NaM1 (tan) resulting in a water-mediated hydrogen bond to catalytic Asp275. Only residues labelled in red belong to TmAcE. All interactions shown are between NaM1 residues and are measured in Å.

4.2.7 Database accession number

The NaM1 structure has been deposited in the Protein Data Bank under the PDB ID, 6FKX.

4.3 CONCLUSION

In conclusion, NaM1 is a roughly spherical homo-hexamer related in structure to other AcXEs. It is the least thermostable member of analyzed CE7 esterases presumably due to several structural elements. Previously, more thermostable CE7 esterases have been investigated, with *Tm*AcE being the most stable. The structure of a thermolabile enzyme may help to identify structural features that reduce its thermostability - the reverse of which could be used to increase thermostability of other members. Residues in NaM1 that could be substituted to investigate this link include Ala122, Ser138, His258, Arg264 and Gln283. Correlating thermostability with substrate specificity may further prove useful as only thermolabile CE7 esterases have been reported to metabolize acetylated xylan. This poses a problem for industry, which needs highly thermostable enzymes. It is important to note the structure of NaM1 is the first for metagenome-derived CEs, indicating that this type of data is now becoming increasingly accessible (Júnior et al., 2017).

4.4 REFERENCES

ADAMS, P. D., AFONINE, P. V., BUNKÓCZI, G., CHEN, V. B., DAVIS, I. W., ECHOLS, N., HEADD, J. J., HUNG, L.-W., KAPRAL, G. J. & GROSSE-KUNSTLEVE, R. W. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography,* 66**,** 213-221.

AFONINE, P. V., GROSSE-KUNSTLEVE, R. W., ECHOLS, N., HEADD, J. J., MORIARTY, N. W., MUSTYAKIMOV, M., TERWILLIGER, T. C., URZHUMTSEV, A., ZWART, P. H. & ADAMS, P. D. 2012. Towards automated crystallographic structure refinement with phenix. refine. *Acta Crystallographica Section D: Biological Crystallography,* 68**,** 352-367.

BIELY, P., WESTERENG, B., PUCHART, V., DE MAAYER, P. & A. COWAN, D. 2014. Recent Progress in Understanding the Mode of Action of Acetylxylan Esterases. *Journal of Applied Glycoscience,* 61**,** 35-44.

CAZY 2017. Carbohydrate Active enZyme Database (CAZy).

CHEN, V. B., ARENDALL, W. B., HEADD, J. J., KEEDY, D. A., IMMORMINO, R. M., KAPRAL, G. J., MURRAY, L. W., RICHARDSON, J. S. & RICHARDSON, D. C. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography,* 66**,** 12-21.

CORPET, F. 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research,* 16**,** 10881-10890.

DE BEER, T. A., BERKA, K., THORNTON, J. M. & LASKOWSKI, R. A. 2014. PDBsum additions. *Nucleic Acids Research,* 42**,** D292-D296.

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.

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.

DELAGENIÈRE, S., BRENCHEREAU, P., LAUNER, L., ASHTON, A. W., LEAL, R., VEYRIER, S., GABADINHO, J., GORDON, E. J., JONES, S. D. & LEVIK, K. E. 2011. ISPyB: an information management system for synchrotron macromolecular crystallography. *Bioinformatics,* 27**,** 3186-3192.

DELANO, W. L. 2002. The PyMOL molecular graphics system. http://pymol.org.

EMSLEY, P. & COWTAN, K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography,* 60**,** 2126-2132.

FIELDS, P. A. 2001. Review: Protein function at thermal extremes: balancing stability and flexibility. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology,* 129**,** 417-431.

HEDGE, M. K., GEHRING, A. M., ADKINS, C. T., WESTON, L. A., LAVIS, L. D. & JOHNSON, R. J. 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.

HOLM, L. 2010. Dali server: conservation mapping in 3D. *Nucleic Acids Research,* 38**,** W545-W549.

JÚNIOR, A. D. N. F., DAMÁSIO, A. R. L., PAIXÃO, D. A. A., ALVAREZ, T. M. & SQUINA, F. M. 2017. Applied Metagenomics for Biofuel Development and Environmental Sustainability. *Advances of Basic Science for Second Generation Bioethanol from Sugarcane.* London: Springer International Publishing.107-129

KLEYWEGT, G. J., HENRICK, K., DODSON, E. J. & VAN AALTEN, D. M. 2003. Pound-wise but penny-foolish: How well do micromolecules fare in macromolecular refinement? *Structure,* 11**,** 1051-1059.

KRISSINEL, E. & HENRICK, K. 2007. Protein interfaces, surfaces and assemblies service PISA at European Bioinformatics Institute. *J Mol Biol,* 372**,** 774-797.

LASKOWSKI, R. A., MACARTHUR, M. W., MOSS, D. S. & THORNTON, J. M. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography,* 26**,** 283-291.

LEVISSON, M., HAN, G. W., DELLER, M. C., XU, Q., BIELY, P., HENDRIKS, S., TEN EYCK, L. F., FLENSBURG, C., ROVERSI, P., MILLER, M. D., MCMULLAN, D., VON DELFT, F., KREUSCH, A., DEACON, A. M., VAN DER OOST, J., LESLEY, S. A., ELSLIGER, M. A., KENGEN, S. W. & WILSON, I. A. 2012. Functional and structural characterization of a thermostable acetyl esterase from *Thermotoga maritima*. *Proteins,* 80**,** 1545-59.

LI, P.-Y., CHEN, X.-L., JI, P., LI, C.-Y., WANG, P., ZHANG, Y., XIE, B.-B., QIN, Q.-L., SU, H.-N. & ZHOU, B.-C. 2015. Interdomain hydrophobic interactions modulate the thermostability of microbial esterases from the hormone-sensitive lipase family. *Journal of Biological Chemistry,* 290**,** 11188-11198.

MARTINEZ-MARTINEZ, I., MONTORO-GARCIA, S., LOZADA-RAMIREZ, J. D., 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.

MCCOY, A. J. 2007. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallographica Section D: Biological Crystallography,* 63**,** 32-41.

MONTORO-GARCÍA, S., GIL-ORTIZ, F., GARCÍA-CARMONA, F., POLO, L. M., 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.

NARDINI, M. & DIJKSTRA, B. W. 1999. α/β Hydrolase fold enzymes: the family keeps growing. *Current Opinion in Structural Biology,* 9**,** 732-737.

PUCCI, F. & ROOMAN, M. 2017. Physical and molecular bases of protein thermal stability and cold adaptation. *Current Opinion in Structural Biology,* 42**,** 117-128.

RAMACHANDRAN, G. N., RAMAKRISHNAN, C. & SASISEKHARAN, V. 1963. Stereochemistry of polypeptide chain configurations. *Journal of molecular biology,* 7**,** 95-99.

ROBERT, X. & GOUET, P. 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research,* 42**,** W320-W324.

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.

SINGH, M. K. & MANOJ, N. 2016a. 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.

SINGH, M. K. & MANOJ, N. 2016b. 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.

SINGH, M. K. & 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.

TERWILLIGER, T. C., GROSSE-KUNSTLEVE, R. W., AFONINE, P. V., MORIARTY, N. W., ZWART, P. H., HUNG, L.-W., READ, R. J. & ADAMS, P. D. 2008. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallographica Section D: Biological Crystallography,* 64**,** 61-69.

VINCENT, F., CHARNOCK, S. J., VERSCHUEREN, K. H., TURKENBURG, J. P., SCOTT, D. J., OFFEN, W. A., ROBERTS, S., PELL, G., GILBERT, H. J. & DAVIES, G. J. 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.

WALLACE, A. C., LASKOWSKI, R. A. & THORNTON, J. M. 1995. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Engineering,* 8**,** 127-134.

CHAPTER FIVE

PROTEIN ENGINEERING: DIRECTED EVOLUTION AND SITE DIRECTED MUTAGENESIS OF NaM1

5.0 INTRODUCTION

Protein engineering is an effective tool for overcoming the limitations as well as improving and/or understanding the function of known proteins (Kazlauskas and Bornscheuer, 2009, Davids et al., 2013). Insights into the evolution of certain protein traits can be gained via protein engineering (Arnold et al., 2001). Also, reconstructing ancestral protein sequences may help understanding the thermal stability (Wijma et al., 2013) and molecular evolution (Gumulya and Gillam, 2017) of proteins. Protein sequence modification is typically achieved rationally by site-directed mutagenesis (SDM) and by random mutagenesis as in directed evolution (DE). A combination of both approaches is termed semi-rational design (Chica et al., 2005, Lutz, 2010). Directed mutagenesis, targets specific nucleotide bases and generally depends on prior knowledge of the protein structure. Random mutagenesis instead generates molecular diversity after which mutants with desired traits will be selected. DE is a powerful tool for engineering enzymes (Porter et al., 2016, Kataoka et al., 2016). It allows specific enzyme properties such as thermal stability, catalytic turnover rate (k_{cat}) substrate specificity, organic solvent tolerance or pH stability to be improved or introduced without requiring structural information. Various methods of random mutagenesis include error-prone PCR (epPCR), *in situ* epPCR (Shao et al., 2017), alanine-scanning mutagenesis (Cunningham and Wells, 1989), DNA and codon shuffling, error-prone rolling circle amplification (epRCA), gene cassette mutagenesis, saturation mutagenesis and others extensively reviewed by Kaur and Sharma (2006) and Kumar and Singh (2013). EpPCR, using non-proof-reading polymerases with reduced fidelity, allows controlled induced mutation which mimics natural selection or evolution. However, other random mutagenesis methods create higher degrees of variation which complicate analysis (Arnold et al., 2001). Recent studies report the use of epPCR for successfully improving enzymes either after a single (Lin et al., 2016, Goomber et al., 2016) or multiple rounds (Hossain et al., 2016, Melzer et al., 2015, Jiang et al., 2015). Increased AcXE activity has also been achieved using epPCR and SDM (DiCosimo et al., 2013). The limitations of epPCR include its

failure to introduce single mutations at all possible positions, failure to allow mutations on consecutive nucleotides and its bias towards substitutions, particularly transitions, over transversions (Currin et al., 2015, Wintrode and Arnold, 2001, Kazlauskas and Bornscheuer, 2009).

Most DE experiments have two-way objectives, such as, achieving improved thermal stability and catalytic efficiency (Figure 5.1). Hence, an efficient screening protocol is required for the created library so as to maximise results and identify the best fit mutant. The screening should make use of well-defined fitness criteria and fluorogenic substrates for easy identification of improved mutants (Kazlauskas and Bornscheuer, 2009, Currin et al., 2015). A colorimetric method was recommended for AcXE and CPC deacetylase screening in DE studies (Martinez-Martinez et al., 2007), but, the complexity of deriving the substrate, AX, is a major setback in using the method. Instead, the use of general, preferably chromogenic, esterase substrates remains the standard method for detecting or quantifying AcXE activities in DE experiments (Adesioye et al., 2016). It is worthy of note that thermostable variants would be typically identified by their thermoactivity in DE screens as not all proteins that are stable at high temperatures are active at the same temperatures. Proteins are capable of losing activity as temperatures increase while maintaining their overall folded structure (Danson et al., 1996).

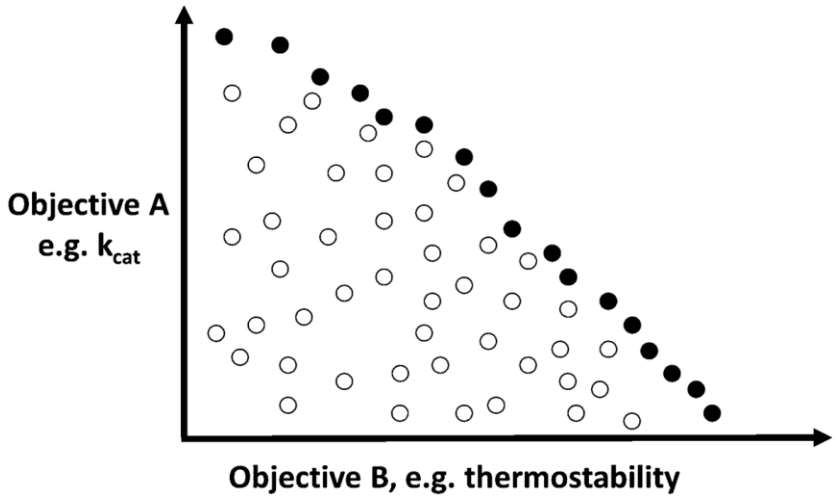


Figure 5.1: A Pareto-front illustration of a two-objective optimisation problem in protein engineering. Each individual symbol is a candidate solution e.g. a clone. Filled circles approximate the Pareto front indicating that there is normally no single solution that is optimal for all objectives (Handl et al., 2007). This illustration was adapted from Currin et al. (2015).

Considering the importance of protein thermal stability for enzymatic lignocellulose deconstruction, several thermolabile CAZymes (Moraïs et al., 2016, Suk-Am et al., 2002, Pei et al., 2015), including AcXEs (Koseki et al., 2005), have been engineered *in-vitro* to evolve more thermostable variants suitable for industrial processes. Efforts have been made to understand the determinants of thermal stability (Singh and Manoj, 2016b), substrate specificity (Singh and Manoj, 2016a, Hedge et al., 2012) and activity (Singh and Manoj, 2017) of CE7 AcXEs. In *T. maritima* TM007 AcE, the deletion of the entire β-loop interface (Singh and Manoj, 2016b) and N-terminal extension (Singh et al., 2017) reduced enzyme thermal stability. However, both elements are also present in less thermostable CE7 enzymes, making it difficult to identify the specific factors responsible for CE7 stability. Thus, DE remained the best approach to improve the thermal stability of NaM1 despite structural and functional data being available for its homologs. Minimizing activity loss over time at temperatures above the enzyme 'temperature optimum (T_{opt})' will fail to maintain enzyme activity at higher temperatures if only the Gibbs energy of activation for thermal inactivation ($\Delta G^{\ddagger}_{inact}$) is increased without changing the equilibrium temperature (T_{eq}) and/or the change in enthalpy (ΔH_{eq}) (Daniel et al., 2001). This is the usual outcome for SDM as opposed to DE which allows the opportunity to explore useful modifications in T_{eq} and/or ΔH_{eq} (Daniel and Danson, 2013). Hence, in this study, random mutagenesis was used to create a library of NaM1 variants. The most thermostable and thermoactive variant obtained was further investigated using SDM.

The aim of the work reported in this chapter was to identify single amino acid changes that affect thermal stability of NaM1. A random library of NaM1 mutants with one or two nucleotide changes was created and screened for thermostable enzyme variants. Kinetic and potential structural characteristics of selected variants were studied in comparison with the wild-type enzyme.

5.1 MATERIALS AND METHODS

5.1.1 Error-prone and site-directed mutagenesis PCR conditions

The epPCR protocol for this study was adapted from published recommendations of Copp et al. (2014). Six mutant libraries were generated and analysed for size and quality. One Taq polymerase (New England Biolabs, MA, USA) was induced to generate errors by adding 0.2-0.4 mM $MnCl_2$ and/or an imbalance in deoxynucleoside triphosphate (dNTP) concentrations in the PCR mix. The epPCR components and cycling conditions for each library are described in Tables 5.1 and 2.3, respectively. The pET vector system T7 forward and reverse primers were used for direct amplification from the pET28a vector.

The SDM PCR conditions are described in Table 2.3. The appropriate annealing temperature was determined by gradient PCR using high fidelity (HiFi) DNA polymerase (Kapa Biosystems, Cape Town South Africa) and following recommendations for primer and PCR conditions from the Stratagene QuikChange™ (Agilent Technologies, CA, USA) SDM protocol. The mutagenic primers were designed using PrimerX (http://www.bioinformatics.org/primerx/documentation.html).

Table 5.1: Six epPCR conditions tested for NaM1 random mutagenesis

EpPCR	**Concentration (mM)**				
Condition	**dATP**	**dGTP**	**dTTP**	**dCTP**	**$MnCl_2$**
0*	0.2	0.2	0.2	0.2	-
1	0.02	0.02	0.2	0.2	-
2	0.2	0.2	0.02	0.02	-
3	0.02	0.02	0.2	0.2	0.2
4	0.02	0.02	0.2	0.2	0.4
5	0.2	0.2	0.02	0.02	0.2
6	0.2	0.2	0.02	0.02	0.4

*Conventional PCR

5.1.2 Cloning, transformation and mutant library construction

Cloning and transformation of epPCR products is described in section 3.1.2. Here, however, the entire PCR product was first treated with DpnI (New England Biolabs, MA, USA) for 1 hr at 37°C to digest the template DNA after which restriction enzyme digestion, cloning and transformation proceeded in batches. A protocol by Sambrook and Russell (2001) was modified for mutant library construction. After transformation, 100 µL of the transformation mix was plated to determine the transformation efficiency and the remaining transformants were inoculated into 5 mL LBB-kan and incubated for 45 min at 37°C. This was then divided into fractions (library units) and stored as 400 µL 25% (v/v) glycerol stocks at -80°C.

SDM PCR products were also transformed as described in section 3.1.2. Five transformants were selected and grown for glycerol stock storage, plasmid miniprep and DNA sequencing.

5.1.3 EpPCR mutant library screening

Mutant library cultivation, expression and screening protocols were adapted from section 3.1.3 and Reetz and Carballeira (2007) (Figure 5.2).

5.1.3.1 Cultivation, pre-culture and expression

Glycerol stocks of the mutant library were plated on LBA-kan and incubated at 37°C overnight. Clones were inoculated into 150 µL LBB-kan in 96-well microplates and incubated overnight with shaking at 250 rpm. Subsequently, 180 µL fresh LBB-kan in duplicate plates was inoculated with 20 µL of this starter culture. This expression culture was supplemented with IPTG after 4 hrs of growth and gene expression was allowed to proceed for ~10 hrs. Wild type NaM1 and three uninoculated wells containing LBB-kan were included in each plate as positive and negative controls, respectively. All pre-culture and expression plates were sealed with Breathe-easier sealing membranes (Sigma Aldrich, MO, USA) during incubation to allow aeration. Following inoculation of the expression culture, the remaining pre-culture was sealed with the Microseal B adhesive film (BioRad Laboratories, UK) to prevent the culture from drying. Sealed plates were stored at 4°C for future reference.

5.1.3.2 Thermal stability screen

Cells were harvested by centrifugation at 4000 x g for 10 min at 4°C and supernatant was discarded. Cells were re-suspended in 20 µL Bacterial protein extraction reagent (Thermoscientific, IL, USA) and incubated for 15 min at room temperature after which the lysate was diluted with 25 mM Tris-HCl and 25 mM NaCl (1:1). The lysate was centrifuged again and the supernatant was used for thermal stability screen. The thermal stability screening temperature (55°C) was determined based on results obtained from NaM1 thermal stability and inactivation profiles in section 3.2.2.3. Duplicate plates containing 10 µL aliquots of the mutant enzymes were incubated at 25°C (control) and 55°C for 15 min and assayed for *p*-NPA activity. Hydrolysis of *p*-NPA by variants possessing residual activity after incubation at 55°C was observed by rapid amber to yellow colour changes, depicting catalytic release of *p*-NP as opposed to auto-hydrolysis (Figure 5.2c). Pre-culture plates were retrieved from storage at 4°C, and potentially thermostable mutants were inoculated into fresh LBB-kan for cultivation. Duplicate glycerol stocks of the selected mutants were stored at –80°C.

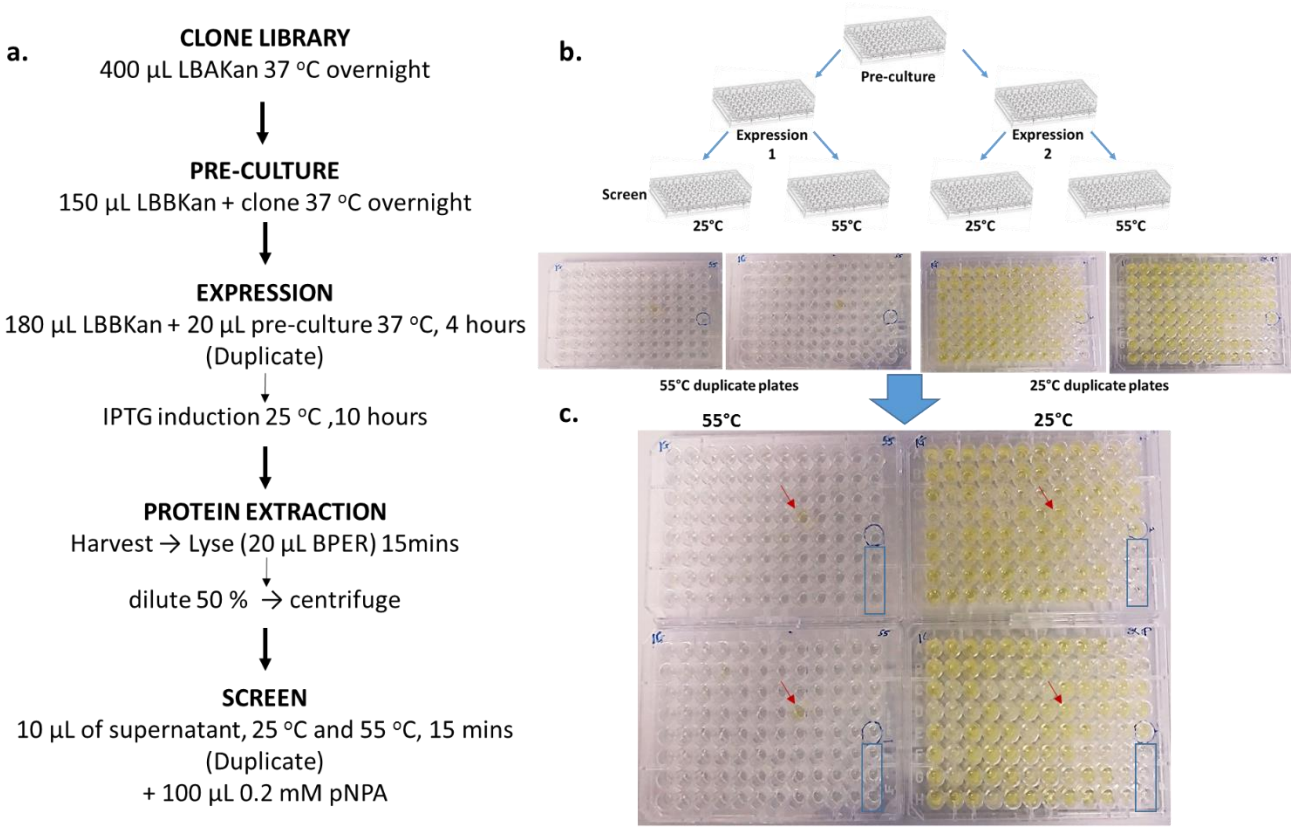


Figure 5.2: NaM1 mutant library screening protocol

a. Flowchart illustrating mutant library cultivation and screening protocols. **b.** Schematic representation of plate layout per 96-well clone library screening process. **c.** DE screen showing light amber color in microplate wells indicating release of *p*-NP by mutant enzymes after incubation at 55°C and 25°C indicated by plates labelled 55 and 25, respectively. Red arrows indicate mutant D8 visibly active at 55°C and 25°C. Blue circles indicate wild type strain active at 25°C and inactive at 55°C. Control uninoculated wells depicted by blue rectangles indicate that wells were not cross contaminated from neighboring active clones.

5.1.4 Confirmation of thermal stability of selected mutants

Cell free extracts of selected variant enzymes were prepared and used for confirmatory thermal stability and thermal inactivation assays as described in section 3.1.7.2. The most thermostable mutant (H2) and its SDM variants ($NaM1_{N96S}$ and $NaM1_{F210L}$) were selected for further studies.

5.1.5 Purification and functional characterization of selected NaM1 variants

Variants H2, $NaM1_{N96S}$ and $NaM1_{F210L}$, were produced on a larger scale and purified (section 3.1.4). The pH and temperature 'optima' of H2 were determined (Section 3.1.7.2).

5.1.6 Kinetics of NaM1 wild type and selected variants

The kinetics of the wild type NaM1 (WT), H2, $NaM1_{N96S}$ and $NaM1_{F210L}$ enzymes on *p*-NPA were determined, using 0.4µg protein, as described in section 3.1.7.5, but at 40°C to accommodate the improved thermal stability of variants. K_M and V_{max} were determined from a Michaelis-Menten non-linear regression curve fit generated by GraphPad Prism 5 software for Windows (GraphPad Software, San Diego California USA, www.graphpad.com). H2 activity on substrates with acyl moieties of more than two carbon atoms (*p*-NPB, *p*-NPO and *p*-NPP), AX and 7-ACA were determined as described in section 3.1.7.5. The release of acetic acid was monitored spectrophotometrically using the K-ACET kit (Megazyme, Ireland) and its associated Mega-Calc software according to the manufacturer's instructions.

5.1.7 Fluorescence and Circular Dichroism Spectroscopy

The absorbance of aromatic residues of NaM1 and its variants was recorded between 280-450 nm (Teale, 1960) using a Shimadzu spectrofluorophotometer RF-5301 PC to detect any changes in the protein fold. Two excitation wavelengths, 280 and 295 nm were used to detect emissions from the aromatic residues.

The thermal stability of all four proteins was further investigated by studying their temperature-induced unfolding patterns between 20 and 65°C at 5°C intervals using far-UV (240-180 nm) circular dichroism (CD) spectroscopy (Kelly and Price, 2000). Each protein at a concentration of 4 µM in a quartz cuvette of path length 2 mm was incubated for 5 min at each interval temperature and the ellipticity value (θ) was recorded from 195 to 250 nm using an Applied Photophysics Chirascan CD spectrophotometer at a 0.2 nm sampling interval. The far-UV CD spectrum was plotted using the Chirascan Pro-Data viewer

and Microsoft Office Excel. The thermal unfolding profile at 222 nm was extracted and fitted to a Boltzmann sigmoidal curve. The proteins were dissolved in 25 mM Tris HCl, 25 mM NaCl buffer (pH 7.0) for both spectroscopic studies. Blank buffer runs were performed to confirm non-interference of the buffer components with absorbance readings.

5.1.8 Sequence analyses and structural modelling of mutants

To identify molecular parameters affecting the thermostability of the NaM1 variants, multiple sequence alignments of CE7 sequences and related α/β- hydrolases were investigated alongside their generated structural models. The proteins were structurally modelled using SwissModel software. The crystal structure of *Thermoanaerobacterium* sp. strain JW/SL YS485 (PDB ID: 3FCY) was used as reference. Structural analysis and superposition on NaM1 was done using COOT (Emsley and Cowtan, 2004) and PyMOL (DeLano, 2002) programs. Structural depictions were generated in PyMOL.

5.2 RESULTS AND DISCUSSION

5.2.1 Error-prone PCR, cloning, transformation and library construction

EpPCR products were visible by agarose gel electrophoresis for three mutagenic conditions (1, 3 and 5). Experiments with high $MnCl_2$ and/or low pyrimidine (dCTP and dTTP) concentrations produced faint or not visible bands indicating poor PCR amplification. This may be due to the strong effect of $MnCl_2$ on the polymerase which resulted in errors during DNA amplification for mutagenic conditions 3, 4, 5 and 6. Further, the mutational bias of *Taq* polymerase, commonly overcome by increasing dCTP and dTTP concentrations (Wang et al., 2006), was enhanced by lowering the concentrations of these nucleotides in epPCR conditions 4, 5 and 6. Pyrimidines constitute 53% of *NaMet1* DNA. Variation in pyrimidine concentrations in the PCR mix will affect PCR amplification more severely than in purine concentrations. The effects of $MnCl_2$ and pyrimidine concentrations explain the poor amplification for epPCR conditions 2, 4 and 6. Condition 1 with lower purine (dATP and dGTP) concentrations and no $MnCl_2$ yielded the brightest band indicating the most successful epPCR amplification.

The transformation efficiency of the *E. coli* BL21 competent cells for the pET28a vector used in this

study was determined to be 9.1 x 10^4 per µg of vector DNA. This is within the acceptable range for this vector size and microbial host (Hanahan, 1983). The suitability of the competent cells was confirmed for transformation as high efficiency competent cells are recommended for effective uptake of mutant DNA (Copp et al., 2014). The epPCR product yield was evident in the transformation results as conditions with fainter agarose gel bands had fewer transformants than those with brighter bands. The same DNA concentration and ligation ratio were used for all transformations. This reflected the quality of the libraries and possibly, the negative effect(s) of Mn^{2+} on downstream restriction digest and ligation steps (Wang et al., 2006). Hence, two conditions which consistently showed DNA amplification and transformants were selected for test library construction. One condition each with (Condition/Library 5) and without (Condition/Library 1) $MnCl_2$ was selected so as to expand the molecular diversity of libraries available for screening.

5.2.2 Library analyses and screening

Following library creation, the mutational diversity was assessed by randomly sequencing 10 clones from each library. Sequencing results indicated several mutations per clone in Library 5 (Lib5) while Library 1 (Lib1) mostly consisted of clones with one or two mutations (Table 5.2). All mutations observed in Lib1 were A/T → G/C substitutions, a known mutational bias which results from an imbalance in dNTP concentrations during epPCR when low fidelity DNA polymerase is used (Kumar and Singh, 2013, Wang et al., 2006). Multiple mutations observed in Lib5 included deletions, insertions and substitutions which would create ambiguities during sequence and functional analysis. This will potentially require step-wise site-directed mutagenesis experiments to identify the cause for any observed effect. There was no significant difference between the percentages of active clones (~70%) per 96-well plate screen of both libraries using *p*-NPA. This indicated an acceptable mutation frequency (Kumar and Singh, 2013) and validated both libraries for screening. Nevertheless, considering the multiple mutations obtained per clone and the small size of Lib5 compared with Lib1 (ratio 1:5), only Lib1 was subsequently screened for thermostable mutants.

More than 3100 variant enzymes were screened for thermal stability. Only 11 clones, or 4% of tested

and 5% of active variants, were selected as putative thermostable mutants. These mutants were named according to the wells from which they were identified during screening and subjected to confirmatory thermal inactivation assays.

Table 5.2: Analysis of mutations from epPCR library one

Mutant	DNA mutation	Substitution type	Protein mutation	Mutation type	Distance from S185 (Å) by Cα
H2	A→G	Transition	N96S	Missense	12.51
	T→C	Transition	F210L	Missense	4.84
E1	A→C	Transversion	T306P	Missense	11.43
F9	A→C	Transversion	E11A	Missense	22.08
D8	A→G	Transition	T94A	Missense	10.92
	C→T	Transition	-	Silent	
B4	A→G	Transition	N228D	Missense	15.08
	T→A	Transversion	-	Silent	
H1	A→T	Transversion	T232S	Missense	20.41
	A→T	Transversion	-	Silent	
B10	A→T	Transversion	T232S	Missense	20.41
H10	A→G	Transition	-	Silent	
	A→C	Transversion	D147A	Missense	30.12
E5N	T→C	Transition	-	Silent	
	T→C	Transition	-	Silent	
4-3	A→C	Transversion	M220L	Missense	18.23
	A→G	Transition	Q319R	Missense	22.16
4-4	T→A	Transversion	L190Q	Missense	8.63
4-11	T→A	Transversion	V39E	Missense	28.76
	A→G	Transition	Y106C	Missense	20.82
	T→A	Transversion	I167N	Missense	23.77
4e	T→C	Transition	F214L	Missense	14.17

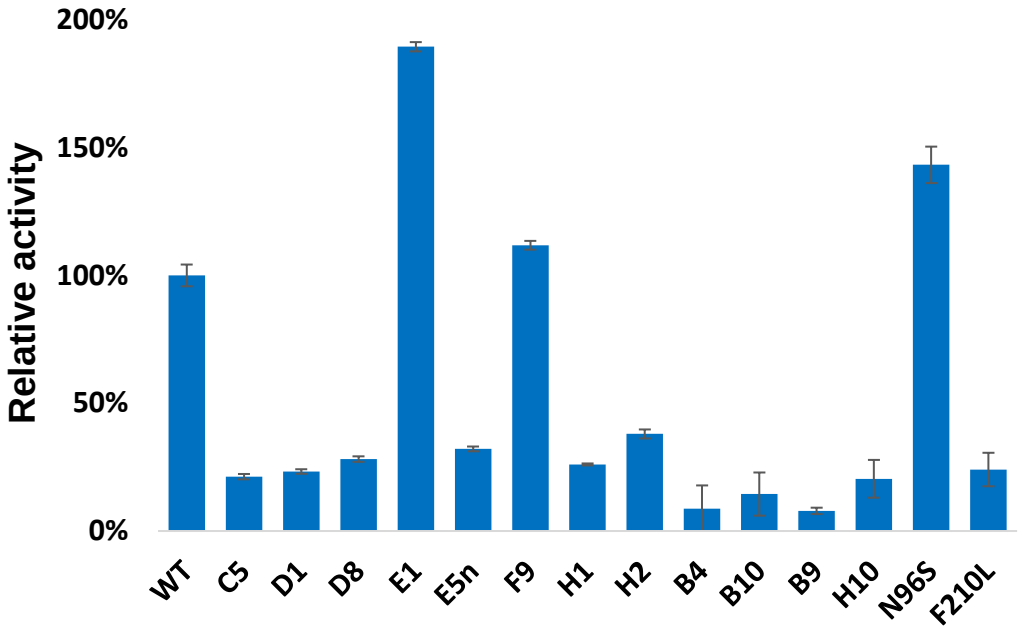


Figure 5.3: Deacetylase activity of selected NaM1 variants relative to the wild type

Deacetylase activity on *p*-NPA was determined at 30°C using cell free extracts of NaM1 variants.

5.2.3 Thermal stability and inactivation assays

5.2.3.1 Directed evolution variants

Thermal stability assays demonstrated that six variants were false positives. This could have been due to thermal insulation from cell debris erroneously transferred together with the enzyme supernatant after cell lysis (Ezemaduka et al., 2014). Three of these variants were completely inactive at 55°C, while the activity and stability of three others were similar to the WT. All six were excluded from subsequent analysis. Five NaM1 variants were further analyzed. Two variants, E1 and F9, retained the thermal stability traits of the WT, but had higher activity. The remaining three variants, B4, D8 and H2, were more thermostable, but less active than the WT (Figures 5.4 and 5.5).

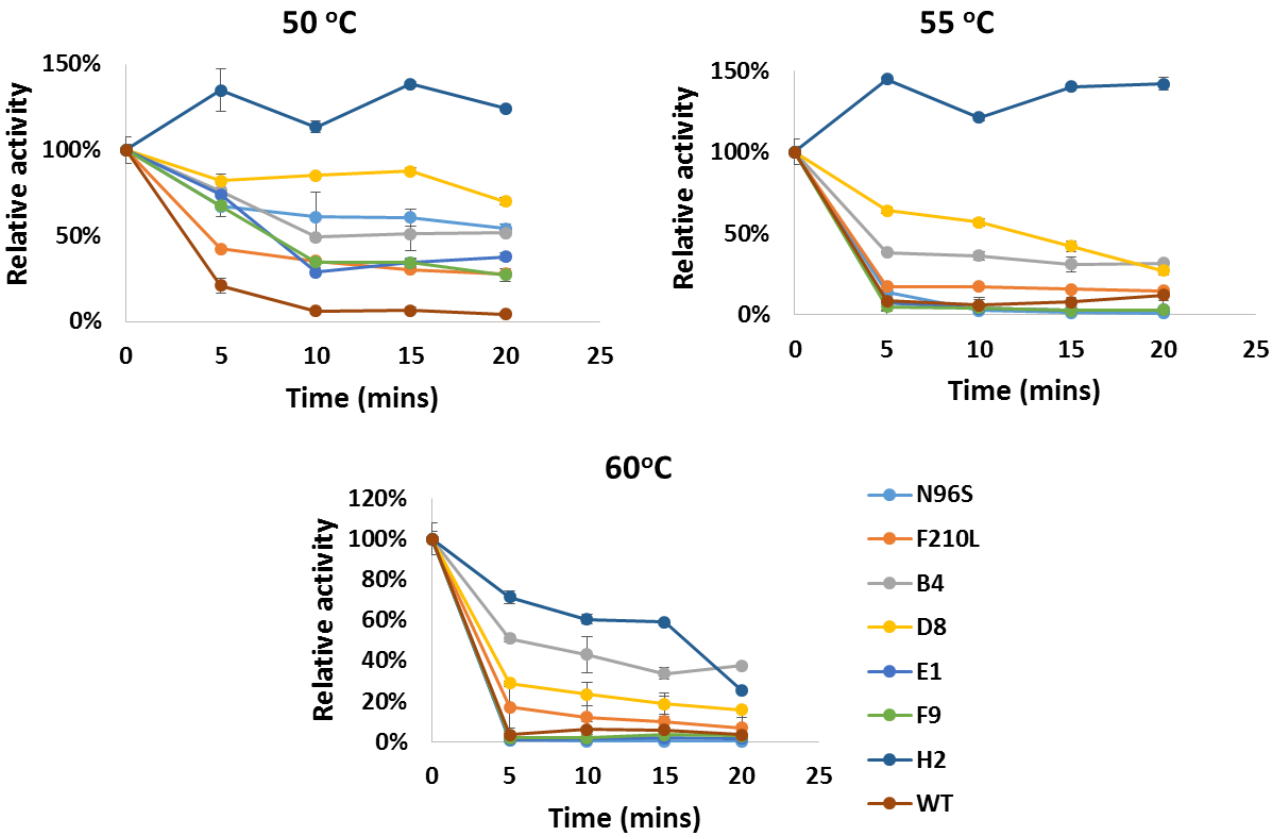


Figure 5.4: Thermal stability profile of NaM1 and its variants using their cell free extracts.

Thermal inactivation assays using cell free extracts of the five mutants revealed that H2 retained 100% activity after 20 min incubation at 50 and 55°C. H2 was consistently activated rather than inactivated at 50 and 55°C (Figure 5.5) suggesting that the enzyme was stabilized and more active than at 4°C. Using purified enzyme, the half-life for both NaM1 (Figure 3.6) and H2 (Figure 5.6b) was 23 min, but at 50 and 65°C, respectively. This confirmed that a NaM1 variant with an improved thermal stability of at least 10°C was obtained from one round of epPCR in this experiment, whereas, DE experiments frequently require tens of thousands of clones or several generations to be screened before a 10°C increase in thermal stability is achieved (Gumulya and Reetz, 2011, Wintrode and Arnold, 2001). Since mutant H2 displayed the most significant thermal stability traits, all further analyses concentrated on this variant. The amino acid(s) responsible for this improved stability were subsequently investigated following SDM.

5.2.3.2 Site-directed mutagenesis variants

Two substitutions, N96S and F210L were identified in the most thermostable variant, H2. These were caused by two point mutations; A293G and T634C on the NaM1 gene (*NaMet1*). Site-directed mutagenesis of *NaMet1* to delineate the two mutations and create variants with single mutations was successful (Appendix 2I). Each variant was designated $NaM1_{N96S}$ and $NaM1_{F210L}$ and further analyzed. Thermal stability studies of $NaM1_{N96S}$ and $NaM1_{F210L}$ indicated that both substitutions contributed to the increase of 10°C in thermal stability of H2. $NaM1_{N96S}$ and $NaM1_{F210L}$ retained only around 50 and 20% activity, respectively, after incubation at 50°C for 20 min. However, both variants were substantially less thermostable than H2. $NaM1_{F210L}$ was six-fold less thermostable than H2 after incubation at 60°C for 10 minutes and retained only ~15% activity after 20 minutes incubation at 55°C while H2 retained 100% activity after similar treatment (Figure 5.5). It is not clear if the improved stability of H2 was due to interactions between the active site, oxyanion hole and their surrounding residues or directly between Phe210 and Asn96 (Kamerzell and Middaugh, 2008).

5.2.4 Optimal activity assays of H2

The 'optimal' temperature and pH for H2 were determined to be 40°C (Figure 5.6a) and pH 8.5 (Figure 5.6c), respectively.

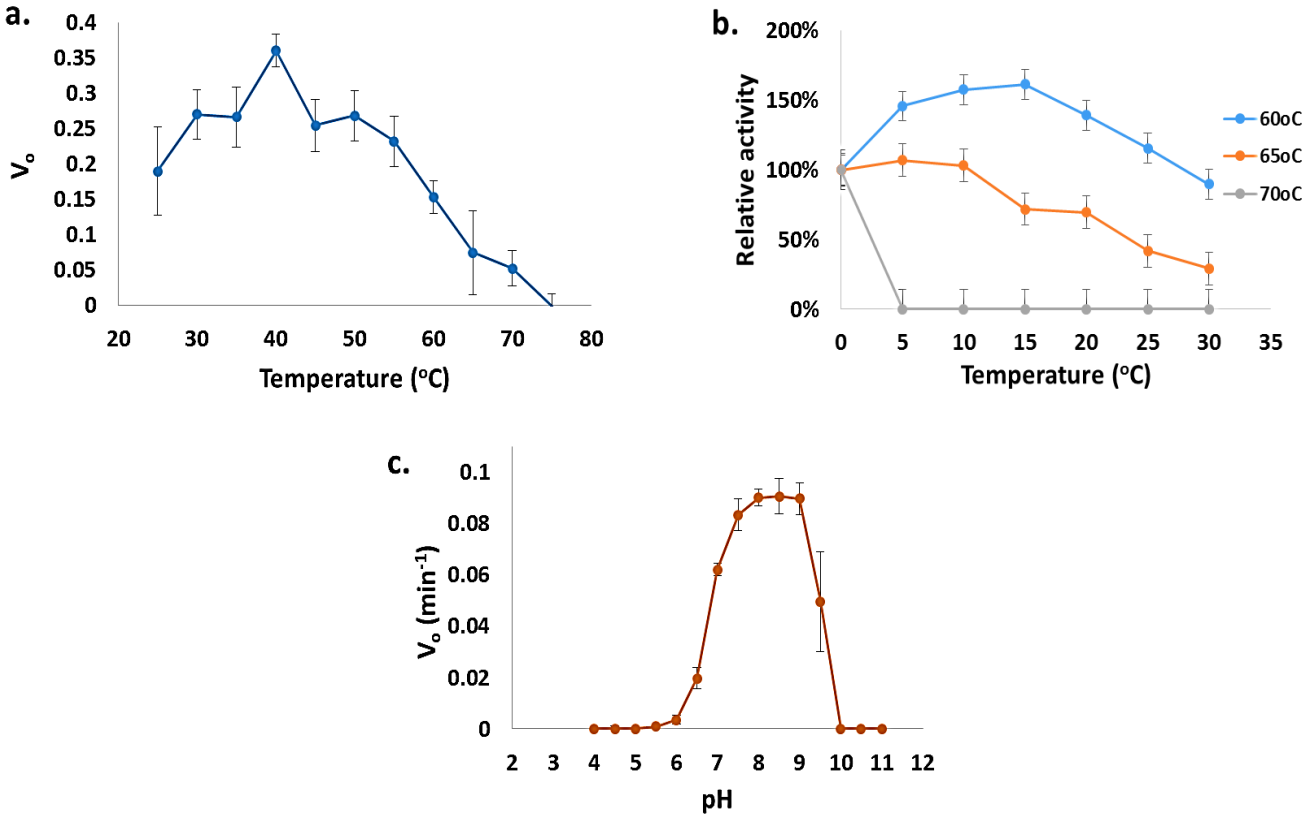


Figure 5.5: Functional characteristics of purified H2

a. H2 deacetylase activity between 25 – 75°C showing 'thermal optimum' at 40°C. **b.** Thermal stability profile of H2 at 60, 65 and 70°C. **c.** H2 activity profile between pH 4 – 11.

5.2.5 Kinetics of NaM1 and thermostable variants

The kinetic properties of $NaM1_{F210L}$, $NaM1_{N96S}$, H2 and WT gave some insights into the catalytic traits of all four enzymes (Table 5.3). $NaM1_{F210L}$ and H2 had the lowest catalytic efficiencies (k_{cat}/K_M) and lost ~80 and 70% activity, respectively. The activity of $NaM1_{N96S}$, by contrast, increased by 30% activity relative to the WT. The trade-off between activity and thermostability is a common (Wintrode and Arnold, 2001), but not absolute (Williams et al., 1999) consequence of improving thermostability through point mutations. However, this relationship between temperature and stability indicates that there was an improved $\Delta G^{\ddagger}_{inact}$, but little productive change in T_{eq} and/or ΔH_{eq} as explained by the equilibrium model theory (Daniel and Danson, 2013).

$NaM1_{N96S}$ had the lowest K_M of 0.42 mM indicating its stronger affinity for the substrate, while H2 had

the lowest affinity for the substrate. This suggests that the N96S mutation remodels the substrate–binding site to stabilize, orient or present the substrate more favorably for catalysis. The catalytic efficiency of $NaM1_{F210L}$ was half that of H2, whereas that of $NaM1_{N96S}$ was 5-fold more efficient than H2. The substitutions, thus, had opposite effects on the overall activity of H2, with one mutation improving and the other decreasing catalysis. The F210L mutation had the most significant cumulative effect. Thus, it seems evident that the stabilizing effect of one substitution (N96S) offsets the destabilizing effect of the other (F210L) (Socha and Tokuriki, 2013). Point mutations have been shown to induce structural changes that may interfere with protein dynamics (Kamerzell and Middaugh, 2008). The reduced activity of H2 incurred for improved stability is not a desirable trait, but confirms the observation that increased thermostability without activity loss is mainly achieved by a combination of substitutions at various locations across the protein (Currin et al., 2015). The kinetic parameters for H2 with other substrates are listed in Appendix 2J.

Lower catalytic efficiency ($8.43x10^5$ $M^{-1}s^{-1}$) was recorded for NaM1 at 40°C (Table 5.3) than at 25°C (Table 3.1). However, $NaM1_{N96S}$ showed comparable catalytic efficiency at 40°C ($1.34x10^6$ $M^{-1}s^{-1}$) to that observed for NaM1 at 25°C ($3.26x10^6$ $M^{-1}s^{-1}$). This confirms that $NaM1_{N96S}$ had an increased k_{cat} and was more thermally stable showing improved substrate binding affinity and catalysis at a temperature above the T_{opt} of the wildtype.

Table 5.3: Kinetic properties of $NaM1_{WT}$ and variants, H2, $NaM1_{N96S}$ and $NaM1_{F210L}$ at 40°C.

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
NaM1	2.75 ± 0.34	832.36 ± 102.5	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.

5.2.6 Thermostability and acetylated xylan specificity

No deacetylase activity was detected after co-incubating H2 with AX, whereas slight activity was observed with the WT. However, H2 does bind substrates with larger acyl groups (up to C8) with higher affinity than NaM1. The relationship between lack of activity on AX and H2 thermal stability is not clear, but it follows the current trend of inactivity on AX by highly thermostable CE7 AcXEs. The results in this study support the view that the preferred substrate of CE7 enzymes in nature is not xylan, but acetylated xylooligosaccharides (AcXOS). Considering the absence of a signal peptide and the inability of AX to access the cell, deacetylation of AX by NaM1 would only be possible if the enzyme were partly secreted. Yet, characterized CE7 AcXEs are known to be intracellularly produced (Adesioye et al., 2016). NaM1 and other CE7 AcXEs presumably only remove peripheral acetate groups from the AX polymer.

5.2.7 Fluorescence spectra of NaM1 and H2

The UV emission spectra for NaM1 and H2 when excited at 280 nm were broadly similar, but a significantly higher absorbance was recorded for NaM1 at 295 nm (Figure 5.7).

The difference in absorbance between NaM1 and H2 at the 295 nm excitation wavelength indicates a difference in the folding pattern of both proteins. The stronger absorbance exhibited by NaM1 suggests that, in comparison with H2, NaM1 is less compactly folded or simply differently folded. This may have resulted in the exposure of more aromatic residues to the protein surface and consequently to the UV source. The variation in folding pattern might be a contributing factor to thermal stability displayed by H2. This assumption was not made without considering the difference in aromatic amino acid residue composition of both proteins. At 295 nm, tryptophan exhibits maximum fluorescence emissions from phenylalanine and tyrosine. The F210L mutation in H2 resulted in a 7% reduction in phenylalanine residues in both proteins which inadequately accounts for the 51% difference in absorbance at 295 nm. Since there was no difference in the number of tryptophan residues present in both proteins, the mutated residues in H2 may have resulted in a slight change in conformation affecting one or more tryptophan residues and consequently resulting in the lower absorbance recorded (Karlish and Yates, 1978).

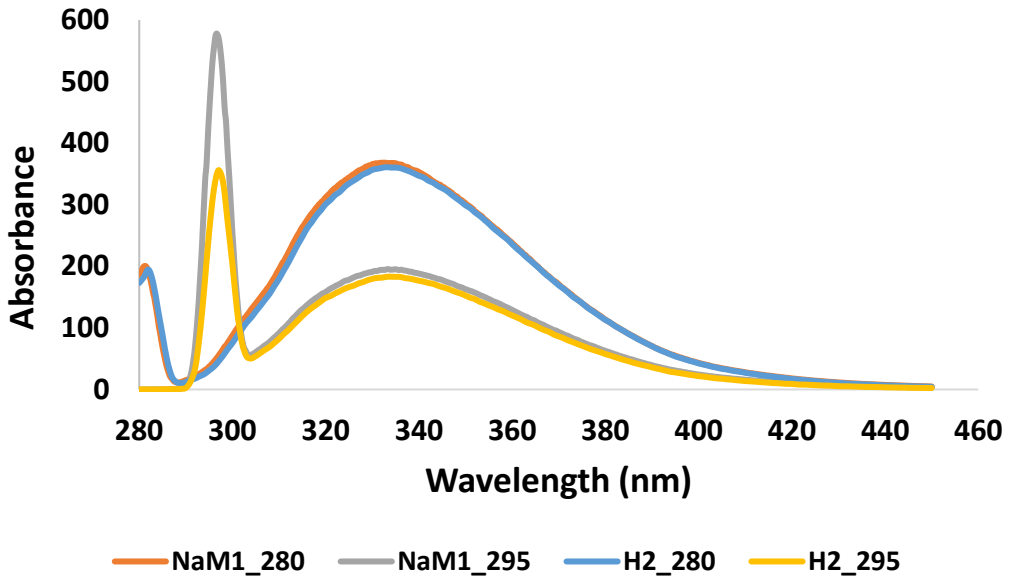


Figure 5.6: Fluorescence spectra of NaM1 and H2 at 280 and 295 nm excitation wavelengths

5.2.8 Thermal-induced unfolding of NaM1 and H2

The unfolding of NaM1 and H2 secondary structure was monitored using far-UV CD spectroscopy. For each variant, the ellipticity at 210 nm, a wavelength particularly sensitive to changes in secondary structure, was plotted against the temperature (Figure 5.8). The CD spectra of NaM1 and H2 at 25°C indicated both proteins were fully folded at this temperature (Appendix 2K) suggesting that the secondary structure of NaM1 was largely unaffected by the substitutions in the variant, H2 (Karlish and Yates, 1978). However, the temperature profile of unfolding is distinctly different for NaM1 and H2, with NaM1 demonstrating a melting temperature (T_m) of ~47 and an unfolding transition at 45°C while H2 has a melting temperature of 63°C, a transition at 60°C and complete loss of secondary structure at 70°C (Figure 5.8a). The sigmoidal shape of the NaM1 plot indicates a Boltzmann two-state model for the transitioning of proteins, from an ordered, folded state to a disordered unfolded state. The transition curves for $NaM1_{N96S}$ and $NaM1_{F210L}$ showed that thermal unfolding began at 50°C and complete structure loss was only apparent at 70°C (Figure 5.8b). However, thermal stability assays of $NaM1_{N96S}$ and $NaM1_{F210L}$ show that they were not active beyond 60°C (Figure 5.5). These suggest that both enzymes were thermostable, but not thermoactive, between 60 and 70°C (Daniel et al., 2001). Thus,

both substitutions contributed to the thermal stability of H2. The unfolding transition curves imply a T_m of 52°C for both $NaM1_{N96S}$ and $NaM1_{F210L}$. These values agree with the results obtained from thermal stability and inactivation studies of the NaM1 variants.

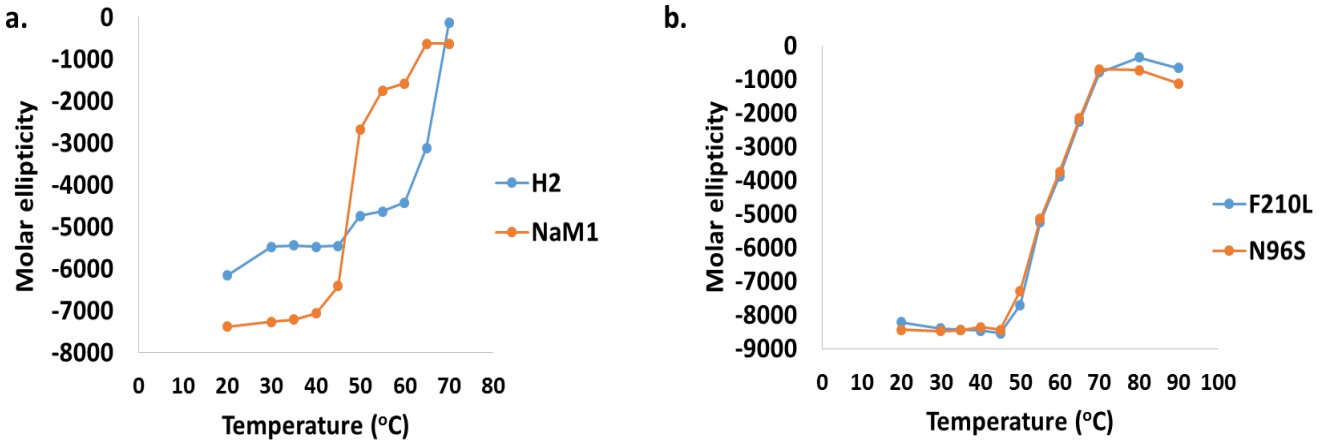


Figure 5.7: Thermal unfolding profiles recorded at 210 nm.

a. Profiles of NaM1 and H2. **b.** Profiles of $NaM1_{N96S}$ and $NaM1_{F210L}$. Inferred melting temperatures are approximately 47, 63, 52 and 52°C, respectively.

5.2.9 Sequence and structure to function analyses of mutants

An analysis of the structural models generated for all NaM1 variants yielded some information on the structural implications of single residue substitutions. PCR mutagenesis is biased towards transitions over transversions, but among mutants selected here, both mutation types occurred equally often. However, the mutations in thermostable variants were mainly transitions (Table 5.2) and were all accompanied with k_{cat} loss. Protein activity, thermostability and flexibility are known to be complementary (Wintrode and Arnold, 2001, Daniel and Danson, 2013, Daniel, 1996). Mutations that improve both activity and thermostability are nevertheless desirable and achievable (Currin et al., 2015). Variants H2, D8 and B4 each had two point mutations. Of these, one mutation was respectively silent in D8 and B4, leaving only H2 with two amino acid substitutions: N96S and F210L. The latter preserves the hydrophobicity at that position, but due to its proximity to the active site may be responsible for the observed loss in activity. This interpretation was confirmed by analyzing $NaM1_{F210L}$ produced by SDM. Substitutions of selected NaM1 variants are discussed below and illustrated in Appendix 2L.

5.2.9.1 F210L

Phe210 in NaM1 is a conserved aromatic residue position in all CE7 esterases, existing either as a tyrosine or phenylalanine. The N-atom of Phe210 maintains H-bonding with the O-atom of the catalytic Ser185 and this interaction is maintained by Leu210 in H2. However, while the aromatic side chain of Phe210 faces away from Ser185 in NaM1, the modelled structure of H2 predicts the side-chain of Leu210 to face towards Ser185 creating unfavorable interactions between C^{γ} and $C^{\delta 1}$ of Leu210 and the O^{γ} of Ser185. These interactions may interfere with substrate positioning during a nucleophilic attack by Ser185 O^{γ}. Further to this, there was a resultant shift in active center orientation due to the F210L mutation. Hence, the distance between the N^{δ} His304 and O^{δ} Asp275 was increased from 3.11 to 3.35 Å, which nullifies or weakens the H-bond (typically 3.2 Å) between both residues (Figure 5.10c). This hydrogen bond critically stabilizes protonated His304 activation of the enzyme nucleophile Ser185 by deprotonation. Loss of this H-bond may explain the reduced catalytic efficiency observed for H2 and $NaM1_{F210L}$ (Dodson and Wlodawer, 1998). A previous study (Montoro-García et al., 2011) suggested that amino acids in the Phe210 position of CE7 esterases may be involved in constraining the size and orientation of the acyl moiety in the active site. A more recent study suggested that the conserved P228 position in *T. maritima* CE7 esterase (P226 in NaM1) served a similar purpose (Singh and Manoj, 2017). Since the P226 position remained unchanged in this study, the results obtained from H2 characterization further suggest that position Phe210 indeed constrains the size of the acyl moiety the most. Hence, leucine being smaller than phenylalanine, should potentially accommodate longer acyl chains (Figure 5.9). However, $NaM1_{F210L}$ actively cleaved *p*-NPB but not *p*-NPO; H2 cleaved *p*-NPB and *p*-NPO, but not *p*-NPP and AX; and NaM1 had lower activity on *p*-NPB than H2, but was inactive on >C4 substrates (Figure 5.11). In view of this, the detectable release of octanoate from *p*-NPO by H2, but not by $NaM1_{F210L}$ was thought to be due to the positive counter-effect of the N96S against the F210L mutation in H2.

Furthermore, Phe210 is important for deacetylation as the aromatic ring contributes to shaping the base of the substrate-binding pocket. Replacing it with an aliphatic residue, as in H2 and $NaM1_{F210L}$, reduced deacetylation activity significantly (Figures 5.4 and 5.9). Sequence alignment with certain GDSL lipases

(Appendix 2M) indicates that the Phe210 position is strictly conserved as phenylalanine or tyrosine in these enzymes. Hence, the results from this study may not only provide information on a combination of amino acids that can be optimized for activity, substrate specificity and/or thermal stability of CE7 enzymes, but also of these α/β hydrolases.

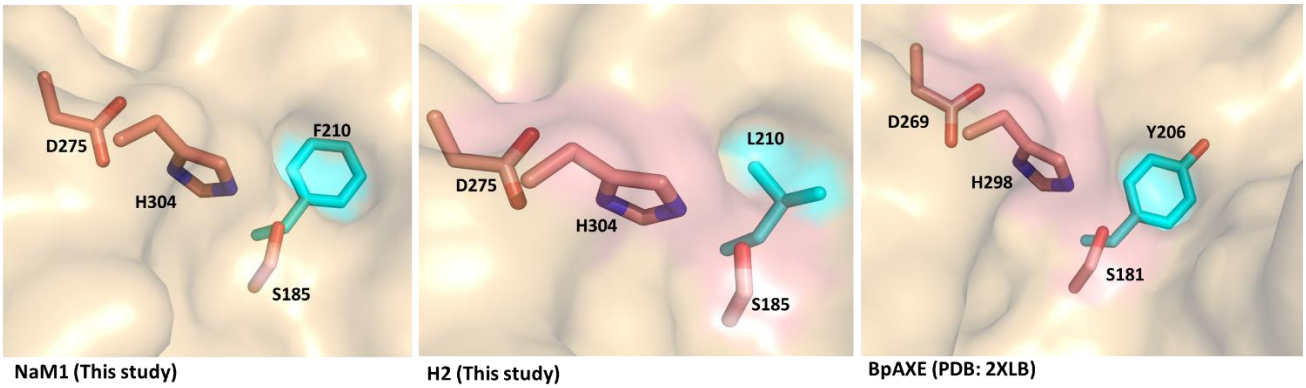


Figure 5.8: Proposed residue position restricting acyl group length at the S2 binding site.

Tyr206 in *Bp*AXE prevents access of substrates with ≥C4 acyl groups. Phe210 in NaM1 allows and restricts substrates with C4 and >C4 acyl groups, respectively. Leu210 in H2 may potentially allow substrates with >C4 acyl moieties access the S2 binding site.

5.2.9.2 N96S

Position 96 in NaM1 is variable in all CE7 enzymes, yet, it is linked to enzyme activity in NaM1. Asn96 is located in a loop near β-strands β4 and β5, faces the inner cavity of the NaM1 hexamer and is partly exposed to solvent (Figure 4.6). It shares a loop with Tyr93, forming part of the oxyanion hole. Asn96 may thus, be help to stabilize the oxyanion hole and the reaction intermediate during catalysis (Simón and Goodman, 2009). The replacement of Asn96 with serine resulted in the loss of stabilizing H-bonds of Asn96 O^{δ} to Gly98 O and a water molecule and Asn96 $N^{\delta 2}$ to Ala122 N (Figure 5.10a). Ala122 is located on the nearby β-interface loop (Gly120 – Leu 140) which is important for thermal stability and activity in the *T. maritima* enzyme (Singh and Manoj, 2016b). Though neither Phe210 nor Asn96 form part of the β-interface loop, the H-bond linking Asn96 to Ala122 may be important for NaM1 stability.

Replacing asparagine with similarly polar but smaller serine, indicate that increased protein flexibility and stability do not necessarily counteract each other (Kamerzell and Middaugh, 2008, Currin et al., 2015). Improved thermal stability may be linked to protein charge distribution (Kamerzell and Middaugh, 2008, Daniel and Danson, 2013) and the location of the mutation (Wintrode and Arnold, 2001). Point mutations in the protein core often disrupt optimal packing. Hence substitutions N96S and F210L may induce minimal, yet significant changes improving hydrogen bonding, loop stabilization and core packing in variant H2 of NaM1. This has been observed for a similar N-S substitutions in a thermostable variant from the DE library of a thermolabile subtilisin E (Wintrode and Arnold, 2001). Thus, thermophilicity may not be essential for this family of enzymes and protein folds may have developed by divergent evolution through adaptation.

Surprisingly, $NaM1_{N96S}$ was found to be active on *p*-NPO while the wild type and mutant $NaM1_{F210L}$ were not. This observation confirmed that the N96S mutation was responsible for expanding the substrate specificity of NaM1. This annuls the initial hypothesis that the F210L mutation allowed this expansion in substrate specificity. Thus, the N96S mutation, with a distance of 12.51 Å from the catalytic serine, significantly expands the specificity of NaM1 from being an esterase, acting on short chain (≤C4) acyl substrates, to one acting on medium length (≥C8) acyl substrates.

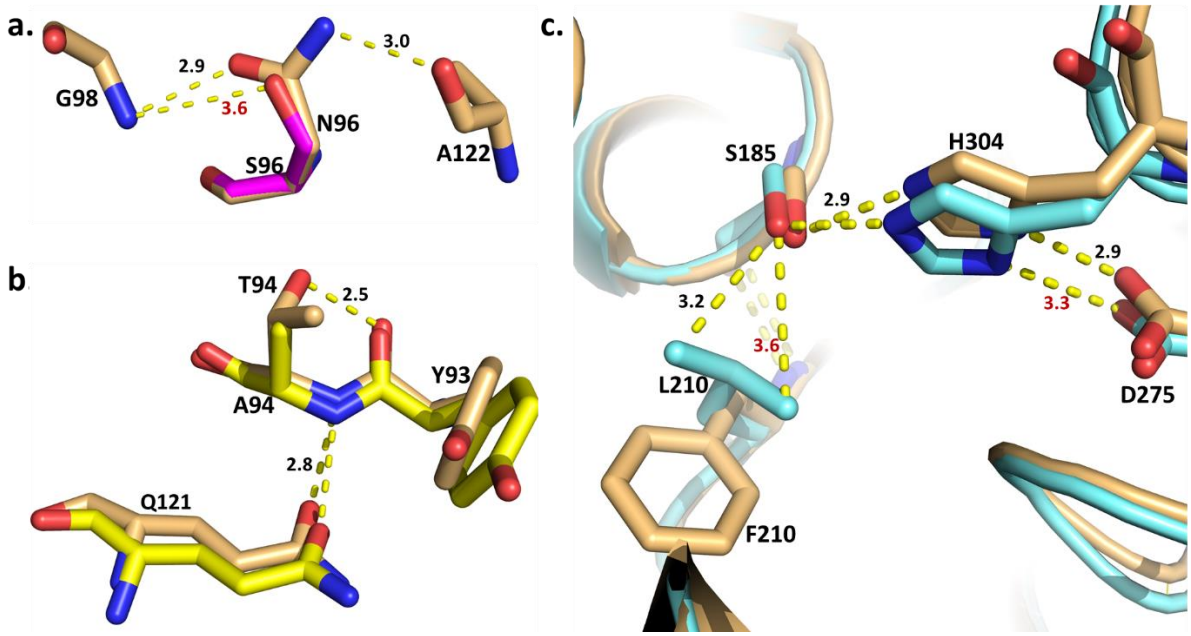


Figure 5.9: Impact of residue substitutions in thermostable NaM1 variants.

NaM1 is tan in color and each modelled variant a unique color. **a.** Loss of H-bond with G98 and A122 due to N96S substitution. **b.** Loss of H-bond with Y93 due to T94A mutation. **c.** The substitution of Phe210 with leucine creates unfavorable interactions between L210 and catalytic S185, and increases the distance between catalytic residues H304 and D275. Distance between atoms is measured in Å; polar interactions longer than typical H-bonds are labeled in red.

5.2.9.3 T94A

NaM1 variants D8 and B4 with improved thermal stability, carried the T94A and N228D substitutions, respectively. Both variants suffered ≥70% loss in activity, and a three and two-fold lower stability than H2, respectively, after incubation at 60°C for five minutes. Substitutions in the D8 and B4 variants are in close proximity, respectively to Asn96 and Pro226, discussed above. Thr94, like Asn96, is not conserved in CE7 hydrolases, is exposed to solvent and can be replaced by alanine or serine (Betts and Russell, 2007). Thr94 is closer to the oxyanion hole than Asn96 and forms one hydrogen and non-directed interactions with the Tyr93 carbonyl. The alanine substitution eliminates the H-bond between O-Tyr93 and $O^{\gamma 1}$-Thr94 (Figure 5.10b) possibly increasing substrate flexibility and hence, reducing activity. How this relates to thermal stability remains unclear. The structural model of D8, which retained

~60% activity after 20 minutes incubation at 55°C, indicates that the T94A substitution, located at a β-turn, enhanced packing and rigidity at higher temperatures (Britton et al., 1995). Furthermore, residue 94 position in NaM1 is always small and hydrophilic (serine, asparagine or threonine) in CE7 enzymes. The replacement of polar threonine with non-polar, hydrophobic alanine may have changed the microenvironment considerably and increased hydrophobicity at this location. This appears to be critical for thermal stability at the protein core (Britton et al., 1995), but unfavorable for the stabilization of the hydrophilic *p*-NPA intermediate (Mattiasson and Ye, 2015). Replacing a hydrophilic oxyanion hole residue by a hydrophobic residue was similarly observed to reduce access of hydrophilic substrates lowering their specificity (higher K_M) but increasing that of more hydrophobic substrates (Lee et al., 2006).

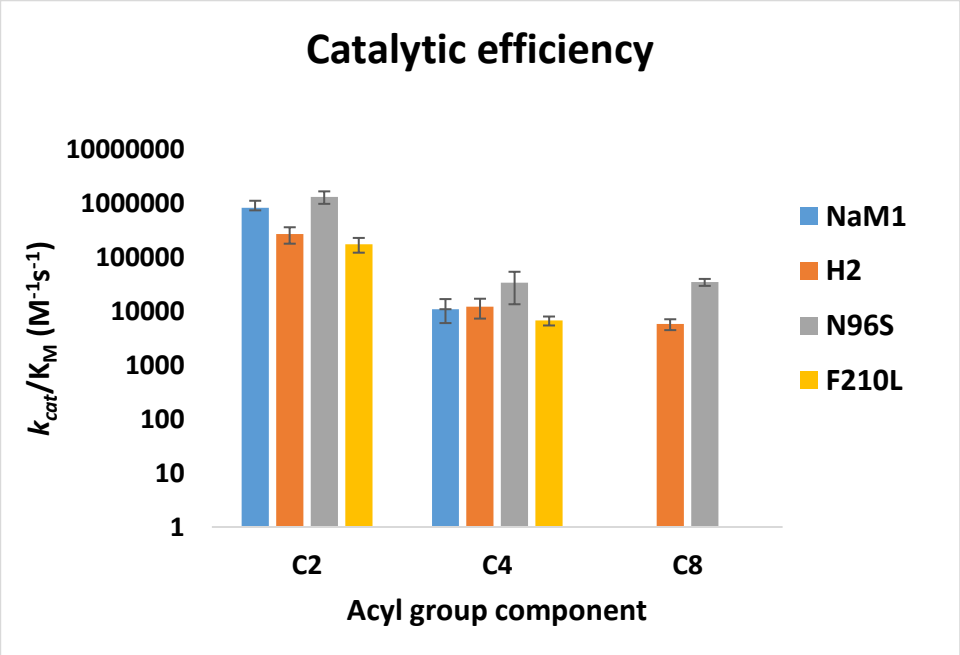


Figure 5.10: Comparison of the catalytic efficiencies of NaM1, H2, $NaM1_{N96S}$ and $NaM1_{F210L}$ on substrates with various lengths of acyl moieties.

5.2.9.4 N228D

In variant B4, Asn228 was replaced by its isosteric, charged counterpart aspartate. This kind of

substitution has been reported in previous studies (Wintrode and Arnold, 2001) to contribute to improved thermal stability. The reverse substitution pattern (e.g. D→N, E→Q) has also been associated with improved thermal stability (Reetz et al., 2006). A single E→Q substitution resulted in 26°C increase in thermal stability without loss in catalytic potency of the protein (Williams et al., 1999). The Nδ2 atom of Asn228 interacts with a water molecule, an interaction that stabilizes the insertion helix (α7) on which the residue is located (Figure 4.3). Furthermore, Asn228 forms non-bonded interactions and a H-bond with Pro226. The loss in activity observed due to this mutation in NaM1 may be related with its proximity to the substrate binding center specifically to Pro226 which serves as a barrier gate shaping the S2 binding site. As asparagine and aspartate essentially have the same size, a N228D substitution would not be expected to affect the substrate binding site. Instead, the introduction of a negative charge near the active site may affect the charge distribution during catalysis, reducing enzyme activity in variant B4 as many enzymatic reactions directly depend on defined charge distribution on certain amino acids (Daniel and Danson, 2013).

5.2.9.5 T306P

Variant E1 involved an A→C transversion resulting in the substitution of hydrophilic Thr306 with hydrophobic Proline (T306P). This resulted in a two-fold increase in activity (Figure 5.4) which might be attributed to a stabilization of NaM1 due to the hydrophobicity and rigidity of Proline. Pro306 is located close to the active center on the same loop as His304, a catalytic triad residue and potentially stabilizes the loop. Conformational changes induced by Pro306 will likely have unmediated effects on His304. Here, the substitution appears to have improved the active site or substrate binding site in favor of catalysis.

5.2.9.6 E11A

The substitution of Glu11 with alanine in variant F9 resulted in a slight (1.12-fold) increase in deacetylase activity (Figure 5.4). The Glu11 position is conserved in all NaM1 homologs as a glutamate (E), glutamine (Q) or lysine (K) residue indicating the requirement of extended side chains with terminal

functional groups (Figure 4.6, Appendix 2H). Hydrogen bonding and non-bonded contacts associated with these E, Q or K residues in this position only involve their backbone amino and carbonyl groups. Hence, it is not immediately clear why a substitution with Alanine would improve k_{cat}. The residue, located near the protein N-terminus is distant from the active site, but possibly contributes to the motions that direct catalysis in NaM1. Changes in k_{cat} have previously been associated with residues remote from the active site (Currin et al., 2015).

Overall, the substitutions improving NaM1 thermal stability involved larger residues (phenylalanine, asparagine and threonine) being replaced with slightly smaller residues (leucine, serine and alanine, respectively). This suggests that improved packing and compactness facilitated by smaller residues, rather than improved rigidity, determined the thermostability of NaM1 mutants. Given the properties of H2, it would seem that a number of thermostabilizing and k_{cat}-improving substitutions of NaM1 would be required to produce a significantly more thermostable variant with a higher k_{cat}.

5.3 CONCLUSION

In this chapter, the successful *in-vitro* evolution by epPCR of a novel thermolabile CE7 AcXE to a more thermostable variant was described. Two amino acid substitutions were found to account for this 10°C increase in thermal stability – albeit, with 70% reduced activity. These results established that residues in CE7 AcXEs corresponding to Phe210, Thr94 and Asn228 in NaM1 critically affect both activity and thermal stability, while positions corresponding with Thr 306 and Glu11 are key players in CE7 family activity. Overall, tighter packing in the core of CE7 enzymes was found to be essential for thermostability and all stabilizing mutations were located within the α/β hydrolase fold of NaM1. This study showed that a single N96S substitution extended the substrate specificity, as well as, improved the thermal stability and catalytic efficiency of NaM1. Furthermore, it is evident that mutations around the oxyanion hole and active site are significant for conformational transition of NaM1 towards modifying both k_{cat} and thermal stability of CE7 enzymes (Britton et al., 1995). Since thermostability is usually known to correspond with evolution of a protein as a result of several mutations (Currin et al., 2015), H2 and $NaM1_{N96S}$ are suitable

templates for further evolution of NaM1 to yield more information on the thermal stability and substrate specificity determinants of the CE7 enzyme family. This knowledge is valuable towards future engineering of CE7 enzymes and possibly other α/β hydrolases.

5.4 REFERENCES

ADESIOYE, F. A., MAKHALANYANE, T. P., BIELY, P. & COWAN, D. A. 2016. Phylogeny, classification and metagenomic bioprospecting of microbial acetyl xylan esterases. *Enzyme and Microbial Technology,* 93**,** 79-91.

ARNOLD, F. H., WINTRODE, P. L., MIYAZAKI, K. & GERSHENSON, A. 2001. How enzymes adapt: lessons from directed evolution. *Trends in Biochemical Sciences,* 26**,** 100-106.

BETTS, M. J. & RUSSELL, R. B. 2007. Amino-acid properties and consequences of substitutions. *Bioinformatics for Geneticists,* 2**,** 311-342.

BRITTON, K. L., BAKER, P. J., BORGES, K. M., ENGEL, P. C., PASQUO, A., RICE, D. W., ROBB, F. T., SCANDURRA, R., STILLMAN, T. J. & YIP, K. S. 1995. Insights into thermal stability from a comparison of the glutamate dehydrogenases from Pyrococcus furiosus and Thermococcus litoralis. *European Journal of Biochemistry,* 229**,** 688-695.

CHICA, R. A., DOUCET, N. & PELLETIER, J. N. 2005. Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Current Opinion in Biotechnology,* 16**,** 378-384.

COPP, J. N., HANSON-MANFUL, P., ACKERLEY, D. F. & PATRICK, W. M. 2014. Error-prone PCR and effective generation of gene variant libraries for directed evolution. *Directed Evolution Library Creation: Methods and Protocols***,** 3-22.

CUNNINGHAM, B. C. & WELLS, J. A. 1989. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science,* 244**,** 1081-1086.

CURRIN, A., SWAINSTON, N., DAY, P. J. & KELL, D. B. 2015. Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. *Chemical Society Reviews,* 44**,** 1172-1239.

DANIEL, R. M. 1996. The upper limits of enzyme thermal stability. *Enzyme and Microbial Technology,* 19**,** 74-79.

DANIEL, R. M. & DANSON, M. J. 2013. Temperature and the catalytic activity of enzymes: A fresh understanding. *FEBS Letters,* 587**,** 2738-2743.

DANIEL, R. M., DANSON, M. J. & EISENTHAL, R. 2001. The temperature optima of enzymes: a new perspective on an old phenomenon. *Trends in Biochemical Sciences,* 26**,** 223-225.

DANSON, M. J., HOUGH, D. W., RUSSELL, R. J., TAYLOR, G. L. & PEARL, L. 1996. Enzyme thermostability and thermoactivity. *Protein Engineering, Design and Selection,* 9**,** 629-630.

DAVIDS, T., SCHMIDT, M., BÖTTCHER, D. & BORNSCHEUER, U. T. 2013. Strategies for the discovery and engineering of enzymes for biocatalysis. *Current Opinion in Chemical Biology,* 17**,** 215-220.

DELANO, W. L. 2002. The PyMOL molecular graphics system. http://pymol.org.

DICOSIMO, R., PAYNE, M. S. & GAVAGAN, J. E. 2013. Perhydrolase variant providing improved specific activity. Google Patents.

DODSON, G. & WLODAWER, A. 1998. Catalytic triads and their relatives. *Trends in Biochemical Sciences,* 23**,** 347-352.

EMSLEY, P. & COWTAN, K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography,* 60**,** 2126-2132.

EZEMADUKA, A. N., YU, J., SHI, X., ZHANG, K., YIN, C.-C., FU, X. & CHANG, Z. 2014. A small heat shock protein enables Escherichia coli to grow at a lethal temperature of 50 C conceivably by maintaining cell envelope integrity. *Journal of Bacteriology,* 196**,** 2004-2011.

GOOMBER, S., KUMAR, A. & KAUR, J. 2016. Disruption of N terminus long range non covalent interactions shifted temp. opt 25° C to cold: Evolution of point mutant Bacillus lipase by error prone PCR. *Gene,* 576**,** 237-243.

GUMULYA, Y. & GILLAM, E. M. 2017. Exploring the past and the future of protein evolution with ancestral sequence reconstruction: the ‘retro’approach to protein engineering. *Biochemical Journal,* 474**,** 1-19.

GUMULYA, Y. & REETZ, M. T. 2011. Enhancing the thermal robustness of an enzyme by directed evolution: least favorable starting points and inferior mutants can map superior evolutionary pathways. *ChemBioChem,* 12**,** 2502-2510.

HANAHAN, D. 1983. Studies on transformation of Escherichia coli with plasmids. *Journal of Molecular Biology,* 166**,** 557-580.

HANDL, J., KELL, D. B. & KNOWLES, J. 2007. Multiobjective optimization in bioinformatics and computational biology. *IEEE/ACM Transactions on Computational Biology and Bioinformatics,* 4 (2).

HEDGE, M. K., GEHRING, A. M., ADKINS, C. T., WESTON, L. A., LAVIS, L. D. & JOHNSON, R. J. 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.

HOSSAIN, G. S., SHIN, H.-D., LI, J., WANG, M., DU, G., LIU, L. & CHEN, J. 2016. Integrating error-prone PCR and DNA shuffling as an effective molecular evolution strategy for the production of α-ketoglutaric acid by L-amino acid deaminase. *RSC Advances,* 6**,** 46149-46158.

JIANG, W., ZHUANG, Y., WANG, S. & FANG, B. 2015. Directed Evolution and Resolution Mechanism of 1, 3-Propanediol Oxidoreductase from Klebsiella pneumoniae toward Higher Activity by Error-Prone PCR and Bioinformatics. *PLOS ONE,* 10 (11)**,** e0141837.

KAMERZELL, T. J. & MIDDAUGH, C. R. 2008. The complex inter-relationships between protein flexibility and stability. *Journal of Pharmaceutical Sciences,* 97**,** 3494-3517.

KARLISH, S. & YATES, D. 1978. Tryptophan fluorescence of (Na++ K+)-ATPase as a tool for study of the enzyme mechanism. *Biochimica et Biophysica Acta (BBA)-Enzymology,* 527**,** 115-130.

KATAOKA, M., MIYAKAWA, T., SHIMIZU, S. & TANOKURA, M. 2016. Enzymes useful for chiral compound synthesis: structural biology, directed evolution, and protein engineering for industrial use. *Applied Microbiology and Biotechnology,* 100**,** 5747-5757.

KAUR, J. & SHARMA, R. 2006. Directed evolution: an approach to engineer enzymes. *Critical Reviews in Biotechnology,* 26**,** 165-199.

KAZLAUSKAS, R. J. & BORNSCHEUER, U. T. 2009. Finding better protein engineering strategies. *Nature Chemical Biology,* 5**,** 526-529.

KELLY, S. M. & PRICE, N. C. 2000. The use of circular dichroism in the investigation of protein structure and function. *Current Protein and Peptide Science,* 1**,** 349-384.

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.

KUMAR, A. & SINGH, S. 2013. Directed evolution: tailoring biocatalysts for industrial applications. *Critical Reviews in Biotechnology,* 33**,** 365-378.

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.

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.

LUTZ, S. 2010. Beyond directed evolution—semi-rational protein engineering and design. *Current Opinion in Biotechnology,* 21**,** 734-743.

MARTINEZ-MARTINEZ, I., MONTORO-GARCIA, S., LOZADA-RAMIREZ, J. D., 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.

MATTIASSON, B. & YE, L. 2015. *Molecularly Imprinted Polymers in Biotechnology*, Springer

MELZER, S., SONNENDECKER, C., FÖLLNER, C. & ZIMMERMANN, W. 2015. Stepwise error-prone PCR and DNA shuffling changed the pH activity range and product specificity of the cyclodextrin glucanotransferase from an alkaliphilicBacillus sp. *FEBS Open Bio,* 5**,** 528-534.

MORAÏS, S., STERN, J., KAHN, A., GALANOPOULOU, A. P., YOAV, S., SHAMSHOUM, M., SMITH, M. A., HATZINIKOLAOU, D. G., ARNOLD, F. H. & BAYER, E. A. 2016. Enhancement of cellulosome-mediated deconstruction of cellulose by improving enzyme thermostability. *Biotechnology for Biofuels,* 9**,** 164.

PEI, H., GUO, X., YANG, W., LV, J., CHEN, Y. & CAO, Y. 2015. Directed evolution of a β-1, 3-1, 4-glucanase from Bacillus subtilis MA139 for improving thermal stability and other characteristics. *Journal of Basic Microbiology,* 55**,** 869-878.

PORTER, J. L., RUSLI, R. A. & OLLIS, D. L. 2016. Directed evolution of enzymes for industrial biocatalysis. *ChemBioChem,* 17**,** 197-203.

REETZ, M. T. & CARBALLEIRA, J. D. 2007. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nature Protocols,* 2**,** 891-903.

REETZ, M. T., CARBALLEIRA, J. D. & VOGEL, A. 2006. Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability. *Angewandte Chemie International Edition,* 45**,** 7745-7751.

SAMBROOK, J. & RUSSELL, D. 2001. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor: Cold Spring Harbor Laboratory Press. Pp 3.17-3.32.

SHAO, W., MA, K., LE, Y., WANG, H. & SHA, C. 2017. Development and Use of a Novel Random Mutagenesis Method: In Situ Error-Prone PCR (is-epPCR). *In Vitro Mutagenesis: Methods and Protocols***,** 497-506.

SIMÓN, L. & GOODMAN, J. M. 2009. Enzyme catalysis by hydrogen bonds: The balance between transition state binding and substrate binding in oxyanion holes. *The Journal of Organic Chemistry,* 75**,** 1831-1840.

SINGH, M. K. & MANOJ, N. 2016a. 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.

SINGH, M. K. & MANOJ, N. 2016b. 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.

SINGH, M. K. & 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.

SINGH, M. K., SHIVAKUMARASWAMY, S., GUMMADI, S. N. & 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.

SOCHA, R. D. & TOKURIKI, N. 2013. Modulating protein stability–directed evolution strategies for improved protein function. *FEBS Journal,* 280**,** 5582-5595.

SUK-AM, K., CHENG, K.-J. & JIN-HAO, L. 2002. A variant of Orpinomyces joyonii 1, 3-1, 4-β-glucanase with increased thermal stability obtained by random mutagenesis and screening. *Bioscience, Biotechnology, and Biochemistry,* 66**,** 171-174.

TEALE, F. 1960. The ultraviolet fluorescence of proteins in neutral solution. *Biochemical Journal,* 76**,** 381.

WANG, T.-W., ZHU, H., MA, X.-Y., ZHANG, T., MA, Y.-S. & WEI, D.-Z. 2006. Mutant library construction in directed molecular evolution. *Molecular Biotechnology,* 34**,** 55-68.

WIJMA, H. J., FLOOR, R. J. & JANSSEN, D. B. 2013. Structure-and sequence-analysis inspired engineering of proteins for enhanced thermostability. *Current Opinion in Structural Biology,* 23**,** 588-594.

WILLIAMS, J. C., ZEELEN, J. P., NEUBAUER, G., VRIEND, G., BACKMANN, J., MICHELS, P. A., LAMBEIR, A.-M. & WIERENGA, R. K. 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.

WINTRODE, P. L. & ARNOLD, F. H. 2001. Temperature adaptation of enzymes: lessons from laboratory evolution. *Advances in Protein Chemistry,* 55**,** 161-225.

CHAPTER SIX

6.0 GENERAL CONCLUSIONS AND RECOMMENDATIONS

This thesis reports the successful *in silico* biomining of novel acetyl xylan esterases, as well as the structural and functional characteristics of one novel AcXE (NaM1) from a Namib hypolith metagenome, and its thermostable mutants. The information on AcXE annotation provided here will be useful in improving accurate functional annotation of AcXEs in the future. Previous studies have reported the presence of carbohydrate active enzyme-encoding genes in the Namib desert soil metagenome (Le et al., 2016, Vikram et al., 2016). This study shows that the Namib Desert soil hypolithic communities produce functional carbohydrate-active enzymes. It is recommended that more bioprospecting studies investigate unconventional metagenomes (Adesioye et al., 2016) for functional enzymes with improved industrially-relevant catalytic traits.

NaM1 was fully characterized and confirmed to be a moderately alkalophilic, halophilic and mesophilic AcXE with 7-ACA deacetylase activity. NaM1 was stable in up to 5M NaCl and 60% organic solvents which suggested adaptation to harsh dry conditions typical of the Namib Desert soil environment (Makhalanyane et al., 2015). To facilitate increased solubility of non-polar substrates, many synthetic reactions in industry are carried out in organic solvents. However, enzymes that are potentially stable in these non-aqueous solvents require bioengineering to achieve desired stability (Lind et al., 2004, Kumar et al., 2016). NaM1 is recommended as a candidate enzyme for biocatalysis, as well as bioengineering for improved catalysis, in organic solvents.

The identification of NaM1 and its activity on AX, similar to moderately thermostable *Bp*AXE, provided a basis for proposing a possible inverse relationship between AX activity and thermal stability of CE7 enzymes. Characterization of more CE7 enzymes will be required to confirm this. NaM1 is suitable for deacetylation at mesophilic temperatures which contrasts current industrial lignocellulose bioconversion conditions, but meets the temperature requirements for the deacetylation of cephalosporins during industrial synthesis of β-lactamases.

To further understand the function of NaM1, the crystal structure was solved using X-ray crystallography. This structure provided a basis for comparison with the crystal structures of thermostable CE7 enzymes.

This comparison is useful for understanding thermostability of enzymes within this family. It was observed from comparative structural and sequence analysis that a substitution of valine, in more thermostable CE7 enzymes, to glutamine in NaM1 and several uncharacterized homologs from mesophilic organisms, resulted in a water-mediated hydrogen bond with catalytic Asp275. This suggests that this non-polar to polar substitution may have interfered with hydrophobic networks and thereby, influenced thermal stability. Hence, site-directed mutagenesis, biochemical and/or structural characterization of known CE7 enzymes is recommended for further understanding of this concept within this enzyme family and consequently, other enzymes.

Structural analyses coupled with protein engineering revealed that the Phe210 position in NaM1 is a key player in determining activity, and contributes to thermal stability. These results show that the placement of aromatic, rather than aliphatic residues, at the base of the S2 binding site is preferable for optimal activity of CE7 enzymes. Several previous studies have indicated that residues surrounding the S2 binding site are key players in determining substrate specificity (Montoro-García et al., 2011, Hedge et al., 2012, Singh and Manoj, 2017), but this study shows that a residue (Asn96) relatively distant from the active center and the S2 binding site, directly influences the substrate specificity of NaM1 without activity loss. A N96S substitution converts NaM1 from an esterase with strict specificity for substrates with ≤C4 acyl groups to a larger chain (≤C8 acyl groups) specificity. Most importantly, this substitution resulted in a simultaneous improvement in thermal stability and catalytic efficiency of NaM1 at higher temperatures. This contributes to understanding the factors influencing the thermal stability, as well as, the broad substrate specificity, yet narrow acyl moiety specificity, of CE7 enzymes which is unique among CE families.

6.1 REFERENCES

ADESIOYE, F. A., MAKHALANYANE, T. P., BIELY, P. & COWAN, D. A. 2016. Phylogeny, classification and metagenomic bioprospecting of microbial acetyl xylan esterases. *Enzyme and Microbial Technology,* 93**,** 79-91.

HEDGE, M. K., GEHRING, A. M., ADKINS, C. T., WESTON, L. A., LAVIS, L. D. & JOHNSON, R. J. 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.

KUMAR, A., DHAR, K., KANWAR, S. S. & ARORA, P. K. 2016. Lipase catalysis in organic solvents: advantages and applications. *Biological Procedures Online,* 18**,** 2.

LE, P. T., MAKHALANYANE, T. P., GUERRERO, L., VIKRAM, S., VAN DE PEER, Y. & COWAN, D. A. 2016. Comparative metagenomic analysis reveals mechanisms for stress response in hypoliths from extreme hyperarid deserts. *Genome Biology and Evolution,* 8**,** 2737-2747.

LIND, P. A., DANIEL, R. M., MONK, C. & DUNN, R. V. 2004. Esterase catalysis of substrate vapour: enzyme activity occurs at very low hydration. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics,* 1702**,** 103-110.

MAKHALANYANE, T. P., VALVERDE, A., GUNNIGLE, E., FROSSARD, A., RAMOND, J. B. & COWAN, D. A. 2015. Microbial ecology of hot desert edaphic systems. *FEMS Microbiology Reviews,* 39**,** 203-21.

MONTORO-GARCÍA, S., GIL-ORTIZ, F., GARCÍA-CARMONA, F., POLO, L. M., 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.

SINGH, M. K. & 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.

VIKRAM, S., GUERRERO, L. D., MAKHALANYANE, T. P., LE, P. T., SEELY, M. & COWAN, D. A.

2016. Metagenomic analysis provides insights into functional capacity in a hyperarid desert soil niche community. *Environmental Microbiology,* 18**,** 1875-1888.

APPENDIX 1: Bioinformatics

1-A: HMMER hmmsearch output per CE family (excluding domain annotation for each sequence and alignments).

		Full sequence			Best domain			Domain number estimation	
CAZy family	**target name**	**E-value**	**score**	**bias**	**E-value**	**score**	**bias**	**exp**	**reg**
CE1	gene_id_106908	5.60E-56	194.2	0.4	6.10E-56	194.1	0.4	1	1
	gene_id_41691	4.20E-50	174.9	3.6	5.00E-50	174.7	3.6	1	1
	gene_id_161653	2.30E-43	152.8	0	2.70E-43	152.6	0	1	1
	gene_id_161768	5.60E-41	145	0	7.00E-41	144.7	0	1.1	1
	gene_id_244571	9.80E-37	131	0.3	1.10E-36	130.8	0.3	1	1
	gene_id_125675	2.00E-32	116.9	4.4	2.50E-32	116.6	4.4	1	1
	gene_id_178165	2.30E-31	113.4	0	2.70E-31	113.2	0	1	1
	gene_id_287039	1.30E-30	110.9	2.8	1.50E-30	110.7	2.8	1	1
	gene_id_335983	2.60E-23	87	0.1	4.00E-23	86.3	0.1	1.2	1
	gene_id_47111	9.60E-23	85.1	0.2	1.70E-22	84.2	0.2	1.4	1
	gene_id_184401	1.60E-22	84.3	0.1	6.60E-22	82.3	0.1	1.8	1
	gene_id_223016	4.50E-22	82.9	0.8	9.30E-22	81.9	0.8	1.4	1
	gene_id_286818	4.90E-22	82.8	0.4	4.90E-22	82.8	0.4	1	1
	gene_id_10715	2.90E-21	80.2	2.4	5.40E-20	76.1	2.4	2	1
	gene_id_183906	2.80E-18	70.4	2.7	1.90E-17	67.7	2.7	1.9	1
	gene_id_77632	9.20E-18	68.7	1.7	1.20E-17	68.3	1.7	1.1	1
	gene_id_93799	6.30E-17	66	0	7.60E-17	65.7	0	1	1
	gene_id_127349	3.30E-14	57.1	0	3.60E-14	57	0	1	1
	gene_id_199178	1.80E-12	51.3	1.3	1.40E-11	48.5	1.3	1.9	1
	gene_id_191390	3.60E-11	47.1	0.2	3.90E-11	47	0.2	1	1
	gene_id_261774	1.70E-09	41.5	0	1.90E-09	41.5	0	1	1
	gene_id_381719	4.80E-09	40.1	0	5.10E-09	40	0	1	1
	gene_id_236856	5.80E-09	39.8	0	6.20E-09	39.7	0	1	1
	gene_id_11547	2.50E-08	37.8	0.3	2.90E-08	37.5	0.3	1.1	1
	gene_id_103091	4.10E-08	37.1	0	4.30E-08	37	0	1	1

gene_id_117276	1.20E-07	35.5	0.1	1.40E-07	35.3	0.1	1.1	1
gene_id_260071	5.00E-07	33.5	0	5.40E-07	33.4	0	1	1
gene_id_154730	9.50E-07	32.6	1	0.00018	25.1	0.6	2.9	1
gene_id_208863	9.90E-06	29.2	0.1	1.00E-05	29.2	0.1	1.2	1
gene_id_206888	1.80E-05	28.4	0	2.20E-05	28.1	0	1.1	1
gene_id_118142	3.40E-05	27.5	0	0.00024	24.7	0	1.9	1
gene_id_234778	7.10E-05	26.4	0.1	0.00011	25.9	0.1	1.3	1
gene_id_122721	0.00021	24.8	0	0.00045	23.8	0	1.5	1
gene_id_125440	0.00026	24.6	0	0.00031	24.3	0	1	1
gene_id_114901	0.00029	24.4	0.2	0.00039	24	0.2	1.2	1
gene_id_83979	0.00032	24.2	0	0.00045	23.8	0	1.2	1
gene_id_156575	0.00046	23.8	0	0.0011	22.5	0	1.5	1
gene_id_281311	0.00092	22.8	0	0.0011	22.4	0	1.1	1
gene_id_241922	0.0011	22.5	1.9	0.0016	22	1.4	1.4	2
gene_id_221381	0.0014	22.1	0.1	0.01	19.3	0.2	1.8	1
gene_id_8288	0.002	21.6	0.1	0.0028	21.2	0.1	1.2	1
gene_id_355812	0.0022	21.5	0.2	0.016	18.7	0.1	1.8	1
gene_id_331919	0.0025	21.3	0.1	0.015	18.8	0	1.9	1
gene_id_248292	0.0048	20.4	0	0.0059	20.1	0	1.1	1
gene_id_246821	0.0051	20.3	1.2	0.011	19.2	1.2	1.5	1
gene_id_334765	0.0058	20.1	0	0.0079	19.7	0	1.1	1
gene_id_40894	0.0062	20	0	0.0087	19.6	0	1.1	1
gene_id_166416	0.0072	19.8	0	0.0092	19.5	0	1.1	1
gene_id_64004	0.0077	19.7	0	0.0077	19.7	0	1	1
gene_id_391839	0.0089	19.5	0	0.0095	19.4	0	1	1
gene_id_202146	0.01	19.3	0.2	0.024	18.1	0.2	1.6	1
gene_id_102417	0.011	19.2	0.4	0.015	18.8	0.4	1.2	1
gene_id_255856	0.013	19	0	0.041	17.4	0	1.7	1
gene_id_66864	0.013	18.9	0.3	0.044	17.2	0.1	1.9	1
gene_id_167448	0.014	18.9	0.1	0.28	14.6	0	2.1	1
gene_id_286664	0.014	18.9	0	0.018	18.5	0	1.1	1

gene_id_126575	0.014	18.8	0	0.023	18.2	0	1.3	1
gene_id_194976	0.014	18.8	0.1	0.047	17.1	0	1.7	1
gene_id_101016	0.018	18.5	0.2	0.038	17.5	0.2	1.5	1
gene_id_4793	0.018	18.5	0	0.028	17.9	0	1.2	1
gene_id_74071	0.019	18.4	0	0.027	17.9	0	1.2	1
gene_id_374856	0.023	18.1	0.2	0.025	18.1	0.2	1.1	1
gene_id_98149	0.026	18	0	0.044	17.2	0	1.3	1
gene_id_146688	0.033	17.6	0	0.12	15.8	0	1.8	1
gene_id_79914	0.034	17.6	0	0.073	16.5	0	1.5	1
gene_id_63203	0.035	17.6	0.1	0.11	16	0	1.7	2
gene_id_267114	0.036	17.5	0	0.11	15.9	0	1.7	1
gene_id_326897	0.042	17.3	0.1	0.048	17.1	0.1	1.1	1
gene_id_207921	0.076	16.5	0	0.13	15.7	0	1.3	1
gene_id_381395	0.087	16.3	0	0.1	16	0	1.1	1
gene_id_78268	0.092	16.2	0.1	0.13	15.7	0.1	1.2	1
gene_id_75102	0.095	16.1	0	0.22	14.9	0	1.4	1
gene_id_180580	0.097	16.1	0.1	0.14	15.6	0.1	1.2	1
gene_id_122795	0.1	16	0.2	0.73	13.2	0	2.3	1
gene_id_355750	0.11	15.9	0	0.11	15.9	0	1.2	1
gene_id_179048	0.13	15.7	0	0.14	15.6	0	1.1	1
gene_id_1784	0.15	15.5	0	0.18	15.2	0	1.1	1
gene_id_104717	0.15	15.5	0.1	0.26	14.7	0.1	1.5	1
gene_id_263641	0.17	15.3	0	0.22	15	0	1.2	1
gene_id_7059	0.17	15.3	0	0.31	14.5	0	1.3	1
gene_id_53067	0.18	15.3	0	0.26	14.7	0	1.2	1
gene_id_160059	0.19	15.2	0	0.27	14.7	0	1.2	1
gene_id_46318	0.2	15.1	0.1	0.29	14.5	0.1	1.3	1
gene_id_191474	0.21	15	0.1	0.27	14.7	0.1	1.1	1
gene_id_381718	0.24	14.8	0	0.24	14.8	0	1	1
gene_id_34194	0.24	14.8	0	1.5	12.2	0	1.9	1
gene_id_214604	0.27	14.7	0	0.52	13.7	0	1.4	1

gene_id_103664	0.27	14.7	0.1	0.38	14.2	0.1	1.1	1
gene_id_341913	0.29	14.6	0	0.4	14.1	0	1.1	1
gene_id_232986	0.32	14.4	0.1	0.65	13.4	0.1	1.5	2
gene_id_265499	0.34	14.3	0.1	0.47	13.9	0.1	1.1	1
gene_id_356308	0.36	14.2	0	0.38	14.2	0	1	1
gene_id_260297	0.37	14.2	0	0.45	13.9	0	1.3	1
gene_id_133558	0.4	14.1	0	0.76	13.2	0	1.4	1
gene_id_211565	0.44	14	0.2	0.53	13.7	0.2	1.1	1
gene_id_46693	0.46	13.9	0	0.82	13.1	0	1.3	1
gene_id_394200	0.49	13.8	0	0.56	13.6	0	1.1	1
gene_id_214851	0.5	13.8	0	0.63	13.5	0	1.1	1
gene_id_338458	0.51	13.7	0	0.55	13.6	0	1.1	1
gene_id_156608	0.53	13.7	0	0.64	13.4	0	1.1	1
gene_id_231523	0.55	13.6	0	0.77	13.2	0	1.1	1
gene_id_280341	0.58	13.6	0	0.79	13.1	0	1.1	1
gene_id_111308	0.64	13.4	0.1	1	12.7	0.1	1.3	1
gene_id_301080	0.69	13.3	0	0.91	12.9	0	1.1	1
gene_id_355002	0.74	13.2	0	0.86	13	0	1.1	1
gene_id_44029	0.77	13.2	0	4.3	10.7	0	2	1
gene_id_222330	0.77	13.2	0	1.1	12.6	0	1.2	1
gene_id_208243	0.85	13	0.1	1.2	12.5	0.1	1.2	1
gene_id_14427	0.89	13	0	1.2	12.5	0	1.1	1
gene_id_4613	1	12.8	0	3.5	11	0	1.7	1
gene_id_180903	1	12.8	0	1.3	12.4	0	1.1	1
gene_id_197888	1	12.8	0.4	12	9.3	0.1	2.4	2
gene_id_250332	1.2	12.6	0	10	9.5	0	1.9	2
gene_id_188734	1.2	12.6	0.1	2	11.8	0.1	1.3	1
gene_id_182371	1.2	12.6	0.2	2.3	11.6	0.2	1.6	1
gene_id_282198	1.2	12.6	0.2	5.4	10.4	0.2	1.8	1
gene_id_133782	1.2	12.5	0	1.5	12.2	0	1.3	1
gene_id_157716	1.6	12.1	0.1	2.2	11.7	0.1	1.2	1

	gene_id_219519	1.6	12.1	0	2.2	11.7	0	1.1	1
	gene_id_109684	1.6	12.1	0	2.1	11.7	0	1.1	1
	gene_id_168917	1.7	12	0	2.5	11.5	0	1.2	1
	gene_id_270964	1.7	12	0	2.2	11.7	0	1.1	1
	gene_id_232492	1.7	12	0	2	11.8	0	1.1	1
	gene_id_190937	1.9	11.9	0	2.5	11.5	0	1.1	1
	gene_id_183357	1.9	11.8	0	2	11.8	0	1.1	1
	gene_id_101375	2	11.8	0	3.1	11.2	0	1.2	1
	gene_id_241169	2.1	11.7	0	3.3	11.1	0	1.3	1
	gene_id_152089	2.2	11.6	0.2	2.8	11.3	0.2	1.1	1
	gene_id_242826	2.3	11.6	0	2.3	11.6	0	1	1
	gene_id_151421	2.4	11.5	0.1	3.5	11	0.1	1.2	1
	gene_id_256507	2.5	11.5	0	3.4	11	0	1.1	1
	gene_id_113863	2.9	11.3	0	30	7.9	0	2	2
	gene_id_200521	3	11.2	0	6	10.2	0	1.4	1
	gene_id_347271	3.1	11.2	0	3.6	11	0	1.1	1
	gene_id_309131	3.3	11.1	0.1	7.6	9.9	0.1	1.7	1
	gene_id_27413	3.5	11	0	5.3	10.4	0	1.2	1
	gene_id_240703	3.5	11	0	4.3	10.7	0	1.2	1
	gene_id_3520	3.6	11	0	6.3	10.2	0	1.3	1
	gene_id_317149	3.8	10.9	0	4.6	10.6	0	1.1	1
	gene_id_268567	4.2	10.8	0	5	10.5	0	1.1	1
	gene_id_244500	4.5	10.6	0.1	8.4	9.7	0.1	1.4	2
	gene_id_229062	5.2	10.4	0	7	10	0	1.1	1
	gene_id_271219	5.2	10.4	0.1	7	10	0.1	1.2	1
	gene_id_32024	5.2	10.4	0.6	16	8.9	0.1	2	2
	gene_id_205789	6.8	10.1	0	8.4	9.7	0	1.1	1
	gene_id_135254	10	9.4	2.7	14	9	2.7	1.2	1
	gene_id_262047	13	9.1	2.3	19	8.6	2.3	1.2	1
CE2	gene_id_10250	4.70E-24	89.3	0	8.70E-24	88.4	0	1.4	1

gene_id_205733	3.30E-08	37.1	0	6.40E-08	36.2	0	1.4	1
gene_id_201281	2.40E-05	27.7	0	0.2	14.8	0	2	2
gene_id_266046	4.50E-05	26.8	0	4.80E-05	26.7	0	1.1	1
gene_id_20333	5.10E-05	26.6	3.2	0.015	18.5	0.7	2.9	2
gene_id_67464	9.30E-05	25.8	0.1	0.029	17.6	0	2.1	2
gene_id_127814	0.00026	24.3	0.2	0.00036	23.8	0.2	1.1	1
gene_id_281834	0.0021	21.3	0.1	0.0031	20.7	0.1	1.3	1
gene_id_212578	0.0021	21.3	0.1	0.0043	20.3	0.1	1.5	1
gene_id_256609	0.0035	20.6	0	0.0043	20.3	0	1.1	1
gene_id_2893	0.0046	20.2	0	4	10.5	0	2.9	2
gene_id_41367	0.0084	19.3	0.2	23	8	0	2.9	2
gene_id_306342	0.013	18.7	0	0.099	15.8	0	1.9	2
gene_id_198378	0.028	17.6	0	4.3	10.4	0	2.1	2
gene_id_199640	0.041	17.1	0	0.041	17.1	0	1.1	1
gene_id_184141	0.058	16.6	0	0.067	16.4	0	1	1
gene_id_311486	0.061	16.5	0	0.073	16.3	0	1.1	1
gene_id_336696	0.09	15.9	0.1	0.11	15.6	0.1	1.1	1
gene_id_226106	0.16	15.1	0	16	8.6	0	2.2	2
gene_id_143708	0.28	14.3	0	0.44	13.7	0	1.3	1
gene_id_344541	0.29	14.3	0.1	0.32	14.1	0.1	1.1	1
gene_id_94295	0.37	13.9	0	3.5	10.7	0	2.1	2
gene_id_101285	0.38	13.9	0	54	6.8	0	2.2	2
gene_id_215440	0.41	13.8	0	0.45	13.7	0	1.1	1
gene_id_78951	0.45	13.6	0	0.51	13.5	0	1	1
gene_id_54585	0.54	13.4	0	0.86	12.7	0	1.3	1
gene_id_321781	0.59	13.3	0.2	0.78	12.9	0.2	1.2	1
gene_id_219190	0.63	13.2	0	0.88	12.7	0	1.2	1
gene_id_205701	0.66	13.1	0	0.75	12.9	0	1.1	1
gene_id_217819	0.8	12.8	0	11	9.1	0	1.9	2
gene_id_238860	0.93	12.6	0.5	1	12.4	0.5	1	1
gene_id_71373	1	12.5	0	6.2	9.9	0	1.9	2

	gene_id_193959	1.1	12.4	0.9	2.1	11.4	0.9	1.5	1
	gene_id_149157	1.3	12.1	0	1.3	12.1	0	1.1	1
	gene_id_162686	1.3	12.1	0	80	6.2	0	2.1	2
	gene_id_260370	1.6	11.9	0	55	6.8	0	2.3	2
	gene_id_276151	2	11.5	0	2.4	11.3	0	1	1
	gene_id_302823	2	11.5	0	2	11.5	0	1	1
	gene_id_253989	2.3	11.3	0	2.6	11.1	0	1.1	1
	gene_id_374407	2.6	11.1	0.1	3.6	10.7	0.1	1.2	1
	gene_id_253839	2.7	11.1	0	91	6.1	0	2.1	2
	gene_id_65101	3	10.9	0	3.70E+02	4.1	0	3	3
	gene_id_247300	3.4	10.8	0.1	9.6	9.3	0	1.7	1
	gene_id_109067	3.4	10.7	0	3.4	10.7	0	1.1	1
	gene_id_270137	3.6	10.7	0	3.6	10.7	0	1	1
	gene_id_227106	6.8	9.8	0	6.8	9.8	0	1	1
CE3									
	gene_id_94295	1.60E-83	285.3	2.5	1.90E-83	285.1	2.5	1	1
	gene_id_127814	1.30E-66	229.7	0.1	1.50E-66	229.5	0.1	1	1
	gene_id_321781	2.40E-57	199.2	0.3	2.60E-57	199.1	0.3	1	1
	gene_id_306342	7.10E-54	187.8	0	8.20E-54	187.6	0	1	1
	gene_id_108306	2.30E-48	169.7	0.4	2.80E-48	169.4	0.4	1.1	1
	gene_id_287410	3.10E-48	169.2	4.4	3.60E-48	169	4.4	1.1	1
	gene_id_253839	9.30E-45	157.8	1	9.30E-45	157.8	1	1.4	1
	gene_id_173248	7.30E-44	154.9	0	9.10E-44	154.6	0	1	1
	gene_id_198378	1.70E-40	143.8	0.1	1.90E-40	143.6	0.1	1	1
	gene_id_212578	1.50E-34	124.2	0.2	1.80E-34	124	0.2	1	1
	gene_id_183094	1.40E-27	101.3	0.3	1.60E-27	101.1	0.3	1	1
	gene_id_140100	1.30E-20	78.4	0.1	1.60E-20	78.1	0.1	1.2	1
	gene_id_336801	1.70E-20	78	0	1.90E-20	77.9	0	1	1
	gene_id_198379	3.80E-15	60.4	0	4.10E-15	60.3	0	1	1
	gene_id_20333	2.80E-14	57.6	0.4	3.30E-11	47.5	0	3	2
	gene_id_359564	2.40E-12	51.2	0	2.50E-12	51.2	0	1	1

gene_id_332089	1.10E-11	49.1	0.2	2.20E-11	48.1	0.1	1.5	1
gene_id_131955	9.00E-11	46	0.2	1.50E-10	45.3	0.2	1.3	1
gene_id_54585	5.30E-10	43.5	0	8.00E-10	42.9	0	1.2	1
gene_id_336696	7.40E-09	39.7	0	1.10E-08	39.2	0	1.2	1
gene_id_281834	7.80E-09	39.7	0.1	7.80E-09	39.7	0.1	1.5	1
gene_id_205733	8.70E-09	39.5	0	1.20E-07	35.8	0	2.3	2
gene_id_10746	1.20E-08	39	0.9	4.50E-07	33.9	0.5	2	2
gene_id_171823	2.80E-08	37.8	0	3.60E-08	37.5	0	1.1	1
gene_id_247300	1.00E-07	36	0.2	3.30E-06	31	0.1	2.3	2
gene_id_156919	3.20E-07	34.4	0	3.30E-07	34.3	0	1	1
gene_id_65101	4.30E-07	33.9	0	0.0009	23	0	2.4	2
gene_id_256609	8.10E-07	33	0	9.40E-07	32.8	0	1.1	1
gene_id_41367	2.10E-06	31.7	0.8	0.00014	25.7	0.8	2.6	1
gene_id_201281	2.20E-06	31.6	0	0.00018	25.3	0	2	1
gene_id_10250	2.30E-06	31.5	0.1	8.20E-05	26.4	0.1	2.6	2
gene_id_143708	4.00E-06	30.8	0	5.50E-06	30.3	0	1.2	1
gene_id_246118	4.80E-06	30.5	0	5.70E-06	30.3	0	1.1	1
gene_id_344541	8.70E-06	29.6	0	9.70E-06	29.5	0	1	1
gene_id_270703	1.60E-05	28.8	0.5	2.20E-05	28.3	0.5	1.2	1
gene_id_67464	6.00E-05	26.9	0.5	0.43	14.2	0.1	3	2
gene_id_355563	8.00E-05	26.5	0	0.005	20.6	0	2	1
gene_id_181231	0.00019	25.3	6.7	0.00023	25	6.7	1.3	1
gene_id_8815	0.0002	25.1	0	0.00035	24.4	0	1.3	1
gene_id_355741	0.00026	24.8	0	0.00026	24.8	0	1	1
gene_id_384164	0.00027	24.7	0	0.085	16.5	0	2.1	2
gene_id_2893	0.00028	24.7	0.1	0.078	16.6	0	2.7	2
gene_id_266046	0.00042	24.1	0	0.00052	23.8	0	1.2	1
gene_id_83169	0.00075	23.3	0.5	0.002	21.9	0.5	1.6	1
gene_id_64061	0.0008	23.2	0	0.0013	22.5	0	1.3	1
gene_id_154141	0.00083	23.1	0	0.00092	23	0	1	1
gene_id_160418	0.0009	23	0.2	0.0011	22.7	0.2	1.1	1

gene_id_217819	0.00098	22.9	0	0.0022	21.8	0	1.5	1
gene_id_226106	0.0015	22.3	0	0.014	19.1	0	1.9	1
gene_id_313059	0.0016	22.2	1.7	0.0016	22.2	1.7	1.6	1
gene_id_147314	0.0028	21.4	0.1	0.0071	20.1	0.1	1.5	1
gene_id_130575	0.0032	21.2	8.9	0.031	18	0.4	2.4	3
gene_id_355740	0.0039	20.9	0	0.0042	20.8	0	1	1
gene_id_376398	0.004	20.9	0	0.0045	20.7	0	1	1
gene_id_205701	0.0045	20.7	0	0.0054	20.5	0	1.2	1
gene_id_101285	0.0046	20.7	0.4	3.1	11.4	0.1	2.2	2
gene_id_248812	0.0063	20.3	0	0.19	15.4	0	2.1	2
gene_id_69045	0.0067	20.2	9.1	0.0086	19.8	8.8	1.4	1
gene_id_21267	0.007	20.1	0	67	7	0	3	3
gene_id_41102	0.0073	20	0.2	0.1	16.3	0	2.5	1
gene_id_9548	0.0077	20	0	0.012	19.3	0	1.2	1
gene_id_71373	0.011	19.5	0	0.086	16.5	0	1.9	1
gene_id_35825	0.012	19.4	0.1	0.077	16.7	0	2.2	1
gene_id_6615	0.012	19.4	0	0.93	13.1	0	2.2	2
gene_id_374407	0.012	19.3	0	0.013	19.2	0	1.2	1
gene_id_239835	0.013	19.2	2.6	0.022	18.5	2.6	1.4	1
gene_id_156823	0.017	18.8	0	0.082	16.6	0	2	1
gene_id_97312	0.017	18.8	2.6	0.017	18.8	2.6	3	3
gene_id_69913	0.02	18.6	1.3	0.043	17.5	0.5	1.7	2
gene_id_114524	0.026	18.2	0	0.68	13.6	0	2	2
gene_id_268129	0.027	18.2	5.5	0.027	18.2	5.5	2.1	1
gene_id_92719	0.028	18.1	0	0.031	18	0	1	1
gene_id_241156	0.034	17.8	3	0.034	17.8	3	2.1	2
gene_id_396352	0.036	17.7	0	0.048	17.4	0	1.2	1
gene_id_243131	0.042	17.5	2.9	0.051	17.3	2.9	1.2	1
gene_id_300176	0.08	16.6	0	0.081	16.6	0	1.1	1
gene_id_218895	0.1	16.3	0	0.1	16.3	0	1.1	1
gene_id_4592	0.12	16.1	0	0.17	15.5	0	1.1	1

gene_id_260370	0.12	16.1	0	4.1	11	0	2	1
gene_id_194409	0.13	15.9	0	5.1	10.7	0	2.1	2
gene_id_30238	0.14	15.8	0	0.59	13.8	0	2.1	1
gene_id_288436	0.14	15.8	0	0.14	15.8	0	1	1
gene_id_376397	0.15	15.8	0	0.16	15.6	0	1.1	1
gene_id_37702	0.16	15.6	0	8.4	10	0	2.1	2
gene_id_300306	0.18	15.4	0.1	52	7.4	0	2.4	2
gene_id_139056	0.18	15.4	0.1	0.19	15.4	0.1	1	1
gene_id_149157	0.18	15.4	0	0.18	15.4	0	1.1	1
gene_id_308567	0.19	15.3	0.1	3.6	11.2	0	2.1	2
gene_id_116634	0.2	15.3	5	0.2	15.3	5	2.5	2
gene_id_9691	0.23	15.1	0.2	4.9	10.7	0	2	2
gene_id_43848	0.24	15	0	2.1	11.9	0	1.9	2
gene_id_349951	0.25	15	0	0.25	15	0	1	1
gene_id_265563	0.25	15	0	0.39	14.3	0	1.3	1
gene_id_56118	0.26	14.9	1.3	0.3	14.7	1.3	1	1
gene_id_162686	0.27	14.9	0	1.3	12.6	0	2	1
gene_id_167871	0.29	14.8	0	1.5	12.4	0	1.9	1
gene_id_167081	0.29	14.8	0.2	6	10.4	0.1	2.2	2
gene_id_394572	0.3	14.7	0	0.31	14.7	0	1	1
gene_id_61908	0.34	14.6	0	1.1	12.9	0	1.7	1
gene_id_183206	0.42	14.3	0	0.71	13.5	0	1.3	1
gene_id_62710	0.44	14.2	0	4	11	0	2.3	1
gene_id_135925	0.47	14.1	0.2	35	7.9	0	2.1	2
gene_id_201282	0.49	14	0	0.49	14	0	1	1
gene_id_233723	0.49	14	0	0.66	13.6	0	1.4	1
gene_id_385251	0.52	13.9	0	0.55	13.9	0	1	1
gene_id_334486	0.54	13.9	0	0.73	13.5	0	1.2	1
gene_id_136054	0.55	13.9	0	1.3	12.6	0	1.6	1
gene_id_114418	0.55	13.8	0	15	9.1	0	2.1	2
gene_id_13386	0.66	13.6	0	0.72	13.5	0	1.1	1

	gene_id_109067	0.67	13.6	0	0.69	13.5	0	1.1	1
	gene_id_359323	0.79	13.3	0.2	31	8.1	0	1.9	2
	gene_id_188176	0.8	13.3	0.1	0.88	13.2	0.1	1.1	1
	gene_id_67564	0.9	13.2	0	2.4	11.8	0	1.6	1
	gene_id_99584	0.93	13.1	0	1.4	12.6	0	1.3	1
	gene_id_71167	0.96	13.1	0	1.2	12.8	0	1.1	1
	gene_id_209063	1.1	12.9	0	1.1	12.9	0	1	1
	gene_id_110962	1.3	12.6	0.2	1.8	12.2	0.1	1.3	1
	gene_id_233069	1.4	12.5	0	1.7	12.2	0	1.1	1
	gene_id_257046	1.5	12.4	0	1.10E+02	6.4	0	2.1	2
	gene_id_227964	1.5	12.4	0	1.7	12.2	0	1	1
	gene_id_133333	1.6	12.3	0	2.1	12	0	1.1	1
	gene_id_259060	1.7	12.3	0	1.9	12.1	0	1	1
	gene_id_351222	1.8	12.2	0	2.3	11.8	0	1.1	1
	gene_id_386454	2	12	0	2	12	0	1	1
	gene_id_43221	2	12	0	2.5	11.7	0	1.1	1
	gene_id_227106	2.1	11.9	0	2.1	11.9	0	1	1
	gene_id_266287	2.1	11.9	0	27	8.3	0	2	2
	gene_id_349017	2.3	11.8	0	2.3	11.8	0	1.1	1
	gene_id_313937	2.5	11.7	0.1	3.4	11.3	0.1	1.2	1
	gene_id_179155	2.5	11.7	0.1	6.7	10.3	0.1	1.7	1
	gene_id_242265	2.8	11.6	0	29	8.2	0	2	2
	gene_id_156447	3.5	11.2	0	3.5	11.2	0	1	1
	gene_id_190775	3.7	11.2	0	3.7	11.2	0	1	1
	gene_id_360214	4.2	11	0	4.2	11	0	1	1
	gene_id_392065	4.2	10.9	0	73	6.9	0	2.1	2
	gene_id_270137	5	10.7	0.3	7.1	10.2	0.2	1.3	1
	gene_id_56371	5	10.7	0	5.4	10.6	0	1.1	1
	gene_id_224901	6	10.5	0	47	7.5	0	2	2
CE4	gene_id_56239	3.90E-75	257.6	14.9	4.40E-75	257.5	14.9	1	1

gene_id_9096	2.50E-67	231.9	10.2	2.80E-67	231.8	10.2	1	1
gene_id_52973	1.00E-25	94.9	1	1.20E-25	94.7	1	1	1
gene_id_251385	2.60E-24	90.3	0.8	2.90E-24	90.1	0.8	1	1
gene_id_256124	1.00E-12	52.1	0.1	1.50E-07	35.1	0	2.1	2
gene_id_7346	4.20E-10	43.5	0	4.10E-09	40.2	0	2.1	1
gene_id_167528	1.50E-07	35.1	0	4.50E-06	30.2	0	2.2	2
gene_id_189043	3.90E-07	33.7	0	5.30E-07	33.3	0	1.2	1
gene_id_7936	5.60E-06	29.9	0	0.0055	20	0	2.2	2
gene_id_58591	2.40E-05	27.8	0	0.031	17.6	0	1.9	2
gene_id_2159	3.90E-05	27.1	0	5.30E-05	26.7	0	1.1	1
gene_id_162360	0.00027	24.4	0	0.0013	22.1	0	2	1
gene_id_292698	0.0009	22.6	0.2	0.011	19.1	0.2	2	1
gene_id_297056	0.0022	21.3	0	0.0029	21	0	1.1	1
gene_id_292882	0.0025	21.2	0	0.0026	21.1	0	1	1
gene_id_226393	0.0044	20.4	0	0.0047	20.3	0	1.1	1
gene_id_139665	0.0048	20.2	0	0.0055	20.1	0	1	1
gene_id_254685	0.012	19	0	0.018	18.3	0	1.2	1
gene_id_160537	0.017	18.5	0	0.023	18	0	1.2	1
gene_id_115406	0.024	17.9	0	0.042	17.1	0	1.3	1
gene_id_162482	0.043	17.1	0	0.45	13.7	0	1.9	2
gene_id_336245	0.084	16.1	0	0.18	15	0	1.5	1
gene_id_165702	0.11	15.7	0	0.11	15.7	0	1	1
gene_id_26235	0.17	15.1	0	5.8	10.1	0	2	2
gene_id_57578	0.2	14.9	0	0.24	14.7	0	1.1	1
gene_id_227520	0.24	14.7	0	0.3	14.3	0	1.1	1
gene_id_263733	0.36	14.1	0	0.4	13.9	0	1	1
gene_id_93710	0.42	13.8	0.2	0.44	13.8	0.2	1	1
gene_id_290682	0.48	13.6	0	0.58	13.4	0	1.1	1
gene_id_275849	1.1	12.5	0.3	1.3	12.2	0.3	1.2	1
gene_id_165541	1.1	12.4	0	1.5	12.1	0	1.1	1
gene_id_311556	1.3	12.3	0	1.6	11.9	0	1.1	1

	gene_id_166750	1.3	12.2	0	1.3	12.2	0	1.2	1
	gene_id_12713	1.4	12.1	0	1.8	11.7	0	1.3	1
	gene_id_73060	1.5	12	5.9	8.6	9.5	5.9	1.8	1
	gene_id_283283	1.7	11.8	0	1.8	11.8	0	1.2	1
	gene_id_332410	1.7	11.8	0.3	2.2	11.5	0.3	1	1
	gene_id_32838	2.2	11.5	0	3	11	0	1.1	1
	gene_id_182844	2.3	11.4	4.1	2.6	11.2	4.1	1	1
	gene_id_269785	2.4	11.3	0.1	2.5	11.3	0.1	1	1
	gene_id_193496	2.5	11.3	3	3.3	10.9	3	1.2	1
	gene_id_143842	2.7	11.2	0	2.8	11.1	0	1	1
	gene_id_175850	5.1	10.3	0.3	5.4	10.2	0.3	1	1
	gene_id_331743	5.6	10.1	3.7	5.8	10.1	3.7	1	1
	gene_id_243894	6.3	10	2.2	6.7	9.9	2.2	1	1
	gene_id_362631	12	9	1.1	13	8.9	1.1	1.1	1
	gene_id_77034	52	6.9	5.8	56	6.8	5.8	1	1
	gene_id_143965	71	6.5	6.7	2.40E+02	4.8	1.8	2.1	2
CE5									
	gene_id_293497	7.20E-10	43.2	0.6	3.30E-09	41.1	0.4	1.7	2
	gene_id_121729	0.22	15.5	0	0.29	15.1	0	1.2	1
	gene_id_66030	3.2	11.6	0	3.5	11.5	0	1.2	1
	gene_id_323149	7.8	10.4	3.3	47	7.8	2.6	1.9	1
CE6									
	gene_id_185715	2.80E-05	27.5	0.1	3.00E-05	27.4	0.1	1	1
	gene_id_125055	8.00E-05	26	0.2	0.79	12.8	0.1	2	2
	gene_id_96208	0.002	21.4	0	4.5	10.3	0	2	2
	gene_id_236825	0.17	15	0	0.22	14.7	0	1.1	1
	gene_id_282516	0.38	13.9	0	0.44	13.7	0	1	1
	gene_id_219519	3.2	10.8	0	3.2	10.8	0	1.1	1
	gene_id_209095	4.8	10.2	0	5	10.2	0	1	1
	gene_id_295217	4.8	10.2	0	6	9.9	0	1	1

CE7

gene_id_39518	3.10E-113	382.4	0	3.70E-113	382.1	0	1	1
gene_id_27345	3.10E-95	323.2	0	3.90E-95	322.9	0	1	1
gene_id_19439	1.50E-78	268.4	0	1.70E-78	268.2	0	1	1
gene_id_294218	1.30E-65	225.9	0	1.40E-65	225.7	0	1	1
gene_id_48524	2.80E-61	211.7	0.1	3.10E-61	211.5	0.1	1	1
gene_id_38890	6.70E-60	207.1	0	7.60E-60	206.9	0	1	1
gene_id_143644	1.40E-55	192.9	0	1.80E-55	192.6	0	1	1
gene_id_216797	2.70E-55	192	0	3.00E-55	191.8	0	1	1
gene_id_106421	5.90E-53	184.3	0	6.40E-53	184.2	0	1	1
gene_id_215763	1.10E-50	176.9	0	1.20E-50	176.7	0	1	1
gene_id_161396	6.10E-47	164.5	0	7.50E-47	164.2	0	1	1
gene_id_12208	5.90E-45	158	0	6.30E-45	157.9	0	1	1
gene_id_291050	1.30E-27	101	0	1.40E-27	100.9	0	1	1
gene_id_277540	2.80E-25	93.4	0	4.20E-25	92.8	0	1.3	1
gene_id_291049	1.70E-24	90.8	0	1.70E-24	90.8	0	1	1
gene_id_58620	1.20E-17	68.2	0	1.40E-17	68	0	1	1
gene_id_216820	1.90E-17	67.6	0	1.90E-17	67.6	0	1	1
gene_id_60892	1.30E-16	64.9	0	1.50E-16	64.7	0	1	1
gene_id_123556	1.90E-14	57.8	0	2.70E-14	57.3	0	1.1	1
gene_id_172765	6.10E-13	52.8	0.1	1.80E-10	44.7	0.1	2	1
gene_id_40982	1.70E-11	48	0	2.00E-11	47.8	0	1	1
gene_id_40964	1.90E-11	47.9	0	2.20E-11	47.7	0	1	1
gene_id_166871	2.70E-11	47.4	0.1	5.60E-05	26.7	0	2	2
gene_id_50998	4.00E-10	43.6	0	2.70E-07	34.2	0	2	1
gene_id_254979	7.90E-09	39.3	0	5.20E-06	30	0	2.1	1
gene_id_394537	1.10E-08	38.8	0	1.30E-08	38.6	0	1	1
gene_id_277238	2.30E-08	37.8	0.5	1.50E-06	31.8	0.5	2.2	1
gene_id_112654	3.60E-08	37.2	0.1	0.012	19.1	0	2	2
gene_id_145052	1.10E-07	35.5	0.1	0.00097	22.6	0	2.2	2
gene_id_312897	1.00E-06	32.4	0.1	1.20E-06	32.2	0.1	1	1
gene_id_76293	2.70E-06	31	0.1	0.00027	24.4	0.1	2.1	1

gene_id_295829	1.20E-05	28.9	0	0.00016	25.2	0	2.5	1
gene_id_251008	1.20E-05	28.9	0.3	0.00049	23.6	0.3	2	1
gene_id_192579	2.10E-05	28	0.4	0.0058	20	0.1	2	2
gene_id_341913	2.60E-05	27.7	0	0.004	20.6	0	2.1	1
gene_id_292418	3.00E-05	27.5	0.2	0.86	12.9	0	2.4	2
gene_id_159172	3.30E-05	27.4	0.1	0.00091	22.7	0.1	2.1	1
gene_id_122475	3.90E-05	27.2	0.2	0.0074	19.7	0.1	2	2
gene_id_322295	5.00E-05	26.8	0.1	0.22	14.8	0.1	2	2
gene_id_18246	5.10E-05	26.8	0.1	0.00011	25.8	0	1.5	1
gene_id_181158	5.30E-05	26.7	0	9.60E-05	25.9	0	1.4	1
gene_id_150212	6.60E-05	26.4	0	0.18	15.1	0	2.7	2
gene_id_208243	7.90E-05	26.2	0.7	0.008	19.6	0.3	2.1	2
gene_id_49820	0.0001	25.8	0.1	0.41	14	0	2	2
gene_id_99144	0.00014	25.4	0.1	0.00099	22.6	0	1.9	2
gene_id_140897	0.00015	25.2	0.1	0.0028	21.1	0.1	2.1	1
gene_id_363640	0.00018	25	0	0.00019	24.9	0	1	1
gene_id_59866	0.00021	24.8	0	1.2	12.4	0	2.1	2
gene_id_77018	0.00023	24.6	0	0.0012	22.3	0.1	1.8	2
gene_id_25013	0.00024	24.6	0	0.00065	23.2	0.1	1.6	1
gene_id_122413	0.00032	24.2	0	0.0012	22.3	0	1.7	1
gene_id_381395	0.00033	24.1	0.1	0.045	17.1	0.1	2	1
gene_id_96679	0.00033	24.1	0.1	0.06	16.7	0.1	2.4	2
gene_id_113863	0.00043	23.8	0	0.053	16.9	0	2.1	2
gene_id_6802	0.00043	23.7	0	0.00053	23.5	0	1.1	1
gene_id_206192	0.00052	23.5	0.8	0.21	14.9	0.2	2	2
gene_id_279036	0.00081	22.9	0.1	0.093	16.1	0	2	2
gene_id_143254	0.00096	22.6	0.1	1.1	12.6	0	2.6	2
gene_id_338458	0.001	22.5	0.1	0.24	14.7	0.1	2.1	1
gene_id_329536	0.001	22.5	0	0.0012	22.3	0	1.2	1
gene_id_34325	0.0011	22.4	0.3	0.61	13.4	0	2.8	3
gene_id_250746	0.0012	22.3	0.2	0.045	17.1	0	2	2

gene_id_156575	0.0012	22.2	0.5	0.16	15.3	0.2	2.1	2
gene_id_270186	0.0013	22.2	0	0.0013	22.1	0	1	1
gene_id_199048	0.0014	22.1	0.3	0.002	21.6	0.3	1.2	1
gene_id_23442	0.0015	22	0	0.0049	20.3	0	1.7	2
gene_id_175431	0.0016	21.9	0	0.0098	19.3	0	1.9	2
gene_id_221886	0.0016	21.9	0.1	0.007	19.8	0.1	1.8	1
gene_id_260297	0.0017	21.8	0.5	0.29	14.4	0.9	2.5	1
gene_id_111269	0.0019	21.6	0.1	2.1	11.6	0	2.7	2
gene_id_224968	0.0021	21.5	0.3	0.0031	20.9	0.3	1.1	1
gene_id_124594	0.0022	21.4	0.2	3.4	10.9	0	2.1	2
gene_id_156964	0.0024	21.3	0	0.52	13.6	0	2.2	1
gene_id_200471	0.0025	21.3	0.1	0.068	16.5	0	2	2
gene_id_163671	0.0029	21	0.8	0.0035	20.8	0	1.4	1
gene_id_360067	0.0033	20.8	0	0.0034	20.8	0	1	1
gene_id_394548	0.0034	20.8	0.4	0.2	15	0	2.1	2
gene_id_83979	0.0034	20.8	1.4	1.7	11.9	0	2.9	2
gene_id_57428	0.0036	20.7	0.1	0.011	19.1	0.1	1.7	1
gene_id_309131	0.004	20.6	0.1	0.91	12.8	0.1	2	1
gene_id_216477	0.0041	20.5	0	0.018	18.5	0	1.7	1
gene_id_347562	0.0045	20.4	0.2	5.7	10.2	0	2.8	3
gene_id_44796	0.005	20.2	0	0.005	20.2	0	1.1	1
gene_id_118405	0.0053	20.2	0.7	3.6	10.8	0.1	2	2
gene_id_268567	0.0057	20.1	0.1	1.4	12.2	0	2	2
gene_id_43977	0.0077	19.6	0.1	0.46	13.8	0.1	2.2	2
gene_id_51928	0.0086	19.5	0.1	3	11.1	0.1	2.1	1
gene_id_168374	0.0086	19.5	0	0.012	19	0	1.1	1
gene_id_161501	0.0091	19.4	0.1	2	11.7	0	2.1	2
gene_id_152422	0.0094	19.3	0.1	0.032	17.6	0	1.6	1
gene_id_368499	0.0098	19.3	0	0.011	19.2	0	1.1	1
gene_id_265499	0.011	19.1	0.4	1	12.6	0.4	2	1
gene_id_218180	0.011	19.1	0	0.89	12.9	0	2.1	1

gene_id_64793	0.012	19.1	0	0.07	16.5	0	2	1
gene_id_75102	0.012	19	0	0.055	16.8	0	1.9	1
gene_id_69987	0.012	19	0	0.051	16.9	0	1.8	1
gene_id_317315	0.015	18.7	0.3	0.13	15.6	0.3	1.9	1
gene_id_133558	0.017	18.5	0.2	0.021	18.2	0.2	1.1	1
gene_id_122795	0.018	18.4	0	1	12.6	0	2.1	1
gene_id_334765	0.018	18.4	0	0.021	18.2	0	1.1	1
gene_id_161395	0.02	18.3	0	0.022	18.1	0	1	1
gene_id_42860	0.021	18.2	0	0.027	17.8	0	1.1	1
gene_id_42398	0.024	18	0	0.032	17.6	0	1.2	1
gene_id_62690	0.026	17.9	0	0.062	16.7	0	1.5	1
gene_id_65452	0.028	17.8	0.1	10	9.4	0	2	2
gene_id_313059	0.03	17.7	0.1	0.063	16.6	0.1	1.4	1
gene_id_69078	0.032	17.6	0	0.035	17.5	0	1.1	1
gene_id_329314	0.033	17.6	0.3	0.067	16.5	0.2	1.4	1
gene_id_158558	0.037	17.4	0	12	9.1	0	2	2
gene_id_255170	0.037	17.4	0.1	1	12.6	0	2	2
gene_id_259671	0.042	17.2	0.1	0.17	15.2	0.1	1.8	1
gene_id_86332	0.043	17.2	0.1	0.42	13.9	0.1	1.9	1
gene_id_95482	0.044	17.1	0	0.053	16.9	0	1.1	1
gene_id_333220	0.044	17.1	0	0.072	16.4	0	1.3	1
gene_id_30443	0.046	17.1	0.5	0.06	16.7	0.5	1.1	1
gene_id_343157	0.047	17.1	0	0.06	16.7	0	1.1	1
gene_id_119746	0.047	17	0	6.6	10	0	2.1	2
gene_id_42563	0.048	17	0	0.068	16.5	0	1.2	1
gene_id_39162	0.05	17	0.1	0.074	16.4	0	1.3	1
gene_id_36422	0.055	16.8	0	0.14	15.5	0	1.6	1
gene_id_25913	0.057	16.8	0	0.33	14.3	0	1.8	1
gene_id_167761	0.057	16.8	0.2	2.4	11.5	0	2	2
gene_id_2287	0.058	16.7	0.1	0.14	15.5	0.1	1.7	1
gene_id_120533	0.059	16.7	0.5	0.99	12.7	0.2	2	2

gene_id_39215	0.059	16.7	0.1	1.3	12.3	0	2.1	2
gene_id_125552	0.063	16.6	0	0.077	16.3	0	1.3	1
gene_id_250245	0.068	16.5	0.9	0.35	14.2	0.2	2.1	2
gene_id_134610	0.072	16.4	0	0.82	13	0	2	2
gene_id_187210	0.078	16.3	0.1	0.18	15.1	0	1.6	1
gene_id_104504	0.08	16.3	0.1	0.11	15.8	0.1	1.2	1
gene_id_258001	0.082	16.3	0	0.2	15	0	1.6	1
gene_id_178087	0.088	16.1	0	0.1	15.9	0	1.1	1
gene_id_133761	0.089	16.1	0	0.23	14.8	0	1.7	1
gene_id_72815	0.093	16.1	0	0.61	13.4	0	1.9	1
gene_id_255481	0.093	16.1	0	0.13	15.6	0	1.1	1
gene_id_16999	0.097	16	0	1.4	12.2	0	2	2
gene_id_274108	0.097	16	0.1	0.63	13.3	0	1.9	2
gene_id_289928	0.1	16	0.1	26	8	0.1	2.2	2
gene_id_237140	0.1	16	0.2	0.13	15.6	0.2	1.1	1
gene_id_160889	0.1	15.9	0.3	5.8	10.2	0	2.1	2
gene_id_103664	0.1	15.9	0	0.12	15.7	0	1.1	1
gene_id_79180	0.1	15.9	0.3	0.11	15.9	0.3	1	1
gene_id_173259	0.11	15.9	0.1	3.6	10.8	0	2	2
gene_id_39834	0.11	15.8	0	45	7.3	0	2.1	2
gene_id_124049	0.11	15.8	0	0.16	15.3	0	1.1	1
gene_id_265470	0.12	15.7	0	2.9	11.2	0	2	2
gene_id_249942	0.12	15.7	0	0.18	15.1	0	1.3	1
gene_id_116020	0.12	15.7	0.1	29	7.9	0.1	2	2
gene_id_323989	0.13	15.6	0.1	12	9.1	0.1	2.1	2
gene_id_207921	0.15	15.4	0	0.55	13.6	0	1.8	1
gene_id_166542	0.15	15.4	0	0.68	13.2	0	1.8	2
gene_id_381719	0.15	15.4	0.1	0.15	15.4	0.1	1	1
gene_id_49358	0.15	15.4	0.2	0.94	12.8	0	1.8	2
gene_id_259188	0.16	15.3	0	0.27	14.6	0	1.3	1
gene_id_292220	0.16	15.3	0	0.17	15.2	0	1	1

gene_id_367347	0.17	15.2	0	0.24	14.7	0	1.3	1
gene_id_374878	0.17	15.2	0	0.18	15.2	0	1.1	1
gene_id_231523	0.18	15.1	0.1	1.6	12	0	2	2
gene_id_270964	0.18	15.1	0	3.6	10.9	0	2.1	1
gene_id_392163	0.18	15.1	0	0.2	15	0	1	1
gene_id_325380	0.19	15.1	0.1	7.9	9.7	0.1	2.1	1
gene_id_372121	0.2	15	0.2	0.28	14.5	0.2	1.1	1
gene_id_317795	0.2	15	0	0.2	15	0	1	1
gene_id_3435	0.21	14.9	0	18	8.6	0	2.2	2
gene_id_293464	0.22	14.8	0.2	30	7.8	0	2.1	2
gene_id_154327	0.23	14.8	0	0.31	14.4	0	1.1	1
gene_id_41602	0.23	14.8	0.1	2	11.7	0	1.9	2
gene_id_92687	0.24	14.7	0	30	7.9	0	2.3	2
gene_id_318831	0.24	14.7	0.2	0.28	14.5	0.2	1.1	1
gene_id_86081	0.24	14.7	0	7.2	9.9	0	2.2	2
gene_id_320245	0.25	14.7	0.1	0.36	14.2	0.1	1.1	1
gene_id_287586	0.25	14.6	0.2	0.31	14.3	0.2	1	1
gene_id_232492	0.26	14.6	0	7.4	9.8	0	2	2
gene_id_342306	0.26	14.6	0	0.3	14.4	0	1.1	1
gene_id_43172	0.27	14.6	0	1	12.6	0	1.8	1
gene_id_110230	0.27	14.5	0.1	0.49	13.7	0	1.4	1
gene_id_242826	0.28	14.5	0.8	16	8.7	0.7	2.1	1
gene_id_161653	0.29	14.4	0.1	0.68	13.2	0.1	1.5	1
gene_id_49067	0.3	14.4	0.2	0.85	12.9	0.1	1.7	1
gene_id_104000	0.3	14.4	0.1	0.44	13.9	0.1	1.2	1
gene_id_47111	0.31	14.4	0.2	1.2	12.5	0.3	1.7	1
gene_id_77877	0.32	14.3	0	40	7.4	0.3	2.1	2
gene_id_145904	0.32	14.3	0	0.48	13.7	0	1.2	1
gene_id_202146	0.34	14.2	0	0.6	13.4	0	1.4	1
gene_id_72117	0.35	14.2	0	10	9.3	0	2.1	2
gene_id_286664	0.35	14.2	0	0.6	13.4	0	1.3	1

gene_id_235769	0.36	14.2	0.8	4.6	10.5	0.8	2	1
gene_id_274265	0.37	14.1	0	2.9	11.2	0	1.9	2
gene_id_63761	0.39	14	0	0.58	13.5	0	1.3	1
gene_id_362937	0.39	14	0	0.43	13.9	0	1.1	1
gene_id_67587	0.39	14	0.2	91	6.3	0	2	2
gene_id_31905	0.41	14	0	2.6	11.3	0	2	1
gene_id_28528	0.41	14	0	4.2	10.6	0	1.9	2
gene_id_326776	0.42	13.9	0	0.46	13.8	0	1.1	1
gene_id_341003	0.42	13.9	0	20	8.4	0	2.4	1
gene_id_145913	0.42	13.9	0	0.64	13.3	0	1.5	1
gene_id_46318	0.42	13.9	0.5	19	8.5	0.5	2.2	1
gene_id_53023	0.43	13.9	0.2	63	6.8	0.1	2.7	2
gene_id_101375	0.43	13.9	0.1	0.84	12.9	0.1	1.4	1
gene_id_99651	0.44	13.9	0.2	10	9.3	0.2	2	2
gene_id_374856	0.46	13.8	0.2	0.48	13.7	0.2	1	1
gene_id_138778	0.46	13.8	0	0.58	13.5	0	1.1	1
gene_id_396452	0.52	13.6	0	0.61	13.4	0	1	1
gene_id_89375	0.52	13.6	0.5	15	8.9	0.6	2	2
gene_id_300316	0.52	13.6	0	0.78	13	0	1.2	1
gene_id_274644	0.53	13.6	0.7	19	8.5	0.1	2.4	2
gene_id_141061	0.54	13.6	0.1	11	9.3	0.1	2	2
gene_id_323339	0.55	13.5	0.1	1.7	11.9	0	1.7	2
gene_id_42271	0.57	13.5	0	0.83	13	0	1.2	1
gene_id_182174	0.58	13.5	0	0.66	13.3	0	1.1	1
gene_id_255892	0.59	13.4	0	0.83	13	0	1.1	1
gene_id_198242	0.59	13.4	0	0.83	13	0	1.1	1
gene_id_341909	0.62	13.4	0	1.1	12.5	0	1.4	1
gene_id_295834	0.63	13.4	0	0.8	13	0	1.2	1
gene_id_7059	0.63	13.3	0	1.2	12.5	0	1.4	1
gene_id_367092	0.65	13.3	0.1	0.67	13.3	0.1	1.1	1
gene_id_210872	0.66	13.3	0	0.7	13.2	0	1.1	1

gene_id_76481	0.67	13.3	0	0.85	12.9	0	1.2	1
gene_id_178399	0.68	13.2	0.4	54	7	0.1	2	2
gene_id_330258	0.68	13.2	0.3	61	6.8	0	2	2
gene_id_251738	0.69	13.2	0.2	0.93	12.8	0.2	1.1	1
gene_id_383227	0.71	13.2	0	0.76	13.1	0	1.1	1
gene_id_134184	0.72	13.1	0	24	8.2	0	2.1	2
gene_id_283848	0.73	13.1	0.1	6	10.1	0.1	2	1
gene_id_49225	0.74	13.1	0.1	0.98	12.7	0.1	1.1	1
gene_id_187158	0.75	13.1	0	40	7.4	0	2	2
gene_id_223811	0.75	13.1	0	1	12.6	0	1.1	1
gene_id_63306	0.76	13.1	0.1	1.6	12.1	0.1	1.4	1
gene_id_50615	0.83	13	0.1	5.9	10.1	0.2	1.9	2
gene_id_313655	0.86	12.9	0.4	12	9.2	0.4	2	1
gene_id_243058	0.86	12.9	0	1.1	12.6	0	1.1	1
gene_id_169835	0.86	12.9	0.4	13	9.1	0.1	2	2
gene_id_191474	0.87	12.9	0	1.3	12.3	0	1.3	1
gene_id_263136	0.88	12.9	0.1	0.96	12.7	0.1	1.1	1
gene_id_33129	0.9	12.8	0	1	12.6	0	1.1	1
gene_id_248623	0.9	12.8	0	1.1	12.6	0	1.1	1
gene_id_317966	0.94	12.8	0	1.1	12.5	0	1	1
gene_id_350273	0.94	12.8	0	1.3	12.3	0	1.1	1
gene_id_180048	0.95	12.8	0	4.4	10.6	0	1.8	1
gene_id_282529	0.97	12.7	0	1.4	12.2	0	1.1	1
gene_id_173656	0.97	12.7	0	2	11.7	0	1.6	1
gene_id_383551	0.99	12.7	0.1	1.1	12.6	0.1	1	1
gene_id_346722	1	12.6	0	1	12.6	0	1.1	1
gene_id_18460	1	12.6	0	2.8	11.2	0	1.5	1
gene_id_23825	1.1	12.6	0	1.3	12.3	0	1.1	1
gene_id_74333	1.1	12.6	0	1.6	12	0	1.2	1
gene_id_183397	1.1	12.6	0	1.3	12.3	0	1.1	1
gene_id_245055	1.1	12.5	0.4	4.8	10.4	0.2	1.8	1

gene_id_216749	1.1	12.5	0	1.8	11.8	0	1.4	1
gene_id_315909	1.1	12.5	0	1.2	12.4	0	1.1	1
gene_id_89654	1.2	12.5	0	9.3	9.5	0	2	1
gene_id_251934	1.2	12.4	0.4	53	7	0	2	2
gene_id_95755	1.2	12.4	0	1.5	12.1	0	1.1	1
gene_id_190937	1.2	12.4	0	2.5	11.4	0	1.5	1
gene_id_55038	1.2	12.4	0.1	30	7.8	0.1	2.2	1
gene_id_43677	1.3	12.3	0.1	3.70E+02	4.2	0	2.9	3
gene_id_346330	1.4	12.3	0.2	1.4	12.2	0.2	1.1	1
gene_id_258351	1.4	12.2	0	5	10.4	0	1.8	1
gene_id_382281	1.4	12.2	0	5.4	10.3	0	1.7	2
gene_id_14314	1.5	12.1	0	1.8	11.8	0	1.2	1
gene_id_296020	1.5	12.1	0.1	2.00E+02	5.1	0.4	2.1	2
gene_id_168182	1.5	12.1	1.3	1.00E+02	6.1	0	2.1	2
gene_id_234811	1.5	12.1	0	1.7	11.9	0	1	1
gene_id_388393	1.5	12.1	0	3.9	10.7	0	1.5	1
gene_id_22930	1.6	12	0	22	8.3	0	2.1	1
gene_id_221381	1.7	12	0	1.7	12	0	1.1	1
gene_id_320873	1.7	11.9	0	1.7	11.9	0	1.1	1
gene_id_78848	1.7	11.9	0.1	1.9	11.8	0.1	1.2	1
gene_id_223576	1.7	11.9	0.1	3.4	10.9	0	1.4	1
gene_id_158210	1.7	11.9	0	3	11.1	0	1.5	1
gene_id_139768	1.7	11.9	0	2.1	11.7	0	1.1	1
gene_id_168917	1.7	11.9	0	12	9.2	0	1.9	2
gene_id_179048	1.8	11.9	0.3	7.3	9.8	0	2	2
gene_id_102224	1.8	11.9	0.1	1.40E+02	5.6	0.1	2	2
gene_id_151104	1.8	11.9	0	1.9	11.8	0	1.1	1
gene_id_141218	1.8	11.8	0.1	26	8	0.1	2	1
gene_id_395973	1.9	11.8	0	1.9	11.8	0	1	1
gene_id_354891	1.9	11.8	0	2.3	11.5	0	1	1
gene_id_114901	1.9	11.7	2.7	45	7.3	2.7	2.1	1

gene_id_326866	2	11.7	0	1.40E+02	5.7	0	2	2
gene_id_71766	2	11.7	0	2.6	11.3	0	1.1	1
gene_id_20326	2	11.7	0	2.7	11.3	0	1.1	1
gene_id_356307	2	11.7	0	2.1	11.6	0	1	1
gene_id_172958	2.1	11.6	0	2.9	11.2	0	1.1	1
gene_id_373069	2.1	11.6	0	3.4	10.9	0	1.4	1
gene_id_146487	2.2	11.5	0	3.1	11.1	0	1.1	1
gene_id_34228	2.2	11.5	0	4.1	10.7	0	1.3	1
gene_id_245667	2.2	11.5	0.2	4.9	10.4	0	1.6	1
gene_id_298982	2.3	11.5	0	2.7	11.3	0	1.1	1
gene_id_218309	2.3	11.5	0	4.1	10.7	0	1.5	1
gene_id_143507	2.3	11.5	0	18	8.6	0	1.9	2
gene_id_26901	2.3	11.5	0	3.5	10.9	0	1.2	1
gene_id_115993	2.4	11.4	0.3	5.6	10.2	0.1	1.6	1
gene_id_246131	2.4	11.4	0	8.6	9.6	0	1.7	1
gene_id_141153	2.4	11.4	0.3	34	7.7	0	2.4	2
gene_id_374753	2.5	11.4	0.1	3.8	10.8	0.1	1.4	1
gene_id_27413	2.6	11.3	0.1	41	7.4	0	2	2
gene_id_130897	2.6	11.3	0.5	50	7.1	0.2	2	2
gene_id_247428	2.6	11.3	0.9	10	9.4	0.1	2	2
gene_id_167676	2.6	11.3	0	20	8.4	0	1.9	2
gene_id_239571	2.6	11.3	0	2.9	11.2	0	1	1
gene_id_72874	2.7	11.3	0	4.4	10.6	0	1.3	1
gene_id_371420	2.7	11.3	0	3	11.1	0	1.1	1
gene_id_340936	2.7	11.3	0.1	6.1	10.1	0	1.5	1
gene_id_152089	2.7	11.2	0	3.5	10.9	0	1.1	1
gene_id_20536	2.8	11.2	0	11	9.3	0	1.7	1
gene_id_240354	2.8	11.2	0.1	3.2	11	0.1	1.1	1
gene_id_124031	2.8	11.2	0	3.3	11	0	1	1
gene_id_137053	2.9	11.1	0	3.7	10.8	0	1.2	1
gene_id_184401	3	11.1	0.1	11	9.3	0.1	1.7	1

gene_id_111120	3.1	11.1	0.1	12	9.1	0.1	1.8	1
gene_id_78211	3.1	11.1	0	4.8	10.4	0	1.2	1
gene_id_43446	3.2	11	0	4.7	10.5	0	1.2	1
gene_id_141986	3.3	11	0	9.1	9.5	0	1.7	1
gene_id_51480	3.3	11	0.3	5.6	10.2	0.3	1.4	1
gene_id_131292	3.4	10.9	0	4.3	10.6	0	1.1	1
gene_id_57521	3.4	10.9	0.1	50	7.1	0.1	2	1
gene_id_60427	3.4	10.9	0.1	80	6.4	0	2	2
gene_id_86100	3.5	10.9	0	35	7.6	0	2	2
gene_id_365375	3.5	10.9	0.1	3.6	10.8	0.1	1	1
gene_id_129170	3.7	10.8	0	5.4	10.3	0	1.1	1
gene_id_258164	3.9	10.7	0	5.6	10.2	0	1.1	1
gene_id_28733	4	10.7	0.2	57	6.9	0	2	2
gene_id_206681	4.1	10.7	0.2	67	6.7	0	2	2
gene_id_268113	4.1	10.7	0	5.9	10.1	0	1.2	1
gene_id_52352	4.3	10.6	0	6.9	9.9	0	1.3	1
gene_id_15882	4.6	10.5	0	7.1	9.9	0	1.4	1
gene_id_158302	4.6	10.5	0	5.7	10.2	0	1.1	1
gene_id_112002	4.6	10.5	0	4.6	10.5	0	1	1
gene_id_210125	4.8	10.4	0	12	9.1	0	1.6	1
gene_id_267628	4.8	10.4	0.7	8.3	9.7	0.5	1.4	1
gene_id_207450	5	10.4	0	5.9	10.2	0	1.1	1
gene_id_124708	5.1	10.4	0.9	99	6.1	0	2.6	2
gene_id_317310	5.2	10.3	0	10	9.4	0	1.3	1
gene_id_174741	5.3	10.3	0.1	11	9.2	0	1.5	1
gene_id_317149	5.6	10.2	0.1	28	8	0.1	1.9	1
gene_id_66030	5.7	10.2	0.2	8.4	9.6	0.1	1.3	1
gene_id_315868	5.7	10.2	0	34	7.7	0	1.9	1
gene_id_123023	5.8	10.2	0	38	7.5	0	1.9	1
gene_id_72600	5.8	10.2	0.1	1.20E+02	5.8	0.1	2	2
gene_id_127349	5.9	10.2	0.2	10	9.4	0.2	1.3	1

gene_id_276813	6	10.1	0	7.5	9.8	0	1.1	1
gene_id_75929	6.3	10.1	0.2	1.20E+02	5.8	0.2	2.2	1
gene_id_143165	6.3	10	0	8.3	9.7	0	1.1	1
gene_id_219119	7.1	9.9	0.1	13	9.1	0	1.3	1
gene_id_186433	8.4	9.6	0.9	87	6.3	0.5	2	2

1-B: NCBI batch CDD search output using 22 full length putative AcXE ORFs from the Namib hypolith metagenome as query (excluding CD descriptions).

Query	Hit type	PSSM-ID	From	To	E-Value	Bitscore	Accession	Short name	Incomplete	Super family
Q#1 >gene_id_41691	non-specific	233593	42	248	3.57E-74	229.296	TIGR01840	esterase_phb	-	cl21494
	superfamily	271914	42	248	3.57E-74	229.296	cl21494	Esterase_lipase	-	-
	non-specific	226040	33	307	1.79E-46	160.688	COG3509	LpqC	-	cl21494
	non-specific	119023	40	235	2.75E-34	125.945	pfam10503	Esterase_phb	-	cl21494
	non-specific	249771	79	165	0.003391	36.8119	pfam00326	Peptidase_S9	C	cl21494
	multi-dom	226584	38	163	7.22E-06	45.6718	COG4099	COG4099	NC	-
	multi-dom	224423	41	167	9.15E-06	45.5982	COG1506	DAP2	N	-
	multi-dom	257231	57	157	3.26E-05	42.9213	pfam12697	Abhydrolase_6	C	-
	multi-dom	223489	38	166	0.000808	38.9008	COG0412	COG0412	C	-
	multi-dom	223669	52	167	0.000843	38.8407	COG0596	MhpC	C	-
	multi-dom	224561	47	170	0.004472	36.544	COG1647	COG1647	C	-
	multi-dom	223477	53	157	0.004956	36.1285	COG0400	COG0400	C	-
Q#2 >gene_id_161768	non-specific	226040	29	317	1.09E-77	244.276	COG3509	LpqC	-	cl21494
	superfamily	271914	29	317	1.09E-77	244.276	cl21494	Esterase_lipase	-	-
	non-specific	233593	48	257	2.21E-77	240.082	TIGR01840	esterase_phb	-	cl21494
	non-specific	119023	49	240	3.62E-42	148.671	pfam10503	Esterase_phd	-	cl21494
	non-specific	249771	127	188	0.001214	38.3527	pfam00326	Peptidase_S9	NC	cl21494
	multi-dom	257229	59	245	0.001082	37.6968	pfam12695	Abhydrolase_5	-	-
	multi-dom	226584	43	172	3.46E-08	53.3757	COG4099	COG4099	NC	-
	multi-dom	224423	36	172	7.56E-07	49.4502	COG1506	DAP2	N	-
	multi-dom	257231	60	243	0.000367	40.2249	pfam12697	Abhydrolase_6	-	-
	multi-dom	223477	75	182	0.001099	38.4397	COG0400	COG0400	C	-
	multi-dom	249959	102	189	0.002186	37.883	pfam00561	Abhydrolase_1	C	-
	multi-dom	223489	43	185	0.004153	36.9748	COG0412	COG0412	C	-

Q#3 >gene_id_47111	non-specific	119023	100	319	1.70E-60	196.821	pfam10503	Esterase_phd	-	cl21494
	superfamily	271914	100	319	1.70E-60	196.821	cl21494	Esterase_lipase	-	-
	non-specific	226040	96	373	4.43E-45	158.762	COG3509	LpqC	-	cl21494
	non-specific	233593	101	299	3.59E-41	145.323	TIGR01840	esterase_phb	-	cl21494
	non-specific	249771	193	292	6.74E-07	48.3679	pfam00326	Peptidase_S9	NC	cl21494
	non-specific	251171	157	229	0.005433	36.5397	pfam02230	Abhydrolase_2	NC	cl21494
	multi-dom	223489	99	371	0.000596	39.6712	COG0412	COG0412	-	-
	multi-dom	226584	101	223	4.79E-11	62.2353	COG4099	COG4099	NC	-
	multi-dom	224423	83	293	1.82E-07	51.3762	COG1506	DAP2	N	-
	multi-dom	223477	104	293	8.58E-05	41.9065	COG0400	COG0400	-	-
	multi-dom	257229	188	301	0.000315	39.2376	pfam12695	Abhydrolase_5	N	-
	multi-dom	250111	187	232	0.001233	38.5959	pfam00756	Esterase	NC	-
	multi-dom	257231	117	299	0.003481	37.1433	pfam12697	Abhydrolase_6	-	-
Q#4 >gene_id_223016	non-specific	119023	27	251	3.56E-70	220.319	pfam10503	Esterase_phd	-	cl21494
	superfamily	271914	27	251	3.56E-70	220.319	cl21494	Esterase_lipase	-	-
	non-specific	226040	22	315	8.70E-48	164.54	COG3509	LpqC	-	cl21494
	non-specific	233593	28	220	1.95E-35	128.759	TIGR01840	esterase_phb	-	cl21494
	non-specific	249771	90	224	1.01E-08	53.3755	pfam00326	Peptidase_S9	C	cl21494
	multi-dom	250111	30	260	6.82E-06	45.5295	pfam00756	Esterase	-	-
	multi-dom	226584	28	152	8.33E-13	67.2429	COG4099	COG4099	NC	-
	multi-dom	223477	39	160	3.08E-08	51.9217	COG0400	COG0400	C	-
	multi-dom	224423	28	157	9.39E-07	49.065	COG1506	DAP2	N	-
	multi-dom	223489	120	165	0.003108	36.9748	COG0412	COG0412	NC	-
	multi-dom	257229	116	230	0.003295	36.156	pfam12695	Abhydrolase_5	N	-
	multi-dom	257231	45	227	0.003437	36.7581	pfam12697	Abhydrolase_6	-	-
Q#5 >gene_id_10715	non-specific	119023	30	236	2.54E-63	201.829	pfam10503	Esterase_phd	-	cl21494
	superfamily	271914	30	236	2.54E-63	201.829	cl21494	Esterase_lipase	-	-
	non-specific	226040	20	302	2.25E-48	165.31	COG3509	LpqC	-	cl21494
	non-specific	233593	31	237	4.04E-38	135.308	TIGR01840	esterase_phb	-	cl21494
	non-specific	249771	117	211	3.47E-05	42.5899	pfam00326	Peptidase_S9	NC	cl21494
	multi-dom	257229	46	223	7.71E-05	40.7784	pfam12695	Abhydrolase_5	-	-

	multi-dom	226584	31	256	7.29E-10	57.9981	COG4099	COG4099	N	-
	multi-dom	224423	30	211	7.68E-06	45.9834	COG1506	DAP2	N	-
	multi-dom	223477	45	214	1.51E-05	43.8325	COG0400	COG0400	C	-
	multi-dom	257231	47	225	0.000238	40.2249	pfam12697	Abhydrolase_6	-	-
	multi-dom	250111	33	225	0.000241	40.5219	pfam00756	Esterase	-	-
	multi-dom	224561	32	148	0.000261	40.3959	COG1647	COG1647	C	-
	multi-dom	223669	39	264	0.003935	36.9147	COG0596	MhpC	C	-
Q#6 >gene_id_154730	non-specific	226040	27	272	3.16E-26	104.834	COG3509	LpqC	-	cl21494
	superfamily	271914	27	272	3.16E-26	104.834	cl21494	Esterase_lipase	-	-
	non-specific	233593	75	219	6.98E-10	56.3416	TIGR01840	esterase_phb	-	cl21494
	non-specific	249771	150	226	3.41E-08	51.4495	pfam00326	Peptidase_S9	NC	cl21494
	non-specific	119023	75	225	0.00012	41.2008	pfam10503	Esterase_phd	-	cl21494
	multi-dom	226584	71	224	1.19E-10	60.3093	COG4099	COG4099	N	-
	multi-dom	224423	50	193	8.43E-07	48.6798	COG1506	DAP2	N	-
	multi-dom	223477	86	225	1.68E-06	46.5289	COG0400	COG0400	C	-
	multi-dom	223489	126	220	4.04E-05	42.7528	COG0412	COG0412	NC	-
	multi-dom	257229	130	225	0.00015	39.6228	pfam12695	Abhydrolase_5	-	-
	multi-dom	223700	36	192	0.002409	37.423	COG0627	COG0627	C	-
Q#7 >gene_id_7346	specific	213022	121	293	4.33E-79	240.984	cd10917	CE4_NodB_like_6s_7s	-	cl15692
	superfamily	271775	121	293	4.33E-79	240.984	cl15692	CE4_SF superfamily	-	-
	non-specific	200584	121	306	2.61E-70	219.085	cd10962	CE4_GT2-like	-	cl15692
	non-specific	200582	121	302	1.88E-66	209	cd10959	CE4_NodB_like_3	-	cl15692
	non-specific	200569	121	299	1.46E-65	206.629	cd10944	CE4_SmPgdA_like	-	cl15692
	non-specific	234001	121	306	4.72E-64	202.953	TIGR02764	spore_ybaN_pdaB	-	cl15692
	non-specific	200578	121	305	1.76E-62	198.579	cd10954	CE4_CtAXE_like	-	cl15692
	non-specific	200571	121	302	3.33E-60	192.6	cd10947	CE4_SpPgdA_BsYjeA_like	-	cl15692
	non-specific	200580	118	307	1.62E-58	188.704	cd10956	CE4_BH1302_like	-	cl15692

non-specific	200574	138	306	1.68E-52	172.846	cd10950	CE4_BsYlxY_like	-	cl15692
non-specific	200575	121	302	2.17E-46	157.431	cd10951	CE4_ClCDA_like	-	cl15692
non-specific	200573	121	308	9.84E-44	150.255	cd10949	CE4_BsPdaB_like	-	cl15692
non-specific	200572	118	302	6.65E-42	146.274	cd10948	CE4_BsPdaA_like	-	cl15692
non-specific	211966	123	303	1.24E-41	144.856	TIGR04243	nodulat_NodB	-	cl15692
non-specific	250683	117	238	3.64E-39	135.826	pfam01522	Polysacc_deac_1	-	cl15692
non-specific	200581	121	305	3.99E-38	135.116	cd10958	CE4_NodB_like_2	-	cl15692
non-specific	200568	123	299	4.09E-38	135.359	cd10943	CE4_NodB	-	cl15692
non-specific	200576	124	293	1.39E-36	130.95	cd10952	CE4_MrCDA_like	-	cl15692
non-specific	200577	123	296	1.47E-36	130.772	cd10953	CE4_SlAXE_like	-	cl15692
non-specific	200579	121	302	2.02E-33	122.81	cd10955	CE4_BH0857_like	-	cl15692
non-specific	200570	121	304	1.58E-18	81.2959	cd10946	CE4_Mll8295_like	-	cl15692
non-specific	213021	139	246	1.62E-17	78.8921	cd10916	CE4_PuuE_HpPgdA_like	C	cl15692
non-specific	200583	123	307	1.31E-16	76.1146	cd10960	CE4_NodB_like_1	-	cl15692
non-specific	200589	123	293	5.54E-16	73.9517	cd10967	CE4_GLA_like_6s	-	cl15692
non-specific	200566	132	219	3.01E-15	72.7097	cd10941	CE4_PuuE_HpPgdA_like_2	C	cl15692
non-specific	200563	139	228	2.44E-14	70.2786	cd10938	CE4_HpPgdA_like	C	cl15692
non-specific	200567	125	225	4.69E-14	69.4295	cd10942	CE4_u11	C	cl15692
non-specific	213020	122	226	4.00E-12	62.0796	cd10585	CE4_SF	-	cl15692
non-specific	200565	139	218	8.43E-12	63.5659	cd10940	CE4_PuuE_HpPgdA_like_1	C	cl15692
non-specific	213028	121	224	4.69E-10	56.5115	cd10973	CE4_DAC_u4_5s	C	cl15692
non-specific	213023	123	237	5.43E-09	53.369	cd10918	CE4_NodB_like_5s_6s	-	cl15692
non-specific	200600	138	226	6.39E-09	54.7665	cd10978	CE4_Sll1306_like	NC	cl15692
non-specific	200599	139	211	2.95E-08	52.7133	cd10977	CE4_PuuE_SpCDA1	NC	cl15692

	non-specific	213024	119	164	2.67E-06	45.3461	cd10966	CE4_yadE_5s	C	cl15692
	non-specific	200602	138	220	1.25E-05	44.8462	cd10980	CE4_SpCDA1	NC	cl15692
	non-specific	213026	96	155	1.52E-05	44.2	cd10969	CE4_Ecf1_like_5s	C	cl15692
	non-specific	200545	122	183	4.05E-05	43.119	cd10919	CE4_CDA_like	C	cl15692
	non-specific	200553	119	218	0.000594	39.3294	cd10927	CE4_u3	C	cl15692
	non-specific	213027	122	277	0.001194	38.0651	cd10970	CE4_DAC_u1_6s	-	cl15692
	non-specific	200555	132	185	0.002323	37.6336	cd10929	CE4_u5	C	cl15692
	non-specific	132086	10	68	0.004479	35.8295	TIGR03042	PS_II_psbQ_bact	C	cl21653
	superfamily	272073	10	68	0.004479	35.8295	cl21653	PsbQ	C	-
	non-specific	200601	140	211	0.005276	36.4503	cd10979	CE4_PuuE_like	NC	cl15692
	multi-dom	223798	84	316	2.12E-55	183.281	COG0726	CDA1	-	-
	multi-dom	131930	108	306	1.77E-40	142.536	TIGR02884	spore_pdaA	-	-
	multi-dom	131920	114	307	6.84E-38	136.923	TIGR02873	spore_ylxY	N	-
	multi-dom	213757	134	218	5.44E-12	63.8949	TIGR03006	pepcterm_polyde	C	-
	specific	213022	25	204	1.82E-63	197.456	cd10917	CE4_NodB_like_6s_7s	-	cl15692
Q#8 >gene_id_7936	superfamily	271775	25	204	1.82E-63	197.456	cl15692	CE4_SF	-	-
	non-specific	234001	20	217	3.77E-54	174.063	TIGR02764	spore_ybaN_pdaB	-	cl15692
	non-specific	200582	25	213	5.18E-54	173.562	cd10959	CE4_NodB_like_3	-	cl15692
	non-specific	200578	26	216	6.77E-51	165.451	cd10954	CE4_CtAXE_like	-	cl15692
	non-specific	200584	27	218	5.12E-49	160.92	cd10962	CE4_GT2-like	-	cl15692
	non-specific	200574	22	217	4.20E-45	150.889	cd10950	CE4_BsYlxY_like	-	cl15692
	non-specific	200571	25	213	5.75E-42	142.139	cd10947	CE4_SpPgdA_BsYjeA_like	-	cl15692
	non-specific	200580	22	218	3.24E-41	140.94	cd10956	CE4_BH1302_like	-	cl15692
	non-specific	200569	25	212	2.26E-39	135.752	cd10944	CE4_SmPgdA_like	-	cl15692
	non-specific	200573	22	218	6.07E-39	134.847	cd10949	CE4_BsPdaB_like	-	cl15692
	non-specific	211966	25	214	5.39E-37	129.833	TIGR04243	nodulat_NodB	-	cl15692
	non-specific	200572	17	213	9.34E-36	127.399	cd10948	CE4_BsPdaA_like	-	cl15692
	non-specific	200579	25	215	1.34E-35	126.277	cd10955	CE4_BH0857_like	-	cl15692

non-specific	200577	27	211	6.41E-34	121.142	cd10953	CE4_SlAXE_like	-	cl15692
non-specific	200575	27	213	8.92E-34	121.223	cd10951	CE4_ClCDA_like	-	cl15692
non-specific	200568	25	210	1.16E-32	118.41	cd10943	CE4_NodB	-	cl15692
non-specific	250683	22	142	2.08E-31	113.1	pfam01522	Polysacc_deac_1	-	cl15692
non-specific	200581	25	216	1.68E-27	104.686	cd10958	CE4_NodB_like_2	-	cl15692
non-specific	200576	27	184	3.22E-21	87.0367	cd10952	CE4_MrCDA_like	-	cl15692
non-specific	213023	27	137	2.19E-17	75.7106	cd10918	CE4_NodB_like_5s_6s	-	cl15692
non-specific	200589	26	143	1.24E-13	65.4774	cd10967	CE4_GLA_like_6s	C	cl15692
non-specific	200570	25	214	1.79E-13	65.5028	cd10946	CE4_Mll8295_like	-	cl15692
non-specific	200583	25	91	5.04E-10	56.0842	cd10960	CE4_NodB_like_1	C	cl15692
non-specific	213028	27	137	6.02E-09	51.8891	cd10973	CE4_DAC_u4_5s	C	cl15692
non-specific	200565	43	122	8.85E-09	53.1656	cd10940	CE4_PuuE_HpPgdA_like_1	C	cl15692
non-specific	200566	43	122	1.10E-08	52.6793	cd10941	CE4_PuuE_HpPgdA_like_2	C	cl15692
non-specific	200563	43	137	5.13E-08	50.6335	cd10938	CE4_HpPgdA_like	C	cl15692
non-specific	213021	43	128	5.75E-08	50.3873	cd10916	CE4_PuuE_HpPgdA_like	C	cl15692
non-specific	213027	27	136	1.01E-07	49.2359	cd10970	CE4_DAC_u1_6s	C	cl15692
non-specific	213020	27	124	6.50E-06	43.2048	cd10585	CE4_SF	C	cl15692
non-specific	213025	25	137	4.12E-05	41.0801	cd10968	CE4_Mlr8448_like_5s	C	cl15692
non-specific	200600	43	137	8.27E-05	41.2846	cd10978	CE4_Sll1306_like	NC	cl15692
non-specific	200594	27	136	0.000183	40.0077	cd10972	CE4_DAC_u3_5s	C	cl15692
non-specific	200567	36	128	0.002155	37.0727	cd10942	CE4_u11	C	cl15692
non-specific	200593	77	146	0.004095	35.7467	cd10971	CE4_DAC_u2_5s	N	cl15692
non-specific	200599	43	126	0.007159	35.3793	cd10977	CE4_PuuE_SpCDA1	NC	cl15692

	multi-dom	223798	22	218	3.21E-39	137.827	COG0726	CDA1	N	-
	multi-dom	131930	22	217	6.85E-37	130.21	TIGR02884	spore_pdaA	-	-
	multi-dom	131920	39	217	4.33E-33	121.515	TIGR02873	spore_ylxY	N	-
	multi-dom	213757	40	137	6.89E-07	47.7165	TIGR03006	pepcterm_polyde	C	-
Q#9 >gene_id_9548	specific	238141	72	241	6.48E-20	84.001	cd00229	SGNH_hydrolase	-	cl01053
	superfamily	260768	72	241	6.48E-20	84.001	cl01053	SGNH_hydrolase	-	-
	non-specific	257796	73	241	1.52E-23	94.1248	pfam13472	Lipase_GDSL_2	-	cl01053
	non-specific	238870	73	241	8.11E-19	80.7755	cd01832	SGNH_hydrolase_like_1	-	cl01053
	non-specific	238874	73	241	2.48E-16	73.839	cd01836	FeeA_FeeB_like	-	cl01053
	non-specific	238872	72	241	9.79E-15	69.2442	cd01834	SGNH_hydrolase_like_2	-	cl01053
	non-specific	238860	73	176	3.22E-13	64.8416	cd01822	Lysophospholipase_L1_like	C	cl01053
	non-specific	239947	73	241	1.16E-11	60.7249	cd04506	SGNH_hydrolase_YpmR_like	-	cl01053
	non-specific	225353	63	241	2.62E-11	60.3034	COG2755	TesA	-	cl01053
	non-specific	250036	73	241	4.26E-10	56.9265	pfam00657	Lipase_GDSL	-	cl01053
	non-specific	238861	72	150	4.16E-08	51.6847	cd01823	SEST_like	C	cl01053
	non-specific	238866	109	241	5.29E-06	44.1926	cd01828	sialate_O-acetylesterase_like 2 -cl01053	-	
	non-specific	258744	72	241	0.000414	38.7678	pfam14606	Lipase_GDSL_3	-	cl01053
	non-specific	238876	73	173	0.00084	38.0023	cd01838	Isoamyl_acetate_hydrolase_like	C	cl01053
	non-specific	238871	114	240	0.005454	35.289	cd01833	XynB_like	-	cl01053
Q#10 >gene_id_20333	specific	257796	24	185	4.61E-24	95.2804	pfam13472	Lipase_GDSL_2	-	cl01053
	superfamily	260768	24	185	4.61E-24	95.2804	cl01053	SGNH_hydrolase	-	-
	non-specific	238872	24	192	1.20E-19	83.1114	cd01834	SGNH_hydrolase_like_2	-	cl01053
	non-specific	238141	25	192	1.38E-17	77.0674	cd00229	SGNH_hydrolase	-	cl01053
	non-specific	225353	19	196	1.10E-16	74.941	COG2755	TesA		-

	non-specific	239945	24	193	7.56E-16	71.9744	cd04501	SGNH_hydrolase_like_4	-	cl01053
	non-specific	238866	26	183	6.45E-15	69.2305	cd01828	sialate_O-acetylesterase_like	-	
	non-specific	250036	14	189	2.44E-14	68.0973	pfam00657	Lipase_GDSL	-	cl01053
	non-specific	238876	16	193	7.76E-14	66.507	cd01838	Isoamyl-acetate hydrolyzing esterase-like	-	
	non-specific	238860	25	194	4.48E-10	55.5968	cd01822	Lysophospholipase_L1_like N	-	cl01053
	non-specific	238863	26	197	9.78E-10	54.9726	cd01825	SGNH_hydrolase_peri1	-	cl01053
	non-specific	238870	48	192	1.17E-09	54.582	cd01832	SGNH_hydrolase_like_1	N	cl01053
	non-specific	238874	25	194	3.03E-09	53.8086	cd01836	FeeA_FeeB_like	N	cl01053
	non-specific	238871	27	186	4.10E-09	52.623	cd01833	XynB_like	-	cl01053
	non-specific	238879	22	193	1.07E-07	48.8648	cd01841	NnaC_like	-	cl01053
	non-specific	238878	119	193	2.87E-06	44.5377	cd01840	SGNH_hydrolase_yrhL_like N	N	cl01053
	non-specific	239947	25	192	3.38E-06	45.3169	cd04506	SGNH_hydrolase_YpmR_like -	-	cl01053
	non-specific	238881	24	193	1.13E-05	43.4646	cd01844	SGNH_hydrolase_like_6	-	cl01053
	non-specific	238865	25	193	5.63E-05	41.2761	cd01827	sialate_O-acetylesterase_like 1	-	cl01053
	non-specific	238858	28	195	0.000407	39.1947	cd01820	PAF_acetylesterase_like	N	cl01053
	non-specific	238869	52	194	0.001146	37.3232	cd01831	Endoglucanase_E_like	N	cl01053
	non-specific	238861	105	192	0.001791	37.4323	cd01823	SEST_like	N	cl01053
	non-specific	238859	131	193	0.002624	36.422	cd01821	Rhamnogalacturan_acetylesterase_like	N	cl01053
	non-specific	238867	42	195	0.003099	36.4861	cd01829	SGNH_hydrolase_peri2	N	cl01053
Q#11	non-specific	249771	158	327	1.52E-06	46.8271	pfam00326	Peptidase_S9	-	cl21494

>gene_id_27345	superfamily	271914	158	327	1.52E-06	46.8271	cl21494	Esterase_lipase	-	-
	multi-dom	253201	5	322	9.81E-124	360.886	pfam05448	Axe1	-	-
	multi-dom	225989	5	327	1.81E-117	344.808	COG3458	COG3458	-	-
	multi-dom	224423	44	211	1.43E-12	67.1694	COG1506	DAP2	N	-
	multi-dom	223489	55	206	1.50E-07	50.4568	COG0412	COG0412	C	-
	multi-dom	251106	65	211	0.00012	41.5479	pfam02129	Peptidase_S15	C	-
	multi-dom	249959	160	279	0.005285	36.3422	pfam00561	Abhydrolase_1	C	-
	multi-dom	254474	162	197	0.005324	36.4116	pfam07859	Abhydrolase_3	NC	-
Q#12 >gene_id_39518	non-specific	249771	100	304	8.08E-12	62.235	pfam00326	Peptidase_S9	-	cl21494
	superfamily	271914	100	304	8.08E-12	62.235	cl21494	Esterase_lipase	-	-
	non-specific	226040	64	196	1.72E-05	44.3575	COG3509	LpqC	C	cl21494
	multi-dom	253201	5	321	2.01E-149	425.985	pfam05448	Axe1	-	-
	multi-dom	225989	4	321	6.80E-122	355.979	COG3458	COG3458	-	-
	multi-dom	224423	28	305	1.53E-15	75.6438	COG1506	DAP2	N	-
	multi-dom	257231	91	304	4.23E-10	57.5589	pfam12697	Abhydrolase_6	-	-
	multi-dom	223489	65	211	8.01E-08	51.2272	COG0412	COG0412	C	-
	multi-dom	257229	90	304	1.57E-06	45.786	pfam12695	Abhydrolase_5	-	-
	multi-dom	225176	61	307	2.81E-06	46.5968	COG2267	PldB	-	-
	multi-dom	251106	67	207	0.000178	41.1627	pfam02129	Peptidase_S15	C	-
	multi-dom	223730	29	314	0.000182	41.074	COG0657	Aes	-	-
	multi-dom	223999	59	304	0.000808	39.3082	COG1073	COG1073	-	-
	multi-dom	234315	85	199	0.000898	38.7336	TIGR03695	menH_SHCHC	C	-
	multi-dom	223477	71	210	0.002094	37.2841	COG0400	COG0400	C	-
Q#13 >gene_id_41102	specific	257796	72	261	1.05E-24	97.5916	pfam13472	Lipase_GDSL_2	-	cl01053
	superfamily	260768	72	261	1.05E-24	97.5916	cl01053	SGNH_hydrolase	-	-
	non-specific	239947	70	265	8.48E-22	90	cd04506	SGNH_hydrolase_YpmR_like	-	cl01053
	non-specific	238141	71	265	5.73E-21	87.4678	cd00229	SGNH_hydrolase	-	cl01053
	non-specific	225353	61	273	1.49E-15	72.2446	COG2755	TesA	-	cl01053
	non-specific	238870	71	265	4.50E-15	70.3752	cd01832	SGNH_hydrolase_li	-	cl01053

								ke_1		
	non-specific	238860	72	265	3.34E-14	67.538	cd01822	Lysophospholipase_L1_like	-	cl01053
	non-specific	250036	71	265	3.59E-13	65.7861	pfam00657	Lipase_GDSL	-	cl01053
	non-specific	238873	72	267	5.86E-11	58.8882	cd01835	SGNH_hydrolase_like_3	-	cl01053
	non-specific	238866	103	267	2.79E-10	56.5189	cd01828	sialate_O-acetylesterase_like 2	-	cl01053
	non-specific	239945	72	263	5.41E-10	56.1812	cd04501	SGNH_hydrolase_like_4	-	cl01053
	non-specific	238872	72	267	2.63E-09	54.2214	cd01834	SGNH_hydrolase_like_2	-	cl01053
	non-specific	238874	74	267	1.33E-08	52.2678	cd01836	FeeA_FeeB_like	-	cl01053
	non-specific	238871	110	263	3.60E-07	47.6154	cd01833	XynB_like	-	cl01053
	non-specific	238861	70	267	8.49E-05	41.6695	cd01823	SEST_like	-	cl01053
	non-specific	238868	72	265	0.000103	41.0733	cd01830	XynE_like	-	cl01053
	non-specific	238863	72	265	0.001775	37.2534	cd01825	SGNH_hydrolase_peri1	-	cl01053
Q#14 >gene_id_43977	non-specific	256874	45	87	0.001361	35.9929	pfam12146	Hydrolase_4	N	cl21494
	superfamily	271914	45	87	0.001361	35.9929	cl21494	Esterase_lipase N	-	
	multi-dom	257231	28	255	9.57E-23	93.3824	pfam12697	Abhydrolase_6	-	-
	multi-dom	223669	8	271	1.38E-15	73.5087	COG0596	MhpC	-	-
	multi-dom	225176	1	241	8.93E-09	53.9156	COG2267	PldB	C	-
	multi-dom	249959	54	264	5.75E-08	50.9798	pfam00561	Abhydrolase_1	-	-
	multi-dom	235628	4	95	2.85E-07	49.9794	PRK05855	PRK05855	C	-
	multi-dom	131480	47	254	3.17E-06	46.1957	TIGR02427	protocat_pcaD	-	-
	multi-dom	257229	27	125	1.88E-05	42.3192	pfam12695	Abhydrolase_5	C	-
	multi-dom	225487	8	111	9.32E-05	42.0363	COG2936	COG2936	C	-
	multi-dom	132100	22	267	0.001466	37.9475	TIGR03056	bchO_mg_che_rel	-	-
	multi-dom	215315	14	146	0.00206	37.8982	PLN02578	PLN02578	NC	-
	multi-dom	253201	94	142	0.002397	37.3186	pfam05448	Axe1	NC	-
	multi-dom	251106	13	114	0.005736	36.1551	pfam02129	Peptidase_S15	C	-

	multi-dom	224561	45	134	0.008077	35.3884	COG1647	COG1647	C	-
Q#15 >gene_id_64061	specific	238866	69	243	1.42E-68	211.369	cd01828	sialate_O-acetylesterase_like 2	-	cl01053
	superfamily	260768	69	243	1.42E-68	211.369	cl01053	SGNH_hydrolase	-	-
	non-specific	238879	68	243	1.82E-36	128.601	cd01841	NnaC_like	-	cl01053
	non-specific	257796	69	236	6.11E-30	111.459	pfam13472	Lipase_GDSL_2	-	cl01053
	non-specific	225353	72	241	5.70E-24	96.1269	COG2755	TesA	-	cl01053
	non-specific	250036	69	233	2.91E-20	85.8165	pfam00657	Lipase_GDSL	-	cl01053
	non-specific	239945	69	235	1.96E-19	82.3748	cd04501	SGNH_hydrolase_like_4	-	cl01053
	non-specific	238858	69	241	3.44E-19	82.337	cd01820	PAF_acetylesterase_like	-	cl01053
	non-specific	238876	69	236	7.26E-19	81.1446	cd01838	Isoamyl_acetate_hydrolase_like	-	cl01053
	non-specific	238874	90	236	3.13E-17	76.1502	cd01836	FeeA_FeeB_like	N	cl01053
	non-specific	238141	69	236	4.22E-17	75.9118	cd00229	SGNH_hydrolase	-	cl01053
	non-specific	238872	92	233	1.24E-14	68.859	cd01834	SGNH_hydrolase_like_2	-	cl01053
	non-specific	239946	68	241	2.53E-14	67.6982	cd04502	SGNH_hydrolase_like_7	-	cl01053
	non-specific	238860	68	233	6.85E-14	66.3824	cd01822	Lysophospholipase_L1_like	-	cl01053
	non-specific	238870	114	240	1.80E-12	62.6712	cd01832	SGNH_hydrolase_like_1	N	cl01053
	non-specific	238871	71	240	5.62E-11	58.0158	cd01833	XynB_like	-	cl01053
	non-specific	238873	84	234	6.72E-08	50.0287	cd01835	SGNH_hydrolase_like_3	-	cl01053
	non-specific	238868	91	233	0.000115	40.6881	cd01830	XynE_like	N	cl01053
	non-specific	182521	91	148	0.002036	37.0494	PRK10528	PRK10528	NC	cl01053
	non-specific	258744	108	241	0.003225	36.0714	pfam14606	Lipase_GDSL_3	N	cl01053
	non-specific	238867	68	233	0.008446	34.9453	cd01829	SGNH_hydrolase_peri2	-	cl01053
	multi-dom	235713	186	215	0.006822	35.8515	PRK06139	PRK06139	NC	-
Q#16 >gene_id_65101	non-specific	238873	5	205	1.49E-28	107.038	cd01835	SGNH_hydrolase_like_3	-	cl01053
	superfamily	260768	5	205	1.49E-28	107.038	cl01053	SGNH_hydrolase	-	-

								superfamily		
	non-specific	257796	5	201	3.09E-28	105.681	pfam13472	Lipase_GDSL_2	-	cl01053
	non-specific	238860	4	209	1.74E-24	95.6575	cd01822	Lysophospholipase_L1_like	-	cl01053
	non-specific	225353	1	210	2.63E-23	93.4305	COG2755	TesA	-	cl01053
	non-specific	238876	3	205	1.62E-21	88.0782	cd01838	Isoamyl_acetate_hydrolase_like	-	cl01053
	non-specific	250036	4	205	4.84E-21	86.9721	pfam00657	Lipase_GDSL	-	cl01053
	non-specific	238141	4	207	6.73E-18	77.8378	cd00229	SGNH_hydrolase	-	cl01053
	non-specific	238872	48	205	1.61E-16	73.8666	cd01834	SGNH_hydrolase_like_2	-	cl01053
	non-specific	239945	4	203	1.63E-16	73.5152	cd04501	SGNH_hydrolase_like_4	-	cl01053
	non-specific	238870	4	205	1.25E-14	68.064	cd01832	SGNH_hydrolase_like_1	-	cl01053
	non-specific	238866	3	205	3.20E-13	63.8377	cd01828	sialate_O-acetylesterase_like 2	-	cl01053
	non-specific	238871	4	203	1.87E-12	61.4826	cd01833	XynB_like	-	cl01053
	non-specific	239947	3	205	5.40E-10	55.3321	cd04506	SGNH_hydrolase_YpmR_like	-	cl01053
	non-specific	238877	3	205	9.59E-08	49.1891	cd01839	SGNH_arylesterase_like	-	cl01053
	non-specific	238874	2	210	2.71E-07	47.6455	cd01836	FeeA_FeeB_like	-	cl01053
	non-specific	238868	4	205	4.07E-07	47.6217	cd01830	XynE_like	-	cl01053
	non-specific	182521	4	210	2.74E-06	45.1386	PRK10528	PRK10528	-	cl01053
	non-specific	238881	4	204	3.35E-05	41.5386	cd01844	SGNH_hydrolase_like_6	-	cl01053
	non-specific	239946	43	209	0.000216	39.1934	cd04502	SGNH_hydrolase_like_7	-	cl01053
Q#17 >gene_id_67464	non-specific	238860	44	206	5.62E-46	152.667	cd01822	Lysophospholipase_L1_like	-	cl01053
	superfamily	260768	44	206	5.62E-46	152.667	cl01053	SGNH_hydrolase	-	-
	non-specific	257796	45	197	1.48E-29	109.533	pfam13472	Lipase_GDSL_2	-	cl01053
	non-specific	239945	44	199	1.61E-29	109.724	cd04501	SGNH_hydrolase_like_4	-	cl01053
	non-specific	182521	48	209	9.15E-22	88.6662	PRK10528	PRK10528	-	cl01053

non-specific	238141	44	203	3.73E-21	87.0826	cd00229	SGNH_hydrolase	-	cl01053
non-specific	225353	43	211	9.82E-18	77.6374	COG2755	TesA	-	cl01053
non-specific	250036	44	201	1.40E-17	77.3421	pfam00657	Lipase_GDSL	-	cl01053
non-specific	238876	43	203	2.98E-15	70.359	cd01838	Isoamyl_acetate_hydrolase_like	-	cl01053
non-specific	238866	43	210	3.37E-15	69.6157	cd01828	sialate_O-acetylesterase_like 2	-	cl01053
non-specific	238872	45	204	5.80E-15	69.2442	cd01834	SGNH_hydrolase_like_2	-	cl01053
non-specific	238870	44	202	9.01E-13	63.0564	cd01832	SGNH_hydrolase_like_1	-	cl01053
non-specific	239947	43	202	1.25E-10	57.2581	cd04506	SGNH_hydrolase_YpmR_like -	-	cl01053
non-specific	238868	44	205	3.94E-10	56.0961	cd01830	XynE_like	-	cl01053
non-specific	238865	48	205	2.20E-09	53.6025	cd01827	sialate_O-acetylesterase_like 1	-	cl01053
non-specific	238871	44	205	6.71E-09	51.8526	cd01833	XynB_like	-	cl01053
non-specific	238874	47	202	2.55E-08	50.727	cd01836	FeeA_FeeB_like	-	cl01053
non-specific	238869	45	206	2.98E-07	47.3384	cd01831	Endoglucanase_E_like	-	cl01053
non-specific	258744	44	206	1.15E-06	46.0866	pfam14606	Lipase_GDSL_3	-	cl01053
non-specific	238881	44	202	1.93E-06	45.3906	cd01844	SGNH_hydrolase_like_6	-	cl01053
non-specific	238861	44	112	2.41E-06	45.9067	cd01823	SEST_like	C	cl01053
non-specific	238879	42	199	1.22E-05	43.0868	cd01841	NnaC_like	-	cl01053
non-specific	238873	41	201	2.51E-05	42.3247	cd01835	SGNH_hydrolase_like_3	-	cl01053
non-specific	238858	48	209	0.000101	40.7355	cd01820	PAF_acetylesterase_like	-	cl01053
non-specific	238863	44	206	0.000122	40.335	cd01825	SGNH_hydrolase_peri1	-	cl01053
non-specific	238859	75	205	0.000204	39.8888	cd01821	Rhamnogalacturan_acetylesterase_like -	-	cl01053
non-specific	239946	72	198	0.000469	38.423	cd04502	SGNH_hydrolase_like_7	-	cl01053

	non-specific	238877	43	136	0.001479	37.2479	cd01839	SGNH_arylesterase_like	C	cl01053
	non-specific	212514	70	144	0.00359	36.3423	cd09988	Formimidoylglutamase	C	cl17011
	superfamily	266496	70	144	0.00359	36.3423	cl17011	Arginase_HDAC	C	-
Q#18 >gene_id_140897	non-specific	249771	287	498	1.42E-22	95.3622	pfam00326	Peptidase_S9	-	cl21494
	superfamily	271914	287	498	1.42E-22	95.3622	cl21494	Esterase_lipase	-	-
	non-specific	226651	254	355	0.003298	38.2012	COG4188	COG4188	C	cl21494
	multi-dom	223489	230	497	9.26E-13	66.6352	COG0412	COG0412	-	-
	multi-dom	224423	2	495	2.75E-50	182.344	COG1506	DAP2	N	-
	multi-dom	225989	243	373	4.68E-10	59.7601	COG3458	COG3458	NC	-
	multi-dom	253201	243	372	3.82E-08	53.8822	pfam05448	Axe1	NC	-
	multi-dom	223730	181	493	4.97E-07	50.3188	COG0657	Aes	-	-
	multi-dom	224422	139	473	8.54E-07	50.0848	COG1505	COG1505	N	-
	multi-dom	250827	288	493	1.09E-06	48.1031	pfam01738	DLH	-	-
	multi-dom	223999	241	493	0.000134	42.3898	COG1073	COG1073	-	-
	multi-dom	257229	288	373	0.00019	40.3932	pfam12695	Abhydrolase_5	C	-
	multi-dom	257231	280	482	0.006493	36.7581	pfam12697	Abhydrolase_6	-	-
	multi-dom	223506	234	372	0.007437	37.3202	COG0429	COG0429	C	-
Q#19 >gene_id_156964	non-specific	249771	123	211	0.000695	39.5083	pfam00326	Peptidase_S9	C	cl21494
	superfamily	271914	123	211	0.000695	39.5083	cl21494	Esterase_lipase	C	-
	non-specific	256874	98	163	0.000792	37.5337	pfam12146	Hydrolase_4	-	cl21494
	non-specific	226040	79	214	0.001771	38.5795	COG3509	LpqC	C	cl21494
	multi-dom	257231	107	357	5.56E-20	86.834	pfam12697	Abhydrolase_6	-	-
	multi-dom	225176	80	390	2.79E-12	65.4716	COG2267	PldB	-	-
	multi-dom	257229	106	352	5.98E-11	59.6532	pfam12695	Abhydrolase_5	-	-
	multi-dom	224423	80	351	6.06E-11	62.547	COG1506	DAP2	N	-
	multi-dom	223669	88	354	2.77E-10	59.6415	COG0596	MhpC	-	-
	multi-dom	223489	80	390	9.58E-09	54.3088	COG0412	COG0412	-	-
	multi-dom	249959	137	351	5.31E-08	51.7502	pfam00561	Abhydrolase_1	-	-
	multi-dom	225989	73	202	1.24E-06	48.5894	COG3458	COG3458	C	-
	multi-dom	251106	104	319	1.55E-06	47.7111	pfam02129	Peptidase_S15	-	-

	multi-dom	131480	130	351	1.57E-06	47.7365	TIGR02427	protocat_pcaD	-	-
	multi-dom	253201	73	202	4.11E-06	46.9486	pfam05448	Axe1	C	-
	multi-dom	223730	97	359	1.98E-05	44.5408	COG0657	Aes	N	-
	multi-dom	132144	78	351	0.000155	41.7216	TIGR03100	hydr1_PEP	-	-
	multi-dom	250827	91	205	0.000292	40.3992	pfam01738	DLH	C	-
	multi-dom	225496	80	218	0.000442	40.0811	COG2945	COG2945	C	-
	multi-dom	250827	300	352	0.000733	39.2436	pfam01738	DLH	N	-
	multi-dom	165193	101	351	0.001534	38.7135	PHA02857	PHA02857	-	-
	multi-dom	234107	98	232	0.002055	38.261	TIGR03101	hydr2_PEP	C	-
Q#20 >gene_id_167528	specific	213022	25	204	6.86E-64	198.612	cd10917	CE4_NodB_like_6s_7s	-	cl15692
	superfamily	271775	25	204	6.86E-64	198.612	cl15692	CE4_SF	-	-
	non-specific	234001	20	217	9.02E-55	175.604	TIGR02764	spore_ybaN_pdaB	-	cl15692
	non-specific	200582	25	213	5.06E-52	168.554	cd10959	CE4_NodB_like_3	-	cl15692
	non-specific	200584	25	218	1.65E-51	167.468	cd10962	CE4_GT2-like	-	cl15692
	non-specific	200578	25	216	1.46E-49	161.985	cd10954	CE4_CtAXE_like	-	cl15692
	non-specific	200580	22	218	4.55E-45	150.955	cd10956	CE4_BH1302_like	-	cl15692
	non-specific	200574	24	217	1.74E-44	148.963	cd10950	CE4_BsYlxY_like	-	cl15692
	non-specific	200571	25	213	5.06E-42	142.524	cd10947	CE4_SpPgdA_BsYjeA_like	-	cl15692
	non-specific	211966	27	213	1.48E-41	141.774	TIGR04243	nodulat_NodB	-	cl15692
	non-specific	200569	25	212	2.22E-40	138.449	cd10944	CE4_SmPgdA_like	-	cl15692
	non-specific	200573	22	217	2.21E-39	136.002	cd10949	CE4_BsPdaB_like	-	cl15692
	non-specific	200568	27	210	3.08E-38	133.048	cd10943	CE4_NodB	-	cl15692
	non-specific	250683	22	142	8.70E-37	126.967	pfam01522	Polysacc_deac_1	-	cl15692
	non-specific	200577	27	210	3.87E-36	127.305	cd10953	CE4_SlAXE_like	-	cl15692
	non-specific	200579	25	215	2.64E-34	122.81	cd10955	CE4_BH0857_like	-	cl15692
	non-specific	200575	22	213	3.93E-34	122.378	cd10951	CE4_ClCDA_like	-	cl15692
	non-specific	200572	17	213	1.07E-32	119.31	cd10948	CE4_BsPdaA_like	-	cl15692
	non-specific	200581	25	216	9.15E-29	107.767	cd10958	CE4_NodB_like_2	-	cl15692
	non-specific	200576	27	178	4.79E-21	86.6515	cd10952	CE4_MrCDA_like	-	cl15692
	non-specific	200589	25	137	2.92E-17	76.2629	cd10967	CE4_GLA_like_6s	C	cl15692

	non-specific	213023	27	141	3.09E-17	75.3254	cd10918	CE4_NodB_like_5s_6s	-	cl15692
	non-specific	200570	25	210	4.03E-15	70.1252	cd10946	CE4_Mll8295_like	-	cl15692
	non-specific	200565	43	122	5.39E-13	65.4919	cd10940	CE4_PuuE_HpPgdA_like_1	C	cl15692
	non-specific	200566	39	122	1.19E-12	63.8501	cd10941	CE4_PuuE_HpPgdA_like_2	C	cl15692
	non-specific	200583	25	215	2.90E-12	62.2474	cd10960	CE4_NodB_like_1	-	cl15692
	non-specific	213021	43	139	3.91E-10	56.5505	cd10916	CE4_PuuE_HpPgdA_like	C	cl15692
	non-specific	200563	43	140	2.76E-09	54.4855	cd10938	CE4_HpPgdA_like	C	cl15692
	non-specific	213028	25	139	4.53E-09	52.2743	cd10973	CE4_DAC_u4_5s	-	cl15692
	non-specific	213020	27	124	6.69E-09	51.6792	cd10585	CE4_SF	C	cl15692
	non-specific	213027	27	136	3.38E-08	50.3915	cd10970	CE4_DAC_u1_6s	C	cl15692
	non-specific	200600	42	138	8.40E-08	50.1441	cd10978	CE4_Sll1306_like	NC	cl15692
	non-specific	200599	43	126	2.56E-07	48.8613	cd10977	CE4_PuuE_SpCDA1	NC	cl15692
	non-specific	213024	24	141	9.13E-05	40.3385	cd10966	CE4_yadE_5s	C	cl15692
	non-specific	213026	24	153	0.000159	40.348	cd10969	CE4_Ecf1_like_5s	-	cl15692
	non-specific	200567	36	128	0.0003	39.3839	cd10942	CE4_u11	C	cl15692
	non-specific	200555	43	100	0.000303	39.5596	cd10929	CE4_u5	C	cl15692
	non-specific	200545	27	140	0.000491	38.8818	cd10919	CE4_CDA_like	C	cl15692
	non-specific	200594	27	124	0.000533	38.4669	cd10972	CE4_DAC_u3_5s	C	cl15692
	non-specific	200564	43	140	0.005354	35.7245	cd10939	CE4_ArnD	C	cl15692
	multi-dom	223798	22	218	5.96E-40	139.753	COG0726	CDA1	N	-
	multi-dom	131930	16	216	1.65E-34	124.047	TIGR02884	spore_pdaA	-	-
	multi-dom	131920	24	218	3.76E-31	116.122	TIGR02873	spore_ylxY	N	-
	multi-dom	213757	40	122	2.15E-08	51.9537	TIGR03006	pepcterm_polyde	C	-
Q#21 >gene_id_173248	specific	238871	41	228	4.90E-35	123.885	cd01833	XynB_like	-	cl01053
	superfamily	260768	41	228	4.90E-35	123.885	cl01053	SGNH_hydrolase	-	-
	non-specific	238860	42	229	1.55E-15	71.0048	cd01822	Lysophospholipase	-	cl01053

								_L1_like		
	non-specific	257796	46	224	2.57E-14	67.546	pfam13472	Lipase_GDSL_2	-	cl01053
	non-specific	238141	46	230	1.29E-13	65.8966	cd00229	SGNH_hydrolase	-	cl01053
	non-specific	238866	88	237	3.42E-13	64.2229	cd01828	sialate_O-acetylesterase_like 2	-	cl01053
	non-specific	250036	43	225	4.79E-12	61.9341	pfam00657	Lipase_GDSL	-	cl01053
	non-specific	225353	46	234	4.91E-09	53.3698	COG2755	TesA	-	cl01053
	non-specific	238874	97	226	8.19E-08	49.5715	cd01836	FeeA_FeeB_like	N	cl01053
	non-specific	238869	119	233	1.85E-07	48.1088	cd01831	Endoglucanase_E_like	N	cl01053
	non-specific	238872	97	228	5.68E-07	47.2878	cd01834	SGNH_hydrolase_like_2	N	cl01053
	non-specific	239946	114	236	7.06E-07	46.5122	cd04502	SGNH_hydrolase_like_7	N	cl01053
	non-specific	239945	97	232	8.41E-07	46.5512	cd04501	SGNH_hydrolase_like_4	N	cl01053
	non-specific	238870	117	226	1.76E-06	45.7224	cd01832	SGNH_hydrolase_like_1	N	cl01053
	non-specific	238881	106	232	2.30E-06	45.3906	cd01844	SGNH_hydrolase_like_6	N	cl01053
	non-specific	238879	97	236	7.70E-06	43.8572	cd01841	NnaC_like	-	cl01053
	non-specific	239947	117	225	9.52E-05	41.0797	cd04506	SGNH_hydrolase_YpmR_like	N	cl01053
	non-specific	238858	118	225	0.00012	40.7355	cd01820	PAF_acetylesterase_like	N	cl01053
	non-specific	238868	43	226	0.000236	39.9177	cd01830	XynE_like	-	cl01053
	non-specific	238877	66	234	0.002223	36.8627	cd01839	SGNH_arylesterase_like	-	cl01053
	multi-dom	223494	32	129	0.003764	37.0187	COG0417	PolB	C	-
	specific	257796	51	211	1.49E-21	88.732	pfam13472	Lipase_GDSL_2	-	cl01053
	superfamily	260768	51	211	1.49E-21	88.732	cl01053	SGNH_hydrolase -	-	SGNH_GDSL
Q#22 - >gene_id_248812	non-specific	238872	47	230	5.74E-20	84.6522	cd01834	SGNH_hydrolase_like_2	-	cl01053
	non-specific	238866	79	194	1.14E-18	80.4013	cd01828	sialate_O-acetylesterase_like 2	C	cl01053

non-specific	238141	48	260	8.84E-16	72.445	cd00229	SGNH_hydrolase	-	cl01053
non-specific	250036	47	258	5.73E-14	68.0973	pfam00657	Lipase_GDSL	-	cl01053
non-specific	239945	75	211	5.98E-13	64.6556	cd04501	SGNH_hydrolase_like_4	-	cl01053
non-specific	225353	39	262	1.45E-12	64.1554	COG2755	TesA	-	cl01053
non-specific	238876	51	260	3.96E-12	62.655	cd01838	Isoamyl_acetate_hydrolase_like	-	cl01053
non-specific	238860	63	262	3.15E-11	59.834	cd01822	Lysophospholipase_L1_like	-	cl01053
non-specific	238879	74	209	1.49E-09	55.028	cd01841	NnaC_like	-	cl01053
non-specific	238870	100	260	1.02E-07	49.5744	cd01832	SGNH_hydrolase_like_1	N	cl01053
non-specific	238873	67	194	2.85E-06	45.7915	cd01835	SGNH_hydrolase_like_3	-	cl01053
non-specific	238874	69	211	2.95E-06	45.7195	cd01836	FeeA_FeeB_like	-	cl01053
non-specific	238871	99	179	5.57E-06	44.1486	cd01833	XynB_like	NC	cl01053
non-specific	239947	47	115	0.002445	36.8425	cd04506	SGNH_hydrolase_YpmR_like	C	cl01053

1-C: Selected putative AcXE conserved domains excluding all non-AcXE and non-specific domain hits as well as hits with incomplete N and/or C termini.

S/N	Query	Hit type	PSSM-ID	From	To	Domain Length (aa)	E-Value	Bit score	Accession	Short name	Super family
1	***27345**	**multi-dom**	**253201**	**5**	**322**	318	**9.81E-124**	**360.886**	**pfam05448**	**Axe1**	**cl21494**
2	***39518**	**multi-dom**	**253201**	**5**	**321**	317	**2.01E-149**	**425.985**	**pfam05448**	**Axe1**	**cl21494**
3	***173248**	**Specific**	**238871**	**41**	**228**	188	**4.90E-35**	**123.885**	**cd01833**	**XynB_like**	**cl01053**
4	20333	Specific	257796	24	185	162	4.61E-24	95.2804	pfam13472	Lipase_GDSL_2	cl01053
5	41102	Specific	257796	72	261	190	1.05E-24	97.5916	pfam13472	Lipase_GDSL_2	cl01053
6	64061	Specific	238866	69	243	175	1.42E-68	211.369	cd01828	sialate_O-acetylesterase_like2	cl01053
7	248812	Specific	257796	51	211	161	1.49E-21	88.732	pfam13472	Lipase_GDSL_2	cl01053
8	9548	Specific	238141	72	241	170	6.48E-20	84.001	cd00229	SGNH_hydrolase	cl01053
9	167528	Specific	213022	25	204	180	6.86E-64	198.612	cd10917	CE4_NodB_like_6s_7s like	cl15692
10	**167528**	**non-specific**	**200578**	**25**	**216**	**192**	**1.46E-49**	**161.985**	**cd10954**	**CE4_CtAXE_like**	**cl15692**
11	**167528**	**non-specific**	**200577**	**27**	**210**	**184**	**3.87E-36**	**127.305**	**cd10953**	**CE4_SlAXE_like**	**cl15692**
12	7346	Specific	213022	121	293	173	4.33E-79	240.984	cd10917	CE4_NodB_like_6s_7s like	cl15692
13	**7346**	**non-specific**	**200577**	**123**	**296**	**174**	**1.47E-36**	**130.772**	**cd10953**	**CE4_SlAXE_like**	**cl15692**
14	**7346**	**non-specific**	**200578**	**121**	**305**	**185**	**1.76E-62**	**198.579**	**cd10954**	**CE4_CtAXE_like**	**cl15692**
15	7936	Specific	213022	25	204	180	1.82E-63	197.456	cd10917	CE4_NodB_like_6s_7s like	cl15692
16	**7936**	**non-specific**	**200578**	**26**	**216**	**191**	**6.77E-51**	**165.451**	**cd10954**	**CE4_CtAXE_like**	**cl15692**
17	**7936**	**non-specific**	**200577**	**27**	**211**	**185**	**6.41E-34**	**121.142**	**cd10953**	**CE4_SlAXE_like**	**cl15692**

In bold: 9 hits with domain models curated by well characterized proteins duly reported to express AcXE activity

***** 3 hits finally selected after further screening

1-D: Contigs containing three predicted AcXE ORFs.

Gene_ids 39518 (966bp), 27345 (990bp) and 173248 (735bp) from the Namib hypolith metagenomic dataset are designated *NaMet1, 2* and *3*, respectively.

NODE_124650_length_2665_cov_4.666417 GeneMark.hmm CDS 1045 2010 . + 0 gene_id 39518

>NODE_124650_length_2665_cov_4.666417

GCCGAGCCGGACGTCGTCGTGGCCAGCGGCCGCTCGACCTACTTCCCCACGGTGGAGCGGCTGCGCGGCGGCGAGCTGGTGGTCGTGTACTACGACGCGCC
CACCCACCTCGGCCCCGACGGGCGCATCGCCATGACCCGCAGCCGGGACGGCGGTCGCACGTGGAGCGCCCCGACGGTCGCGATCGACACGCCCATCGAC
GACCGCGACCCGTCGCTCGTCCAGCTGGCCGACGGCACGCTGCTGATGAGCTGGTTCAGGACCGACTGGTCGACCAGCCCGCTGCCCACCCCGGGCGGGGT
GTGGACGGCCCGGTCGCGCGACGGCGGCCGCAGCTGGTCGGAGCCGGTGCGCGCCCGGTCGGCCACCTACGGCGAGGGCGACTGGGCCATCCCGCCGGT
CGGCTACGCCACCCCGTGGCTCGCGACCTCGGCGAAGGTGCTGGAGCTGCCGGACCGCACGCTGCTGCTGCCCGTCTACGGGAACTCGCCGACCGACTCGC
GGTCGTCGATCACCGTGCTGCGCTCGCGGGACGGCGGCCGGACCTGGCCGGCGAGCGAGGAGGTCCACGTCGCCGCGCACCCGCCGAACAACGCCCTCAG
CGAGCCGGCCCTCGCGCTGCTCGGCGACGGATCGATCCAGATGGCCATCCGCTCGCTGAACGTCGGCTACTGGTCCTCCTCCCGCGACGGCGGCCGGACCT
GGACGCCCGCGACCACCGAGAACCGCTACCACGAGCAGGCGTCGGACCTGCTGCCCGTGCGCGAGGGCGGACGGCGGATCGTGCTGCACACCTGGGGTGA
CCGGTCGGGCGAGTTCGGTCCCGGCCGGCCCACGGTCGGCCGGTTGATCCTGCCGGACGGCACGCGGACGGCGCCAAAGGTGATCTACGCGGGCGGCGCC
AACGACGAGAGCTACCCGACCAGCGTCCAGGTCGCGCCCGGCACGTTCTTCACCGTCTACTACGACGCGGCACGCGGGATCATCGGCGGCACGTACAGCAGA
CTCTCCGAGTACCGGAAACCCTGACCCCGAGGAGCGCACGTGCCGCTGACGTTCGACCTGCCGTTCGAGGAGCTGCTCACCTACCCCGGGCGCACCCCCCGG
CCGGCCGACCACGACGAGTACTGGGACAGGGGCCTCGCCGACCTGGCGGCCGTGCCGGCCGACGTGGTGATCGAGCCCGCCGAGTTCACCACGCCGCTCG
CGCGGTGCTCGCACCTGTGGTTCACCGGGACCGGCGGCGTGCGCGTGCACGCGAAGCTGCTGCGGCCGGTGGCGCCGGTCGAGCCGCACCCGGCGCTGCT
GCAGTTCCACGGCTACACCGGCAACTCGGGCGACTGGTCGAGCAGGCTCCACTACGTGGCGCTCGGGTACACGGTCGCCGCGCTCGACTGTCGCGGCCAGG
CAGGCCTGTCGGTCGGCGAGGCGCCGGTGGAGAACTGGTCGATGGCGAGCTACCTGCTGCGCGGCATCGACGACGACGCGGCCGACAACCTGGCGCTGCG
GCACCTGTTCCTGGACACCGCCCGCCTGGCCCAGATCGTGCTGGCGATGGACGACGTGGACCCCGACCGGGTGGCGGCGACGGGCTACAGCCAGGGCGGC
GGGCTCACGCTCGCGTGTGCCGCGCTGGAGCCGCGCATCCGGCTGGCCGCACCGGTGTACCCGTTCCTGTGCGACTTCCGGCGGGCGTGGGAGATGGACCT
GGAGAAGGGCCCGTACAACGAGATCACCACGTACTTCCGGGCTCGGGACCCGCGGCACCTGCGCGAGGAGGAGATCTTCTCCCGGCTCGGCTACGTCGACGT
GCAGCACCTGGCGCCACGCGTGCGGGCCGAGGTGCTCATGACCGTGTCCCTCGCCGACAAGATCTGTCCGCCGTCCACCCAGTTCGCCGCGTACAACAAGCT
GGGCGGCCCCAAGGACTACCGGCTGTACCCAGACTTCGCCCATGAGACCCTGCCCGGCACCGACGACGCGATCTTCACGTTCCTGCAAGGGTTGTAGGCCTC
CGCACATTTCCGGCACAGCCACATGTACATGGAACGGAGTTTCACATATAACGTTTCAAGGGGGACGTGATGCAAGGAGGCACACGATGACCCGACCACGACG
GCATCCCTGGCTGTTAGCCGCGCTCGTCCTACCCCTGCTCACCGCTCCGGCCCCCGCGGCCGCCGACAGCGTGACCGTCCCACCCGGACAGATCTTCTACAA
CGCCACGTCCACCGAGCCGTTCGCCGGCTACCCCGCGGTGTTCGACGCCGGCTCGACGGCGGCCGACGGCACCCGCAGCCAGCGGATCATGTTGACCTACA
ACGCCAAGCAGGACCTGCCCGGCGAGAAGGGCAGCCGGGTCCGGCAGTCGACCGACGGCGGGGTGACCTGGCCGACCTACGAGTCGCACGACGGGACCAT
CCAGGCGTCCAACATGGTCCGGCTGCCGATCGCGTCCGGCGGCGTGATCACCATCAACTACGAGGATCTGAGCGTCACCAACCCGAGCGACCCGCTGTGCAG
CTCGGAGCCACCCTGCCGCCACGTGTTCAACCGCTGGATGATGACCCCGACGACGTTCGTCGACGCGGGCACCGCGCACGTGACCTTCCTGCCGGGCAAAGC
CATCGCCTGGGCCAGGTTCGGCCAGGGACCTATCGTGCTGGACGACGGCAAGACCATGCTGTCGGCCATGTACGGCGTGCGCAACGGCAC

NODE_84744_length_1851_cov_6.385197 GeneMark.hmm CDS 288 1277 . - 0 gene_id 27345

>NODE_84744_length_1851_cov_6.385197

CTGGCGCTGGTGGCGATGGGTGTGGTGAACCTGGCGTTCCAGGCCATGGCGAACTCGTCGGTGCAGTCGTGGGTGGACCCGGAGGTCCGCGGGCGGGTCAT
GGGGCTCTACATGCTCGCGTTCGTCGGCGGCACGCCGATCGGCGGCCCGGCGGTCGGCTGGGTCACCGGTGCGTTCGGCGCCCGGGCCGGCATGGCCGTG
TGCGGCCTCATCCCGCTGCTCGCCGCGGCGGTGGTGGCGATCGTGCTCGCCCGGCAGGTCCGCGCGGACCGGGCGGGCCGGGCTCGTCAGGATCGGCCCA
GCACGCCGTTCAGCCAACGGATCTGGGCCTGGCGCTGGTAGGACTCGCCGCCCTCGTGGTCGTTGTAGTGGTAGACCGCGATGTCCGTGCCGATGTCGGCGC
GCCCGCCCGCGTAGTGGTTGTACGCCGCGTACACGGTCGAGGGCGGGCAGACCAGGTCCATCAGCGCCACCGAGAACAGGGTCGGCGCCGTGGCGCGCCT
GGCGAAGTTGACCGCGTCGAAGTACGACAGGGTGCGCAGTACGTCGTCCACCCGGGGGCGGTGCACCGCCAGGTAGCGCGTGATCTCCGCGAACGGATCCC
TGTCGCAGACGTCCAGCGCCCTGGCGAAGTGGCAGAGGAACGGGACGTCGGGCATGACCGCCACGAGATCGGGCACCAGCCCGGCTACCGCGAGGGTGATC
CCGCCGCCCTGGCTGATGCCCTGCACGGCCACCCGGGCTGGATCGACGCCGGGCAGCGTGCGCACCGCCTCGACGGCGCGCACCGCATCGGTGAACACGC
GGCGGTAGAAGTAGCTGTCCGGGTGCCGGATGCCGCGCGTCATGAACCCGGGCGTGCTGGGCTCCGAGCCGTGCGGGTCGGCCGTGCTGCCTCCGCCGCC
CCACCGGGCGCCTTGACCCCGGGTGTCCACCAGCAGGTGCGCGTACCCCGCGCTGGCCCACCGCAGCGCCTCGCCGGGCAGGTCGCGCCCGCCGTTGTAG
CCGATGAACTCCACGACGGCGGGCAGCGGGACGTCCTCTGCCCCCGGCACCGTGAACCACGCGCCGATCCGGTCACCGTCGAACCCGCGGAAGCTGACGTC
CTGCACCGCGACGGTCCGGTACGGCGTGACCGCCGCCTCGGCGCGGACGTCCAGGTCGTGTGCCCGGGCCTCGGCGAGCGTGCGGTTCCAGAACAGGTCG
AAGTCGTCGGGTTCGCGGACGACCGGGCGGTAGGCGACGAGCTCGTCAAGGGGCAGATCGGCGCGGGCCACGAATCTCCTGTCCTCGGGTGGTCGGGCCG
GGAGAGGATCCCACGGGTGCCGTCAGAGATCGGCCAGAGCGCTGTGCAGGTCGGGGTGCAGCGGGAGGATCTGCTGCAGGTTCATCGTCTCCCACGGTCGG
ACGACGGCGTTGTTGGTGGGTGGTGCCACCAGCCGCAGTGGCGGGAGCGGCTTCGGCGGCTCCGACCGGTAACCCGGGCCGGTGGGCGCGGTGGTGCGGG
TGGCCAGTTCGGCGAGCGTGCCGAGCCCCGAGGAGCCGAGGAAGCTCACCGCGGTCAGGTCGACGACGACGGCGCTGAGATCCGGCCACTGCAGCGTCTCG
TCGACCATCGACATCAGGTTCCGCGCGGTGTGCAGGTCCAGCTCGCCGGTCACCGTCACCACCGCCGCCGAGTCGACGCGGGTGGCGGCCAGCGTCATCAC
GTCGGCGGGCGCGGCGGGCGCATCCGAGGCGCTGCGGTCGCTGTCCTTCATCGGGGCACCGTATCCGGCGCACCCGTGCGGCCGCGACGCCACGGGGGCC
GGATCTGCTACGGGCGGTAGGACAGCGGGCCCCGGGTAACCCGTCCGGGGGGCCTGCGCCCGGAGCGGTGTAGCGGCAGGGGCCCTCGGGGT

>#NODE_84744_length_1851_cov_6.385197_Reverse_complement

ACCCCGAGGGCCCCTGCCGCTACACCGCTCCGGGCGCAGGCCCCCCGGACGGGTTACCCGGGGCCCGCTGTCCTACCGCCCGTAGCAGATCCGGCCCCCGT
GGCGTCGCGGCCGCACGGGTGCGCCGGATACGGTGCCCCGATGAAGGACAGCGACCGCAGCGCCTCGGATGCGCCCGCCGCGCCCGCCGACGTGATGACG
CTGGCCGCCACCCGCGTCGACTCGGCGGCGGTGGTGACGGTGACCGGCGAGCTGGACCTGCACACCGCGCGGAACCTGATGTCGATGGTCGACGAGACGCT
GCAGTGGCCGGATCTCAGCGCCGTCGTCGTCGACCTGACCGCGGTGAGCTTCCTCGGCTCCTCGGGGCTCGGCACGCTCGCCGAACTGGCCACCCGCACCA
CCGCGCCCACCGGCCCGGGTTACCGGTCGGAGCCGCCGAAGCCGCTCCCGCCACTGCGGCTGGTGGCACCACCCACCAACAACGCCGTCGTCCGACCGTGG
GAGACGATGAACCTGCAGCAGATCCTCCCGCTGCACCCCGACCTGCACAGCGCTCTGGCCGATCTCTGACGGCACCCGTGGGATCCTCTCCCGGCCCGACCA
CCCGAGGACAGGAGATTCGTGGCCCGCGCCGATCTGCCCCTTGACGAGCTCGTCGCCTACCGCCCGGTCGTCCGCGAACCCGACGACTTCGACCTGTTCTGG
AACCGCACGCTCGCCGAGGCCCGGGCACACGACCTGGACGTCCGCGCCGAGGCGGCGGTCACGCCGTACCGGACCGTCGCGGTGCAGGACGTCAGCTTCC
GCGGGTTCGACGGTGACCGGATCGGCGCGTGGTTCACGGTGCCGGGGGCAGAGGACGTCCCGCTGCCCGCCGTCGTGGAGTTCATCGGCTACAACGGCGG
GCGCGACCTGCCCGGCGAGGCGCTGCGGTGGGCCAGCGCGGGGTACGCGCACCTGCTGGTGGACACCCGGGGTCAAGGCGCCCGGTGGGGCGGCGGAGG
CAGCACGGCCGACCCGCACGGCTCGGAGCCCAGCACGCCCGGGTTCATGACGCGCGGCATCCGGCACCCGGACAGCTACTTCTACCGCCGCGTGTTCACCG
ATGCGGTGCGCGCCGTCGAGGCGGTGCGCACGCTGCCCGGCGTCGATCCAGCCCGGGTGGCCGTGCAGGGCATCAGCCAGGGCGGCGGGATCACCCTCGC
GGTAGCCGGGCTGGTGCCCGATCTCGTGGCGGTCATGCCCGACGTCCCGTTCCTCTGCCACTTCGCCAGGGCGCTGGACGTCTGCGACAGGGATCCGTTCGC

GGAGATCACGCGCTACCTGGCGGTGCACCGCCCCCGGGTGGACGACGTACTGCGCACCCTGTCGTACTTCGACGCGGTCAACTTCGCCAGGCGCGCCACGG
CGCCGACCCTGTTCTCGGTGGCGCTGATGGACCTGGTCTGCCCGCCCTCGACCGTGTACGCGGCGTACAACCACTACGCGGGCGGGCGCGCCGACATCGGC
ACGGACATCGCGGTCTACCACTACAACGACCACGAGGGCGGCGAGTCCTACCAGCGCCAGGCCCAGATCCGTTGGCTGAACGGCGTGCTGGGCCGATCCTGA
CGAGCCCGGCCCGCCCGGTCCGCGCGGACCTGCCGGGCGAGCACGATCGCCACCACCGCCGCGGCGAGCAGCGGGATGAGGCCGCACACGGCCATGCCG
GCCCGGGCGCCGAACGCACCGGTGACCCAGCCGACCGCCGGGCCGCCGATCGGCGTGCCGCCGACGAACGCGAGCATGTAGAGCCCCATGACCCGCCCGC
GGACCTCCGGGTCCACCCACGACTGCACCGACGAGTTCGCCATGGCCTGGAACGCCAGGTTCACCACACCCATCGCCACCAGCGCCAG

NODE_669622_length_1436_cov_4.759749 GeneMark.hmm CDS 100 834 . - 0 gene_id 173248

>NODE_669622_length_1436_cov_4.759749
AGGCCTACCTTGAGGCCGTGGAGCCGGTGCCGGAGGCCGCGGCGCTGCCCGCGGCGCGGGCCGCCGCCCAGGAGCGGCTGGCCGCCCTGCGGGCCGAGG
TCACCGCGCTGCTGGGCGCGTGACCTGCAGGTACGCCCGGATCTCCCGCTCCCACATCGCGGCGATGCGGCGGTAGCCCTCGGCGTTGGGGTGCAGCCCGT
CGGCAAGCTCGTCGGGGGCCAGCAGCTCGGAGGTGTCCACGAACGTGGCGGAGCGCCCGCGCGCCCGCTGCTCGTCGACGACCACCCGACTCAGCCGGGT
GAACTCCGCCCGGTCCCGCTGGTGGCCGGGTAGCGGCGCCCACACGCCCGCGACGACGACGTGCGCGTCGGACACGCCGTAGATGGCGTCCAGCACTGCG
GTGAGCCGTTCGGCCGTTTCCGCGGCCGAGGCGCCCAGCAGCAGGTCGTTGGTGCCGTTGTGGAGCAGGACGACGTCCGGCTGCGCCTGCCGCACCCATGC
GGCGACGCGGTCCTCCATCGCCTCCAGCGTGATGCCGCTGCGGCCCTCGTGATCGCGGTCCGGCACGGAGTCGGGGCCGTCGCGCAGCGACCCGACGAGAT
CGACGGCCACCCCGTCGCGGGCGAGCAGCTCCTGCAGCGGCCCCCGGTAGCCCGCCGTCTCCGGGCTGCCCTTCCCCACCGTGCTGGACGCGCCGAGCGG
CATGACCCGCAGCACCCCGGGCACGGTCTCCGCGCTGCGCTGGACCGGAACGGCGCCCGAGCGGACGCCCGCCAGCACCGTGACCACGCAGGCCAGCACC
AGCAGGGCTCGTCCGAACCGGATCGTTCCCACCCGTCCCCCTCAGGCCAGCGCCGCGGCCAGCGTATCGAGCGCGCGGGCCGGCCGGGCGTCGAGCGCCA
TGCGCGGCGCAGGCGAGCCCCGCGGCGTACTCGGCGCGGAAGGCGGCGACCGGCTCCGCGGGATCGTCCAGGACGGAACTTCTGCCGGGCACCGGCCTCG
CGCGCGCTCGCCGGCGTGGCAGCATCCGCCACGTGGCCCGACTCCTGATCGTCCACCACACGCCGTCGCCCGCGACGGCGGAGCTCCTCGACGCCGTCGTG
CGCGGCGCGTCCGACCCGGAGATCGTCGGCGTCGAAGTGGTGCGCCGGGCCGCGCTCGCCGCCACGGCGTCGGACCTGCTCGCCGCCGACGCGGTCGTGA
TCGGCACGCCGGCCAACATCGGCTACATGTCGGGAGCCGTCAAACACTTCTTCGACACGGTCTACTACGTGTGCCGTGACGACACCCGCGGGTTGCCCTACGG
CCTCTACGTGCACGGCAACCTGGGGGCCGACGGTGCCGTCCGCGCCGTCGGCTCGATCGCCGAGGGCATGGGCTGGACCCGCGCTGCTGCACCGGTCGAG
GTCACCGGCGCGCCGGACCGCGCCGCCCTGGAGGCGTGCTGGGAGCTGGGCGCCGTCGTCGCCGCGGGCATCGCCCCCGA

>NODE_669622_length_1436_cov_4.759749_Reverse_complement

TCGGGGGCGATGCCCGCGGCGACGACGGCGCCCAGCTCCCAGCACGCCTCCAGGGCGGCGCGGTCCGGCGCGCCGGTGACCTCGACCGGT
GCAGCAGCGCGGGTCCAGCCCATGCCCTCGGCGATCGAGCCGACGGCGCGGACGGCACCGTCGGCCCCCAGGTTGCCGTGCACGTAGAGGC
CGTAGGGCAACCCGCGGGTGTCGTCACGGCACACGTAGTAGACCGTGTCGAAGAAGTGTTTGACGGCTCCCGACATGTAGCCGATGTTGGCCG
GCGTGCCGATCACGACCGCGTCGGCGGCGAGCAGGTCCGACGCCGTGGCGGCGAGCGCGGCCCGGCGCACCACTTCGACGCCGACGATCTC
CGGGTCGGACGCGCCGCGCACGACGGCGTCGAGGAGCTCCGCCGTCGCGGGCGACGGCGTGTGGTGGACGATCAGGAGTCGGGCCACGTG
GCGGATGCTGCCACGCCGGCGAGCGCGCGCGAGGCCGGTGCCCGGCAGAAGTTCCGTCCTGGACGATCCCGCGGAGCCGGTCGCCGCCTTC
CGCGCCGAGTACGCCGCGGGGCTCGCCTGCGCCGCGCATGGCGCTCGACGCCCGGCCGGCCCGCGCGCTCGATACGCTGGCCGCGGCGCT
GGCCTGAGGGGGACGGGTGGGAACGATCCGGTTCGGACGAGCCCTGCTGGTGCTGGCCTGCGTGGTCACGGTGCTGGCGGGCGTCCGCTCG
GGCGCCGTTCCGGTCCAGCGCAGCGCGGAGACCGTGCCCGGGGTGCTGCGGGTCATGCCGCTCGGCGCGTCCAGCACGGTGGGGAAGGGCA
GCCCGGAGACGGCGGGCTACCGGGGGCCGCTGCAGGAGCTGCTCGCCCGCGACGGGGTGGCCGTCGATCTCGTCGGGTCGCTGCGCGACG
GCCCCGACTCCGTGCCGGACCGCGATCACGAGGGCCGCAGCGGCATCACGCTGGAGGCGATGGAGGACCGCGTCGCCGCATGGGTGCGGCA
GGCGCAGCCGGACGTCGTCCTGCTCCACAACGGCACCAACGACCTGCTGCTGGGCGCCTCGGCCGCGGAAACGGCCGAACGGCTCACCGCAG
TGCTGGACGCCATCTACGGCGTGTCCGACGCGCACGTCGTCGTCGCGGGCGTGTGGGCGCCGCTACCCGGCCACCAGCGGGACCGGGCGGA
GTTCACCCGGCTGAGTCGGGTGGTCGTCGACGAGCAGCGGGCGCGCGGGCGCTCCGCCACGTTCGTGGACACCTCCGAGCTGCTGGCCCCC
GACGAGCTTGCCGACGGGCTGCACCCCAACGCCGAGGGCTACCGCCGCATCGCCGCGATGTGGGAGCGGGAGATCCGGGCGTACCTGCAGG
TCACGCGCCCAGCAGCGCGGTGACCTCGGCCCGCAGGGCGGCCAGCCGCTCCTGGGCGGCGGCCCGCGCCGCGGGCAGCGCCGCGGCCTC
CGGCACCGGCTCCACGGCCTCAAGGTAGGCCT

Note: Red and blue fonts indicate selected ORFs on leading (+) and lagging (-) strand contigs respectively. The reverse complements of lagging strand (-) contigs are shown.

1-E: Translated amino acid sequences of three selected putative AcXEs.

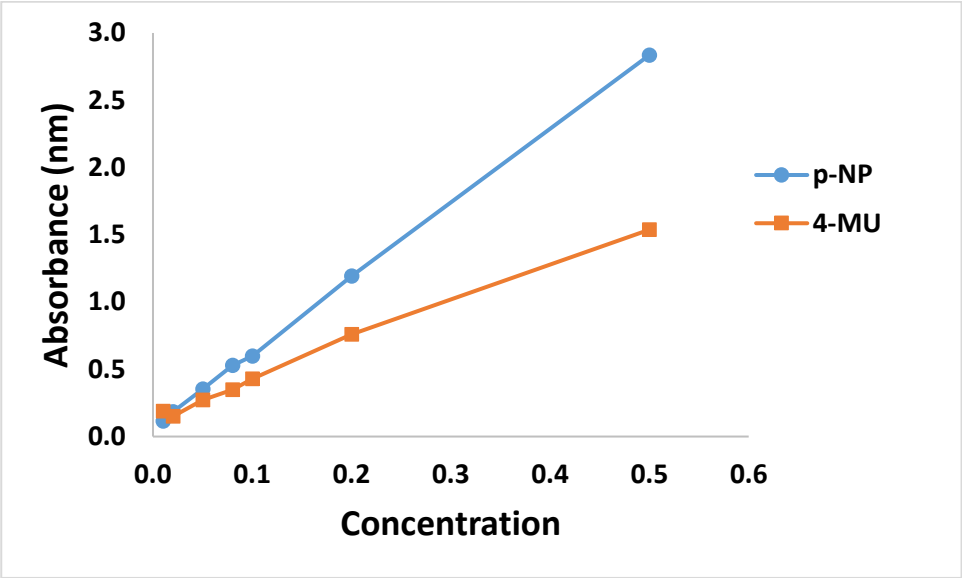
ORF_ID	Amino acid (aa) Sequence
gene_id 39518 (321aa)	VPLTFDLPFEELLTYPGRTPRPADHDEYWDRGLADLAAVPADVVIEPAEFTTPLARCSHLWFTGTGGVRVHAKLLRPVAPVEPHPALLQFHGYTGN SGDWSSRLHYVALGYTVAALDCRGQAGLSVGEAPVENWSMASYLLRGIDDDAADNLALRHLFLDTARLAQIVLAMDDVDPDRVAATGYSQGGGL TLACAALEPRIRLAAPVYPFLCDFRRAWEMDLEKGPYNEITTYFRARDPRHLREEEIFSRLGYVDVQHLAPRVRAEVLMTVSLADKICPPSTQFAAY NKLGGPKDYRLYPDFAHETLPGTDDAIFTFLQGL
gene_id2 7345 (329aa)	VARADLPLDELVAYRPVVREPDDFDLFWNRTLAEARAHDLDVRAEAAVTPYRTVAVQDVSFRGFDGDRIGAWFTVPGAEDVPLPAVVEFIGYNGG RDLPGEALRWASAGYAHLLVDTRGQGARWGGGGSTADPHGSEPSTPGFMTRGIRHPDSYFYRRVFTDAVRAVEAVRTLPGVDPARVAVQGISQ GGGITLAVAGLVPDLVAVMPDVPFLCHFARALDVCDRDPFAEITRYLAVHRPRVDDVLRTLSYFDAVNFARRATAPTLFSVALMDLVCPPSTVYAAY NHYAGGRADIGTDIAVYHYNDHEGGESYQRQAQIRWLNGVLGRS
gene_id 173248 (244aa)	VGTIRFGRALLVLACVVTVLAGVRSGAVPVQRSAETVPGVLRVMPLGASSTVGKGSPETAGYRGPLQELLARDGVAVDLVGSLRDGPDSVPDRD HEGRSGITLEAMEDRVAAWVRQAQPDVVLLHNGTNDLLLGASAAETAERLTAVLDAIYGVSDAHVVVAGVWAPLPGHQRDRAEFTRLSRVVVDE QRARGRSATFVDTSELLAPDELADGLHPNAEGYRRIAAMWEREIRAYLQVTRPAAR

1-F: Codon-optimised NaMet1-3 including EcoR1 and Nde1 restriction sites (red font).

ORF	Sequence
NaMet1 (978bp)	GAATTCGTCCCGCTGACCTTTGATCTGCCGTTTGAAGAACTGCTGACCTATCCGGGCCGCACCCCGCGTCCGGCTGACCACGATGAATACT GGGATCGTGGCCTGGCAGATCTGGCCGCGGTGCCGGCTGATGTGGTTATTGAACCGGCCGAATTTACCACGCCGCTGGCACGTTGCAGTC ATCTGTGGTTTACCGGTACGGGCGGTGTCCGTGTGCACGCAAAACTGCTGCGCCCGGTGGCTCCGGTTGAACCGCATCCGGCGCTGCTGC AGTTTCACGGTTATACCGGCAACAGCGGTGATTGGAGCTCTCGTCTGCATTATGTTGCCCTGGGCTACACCGTGGCAGCTCTGGATTGTCGT GGTCAAGCGGGTCTGTCTGTCGGTGAAGCACCGGTGGAAAACTGGAGTATGGCTTCCTATCTGCTGCGTGGTATTGATGACGATGCGGCCG ACAATCTGGCCCTGCGTCACCTGTTCCTGGATACCGCACGCCTGGCTCAGATCGTTCTGGCCATGGACGATGTTGACCCGGATCGCGTCGC AGCTACCGGTTATTCCCAAGGCGGTGGCCTGACCCTGGCATGTGCCGCACTGGAACCGCGTATTCGCCTGGCAGCTCCGGTGTACCCGTT TCTGTGTGACTTCCGTCGCGCATGGGAAATGGATCTGGAAAAAGGCCCGTATAACGAAATTACCACGTACTTTCGCGCGCGCGACCCGCGT CATCTGCGCGAAGAAGAAATCTTCAGCCGTCTGGGTTATGTTGATGTTCAGCACCTGGCGCCGCGTGTCCGTGCCGAAGTGCTGATGACCG TTTCACTGGCGGATAAAATCTGCCCGCCGTCGACGCAATTTGCGGCCTACAATAAACTGGGTGGCCCGAAAGACTATCGCCTGTACCCGGA TTTCGCACACGAAACCCTGCCGGGCACCGATGACGCTATCTTCACGTTCCTGCAAGGTCTGTAACTCGAG
NaMet2 (1002bp)	GAATTCGTGGCGCGTGCCGATCTGCCGCTGGACGAACTGGTTGCGTACCGTCCGGTGGTTCGCGAACCGGATGACTTTGACCTGTTTTGGA ACCGTACCCTGGCAGAAGCTCGCGCGCATGATCTGGATGTCCGTGCAGAAGCCGCGGTGACCCCGTATCGTACCGTTGCCGTCCAAGATG TGAGCTTTCGTGGTTTCGATGGCGACCGCATTGGTGCATGGTTTACCGTTCCGGGTGCTGAAGATGTCCCGCTGCCGGCCGTCGTGGAATT TATCGGTTACAATGGCGGTCGTGACCTGCCGGGTGAAGCTCTGCGTTGGGCAAGCGCAGGTTATGCACATCTGCTGGTTGATACCCGTGGT CAGGGTGCCCGTTGGGGCGGTGGCGGTTCTACCGCCGACCCGCATGGCAGCGAACCGTCTACCCCGGGTTTTATGACGCGTGGCATTCG CCACCCGGATAGTTATTTTTACCGTCGCGTGTTCACCGACGCCGTTCGTGCGGTTGAAGCTGTTCGCACCCTGCCGGGCGTTGATCCGGCG CGCGTGGCCGTTCAAGGTATTTCACAGGGCGGTGGCATCACCCTGGCTGTGGCCGGTCTGGTTCCGGATCTGGTCGCCGTGATGCCGGAC GTCCCGTTTCTGTGCCATTTCGCCCGTGCACTGGATGTTTGTGATCGCGACCCGTTTGCCGAAATCACCCGCTACCTGGCAGTGCACCGTC CGCGCGTTGATGACGTCCTGCGCACGCTGTCCTATTTTGATGCAGTTAATTTTGCTCGTCGCGCTACCGCCCCGACCCTGTTCAGTGTGGC GCTGATGGATCTGGTTTGCCCGCCGTCCACCGTCTATGCAGCTTACAATCATTATGCGGGTGGCCGTGCCGATATTGGCACGGACATCGCC GTGTATCATTACAACGATCACGAAGGTGGCGAATCATATCAGCGTCAAGCGCAGATTCGCTGGCTGAATGGTGTTCTGGGCCGCTCGTAACT CGAG
NaMet3 (747bp)	GAATTCGTTGGCACCATTCGCTTTGGTCGCGCCCTGCTGGTGCTGGCATGTGTGGTTACCGTCCTGGCGGGCGTGCGCAGCGGTGCAGTT CCGGTCCAGCGTTCTGCCGAAACCGTCCCGGGCGTGCTGCGTGTTATGCCGCTGGGTGCGAGCTCTACCGTTGGCAAAGGTAGCCCGGAA ACCGCGGGTTATCGTGGTCCGCTGCAAGAACTGCTGGCACGTGATGGCGTGGCTGTTGACCTGGTCGGCAGTCTGCGTGATGGTCCGGAC TCCGTGCCGGATCGCGACCATGAAGGCCGTTCTGGTATTACCCTGGAAGCAATGGAAGATCGCGTTGCGGCCTGGGTCCGTCAGGCTCAA CCGGATGTCGTGCTGCTGCACAACGGCACGAATGACCTGCTGCTGGGTGCAAGTGCAGCTGAAACCGCTGAACGCCTGACGGCGGTGCTG GATGCCATCTACGGCGTTTCCGACGCGCATGTTGTTGTGGCAGGCGTTTGGGCTCCGCTGCCGGGTCACCAGCGTGATCGTGCAGAATTTA CCCGCCTGAGCCGTGTTGTCGTGGATGAACAACGTGCGCGCGGTCGTTCAGCCACCTTCGTGGATACGTCGGAACTGCTGGCACCGGATG AACTGGCTGACGGCCTGCATCCGAACGCCGAAGGTTATCGTCGCATTGCGGCCATGTGGGAACGCGAAATTCGCGCATACCTGCAGGTGA CCCGTCCGGCAGCTCGTTAACTCGAG

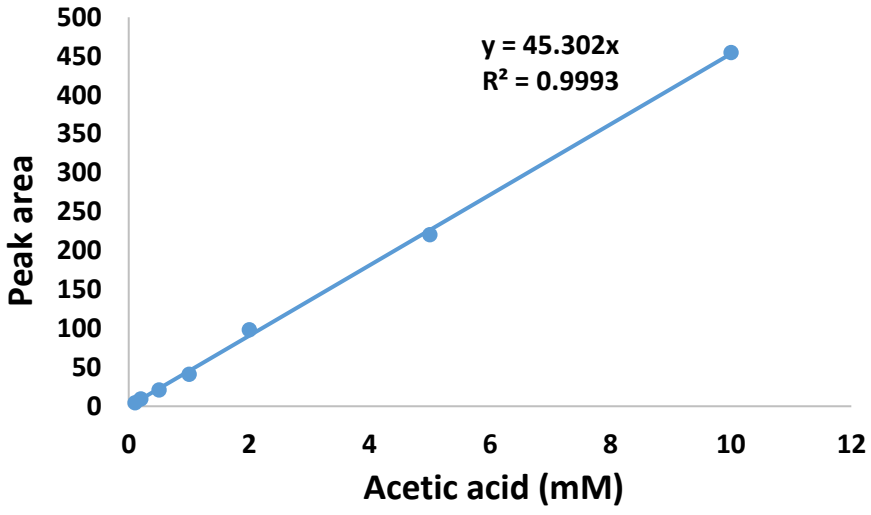
APPENDIX 2: Functional and structural characterization

2-A: Standard curves for products detected during assays.



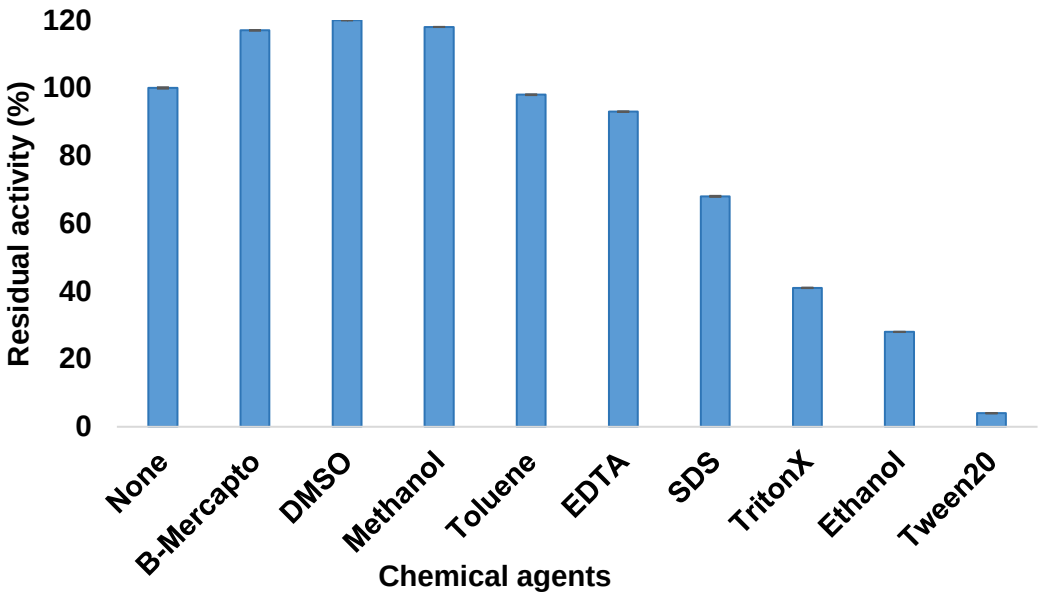
2A-1: Standard curve indicating absorbance of *p*-NP and 4-MU at 354 and 405 nm, respectively and at pH 8.

Error bars are barely visible.



2A-2: HPLC acetic acid calibration curve.

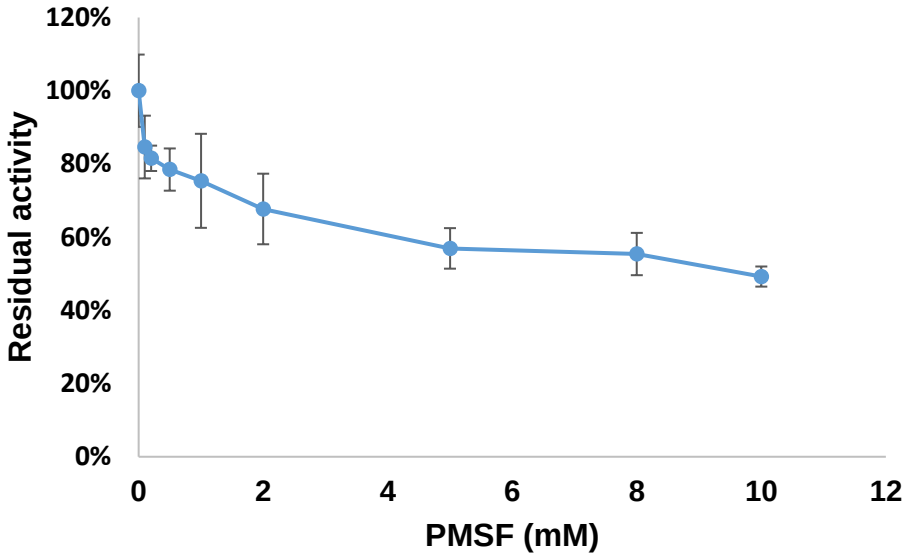
Error bars are barely visible.



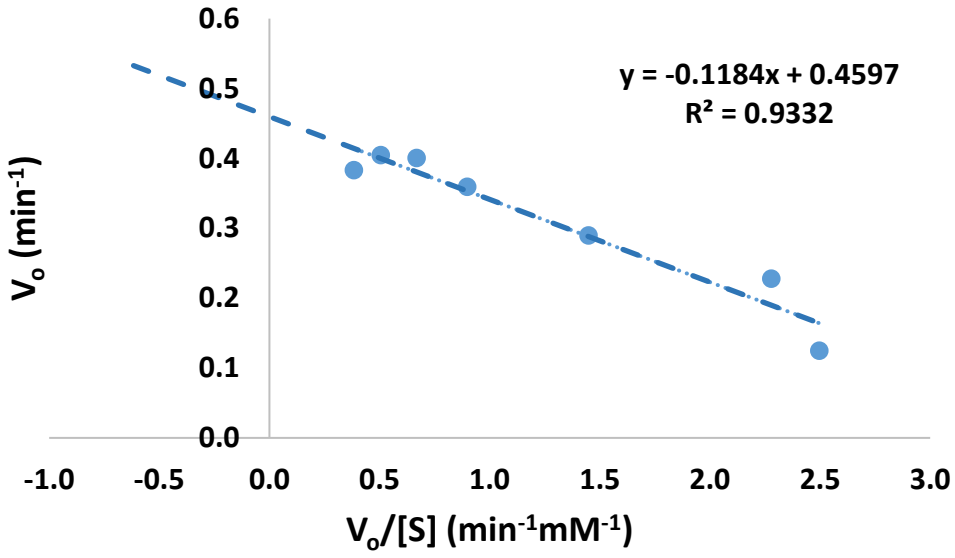
2-B: Influence of chemical agents on NaM1.

1% organic solvents (di-methyl sulphoxide, ethanol, methanol and toluene), 1% non-ionic detergents (triton X-100, Tween 20), ionic detergent (2 mM sodium dodecyl sulphate), 2mM ethylenediamine tetraacetic acid and 2 mM β-mercaptoethanol.

Error bars are barely visible.



2-C: PMSF inhibition of NaM1.

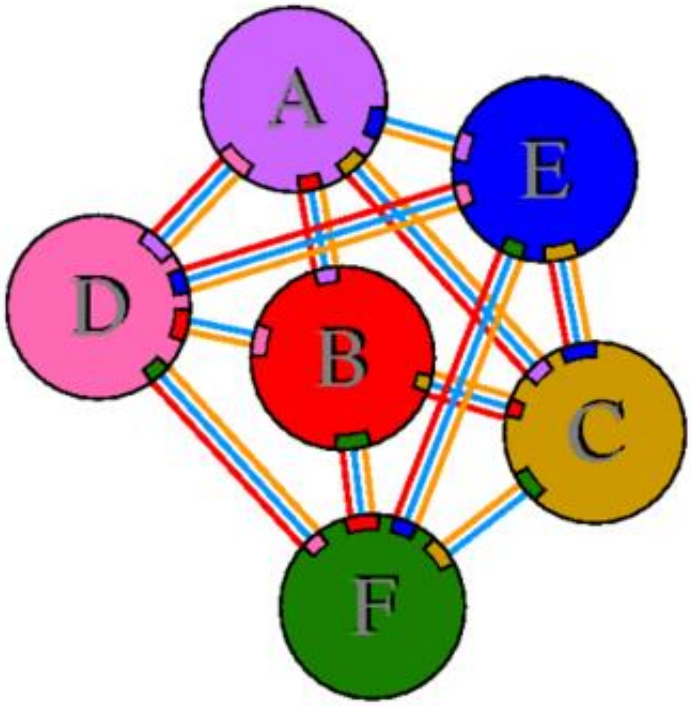


2-D: Eadie-Hofstee plot showing the relationship between initial velocity (Vo) of NaM1 activity and p-NPA substrate concentrations [S].

The slope is $-K_M$ and the y-intercept is V_{max}.

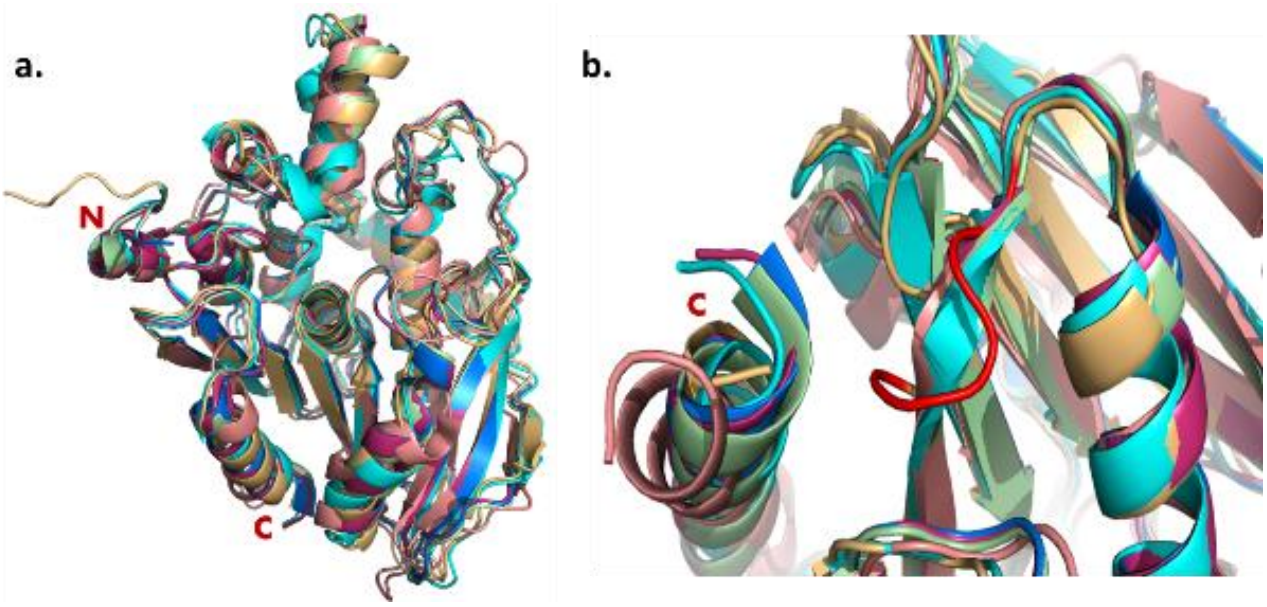
2-E: Chemical composition of NaM1.

Chain	Component	Quantity	Resolved Histidine(s) of His_6-tag
A	Amino acids	324	3
B	Amino acids	322	1
C	Amino acids	319	-
D	Amino acids	322	1
E	Amino acids	322	1
F	Amino acids	323	2
G	Na^+	6	-
K	MES (head group)	15	-
L	$CHOO^-$	1	-
S	H_2O	2238	-



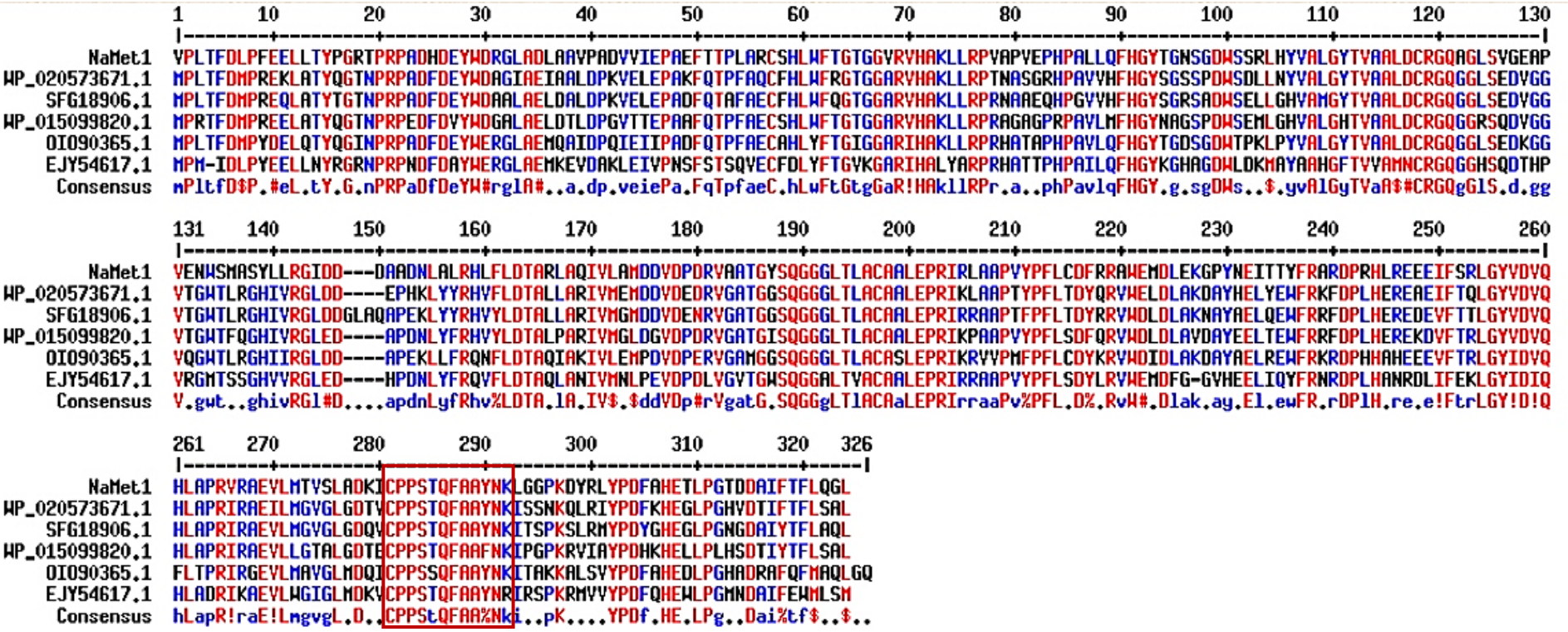
2-F: Schematic diagram of interactions between NaM1 subunits.

Interacting subunits are linked by colored lines, each representing a different type of interaction. Salt bridges, hydrogen bonds and non-bonded contacts are indicated by red, blue and orange colors respectively. A colored wedge on each subunit corresponding to the color of the other subunit indicates the degree of interaction and size of the interface area. Diagram was generated by PDBsum (de Beer, Berka et al. 2014)



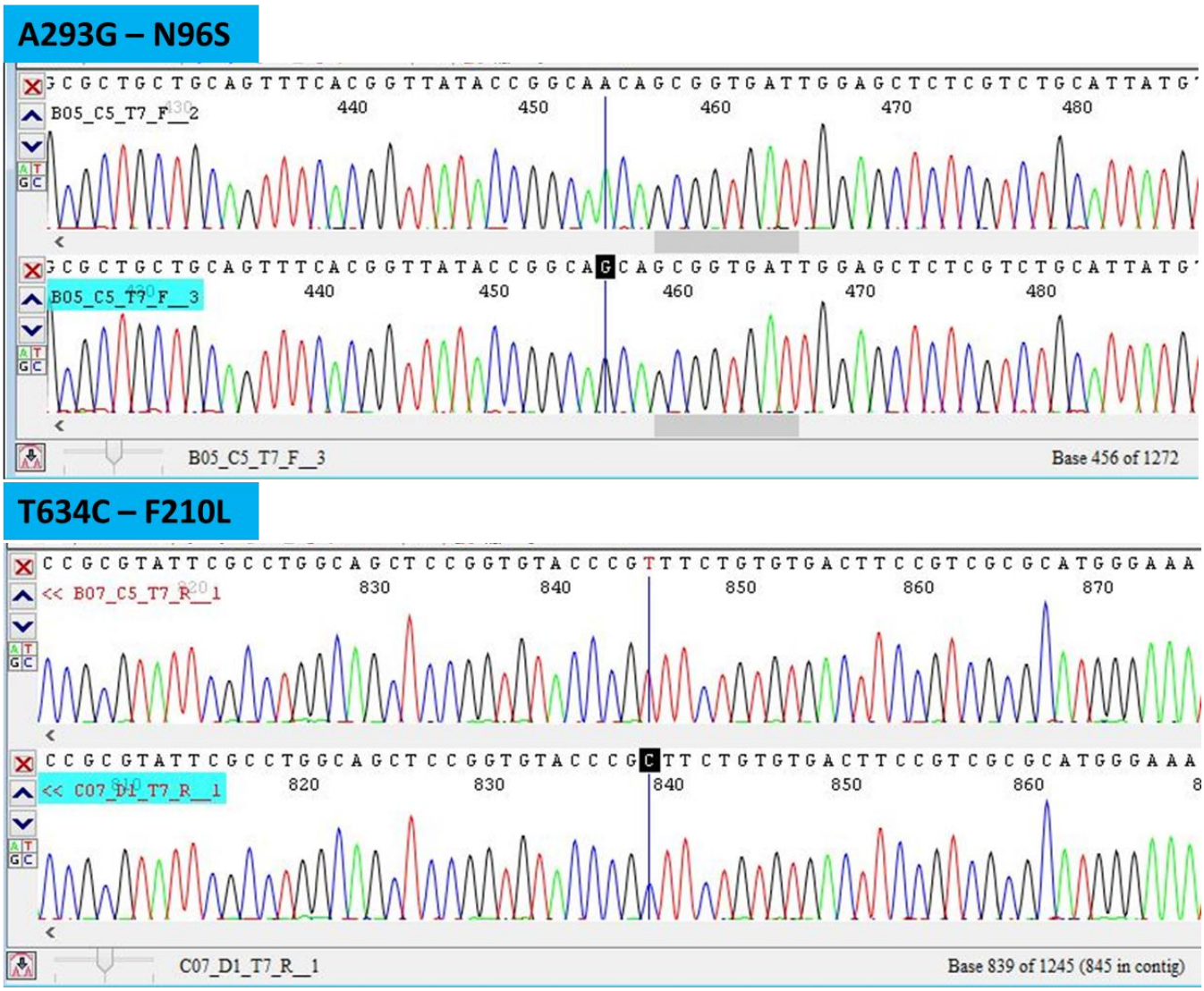
2-G: Superposition of NaM1 with CE7 homologs.

a. Superposition of NaM1 (light brown) with from *Bacillus subtilis* 168 (PDB: 1ODS, pink), *B. pumilus* CECT5072 (2XLB, blue), *Thermoanaerobacterium* sp. (3FCY, cyan), *B. pumilus* PS213 (3FVT, pink), and *Thermotoga maritima* (3M81, salmon) showing structural conservation. **b.** Close-up view of a difference in structural topology involving a loop (red) formed by residues 174-178 in NaM1, but which occur in the other CEs as a β-strand. N and C-termini are indicated.



2-H: Multiple sequence alignment of NaM1 with CE7 homologs.

(WP_020573671.1 – α/β hydrolase *Actinopolymorpha alba*, SFG18906.1 – cephalosporin-C deacetylase *Actinopolymorpha cephalotaxi*, WP_015099820.1 – AcXE *Saccharothrix espanaensis*, OIO90365.1 – AcE *Anaerolineae* sp. CG2_30_64_16, EJY54617.1 – peptidase S15 *Alicyclobacillus hesperidum* URH17-3-68). The PPST**Q**FAAYN conserved region is highlighted with a red rectangle. The MSA was created using Multalin 5.4.1 (Corpet, 1988).

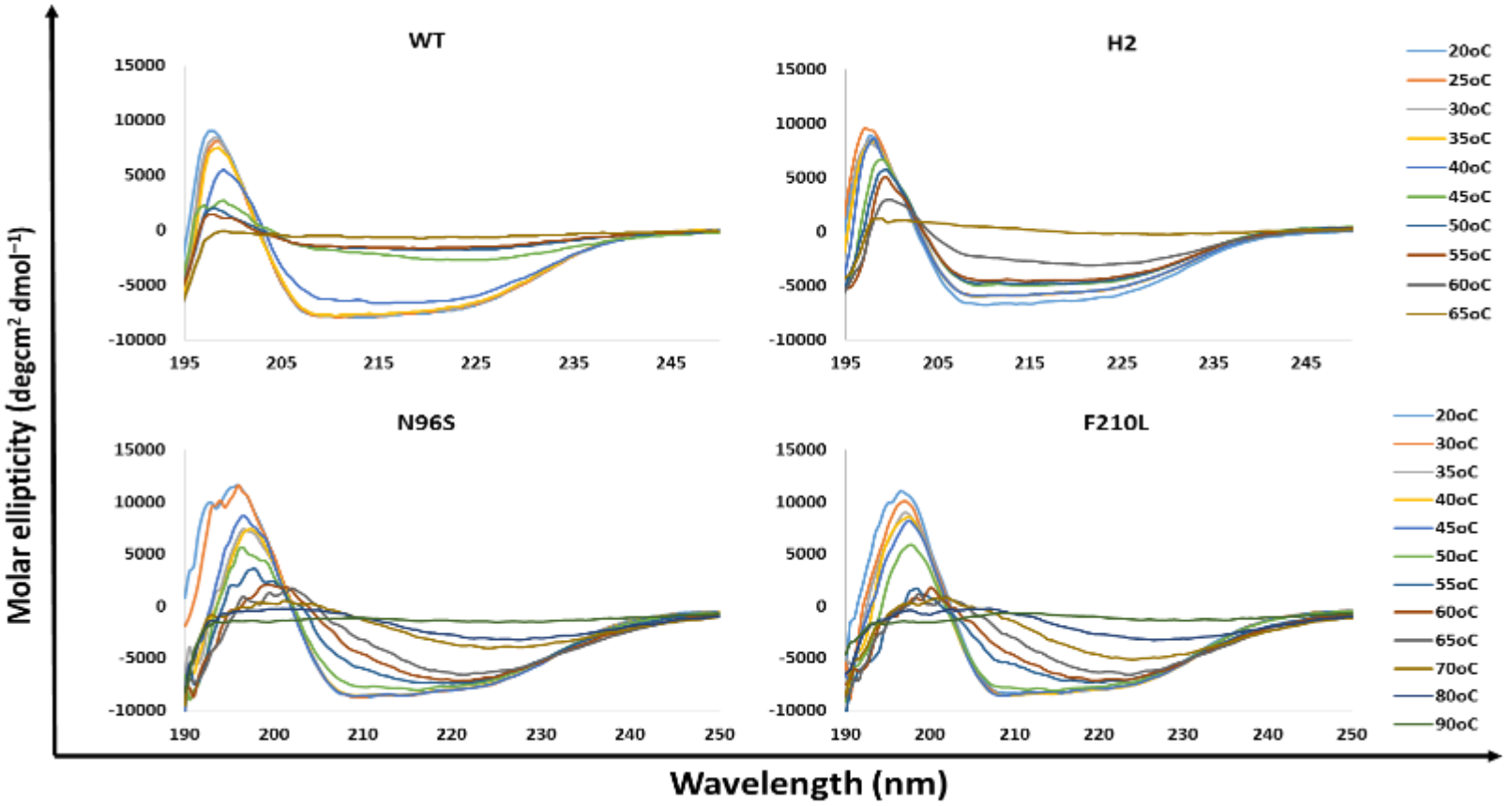


2-I: Sequence data confirming A293G and T634C substitutions which resulted in N96S and F210L mutations in NaM1 (H2).

2-J: Substrate specificity and enzyme kinetics of H2.

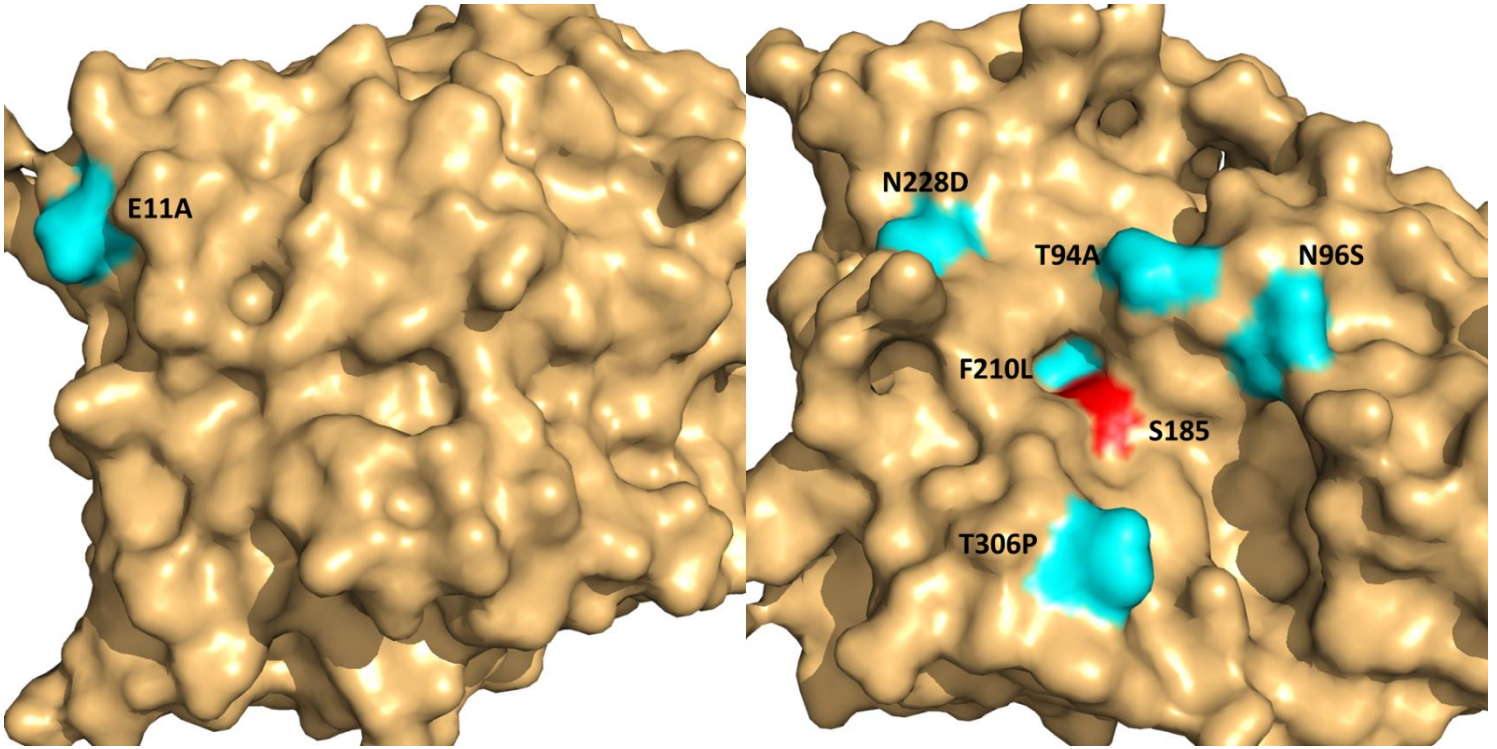
Substrate	V_{max} (Uml^{-1})*	Specific activity (Umg^{-1})*	K_M (mM)	k_{cat} (s^{-1})	Catalytic Efficiency ($M^{-1}s^{-1}$)
***p*-NPA**	1.15 ± 0.11	348.25 ± 34.17	0.76 ± 0.22	205.96 ± 20.21	2.72×10^5
***p*-NPB**	0.10 ± 0.01	20.46 ± 2.23	1.34 ± 0.27	12.18 ± 1.33	9.09×10^3
p-NPO	0.01 ± 0.003	1.86 ± 0.56	0.19 ± 0.25	1.11 ± 0.33	5.83×10^3
4-MUA	1.06 ± 0.06	211.67 ± 12.65	0.47 ± 0.08	125.99 ± 7.53	2.68×10^5
2-NA	0.39 ± 0.05	78.51 ± 9.30	0.41 ± 0.14	46.73 ± 5.54	1.14×10^5
7-ACA	5.84 ± 1.84	106.25 ± 33.40	0.51 ± 0.39	62.83 ± 19.75	1.22×10^2
0.5% AX	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.



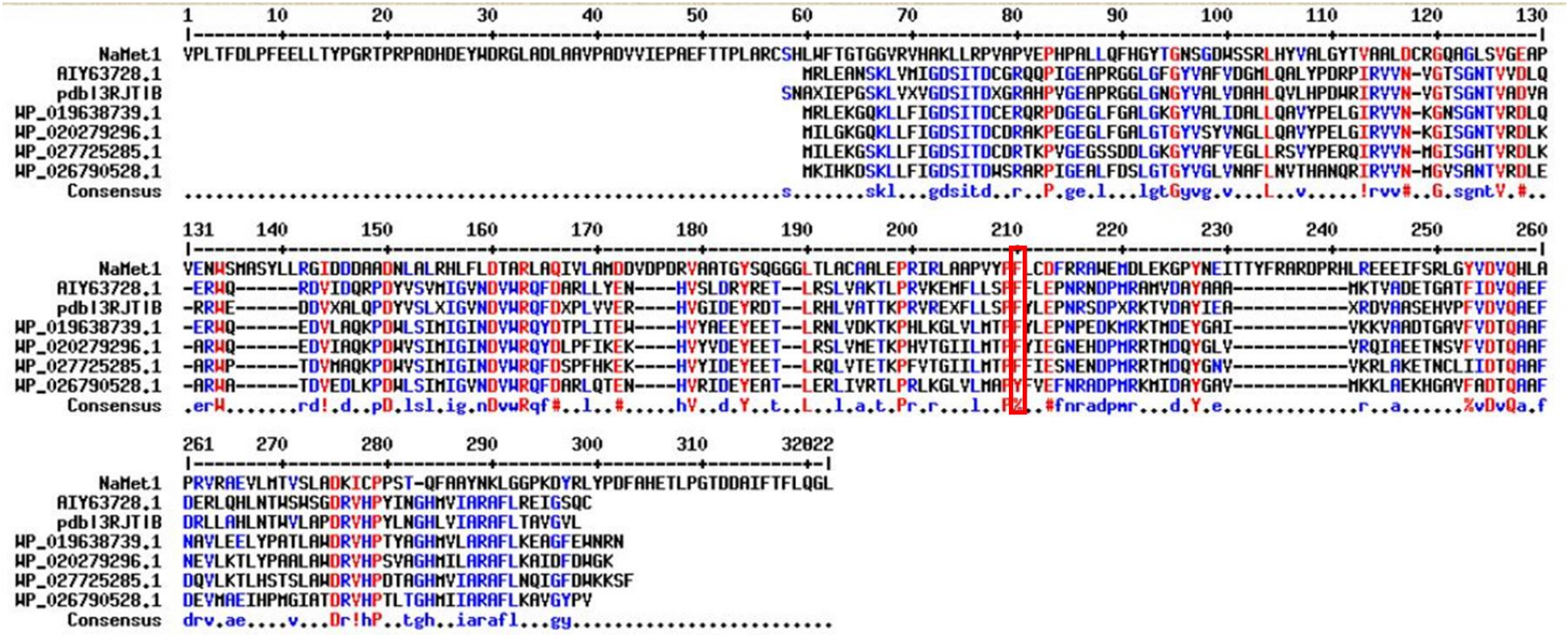
2-K: Far-uv CD spectra of NaM1, H2. $NaM1_{N96S}$ and $NaM1_{F210L}$.

CD spectra of wildtype (WT) NaM1 showing denaturation at temperatures >40 ^{o}C; mutant H2 showing denaturation only at temperatures >60 ^{o}C and indicating >10 ^{o}C increase in thermal stability of the wildtype; and mutants $NaM1_{N96S}$ and $NaM1_{F210L}$ showing gradual denaturation from 55^{o}C, indicating co-contribution to thermal stability of H2.



2-L: NaM1 mutation map.

All mutations selected are located in the core except one (E11A) located on the outer surface, close to the N-terminus of NaM1.



2-M: Multiple sequence alignment of NaMet1 with α/β hydrolases

(AIY63728.1 – SGNH hydrolase-like *Alicyclobacillus* sp., PDB-3RJT – GDSL lipase *A. acidocaldarius* DSM 446, WP_019638739.1 – lipase *Paenibacillus fonticola*, WP_020279296.1 – lipase *Geobacillus kaustophilus*, WP_027725285.1 – lipase *Tuberibacillus calidus*, WP_026790528.1 – lipase *Pleomorphomonas oryzae*). The conserved Phe210 position (NaM1) is highlighted with a red rectangle. The MSA was created using Multalin 5.4.1 (Corpet, 1988).

APPENDIX 3: Publications during PhD study

3A: Type: Review article - Published

Article: **Adesioye F.A.**, Makhalanyane, T.P., Biely, P. and Cowan, D.A. 2016. Phylogeny, classification, metagenomic mining and applications of microbial acetyl xylan esterases. Enzyme and Microbial Technology 93: 79-91. (impact factor – 2.962). Elsevier ©

Author contributions: FAA performed the bioinformatic and phylogenetic analyses and wrote the first draft of the manuscript. All other authors edited subsequent versions of the article.

3B: Type: Research article - Published

Article: Bezuidt, O.K.I., Pierneef, R., Gomri, A.M., **Adesioye, F.A.**, Makhalanyane, T.P., Kharroub, K. and Cowan, D.A. 2016. The pan-genome structure of Geobacillus: implications for the evolution of the genus. Frontiers of Microbiology 7: 723. (impact factor – 4.165)

Author contributions: FAA conducted parts of the bioinformatic analysis involving CAZymes and wrote sections of the first draft of the paper

3C: Type: Research article – Accepted for publication

Article: **Fiyinfoluwa A. Adesioye**, Thulani P. Makhalanyane, Surendra Vikram, Trevor Sewell, Wolf-Dieter Schubert, Don A. Cowan. 2018. Structural characterization and directed evolution of a novel acetyl xylan esterase reveals thermostability determinants of the Carbohydrate Esterase 7 family. Applied and Environmental Microbiology. (impact factor – 3.807)

Author contributions: FAA conducted all the laboratory experiments, structure modelling, substantial parts of the bioinformatics, x-ray crystallography and analyses and wrote the first draft of the manuscript.

APPENDIX 4: List of abbreviations

- 2-NA – 2-naphthyl acetate or α-naphthyl acetate
- 4-MU – 4-methylumbelliferrone
- 4-MUA – 4-methylumbelliferryl acetate
- 7-ACA – 7-aminocephalosporanic acid
- AcE – acetyl esterase
- AcXE - acetyl xylan esterase
- AcXOS – acetylated xylooligosaccharides
- AX – acetylated xylan
- Axe1 – Database annotation for the CE7 family domain
- *Bp*AXE – Axe1 from *Bacillus pumilus*
- BSA – bovine serum albumin
- CAZy or CAZyme – carbohydrate active enzyme
- CCD – cephalosporin C deacetylase
- CD spectroscopy – circular dichroism spectroscopy
- CDD – conserved domain database
- CE - carbohydrate esterase
- CPC – cephalosporin C
- dATP – deoxyadenosine triphosphate
- dCTP – deoxycytidine triphosphate
- DE – directed evolution
- dGTP – deoxyguanosine triphosphate
- DMSO – dimethy sulfoxide
- dNTP – deoxynucleoside triphosphate
- dTTP – deoxythymidine triphosphate

- EDTA – ethylenediaminetetraacetic acid
- epPCR – error-prone polymerase chain reaction
- ESRF – European Synchrotron Radiation Facility
- FPLC – fast pressure liquid chromatography
- hmm – hidden Markov model
- HPLC – high performance liquid chromatography
- IMAC – immobilised metal affinity chromatography
- IPTG – isopropyl β-D-1-thiogalactopyranoside
- Kan – Kanamycin
- K_{cat} – catalytic turnover
- K_M – Michaelis-Menten constant
- LBA – Luria Bertani agar
- LBB – Luria Bertani broth
- MES – 2-(N-morpholino) ethanesulfonic acid
- MSA – multiple sequence alignment
- MWCO – molecular weight cut-off
- NaM1 – Axe1 from Namib Desert soil hypolith metagenome
- NCBI – National Center for Biotechnology Information
- ORF – open reading frame
- PDB – protein data bank
- PEG – polyethylene glycol
- PHENIX – Python-based Hierarchical ENvironment for Integrated Xtallography
- PMSF – phenylmethylsulfonyl fluoride
- *p*-NP – *para*-nitrophenol
- *p*-NPA – *para*-nitrophenyl acetate

- *p*-NPB – *para*-nitrophenyl butyrate
- *p*-NPO – *para*-nitrophenyl octanoate
- *p*-NPP – *para*-nitrophenyl palmitate
- RMSD – root mean square deviation
- SDM – site-directed mutagenesis
- SDS-PAGE – sodium dodecyl sulphate – polyacrylamide gel electrophoresis
- T_{eq} – equilibrium temperature
- *Tm*AcE – Axe1 from *Thermotoga maritima*
- T_{opt} – optimum temperature
- V_{max} – maximum velocity
- V_o – initial velocity
- $\Delta G^{\ddagger}_{inact}$ – Gibbs energy of activation for thermal inactivation
- ΔH_{eq} – change in enthalpy