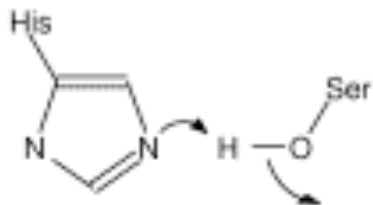


Bartlett, Porter, Borkakoti & Thornton 2002. *J Mol Biol* **324**:105-121

Analysis of Catalytic Residues in Enzyme Active Sites

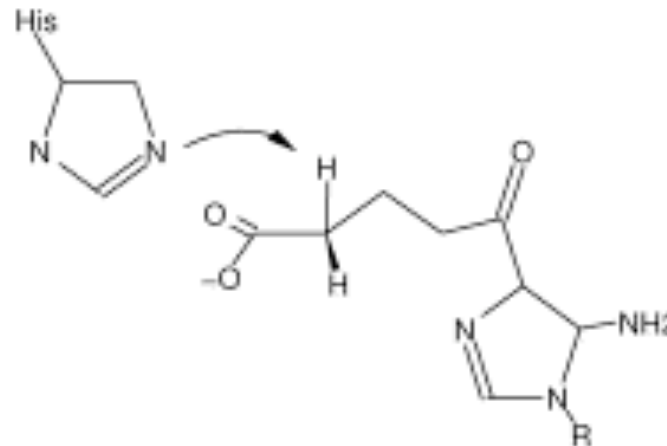
Distinction between “acid/base” and “substrate activation” definitions

a) Carboxypeptidase D



Primes residue for nucleophilic attack

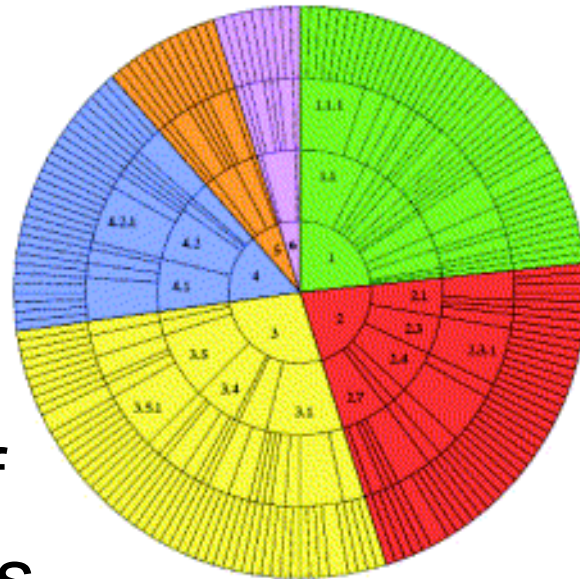
b) Adenylosuccinate lyase



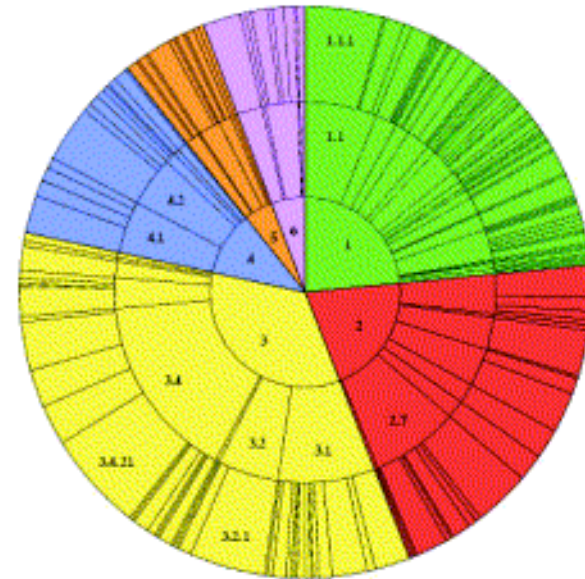
Deprotonation of substrate

Figure 2. The role of histidine in the first step of serine protease and adenylosuccinate lyase reactions.[10. and 38.] (a) Carboxypeptidase D, histidine primes serine residue for nucleophilic attack on the substrate; (b) Adenylosuccinate lyase, histidine residue directly deprotonates the substrate.

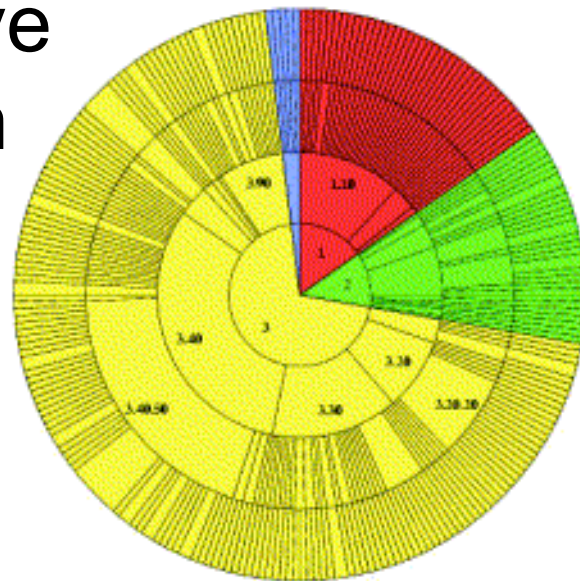
Is the set of 178 enzymes representative of all known structures?



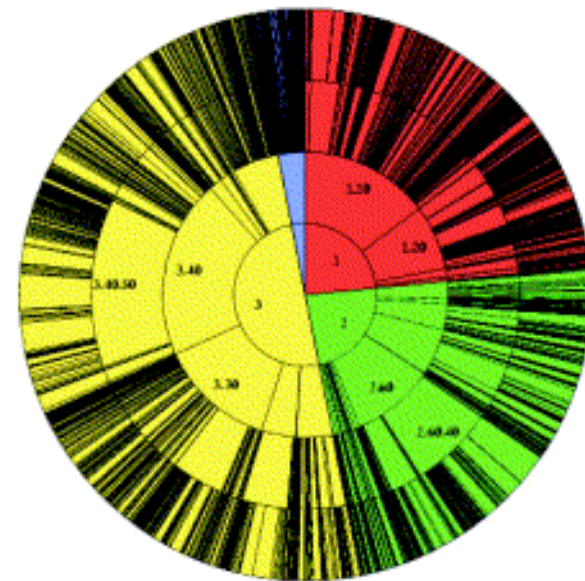
a. EC wheel functional classification of dataset



b. EC numbers of all known enzymes⁴⁰



c. CATH wheel structural classification of enzyme dataset



d. CATH wheel of all structures in PDB

Table 3. Catalytic residue types and their secondary structure compared with all residues in the dataset

	Catalytic residue type ^a			Secondary structure environment		
	Charged (%)	Polar (%)	Hydrophobic (%)	Alpha helix (%)	Beta sheet (%)	Coil (%)
Catalytic residues	65	27	8	28	22	50
All residues	25	25	50	47	23	30

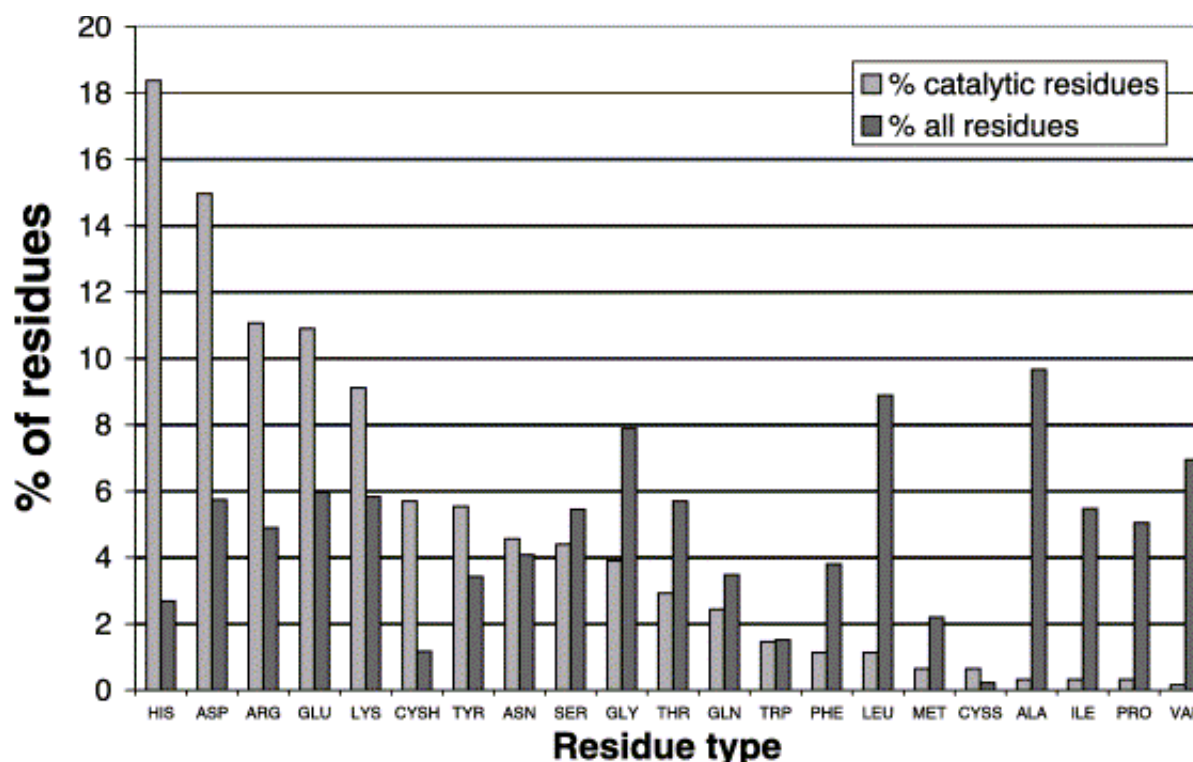


Figure 4. Observed frequency distribution of catalytic residue types compared with all residues in the dataset. CYSH indicates free cysteine residues. CYSS indicates disulphide-bridged cysteine residues. Catalytic residues were taken from each structure. In the case of structures with multiple subunits, the smallest possible unique unit was taken, e.g. in a homodimer with catalytic residues on one subunit only, one subunit was used for the all residue calculation. If the catalytic residues were split across two subunits, and there were two active sites in the homodimer, only one subunit was used for the all residue calculation. However, if catalytic residues were split across two subunits, and there was only one active site in the dimer, both subunits were used for the all residue calculation.

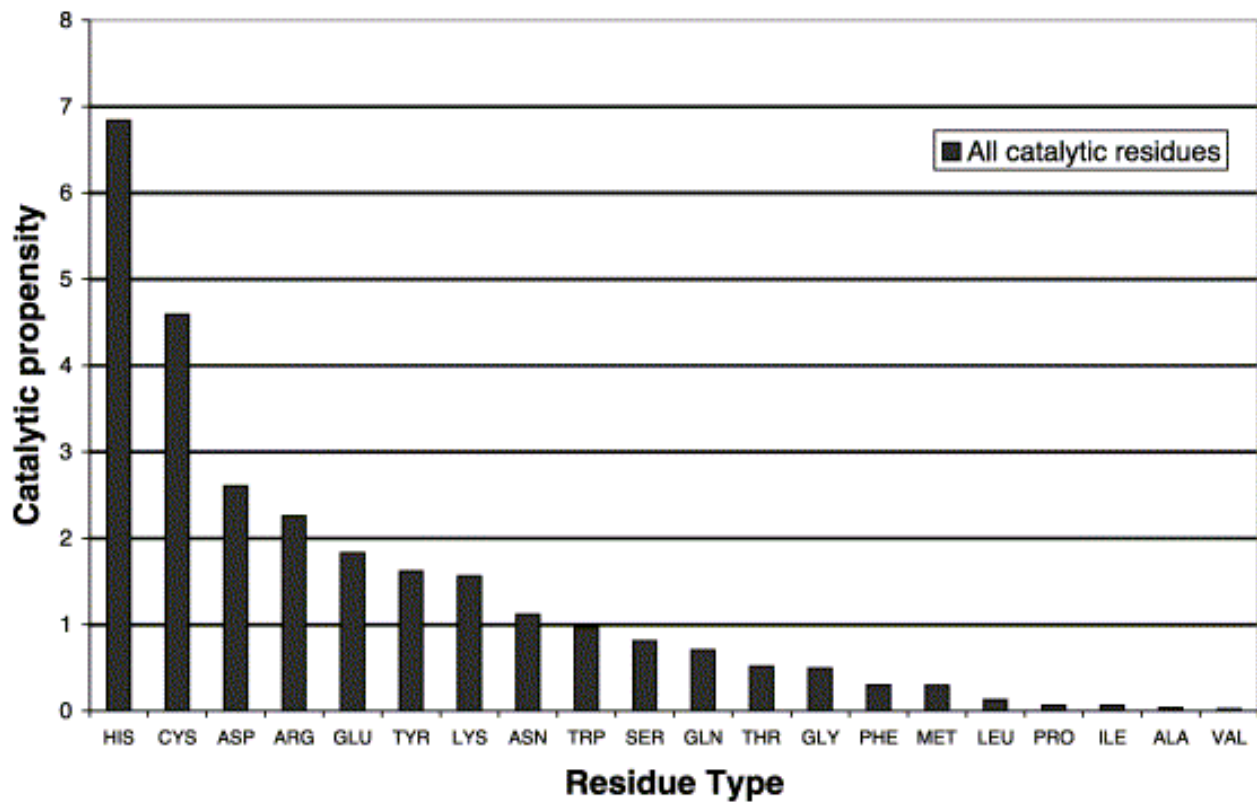


Figure 5. Catalytic propensity of residue types. Catalytic propensity is defined as the percentage of catalytic residues constituted by a particular residue type, divided by the percentage of all residues constituted by the same particular residue type.

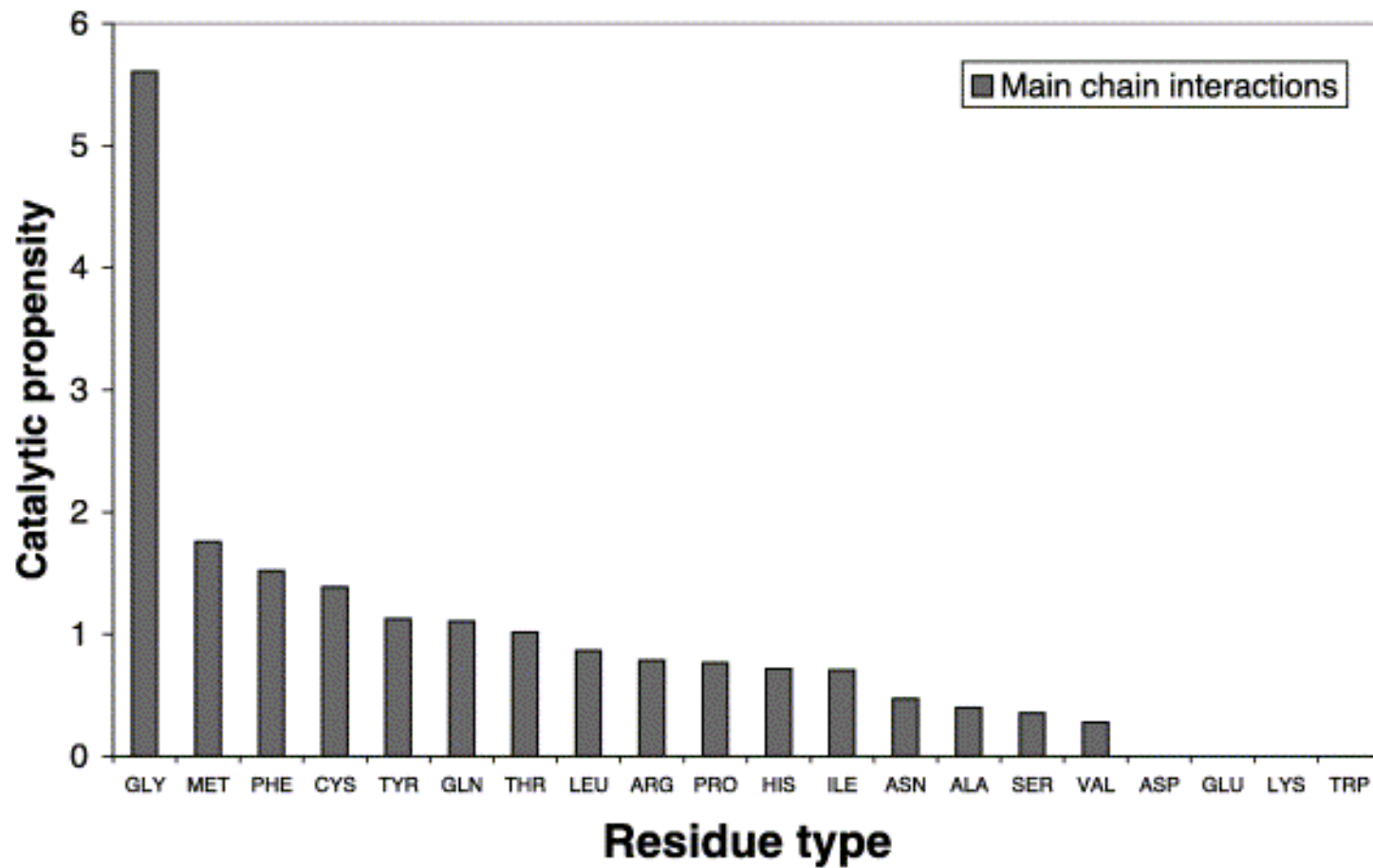


Figure 6. Catalytic propensity of residues interacting *via* their main-chain N–H or C=O groups.

Table 4. Hydrophobic residues aiding catalysis via their side-chain as opposed to their main-chain

Residue	Enzyme	Description of function
Met219	Human fibroblast stromelysin-1	Enhances effective concentration of Zn^{2+} cofactor, which coordinates and enhances the nucleophilicity of a hydroxyl nucleophile. ^{41,42}
Met20	Dihydrofolate reductase	Provides a hydrophobic region pushing positive charge from N5 of folate to C6 where it can accept hydride from NADPH. ³⁷
Leu28	Dihydrofolate reductase	Constrains folate ring in optimum position to receive hydride. ³⁷
Leu54	Dihydrofolate reductase	Constrains folate ring in optimum position to receive hydride. ³⁷
Leu20	D-amino-acid aminotransferase	Aids PLP cofactor catalysis by supporting the ring orientation without disturbing oscillating motions. ⁴³
Gly734	Pyruvate-formate lyase	Radical formation at the C- α position. ^{44,45}
Phe31	Dihydrofolate reductase	Forces proximity between folate and NADPH optimising hydride transfer. ³⁷
Phe175	L-2-haloacid dehalogenase	Forms a halide stabilising cradle which makes the halide a better leaving group. ⁴⁶
Phe77	Pentalenene synthase	Stabilises carbocation intermediate with Asn219, by cation- π interactions. ⁴⁷
Phe50	4-oxalocrotonate tautomerase	Provides a hydrophobic environment to lower the pK_a of the N-terminal nucleophile. ⁴⁸

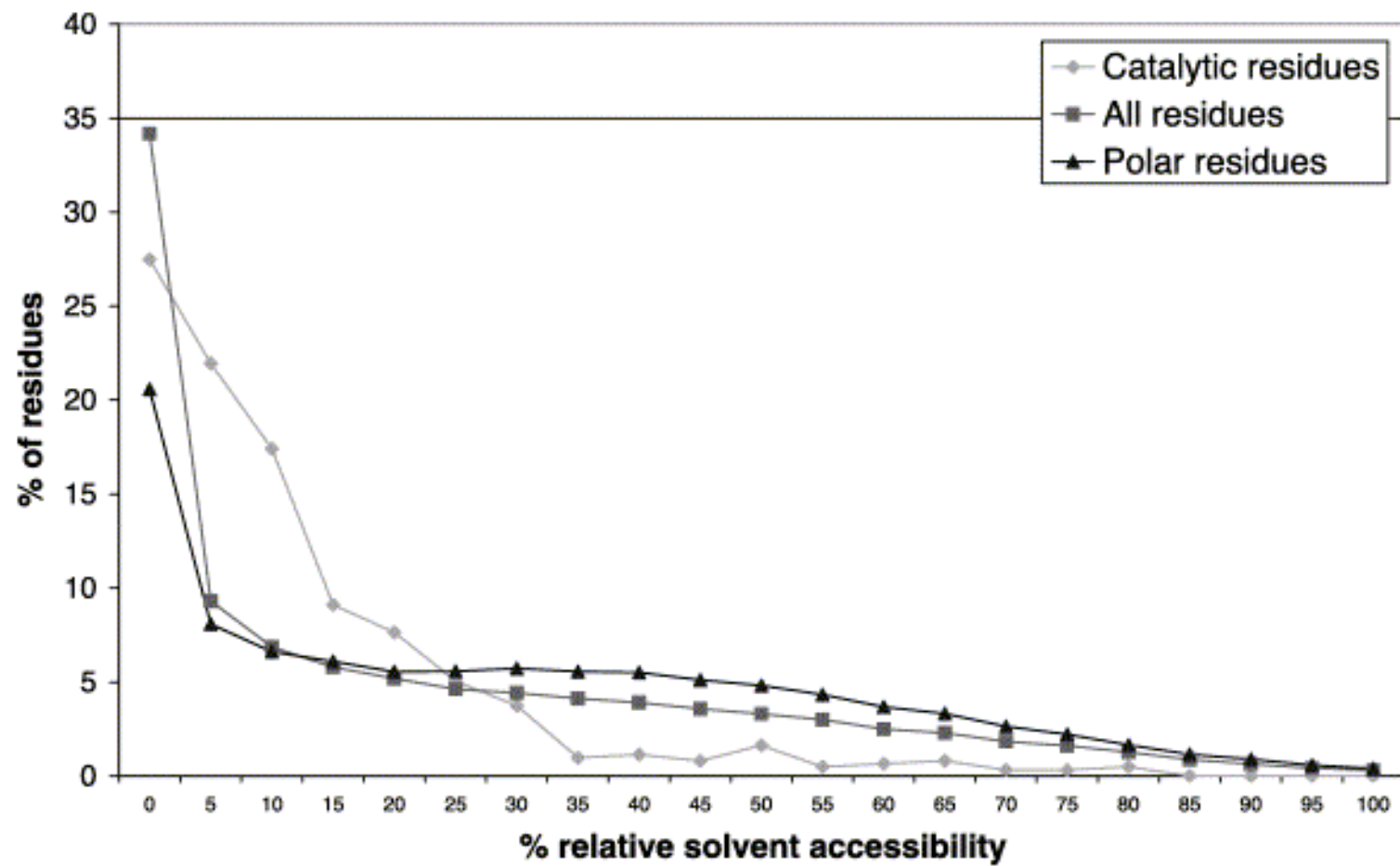


Figure 7. Residue solvent accessibility in the absence of ligands.

Table 5. Occurrence of catalytic residues in clefts in the enzyme, as calculated by SURFNET[12.]

	Number of enzymes (%)
≥50% of catalytic residues in three largest clefts	151 (85%)
≥50% of catalytic residues in any cleft	160 (90%)
At least one catalytic residue in any cleft	165 (93%)
No catalytic residues in any cleft	12 (7%)

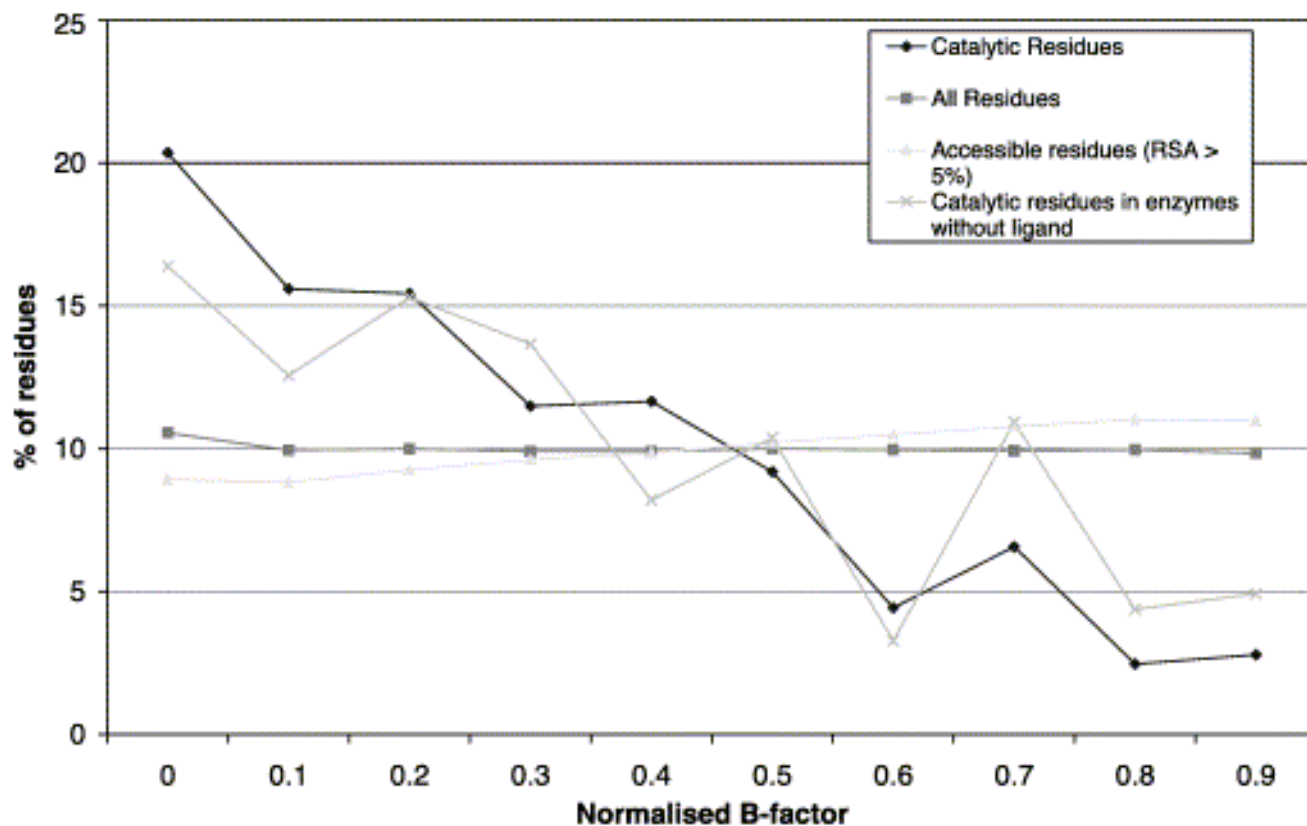


Figure 9. Catalytic residue *B*-factors—a measure of residue flexibility. Absolute *B*-factors were taken from the PDB file for each enzyme and normalised over the whole protein. Enzyme structures determined by NMR (1mek and 1adn) were excluded. Normalised *B*-factor values were placed into bins and the percentage of residues in each bin displayed.

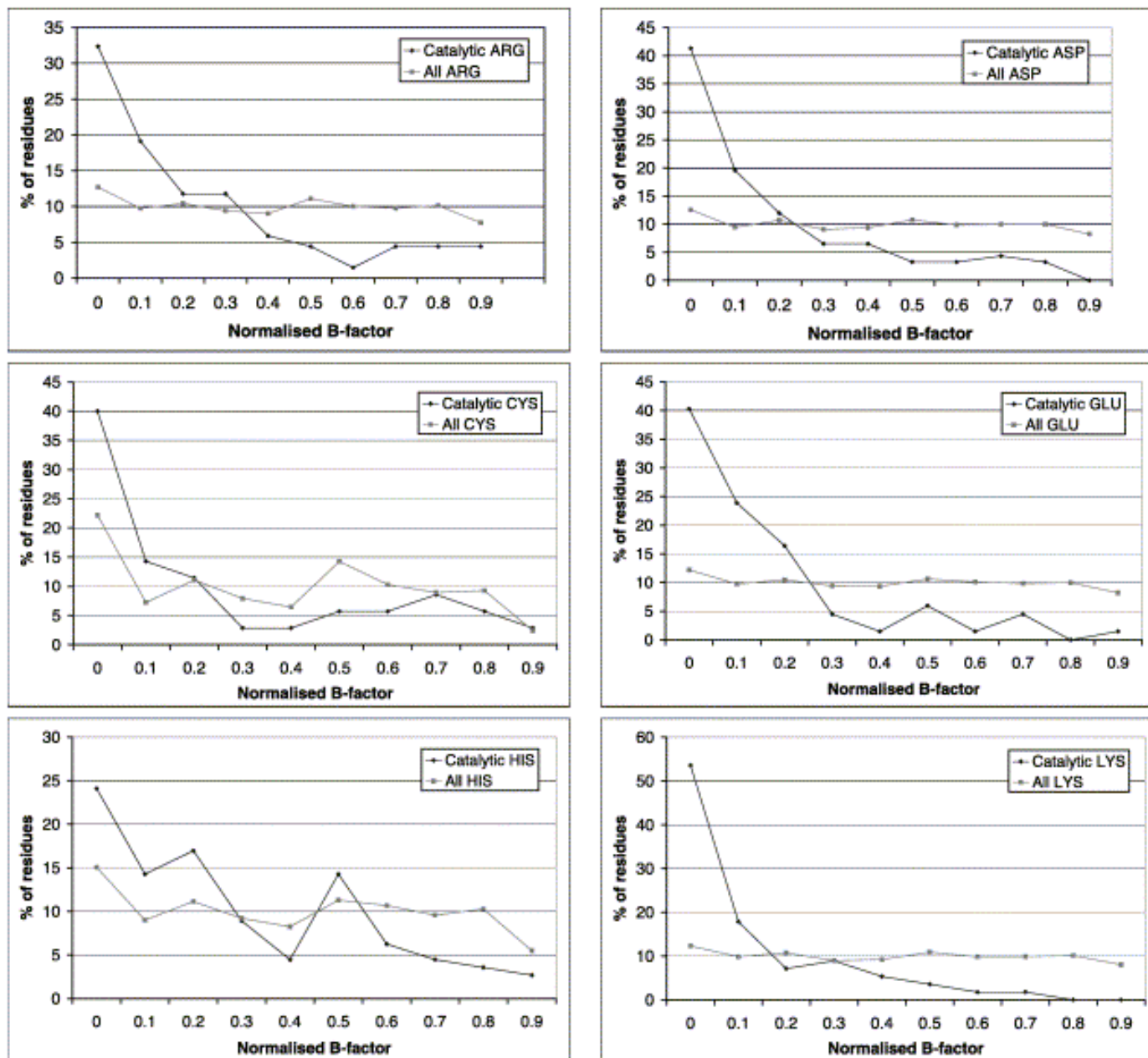
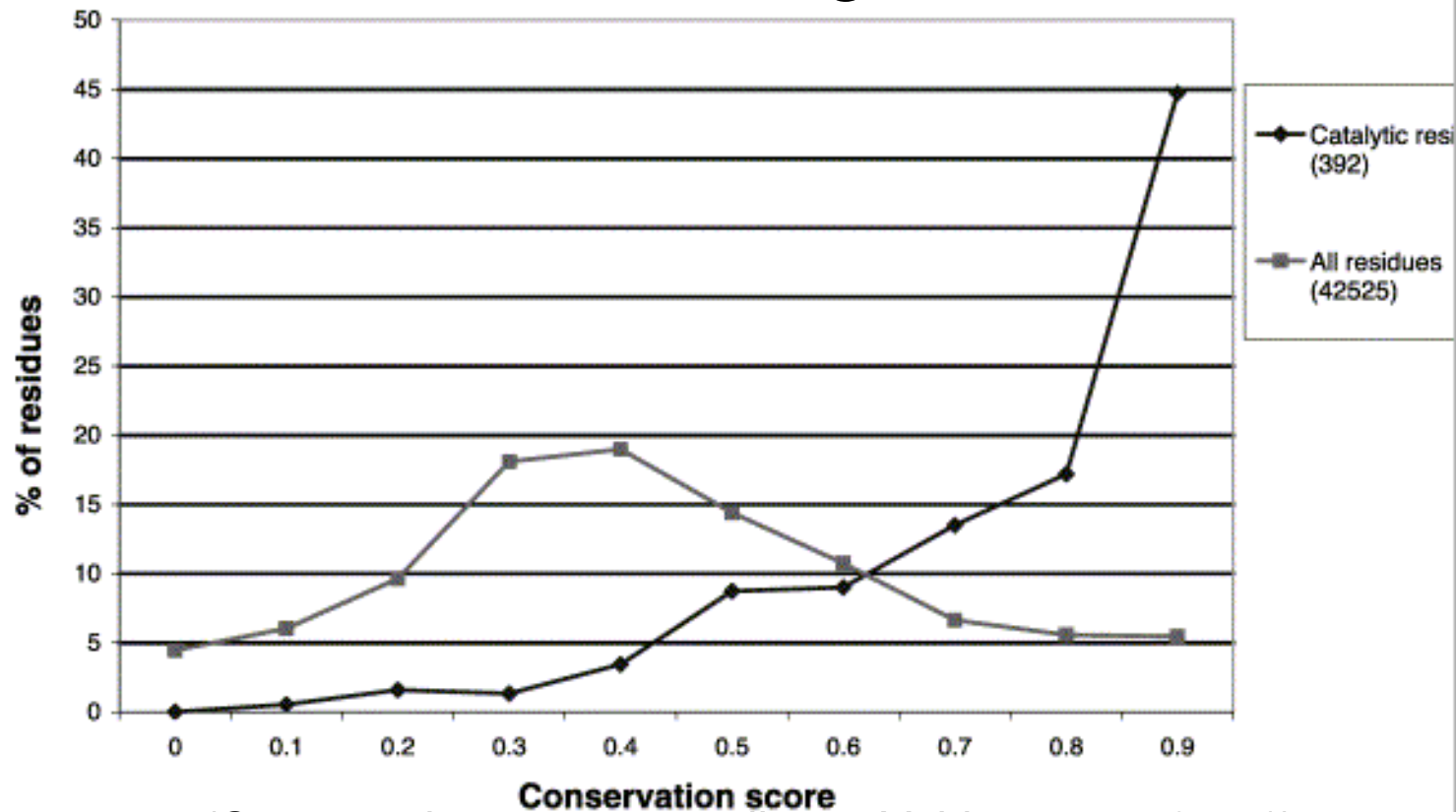


Figure 10. Normalised B -factors for individual catalytic residues (Arg, Asp, Cys, Glu, His and Lys) compared with normalised B -factors for all residues of the same type.

Table 6. Catalytic residue hydrogen bonds

Residue (number analysed)	Number making ≥ 1 H-bond						Total
	<i>Via</i> -N-H or -C=O group			<i>Via</i> side-chain atoms			
	To protein	To ligand	Total	To protein	To ligand	Total	
His (107)	86	1	87	81 (96%)	20 (24%)	84	100
Asp (93)	74	6	77	73 (96%)	10 (13%)	76	90
Arg (67)	55	3	55	54 (92%)	30 (51%)	59	65
Glu (65)	59	1	59	44 (94%)	6 (12%)	47	62
Lys (55)	49	3	50	40 (89%)	17 (38%)	45	54
Cys (38)	32	3	32	16	2	16	33
Tyr (32)	27	2	27	21	5	22	31
Asn (26)	18	1	19	20	3	20	23
Ser (26)	22	5	25	20	4	21	26
Gly (24)	10	7	14	0	0	0	14
Thr (18)	16	3	16	13	3	14	18
Gln (14)	13	1	13	11	5	13	14
Trp (9)	8	0	8	5	0	5	8
Phe (7)	5	2	6	–	–	–	6
Leu (7)	6	1	6	–	–	–	6
Met (4)	2	0	2	–	–	–	2
Ala (1)	1	0	1	–	–	–	1
Ile (2)	2	0	2	–	–	–	2
Pro (2)	1	0	1	–	–	–	1
Val (1)	1	0	1	–	–	–	1
All (598)	487	39	501	498	105	422	557 (93%)
M/C ^a (48)	28	14	34	6	3	7	34 (71%)

Catalytic residues are more conserved in general



(Conservation score according to Valdar; range 0 to 1)

(Note: this analysis could have been more refined/advanced/thorough in various aspects)

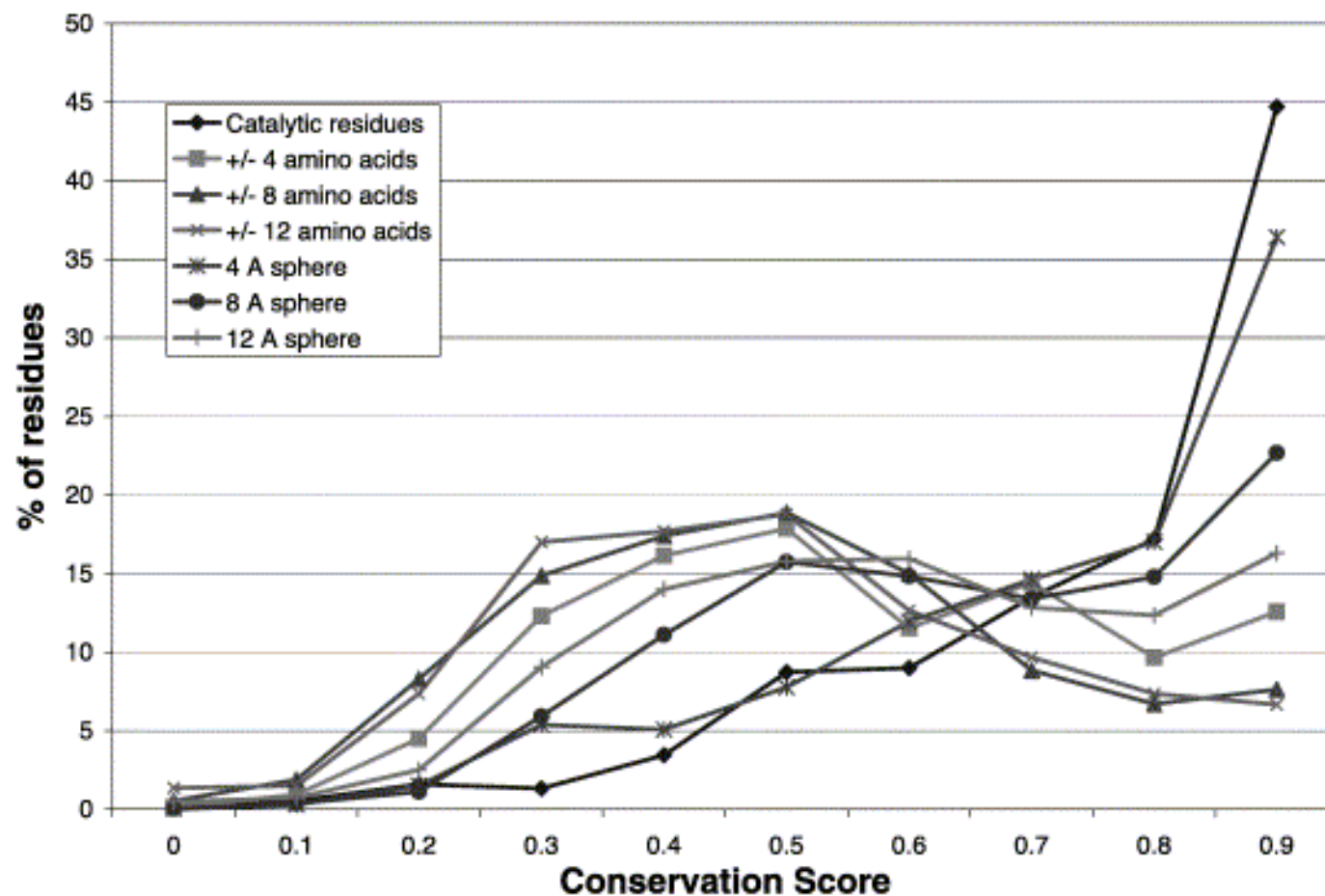


Figure 11. Residue conservation scores. (a) Catalytic residue conservation scores compared with conservation scores for all residues in the dataset. The conservation score ranges from 0 (least conserved) to 1 (most conserved). (b) Conservation scores in sequence and structural locality. The centre of gravity of the catalytic residues in each enzyme was calculated and the conservation score of any residue falling within a sphere of 4 Å, 8 Å, and 12 Å of the centre of gravity was recorded. Additionally the conservation scores of residues at sequence positions ± 4 , 8 and 12 amino acid residues from each catalytic residue were recorded.

Table 2. Functional classification of catalytic residues

Catalytic function	Description
Acid–base	Involved in proton abstraction, donation or both, to or from a substrate, as a direct part of the catalytic mechanism. Excludes residues which affect other residues or water molecules in this manner
Nucleophile ^a	Forms a covalent intermediate with the substrate <i>via</i> nucleophilic attack
Transition state stabiliser	Stabilises the transition state in some way (e.g. by stabilising an oxyanion hole formed during ester hydrolysis), lowering the activation energy of the reaction
Activate water	Alter the pK_a of or deprotonate a water molecule which is directly involved in the reaction
Activate cofactor	Exerts a favourable effect on a cofactor (could be metal or minor substrate such as FAD) through various means (e.g. by altering redox potential, or increasing effective charge)
Primer	Exerts a favourable effect on another residue directly involved in the catalytic mechanism, e.g. by acting as an acid or base, or through electrostatic effects
Activate substrate	Exerts a favourable effect on the substrate (e.g. by polarising a bond to be broken)
Radical	Forms a radical which is involved in the catalytic reaction
Modified	Modified in some way in order to perform catalysis during the reaction, e.g. carbamylated lysine residue

Table 7. Catalytic residue functions

Residue (total)	Acid/base	Nucleophile	Transition state stabiliser	Activates wa- ter/cofactor/residue	Activates substrate	Other (radical/modified)
Histidine (113)	58	4	18	37	13	1
Aspartate (92)	31	6	10	45	6	–
Arginine (68)	6	–	51	9	5	–
Glutamate (67)	30	–	7	31	5	–
Lysine (56)	13	1	24	11	9	4
Cysteine (39)	6	21	2	5	1	7
Tyrosine (34)	17	1	7	4	5	2
Asparagine (28)	1	–	19	6	5	–
Serine (27)	6	9	4	9	1	–
Glycine (24)	–	–	19	3	1	1
Threonine (18)	3	1	10	4	–	–
Glutamine (15)	–	–	7	5	3	–
Tryptophan (9)	–	–	3	3	–	3
Phenylalanine (7)	–	–	5	1	1	–
Leucine (7)	–	–	3	2	2	–
Methionine (4)	–	–	2	1	1	–
Alanine (2)	–	–	2	–	–	–
Isoleucine (2)	–	–	–	–	2	–
Proline (2)	1	–	–	–	1	–
Valine (1)	–	–	–	1	–	–
All (615)	172 (28%)	44 (7.2%)	193 (31%)	178 (29%)	60 (9.8%)	17 (2.7%)
No. of enzymes with at least one residue performing this function	106	44	96	104	41	11
Variation in no. of residues performing this function in any one enzyme	1–7	1	1–6	1–6	1–6	1–2