Multiple nucleophilic elbows leading to multiple active sites in a single module esterase from *Sorangium cellulosum*

D.B.R.K. Gupta Udatha[1,*], Karina Marie Madsen[2], Gianni Panagiotou[3], Lisbeth Olsson[4]

[1]Section for Genetics and Evolutionary Biology, Department of Biosciences, University of Oslo, Oslo, Norway

[2]Center for Biological Sequence Analysis, DTU-Systems Biology, Building 208, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark.

[3]School of Biological Sciences, The University of Hong Kong, Kadoorie Biological, Sciences Building, Hong Kong

[4]Department of Chemical and Biological Engineering, Industrial Biotechnology, Chalmers University of Technology, Kemivägen 10, SE-412 96, Gothenburg, Sweden.

Correspondence and requests to materials should be addressed to D.G.U. (email: gupta.udatha@outlook.com).

Abstract

The catalytic residues in carbohydrate esterase enzyme families constitute a highly conserved triad: serine, histidine and aspartic acid. This catalytic triad is generally located in a very sharp turn of the protein backbone structure, called the nucleophilic elbow and identified by the consensus sequence GXSXG. An esterase from *Sorangium cellulosum* Soce56 that contains five nucleophilic elbows was cloned and expressed in *Escherichia coli* and the function of each nucleophilic elbowed site was characterized. In order to elucidate the function of each nucleophilic elbow, site directed mutagenesis was used to generate variants with deactivated nucleophilic elbows and the functional promiscuity was analysed. *In silico* analysis together with enzymological characterization interestingly showed that each nucleophilic elbow formed a local active site with varied substrate specificities and affinities. To our knowledge, this is the first report presenting the role of multiple nucleophilc elbows in the catalytic promiscuity of an esterase. Further structural analysis at protein unit level indicates the new evolutionary trajectories in emerging promiscuous esterases.

In enzymology, the interplay between the substrate structure and enzyme structure may be unclear due to the promiscuous behaviour of enzymes that have activities on multiple substrates[1]. Many enzymes are highly evolved for a specialized function and it has been proposed that evolving enzyme species possess multiple substrate specificities while its ancestral function is maintained during the whole path of its evolution[2]. Several enzyme groups that act on plant biomass, that constitute a structural diverse set of substrates, can generally hydrolyze several alternative substrates and therefore possess the promiscuous behaviour of multiple substrate specificity[1,3,4].

Enzymes that act on complex carbohydrates and glycoconjugate substrates have been designated as Carbohydrate-Active enZymes (CAZymes) classified into various families based on semi-automatic modular assignment[3]. The feature used to classify the enzymes in CAZy is protein sequence similarity to experimentally characterized enzymes, which serves as a seed for the family that is gradually extended with sequences that share statistically significant sequence similarity. The fact that members of some carbohydrate esterase families in the CAZy (Carbohydrate Active EnZymes) database are able to hydrolyze the substrates specific for other carbohydrate esterase families raises questions on the accuracy of the classification of carbohydrate esterases[5].

Our recent work[1,6] on overlapping substrate specificity among feruloyl esterases (FAEs), an enzyme group of carbohydrate esterase family 1, have indicated the possibility of considering FAEs as 'non-homologous isofunctional enzyme superfamily'. FAEs seem to have evolved from a common ancestor with the classic constellation of the Ser-His-Asp catalytic triad[7]. However, the presence of a common domain with the same catalytic triad among different FAEs does not imply that they have the same function and can act on the same substrates[8].

In esterases that use the classical Ser-His-Asp catalytic triad mechanism, serine acts as nucleophile, histidine as the general acid-base, and the aspartic acid helps to neutralize the charge that forms on histidine during the catalytic process[9]. The nucleophile of this catalytic triad is generally located in a very sharp turn of the protein backbone structure, called the nucleophilic elbow and identified by the consensus sequence GXSXG (X=any amino acid residue; The serine residue is located at the centre of this conserved pentapeptide)[1,8,10]. In our earlier study[8], a total of 70 FAE protein sequences have been found with 2-3 nucleophilic elbows, but only one, had as many as five nucleophilic elbows. The putative

esterase with five nucleophilic elbows subject to this study is from *Sorangium cellulosum* Soce56[11] and will be referred as 'Multiple Nucleophilic Elbowed Esterase' (MNEE).

S. cellulosum is a soil-dwelling gram-negative bacterium and a producer of several different secondary metabolites. The substrate matrix present in the habitat or the chemical environment will influence the activity trajectories of the evolutionarily adapting enzymes that are essential contributors to cell survival and proliferation in altered ambient conditions[12]. Modern understanding of enzyme evolution indicates that mutations in the protein sequence can have effect on catalytic activity through small changes in local structure of the active site. Some of the mutations in the evolutionary path of enzymes may affect the local structure, that do not change the catalytic activities, but may change the catalytic parameters of the enzyme and thus creating merely an enzyme variant with different substrate affinity[13]. Furthermore, the rearrangements of DNA sequences encoding different enzymes results in redesigning of entire structure to form proteins with promiscuous activity[14,15]. DNA recombinations can result into major changes in the structure and multiple nucleophilic elbows in a single protein can only be accomplished by the fusion of genes.

The aim of the present study is to characterize the enzyme MNEE from *S. cellulosum* Soce56, investigating the role of each nucleophilic elbow and search for novel functionalities. The gene sce3277 from *S. cellulosum* SoCe56 encoding MNEE, was cloned, expressed in *Escherichia coli*, purified, characterized, and compared to other FAE and CAZy family members. In order to elucidate the function of each nucleophilic elbow, site directed mutagenesis was used to generate variants. To the best of our knowledge this is the first time that a predicted feruloyl esterase with multiple nucleophilic elbows is isolated and characterized.

Results

Identification and probing the function of nucleophilic elbows

Even though MNEE possess five nucleophilic elbows (see Table 1) with the classic constellation of the Ser-His-Asp triad forming the active site, variations in amino acid sequences forming surface loops and additional domains allow them to accommodate different aromatic substrates. The first two nucleophilic elbows are overlapping with the

presence of two successive nucleophiles (serine residues) at positions 123 and 124 of the MNEE's amino acid sequence. To demonstrate the presence of catalytic promiscuity or multiple substrate specificity of MNEE, the relevant substrates defined by the results of amino acid sequence alignments with CAZy family members were considered for analysis. MNEE and its variants were tested for the presence of enzyme activities viz., feruloyl esterase, aryl esterase, acetyl esterase, carboxyl esterase, paraben esterase, paraoxonase and PHB depolymerase.

Feruloyl esterase activity

The wild type enzyme MNEE-WT has high protein sequence similarity with feruloyl esterases, which has become evident from our previous work[8]. Based on this, the enzyme has been annotated as a putative feruloyl esterase. In order to validate this annotation, the feruloyl esterase activity was tested against 15 different methyl esters. The only substrate that the MNEE-WT does not display any activity against was M43PP (see Table 2a). Other than that, the activity varies from 0.34 U/mg against M34DC to almost 10-fold higher at 3.00 U/mg against M35DC. Only the substrates M3C, MTM, M3M and M35DC yield an activity markedly above 1 U/mg (Table 2a).

Variants with one mutated nucleophilic elbow: The *variants 1* to *5* of MNEE contains one mutated nucleophilic elbow i.e., the catalytic residue serine is mutated to alanine to make the respective elbow nonfunctional. In addition, these variants possess the four unmutated nucleophilic elbows which are active. The results from the activity analysis of these variants with one mutated nucleophilic elbow were surprising, as the mutation proved to be an advantage over the WT in several cases (Table 2b). All the variants showed different FAE activity profile.

- *Variant 1* has activity against all substrates except two, M2M and M4M. The activity against M43PP is striking as the WT had no activity against this substrate.
- *Variant 2* seemed to suffer more from the mutation, as that variant was inactive for as many as 7 substrates.
- *Variant 3* is the only case where no activity is observed against MFA.
- The highest activity against M3M among all mutants was observed for *Variant 4*. All other activities are distinctly lower than the performance of the WT.

- *Variant 5* is most active among the variants with one mutated nucleophilic elbow. Activities >5.0 is U/mg observed for both M3C and M43PP, and in general, the activity is typically slightly higher for the remaining substrates, compared to the other mutants.

Variants with four mutated nucleophilic elbows: Variants 6-10 each have only one of the nucleophilic elbows unchanged; the remaining four has been mutated by single amino acid substitutions (Table 2c).

- *Variant 6* displays maximum activity against M3C. This mutant is inactive against 4 different substrates, M2M, M3M, M34MC and M43PP.
- Having only nucleophilic elbow E2 unchanged, *Variant 7* is markedly less active than other mutants. Total inactivity is observed against 3 substrates, and maximum activity is observed against M3C.
- *Variant 8,* is one of the most active variants overall. It has an activity above 2.5 U/mg against MCA, M2C, M3C and MTM.
- *Variant 9* has a well distributed activity as well, being inactive against 3 substrates, and reaching maximum activity against M3M.
- *Variant 10* with an unchanged nucleophilic elbow 5 is inactive against 5 of the 15 substrates.

All the mutants display feruloyl esterase activity, but the profile against the 15 methyl cinnamate substrates is very different. There are three substrates against which all mutants have activity; M2C, M3C and M35DC. The specific activity towards the substrates changes with the mutations of the nucleophilic elbows, indicating that the binding of the substrates is not specific to one nucleophilic elbow, but can happen at multiple elbows, with different specificity. The enzymatic activity is generally higher against the substrates M2C and M3C in variants having only one active or unmutated nucleophilic elbow.

Another interesting observation in case of M43PP, against which the WT is inactive, is that variant 1 and variant 5 has significant activity against this substrate and to a less extent for variant 9 and variant 10. Interestingly, *variant 10* have all nucleophilic elbows mutated except E5; whereas in *variant 5*, E5 is the only mutated elbow. Thus the substrate is not specific to one site, but do however depend on mutations of the WT in order to obtain proper binding and subsequent hydrolysis.

Table 1. The five nucleophilic elbows of *S. cellulosum* MNEE. The catalytic triad residues were found at appropriate distances in the respective binding pockets of MNEE[74].

Nucleophilc Elbow No.	Nucleophilc Elbow position in the amino acid sequence	Nucleophilic Elbow	Position of SER	Position of HIS	Position of ASP
E1	121 - 125	GG**S**SG	123	271	212
E2	122 - 126	GS**S**GG	124	271	212
E3	359 - 363	GG**S**AG	361	434	308
E4	385 - 389	GD**S**TG	387	43	404
E5	392 - 396	GS**S**SG	394	90	379

Table 2. Feruloyl esterase activity of *S. cellulosum* MNEE and its variants against 15 methyl esters. (**a**) Feruloyl esterase activity MNEE-WT. (**b**) Feruloyl esterase activity of variants 1-5. (**c**) Feruloyl esterase activity of variants 6-10.

(**a**)

	MNEE-WT
non-mutated Elbow/s	E1-E2-E3-E4-E5
Mutated Elbow/s	None
Substrate	**Activity (U/mg)**
MFA	1.11 ± 0.10
MCA	0.85 ± 0.08
MPC	0.75 ± 0.04
MSA	0.84 ± 0.10
M2C	1.26 ± 0.06
M3C	1.69 ± 0.09
MC	1.11 ± 0.05
MTM	2.53 ± 0.36
M2M	0.69 ± 0.09
M3M	2.58 ± 0.11
M4M	0.57 ± 0.02
M34DC	0.34 ± 0.02
M35DC	3.00 ± 0.07
M34MC	0.82 ± 0.03
M43PP	N.D.

n=3; errors indicate standard deviation
N.D. = Not Detected

Methyl ferulate or Methyl 4-hydroxy-3-methoxy cinnamate (MFA), Methyl caffeate or Methyl 3,4-dihydroxy cinnamate (MCA), Methyl p-coumarate or Methyl 4-hydroxy cinnamate (MPC), Methyl sinapate or Methyl 4-hydroxy-3,5-dimethoxy cinnamate (MSA), Methyl 2-hydroxy cinnamate (M2C), Methyl 3-hydroxy cinnamate (M3C), Methyl cinnamate (MC), Methyl 3,4,5-trimethoxy cinnamate (MTM), Methyl 2-methoxy cinnamate (M2M), Methyl 3-methoxy cinnamate (M3M), Methyl 4-methoxy cinnamate (M4M), Methyl 3,4-dimethoxy cinnamate (M34DC), Methyl 3,5-dimethoxy cinnamate (M35DC), Methyl 3-hydroxy-4-methoxy cinnamate (M34MC),Methyl 4-hydroxy-3-methoxy phenyl propionate (M43PP).

(**b**)

	Variant 1	**Variant 2**	**Variant 3**	**Variant 4**	**Variant 5**
non-mutated Elbow/s	E2-E3-E4-E5	E1-E3-E4-E5	E1-E2-E4-E5	E1-E2-E3-E5	E1-E2-E3-E4
Mutated Elbow/s	E1	E2	E3	E4	E5
Substrate	**Activity (U/mg)**	**Activity (U/mg)**	**Activity (U/mg)**	**Activity (U/mg)**	**Activity (U/mg)**
MFA	2.26 ± 0.07	1.26 ± 0.08	N.D.	0.35 ± 0.02	1.28 ± 0.26
MCA	0.69 ± 0.15	N.D.	0.43 ± 0.03	0.34 ± 0.05	0.48 ± 0.06
MPC	0.70 ± 0.07	N.D.	N.D.	0.23 ± 0.04	0.39 ± 0.04
MSA	0.21 ± 0.02	N.D.	0.22 ± 0.04	0.55 ± 0.03	0.62 ± 0.08
M2C	0.43 ± 0.06	0.47 ± 0.02	1.61 ± 0.02	0.34 ± 0.04	1.18 ± 0.12
M3C	1.65 ± 0.05	1.11 ± 0.16	0.35 ± 0.04	0.38 ± 0.01	5.45 ± 0.08
MC	0.43 ± 0.07	0.39 ± 0.09	0.69 ± 0.04	0.52 ± 0.03	1.87 ± 0.06
MTM	0.54 ± 0.07	N.D.	0.28 ± 0.01	N.D.	0.84 ± 0.11
M2M	N.D.	N.D.	N.D.	N.D.	0.68 ± 0.10
M3M	2.26 ± 0.14	2.02 ± 0.07	1.27 ± 0.09	6.43 ± 0.74	0.76 ± 0.02
M4M	N.D.	N.D.	0.20 ± 0.03	0.13 ± 0.02	N.D.
M34DC	0.28 ± 0.03	0.17 ± 0.04	N.D.	0.18 ± 0.01	0.43 ± 0.04
M35DC	1.51 ± 0.14	0.29 ± 0.01	4.37 ± 0.21	0.35 ± 0.04	1.81 ± 0.12
M34MC	0.46 ± 0.02	N.D.	1.27 ± 0.07	1.16 ± 0.07	1.61 ± 0.05
M43PP	4.90 ± 0.26	1.03 ± 0.05	N.D.	N.D.	5.46 ± 0.42

n=3; errors indicate standard deviation
N.D. = Not Detected

The MNEE variants and their respective amino acid substitutions:
Variant 1: S123A
Variant 2: S124A
Variant 3: S361A
Variant 4: S387A
Variant 5: S394A

(**c**)

	Variant 6	**Variant 7**	**Variant 8**	**Variant 9**	**Variant 10**
non-mutated Elbow/s	E1	E2	E3	E4	E5
Mutated Elbow/s	E2-E3-E4-E5	E1-E3-E4-E5	E1-E2-E4-E5	E1-E2-E3-E5	E1-E2-E3-E4
Substrate	**Activity (U/mg)**	**Activity (U/mg)**	**Activity (U/mg)**	**Activity (U/mg)**	**Activity (U/mg)**
MFA	0.65 ± 0.04	0.56 ± 0.03	0.52 ± 0.11	1.22 ± 0.07	0.73 ± 0.03
MCA	0.19 ± 0.02	0.18 ± 0.05	2.70 ± 0.08	N.D.	N.D.
MPC	0.24 ± 0.05	N.D.	1.50 ± 0.03	0.61 ± 0.02	N.D.
MSA	0.60 ± 0.09	N.D.	N.D.	0.19 ± 0.04	N.D.
M2C	1.22 ± 0.07	0.35 ± 0.06	2.72 ± 0.11	1.17 ± 0.10	0.42 ± 0.05
M3C	3.65 ± 0.13	1.48 ± 0.04	3.33 ± 0.08	1.91 ± 0.04	1.66 ± 0.04
MC	0.53 ± 0.03	0.50 ± 0.03	N.D.	N.D.	0.40 ± 0.01
MTM	0.92 ± 0.08	0.19 ± 0.02	4.54 ± 0.08	1.03 ± 0.04	1.69 ± 0.07
M2M	N.D.	N.D.	N.D.	N.D.	N.D.
M3M	N.D.	1.14 ± 0.04	0.80 ± 0.05	4.37 ± 0.23	3.45 ± 0.04
M4M	0.50 ± 0.02	0.17 ± 0.06	0.24 ± 0.05	0.23 ± 0.04	0.14 ± 0.03
M34DC	0.16 ± 0.01	0.19 ± 0.05	1.06 ± 0.14	N.D.	0.34 ± 0.02
M35DC	1.73 ± 0.07	1.45 ± 0.10	1.14 ± 0.06	1.94 ± 0.03	1.41 ± 0.04
M34MC	N.D.	0.65 ± 0.05	0.21 ± 0.07	0.22 ± 0.03	N.D.
M43PP	N.D.	N.D.	N.D.	1.57 ± 0.07	2.23 ± 0.13

n=3; errors indicate standard deviation
N.D. = Not Detected

The MNEE variants and their respective amino acid substitutions:
Variant 6: S124A; S361A; S387A; S394A
Variant 7: S123A; S361A; S387A; S394A
Variant 8: S123A; S124A; S387A; S394A
Variant 9: S123A; S124A; S361A; S394A
Variant 10: S123A; S124A; S361A; S387A

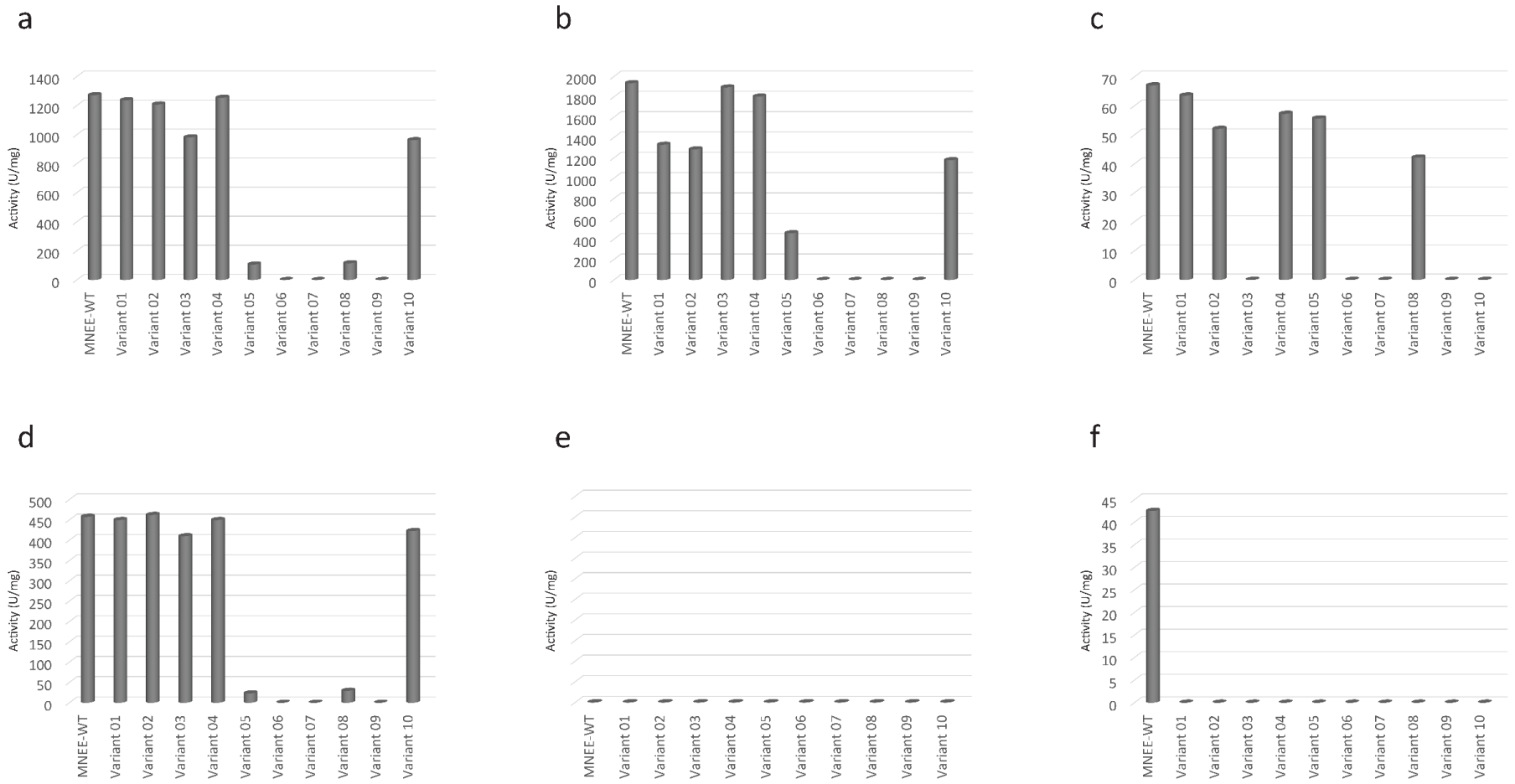


Figure 1. Non-feruloyl esterase activities of *S. cellulosum* MNEE and its variants. (**a**) Aryl esterase activity. (**b**) Acetyl esterase activity. (**c**) Carboxyl esterase activity. (**d**) Paraben Esterase activity. (**e**) Paraoxonase activity. (**f**) PHB depolymerase activity.

Non-feruloyl esterase activities

The wild type MNEE and its 10 variants were assayed against a range of substrates for activities other than feruloyl esterase activity. Figure 1 summarizes the results of the non-FAE activities.

***Aryl esterase activity*:** The aryl esterase activity assay was carried out using phenyl acetate (PA) as a substrate. No enzymatic activity against PA was observed for variant 6, variant 7 and variant 9, where E1, E2 and E4, respectively, were the only active nucleophilic elbows.

Variant 5, which possess the amino acid substitution in Elbow 5, had the most significant decrease in activity from the WT activity. The activity of variant 3 is a noticeable decrease compared to the WT. The aryl esterase activities of variant 8 and variant 10 corresponds well to these observations, as each of the nucleophilic elbows E3 and E5 are able to uphold aryl esterase activity on their own. Elbow 3 contributes to a tenth of the aryl esterase activity, while the majority of aryl esterase activity can be assigned to E5.

***Acetyl esterase activity*:** The acetyl esterase activity assay was determined using p-nitrophenyl acetate (PNA) as a substrate. Variants 1-5, with one mutated elbow, showed activity against PNA, but lower than the WT activity. This shows that one nucleophilic elbow alone is not responsible for the acetyl esterase activity.

Of the variants 6-10, acetyl esterase activity was observed only for variant 10, where the only unmutated nucleophilic elbow is E5. From this it can be concluded, that the functionality of E5 is crucial for the enzymatic activity against PNA. E1 and E2 both play a role in the acetyl esterase activity as well, but the results indicate that even a single amino acid mutation of this overlapping nucleophilic elbow disturbs the catalytic function of the respective active site.

***Carboxyl esterase activity*:** The carboxyl esterase activity assay was determined using p-nitrophenyl palmitate (PNPM) as a substrate. In comparison with feruloyl esterase activity, the carboxyl esterase activity among MNEE and its variants are markedly higher. The most interesting result regarding the carboxyl esterase activity is that variant 3 is the only mutant of one elbow mutated variants (variants 1-5) which has not showed activity against PNPM. Correspondingly, variant 8 is the only active mutant of the one elbow unmutated variants (variants 6-10). Consequently it can be deduced that E3 is the only binding pocket playing an important role in the carboxyl esterase activity.

***Paraben esterase activity*:** The paraben esterase activity was determined using methyl paraben (MP) as a substrate. Variants 1, 2 and 4 showed paraben esterase activity almost measuring up to the activity of the MNEE-WT enzyme. The same is the case for variant 10, where the only unmutated elbow is E5. For variant 5 and variant 8, very small activities are observed. The results indicated that the paraben esterase activity relies on nucleophilic elbow 5 to a great extent in addition to a slight depending on nucleophilic elbow 3.

***Paraoxonase activity*:** Paraoxonases are involved in the hydrolysis of organophosphates, esters of phosphoric acid. The paraoxonase activity was tested against the substrate paraoxon (PO). In all cases, the activity was zero, consequemtly MNEE does not possess any paraoxonase activity irrespective of its amino acid sequence similarity with paraoxonases.

***PHB depolymerase activity*:** The activity against poly-3-hydroxybutyric acid (PHB) was tested for MNEE and its variants, to determine PHB depolymerase activity. The WT enzyme displayed PHB depolymerase activity, but no PHB depolymerase activity was observed for any of the variants, which indicates that the structural changes induced by even one amino acid substitution deactivates the enzyme activity against PHB completely. Due to this high sensitivity, it is unclear which nucleophilic elbows are involved in the PHB depolymerase activity in the WT enzyme.

Variant with all five elbows mutated

We have utilized the custom gene synthesis service (NZYTech Ltd, Portugal) for cloning the variant with all the five elbows mutated. We used the gene of variant 10 (with four mutated elbows) as a starting material and introduced the fifth mutation in it. We carried out the expression and purification of the new variant as per the protocols followed for the other variants of the enzyme. The activities that were observed in wild type MNEE are not present in the variant with the five elbows mutated.

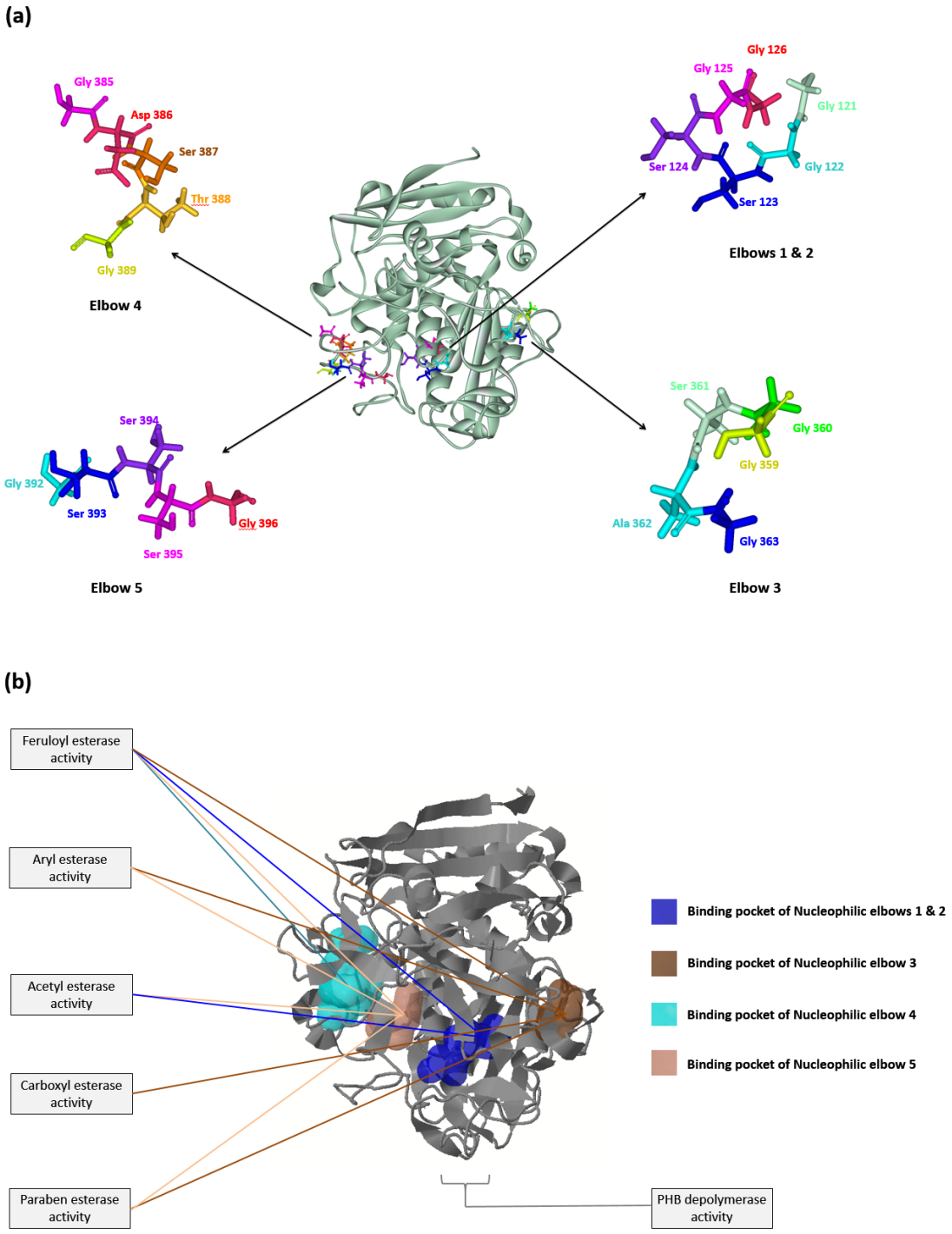


Figure 2. Structure of the modeled MNEE and its five nucleophilic elbows. (**a**) The consensus sequence GXSXG of each nucleophilic elbow in MNEE was represented in stick model. The four binding pockets of MNEE displayed in different colours. The entire protein volume is 34766 Cubic Angstroms and each nucleophilic elbow is a part of a distinct binding pocket. The overlapping nucleophilic elbows 1 & 2 are present in a binding pocket with a site volume of 309 Cubic Angstroms. Elbow 3, Elbow 4 and Elbow 5 are part of three different binding pockets with site volumes of 179, 487 and 183 Cubic Angstroms, respectively. The modeled MNEE structure was deposited in PMDB and its accession code is PM0078343. (**b**) MNEE showed hydrolytic activity on the substrates of six different enzymes and mutational analysis revealed that each binding pocket possess both unique and overlapping substrate specificities.

Kinetic promiscuity of MNEE's nucleophilic elbowed binding pockets

Along with catalytic promiscuity (as shown in Figure 2), MNEE's nucleophilic elbowed binding pockets not only showed different K_m and V_{max} values but also follow different kinetic equations, a property we refer to *kinetic promiscuity*. To probe the kinetic promiscuity of MNEE, the kinetic parameters for the wild type and the 10 variants were determined independently. As a case study, we presented below the kinetic behavior showed by MNEE and its variants for aryl esterase activity.

Aryl esterase kinetics of MNEE and its variants

The nucleophilic elbowed binding pockets of MNEE display non-Michaelis-Menten or atypical steady state kinetic patterns. MNEE's ability to bind to multiple substrate molecules simultaneously leads to atypical kinetic phenomena with cooperativity that can be either homotropic or heterotropic. Homotropic cooperativity describes the interactions of molecules of the same substrate, whereas heterotropic cooperativity describes the interactions of different substrates[16]. Based on the specific activity data we conclude that Elbow 3 and Elbow 5 are responsible for the aryl esterase activity of the MNEE. The kinetic situation we are investigating is that two PA molecules can bind to two different binding pockets present in MNEE and thus discuss the homotropic effects in among the nucleophilic elbowed binding pockets of MNEE. Therefore, the following kinetic equation was derived as applicable for an enzyme with two binding sites[16,17]. Figure 3 depicts the simplest model for homotropic effects with rapid equilibrium assumption.

$$\frac{v}{V_{max}} = \frac{\frac{[S]}{K_S} + \frac{\beta[S]^2}{\alpha K_S^2}}{1 + \frac{2[S]}{K_S} + \frac{[S]^2}{\alpha K_S^2}} \quad \textbf{(1)}$$

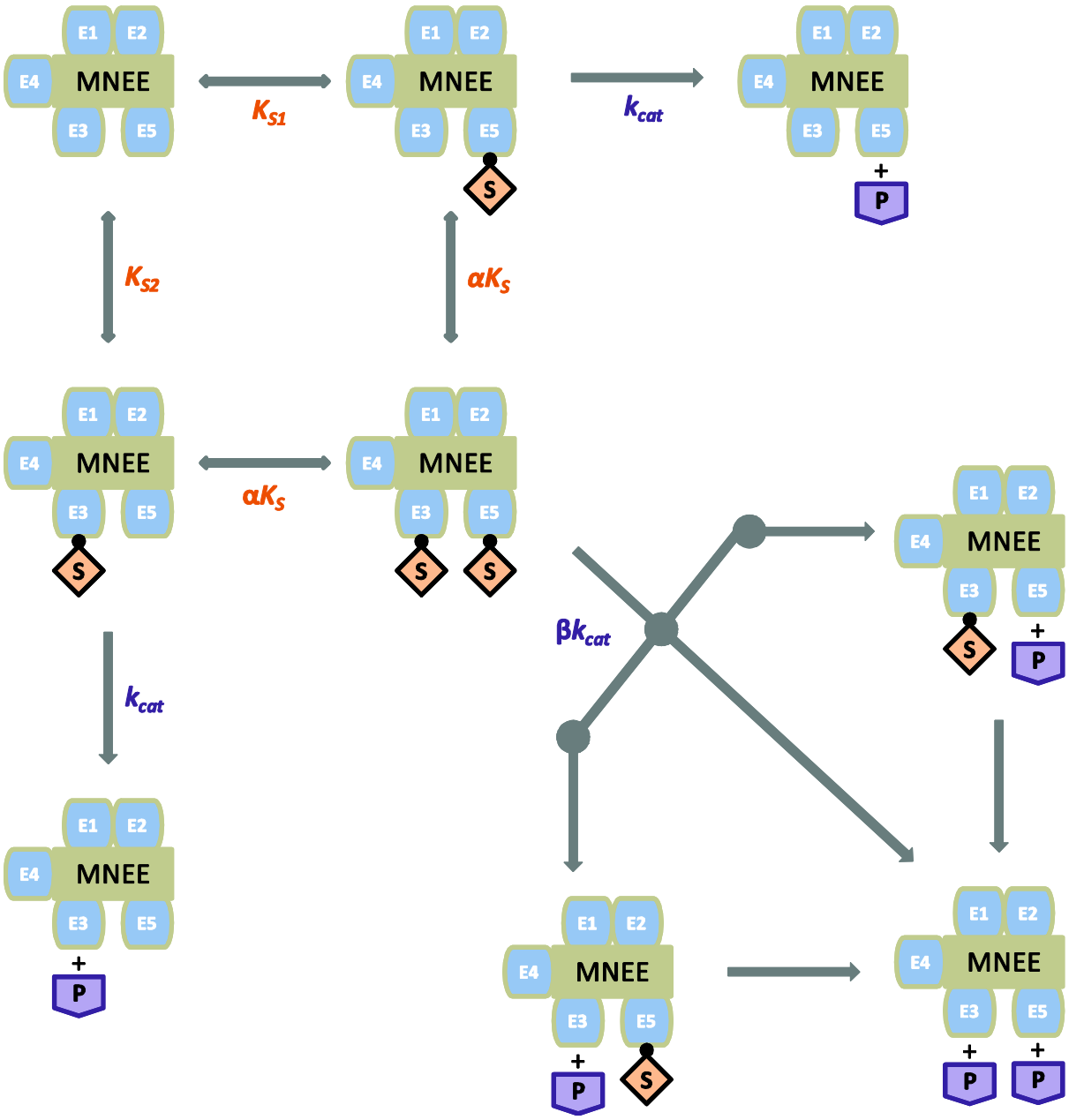


Figure 3. Kinetic scheme for *One substrate-Two binding sites* model enzyme. Elbows 1, 2 and 3 were not considered as they are not involved in aryl esterase activity.

In the model shown in Figure 3, the substrate can bind either to Elbow 3 or Elbow 5, and the two respective enzyme-substrate complexes have identical substrate dissociation constants, K_S and identical k_{cat} values for product formation. The equation 1 is an oversimplified form of homotropic allosteric kinetics for MNEE due to the kinetic equivalence of [MNEE(E3)•S] and [MNEE(E5)•S]. Positive homotropic cooperativity will be obtained when $\alpha < 1$ or $\beta > 1$ and velocity curves will be sigmoidal. Biphasic velocity curves will be obtained if $\alpha > 1$ or $\beta < 1$ indicating substrate inhibition. The effect that has the binding of the substrate to Elbow 3 on the binding of the substrate to Elbow 5 is depicted as α, whereas β is the effect of presence of the substrate in elbow 3 on the k_{cat} of the Elbow 5. Even though the equation 1 and scheme presented in Figure 3 reduces the complexity of the system, it may not be sufficient to describe the K_m values that we obtained experimentally. An equation[18] that accommodates K_m values is given below and the modified scheme is shown in Figure 4.

$$\frac{v}{[E_t]} = \frac{K_{Cx}\left({}^{[S]}/_{K_{m1}}\right) + K_{Fx}\left({}^{[S]^2}/_{K_{m1}K_{m2}}\right)}{1 + {}^{[S]}/_{K_{m1}} + {}^{[S]^2}/_{K_{m1}K_{m2}}} \qquad \textbf{(2)}$$

Where,

$K_{m1} = (K_{Bx} + K_{Cx})/K_{Ax}$ [or] $(K_{By} + K_{Cy})/K_{Ay}$

$K_{m2} = (K_{Dx} + K_{Fx})/K_{Ex}$ [or] $(K_{Dy} + K_{Fy})/K_{Ey}$

The difference between Equation 2 and 1 is the replacement of equilibrium constants (K_S) by K_m constants (For example, K_S shown in Figure 3 = K_{Bx}/K_{Ax} or K_{By}/K_{Ay} shown in Figure 4). As shown before, aryl esterase activity was observed in the native MNEE and its variants 1, 2, 3, 4, 5, 8 and 10; hence their kinetics were discussed below.

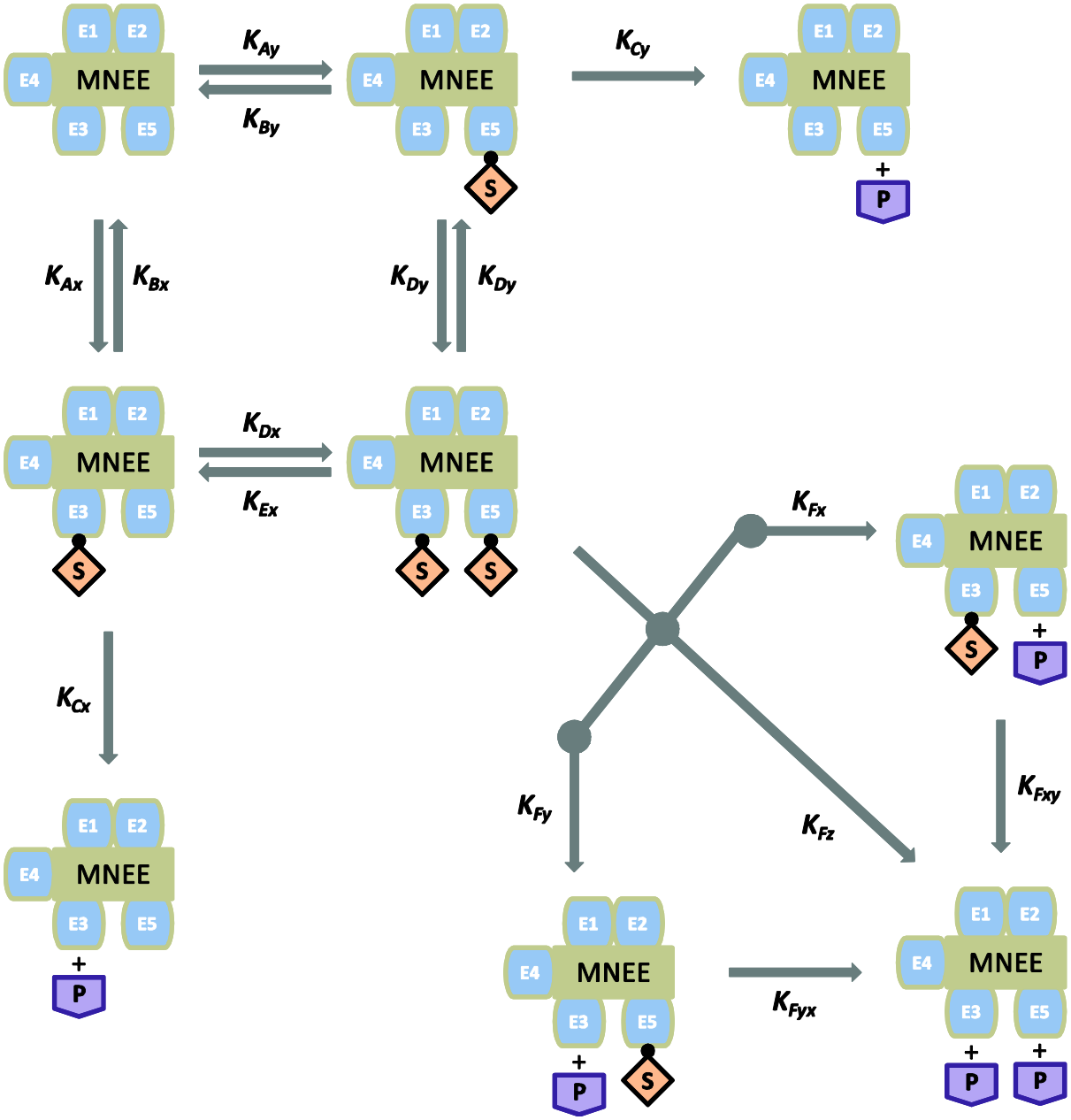
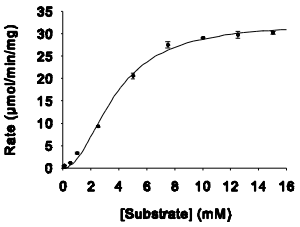


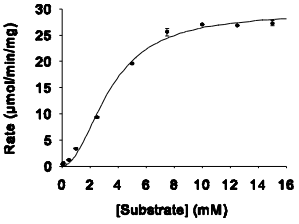
Figure 4. Kinetic scheme that accommodates kinetic constants for *One substrate-Two binding sites* model enzyme. Elbows 1, 2 and 3 were not considered as they are not involved in aryl esterase activity.

a



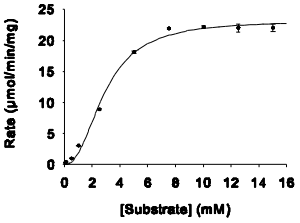
Equation to fit: Hill
Km = 3.72 ± 0.14
Vmax = 32.56 ± 0.74
n = 2.05 ± 0.13

b



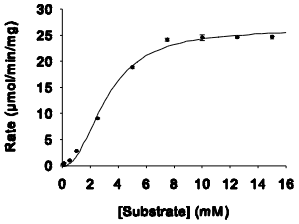
Equation to fit: Hill
Km = 3.43 ± 0.13
Vmax = 29.21 ± 0.63
n = 2.06 ± 0.13

c



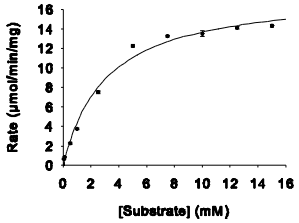
Equation to fit: Hill
Km = 2.89 ± 0.11
Vmax = 23.27 ± 0.44
n = 2.24 ± 0.17

d



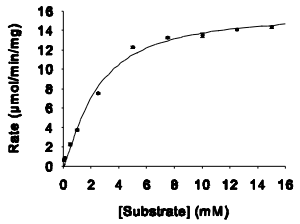
Equation to fit: Hill
Km = 3.17 ± 0.10
Vmax = 26.16 ± 0.48
n = 2.24 ± 0.15

e



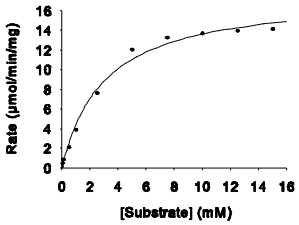
Equation to fit: Michaelis-Menten
Km = 3.06 ± 0.25
Vmax = 17.85 ± 0.45

f



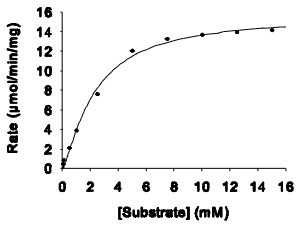
Equation to fit: Hill
Km = 2.37 ± 0.18
Vmax = 15.88 ± 0.50
n = 1.28 ± 0.09

g



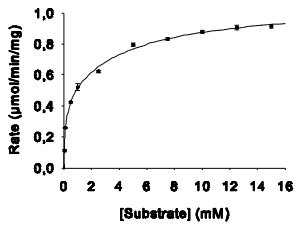
Equation to fit: Michaelis-Menten
Km = 3.01 ± 0.22
Vmax = 17.69 ± 0.39

h



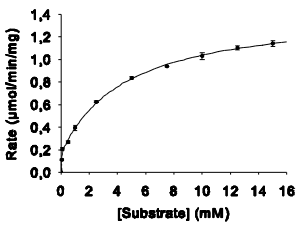
Equation to fit: Hill
Km = 2.31 ± 0.13
Vmax = 15.68 ± 0.37
n = 1.3 ± 0.07

i



Equation to fit: Isoenzyme
Km1 = 0.10 ± 0.02
Km2 = 4.24 ± 1.19
Vmax1 = 0.43 ± 0.04
Vmax2 = 0.63 ± 0.03

j



Equation to fit: Isoenzyme
Km1 = 0.02 ± 0.01
Km2 = 4.72 ± 0.63
Vmax1 = 0.17 ± 0.03
Vmax2 = 1.26 ± 0.03

Figure 5. Aryl esterase activity kinetics of MNEE and its variants. (**a**) MNEE-WT. (**b**) Variant 1. (**c**) Variant 2. (**d**) Variant 4. (**e**) & (**f**) Variant 3. (**g**) & (**h**) Variant 10. (**i**) Variant 5. (**j**) Variant 8.

Native MNEE kinetics: A sigmoidal curve (see Figure 5a) was obtained for the wild MNEE, which essentially means that the singly bound substrate increases the affinity of MNEE for the second substrate molecule. As shown in Figure 5a, the Hill coefficient of > 1 also indicates (n = 2.05 ± 0.13) the positive cooperativity where $K_{m2} < K_{m1}$ and $V_{max2} > V_{max1}$. Since V_{max1} cannot be determined independently, it is not possible to explain whether the sigmoidal kinetics of the native MNEE is due to increased binding affinity (MNEE-E3•S or MNEE-E5•S) or increased velocity from S•E5-MNEE-E3•S or both.

Variants 1, 2 and 4 fit well to Hill equation and follows sigmoidal kinetics (Figure 5b, c and d) with n values > 1, indicating positive cooperativity

Variants 3 and 10 showed hyperbolic curve (Figure 5e and g) and fits well to Michaelis-Menten equation with low K_m and high V_{max} values. As shown in Figure 5f and h, under these conditions it is not possible to distinguish between a two-binding site model from a single-binding site model.

Variant 5 fit well to Isoenzyme equation. A biphasic saturation profile was obtained for variant 5 indicating that $V_{max2} > V_{max1}$ and $K_{m2} >> K_{m1}$. It is evident from Figure 5i, the presence of one low affinity elbow for the substrate and another high affinity elbow for the substrate. As shown in Figure 5j, ***Variant 8*** follows similar kinetic curve as variant 5.

From the kinetic data, we can conclude that the Elbow 3 has low affinity for the substrate whereas Elbow 5 has high affinity. Both the Elbows (E3 and E5) were active in wild, variant 1, variant 2 & variant 4; and all of them showed sigmoidal kinetics resulting from positive cooperativity. In variant 3 and variant 10, the Elbow 3 is inactivated and the kinetics of Elbow 5 can be seen with low K_m and high V_{max} values. In variant 5 and variant 8, the Elbow 5 is inactivated and the kinetics of Elbow 3 can be seen with high K_m and low V_{max} values.

The results of aryl esterase activity discussed above for MNEE provides shows the kinetics profiles with the assumption (which was based on the specific activity data obtained from wild type and 10 variants) that two similar substrate molecules bind to two binding sites of MNEE. The kinetic scheme for two different substrates binding to different binding pockets is quite complex, especially for MNEE with four binding pockets with broad substrate acceptance. The kinetic equations for such heterotropic effects cannot be easily interpreted and is beyond the scope the current study. From the evolutionary angle, MNEE may be the result of genetic recombination of different DNA sequences encoding proteins with different kinetic capabilities. This molecular breeding resulted in the kinetic variability

among the nucleophilic elbowed binding pockets of MNEE emerged from the kinetic spectrum of the molecular parents.

Discussion

Our results showed that each nucleophilic elbow of *S. cellulosum's* MNEE forms a local active site possessing one or more enzyme activities. To the best of our knowledge, this is the first study in the literature that shows the presence of four binding pockets in a single protein domain and further proves the interplay of multiple nucleophilic elbows and catalytic promiscuity of esterases. Our analysis also showed that broad substrate specificity acquired by MNEE comes at the price of low reaction turnover number for its original feruloyl esterase activity, whereas the nature of the reaction catalyzed is unchanged. Even though enzymes cannot freely be mutated for acquiring novel substrate specificities without interruption of its original or starting substrate specificity, it has been proposed that evolutionarily adapting enzymes possess promiscuous activities[2]. Generally enzyme evolution results into the acquisition of novel activity and during this process suppression of the original activity is important to be selective in their action. Being highly selective is an advantage for the enzymes involved in metabolic pathways, but for the enzymes that are involved in plant polysaccharide degradation being promiscuous with broad substrate specificity can be an advantage. From the specific activity data (see Table 2 and Figure 1), it can be speculated that MNEE subsequently sacrificed its efficiency of FAE activity with the emerging new additional binding pockets in its protein scaffold with non-FAE activities.

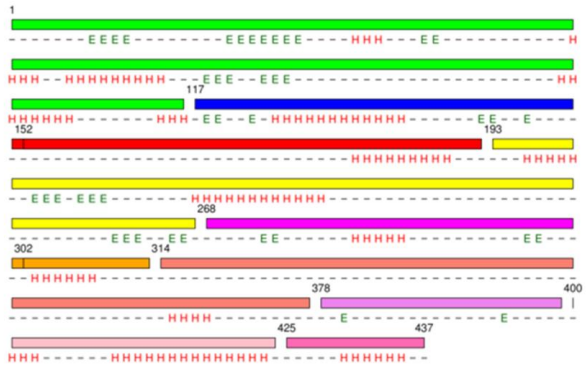
From specific activity data, it is also evident that binding pocket with nucleophilic elbow 5 is the most promiscuous site in MNEE. Presence of three consecutive serine residues in elbow 5 (see Figure 2) might have an effect on the flexibility for the role of nucleophile during the catalytic process and thus may have the structural advantage for activity on wide range of substrates. It has been shown that substrate specificity towards nonlipidic polar substrates in an esterase from *Penicillium purpurogenum* is modulated by multiple conformations of the catalytic serine residue that confers flexibility to the binding pocket interactions[19] and the presence of three consecutive serine residues in elbow 5 of *S. cellulosum* MNEE further provides much flexibility for the binding pocket interactions. In addition, catalytic promiscuity can be due to the presence of more than one conformation

for the entire active site. Ben-David *et al* have shown that individual active site residues perform multiple tasks that results from different catalytic modes offered by multiple conformations of the binding pocket[20].

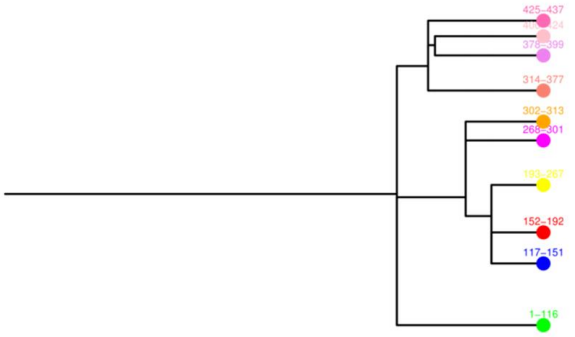
Substrate selectivity of an enzyme is dependent on the constellation of amino acid residues forming the active site and can be changed by the mutations that occur during the evolutionary trajectory. All enzymes that fall under α/β-hydrolase superfamily catalyze reactions involving a nucleophilic attack[21]. For example, serine acts as nucleophile in the reaction cycle of esterases. The 3D structures of α/β-hydrolase superfamily enzymes and their biochemical studies suggest that changes in the enzymatic reaction steps requires only a few amino acid substitutions in the active site[22]. Small changes in the constellation of amino acid residues in the binding pocket dictates the mechanistic step of the nucleophilic attack accordingly that results into the hydrolysis of ester[23], amide[24], alkyl halide[25], epoxide[26] and cyanide addition to aldehydes. In case of MNEE, all the four binding pockets showed ester hydrolysis capability with different substrate specificities. From our study, it is clear that enzymes with multiple nucleophilic elbows may not be specific in their function, but actually the elbows have the structural possibility to form distinct binding pockets that can act on different substrates.

Even though *S. cellulosum* produces different enzymes hydrolysing lignocellulosic materials, it would be an advantage to the organism to produce enzymes that displays broad substrate acceptance especially in the adaptation to novel habitats with a myriad of substrates and altered environmental conditions. A reasonable assumption to explain the presence of multiple nucleophilic elbows, forming distinct active sites, is that homologous recombinations or gene duplications resulted into a gene encoding MNEE and also the interactions of different substrates with the precursor of MNEE.

(a)



(b)



(c)

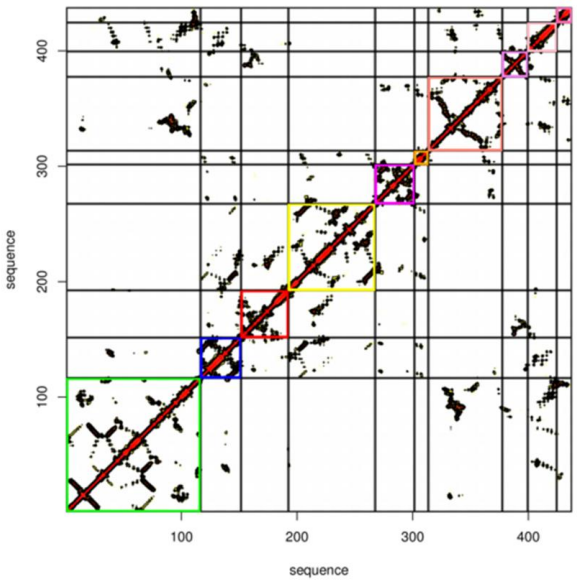


Figure 6. Suggested protein units in MNEE. (**a**) The ten PUs of MNEE shown in different colours. The secondary structure is represented by extended (E), helical (H) and other (-) types of secondary structure. (**b**) Hierarchical tree of splitting for PUs in MNEE. (**c**) Contact probability map for protein units of MNEE. Each splitting event is represented by vertical and horizontal lines. The map is based on the basic principle that each PU must have a high number of intra-PU contacts and a low number of inter-PU contacts.

Multi-functionality, generally seen in multi-domain proteins, is determined by the domain composition of the protein and their interactions[27]. Recombination results into shuffling of the domains to create proteins with new arrangements of domains. Thus, domains are the evolutionary units of proteins and their combinations have functional and spatial relationship of the protein structure[28]. Each domain can have an independent function or contribute to the common function of the protein. MNEE is a small protein comprised of a single domain and such proteins represent about one-third of proteins in a prokaryote proteome[29]. Investigation of MNEE protein 3D structure using Domain Reconstruction Algorithm[30] showed that MNEE is comprised of ten small protein units. A protein unit (PU) is an intermediate level of organization between secondary structure elements and domains. The ten PUs of MNEE and their hierarchical splitting is shown in Figure 6. We therefore consider MNEE as an intermediate enzyme with broad substrate acceptance in the evolutionary process that may be a result of the recombination of protein coding DNA sequences.

Does the protein evolution have bias between metabolic enzymes which are generally assumed as 'specialists' and secreted biomass degrading enzymes which are generally assumed as 'generalists' [31,32]? How common is the promiscuous behaviour in metabolic enzymes? A very recent study by *Nam et al* showed that an estimated 37% of enzymes in *Escherichia coli* are generalists and exhibit substrate promiscuity [31]. Why a fraction of generalist metabolic enzymes are maintained in the evolutionary path? It might be the flux of metabolites that renders selective pressure on the organism to carry out the different catalytic processes while maintaining the low levels of total enzyme concentration. The same assumption can be applied to the activities observed in the MNEE of *S. cellulosum* Soce56. It is evident from the enzyme activity data, that MNEE has low feruloyl esterase activity. This might be due to the presence of low amount of feruloyl groups compared to other ester bonds in the plant biomass [33-35] present in the habitat of *S. cellulosum* Soce56. This might have provided selective pressure to retain low amount of the feruloyl esterase activity in MNEE. Furthermore, the classification of enzymes with multiple active sites arisen from the selective pressure remains challenging and it would be counterproductive to attempt a classification based on function or structure. The points discussed above regarding the evolutionary space of protein sequence-structure-function is complex and in many ways defy classification systems based on only sequence or structural similarity[8]. As

shown in Figure 7, difficulties in defining the function or substrate specificity of an enzyme occur at all levels of classification hierarchy, due to the promiscuous nature of proteins in the evolutionary path.

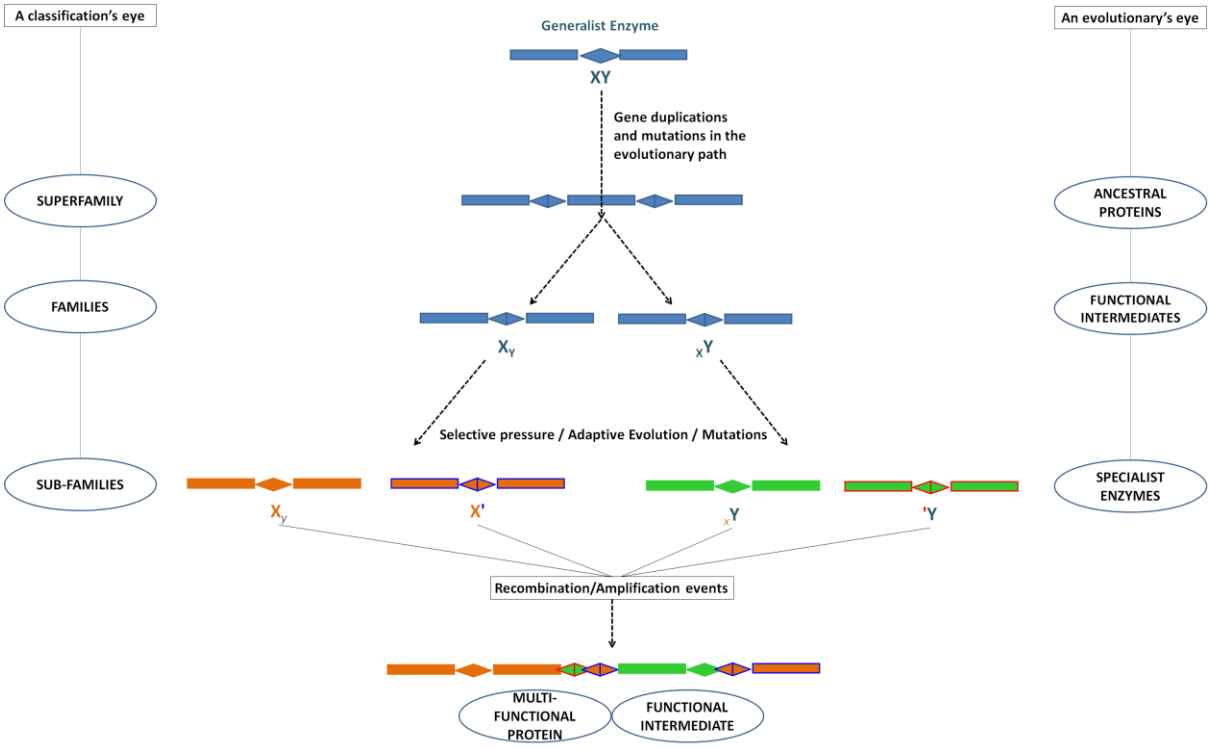


Figure 7. The emergence of specialist enzymes from the ancestral generalist enzymes in the view of classification and evolution[70-73]. Classification is feasible or functionally meaningful till the evolutionary path of specialist enzymes with defined substrate specificity. A challenge is posed for the classification system, when the evolutionary path gives rise to proteins with multiple active sites through duplication and divergence of genes. Novel enzymes can be further evolved from the specialized enzymes by entering again the phase of functional intermediates.

The scheme depicted is for a hypothetical generalist enzyme that possesses two different activities. The starting point in the generation of specialist enzymes is a generalist enzyme (**XY**), where duplication of genes encoding it leads to division of its ancestral functions and generate enzymes ($\mathbf{X}_{\mathbf{Y}}$, $_{\mathbf{X}}\mathbf{Y}$) which are catalytically promiscuous with varied affinity towards substrates. Further mutations guided by adaptive evolution may give rise to specialist enzymes. Even though the active sites of few specialist enzymes ($\mathbf{X}_{y}$, $_{x}\mathbf{Y}$) are very specific to a set of substrates, molecules that bear resemblance to their natural substrates can bind with lower affinity. When such molecules bind in correct orientations the reactive functional groups of the enzymes active site catalyze chemical reactions, which gives the promiscuous property to specialist enzymes. Further mutations from this point may also give rise to more specific enzymes (**X'**, **'Y**) evolved to catalyze a reaction with more specificity and catalytic efficiency. Still, multi-functional enzymes can be emerged from specialist enzymes by gene duplication and recombination events. These multi-functional enzymes further may repeat the entire cycle described above giving rise to novel and multiple promiscuous enzymes.

To reveal whether the mutants represent different conformers, or are identical to the wild type MNEE, we performed the Circular Dichroism spectroscopy (as a function of temperature) of the variant with five mutated elbows and the wild type MNEE. As shown in Figure 8, the deconvolution of CD spectra indicates that the temperature stability of the wild type MNEE and the variant with five mutated elbows is similar; and further it shows that they are not representing different conformers. It is worth to mention at this point that we have not observed any enzyme activity in the variant with five mutated elbows.

Multidomian multifunctional enzymes can be found in nature[36] and successful attempts have been made by researchers to synthesize chimeric, bifunctional or multifunctional enzymes[37-39] for efficient degradation of plant biomass. Several multidomain multifunctional enzymes have been discovered in microorganisms like *Caldicellulosiruptor saccharolytics* that live in extreme conditions of hot springs[40]. Domain-shuffling is the generally assumed reason in the evolution of such enzymes to obtain selective growth advantage[41,42]. Recently Cheng *et al* collected information of 6,799 multi-functional enzymes reported in the literature and observed that bacteria have relatively more multi-functional enzymes[43]. Bacteria, which are simple life forms (e.g. *S. cellulosum*), have limited protein-coding capacity. None of the enzymes analyzed in the study by Cheng *et al* possess four active sites in a single domained protein. As shown in the present study, identification of multiple nucleophilic elbows that forms distinct binding pockets in enzymes helps to identify new catalytic sites and further understanding of multi-dimensional nature of enzyme evolution to apply in enzyme engineering.

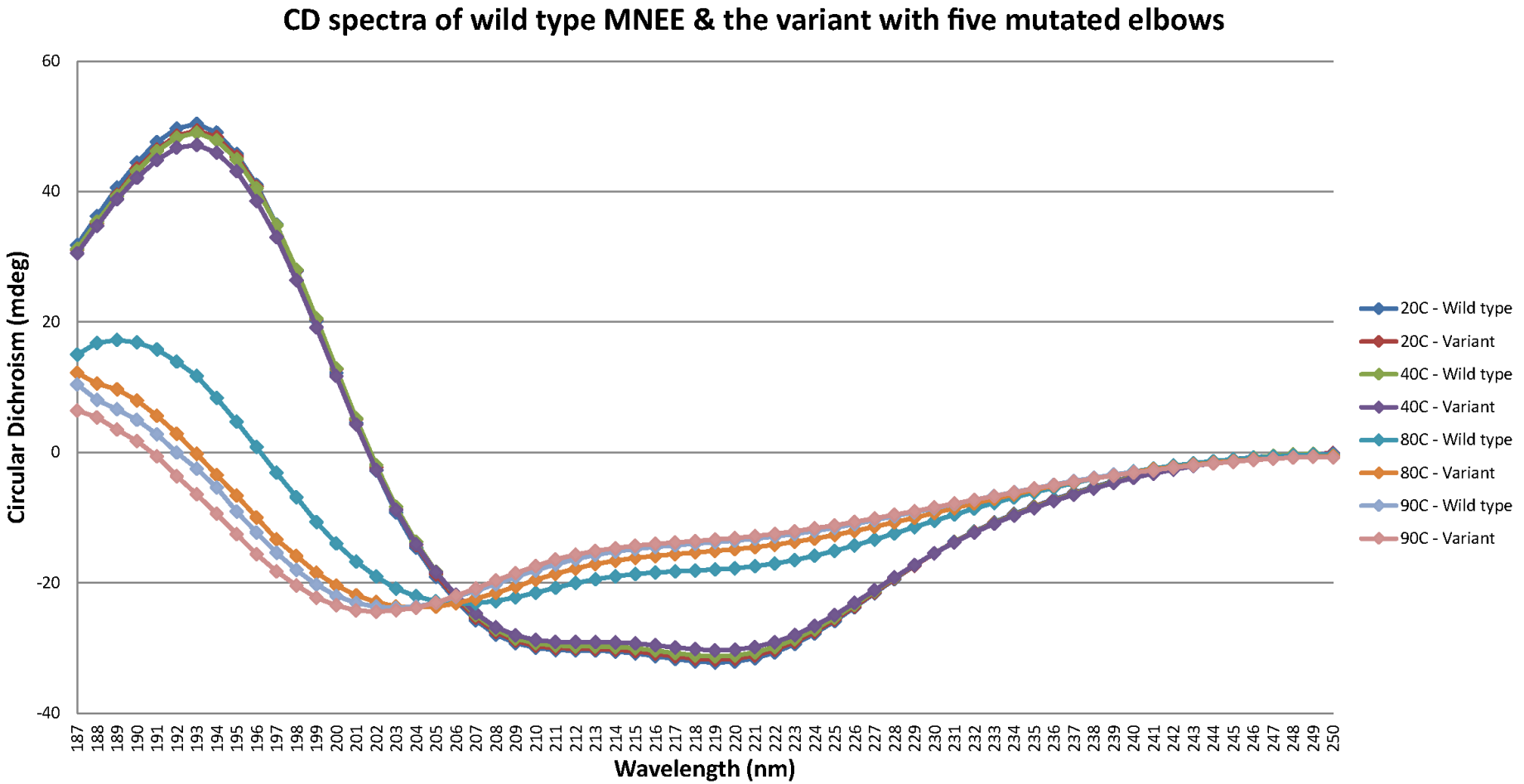


Figure 8. CD spectra of wild type MNEE & the variant with five mutated elbows. The deconvolution CD spectra indicates that the temperature stability (structural) of the wild type MNEE and the variant with five mutated elbows is similar; and further it shows that they are not representing different conformers.

Methods

Cloning and expression

The gene sce3277 encoding MNEE was commercially synthesized using codon optimization for efficient protein expression in *Escherichia coli* and incorporating six histidine residues at the carboxyl- terminus to facilitate protein purification by nickel-sepharose (Ni-Sp) chromatography (DNA2.0 Inc., Menlo Park, CA 94025) and inserted into the expression vector pJexpress401. The genes encoding 10 variants of MNEE were commercially synthesized (NZYTech Ltd, Portugal) and inserted into the respective expression vector pJexpress401. The expression constructs were transformed into *E. coli* strain BL21-Gold DE3 (Agilent Technologies, United States), that incorporate major improvements over the original BL21 strain and feature the Hte phenotype present in the highest efficiency Stratagene competent cell strain, XL10-Gold.

Enzyme activity assays

In every measurement, the effect of nonenzymatic hydrolysis of substrates was taken into consideration and subtracted from the value measured when the enzyme was added. Measurements were carried out in at least triplicates.

***Feruloyl esterase*:** The expressed recombinant enzymes were tested against 15 methyl cinnamate esters (obtained from Aapin Chemicals, Oxon, UK) viz., Methyl ferulate or Methyl 4-hydroxy-3-methoxy cinnamate (MFA), Methyl caffeate or Methyl 3,4-dihydroxy cinnamate (MCA), Methyl p-coumarate or Methyl 4-hydroxy cinnamate (MPC), Methyl sinapate or Methyl 4-hydroxy-3,5-dimethoxy cinnamate (MSA), Methyl 2-hydroxy cinnamate (M2C), Methyl 3-hydroxy cinnamate (M3C), Methyl cinnamate (MC), Methyl 3,4,5-trimethoxy cinnamate (MTM), Methyl 2-methoxy cinnamate (M2M), Methyl 3-methoxy cinnamate (M3M), Methyl 4-methoxy cinnamate (M4M), Methyl 3,4-dimethoxy cinnamate (M34DC), Methyl 3,5-dimethoxy cinnamate (M35DC), Methyl 3-hydroxy-4-methoxy cinnamate (M34MC) and Methyl 4-hydroxy-3-methoxy phenyl propionate (M43PP).

The assay was based on a previously reported spectrophotometric method for determination of FAE activity recommended by Biocatalysts Limited, Wales, UK [44,45]. Enzyme activity was calculated using standard curves of FA and MFA according to the equation:

$$\boldsymbol{Activity}\ (\boldsymbol{U}) = \frac{[(\mathrm{OD0}\ min - \mathrm{OD30}\ min) - (\mathrm{ODblank\,0}\ min - \mathrm{ODblank\,30}\ min)] \times V\mathrm{reaction} \times \mathrm{Dilution}}{[(\varepsilon\mathrm{MFA} \times l - \varepsilon\mathrm{FA} \times l) \times V\mathrm{sample}]}$$

In which, $OD_{0\ min}$ is absorbance of reaction system at 0 min, $OD_{30\ min}$ is an absorbance of reaction system at 30 min, $OD_{blank\ 0\ min}$ is an absorbance of substrate-containing at 0 min, $OD_{blank\ 30\ min}$ is an absorbance of substrate-containing at 30 min, $V_{reaction}$ is a volume of reaction system, V_{sample} is a volume of sample, ε_{MFA} is the extinction coefficient of MFA; ε_{FA} is the extinction coefficient of FA, and *l* is the pathlength.

The absorption spectra (FLUOstar Omega, BMG LABTECH, Germany) of the methyl esters of cinnamic acids and their hydrolysis products was monitored in 240 – 480 nm and their absorption maxima were determined [46]. FAE activity was expressed in Units (U); 1 U is equal to 1 nmol of ferulic acid or respective cinnamic acid released in 1 ml of the reaction medium after 1 min of incubation[44].

***Aryl esterase*:** Aryl esterse activity was determined by measuring the amount of phenol released from phenyl acetate[47] at 270nm at 45°C using a spectrophotometer fitted with constant-temperature controlled cell holder. The substrate solution was prepared by mixing 0.1 ml of a methanolic solution of Phenyl acetate and 0.8 ml of prewarmed 100mM sodium phosphate buffer in cells held in cell holder at 45°C. After 1 minute, 0.1 ml of enzyme solution was added to this substrate solution. The same substrate buffer solution without the enzyme was used in a blank. One unit is defined as the amount of enzyme releasing 1 µmol of phenol from Phenyl acetate.

Acetyl esterase: Acetyl esterase activity spectrophotometric assay was based on the liberation of p-nitrophenol from pNP-acetate (2mM) at 405nm[48,49]. The reactions were performed at 45°C and pH 6.5 in 100mM phosphate buffer containing enzyme and substrate. One unit was defined as the amount of enzyme that liberated 1 µmol pNP per minute.

***Carboxyl esterase*:** A quantitative standard enzyme assay was carried out using an artificial chromogenic substrate, PNP-palmitate (C_{16})[50]. The enzyme reaction was started by the addition of 0.1 ml of freshly prepared and prewarmed PNP-palmitate solution as a substrate to 0.1 ml enzyme solution and 0.8 ml of prewarmed 100 mM sodium phosphate buffer (pH 7.0) containing sodium taurocholate (2 mg/ml) and gum arabic (1 mg/ml) as

emulsifiers at 45 °C. One unit of carboxylesterase activity was defined as the amount releasing 1 µmol of *p*-nitrophenol per min from PNP esters at 405nm.

***Paraben Esterase*:** Paraben esterase activity was assayed using HPLC method[51]: 0.05 ml of 10 mM ethyl 4-hydroxybenzoate (EHB) or Methyl Paraben (MP) dissolved in dimethyl sulfoxide was added to 0.9 ml of 50 mM sodium phosphate buffer (pH 7.0). Following this, 0.1 ml of enzyme source was added. The reaction mixture was incubated for 10 min at 45°C, and the reaction was terminated by incubating in boiling water for 5 min. We detected HBA released during this process using high-performance liquid chromatography (Dionex–Ultimate 3000) with ultraviolet detection (280 nm) using a Kinetex 2.6 u C18 100 × 4.6 mm column with a flow rate of 1 ml/minute at 40^{o}C . The mobile phase was 50 mM acetate buffer (pH 4.0) prepared in water and acetonitrile (70:30). One unit (U) of enzyme activity was defined as the amount of enzyme that releases 1 µmol of HBA/min under the assay conditions described above.

***Paraoxonase*:** Paraoxonase activity was determined by measuring the hydrolysis of an organophosphate substrate, paraoxon[47] to yield *p*-nitrophenol at 405nm at 45^{o}C. The basal reaction mixture contained 0.1ml of paraoxon solution (1mM), 0.1 ml of enzyme solution and 0.8ml of 100mM sodium phosphate buffer (pH 7.0). One unit of Paraoxonase is defined as the amount of the enzyme required to liberate 1 µmol of *p*-nitrophenol per minute.

***PHB depolymerase*:** The measurement of PHB depolymerase activity using PHB powder was performed as described previously[52]. PHB powder, at a concentration of 300 µg PHB/mL suspended in 50 mM Tricine−NaOH containing 0.05 mM $CaCl_2$ at pH = 8.0, was suspended by sonication for 20 min. The resulting suspension was used for determining depolymerase activity as a standard method. The reaction was initiated by the addition of enzyme to 1.0 mL of PHB suspension in a 1.5 mL (10 mm) disposable cuvette at room temperature. After an initial measurement of absorption at 660 nm, the assay mixture was incubated at 45 °C with the enzyme solution. The change in absorption was measured as a function of time using a spectrophotometer fitted with constant-temperature controlled cell holder. A unit of enzyme activity is defined under these conditions as a change in absorption of 0.001 units per minute.

Circular dichroism spectroscopy

The far-UV Circular Dichroism (CD) measurements were carried on a Chirascan™ CD Spectrometer equipped with a thermostated cell holder (Applied Photophysics Limited, UK). Chirascan™ CD Spectrometer is endowed with photon flux for a 1nm bandwidth in excess of 10^{13} per second for all UV wavelengths from 360 nm to 180 nm. During CD spectroscopy analysis, respective purified 6×His N-terminally tagged recombinant proteins (0.5 mg/mL) were resuspended in the buffer and analysed at respective temperature (20°C, 40°C and 90°C). UV CD spectra between 185 and 250 nm were collected with a data pitch of 0.1 nm, bandwidth of 2.0 nm and scanning speed of 50 nm/min[53]. Each sample was measured in triplicate and the data points between 190 – 240 nm were analyzed using the K2D2 method. K2D2 method uses a self-organized map of spectra from proteins with known structure to deduce a map of protein secondary structure that is used to develop the predictions[54-56].

Protein structure prediction

In the absence of any resolved X-ray or NMR structures, the three-dimensional atomic models for *S. cellulosum* MNEE were modeled from multiple threading alignments[57] and iterative structural assembly simulations using the I-TASSER algorithm, an extension of the previous TASSER method[58-62]. Structure refinement of the modeled structures was carried out using the Discovery Studio software suite version 3.0 (Accelrys Inc, USA). The Prepare Protein protocol package in Discovery Studio suite was used for inserting missing atoms in incomplete residues, modeling missing loop regions[63], deleting alternate conformations (disorder), standardizing atom names, and protonating titratable residues using predicted pKs[64]. The Side-Chain Refinement protocol was used for each structure to optimize the protein side-chain conformation based on systematic searching of side-chain conformation and CHARMM Polar H energy minimization[65] using the ChiRotor algorithm[66]. Smart Minimizer algorithm was used for the minimization process which performs 1000 steps of Steepest Descent with a RMS gradient tolerance of 3, followed by Conjugate Gradient minimization for faster convergence towards a local minimum[67]. Structure evaluations were carried out using DOPE, which is an atomic based statistical potential in MODELER package for model evaluation and structure prediction[68]. Structure verifications were carried out using VerifyProtein-Profiles-3D that allows evaluating the fitness of a protein sequence in its current 3D environment[69].

References

1 Udatha, D. B., Mapelli, V., Panagiotou, G. & Olsson, L. Common and Distant Structural Characteristics of Feruloyl Esterase Families from Aspergillus oryzae. *PloS one* **7**, e39473, doi:10.1371/journal.pone.0039473 (2012).

2 Khersonsky, O., Roodveldt, C. & Tawfik, D. S. Enzyme promiscuity: evolutionary and mechanistic aspects. *Current opinion in chemical biology* **10**, 498-508, doi:10.1016/j.cbpa.2006.08.011 (2006).

3 Cantarel, B. L. *et al.* The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic acids research* **37**, D233-238, doi:10.1093/nar/gkn663 (2009).

4 Turcot-Dubois, A. L. *et al.* Long-term evolution of the CAZY glycosyltransferase 6 (ABO) gene family from fishes to mammals--a birth-and-death evolution model. *Glycobiology* **17**, 516-528, doi:10.1093/glycob/cwm016 (2007).

5 Cantarel, B. L. *et al.* The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* **37**, D233-D238, doi:Doi 10.1093/Nar/Gkn663 (2009).

6 Udatha, D. B., Sugaya, N., Olsson, L. & Panagiotou, G. How well do the substrates KISS the enzyme? Molecular docking program selection for feruloyl esterases. *Scientific reports* **2**, 323, doi:10.1038/srep00323 (2012).

7 McAuley, K. E., Svendsen, A., Patkar, S. A. & Wilson, K. S. Structure of a feruloyl esterase from Aspergillus niger. *Acta Crystallogr D* **60**, 878-887, doi:Doi 10.1107/S0907444904004937 (2004).

8 Udatha, D. B., Kouskoumvekaki, I., Olsson, L. & Panagiotou, G. The interplay of descriptor-based computational analysis with pharmacophore modeling builds the basis for a novel classification scheme for feruloyl esterases. *Biotechnology advances* **29**, 94-110, doi:10.1016/j.biotechadv.2010.09.003 (2011).

9 Ekici, O. D., Paetzel, M. & Dalbey, R. E. Unconventional serine proteases: variations on the catalytic Ser/His/Asp triad configuration. *Prot Sci* **17**, 2023-2037, doi:10.1110/ps.035436.108 (2008).

10 Lai, K. K. *et al.* An inserted alpha/beta subdomain shapes the catalytic pocket of Lactobacillus johnsonii cinnamoyl esterase. *PloS one* **6**, e23269, doi:10.1371/journal.pone.0023269 (2011).

11 Schneiker, S. *et al.* Complete genome sequence of the myxobacterium Sorangium cellulosum. *Nature biotechnology* **25**, 1281-1289, doi:10.1038/nbt1354 (2007).

12 Depristo, M. A. The subtle benefits of being promiscuous: adaptive evolution potentiated by enzyme promiscuity. *HFSP journal* **1**, 94-98, doi:10.2976/1.2754665 (2007).

13 Kurtovic, S., Shokeer, A. & Mannervik, B. Emergence of novel enzyme quasi-species depends on the substrate matrix. *Journal of molecular biology* **382**, 136-153, doi:DOI 10.1016/j.jmb.2008.07.003 (2008).

14 Gerlt, J. A. & Babbitt, P. C. Enzyme (re)design: lessons from natural evolution and computation. *Current opinion in chemical biology* **13**, 10-18, doi:10.1016/j.cbpa.2009.01.014 (2009).

15 Voigt, C. A., Kauffman, S. & Wang, Z. G. Rational evolutionary design: The theory of in vitro protein evolution. *Adv Protein Chem* **55**, 79-160 (2001).

16 Atkins, W. M. Non-Michaelis-Menten kinetics in cytochrome P450-catalyzed reactions. *Annual Review of Pharmacology and Toxicology* **45**, 291-310, doi:DOI 10.1146/annurev.pharmtox.45.120403.100004 (2005).

17 Cornish-Bowden, A. & Cardenas, M. L. Specificity of Non-Michaelis-Menten Enzymes: Necessary Information for Analyzing Metabolic Pathways. *J Phys Chem B* **114**, 16209-16213, doi:Doi 10.1021/Jp106968p (2010).

18 King, E. L. & Altman, C. A Schematic Method of Deriving the Rate Laws for Enzyme-Catalyzed Reactions. *J Phys Chem-Us* **60**, 1375-1378 (1956).
19 Ghosh, D. *et al.* Multiple conformations of catalytic serine and histidine in acetylxylan esterase at 0.90 A. *The Journal of biological chemistry* **276**, 11159-11166, doi:10.1074/jbc.M008831200 (2001).
20 Ben-David, M. *et al.* Catalytic versatility and backups in enzyme active sites: the case of serum paraoxonase 1. *Journal of molecular biology* **418**, 181-196, doi:10.1016/j.jmb.2012.02.042 (2012).
21 Ollis, D. L. *et al.* The alpha/beta hydrolase fold. *Protein engineering* **5**, 197-211 (1992).
22 Smith, A. J. *et al.* Structural reorganization and preorganization in enzyme active sites: comparisons of experimental and theoretically ideal active site geometries in the multistep serine esterase reaction cycle. *J Am Chem Soc* **130**, 15361-15373, doi:10.1021/ja803213p (2008).
23 Cygler, M. *et al.* Relationship between Sequence Conservation and 3-Dimensional Structure in a Large Family of Esterases, Lipases, and Related Proteins. *Protein Science* **2**, 366-382 (1993).
24 Carter, P. & Wells, J. A. Dissecting the Catalytic Triad of a Serine Protease. *Nature* **332**, 564-568 (1988).
25 Janssen, D. B. Evolving haloalkane dehalogenases. *Current opinion in chemical biology* **8**, 150-159, doi:DOI 10.1016/j.cpba.2004.02.012 (2004).
26 Archelas, A. & Furstoss, R. Epoxide hydrolases: new tools for the synthesis of fine organic chemicals. *Trends in biotechnology* **16**, 108-116 (1998).
27 Apic, G., Huber, W. & Teichmann, S. A. Multi-domain protein families and domain pairs: comparison with known structures and a random model of domain recombination. *Journal of structural and functional genomics* **4**, 67-78 (2003).
28 Vogel, C., Berzuini, C., Bashton, M., Gough, J. & Teichmann, S. A. Supra-domains: evolutionary units larger than single protein domains. *Journal of molecular biology* **336**, 809-823, doi:10.1016/j.jmb.2003.12.026 (2004).
29 Teichmann, S. A., Park, J. & Chothia, C. Structural assignments to the Mycoplasma genitalium proteins show extensive gene duplications and domain rearrangements. *Proc Natl Acad Sci U S A* **95**, 14658-14663 (1998).
30 Gelly, J. C., de Brevern, A. G. & Hazout, S. 'Protein Peeling': an approach for splitting a 3D protein structure into compact fragments. *Bioinformatics* **22**, 129-133, doi:10.1093/bioinformatics/bti773 (2006).
31 Nam, H. *et al.* Network context and selection in the evolution to enzyme specificity. *Science* **337**, 1101-1104, doi:10.1126/science.1216861 (2012).
32 Copley, S. D. Toward a systems biology perspective on enzyme evolution. *The Journal of biological chemistry* **287**, 3-10, doi:10.1074/jbc.R111.254714 (2012).
33 Caffall, K. H. & Mohnen, D. The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydrate research* **344**, 1879-1900, doi:DOI 10.1016/j.carres.2009.05.021 (2009).
34 Heredia, A., Jimenez, A. & Guillen, R. Composition of Plant-Cell Walls. *Z Lebensm Unters For* **200**, 24-31, doi:Doi 10.1007/Bf01192903 (1995).
35 Faulds, C. B. What can feruloyl esterases do for us? *Phytochem Rev* **9**, 121-132, doi:DOI 10.1007/s11101-009-9156-2 (2010).
36 Cepeljnik, T., Rincon, M. T., Flint, H. J. & Marinsek-Logar, R. Xyn11A, a multidomain multicatalytic enzyme from Pseudobutyrivibrio xylanivorans Mz5T. *Folia microbiologica* **51**, 263-267 (2006).
37 Levasseur, A. *et al.* Construction of engineered bifunctional enzymes and their overproduction in Aspergillus niger for improved enzymatic tools to degrade agricultural by-

products. *Applied and environmental microbiology* **71**, 8132-8140, doi:10.1128/AEM.71.12.8132-8140.2005 (2005).
38 Dodd, D. *et al.* Biochemical analysis of a beta-D-xylosidase and a bifunctional xylanase-ferulic acid esterase from a xylanolytic gene cluster in Prevotella ruminicola 23. *J Bacteriol* **191**, 3328-3338, doi:10.1128/JB.01628-08 (2009).
39 Shi, P. *et al.* Paenibacillus sp. strain E18 bifunctional xylanase-glucanase with a single catalytic domain. *Applied and environmental microbiology* **76**, 3620-3624, doi:10.1128/AEM.00345-10 (2010).
40 Gibbs, M. D. *et al.* Multidomain and multifunctional glycosyl hydrolases from the extreme thermophile Caldicellulosiruptor isolate Tok7B.1. *Current microbiology* **40**, 333-340 (2000).
41 Millward-Sadler, S. J. *et al.* Novel cellulose-binding domains, NodB homologues and conserved modular architecture in xylanases from the aerobic soil bacteria Pseudomonas fluorescens subsp. cellulosa and Cellvibrio mixtus. *Biochem J* **312 (Pt 1)**, 39-48 (1995).
42 Birch, P. R., Sims, P. F. & Broda, P. Substrate-dependent differential splicing of introns in the regions encoding the cellulose binding domains of two exocellobiohydrolase I-like genes in Phanerochaete chrysosporium. *Applied and environmental microbiology* **61**, 3741-3744 (1995).
43 Cheng, X. Y. *et al.* A Global Characterization and Identification of Multifunctional Enzymes. *PloS one* **7**, e38979 (2012).
44 Szwajgier, D. & Jakubczyk, A. BIOTRANSFORMATION OF FERULIC ACID BY LACTOBACILLUS ACIDOPHILUS KI AND SELECTED BIFIDOBACTERIUM STRAINS. *Acta Scientiarum Polonorum Technologia Alimentaria* **9**, 45-59 (2010).
45 Yue, Q. *et al.* Feruloyl and acetyl esterase production of an anaerobic rumen fungus Neocallimastix sp YQ2 effected by glucose and soluble nitrogen supplementations and its potential in the hydrolysis of fibrous feedstuffs. *Anim Feed Sci Tech* **153**, 263-277, doi:DOI 10.1016/j.anifeedsci.2009.06.019 (2009).
46 Fazary, A. E. & Ju, Y. H. Production, partial purification and characterization of feruloyl esterase by Aspergillus awamori in submerged fermentation. *Biotechnol J* **3**, 1264-1275, doi:10.1002/biot.200800101 (2008).
47 Park, Y. J., Yoon, S. J. & Lee, H. B. A novel thermostable arylesterase from the archaeon Sulfolobus solfataricus P1: purification, characterization, and expression. *J Bacteriol* **190**, 8086-8095, doi:10.1128/JB.00803-08 (2008).
48 Drzewiecki, K. *et al.* Hyperthermostable acetyl xylan esterase. *Microb Biotechnol* **3**, 84-92, doi:10.1111/j.1751-7915.2009.00150.x (2010).
49 Degrassi, G., Okeke, B. C., Bruschi, C. V. & Venturi, V. Purification and characterization of an acetyl xylan esterase from Bacillus pumilus. *Appl Environ Microb* **64**, 789-792 (1998).
50 Park, Y. J., Choi, S. Y. & Lee, H. B. A carboxylesterase from the thermoacidophilic archaeon Sulfolobus solfataricus P1; purification, characterization, and expression. *Biochim Biophys Acta* **1760**, 820-828, doi:10.1016/j.bbagen.2006.01.009 (2006).
51 Koseki, T., Mihara, K., Murayama, T. & Shiono, Y. A novel Aspergillus oryzae esterase that hydrolyzes 4-hydroxybenzoic acid esters. *Febs Letters* **584**, 4032-4036, doi:DOI 10.1016/j.febslet.2010.08.021 (2010).
52 Scherer, T. M., Fuller, R. C., Goodwin, S. & Lenz, R. W. Enzymatic hydrolysis of oligomeric models of poly-3-hydroxybutyrate. *Biomacromolecules* **1**, 577-583, doi:Doi 10.1021/Bm000012c (2000).
53 Guo, J. *et al.* Characterization of a novel esterase Rv0045c from Mycobacterium tuberculosis. *PLoS One* **5**, doi:10.1371/journal.pone.0013143 (2010).
54 Perez-Iratxeta, C. & Andrade-Navarro, M. A. K2D2: estimation of protein secondary structure from circular dichroism spectra. *BMC Struct Biol* **8**, 25, doi:10.1186/1472-6807-8-25 (2008).

55 Andrade, M. A., Chacon, P., Merelo, J. J. & Moran, F. Evaluation of Secondary Structure of Proteins from Uv Circular-Dichroism Spectra Using an Unsupervised Learning Neural-Network. *Protein Engineering* **6**, 383-390 (1993).
56 Merelo, J. J., Andrade, M. A., Prieto, A. & Moran, F. Proteinotopic Feature Maps. *Neurocomputing* **6**, 443-454 (1994).
57 Wu, S. & Zhang, Y. LOMETS: a local meta-threading-server for protein structure prediction. *Nucleic Acids Res* **35**, 3375-3382, doi:10.1093/nar/gkm251 (2007).
58 Ananthan, S., Zhang, W. & Hobrath, J. V. Recent advances in structure-based virtual screening of G-protein coupled receptors. *AAPS J* **11**, 178-185, doi:10.1208/s12248-009-9094-3 (2009).
59 Zhang, Y. I-TASSER: Fully automated protein structure prediction in CASP8. *Proteins-Structure Function and Bioinformatics* **77**, 100-113, doi:Doi 10.1002/Prot.22588 (2009).
60 Zhang, Y. I-TOSSER server for protein 3 D structure prediction. *Bmc Bioinformatics* **9**, -, doi:10.1186/1471-2105-9-40 (2008).
61 Zhang, Y. Template-based modeling and free modeling by I-TASSER in CASP7. *Proteins-Structure Function and Bioinformatics* **69**, 108-117, doi:Doi 10.1002/Prot.21702 (2007).
62 Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* **5**, 725-738, doi:Doi 10.1038/Nprot.2010.5 (2010).
63 Spassov, V. Z., Flook, P. K. & Yan, L. LOOPER: a molecular mechanics-based algorithm for protein loop prediction. *Protein Eng Des Sel* **21**, 91-100, doi:10.1093/protein/gzm083 (2008).
64 Spassov, V. Z. & Yan, L. A fast and accurate computational approach to protein ionization. *Protein science : a publication of the Protein Society* **17**, 1955-1970, doi:10.1110/ps.036335.108 (2008).
65 Neria, E., Fischer, S. & Karplus, M. Simulation of activation free energies in molecular systems. *J Chem Phys* **105**, 1902-1921 (1996).
66 Spassov, V. Z., Yan, L. & Flook, P. K. The dominant role of side-chain backbone interactions in structural realization of amino acid code. ChiRotor: A side-chain prediction algorithm based on side-chain backbone interactions. *Protein Science* **16**, 494-506, doi:Doi 10.1110/Ps.062447107 (2007).
67 Fletcher, R. & Reeves, C. M. Function Minimization by Conjugate Gradients. *Comput J* **7**, 149-154 (1964).
68 Shen, M. Y. & Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Science* **15**, 2507-2524, doi:Doi 10.1110/Ps.062416606 (2006).
69 Eisenberg, D., Luthy, R. & Bowie, J. U. VERIFY3D: Assessment of protein models with three-dimensional profiles. *Method Enzymol* **277**, 396-404 (1997).
70 Park, H. S. *et al.* Design and evolution of new catalytic activity with an existing protein scaffold. *Science* **311**, 535-538, doi:DOI 10.1126/science.1118953 (2006).
71 Hughes, A. L. The Evolution of Functionally Novel Proteins after Gene Duplication. *P Roy Soc B-Biol Sci* **256**, 119-124, doi:DOI 10.1098/rspb.1994.0058 (1994).
72 Bergthorsson, U., Andersson, D. I. & Roth, J. R. Ohno's dilemma: Evolution of new genes under continuous selection. *P Natl Acad Sci USA* **104**, 17004-17009, doi:DOI 10.1073/pnas.0707158104 (2007).
73 Freilich, S. *et al.* The complement of enzymatic sets in different species. *Journal of molecular biology* **349**, 745-763, doi:DOI 10.1016/j.jmb.2005.04.027 (2005).
74 Pandit, S. B. & Srinivasan, N. Identification and analysis of a new family of bacterial serine proteinases. *In silico biology* **4**, 563-572 (2004).

Acknowledgements

Funding support for this research was provided by The Swedish Research Council (Vetenskapsrådet Grant no: 2008-2955). GPA would like to thank eSRT of Integrative Biology, The University of Hong Kong

Author contributions

DGU, GPA and LO conceived and designed the research. DGU performed the experiments. KMM assisted DGU in performing only the enzyme assays. DGU wrote the manuscript and LO thoroughly assisted in writing and editing the manuscript. KMM wrote the enzyme assays part of the manuscript as part of her master's thesis supervised by DGU. All authors analyzed the data, commented on the results and contributed to the scientific discussions.

Competing financial interests

The authors declare that no competing financial interests exists.