

## Research article

# Paradigms of convergent evolution in enzymes

Ioannis G. Riziotis<sup>1\*†</sup>, Jenny C. Kafas<sup>1†</sup>, Gabriel Ong<sup>1</sup>, Neera Borkakoti<sup>1</sup>, Antonio J.M. Ribeiro<sup>1</sup>, Janet M. Thornton<sup>1</sup>

<sup>1</sup>European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, CB10 1SD, Cambridge, UK

*\*To whom correspondence should be addressed: [riziotis@ebi.ac.uk](mailto:riziotis@ebi.ac.uk), tel.: 0044 77 490 32067*

*†Equal contribution*

## Abstract

There are many occurrences of enzymes catalysing the same reaction but having significantly different structures. Leveraging the comprehensive information on enzymes stored in the Mechanism and Catalytic Site Atlas (M-CSA), we present a collection of 38 cases for which there is sufficient evidence of functional convergence without an evolutionary link. For each case, we compare enzymes which have identical Enzyme Commission numbers (i.e. catalyse the same reaction), but different identifiers in the CATH data resource (i.e. different folds). We focus on similarities between their sequence, structure, active site geometry, cofactors and catalytic mechanism. These features are then assessed to evaluate whether all the evidence on these structurally diverse proteins supports their independent evolution to catalyse the same chemical reaction. Our approach combines literature information with knowledge-based computational resources from, amongst others, M-CSA, PDBe and PDBsum, supported by tailor made software to explore active site structure and assess mechanism similarity. We find that there are multiple varieties of convergent functional evolution observed to date and it is necessary to investigate sequence, structure, active site geometry and enzyme mechanisms to describe such convergence accurately.

## Keywords

catalysis, enzyme, evolution, convergence, divergence, domain, active site, mechanism

## Introduction

Nature shapes biological macromolecules during evolution, allowing mutable elements to change, or conserving them by natural selection. The extent of selective pressure is variable, and this contributes to functional divergence in proteins of common ancestry. We have recently reviewed some of the concepts in enzyme evolution, especially functional divergence from a mechanistic viewpoint[1]. All known enzyme reactions are performed by a relatively limited number of structural folds. However, the chemical reaction space is vast and there may be many biological reactions we have yet to discover. Despite this, amongst the many reactions we know, many are catalysed by more than one family of enzymes, that are not related by evolution. Surprisingly, this is quite common, with a reaction on average being catalysed by ~2 evolutionarily unrelated enzymes[2,3]. Such analogues are called isofunctional enzymes or isozymes[4] and the evolutionary phenomenon that describes them is convergent evolution[5].

It is not uncommon for a catalytic process, or part of it, to be facilitated by a set of chemical groups in a well-defined geometry that occur in multiple enzymes[6]. These 3D constellations often drive a unique sequence of events of the catalytic mechanism. Popular examples are the various catalytic triads in proteinases[7,8] that hydrolyse peptide bonds by nucleophilic substitution. Although the nucleophile position might be occupied by different residues, usually Ser or Cys, the spatial arrangement of the triad (Nucleophile-His-Acid) is highly similar among most proteinases and has evolved independently in analogues[9]. In all cases, the same sequence of catalytic steps takes place. Similarly, the binding affinity for a substrate depends on the geometry of residues in the binding pocket, where any conformational changes can dramatically affect ligand selectivity. In the case of small ligands (especially metal ions), binding constraints can be strong, leading to a common geometry in unrelated proteins (e.g. Fe-SO metal cluster binding site).

In the context of convergent evolution, we can ask questions such as:

- 1) Since functional convergence can occur in multiple ways, what are the different facets of such convergence?
- 2) Can we distinguish different paradigms, and which occur most often in nature?
- 3) Why has nature evolved the same function more than once?

Leveraging the integrated information within the Mechanism and Catalytic Site Atlas (M-CSA) –i.e. sequence, structure, homologues, functional annotation of residues, curated annotation of

catalytic residues and explicit description of the mechanism – herein we aim to address these questions and define broad paradigms of enzyme convergent evolution.

We present 38 detailed examples of isozymes, characterising their catalytic sites and mechanisms to illustrate convergent evolution of function. We also provide evidence that categorisation into fine-grained paradigms is not straightforward, and in several cases, enzymes adhere to multiple paradigms. Our observations highlight some limitations in enzyme function and structure classification systems, and the need to consider additional parameters such as local active site geometry, cofactor selectivity and chemistry to classify ambiguous evolutionary relationships. This systematic analysis of multiple enzyme features (sequence, local and global structure, catalysed reaction, catalytic mechanism, substrate selectivity, promiscuity etc.), provides extra knowledge, useful for enzyme design by computational methods and directed evolution[10].

## Results

The Mechanism and Catalytic Site Atlas (M-CSA) aims to include entries for all known enzyme reactions for which sequence, structure and functional data are available, and so inevitably the content reflects the protein coverage in UniProt and PDB, which are both biased towards well studied enzymes. The rationale of M-CSA content and curation is broadly based on including one reference entry for every unique EC number, supplemented by enzymes with the same EC number but with a different mechanism. We took advantage of this redundancy to identify and compare pairs of enzymes, which appear to perform the same reaction but belong to different structural families (i.e. examples of convergent evolution). To this end, we extract two datasets of paired enzymes: 1) all pairs of entries in M-CSA (2021 update) with the same EC number at reaction level (sub-sub class - EC x.x.x.-), and 2) all pairs of entries with the same EC number performing the same reaction on the same substrate (sub-sub-subclass - EC x.x.x.x). In a pair, the two enzymes might be homologues (their catalytic domains belonging to the same CATH superfamily) or analogues (when the proteins are from different CATH superfamilies, thus are likely to have evolved independently). The whole dataset consisted of 6,345 pairs, many of which occurring more than once, so we further filtered for unique EC pairs, reducing this number to 2,209. Using protein CATH numbers as filters we distinguished homologous pairs (identical CATH number) from analogous pairs (different CATH numbers). These results are presented in Figure 1, which compares the two datasets.

We find many more analogous pairs than homologues (only 9-15% are homologues, covering 41 EC numbers in total) reflecting the content of the M-CSA. Most of the homologous functional duplicates differ in their mechanism, and a few are attributed to curation mistakes. We also find that considering only the reaction at the third level (i.e. ignoring substrate specificity by removing the 4<sup>th</sup> EC level filter) leads to a steep increase of the number of enzyme pairs. This increase remains significant even when redundancy is removed (All vs Unique pairs in Figure 1). However, we already know that in nature, most new functions evolve from old functions[11] by simply changing the substrate – but here we are deliberately targeting pairs of enzymes with the same function but in different structural families – i.e. analogues) Out of these sets, we selected the unique analogous pairs matching at the 4<sup>th</sup> EC level (83 pairs in total, with both members of the pair catalysing the same reaction with identical substrates). After filtering (see Methods) we ended up with 38 pairs, to analyse and exemplify three paradigms of convergence (Figure 2).

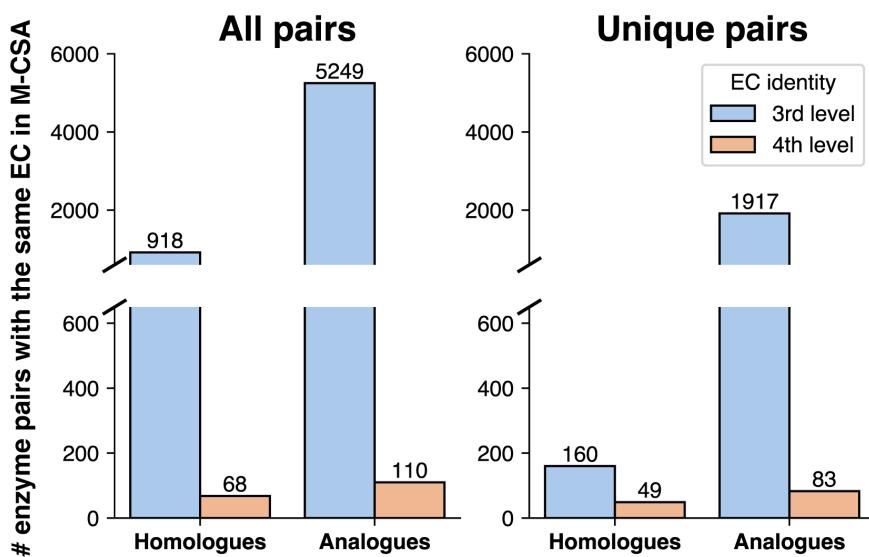


Figure 1: Number of enzyme pairs in M-CSA sharing the same EC number, when looking at the sub-subclass (3<sup>rd</sup> number) or sub-sub-subclass (4<sup>th</sup> number) level. Homologues have identical catalytic domain CATH numbers and analogues have different CATH numbers.

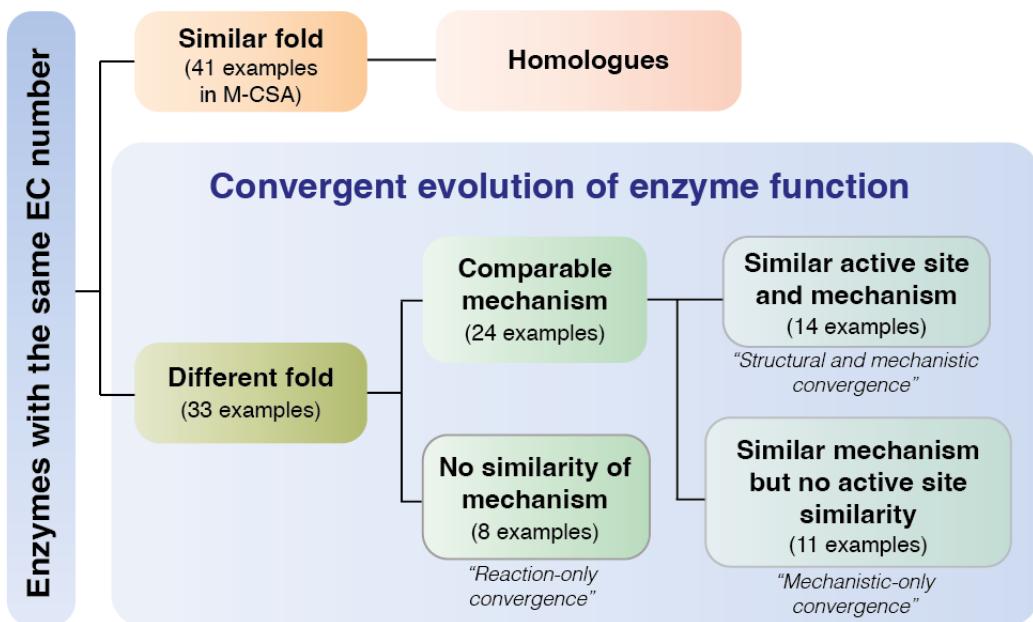


Figure 2: Evolutionary relationships of enzymes of the same function. Paradigms of convergence are quoted.

These 38 pairs were examined for similarities in different facets, some of which we discussed in the context of bioinformatics in our recent review[1]. Using computational tools and/or literature reviewing, we looked at the following:

- 1) overall structural similarity (CATH superfamily assignment and 3D fitting)
- 2) active site structural similarity (quantified by RMSD of aligned residues)
- 3) mechanistic features (catalytic steps[12,13] or/and residue roles)
- 4) substrate similarity
- 5) cofactor presence/similarity

Our results are organised as pairs or triplets of analogues in Table 1-3. Three selection rules are implicit in this survey of convergent evolution: 1) catalytic domains should always belong to different CATH superfamilies, 2) All 4 E.C. numbers should be identical and 3) isozymes should share at least one identical product and substrate. This means that different paradigms of evolutionary convergence are defined by similarities in the active site composition and geometry, catalytic mechanism and cofactor selectivity. Focussing solely on mechanism, three general ‘cases’ are observed. Firstly, analogues that are completely different in their mechanism of catalysis. In the second ‘case’, analogues that are similar but only part of the reaction is the same (e.g. phospholipases (EC 3.1.1.4) had similarities in the steps 1 & 2 of one enzyme and steps 3 & 4 of the other respectively). In the 3<sup>rd</sup> case, mechanisms between the analogues are essentially the same, with only minor differences. For these, the mechanisms are

nearly identical except for differences in the catalytic residues being used (e.g. both analogous Papain-like cysteine proteases (EC 3.4.22.28) use a Cys-containing catalytic triad, but the His activator can be Asp or Asn). Combining these mechanistic paradigms with any localised structural similarities in the active site[14], led us to define three general paradigms of convergence in isozymes, described in the following section.

Table 1: Isoenzyme examples with structural and mechanistic convergence

Enzyme	EC	CATH domain	M-CSA ID	PDB ID	Residue composition	3D-aligned residues	RMSD (Å)	Mech. sim. (calc. perc.)	Cofactor	Notes	Refs.
Carbonate dehydratase	4.2.1.1 (a)	2.160.10.10	516	1QRM	EEHHHNQ	His117 Glu62 His122 His81	1.45	Yes (30%)	Co <sup>3+</sup>	Similar metal binding site. Sulfate moieties bind close to HCO <sub>3</sub> positions in both analogues.	[15,16]
		3.10.200.10	216	5GMN	EEHHHT	His94 Glu106 His119 His96			Zn <sup>2+</sup>		
Protein tyrosine phosphatase	3.1.3.48	3.40.50.2300	462	2P4U	CCDNRS	Ser20 Cys13 Arg19	1.30	Yes (n/a)	-	Both analogues are mammalian. Significant catalytic core 3D similarity in completely different fold contexts. Divergence followed by convergence is plausible.	[17,18]
		3.90.190.10	469	2GJT	CDQRT	Thr1143 Cys1136 Arg1142			-		
3-hydroxydecanoyl dehydratase	4.2.1.59	3.10.129.110	972	2VZ9	DGHHLY	Any885 Asp1033 Any888 His878	0.62	Yes (27%)	-	One analogue is mammalian, the other is bacterial. Significant catalytic core 3D similarity. Either a case of extreme convergence or distant homology.	[19]
		3.10.129.10	10	1MKA	CDGHV	Any76 Asp84 Any79 His70			-		
Endo-1,4- $\beta$ -xylanase	3.2.1.8	2.60.120.180	432	2B42	EENYY	Glu172 Glu78 Asn35	1.67	Yes (n/a)	-	Sugar is cleaved with very similar mechanisms using a Glu-Glu dyad.	[20,21]
		3.20.20.80	548	1FHD	DEEHN	Glu233 Glu127 Asn169			-		
Peptidylprolyl isomerase	5.2.1.8	2.40.100.10	189	1XO7	FFHLNQR	Any103 Phe62 Phe114	2.56	Yes (n/a)	-	Two Phe residues form a hydrophobic pocket for substrate side chain rotation. Main chain residue is H-bonded to substrate carbonyl. Both human-derived, but 362 induces reversible protein folding.	[22,23]
		3.10.50.40	362	1NSG	DFFIYY	Any56 Phe36 Phe99			-		
Lysozyme	3.2.1.17	1.10.530.10	203	3IJU	DDENNS	Glu35 Asp52 Asp48	2.65	Yes (27%)	-	Hydrolysis mechanism is similar, with slightly different spatial arrangement of residues.	[24,25]
		3.20.20.80	774	1H09	DDDE	Glu94 Asp92 Asp10			-		
Protein-methionine-S-oxide reductase	1.8.4.11	3.30.1060.10	122	4D7L	CCCDEYY	Asp130 Cys52 Cys206	3.14	Yes (23%)	-	Cys residues have slightly different roles in the two enzymes, but two catalytic steps are similar.	[26,27]
		2.170.150.20	715	1L1D	CCDHHR	Asp484 Cys440 Cys495			-		
Ubiquitinyl hydrolase 1	3.4.19.12	3.40.532.10	597	4I6N	CDHQ	Any243 His340	1.18	Yes (n/a)	-	Common Cys deprotonation by His with Cys also contributing its backbone in oxyanion hole	[28,29]
		2.40.10.10	830	2GZ9	CEGH	Any147 His40			-		
Papain-like cysteine protease	3.4.22.28	3.90.70.10	805	1Y4H	CHNQ	Any243 His340	1.26	Yes (n/a)	-	Very similar active sites, oxyanion hole is Cys backbone and Ala in one analogue and Cys backbone and Asn in the other. Asn and Glu modify the pKa of the His in both.	[30,31]
		2.40.10.10	477	6FFS	CEGH	Any147 His40			-		
Ornithine decarboxylase	4.1.1.17	3.40.640.10	860	1ORD	DHK	Lys355 His223	1.75	Yes (n/a)	PLP	Identical mechanism between bacterial/mammalian analogues. Their only difference is the mirrored arrangement of residues.	[32,33]
		3.20.20.10	937	5GJO	EHK	Lys51 His179			PLP		
Phosphohistidine-D-mannose phosphotransferase	2.7.1.191	1.20.58.80	514	1E2A	HHQ	His78 His82	0.44	Yes (n/a)	-	Two His are at almost identical positions, but reside on different secondary structure elements. Roles in mechanism are similar.	[34,35]
		2.70.70.10	513	3OUR	HHT	His91 His76			-		
Cytosine deaminase	3.5.4.1	3.20.20.140	710	1K6W	DEHHHQ	His214 Glu217	0.69	Yes (35%)	Zn <sup>2+</sup>	In both analogues, His ligates a metal. Glu deprotonates a water molecule and protonates cytosine N3. Difference lies in the combination of metal-ligating residues.	[36,37]
		3.40.140.10	636	1UAQ	CCEHS	His62 Glu64			Zn <sup>2+</sup>		
Endonuclease	3.1.21.1	3.90.540.10	791	2GZ1	EEHHHRR	His103 His127 Glu100	1.43	Yes (30%)	Zn <sup>2+</sup>	Similar initial steps of nucleophilic attack on phosphate via a proton relay. A Glu-His dyad is present in both analogues.	[38]
		3.60.10.10	41	2A40	DDEEHHY	His252 His134 Glu78			Mg <sup>2+</sup>		
3-dehydroquinate dehydratase	4.2.1.10	3.40.50.9100	55	4K17	EGHNNPRRY	Glu99 His101	0.83	No (0%)	-	Glu-His proton relay motif present in both analogues. Glu activates His, but the overall mechanistic context is different.	[39,40]
		3.20.20.70	54	1QFE	EHK	Glu86 His143			-		

Table 2: Isoenzyme examples with mechanistic-only convergence

Enzyme	EC	CATH domain	M-CSA ID	PDB ID	Residue composition	3D-aligned residues	RMSD (Å)	Mech. sim. (calc. perc.)	Cofactor	Notes	Refs.					
Aminopeptidase	3.4.11.19	3.60.70.12	676	1B65	GNNSY	Ser288 Asn218	1.63	Yes (n/a)	-	Spurious active site similarities, but peptide bond hydrolytic mechanisms share common features.	[41,42]					
		3.40.710.10	782	1HVB	HKNSY	Ser62 Asn161										
Chorismate mutase	5.4.99.5	3.30.1330.40	474	1DBF	CERRRY	Arg116 Arg63 Glu78	4.04	Yes (100%)	-	Active site electrostatic environments and proposed mechanisms are similar, but residues do not superimpose well.	[43,44]					
		1.10.590.10	81	2CSM	EEKNRRRT	Arg16 Arg157 Glu198										
Ser-Thr-Tyr phosphatase	3.1.3.16	3.60.21.10	406	3EGG	DDDHHHHNRR	Asp95 Arg96	1.41	No (0%)	-	Spurious active site similarities but the role of Arg is the same in the two analogues.	[45,46]					
		3.90.190.10	456	5BZX	CDR	Asp92 Arg130										
Cellulase	3.2.1.4	1.50.10.10	559	1KS8	DDEY				-	General mechanism of sugar cleavage is the same in all three analogues, involving a water molecule activated by an Asp or Glu residue. Although all three have an acid-acid dyad, active sites do not superpose well probably due to the presence of different bound ligands. Catalytic step order is slightly different between m560 and m561, thus they appear mechanistically different.	[47-49]					
		2.40.40.10	561	3ENG	DD											
		2.60.120.180	560	3VL8	EE											
Protein-glutamate methyl esterase	3.1.1.61	3.40.50.180	337	1A2O	DHMST				-	Methyl-ester hydrolysis mechanism involves a catalytic triad in both cases, but the nucleophile and stabiliser residues are different.	[50,51]					
		3.30.1330.200	729	2F9Z	CHT											
Licheninase	3.2.1.73	3.20.20.80	400	1AQ0	EEEK				-	Both analogues use an acid dyad to cleave the sugar substrate, but active sites do not superpose well.	[52,53]					
		2.60.120.200	924	1CPM	EE											
Penicillin amidase	3.5.1.11	3.60.60.10	241	3HBC	CFNR				-	Similar hydrolysis mechanisms, involving a self-activated nucleophile and an oxyanion hole in both analogues. Differences lie on the nucleophile (Ser vs. Cys) and the overall catalytic residue arrangement.	[54,55]					
		3.60.20.10	841	1GK9	ANS											
Alcohol dehydrogenase	1.1.1.1	3.40.50.720	255	3W8F	KNSY				-	Highly similar mechanisms, but without detectable structural similarities.	[56,57]					
		3.90.180.10	256	5VKR	CCHHS											
Histone acetyltransferase	2.3.1.48	3.90.226.10	344	3HE2	AAD				-	Acetyl-CoA	Highly similar mechanisms with several overlapping steps. However, no geometrical similarities are detected	[58]				
		3.40.630.30	224	4PSX	EF											
Carbonate dehydratase	4.2.1.1 (b)	3.10.200.10	216	5GMN	EHHHHT				-	Yes (33%)	Zn <sup>2+</sup> Zn <sup>2+</sup>	High mechanistic similarities, no geometrical similarities in active sites.	[16,59]			
		3.40.1050.10	517	5CXK	CCDHR											
Carbonate dehydratase	4.2.1.1 (c)	2.160.10.10	516	1QRM	EEHHHNQ				-	Yes (44%)	Co <sup>3+</sup> Zn <sup>2+</sup>	High mechanistic similarities, no geometrical similarities in active sites.	[15,59]			
		3.40.1050.10	517	5CXK	CCDHR											

Table 3: Isoenzyme examples with reaction-only convergence

Enzyme	EC	CATH domain	M-CSA ID	PDB ID	Residue composition	3D-aligned residues	RMSD (Å)	Mech. sim. (calc. perc.)	Cofactor	Notes	Refs.
Acid phosphatase	3.1.3.2	3.60.21.10	43	1KBP	DDHHHHHHNY	His296 Asp164 His325 His59 Asp335 His334 Arg168 His334 His59 Arg183 His189 His150	1.40	No (0%)	Zn <sup>2+</sup> , Fe <sup>3+</sup>	Every pairing of analogues has active site similarities and in every case it involves ligand/cofactor binding residues (different residues match in every pair). Purple acid analogue uses metal cofactors, and the other two were crystallised with metal ions to substitute the phosphate substrate. Mechanisms share some similarities within individual pairs but are generally different.	[60–62]
		3.40.50.1240	454	4JOC	DHHRRR		1.98	Yes (n/a)	-		
		1.20.144.10	558	1EOI	DHHR	His189 Asp193 His150	1.48	No (0%)	Zn <sup>2+</sup> , Fe <sup>3+</sup>		
		3.60.21.10	43	1KBP	DDHHHHHHNY	His325 Asp135 His295					
Phospholipase A2	3.1.1.4	3.40.1090.10	529	5IZR	DGGRS	Any330 Asp647	2.56	Yes (23%)	-	Spurious active site similarities. Mechanisms are different.	[63,64]
		1.20.90.10	528	1FAZ	DDDHDL	Any44 Asp85			Ca <sup>2+</sup>		
Catalase	1.11.1.6	1.20.1260.10	572	1JKU	EEEEHH	His69 His181	4.31	No (n/a)	Mn <sup>3+</sup>	Spurious active site similarities. Mechanisms are different.	[65,66]
		2.40.180.10	573	3VU3	HHN	His392 His128			Heme		
Chloride peroxidase	1.11.1.10	1.10.606.10	14	IIDQ	GHHKRRS	Ser402 His404 Any403 Ser98 His257 Any32	2.98	No (0%)	VO <sub>4</sub>	Any calculated active site similarities are spurious among the three analogues. Mechanisms are completely different, and the only commonality is Cl <sup>-</sup> selectivity, which is one of three substrates (Cl <sup>-</sup> , H <sub>2</sub> O <sub>2</sub> and a variable organic compound that is chlorinated).	[67–69]
		3.40.50.1820	248	1BRT	DFHST	Asp228 His257	2.86	No (0%)	-		
		1.10.489.10	250	2CJ0	CDEH	Asp106 His105			Heme		
Cellulose 1,4- $\beta$ -cellobiosidase	3.2.1.91	3.20.20.40	440	1QK0	DDDRSY	-	-	No (n/a)	-	Both analogues cleave cellulose to cellobiose, with different mechanisms (invertin vs. retaining).	[70,71]
		2.70.100.10	444	6GRN	DEEH				-		
B-lactamase	3.5.2.6	3.60.15.10	258	6U13	DHHHHHY	-	--	No (0%)	Zn <sup>2+</sup>	Catalysis is metal-dependent in one analogue, in contrast to the other.	[72,73]
		3.40.710.10	210	1M6K	ASSSWX				-		
Dihydrofolate reductase	1.5.1.3	2.30.30.60	752	1VIF	1KQY	-	-	Yes (n/a)	NADPH	No similarities in mechanism detected, however both analogues use NADPH for hydride transfer to dihydrofolate.	[74,75]
		3.40.430.10	490	1DHF	EL				NADPH		
		2.160.20.10	896	1O8H	DDDER	-			Ca <sup>2+</sup>		
Pectate lyase	4.2.2.2	1.50.10.20	509	1GXN	NR	-	-	No (n/a)	-	The only mechanistic similarity lies in the presence of an Arg proton shuttle, however the overall mechanism differs. The bacterial analogue also requires a calcium activator of Arg.	[76,77]
ADP-ribosyltransferase	2.4.2.36	3.90.175.10	773	1TOX	E	-	-	No (n/a)	-	Toxin mechanisms are not well studied. The analogues both use NAD and diphthamide as substrates but are likely unrelated.	[78,79]
		3.90.210.10	919	1XTC	EERS				-		

### ***Paradigms of convergent evolution***

38 cases of isozymes in M-CSA reveal three paradigms of functional convergence:

1. **Structural and mechanistic (Table 1).** Two or more catalytic residues align in 3D, and there are similarities in the mechanism. This is a very common paradigm, with 14 examples. In some cases, a subset of catalytic residues may align in 3D, but the mechanisms are nevertheless different. This sub-paradigm is not common here (2 examples), and active site similarities, if found, are usually located on a ubiquitous motif (e.g. an activator stabilised by an acid or a common motif binding a catalytic metal).
2. **Mechanistic-only (Table 2).** Similarities in mechanism are observed, with no aligned catalytic residues. This is also a very common paradigm (11 examples). In such cases, a general similarity in the active site might exist, but this is not detectable in superposition.
3. **Reaction-only (Table 3).** Only the overall reaction is similar, with no similarities in mechanism or active site geometry. A moderately common paradigm for which 8 examples are found.

Paradigms can be divided further according to cofactor selectivity: a) Similar cofactors are used and b) different cofactors are used, or one analogue uses a cofactor while the other does not.

There are also cases in this dataset that could be assigned to more than one paradigm or sub-paradigm. For instance, the two unrelated 3-dehydroquinate dehydratase analogues (EC 4.2.1.10) use a His-Glu dyad, where Glu as an activator of His, but these occur in different mechanistic contexts. Chorismate mutases (EC 5.4.99.5) are also ambiguous and can be assigned to either paradigm 1 or 2, depending on how structural and functional similarities are interpreted.

### ***Active site 3D similarities accompany mechanistic similarities (Table 1)***

Several examples in our data correspond to mechanistic analogues, with similarities in mechanism being reflected in three dimensions (structural and mechanistic convergence).

To illustrate this paradigm, we selected the human and yeast carbonate dehydratase analogues. These classes of carbonate dehydratases (alpha and gamma) have converged to catalyse the

reversible conversion of carbon dioxide and water into hydrogencarbonate (Figure 3). They use a similar mechanism where the most important catalytic step is a nucleophilic attack on the carbon dioxide by a hydroxide ion, which is being stabilized by a positively charged metal ion (mechanism step picture in Figure 3). Interestingly, although the enzymes use a different metal ion,  $\text{Co}^{3+}$  in the case of the gamma class and  $\text{Zn}^{2+}$  for the alpha class, both active sites provide three histidine residues to hold the metal ion in place. While the coordination sphere of  $\text{Zn}^{2+}$  is satisfied by the three histidines and the hydroxide,  $\text{Co}^{3+}$  binds an additional water molecule and as the reaction proceeds, the hydrogencarbonate product. The active sites also differ in the residue that deprotonates the water molecule to generate the hydroxide (a His for alpha vs. a Glu for gamma) and the residues that orient the carbon dioxide (Glu and Thr for alpha class vs. Asn for gamma class).

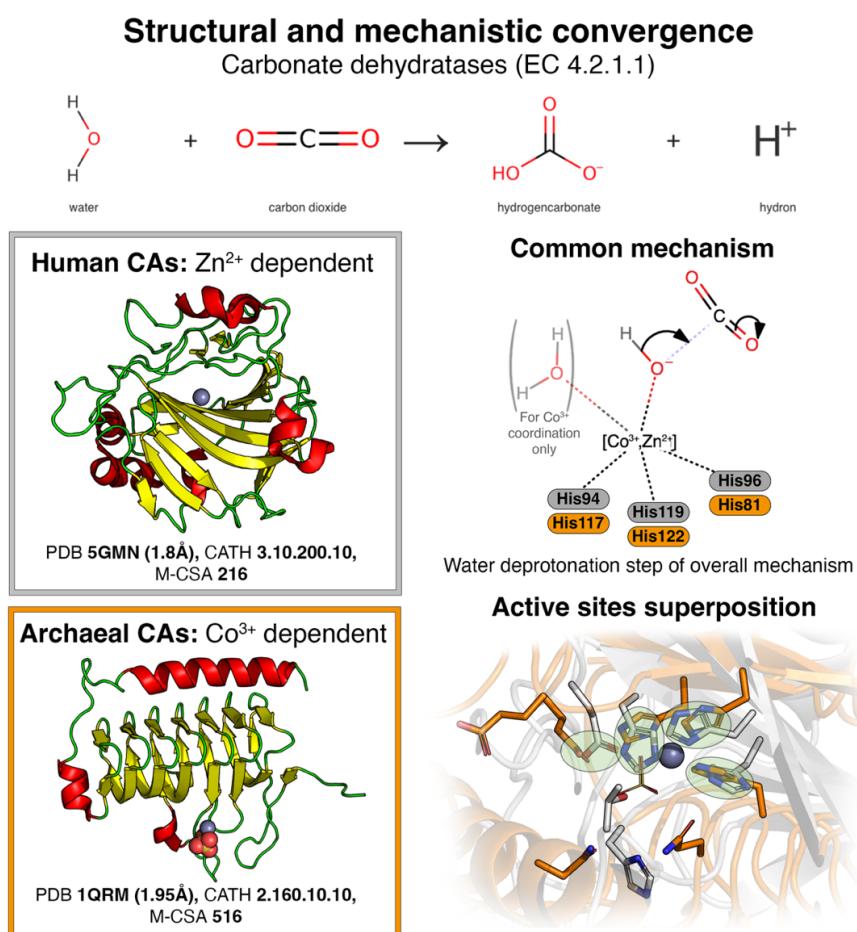


Figure 3: Structural and mechanistic convergence between human (grey) and archaeal (orange) carbonate dehydratases. Green circles in the active sites superposition indicate pairs of 3D-aligned catalytic residues.

Another illustrative example is the pair of mammalian/bacterial ornithine decarboxylases (EC 4.1.1.7) that both use an (Asp/Glu)-His-Lys residue triad to cleave a carboxyl group from the

ornithinium substrate (Figure 4a). A PLP cofactor is involved in both mechanisms, that is initially covalently bonded to the catalytic Lys. Global sequence alignment showed low similarity although the local alignment has relatively high coverage. The enzymes have no overall structural similarity, belonging only to the same primary CATH class (1<sup>st</sup> level). In their active sites, the two analogues have mirrored but similar arrangement of catalytic residues, with Lys and His endpoint atoms superposing at RMSD 1.75Å. By these observations, homology is unlikely; instead, convergence in function, local structure and mechanism is inferred.

Structural constraints may be relaxed in some cases, with slightly different functional atom 3D configurations facilitating the same mechanistic sequence. Mammalian and viral lysozyme analogues (EC 3.2.1.17) represent this paradigm variant (Figure 4b), between which the carbonyl groups of three acidic residues (one Glu and two Asp) loosely superpose at RMSD 2.65Å. Glycosyl hydrolases GH22 (viral) and GH25 (mammalian) break down  $\beta$ -MurNAc-(1->4)- $\beta$ -D-GlcNAc to N-acetyl- $\beta$ -D-muramic acid and N-acetyl- $\beta$ -D-glucosamine via hydrolysis, with two proposed mechanisms in each analogue. They have virtually no structural similarity, and GH25 uses a choline ion to facilitate cell wall binding, which is not directly involved in catalysis but does increase catalytic activity, while GH22 does not use any cofactors. Global and local sequence alignments showed low similarity, and except for Glu35 in GH22 lining up with a non-catalytic Glu in GH25 in the global alignment, no active site residues line up with an identical (or near-identical) residue in any alignment. The only differences between them are in which atoms serve as donors/acceptors, and in one case the active site regeneration chemistry is different. Every proposed mechanism directly involves a Glu with an intramolecular O atom (the latter is present in three out of total four proposals) to break the glycosidic bond. Both have the presence of two acidic residues hugging the substrate, similar to cellulases. The analogues show 27% mechanistic similarity, and given the low sequence and structural similarity, it is likely that they evolved this function separately.

Evolution can also shape active sites in a modular way, with resulting enzymes having similarities in the mechanism, but differing at a technicality. A good example is cytosine deaminases (EC 3.5.4.1) where both analogues use a catalytic metal, coordinated by different residues sets (Figure 4c): 3 His and an Asp binding a Fe ion in the yeast enzyme, while 1 His and 2 Cys bind a Zn ion in its bacterial analogue. In both cases, the metal binding His and the

proton donor/acceptor Glu have similar 3D arrangement, with the latter protonating the cytosine N3 by obtaining a proton from a water molecule.

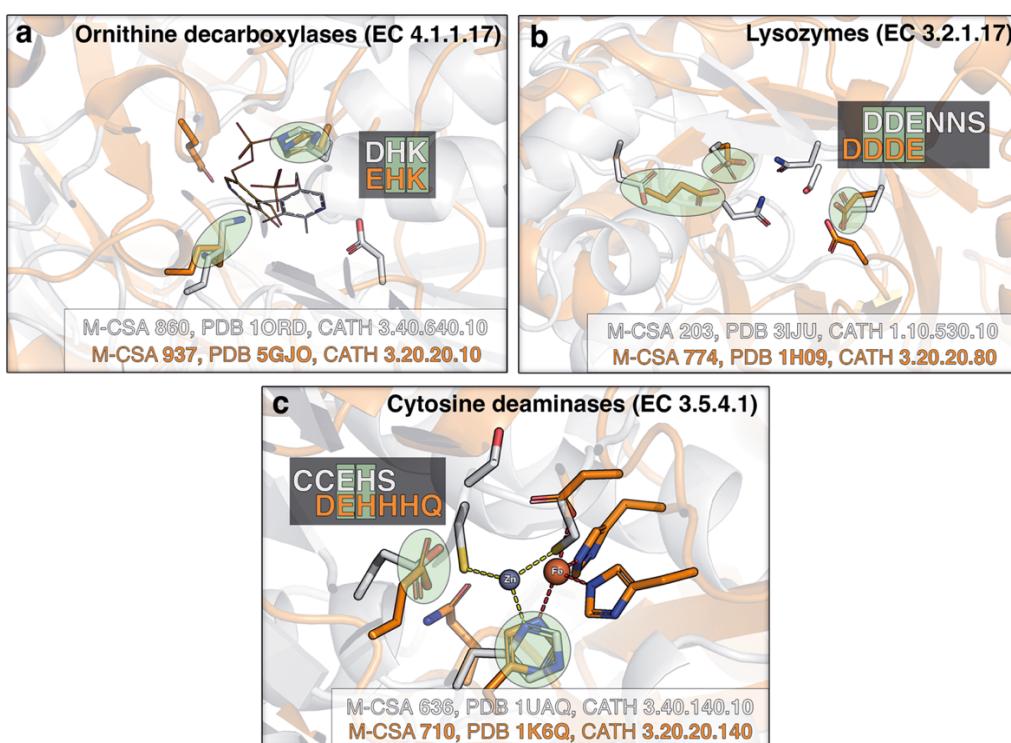


Figure 4: Active site similarities in three pairs of analogues with structural and mechanistic convergence. a) Bacterial (grey) and mammalian (orange) ornithine decarboxylases. b) Chicken (grey) and viral (orange) lysozymes. c) Yeast (grey) and bacterial (orange) cytosine deaminases. Green circles indicate 3D-aligned catalytic residue pairs. Residue alignment is also shown on the active site residue composition (pseudo-sequence as in Table 1) with the aligned residues highlighted in green.

Within this paradigm, we also distinguish analogue pairs that can be characterised alternatively, by a special case we name as “structural-only convergence”. This can be defined as structural convergence in the active site, but with similarities only observable in small motifs such as catalytic metal binding sites or activator-stabiliser dyads (e.g. His-Glu or His-Asp). This is clearly seen in two examples that both contain a His-Acid dyad: 3-dehydroquinate dehydratases (EC 4.2.1.10) and endonucleases (EC 3.1.21.1). In all four enzymes, the acid (Asp/Glu) interacts with a His via an H-bond, to activate and/or sterically orientate it, however in different mechanistic contexts. The same dyad is also present in catalytic triads of Ser/Cys proteases with the acid playing the same role.

#### ***Mechanistic similarities with different residue toolkits (Table 2)***

We observed cases where mechanistic similarities might exist, without detectable active site similarities (mechanistic-only convergence). In such cases, any structurally aligned residues

found by our alignment algorithm are most likely spurious or superpose loosely. This is demonstrated in fungal and bacterial chorismate mutases (EC 5.4.99.5), between which 3 catalytic residues superpose at a very high RMSD (4.04Å – considered dissimilar), but the overall electrostatic environment that facilitates isomerization of the chorismate substrate is similar, after manual inspection.

A representative example for the “mechanistic-only” paradigm is licheninases (EC 3.2.1.73) produced by both plant and bacteria for cell wall degradation. These are unrelated in primary sequence (20% sequence identity) and tertiary structure (Figure 5). Despite the differences in protein fold and local geometry they retain a similar mechanism for cleaving cereal  $\beta$ -D glucans and lichenin with the formation of an intermediate covalent bond between the substrate and a glutamic acid in the enzyme during catalysis[80].

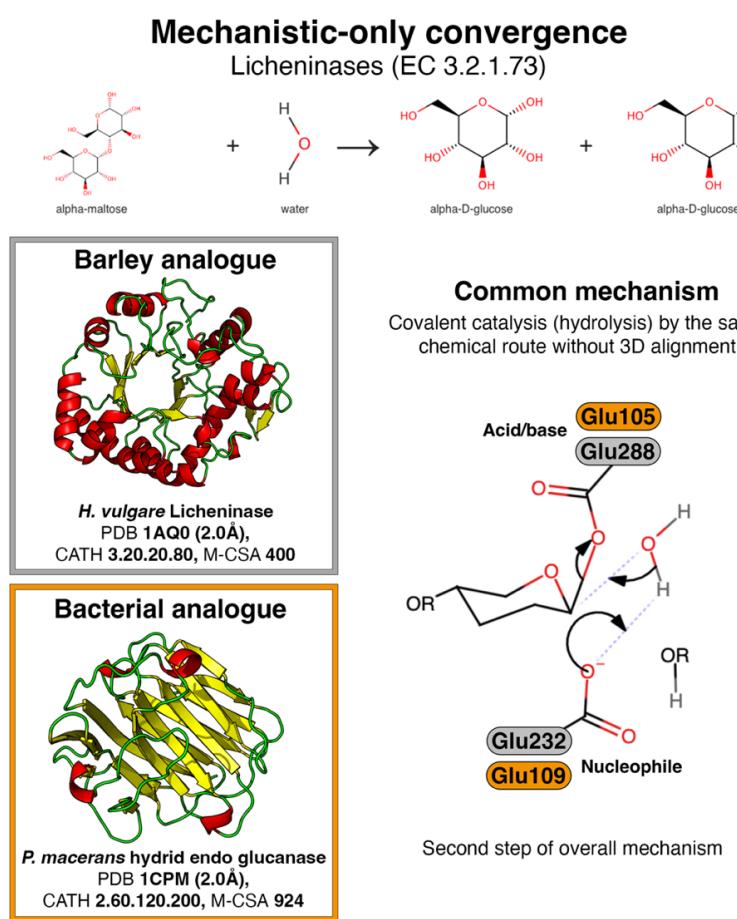


Figure 5: Mechanistic-only convergence between plant (grey) and bacterial (orange) licheninases.

Aminopeptidase isoenzymes from *Ochrobacterium anthropi* (EC 3.4.11.19) also conform to this paradigm. They are also a great example of a duplicate function within the same

organism/cell/compartment, presenting the evolutionary question of why two isoenzymes would co-exist. The two analogues are known as DmpA (M-CSA entry 676) and DmpB (M-CSA entry 782) with no evidence of a common ancestor between them. However, their mechanisms are similar without geometrical convergence in the active site. DmpA is poorly characterised, mostly because it is expressed in low quantities[81]. DmpB is more abundant and better characterised, and is also an ancestor of penicillin resistance conferring enzymes ( $\beta$ -lactamase activity)[42]. It is plausible that the two analogues co-exist in the same organism, because they are regulated differently. DmpA is also functionally promiscuous, with D-esterasic and D-amidasic activities as well as autoproteolysis[43], suggesting that aminopeptidase activity is not the primary one, compared to DmpB, whose function is more specific.

### ***Convergent evolution with different residue toolkits and mechanism (Table 3)***

Enzymes can converge in the overall catalysed reaction, with different stepwise chemistries, with no or limited similarities (reaction-only convergence – no mechanism or structure similarity). Reaction-only convergence is moderately abundant in M-CSA (9 examples) and refers to enzymes that have the same EC number (i.e. they catalyse the same reaction) but have no detectable similarities in their 3D structure. These cases (Table 3) are particularly interesting when enzyme analogues of different CATH families catalyse the same reaction in a given phylum.

In the case of fungi for example, the two cellobiosidases (EC 3.2.1.91) act on the cellobiose substrate using a different ensemble of residues and mechanisms (inverting vs. retaining[82]) for bond cleavage, giving isomeric products of opposite stereochemistry[83]. Similarly, structurally different  $\beta$ -lactamases (EC 3.5.2.6) in bacteria perform the same catalytic function using either a protein-water or a distinct protein-metal mediated process (Figure 6). Reaction-only convergence can also be observed across taxa, for example the peroxidases (EC 1.11.1.10) from fungus and bacteria (Table 3). In this case, the oxidation of the substrate is achieved by utilising significantly different mechanisms through the selective exploitation of non-proteinous components like Chlorine, Heme or Vanadium.

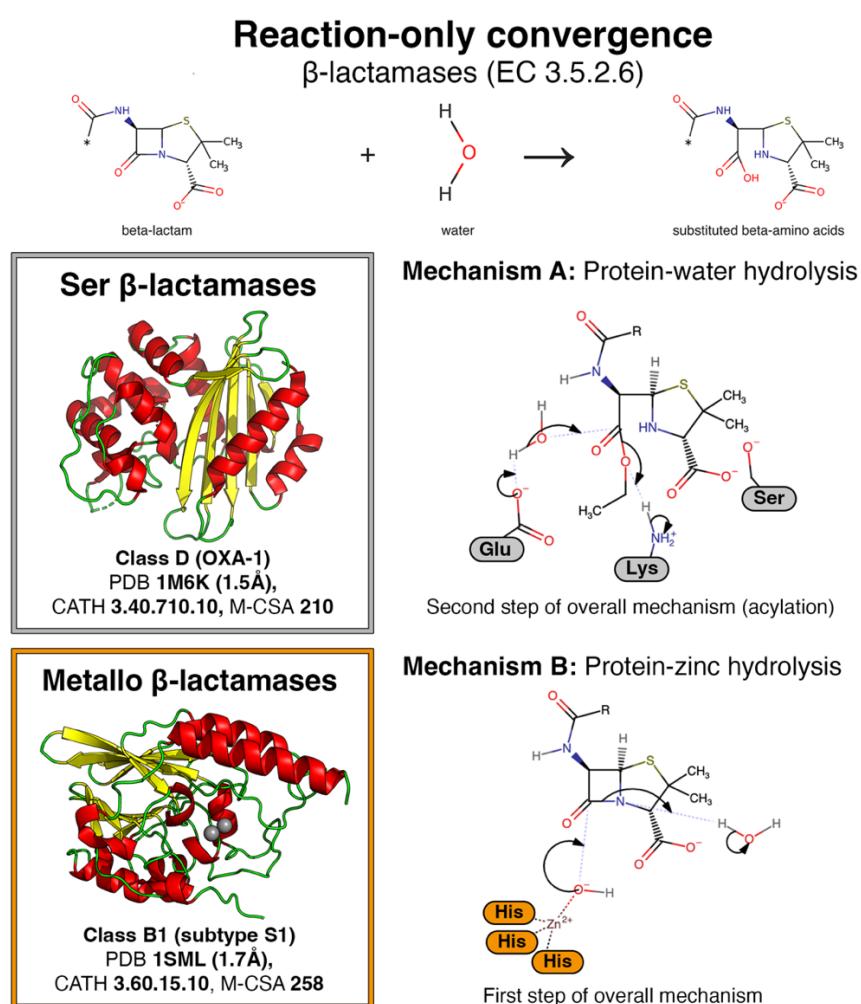


Figure 6: Reaction-only convergence in  $\beta$ -lactamases. The exact acid and bases, and the sequence of catalytic events vary in different Ser  $\beta$ -lactamases. Metallo  $\beta$ -lactamases bind two metal ions in the active site, but only the one participating in the described steps is shown in the mechanism diagram.

### Convergence and distant divergence are often indistinguishable

Convergence is not easily distinguished from distant divergence, especially in the case of generic metabolic enzymatic activities. Such isoenzymes have little sequence and overall structural similarity, belonging to different CATH superfamilies (H level) or even having different topologies (T level) and secondary structure architecture (A level). However, their active/binding sites and overall conformation of the catalytic core occasionally retain structural similarity.

This is exemplified in two mammalian tyrosine phosphatases (EC 3.1.3.48) that have a remarkably similar catalytic core but completely different overall folds (CATH 3.40.50.2300 for the bovine low molecular weight phosphatase, vs. CATH 3.90.190.10 for the human tyrosine phosphatase). For these two analogues, Gherardini et al. suggest convergence[3],

however, the striking similarities in the vicinity of the active site could infer two additional plausible scenarios: a) divergence to accommodate different metabolic needs and subcellular location (endoplasmic reticulum vs. cytoplasm), with conservation of the catalytic core, or b) divergence towards a different function, followed by re-convergence to the initial one. Another example of evolutionary ambiguity is the pair of two bacterial/mammalian 3-hydroxydecanoyl dehydratases (EC 4.2.1.59) that share high similarity in the extended active site, with four catalytic residues superposing at very low RMSD (0.62Å). This is accompanied by high mechanistic similarity. Our observations lead us to believe these analogues are either an extreme case of convergent evolution, or a case of horizontal gene transfer from symbiotic bacteria to the host, followed by widespread structural divergence, retaining the structure of the first residue shells.

Evolutionary ambiguities are expected to be common in activities that are: a) metabolically ubiquitous –i.e. found in multiple different organisms and metabolic pathways, and b) mechanistically specific –i.e. demanding a specific chemical toolkit and catalytic steps to happen.

### ***Limitations of EC classification***

We identified enzymes of similar EC classification that have substantial functional differences. Although we are looking at identity in all 4 EC levels, which should imply identity in substrate selectivity, this is fuzzy in some enzymes, such as polymer-acting ones, or ones using multiple substrates, like chloride peroxidases (EC 1.11.1.10). Three isozymes on our set share this EC and are totally unrelated. Structures, mechanisms, and metabolic contexts are completely different, and their only commonality is selectivity for Cl<sup>-</sup> ions and H<sub>2</sub>O<sub>2</sub>. The third cognate substrate is a variable organic compound that is chlorinated.

Reactions like this are generic and involve several components, one of which is simplistic (like an ion). Also, Markush structures are often used to represent ensembles of substrates that share a common ‘catalysable’ moiety. In such cases, EC often classifies the enzyme based on the simplest cognate ligand. This means that functional differences could be concealed by the EC classification process. Similar classification ambiguities are observed in promiscuous enzymes, such as proteases or DNases, leading to the same problem. Additional complexity in EC classification can derive from the type of cognate ligand considered for defining the 4<sup>th</sup> number.

In most hydrolases, for example phosphatases from the 3.1.3.x sub-subclass, the sub-sub-subclass is defined by substrate selectivity whereas in triterpene synthases (5.4.99.x), it is defined by product selectivity.

## Discussion

We have collected a sample of isozymes with sufficient structural and functional evidence, to reveal some “rules” or paradigms of convergent evolution. These data, albeit limited, have allowed us to identify cases of chemical convergence in enzymes (catalysis of the same overall reaction), accompanied by structural convergence in the active site and mechanistic convergence. We showed that these three levels of similarities might not necessarily co-exist, and based on this, we defined three paradigms of convergence: structural and mechanistic (13 cases), mechanistic-only (11 cases) and reaction-only (9 cases).

Our methodology for comparing isozymes, among others, included examination of cofactor selectivity. If we investigate the paradigms further by considering cofactor selectivity, there were 4 unobserved sub-paradigms in this survey. One sub-paradigm refers to enzymes with similar active site geometries, dissimilar mechanisms, and the same cofactor selectivity. This might also not be observed because having these traits in common is usually indicative of a homologous relationship. Conversely, we did not observe any enzymes that had similar active site geometry and mechanism, but significantly different cofactors (e.g. a metal vs. organic cofactor); this is also expected as the presence of cofactors would affect the geometry of the active site. We also do not see a pair of enzymes with similar active site geometries and mechanisms, where one enzyme uses a cofactor, and one does not. This is also expected for the same reason as the previous sub-paradigm; however, we observe this when mechanisms differ (e.g. in acid phosphatases). Although there are examples where a reaction can be performed under both a cofactor dependent and independent mechanism, these are mostly classified into the “reaction-only” paradigm. This is probably because it is quite unlikely that enzymes would have the same geometry and cofactors, without having any mechanistic commonality. Overall, these results could imply some level of rigidity that comes with active site geometries being similar. Last, we do not see any cases where two enzymes have significantly different cofactors that are both present, and similar mechanisms. This could be attributed to the limited variety of cofactors in nature, where each cofactor has evolved to serve

a specific role in catalysis[84]; thus, it is unlikely for two enzymes to evolve to bind different cofactors and still have a similar mechanism.

Classification of isoenzyme groups into paradigms was performed in a semi-manual way, by calculating some quantitative measures of similarity (catalytic residue superposition, mechanistic similarity), and by referring to available literature. In this process, we realised that the quantitative measures alone are not sufficient for assignment into paradigms, therefore, each case was examined individually and, if necessary, the quantitative results were overridden. For example, there were several cases where geometrical similarities were found to be spurious, and the active sites had to be examined manually to ensure biological relevance. Interestingly in the context of template matching: we have found that smaller templates, consisting of a few atoms, may yield spurious matches, even if the search is restricted to active sites. Therefore, one needs to be aware of false positive results in template-based searches, especially when these are performed at a high-throughput level.

To address the question of why solutions to chemical functions have evolved several times, we have discussed several examples of comparable chemical reactions by analogous enzymes. These data indicate that the factors influencing convergent chemistry between non-homologous enzymes are: 1) a common need for a metabolic reaction, but independent evolution has led to the emergence of different molecules. This is mostly observable in analogues from different kingdoms of life. 2) a different metabolic context demanding different enzymes. 3) different regulation of isoenzymes leading to enzymes of functional promiscuity. 4) Environmental pressures to solve biological challenges with the limited raw material available (i.e. amino acids, water, metals, etc.) lead to similar solutions, with the more efficient being retained.

Analysing how proteins with completely different folds maintain identical catalytic activity, we found many cases where a reaction might be mechanistically and/or structurally constrained. The more constrained the process, the higher the similarities between analogues (e.g. in phosphatases and in proteases). Other processes are less constrained, and convergence is only observable at the end of the reaction. Our approach to examining convergence from various points of view was hybrid, including both manual inspection and automated/quantitative methods. The latter can be scaled in a reproducible pipeline, allowing a more systematic and comprehensive analysis of convergence in the future, not necessarily restricted within the limits of M-CSA. In the interest of clarity and conceptualisation we have

generalised the observations from the data, where each of the pairs of enzymes evaluated provides unique insights into how proteins can maintain the same catalytic potential on a completely different overall shape.

## Methods

### *Data Preparation*

Groups of enzymes catalysing the same reaction and having different folds were collected from M-CSA[85] using the public API, by querying for entries with identical EC numbers and different CATH[86] numbers in the catalytic domain of their reference structure. Entries were grouped by EC number, and those groups having redundant CATH numbers, multiple CATH numbers (e.g. active site formed in the interface of two domains), or incomplete information (not all 4 EC levels determined, no CATH mapping, etc) were removed. In cases of EC groups having more than two enzymes after filtering, relationships were analysed in pairs, so that groups with 3 enzymes would result in 3 distinct analyses. This first filtering resulted in 46 pairs of enzymes that perform the same function and have different folds, across 34 unique EC numbers. Some extra filtering was done to ensure that this survey focuses solely on pairs that have similarities due to functional convergence. From the original set of 46 pairs, 4 were excluded as sequence or structural similarity implied potential homology. Pairs from the restriction enzyme group (EC 3.1.21.4, accounting for six pairs) were also excluded for the same reason [67-69]. 5 pairs were taken out due to limited available information. In several cases, two enzymes may not bind identical substrates or products, but those are still in scope since some of them are substrate promiscuous. Most of these were due to unspecified -R groups, different substrate specificities, or both enzymes manipulating the same functional groups on different molecules. The final, reduced set included 31 pairs across 23 EC numbers. 21 of the 31 pairs were found to have mechanism similarity, however none of the pairs showed complete identity.

### *Literature analysis*

Each pair was analysed using information found on M-CSA (enzyme mechanism and catalytic residue reaction roles), PDBe[87] (representative crystal structure selection, assembly composition), PDBsum[88] (bound ligand information), and the enzymes associated literature from these databases.

### ***Sequence and structure comparison***

Local and global sequence alignments were performed using the Needleman-Wunsch and Smith-Waterman algorithms respectively via the EMBOSS service[88,89]. Alignments were visualized in JalView[90].

### ***Active site structure comparison***

Functional atom templates[9] (three atoms per residue) were generated for each active site, taking all possible combinations of 2, 3 and 4 residues, using the CSA-3D package of our previous work[91]. Template definitions are mechanism-informed (e.g. three backbone atoms are selected for residues known to contribute their backbone to the mechanism) and allow for alternative matching of chemically similar atoms and residues, as described in ref. [6]. Templates of the same size were structurally compared using the template matching program Jess[92], in an “all against all” fashion. The match with the lowest RMSD and highest number of residues was selected as a reference atom-atom mapping frame for dynamic active site fitting.

### ***Dynamic active site fitting***

Using the active site atom-atom mapping method described above, enzyme pairs were fitted on their functional atoms, using a Gaussian-weighted version of the Kabsch algorithm[93,94], as described in ref. [91]. The algorithm outputs the two enzymes with their coordinates transformed accordingly, along with a weighted RMSD (wRMSD), an unweighted RMSD and a coverage value. The latter corresponds to the proportion of mapped catalytic residues over the number of residues of the largest of the two active sites.

### ***Catalytic mechanism comparison***

Mechanism comparison was conducted, which resulted in pairs being categorised as having “similar” or “not similar” mechanisms which was determined by seeing if the enzymes had *any* similarities in their mechanism. Similarities included near identical steps (ex: nucleophilic attack on the same atom, not necessarily the same nucleophile) or identical order of some or all events. This was initially done by manual inspection and comparison of the mechanisms found on M-CSA, and in cases where both enzymes in a pair had detailed mechanism description, a mechanism similarity score was calculated as follows: the data on the curly arrow diagrams of each catalytic step are first abstracted into a set of curly arrows and their chemical

environments. This representation of each curly arrow is called an arrow-environment and it includes the atoms interacting directly with the curly arrow and two shells of atoms and bonds around those. The similarity score for each mechanism pair is then calculated as the Jaccard index of the two sets of arrow-environments (*Ribeiro et al. [95] in preparation*).

### ***Cofactor binding comparison***

Cofactor preferences were compared for each enzyme pair, focusing only on cofactors directly involved in catalysis. The result of this analysis was placement into two broad categories that addressed identity, with pairs having the same cofactors being marked “yes” and ones with different cofactors being marked “no”. Within the “yes” and “no” categories there were subcategories. In the “yes” category pairs were marked as both having present cofactors that were identical, or both having no cofactor present. In the “no” category pairs were marked as both having cofactors that are different, or one having a cofactor present and one not having a cofactor present. We considered NAD and CoA cofactors if they were in the reaction regardless if their original chemical composition was regenerated at the end of the catalytic cycle.

## **References**

- 1 Ribeiro AJM, Riziots IG, Borkakoti N & Thornton JM (2023) Enzyme function and evolution through the lens of bioinformatics. *Biochemical Journal* **480**, 1845–1863.
- 2 Furnham N, Dawson NL, Rahman SA, Thornton JM & Orengo CA (2016) Large-Scale Analysis Exploring Evolution of Catalytic Machineries and Mechanisms in Enzyme Superfamilies. *J Mol Biol* **428**, 253–267.
- 3 Gherardini PF, Wass MN, Helmer-Citterich M & Sternberg MJE (2007) Convergent Evolution of Enzyme Active Sites Is not a Rare Phenomenon. *J Mol Biol* **372**, 817–845.
- 4 Omelchenko M V., Galperin MY, Wolf YI & Koonin E V. (2010) Non-homologous isofunctional enzymes: A systematic analysis of alternative solutions in enzyme evolution. *Biol Direct* **5**.
- 5 Galperin MY & Koonin E V. (2012) Divergence and convergence in enzyme evolution. *Journal of Biological Chemistry* **287**, 21–28.
- 6 Riziots IG, Ribeiro AJM, Borkakoti N & Thornton JM (2023) The 3D modules of enzyme catalysis: deconstructing active sites into distinct functional entities. *J Mol Biol* **435**.
- 7 Polgár L (2005) The catalytic triad of serine peptidases. *Cellular and Molecular Life Sciences* **62**.
- 8 Kubiak RJ, Yue X, Hondal RJ, Mihai C, Tsai MD & Bruzik KS (2001) Involvement of the Arg - Asp - His catalytic triad in enzymatic cleavage of the phosphodiester bond. *Biochemistry* **40**.
- 9 Riziots IG & Thornton JM (2022) Capturing the geometry, function, and evolution of enzymes with 3D templates. *Protein Science* **31**, e4363.
- 10 Arnold FH (2018) Directed Evolution: Bringing New Chemistry to Life. *Angewandte Chemie - International Edition* **57**.

- 11 Sillitoe I & Furnham N (2016) FunTree: Advances in a resource for exploring and contextualising protein function evolution. *Nucleic Acids Res* **44**, D317–D323.
- 12 Ribeiro AJM, Riziotis IG, Tyzack JD, Borkakoti N & Thornton JM (2023) EzMechanism: An Automated Tool to Propose Catalytic Mechanisms of Enzyme Reactions. *Nat Methods* **20**, 1516–1522.
- 13 Holliday GL, Mitchell JBO & Thornton JM (2009) Understanding the Functional Roles of Amino Acid Residues in Enzyme Catalysis. *J Mol Biol* **390**, 560–577.
- 14 Bartlett GJ, Borkakoti N & Thornton JM (2003) Catalysing new reactions during evolution: Economy of residues and mechanism. *J Mol Biol* **331**, 829–860.
- 15 Cronk JD, Endrizzi JA, Cronk MR, O’neill JW & Zhang KYJ (2001) Crystal structure of *E. coli*  $\beta$ -carbonic anhydrase, an enzyme with an unusual pH-dependent activity. *Protein Science* **10**.
- 16 Eriksson AE, Jones TA & Liljas A (1988) Refined structure of human carbonic anhydrase II at 2.0 Å resolution. *Proteins: Structure, Function, and Bioinformatics* **4**.
- 17 Groves MR, Yao ZJ, Roller PP, Burke TR & Barford D (1998) Structural basis for inhibition of the protein tyrosine phosphatase 1B by phosphotyrosine peptide mimetics. *Biochemistry* **37**.
- 18 Zhang M, Van Etten RL & Stauffacher C V. (1994) Crystal Structure of Bovine Heart Phosphotyrosyl Phosphatase at 2.2-Å Resolution. *Biochemistry* **33**.
- 19 Leesong M, Henderson BS, Gillig JR, Schwab JM & Smith JL (1996) Structure of a dehydratase-isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: Two catalytic activities in one active site. *Structure* **4**.
- 20 White A, Tull D, Johns K, Withers SG & Rose DR (1996) Crystallographic observation of a covalent catalytic intermediate in a  $\beta$ -glycosidase. *Nat Struct Biol* **3**.
- 21 Sidhu G, Withers SG, Nguyen NT, McIntosh LP, Ziser L & Brayer GD (1999) Sugar ring distortion in the glycosyl-enzyme intermediate of a family G/11 xylanase. *Biochemistry* **38**.
- 22 Burkhard P, Taylor P & Walkinshaw MD (2000) X-ray structures of small ligand-FKBP complexes provide an estimate for hydrophobic interaction energies. *J Mol Biol* **295**.
- 23 Howard BR, Vajdos FF, Li S, Sundquist WI & Hill CP (2003) Structural insights into the catalytic mechanism of cyclophilin A. *Nat Struct Biol* **10**.
- 24 Weiss MS, Palm GJ & Hilgenfeld R (2000) Crystallization, structure solution and refinement of hen egg-white lysozyme at pH 8.0 in the presence of MPD. *Acta Crystallogr D Biol Crystallogr* **56**.
- 25 Hermoso JA, Monterroso B, Albert A, Galán B, Ahrazem O, García P, Martínez-Ripoll M, García JL & Menéndez M (2003) Structural basis for selective recognition of pneumococcal cell wall by modular endolysin from phage Cp-1. *Structure* **11**.
- 26 Lowther WT, Brot N, Weissbach H & Matthews BW (2000) Structure and mechanism of peptide methionine sulfoxide reductase, an “anti-oxidation” enzyme. *Biochemistry* **39**.
- 27 Lowther WT, Weissbach H, Etienne F, Brot N & Matthews BW (2002) The mirrored methionine sulfoxide reductases of *neisseria gonorrhoeae* pilb. *Nat Struct Biol* **9**.
- 28 Johnston SC, Larsen CN, Cook WJ, Wilkinson KD & Hill CP (1997) Crystal structure of a deubiquitinating enzyme (human UCH-L3) at 1.8 Å resolution. *EMBO Journal* **16**.
- 29 Tan J, Verschueren KHG, Anand K, Shen J, Yang M, Xu Y, Rao Z, Bigalke J, Heisen B, Mesters JR, Chen K, Shen X, Jiang H & Hilgenfeld R (2005) pH-dependent conformational flexibility of the SARS-CoV main proteinase (Mpro) dimer: Molecular dynamics simulations and multiple X-ray structure analyses. *J Mol Biol* **354**.
- 30 Filipek R, Szczepanowski R, Sabat A, Potempa J & Bochtler M (2004) Prostaphopain B structure: A comparison of proregion-mediated and staphostatin-mediated protease inhibition. *Biochemistry* **43**.

31 Matthews DA, Dragovich PS, Webber SE, Fuhrman SA, Patock AK, Zalman LS, Hendrickson TF, Love RA, Prins TJ, Marakovits JT, Zhou R, Tikhe J, Ford CE, Meador JW, Ferre RA, Brown EL, Binford SL, Brothers MA, DeLisle DM & Worland ST (1999) Structure-assisted design of mechanism-based irreversible inhibitors of human rhinovirus 3C protease with potent antiviral activity against multiple rhinovirus serotypes. *Proc Natl Acad Sci U S A* **96**.

32 Kern AD, Oliveira MA, Coffino P & Hackert ML (1999) Structure of mammalian ornithine decarboxylase at 1.6 Å resolution: Stereochemical implications of PLP-dependent amino acid decarboxylases. *Structure* **7**.

33 Momany C, Ernst S, Ghosh R, Chang NL & Hackert ML (1995) Crystallographic structure of a PLP-dependent ornithine decarboxylase from *Lactobacillus 30a* to 3.0 Å resolution. *J Mol Biol* **252**.

34 Sliz P, Engelmann R, Hengstenberg W & Pai EF (1997) The structure of enzyme IIA(lactose) from *Lactococcus lactis* reveals a new fold and points to possible interactions of a multicomponent system. *Structure* **5**.

35 Herzberg O (1992) An atomic model for protein-protein phosphoryl group transfer. *Journal of Biological Chemistry* **267**.

36 Mahan SD, Ireton GC, Knoeber C, Stoddard BL & Black ME (2004) Random mutagenesis and selection of *Escherichia coli* cytosine deaminase for cancer gene therapy. *Protein Engineering, Design and Selection* **17**.

37 Ko TP, Lin JJ, Hu CY, Hsu YH, Wang AHJ & Liaw SH (2003) Crystal structure of yeast cytosine deaminase: Insights into enzyme mechanism and evolution. *Journal of Biological Chemistry* **278**.

38 Weston SA, Lahm A & Suck D (1992) X-ray structure of the DNase I-d(GGTATACC)2 complex at 2.3 Å resolution. *J Mol Biol* **226**.

39 Roszak AW, Robinson DA, Krell T, Hunter IS, Fredrickson M, Abell C, Coggins JR & Lapthorn AJ (2002) The structure and mechanism of the type II dehydroquinase from *Streptomyces coelicolor*. *Structure* **10**.

40 Gourley DG, Shrive AK, Polikarpov I, Krell T, Coggins JR, Hawkins AR, Isaacs NW & Sawyer L (1999) The two types of 3-dehydroquinase have distinct structures but catalyze the same overall reaction. *Nat Struct Biol* **6**.

41 Bompard-Gilles C, Villeret V, Davies GJ, Fanuel L, Joris B, Frère JM & Van Beeumen J (2000) A new variant of the Ntn hydrolase fold revealed by the crystal structure of L-aminopeptidase D-Ala-esterase/amidase from *Ochrobactrum anthropi*. *Structure* **8**.

42 Bompard-Gilles C, Remaut H, Villeret V, Prangé T, Fanuel L, Delmarcelle M, Joris B, Frère JM & Van Beeumen J (2000) Crystal structure of a D-aminopeptidase from *Ochrobactrum anthropi*, a new member of the “penicillin-recognizing enzyme” family. *Structure* **8**.

43 Burschowsky D, Van Eerde A, Okvist M, Kienhöfer A, Kast P, Hilvert D & Krenzel U (2014) Electrostatic transition state stabilization rather than reactant destabilization provides the chemical basis for efficient Chorismate mutase catalysis. *Proc Natl Acad Sci U S A* **111**.

44 Sträter N, Schnappauf G, Braus G & Lipscomb WN (1997) Mechanisms of catalysis and allosteric regulation of yeast chorismate mutase from crystal structures. *Structure* **5**.

45 Kissinger CR, Parge HE, Knighton DR, Lewis CT, Pelletier LA, Tempczyk A, Kalish VJ, Tucker KD, Showalter RE, Moomaw EW, Gastinel LN, Habuka N, Chen X, Maldonado F, Barker JE, Bacquet R & Villafranca JE (1995) Crystal structures of human calcineurin and the human FKBP12–FK506–calcineurin complex. *Nature* **378**.

46 Lee JO, Yang H, Georgescu MM, Cristofano A Di, Maehama T, Shi Y, Dixon JE, Pandolfi P & Pavletich NP (1999) Crystal structure of the PTEN tumor suppressor: Implications for its phosphoinositide phosphatase activity and membrane association. *Cell* **99**.

47 Davies GJ, Tolley SP, Henrissat B, Hjort C & Schiilein M (1995) Structures of Oligosaccharide-Bound Forms of the Endoglucanase V from *Humicola insolens* at 1.9 Å Resolution. *Biochemistry* **34**.

48 Sakon J, Irwin D, Wilson DB & Andrew Karplus P (1997) Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. *Nat Struct Biol* **4**.

49 Sulzenbacher G, Mackenzie LF, Wilson KS, Withers SG, Dupont C & Davies GJ (1999) The crystal structure of a 2-fluorocellotriosyl complex of the *Streptomyces lividans* endoglucanase CelB2 at 1.2 Å resolution. *Biochemistry* **38**.

50 Chao X, Muff TJ, Park SY, Zhang S, Pollard AM, Ordal GW, Bilwes AM & Crane BR (2006) A receptor-modifying deamidase in complex with a signaling phosphatase reveals reciprocal regulation. *Cell* **124**.

51 West AH, Martinez-Hackert E & Stock AM (1995) Crystal structure of the catalytic domain of the chemotaxis receptor methylesterase, cheb. *J Mol Biol* **250**.

52 Hahn M, Keitel T & Heinemann U (1995) Crystal and Molecular Structure at 0.16-nm Resolution of the Hybrid *Bacillus* Endo-1,3-1,4-beta-D-Glucan 4-Glucanohydrolase H(A16-M). *Eur J Biochem* **232**.

53 Müller JJ, Thomsen KK & Heinemann U (1998) Crystal structure of Barley 1,3-1,4- $\beta$ -glucanase at 2.0-Å resolution and comparison with *Bacillus* 1,3-1,4- $\beta$ -glucanase. *Journal of Biological Chemistry* **273**.

54 Suresh CG, Pundle A V., SivaRaman H, Rao KN, Brannigan JA, McVey CE, Verma CS, Dauter Z, Dodson EJ & Dodson GG (1999) Penicillin V acylase crystal structure reveals new Ntn-hydrolase family members. *Nat Struct Biol* **6**.

55 Duggleby HJ, Tolley SP, Dodson EJ, Dodson G, Moody PCE & Hill CP (1995) Penicillin acylase has a single-amino-acid catalytic centre. *Nature* **373**.

56 Ramaswamy S, Park DH & Plapp B V. (1999) Substitutions in a flexible loop of horse liver alcohol dehydrogenase hinder the conformational change and unmask hydrogen transfer. *Biochemistry* **38**.

57 Benach J, Winberg JO, Svendsen JS, Atrian S, González-Duarte R & Ladenstein R (2005) *Drosophila* alcohol dehydrogenase: Acetate-enzyme interactions and novel insights into the effects of electrostatics on catalysis. *J Mol Biol* **345**.

58 Dutnall RN, Tafrov ST, Sternglanz R & Ramakrishnan V (1998) Structure of the histone acetyltransferase Hat1: A paradigm for the GCN5-related N-acetyltransferase superfamily. *Cell* **94**.

59 Iverson TM, Alber BE, Kisker C, Ferry JG & Rees DC (2000) A closer look at the active site of  $\gamma$ -class carbonic anhydrases: High-resolution crystallographic studies of the carbonic anhydrase from *methanosarcina thermophila*. *Biochemistry* **39**.

60 Lindqvist Y, Schneider G & Vihko P (1994) Crystal structures of rat acid phosphatase complexed with the transition-state analogs vanadate and molybdate: Implications for the reaction mechanism. *Eur J Biochem* **221**.

61 Ishikawa K, Mihara Y, Gondoh K, Suzuki EI & Asano Y (2000) X-ray structures of a novel acid phosphatase from *Escherichia blattae* and its complex with the transition-state analog molybdate. *EMBO Journal* **19**.

62 Klabunde T, Sträter N, Fröhlich R, Witzel H & Krebs B (1996) Mechanism of Fe(III)-Zn(II) purple acid phosphatase based on crystal structures. *J Mol Biol* **259**.

63 Dessen A, Tang J, Schmidt H, Stahl M, Clark JD, Seehra J & Somers WS (1999) Crystal structure of human cytosolic phospholipase A2 reveals a novel topology and catalytic mechanism. *Cell* **97**.

64 Sugiyama M, Ohtani K, Izuhara M, Koike T, Suzuki K, Imamura S & Misaki H (2002) A novel prokaryotic phospholipase A2. Characterization, gene cloning, and solution structure. *Journal of Biological Chemistry* **277**.

65 Bravo J, Verdague N, Tormo J, Betzel C, Switala J, Loewen PC & Fita I (1995) Crystal structure of catalase HPII from *Escherichia coli*. *Structure* **3**.

66 Whittaker MM, Barynin V V., Igarashi T & Whittaker JW (2003) Outer sphere mutagenesis of *Lactobacillus plantarum* manganese catalase disrupts the cluster core: Mechanistic implications. *Eur J Biochem* **270**.

67 Hofmann B, Tölzer S, Pelletier I, Altenbuchner J, Van Pee KH & Hecht HJ (1998) Structural investigation of the cofactor-free chloroperoxidases. *J Mol Biol* **279**.

68 Sundaramoorthy M, Terner J & Poulos TL (1995) The crystal structure of chloroperoxidase: a heme peroxidase-cytochrome P450 functional hybrid. *Structure* **3**.

69 Messerschmidt A & Wever R (1996) X-ray structure of a vanadium-containing enzyme: Chloroperoxidase from the fungus *Curvularia inaequalis*. *Proc Natl Acad Sci U S A* **93**.

70 Divne C, Ståhlberg J, Reinikainen T, Ruohonen L, Pettersson G, Knowles JKC, Teeri TT & Jones TA (1994) The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science* (1979) **265**.

71 Zou JY, Kleywegt GJ, Ståhlberg J, Driguez H, Nerinckx W, Claeysseens M, Koivula A, Teeri TT & Jones TA (1999) Crystallographic evidence for substrate ring distortion and protein conformational changes during catalysis in cellobiohydrolase Cel6A from *Trichoderma reesei*. *Structure* **7**.

72 Sun T, Nukaga M, Mayama K, Braswell EH & Knox JR (2003) Comparison of  $\beta$ -lactamases of classes A and D: 1.5-Å crystallographic structure of the class D OXA-1 oxacillinase. *Protein Science* **12**, 82–91.

73 Ullah JH, Walsh TR, Taylor IA, Emery DC, Verma CS, Gamblin SJ & Spencer J (1998) The crystal structure of the L1 metallo- $\beta$ -lactamase from *Stenotrophomonas maltophilia* at 1.7 Å resolution. *J Mol Biol* **284**.

74 Davies JF, Delcamp TJ, Prendergast NJ, Ashford VA, Freisheim JH & Kraut J (1990) Crystal Structures of Recombinant Human Dihydrofolate Reductase Complexed with Folate and 5-Deazafolate. *Biochemistry* **29**.

75 Narayana N, Matthews DA, Howell EE & Xuong NH (1995) A plasmid-encoded dihydrofolate reductase from trimethoprim-resistant bacteria has a novel d2-symmetric active site. *Nat Struct Biol* **2**.

76 Herron SR, Benen JAE, Scavetta RD, Visser J & Jurnak F (2000) Structure and function of pectic enzymes: Virulence factors of plant pathogens. *Proc Natl Acad Sci U S A* **97**.

77 De Armas HN, Verboven C, De Ranter C, Desair J, Vande Broek A, Vanderleyden J & Rabijns A (2004) *Azospirillum irakense* pectate lyase displays a toroidal fold. *Acta Crystallogr D Biol Crystallogr* **60**.

78 Bell CE & Eisenberg D (1996) Crystal structure of diphtheria toxin bound to nicotinamide adenine dinucleotide. *Biochemistry* **35**.

79 Zhang RG, Westbrook ML, Nance S, Spangler BD, Westbrook EM, Scott DL & Shipley GG (1995) The three-dimensional crystal structure of cholera toxin. *J Mol Biol* **251**.

80 Hahn M, Keitel T & Heinemann U (1995) Crystal and Molecular Structure at 0.16-nm Resolution of the Hybrid *Bacillus* Endo-1,3-1,4-beta-D-Glucan 4-Glucanohydrolase H(A16-M). *Eur J Biochem* **232**.

81 Fanuel L, Thamm I, Kostanjevecki V, Samyn B, Joris B, Goffin C, Brannigan J, Van Beeumen J & Frère JM (1999) Two new aminopeptidases from *Ochrobactrum anthropi* active on D-alanyl-p-nitroanilide. *Cellular and Molecular Life Sciences* **55**.

82 Cairns JRK & Esen A (2010)  $\beta$ -Glucosidases. *Cellular and Molecular Life Sciences* **67**.

83 Davies G & Henrissat B (1995) Structures and mechanisms of glycosyl hydrolases. *Structure* **3**.

84 Goldman AD & Kacar B (2021) Cofactors are Remnants of Life's Origin and Early Evolution. *J Mol Evol* **89**.

85 Ribeiro AJM, Holliday GL, Furnham N, Tyzack JD, Ferris K & Thornton JM (2018) Mechanism and Catalytic Site Atlas (M-CSA): A database of enzyme reaction mechanisms and active sites. *Nucleic Acids Res* **46**, D618–D623.

86 Sillitoe I, Dawson N, Lewis TE, Das S, Lees JG, Ashford P, Tolulope A, Scholes HM, Senatorov I, Bujan A, Ceballos Rodriguez-Conde F, Dowling B, Thornton J & Orengo CA (2019) CATH: Expanding the horizons of structure-based functional annotations for genome sequences. *Nucleic Acids Res* **47**, D280–D284.

87 Velankar S, Best C, Beuth B, Boutselakis CH, Cobley N, Sousa da Silva AW, Dimitropoulos D, Golovin A, Hirshberg M, John M, Krissinel EB, Newman R, Oldfield T, Pajon A, Penkett CJ, Pineda-Castillo J, Sahni G, Sen S, Slowley R, Suarez-Uruena A, Swaminathan J, Van Ginkel G, Vranken WF, Henrick K & Kleywegt GJ (2009) PDBe: Protein Data Bank in Europe. *Nucleic Acids Res* **38**.

88 Laskowski RA, Jabłońska J, Pravda L, Vařeková RS & Thornton JM (2018) PDBsum: Structural summaries of PDB entries. *Protein Science* **27**.

89 Rice P, Longden L & Bleasby A (2000) EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics* **16**.

90 Waterhouse AM, Procter JB, Martin DMA, Clamp M & Barton GJ (2009) Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**.

91 Riziotis IG, Ribeiro AJM, Borkakoti N & Thornton JM (2022) Conformational Variation in Enzyme Catalysis: A Structural Study on Catalytic Residues. *J Mol Biol* **434**.

92 Barker JA & Thornton JM (2003) An algorithm for constraint-based structural template matching: Application to 3D templates with statistical analysis. *Bioinformatics* **19**, 1644–1649.

93 Kabsch W (1976) A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A* **32**, 922–923.

94 Damm KL & Carlson HA (2006) Gaussian-weighted RMSD superposition of proteins: A structural comparison for flexible proteins and predicted protein structures. *Biophys J* **90**, 4558–4573.

95 Ribeiro AJM & et al. (2024) Measuring Catalytic Mechanism Similarity - A New Approach to Study Enzyme Function and Evolution. *In Preparation*.

## Conflict of interest

Authors declare no conflict of interest.

## Acknowledgements

The work was supported by the EMBL International PhD Programme (IGR) and the European Molecular Biology Laboratory (JCK, GO, AJMR, NB, JMT)

## Author contributions

**IGR:** Conceptualisation, Methodology, Software, Validation, Writing – Original Draft preparation

**JCK:** Methodology, Investigation, Visualisation

**GO:** Software, Methodology

**AJMR:** Data Curation, Validation

**NB:** Conceptualisation, Validation, Writing – Review & Editing

**JMT:** Conceptualisation, Supervision, Resources, Funding Acquisition, Project Administration, Writing – Review & Editing