Esterase Sequence Composition Patters As Inspiration For The Design Of Short Catalytic Peptides

Daniela Kalafatovic[1], Patrizia Janković[1], Marko Babić[1], Ana Sofia Pina[2], Goran Mauša[3]

[1]University of Rijeka, Department of Biotechnology, Rijeka, Croatia
[2]Instituto de Tecnologia Química e Biológica António Xavier (ITQB), Universidade Nova de Lisboa, Oeiras, Portugal
[3]University of Rijeka, Faculty of Engineering, Rijeka, Croatia

UIP-2019-04-7999

Statistical analysis of 22 EC 3.1 esterases with unique mechanisms based on the catalytic triads and their 974 homologues that led to the identification of catalytic triad microenvironments [1]

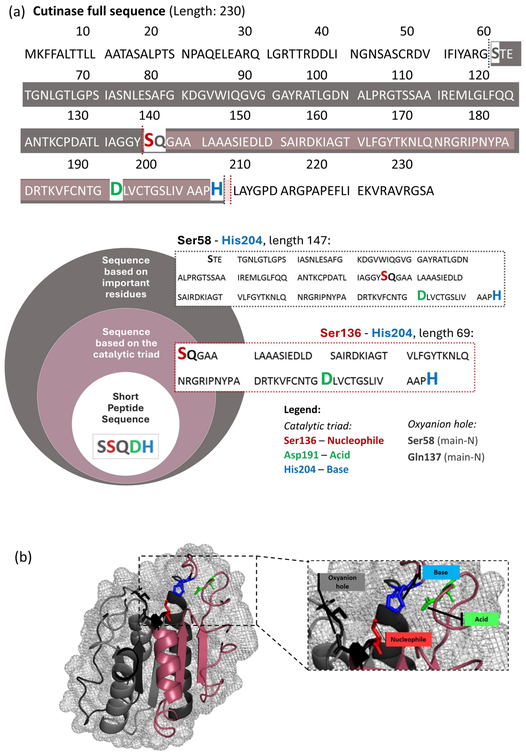


Fig 1. (a) Schematic representation of the full (Met1-Ala230) sequence (Uniprot P00590), long (Ser58-His204), medium (Ser136-His204), and short (Ser58, Ser136, Gln137, Asp191, His204) fragments of cutinase (1AGY). (b) 1AGY (3D) with highlighted triad members Ser, Asp, and His, obtained in PyMol.

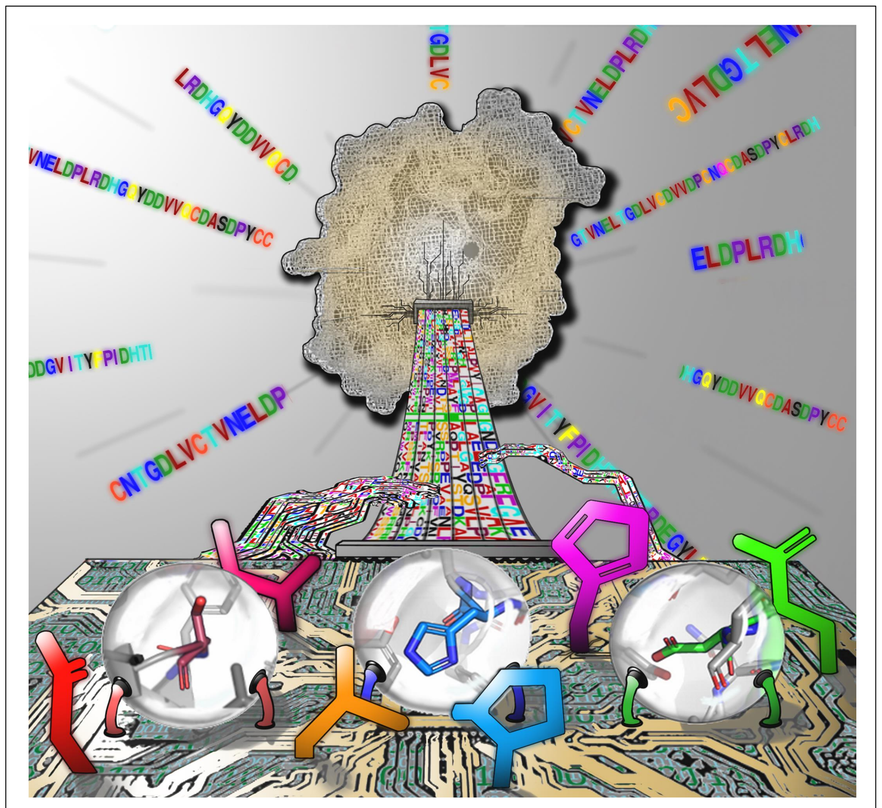


Fig 2. Esterase sequence composition patterns obtained through statistical analysis of EC 3.1 enzymes (22 with unique mechanisms and their 974 homologues).

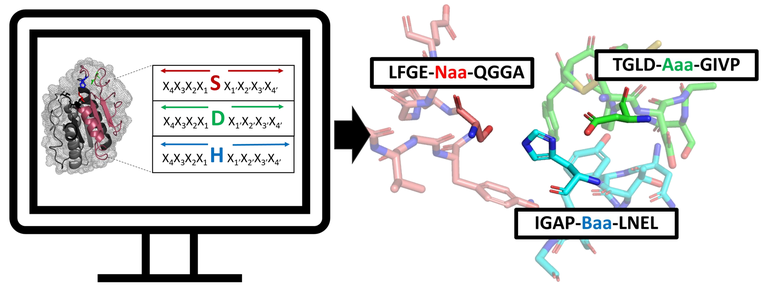


Fig 3. Triad amino acids microenvironment determination

Table 1. Consensus Sequences Obtained from *D2* Data Set and Their Homologues (*D3*)[a]

	X_4	X_3	X_2	X_1	Naa	$X_{1'}$	$X_{2'}$	$X_{3'}$	$X_{4'}$	Chi-square test	p-value
***D2*-consensus**	Leu	Phe	Gly	Glu	Ser	Gln/Ala	Gly	Gly	Ala	**113.93**	**1.41 X 10^{-21}**
Frequency (%)	27.3	27.3	90.9	27.3	95.5	22.7	72.7	63.6	22.7		
***D3*-consensus**	Leu	Val/Phe	Gly	Asp	Ser	Gln/Ala	Gly	Gly	Ala	/	/
Frequency (%)	31.8	18.2	100	27.3	95.5	27.3	72.7	50.0	27.3		
	++	+	++	+		++	+++	++	++		

	X_4	X_3	X_2	X_1	Aaa	$X_{1'}$	$X_{2'}$	$X_{3'}$	$X_{4'}$	Chi-square test	p-value
***D2*-consensus**	Thr/Gly	Gly	Leu/Gly	Asp	Asp	Gly	Ile	Val	Pro	**51.09**	**8.84 X 10^{-9}**
Frequency (%)	13.6	27.3	13.6	27.3	81.8	36.4	18.2	22.7	27.3		
***D3*-consensus**	Val	Gly	Pro/Gly	Asp	Asp	Gly	Val/Gly	Trp/His	Pro	/	/
Frequency (%)	27.3	22.7	18.2	31.8	72.7	36.4	18.2	18.2	27.3		
	+	++	+	++		+++	+		+++		

	X_4	X_3	X_2	X_1	Baa	$X_{1'}$	$X_{2'}$	$X_{3'}$	$X_{4'}$	Chi-square test	p-value
***D2*-consensus**	Ile	Gly	Ala	Pro/Gly	His	Leu/Gly	Asn	Xaa*	Leu	**37.29**	**4.12 X 10^{-6}**
Frequency (%)	13.6	36.4	22.7	13.6	100	18.2	18.2	13.6	27.3		
***D3*-consensus**	Val	Gly	Ala	Asp	His	Gly	Asp	Glu	Ile	/	/
Frequency (%)	27.3	40.9	31.8	22.7	90.9	27.3	18.2	22.7	18.2		
	+	++	++	+		+	+	+	+		

[a]The χ^2 test results and *p*-values are provided for each microenvironment-based consensus sequence. Additional observations: + shared chemical property; + + identical in aa, different frequency; + ++ identical in aa and frequency; * Xaa corresponds to Glu, Asp, Leu and Pro, sharing the highest frequency.

Experimental evaluation of the protein-glutamate methylesterase (CheD) active site-inspired peptide containing Cys as nucleophile. Preferential intramolecular cyclization over dimerization is observed, controlled by the environment (pH, salts), but it remains unaffected by concentration. Heterodimerization can be induced through sequence manipulation (Cys→ Pen). The catalytic activity of CG11 is correlated to Cys oxidation state is correlated. MD simulations show that CG11 can temporarily access favorable distances for disulfide bridge formation.

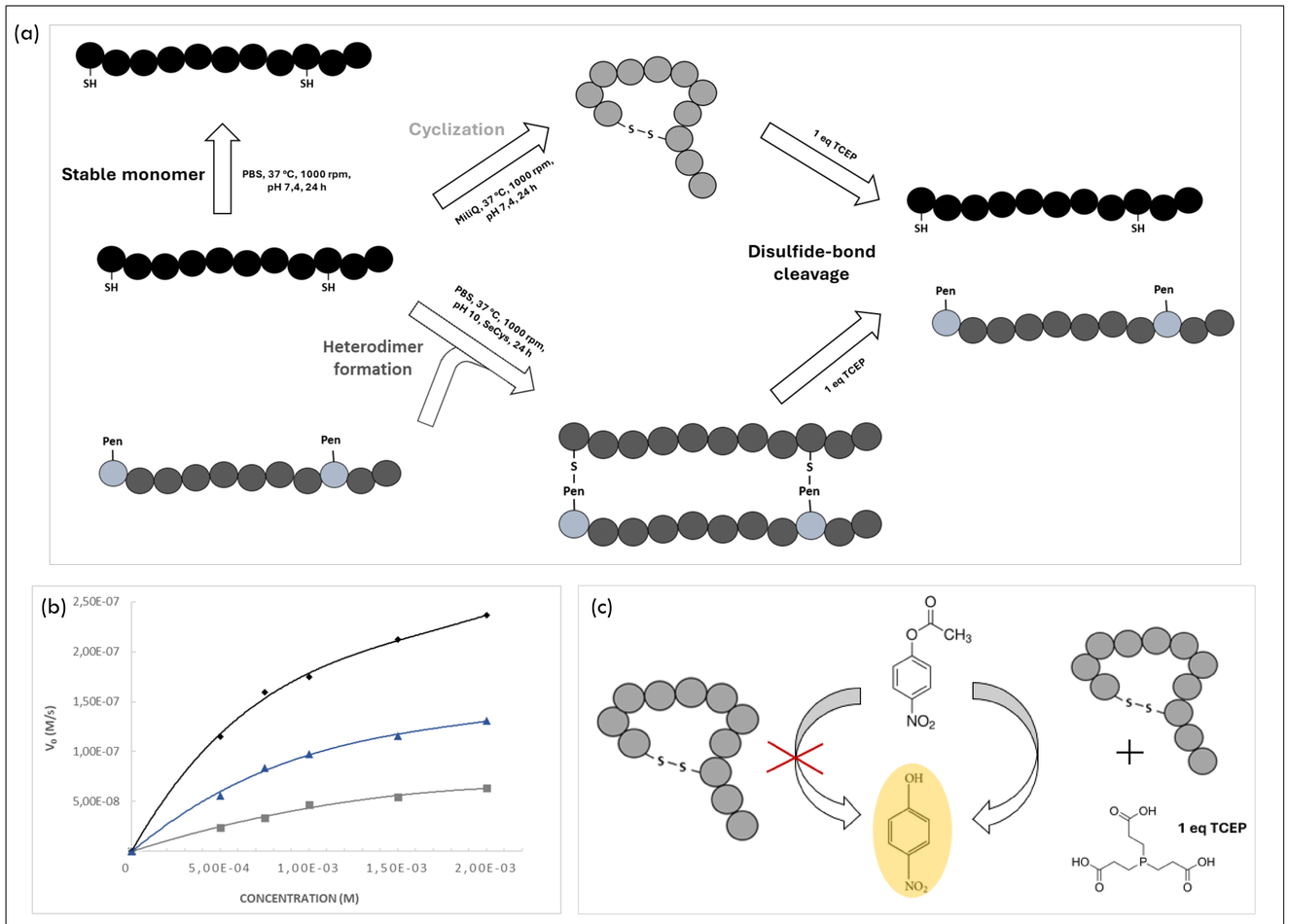
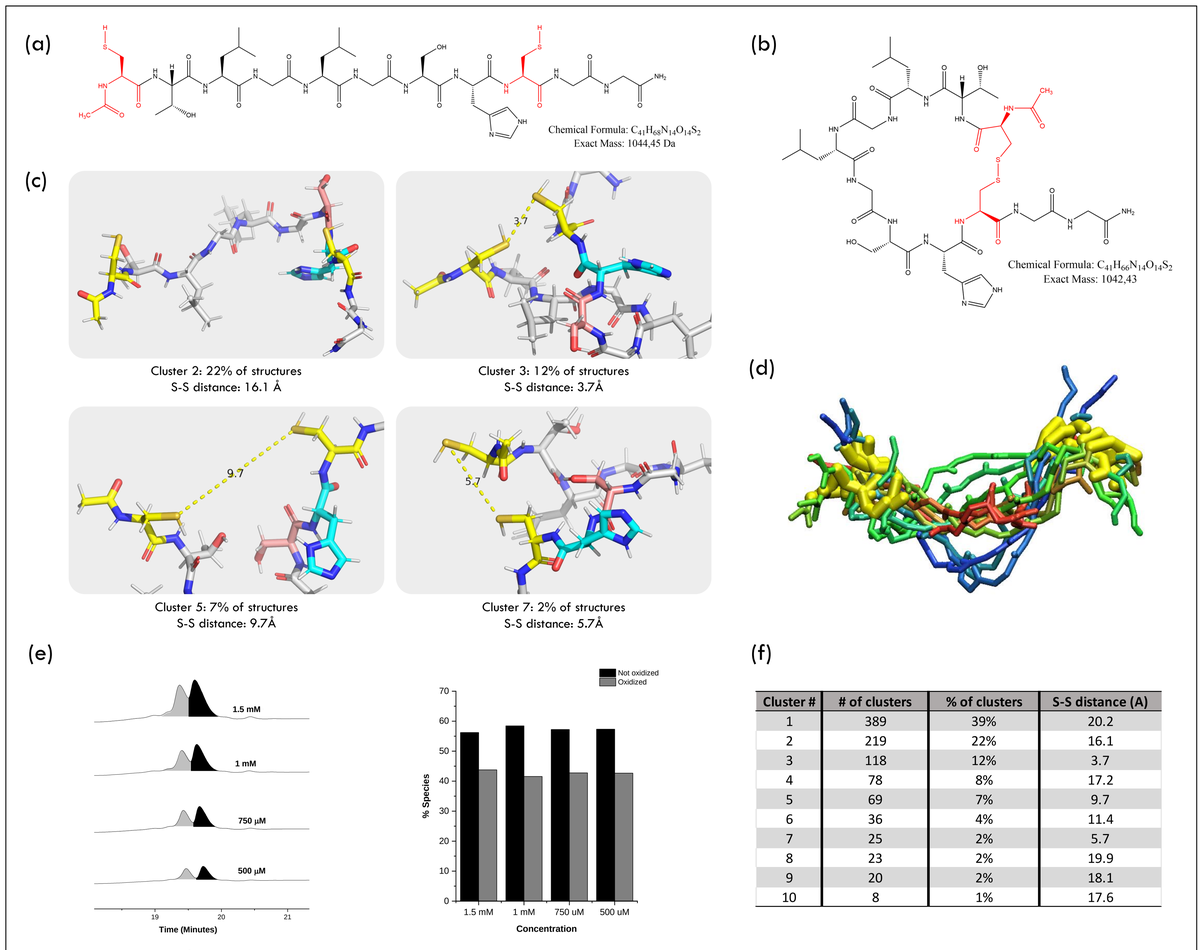


Fig 4. **Control of cyclization/dimerization and catalytic activity through the environment and sequence composition.** (a) The linear CG11 spontaneously forms intramolecular disulfide bridges giving rise to its cyclic form in MiliQ water (24 h), however if dissolved in PBS at the same conditions it remains stable (linear). To induce intermolecular disulfide bridges and controlled dimer formation CG11 was modified to include Penicillamines (Pen) instead of Cys, consequently it forms heterodimers in the presence of the parent sequence and SeCys. Disulfide-bond cleavage can be achieved with the addition of TCEP. (b) Michaelis-Menten plots showing reaction rates of p-NPA hydrolysis for CG11 (black), oxidized CG11 (gray) and oxidized CG11 after the addition of TCEP (blue). (c) Oxidation of cysteines diminishes catalytic activity that is recuperated with the disulfide-bond cleavage.



Cluster #	# of clusters	% of clusters	S-S distance (A)
1	389	39%	20.2
2	219	22%	16.1
3	118	12%	3.7
4	78	8%	17.2
5	69	7%	9.7
6	36	4%	11.4
7	25	2%	5.7
8	23	2%	19.9
9	20	2%	18.1
10	8	1%	17.6

Fig 5. **Cyclization/dimerization balance is not affected by concentration.** Chemical structures of CG11 in its linear (a) and cyclic (b) forms. (c) **Cluster analysis of a 100 ns MD simulation of CG11** made in GROMACS (1001x1001 matrix, backbone atom clusters, cutoff of 3Å, CHARMM36m force field) (d) Overlap (with trajectory smoothing) of the 15 discovered clusters, Cys backbone atoms are marked yellow. (e) HPLC traces of CG11 in miliQ water (24 h, 37 °C, 1000 rpm): oxidized form (*1042,4 Da*) in grey and non-oxidized form (*1044,45 Da*) in black. (f) Percentage of species present in the mixture calculated from the AUC.

References
[1] M Babić, P Janković, S Marchesan, G Mauša, D Kalafatovic, J. Chem. Inf. Model. 2022, doi.org/10.1021/acs.jcim.2c00977