OPEN ACCESS PEER REVIEW
DOIttt:https://doi.org/10.52756/lbsopf.2024.e03.003
Chapter- 3

Discovery of Rim Region Between Core and Surface of Proteins

Amal Kumar Bandyopadhyay[1#]*, Sahini Banerjee[2#] and Somnath Das[3#]

Keywords: Rim-region, topological pattern, interior cavity, non-bonded interaction, protein compaction

Abstract:

The crystal structure reveals the complexities of protein component organization from the core to the surface. In general, the existence of core and surface in protein structure is well known. Here, we raise the question of how this core and surface, respectively, end and begin. To further understand the concerns, we did a comprehensive investigation on high-resolution structures of three protein families [ADH (Alcohol Dehydrogenase), Glyceraldehyde 3-Phosphate Dehydrogenase (GDH), and Malate Dehydrogenase (MDH)] from various domains of life utilizing authentic methods. The results demonstrate the presence of a zone (designated as the rim), which has separate properties from the core and the surface. The Kyte Doolittle grand average hydrophobicity (KD) of the core, surface, and rim are positive, negative, and neutral, respectively. Compared to the rim-zone, the core and surface have more hydrophobic residues and beta-strand structure, and charged residues and coil structure, respectively. In terms of polar residues and helix structure, compared to the rim, the core and surface have essentially similar but lower contents. The core's long β-sheets and shell-water-filled cavities may restrict residue compaction in this location. These analyses, thus, demonstrate that the archaea employed a distinct approach to fine-tune the compaction of their core compared to the bacteria and eukaryotes. Simultaneously, the dominance of the coil at the surface appears to produce a similar result. Overall, our study provides evidence that the rim-zone has a very distinct structure from the core and surface. The study that applies to other proteins finds application in protein structure bioinformatics.

Introduction:

The crystal structure reveals how the codes in a protein's primary sequence are expressed in the functional state of the structure. It is really interesting to know the pattern that may exist in such a complex structure (Kendrew et al., 1958; Dill and MacCallum, 2012; Rovers& Schapira, 2022). Orthologous proteins from Archaea (ar), Bacteria (ba), and Eukaryotes (eu) have comparable topologies and biological functions despite differing in basic sequences (Bandyopadhyay et al., 2007; Bottini et al., 2013; Bandyopadhyay et al., 2019; Flower et al., 2000; Petitjean et al., 2015). In true solvent conditions, the primary sequence generates the tertiary structure through weak forces and intermediate secondary structures (Anfinsen, 1973; Dill, 1990; Bandyopadhyay et al., 2020). In terms of protein structure, amino acid residues are preferentially arranged from the core to the surface. The core and surface of a globular protein are typically enriched with hydrophobic (A, V, F, I, L, M, C) and hydrophilic residues,

Amal Kumar Bandyopadhyay
Department of Biotechnology, The University of Burdwan, Burdwan, West Bengal, India
E-mail: akbanerjee@biotech.buruniv.ac.in
Sahini Banerjee
Department of Biological Sciences, Indian Statistical Institute, Kolkata, West Bengal, India
Somnath Das
Department of Education, CDOE, The University of Burdwan, West Bengal, India
#Authors equally contributed
***Corresponding Author**: akbanerjee@biotech.buruniv.ac.in

Dr. Somnath Das, Dr. Jayanta Kumar Das, Dr. Mayur Doke, Dr. Vincent Avecilla (eds.), Life as Basic Science: An Overview and Prospects for the Future Volume: 3. ISBN: 978-81-978955-7-9;
pp. 41-96; Published online: 30thNovember, 2024

respectively (Lim, 1974; Rose et al., 1985; Stollar& Smith, 2020). Hydrophilic amino acids are classified into two types: ionizable (D, E, H, R, K) and non-ionizable (S, T, N, Q, Y, W) (Betts & Russell, 2003). Ser and charged residues are particularly important in determining protein solubility (Trevino et al., 2007). These hydrophobic and charged residues, in particular, tend to form helix and strand configurations (Williams et al., 1987). Aside from amino acid residues, SW plays an essential role in globular protein stability and specificity (Finney, 1987). Although there is no SW in the globular protein's core, SWs can be found in the inner cavity, which is generated by the interaction of several protein atoms (Ernst et al., 1995). These SWs are found inside and on the cavity's entry face, but the first type has a significantly larger interaction multiplicity (Bandyopadhyay et al., 2019). It would be interesting to learn more about the atomic composition, SW content, and region-specific distribution of these cavities. SW and protein atoms interact by non-covalent interactions such as hydrogen bonding, electrostatic (ion-pair, cation-π, anion-π), and hydrophobic (alkyl-alkyl, π-π, π-σ, etc.) to contribute to stability (Baker & Hubbard, 1984; Mecozzi et al., 1996; Martinez & Iverson, 2012; Islam et al., 2019; Surti et al., 2020). There were structural differences between the protein's core and surface. This region of accessibility is known as the in-between or rim (Lins et al., 2003; Sen Gupta et al., 2017). Alcohol dehydrogenase (ADH), glyceraldehyde-3-phosphate dehydrogenase (GDH), and malate dehydrogenase (MDH) are oxydoreductases that use NAD+ as a cofactor to transform an alcoholic substrate into an oxidized product (Williamson & Paquin, 1987; Soukri et al., 1989; Song et al., 1990; Gutheil et al., 1992; Madern, 2002; Ceccarelli et al., 2004; Xiao et al., 2023). All of these enzyme types create multimeric structural complexes to enhance NAD+ and substrate binding at the interface (Song et al., 1999; Dalhus et al., 2002; Moon et al., 2011). In the case of these three protein families, the structural database is extensive. There is little known about the rim region's structure and interactions. A comparative analysis of the distributions and non-bonded interactions of amino acid residues, residue classes, secondary structures, cavities, and SWs in these three accessible zones appears to be useful.

In the current study, we addressed and provided evidence on the aforementioned concerns using representative crystal structures of three protein families from each of the three domains of life. We have done this analysis by using our and others' authentic automated procedures. Our study, based on extensive studies of three protein structures, demonstrates that the rim region exists between the core and the surface. We hypothesize that this rim region acts as a container and carrier of protein structure and function. We believe that this rim effect is present in nearly all structures. Tour study finds application in the field of protein bioinformatics.

Materials and Methods

Retrieval of Dataset

The 3D structures (Table 1) of ADH, GDH, and MDH of different domains of life (ar, ba, and eu) were procured from the Research Collaboratory for Structural Bioinformatics (RCSB) protein data bank (PDB) (Berman et al., 2000). Each of the structures was minimized using Autominv1.0 (Islam et al., 2018) in the presence of SW interacting with protein atoms within

4.0Å distance as in our earlier studies (Nayek et al., 2014; Biswas et al., 2020; Banerjee et al., 2021; Roy et al., 2024).

Table 1. Details on protein structures used in the study.

ADH			GDH			MDH		
ar	**ba**	**eu**	**ar**	**ba**	**eu**	**ar**	**ba**	**eu**
1h2b [16]	1rjw [15]	1adb [17]	1b7g [16]	1euh [28]	1dss [23]	1o6z [13]	1b8p [22]	1civ [30]
1jvb [15]	3ox4 [23]	1b15 [17]	1cf2 [9]	1gad [19]	1j0x [21]	1v9n [20]	1bdm [21]	1mld [18]
1rhc [22]	3uog [16]	1cdo [18]	1uxt [39]	1gd1 [19]	1k3t [22]	2d4a [17]	1emd [26]	1sev [18]
2eer [13]	4cpd [23]	1d1t [16]	2czc [12]	1obf [17]	1vsu [26]	2x0i [16]	1guz [6]	2dfd [18]
2h6e [16]	4eez [15]	1ee2 [17]	2yyy [28]	2d2i [14]	1ywg [18]	4bgu [18]	1gv1 [15]	2fn7 [21]
4jbg [15]	4gkv [10]	1ht0 [21]		2g82 [11]	2i5p [22]	4bgv [8]	1ur5 [17]	2g76 [9]
6c75 [32]	4j6f [18]	1mc5 [14]		3gnq [19]	2vyn [20]	6ihd [23]	1z2i [24]	2hjr [11]
	4z6k [13]	1mg5 [17]		4dib [17]	3cps [23]		3d5t [21]	2i6t [17]
	5yln [11]	1u3t [14]		5ld5 [19]	3pym [24]		3fi9 [16]	3i0p [20]
		3wle [26]		5utm [18]	3qv1 [18]		3flk [18]	4h7p [15]
		4jji [19]		6fzh [18]	3sth [24]		3gvh [19]	4mdh [25]
		4rqt [16]			4k9d [19]		3nep [7]	4plh [14]
		4w6z [20]			4lsm [14]		3p7m [12]	4plt [20]
		5ilg [12]			5c7i [18]		3tl2 [16]	4uuo [19]
					5tso [15]		4e0b [19]	5nue [19]
							4ror [13]	5zi2 [17]
							4tvo [15]	6um4 [11]
							5ujk [20]	7mdh [21]
							6aoo [19]	
							6bal	

							[20]	
							6itk [19]	
							6itl [14]	

Protein data bank files of Alcohol Dehydrogenase (ADH), Glyceraldehyde 3-Phosphate Dehydrogenase (GDH), and Malate Dehydrogenase (MDH) from archaea (ar), bacteria (ba) and eukaryote (eu) domains. Only the first chain was processed in each case. Each of these structures was minimized in the presence of shell-water interacting with protein atoms within 4.0Å distance. The value of the third bracket indicates the number of cavities in this protein.

Protein residue accessibility (ASA) and accessibility regions

We used the novel Surface Racer v5.0 program (Tsodikov et al., 2002; Banerjee et al., 2018) to extract the accessibility and cavity details of the protein atom. Absolute accessibility of protein atoms has been converted into relative accessibility of residue. In this case, N, CA, C, and O are considered as the main-chain and the rest of the atoms of a residue are taken as side-chain.The relative accessibility of the residue's main-chain, side-chain, and all-atoms was determined by the absolute value of their folded (Tsodikov et al., 2002) and unfolded (Zielenkiewicz & Saenger, 1992) states as earlier (Bandyopadhyay et al., 2019). Based on overall KD value, the accessibility of proteins was divided into three regions namely core (ASA≤15Å^2), rim (ASA>15 Å^2 and ≤35 Å^2), and surface (ASA>35 Å^2) (Sen Gupta et al., 2017; Banerjee et al., 2018).

Details on the residue, cavity, shell-water, and non-covalent interactions

Relative to these three regions (core, rim, and surface), domains of life (ar, ba, eu) specific, protein-family (ADH, GDH, MDH) specific structures were compared to the average properties of protein residue composition, residue-class composition, Kyte-Doolittle hydrophobicity (Kyte & Doolittle, 1982; Islam et al., 2018), secondary structure type (Helix/Strand/Coil) (Islam et al., 2018), cavity details (Sen Gupta et al., 2017), non-bonded interactions (Islam et al., 2019), SW and its interactions, etc. The composition, secondary-structure details, accessibility-details of cavity residues, and their association and interaction with SWs were also extracted for comparison. Residues, A, V, L, I, F, M, C, P, G, residues, S, T, N, Q, Y, W and residues, H, R, K, E, D were taken as hydrophobic, hydrophilic, and charged types, respectively (Betts & Russell, 2012). The secondary structure details of the amino acid residue were taken from the crystal structure.

Protein residue and shell-water interaction

Biovia Discovery Studio 2020 was used to determine all atomic level non-bonded interactions that include hydrogen bond, ion-pair, π-π, π-cation, π-anion, π-amide, π-sigma, π-alkyl, and alkyl-alkyl, etc. for the above-mentioned accessibility regions of a protein. These

interactions were processed using an automated AWK script to obtain different inter-residue interactions (Islam et al., 2018; Mitra et al., 2019).

Automated extraction of accessibility (ASA) region-specific details

Our database has 108 protein PDB structures in it. Thus, the total number of proteins to analyze is 324, as each protein has three accessibility (ASA) regions. We used a fully automated AWK script (that makes use of the Surface Racer program to determine accessibility regions and cavities) to properly and accurately extract the properties of all the items mentioned above (Nayek et al., 2014; Gupta et al., 2014a; Gupta et al., 2014b; Nayek et al., 2015a; Nayek et al., 2015b; Banerjee et al., 2015). The script runs in a Cygwin environment in a user-friendly and error-free manner. It takes all PDB structures as an input that is present in the working directory. Outputs are redirected in Excel format for the ease of post-run analyses. The binary version of the script can be obtained upon request.

Statistical Analysis

All statistical tests were performed using paired t-test if not stated otherwise. P-value, thus obtained, was used to judge against the null hypothesis at the 5% level using the standard two-tailed procedure. The *p*-value is included with the relevant data in the results (Banerjee et al., 2021; Roy et al., 2023).

Results

Residue, residue-class, and secondary structure distribution in co, rm, and su zones

The primary goal of our research is to understand the topological characteristics and patterns of enzymes as they work in various earths' conditions (Bandyopadhyay & Sonawat, 2000; Gupta et al., 2015). We investigated crystal structures from ADH, GDH, and MDH enzymes (Table 1). A typical crystal structure shows how the co, rm, and su residues and a typical cavity are arranged (Figure 1a). The properties of archaea's primary sequence differ from those of ba and eu (Figure 1b and Figure 2a-h). The overall KD (GRAVY) of co, rm and su are positive, neutral and negative, respectively (Figure 1c for MDH's ba and Figure 3a-c for ADH's ar, ba, and eu; d-f for GDH's ar, ba, and eu; g-h for MDH's ar and eu, respectively). Further, co and su regions are significantly abundant in hb and cr residue classes, respectively. Again, these two regions have in common the abundance of po class (Figure 1d, e). These observations of ba MDH's are also true in other domains of life specific enzyme classes (Figure 4a-c for ADH's ar, ba and eu; d-f for GDH's ar, ba and eu; and g-h for MDH'a ar and eu; and Figure 5a-c for ar ADH's hb, hl, and cr; d-f for ba ADH's hb, hl, and cr; g-i for eu ADH's hb, hl, and cr; j-l for archaeal GDH's hb, hl, and cr; Figure **6**a-c for ba GDH's hb, hl, and cr; d-f for eu GDH's hb, hl, and cr; g-i for archaeal MDH's hb, hl and cr and j-k for eu MDH's hb, hl, and cr, respectively). Notably, for rm-region, which is between the co and su, these classes of amino acids are negligible. Relative to this region, these features of the amino acid compositions and classes of the co and the su regions are statistically significant.

The secondary structure, which is between the primary and tertiary structures, determines the topology and conformation of enzyme classes (Bandyopadhyay et al., 2001). Here, we raise the question as to how these secondary structures are distributed in these regions (core, rim and surface). Here, we see that when normalized, and average helix-content is almost equivalent to co and su, and much higher than that of rm. The strand and coil, in the core and surface are uniquely, and significantly dominating over the rim region, respectively (Figure 1f for ba MDH's helix, strand and coil; and Figure 7a-c for archaeal ADH's helix, strand, and coil; d-f for ADH's ba helix, strand, and coil; g-i for ADH's eu helix, strand and coil; j-l for GDH's archaeal helix, strand, and coil; Figure 8a-c for GDH's ba helix, strand and coil; d-f for GDH's eu helix, strand, and coil; g-i for archaeal helix, strand and coil and j-l for MDH's eu helix, strand, and coil, respectively). Here, it is important to mention that although the accessibility region-specific differential pattern in amino acid compositions, residue classes, and secondary structures described above are true and statistically significant in all domains of life and enzyme classes, their normalized scales show domains of life specific and also enzyme class specific variation (Table 2 and 3). It should be noted here that although the rm's preference for the protein structure components is low, this region has a strong preference towards shell-water (see below).

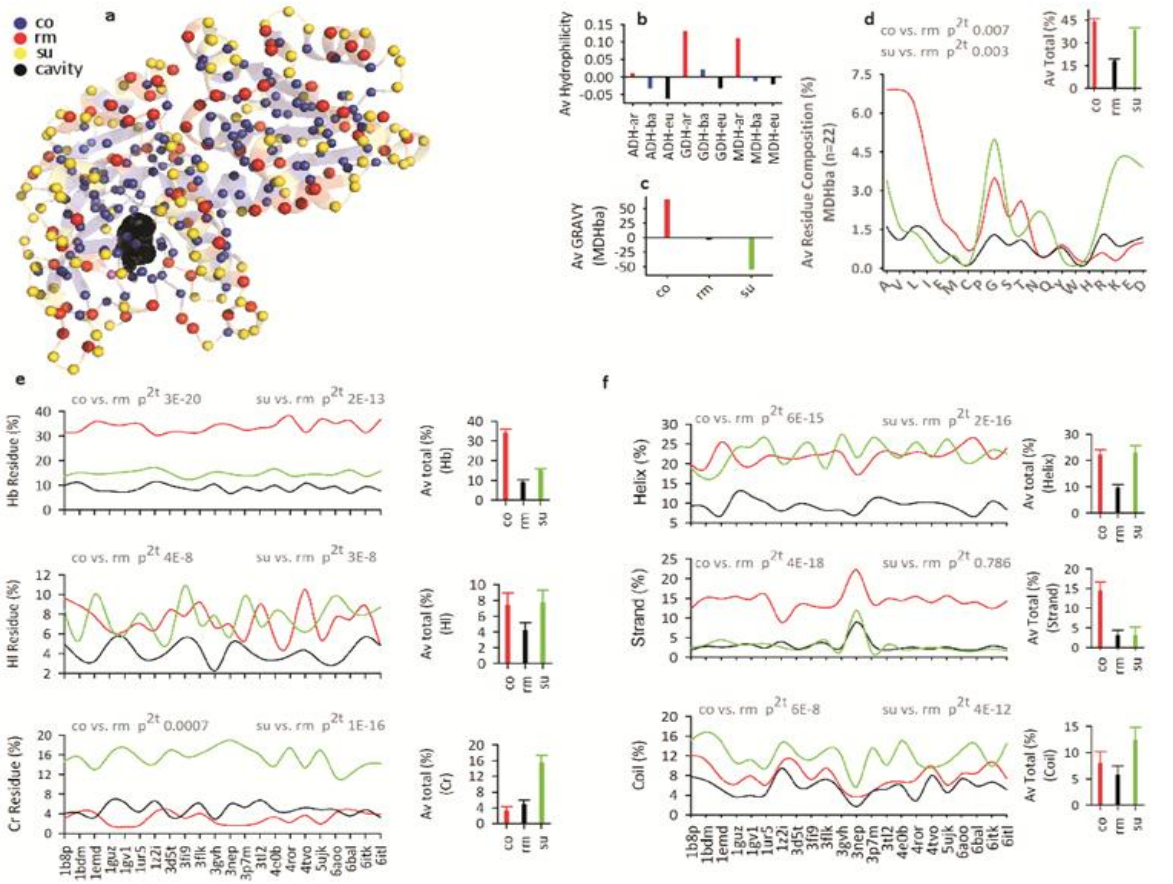


Figure 1. Details on residue composition, residue-class composition and secondary structure types.

Variation of amino acid composition, class and secondary structure in different accessibility regions along with sequence hydrophobicity relative to domains of life. **a**, A general presentation of a crystal structure where the residues (Cα) in co (blue), rm (red) and su (yellow) are shown with a typical cavity (black). **b**, Av KD of sequences in ar, ba, and eu. **c**, Av KD (n=22 PDBs) of ba MDH's co, rm and su regions. **d**, The average residue composition in the co (red), rm (black) and su (green) regions are shown with *p*-values (p^{2t}) for test of significance in co vs. rm, and su vs. rm format. **e**, The average residue class (hb, hl and cr) composition in the co (red), rm (black) and su (green) regions with *p*-values. **f**, The average secondary structure type in the co (red), rm (black) and su (green) regions with *p*-values. In each case, the total value of each of the three regions is shown as inset.

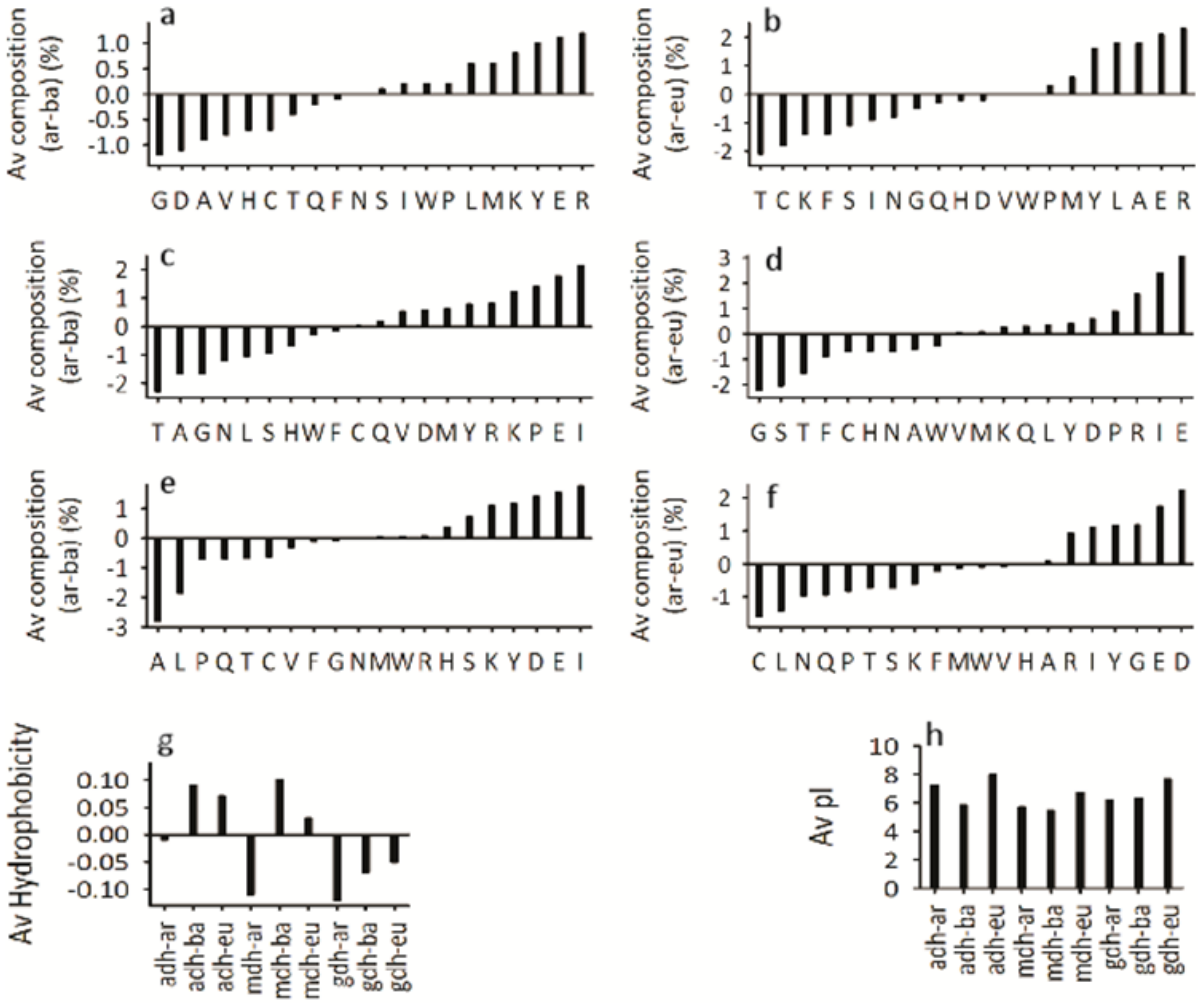


Figure 2. Relative residue compositions of archaeal proteins.

Average (av) normalized (%) relative composition of ar in reference to ba, and eu for ADH (**a** and **b**), GDH (**c** and **d**), and MDH (**e** and **f**), respectively. Grand average Kyte-Doolittle hydrophobicity of AD, GDH and MDH for ar ba and eu (**g**). Average pI value ADH, GDH and MDH for ar ba and eu (**h**).

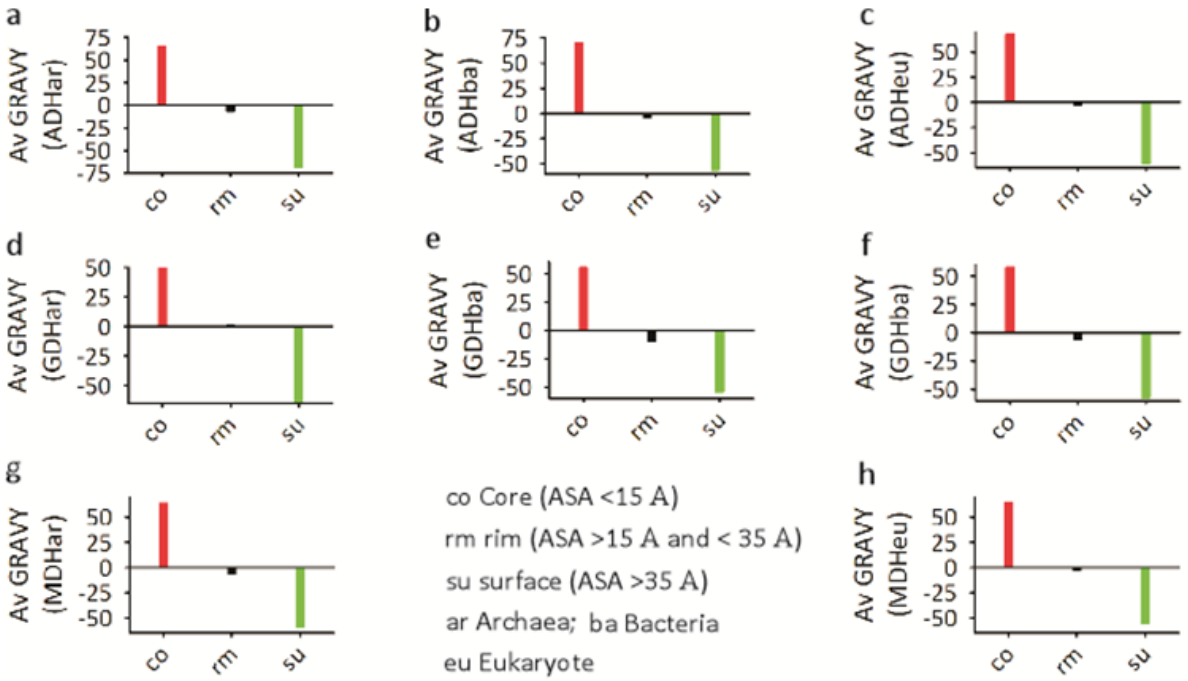


Figure 3. Details on ASA-region-specific GRAVY of proteins in our dataset.

Kyte-Doolittle overall hydrophobilicy (GRAVY) for the co, rm and su regions of ar (**a**), ba (**b**) and eu (**c**) enzyme, ADH. Similarly the GRAVY of co, rm and su regions of ar (**d**), ba (**e**) and eu (**f**) enzyme, GDH and ar (**g**), and eu (**h**) enzyme, MDH.

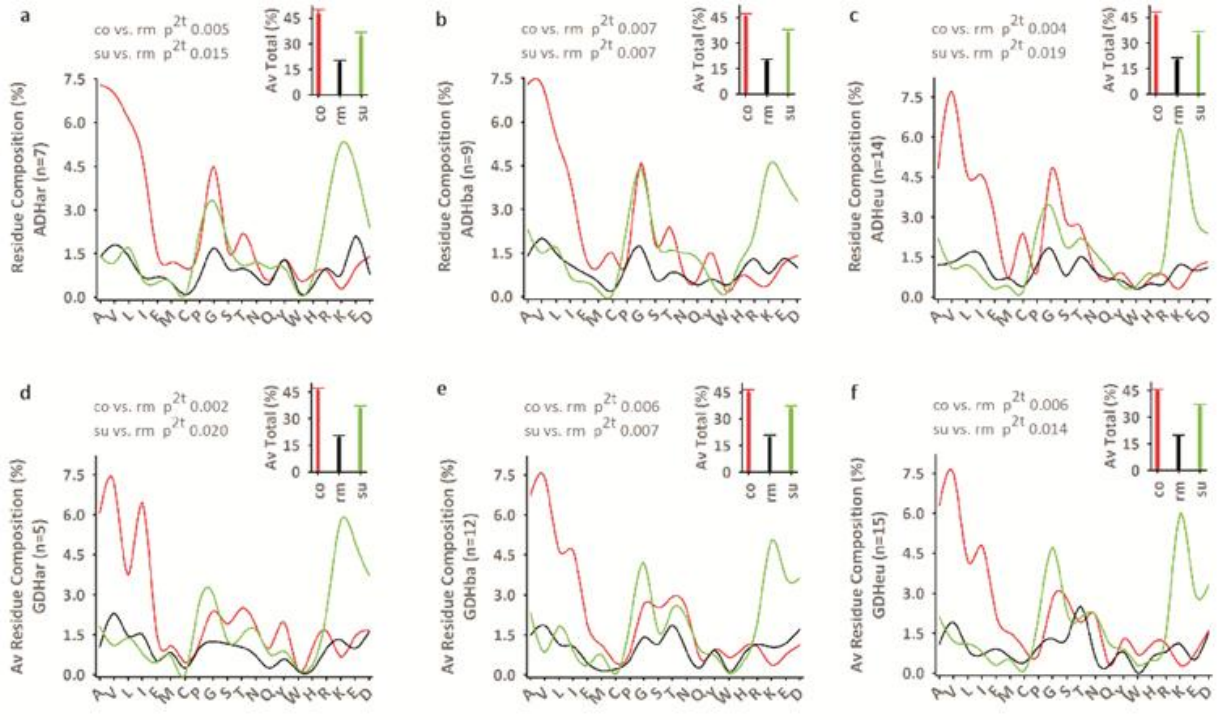
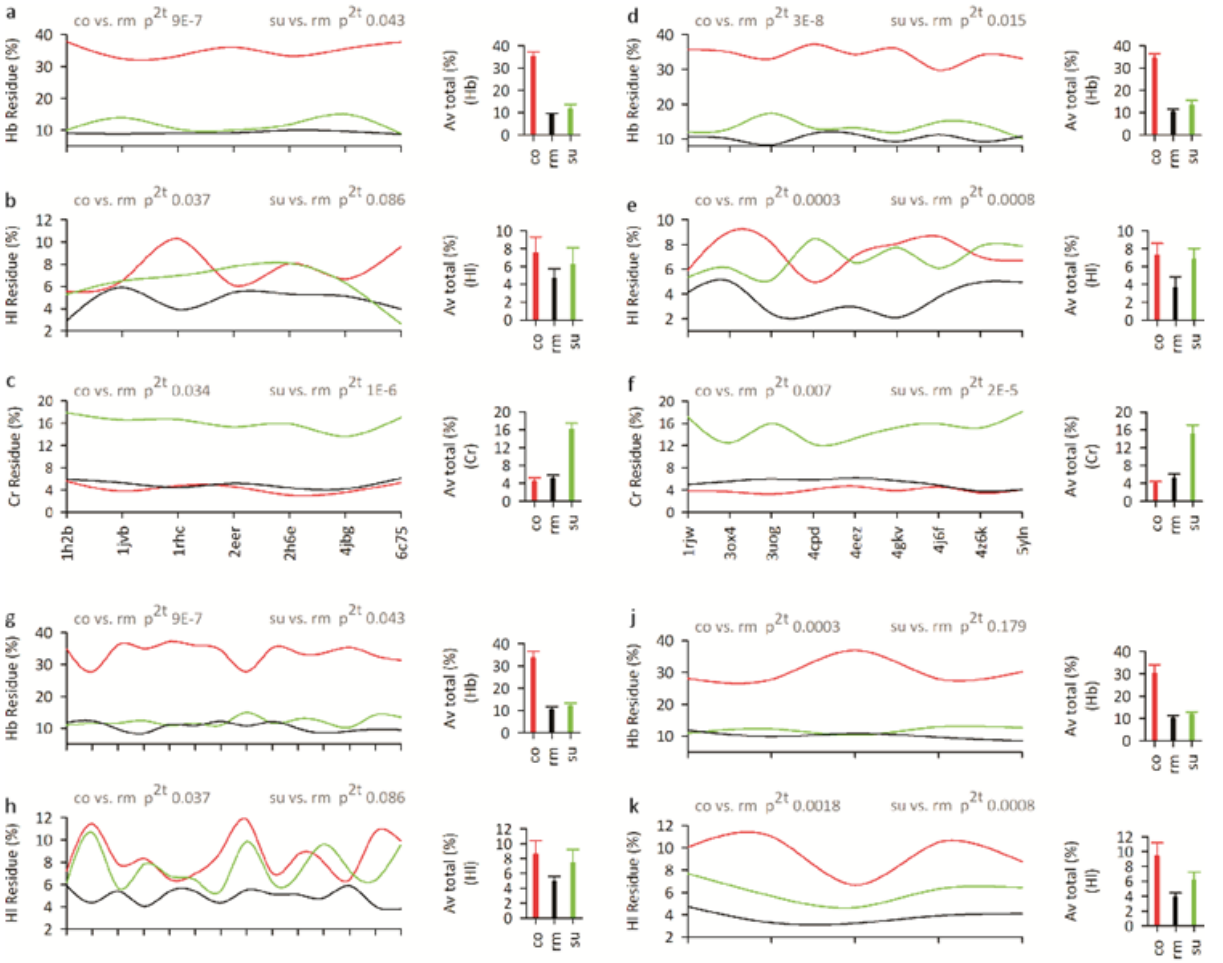


Figure 4. ASA-region-specific normalized average residue compositions of proteins of our dataset.

Average (av) normalized (%) amino acid residue composition of core (co, red), rim (rm, black) and surface (su, green) for ADH (**a** for ar, **b** for ba and **c** for eu), GDH (**d** for ar, **e** for ba and **f** for eu) and MDH (**g** for ar, **h** for eu). Individual and total values of these three regions are shown in each case. Significance of the difference of mean observation is tested using paired t-test using two tailed procedures.



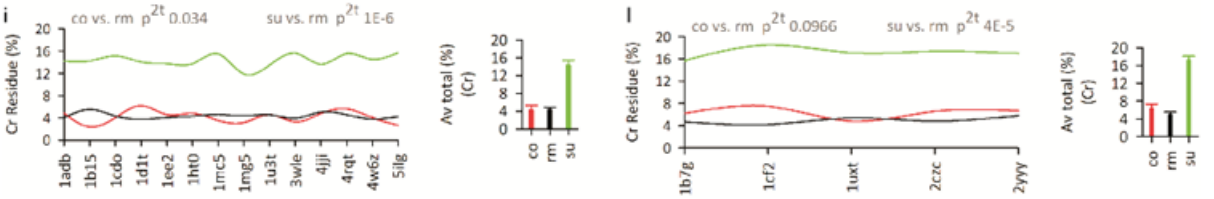


Figure 5. ASA-region-specific normalized average residue class compositions of proteins of our dataset.

Average (av) normalized (%) amino acid classes composition of core (co, red), rim (rm, black) and surface (su, green) regions of ar's ADH (**a,** hb; **b,** hl; **c,** cr), ba's ADH (**d,** hb; **e,** hl; **f,** cr), eu's ADH (**g,** hb; **h,** hl; **i,** cr) and ar's GDH (**j,** hb; **k,** hl; **l,** cr). The region-specific average plots of all the proteins are shown as bars on the side of each protein-specific profile plot. The significance of the difference of mean observation is tested using a paired t-test using a tailed procedure.

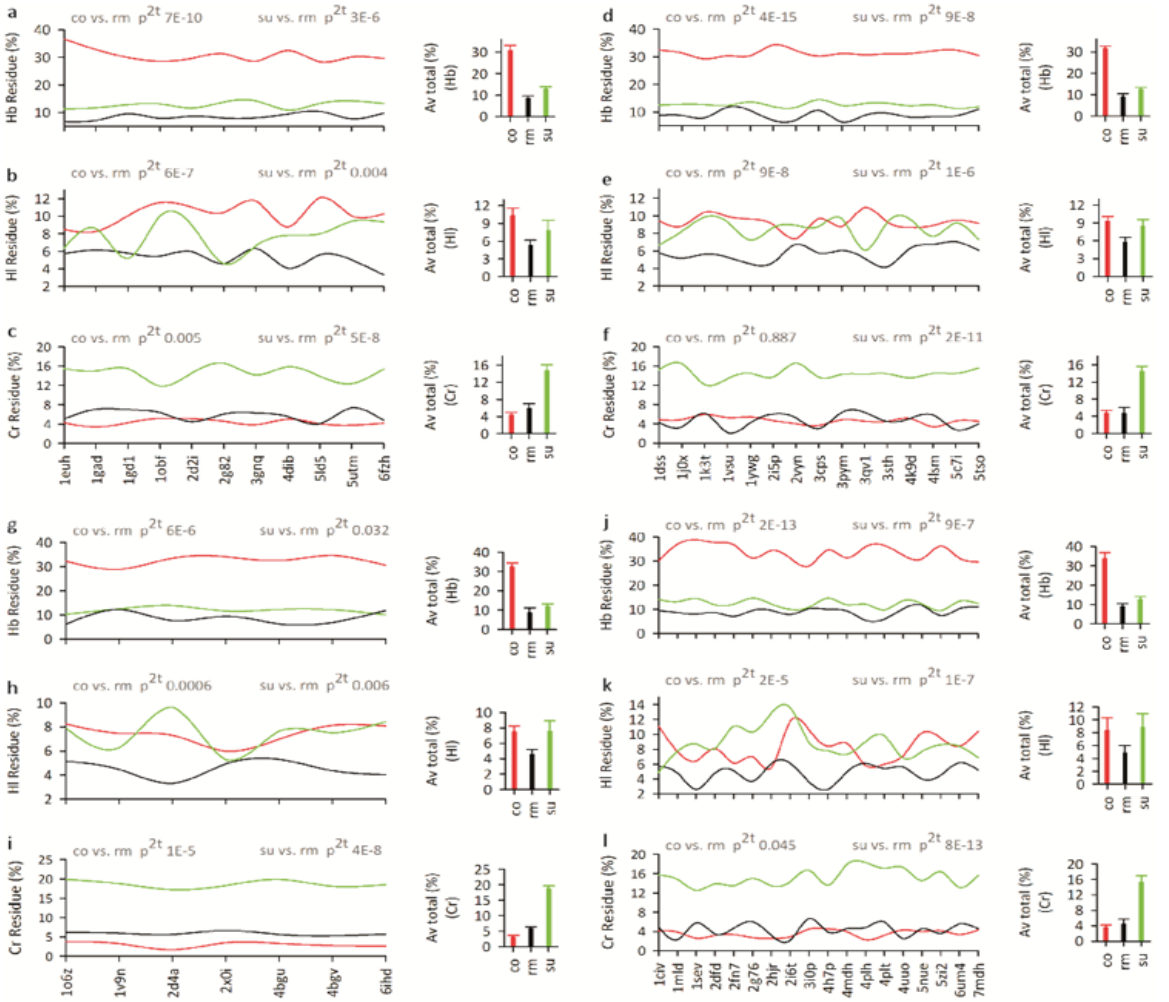


Figure 6. ASA-region-specific normalized average residue class compositions of proteins of our dataset (continued).

Average (av) normalized (%) amino acid classes composition of core (co, red), rim (rm, black) and surface (su, green) regions of ba's GDH (**a,** hb; **b,** hl; **c,** cr), eu's GDH (**d,** hb; **e,** hl; **f,**

cr), ar's MDH (**g,** hb; **h,** hl; **i,** cr) and eu's MDH (**j,** hb; **k,** hl; **l,** cr). The region-specific average plots of all the proteins are shown as bars on the side of each protein-specific profile plot. The significance of the difference of mean observation is tested using a paired t-test using a tailed procedure.

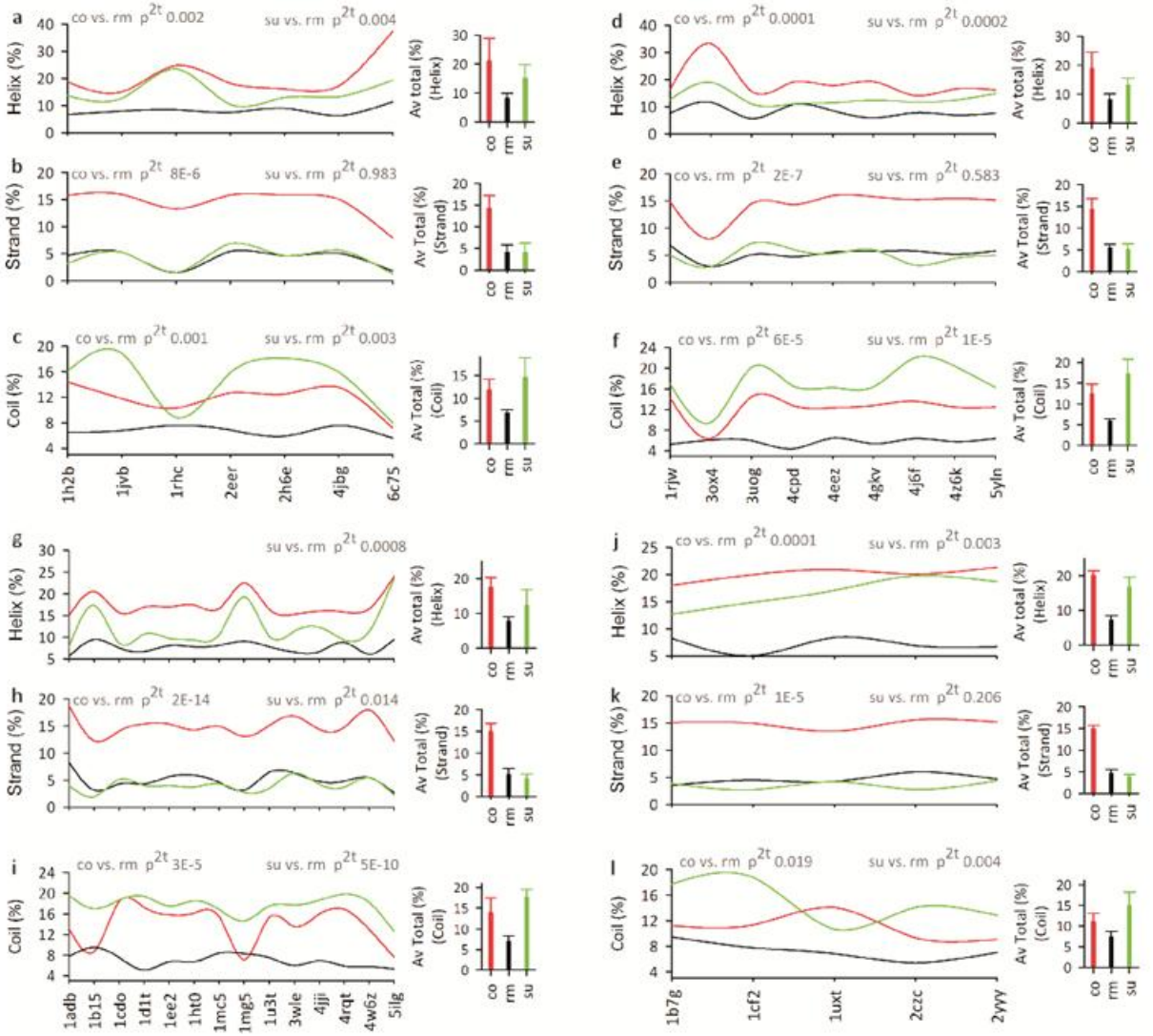


Figure 7. ASA-region-specific normalized average secondary structure types of proteins of our dataset.

Average (av) normalized (%) secondary structure types of core (co, red), rim (rm, black) and surface (su, green) regions of ar's ADH (**a,** hb; **b,** hl; **c,** cr), ba's ADH (**d,** hb; **e,** hl; **f,** cr), eu's ADH (**g,** hb; **h,** hl; **i,** cr) and ar's GDH (**j,** hb; **k,** hl; **l,** cr). The region-specific average plots of all the proteins are shown as bars on the side of each protein-specific profile plot. The significance of the difference of mean observation is tested using a paired t-test using a tailed procedure.

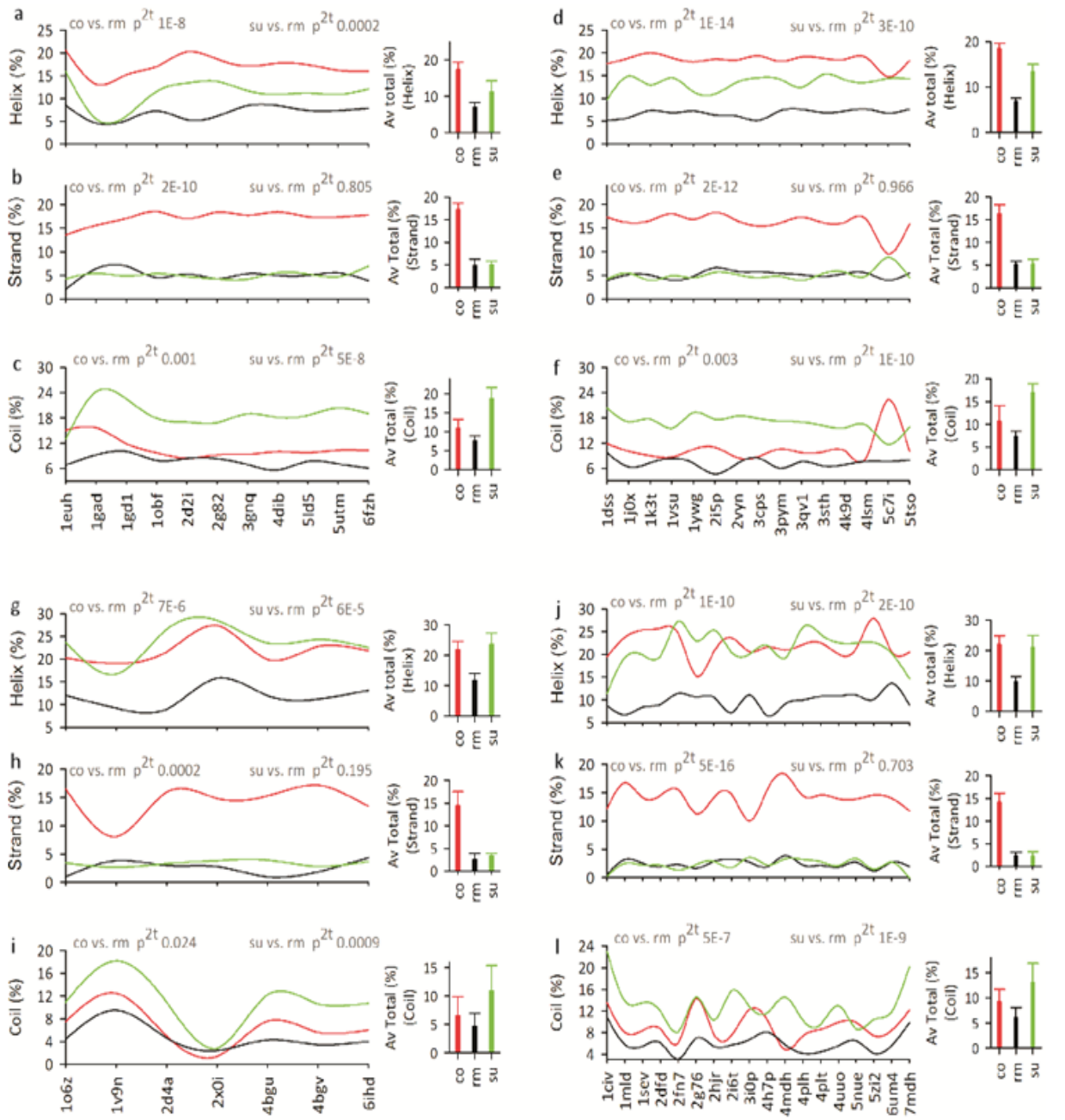


Figure 8. ASA-region-specific normalized average secondary structure types of proteins of our dataset (continued).

Average (av) normalized (%) secondary structure types of cores (co, red), rim (rm, black) and surface (su, green) regions of ba's GDH (**a,** hb; **b,** hl; **c,** cr), eu's GDH (**d,** hb; **e,** hl; **f,** cr), ar's MDH (**g,** hb; **h,** hl; **i,** cr) and eu's MDH (**j,** hb; **k,** hl; **l,** cr). The region specific average plots of all the proteins are shown as bars on the side of each protein-specific profile plot. The significance of the difference of mean observation is tested using a paired t-test using a tailed procedure.

Table 2. Domain-specific, enzyme-class-specific and ASA-specific Normalized (%) residue class.

	cohb	rmhb	suhb	cohl	rmhl	suhl	cocr	rmcr	sucr
MDH-ar	32.4	8.6	11.9	7.5	4.5	7.5	3.0	5.9	18.7
MDH-ba	33.9	8.9	14.8	7.3	4.2	7.7	3.1	4.8	15.4
MDH-eu	33.6	8.9	12.4	8.3	4.8	8.8	3.5	4.4	15.2
ADH-ar	35.2	9.2	11.5	7.5	4.7	6.2	4.4	5.1	16.1
ADH-ba	34.3	10.3	13.4	7.3	3.6	6.8	4.0	5.2	15.1
ADH-eu	33.7	10.4	12.1	8.6	4.9	7.5	4.2	4.4	14.3
GDH-ar	30.2	10.1	11.8	9.4	3.8	6.1	6.4	5.0	17.1
GDH-ba	30.7	8.4	12.7	10.3	5.3	7.7	4.4	5.9	14.6
GDH-eu	31.4	8.9	12.6	9.3	5.7	8.4	4.7	4.6	14.4

Domains of life specific, enzyme class specific and accessibility region (co, rm and su) specific percentage of residue classes such as hydrophobic (hb), hydrophilic (hl) and charged (cr).

Table 3. Domain-specific, enzyme-class-specific and ASA-specific Normalized (%) secondary structure types.

	co-h%	rm-h%	su-h%	co-s%	rm-s%	su-s%	co-c%	rm-c%	su-c%
MDH-ar	21.9	11.8	23.7	14.5	2.6	3.4	6.5	4.7	11.0
MDH-ba	22.0	9.3	22.6	14.4	2.9	3.0	7.9	5.6	12.2
MDH-eu	22.1	9.7	21.2	14.1	2.3	2.3	9.2	6.1	13.0
ADH-ar	21.1	8.2	15.2	14.3	4.1	4.1	11.8	6.7	14.5
ADH-ba	18.8	8.1	13.0	14.4	5.3	5.0	12.4	5.8	17.2
ADH-eu	17.6	7.7	12.3	14.9	5.0	4.0	13.9	7.0	17.6
GDH-ar	20.1	7.1	16.7	14.9	4.6	3.6	11.0	7.3	14.8
GDH-ba	17.3	7.0	11.2	17.2	5.0	5.1	10.9	7.6	18.8
GDH-eu	18.5	6.7	13.4	16.2	5.1	5.1	10.7	7.3	16.9

Domains of life specific, enzyme class specific and accessibility region (co, rm and su) specific percentage of secondary structure types such as helix (h), strand (s) and coil (c)

Table 4. Protein-specific shell-water-specific cavity details.

ADH_ar	Prot res	*fCv*	*fCv woW*	ADH_ba	Prot res	*fCv*	*fCv woW*	ADH_eu	Prot res	*fCv*	*fCv woW*
1h2b	341	16	1	1rjw	338	15	7	1adb	373	17	7
1jvb	338	15	6	3ox4	377	23	8	1b15	253	17	10
1rhc	330	22	4	3uog	332	16	4	1cdo	370	18	5
2eer	346	13	4	4cpd	343	23	8	1d1t	371	16	7
2h6e	321	16	6	4eez	338	15	2	1ee2	371	17	2

4jbg	331	15	2	4gkv	335	10	1	1ht0	372	21	0
6c75	377	32	19	4j6f	346	18	5	1mc5	368	14	2
				4z6k	343	13	1	1mg5	253	17	2
				5yln	343	11	8	1u3t	371	14	3
								3wle	333	26	3
								4jji	375	19	0
								4rqt	373	16	3
								4w6z	345	20	7
								5ilg	262	12	3

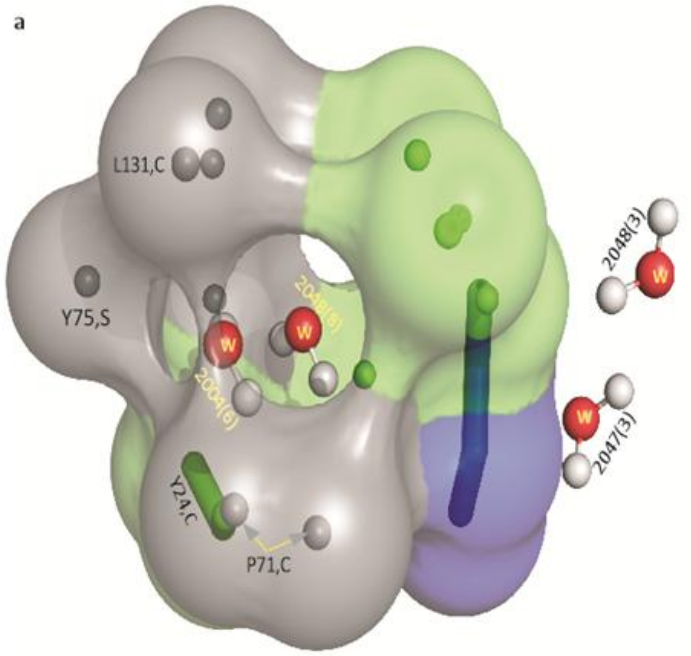
GDH_ar	Prot res	*fCv*	*fCv woW*	GDH_ba	Prot res	*fCv*	*fCv woW*	GDH_eu	Prot res	*fCv*	*fCv woW*
1b7g	338	16	6	1euh	470	28	6	1dss	329	23	6
1cf2	335	9	1	1gad	326	19	6	1j0x	328	21	12
1uxt	497	39	10	1gd1	328	19	1	1k3t	355	22	9
2czc	333	12	3	1obf	329	17	5	1vsu	322	26	10
2yyy	342	28	12	2d2i	335	14	3	1ywg	332	18	8
				2g82	326	11	3	2i5p	300	22	9
				3gnq	332	19	5	2vyn	325	20	6
				4dib	320	17	14	3cps	330	23	7
				5ld5	338	19	7	3pym	330	24	10
				5utm	339	18	3	3qv1	329	18	8
				6fzh	331	15	2	3sth	334	24	5
								4k9d	333	19	6
								4lsm	326	14	3
								5c7i	326	18	7
								5tso	328	15	4
MDH_ar	Prot res	*fCv*	*fCv woW*	MDH_ba	Prot res	*fCv*	*fCv woW*	MDH_eu	Prot res	*fCv*	*fCv woW*
1o6z	291	13	2	1b8p	322	22	5	1civ	369	30	7
1v9n	335	20	5	1bdm	316	21	5	1mld	311	18	5
2d4a	301	17	11	1emd	309	26	10	1sev	311	18	6
2x0i	284	16	12	1guz	304	6	1	2dfd	310	18	3
4bgu	301	18	4	1gv1	304	15	8	2fn7	296	21	9

4bgv	320	8	2	1ur5	297	17	2	2g76	301	9	4
6ihd	297	23	21	1z2i	348	24	7	2hjr	311	11	5
				3d5t	319	21	7	2i6t	279	17	8
				3fi9	320	16	4	3i0p	359	20	3
				3flk	346	18	6	4h7p	309	15	1
				3gvh	316	19	8	4mdh	329	25	5
				3nep	301	7	2	4plh	311	14	6
				3p7m	302	12	4	4plt	315	20	5
				3tl2	309	16	2	4uuo	323	19	16
				4e0b	299	19	5	5nue	324	19	3
				4ror	316	13	4	5zi2	336	17	3
				4tvo	323	15	0	6um4	321	11	0
				5ujk	312	20	7	7mdh	347	21	11
				6aoo	313	19	5				
				6bal	308	20	2				
				6itk	315	19	7				
				6itl	310	14	1				

Frequency of total water, total cavity and number of water-unfilled cavities in each representative protein in each of the three domains (ar, ba and eu) of each of the three classes of enzyme (ADH, GDH and MDH). Cv, cavity; *f*, frequency; wo, without; W, shell-water; freq, frequency.

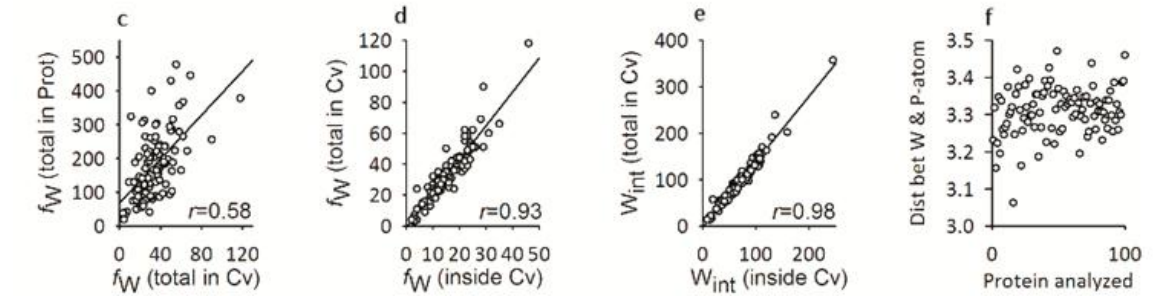
The interior cavity of the protein and its properties. (**a)** The typical cyrysy type cavity is constituted by atoms of co (grey), rm (green) and su (blue). It also has four moles of SWs. Cavity atoms in coil (C) and strand (S) are labelled and helix atoms are left unlabelled. (**b)** The grand average propensity of amino acid residues in the cavity. (**c**) Correlation between the total SW of the protein and the total SW of the cavity. (**d)** Correlation between the total SW and inside SW of the cavity. (**e)** Correlation between the total and inside interactions of the cavity atom. (**f**) Average distance of interaction of SW inside the cavity with its atoms. (**g)** Types of cavity and its frequency. (**h)** Cavity type specific water content. (**i**) Correlation between the cavity type specific frequency and their SW frequency. (**j)** Correlation between the cavity type specific frequency and their SW interaction frequency.

TW, total water in protein; Cv TW, cavity's total water; Cv Wins, cavity's inside water; Wint ins, water interaction with cavity atoms from inside; Wint T, water (inside and outside) interaction with cavity atoms; CvW:TW, total cavity water : total protein water in %; Cv_atom, protein specific total unique atoms that are present in all cavity; T_rs, total residue in protein; Cv_hT, helix% in cavity, Cv_sT, Strand% in cavity; Cv_cT, coil% in cavity; Cv_coT, cavity% in core region; Cv_rmT, cavity% in rim region; Cv_suT, cavity % in surface region; Dist, distance in Angstrom between the water and cavity atom interaction.



b

	ADH			GDH			MDH		
	ar	ba	eu	ar	ba	eu	ar	ba	eu
A	1.0	0.9	0.7	0.9	0.7	0.9	0.8	0.8	0.7
V	1.2	1.2	1.0	1.2	1.1	1.2	1.4	1.3	1.3
L	1.5	1.5	1.6	1.6	1.7	1.8	1.5	1.3	1.5
I	1.3	1.2	1.4	1.8	1.4	1.6	2.0	1.8	1.6
F	1.6	1.9	1.8	1.1	1.4	1.3	1.3	1.5	1.5
M	0.8	1.2	1.7	0.6	1.1	1.3	1.7	1.2	1.3
C	0.4	1.7	1.1	0.7	2.2	1.8	1.9	1.4	1.2
P	1.1	0.7	0.6	0.7	0.7	0.3	0.8	0.8	0.8
G	0.6	0.8	0.7	0.4	0.7	0.6	0.9	0.7	0.7
S	1.1	0.8	1.1	0.7	1.0	0.6	0.7	0.8	0.7
T	1.0	1.1	0.9	0.8	0.5	0.7	0.5	1.2	1.0
N	1.0	0.8	1.2	1.2	1.0	1.0	0.7	0.6	1.0
Q	1.4	0.9	1.1	0.8	1.2	1.4	0.5	0.7	0.6
Y	1.4	1.6	1.2	1.5	2.2	1.4	1.2	1.6	1.3
W	2.5	1.5	1.9	2.6	1.7	1.0	1.1	1.1	1.6
H	1.3	1.0	1.0	1.9	0.6	0.8	0.8	1.5	1.5
R	0.5	1.1	0.9	1.6	1.5	1.9	1.1	0.8	1.2
K	0.5	0.6	0.6	0.5	0.6	0.5	0.4	0.5	0.5
E	0.6	0.5	0.6	0.6	0.5	0.4	0.4	0.6	0.6
D	0.7	0.5	0.5	0.7	0.7	0.8	0.6	0.7	0.7



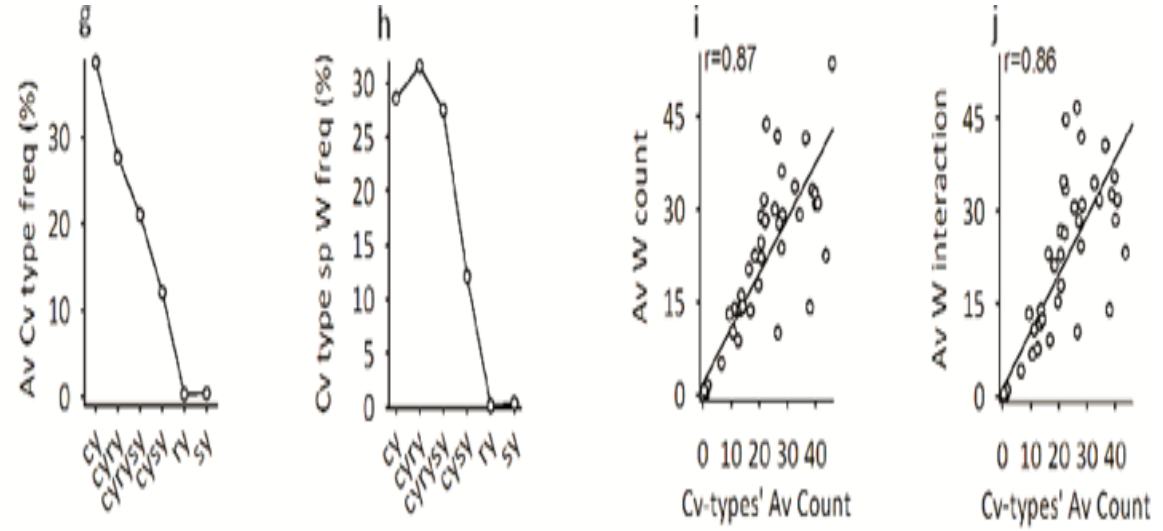


Figure 9. Cavity, its interaction and types.

Table 5. Domain-specific, enzyme-specific, cavity details.

Domain	protein	TW	Cv TW	Cv Wins	Wint ins	Wint T	CvW:TW	Cv_atom	T_rs	Cv_hT	Cv_sT	Cv_cT	Cv_coT	Cv_rmT	Cv_suT	Dist
ADH-ar	1h2b	286	51	26	108	147	17.8	131	341	44	22	65	92	19	20	3.23
	1jvb	187	27	10	33	56	14.4	154	338	50	62	42	108	27	19	3.32
	1rhc	143	37	19	90	115	25.9	155	330	74	31	50	102	33	20	3.16
	2eer	205	19	10	41	54	9.3	102	346	45	28	29	82	11	9	3.22
	2h6e	205	30	11	59	85	14.6	136	321	55	35	46	91	32	13	3.35
	4jbg	141	27	16	62	75	19.1	82	331	23	28	31	65	11	6	3.20
	6c75	98	25	16	86	97	25.5	221	377	168	20	33	174	24	23	3.34
GDH-ar	1b7g	265	25	8	34	53	9.4	108	338	36	36	36	79	11	18	3.26
	1cf2	231	25	12	46	65	10.8	81	335	44	14	23	70	4	7	3.25
	1uxt	256	90	29	136	239	35.2	314	497	109	79	126	263	37	14	3.27
	2czc	139	19	9	27	38	13.7	69	333	31	23	15	51	11	7	3.28
	2yyy	191	29	15	50	71	15.2	149	342	70	41	38	113	21	15	3.37
MDH-ar	1o6z	157	38	16	75	109	24.2	114	291	59	33	22	100	13	1	3.30
	1v9n	183	44	19	90	120	24	152	335	42	10	100	98	21	33	3.31
	2d4a	42	7	3	13	19	16.7	109	301	44	44	21	82	21	6	3.32
	2x0i	38	4	4	17	17	10.5	109	284	66	39	4	87	15	7	3.06
	4bgu	478	55	22	90	135	11.5	132	301	75	31	26	111	15	6	3.25
	4bgv	313	24	10	30	46	7.7	54	320	23	28	3	52	2		3.36
ADH-ba	1rjw	74	13	9	37	45	17.6	104	338	27	30	47	85	15	4	3.42
	3ox4	121	29	13	71	95	24	196	377	147	18	31	142	27	27	3.38
	3uog	85	33	16	83	111	38.8	162	332	55	55	52	141	10	11	3.31
	4cpd	60	20	10	37	55	33.3	159	343	76	34	49	132	14	13	3.16
	4eez	223	54	24	94	147	24.2	130	338	64	30	36	110	16	4	3.26
	4j6f	79	32	10	56	91	40.5	200	346	69	52	79	159	28	13	3.38
	4z6k	253	34	15	57	88	13.4	113	343	33	34	46	85	13	15	3.32
	5yln	32	5	3	15	17	15.6	77	343	17	34	26	54	17	6	3.28
GDH-ba	1euh	378	118	46	245	357	31.2	270	470	89	63	118	203	37	30	3.40
	1gad	132	27	12	62	81	20.5	149	326	42	57	50	104	27	18	3.30
	1gd1	170	37	15	72	110	21.8	156	328	51	50	55	98	38	20	3.30
	1obf	241	39	15	62	92	16.2	135	329	44	61	30	99	26	10	3.33
	2d2i	101	16	8	42	51	15.8	118	335	47	38	33	99	13	6	3.27
	2g82	170	25	13	62	76	14.7	82	326	46	26	10	55	11	16	3.35
	3gnq	130	35	16	70	105	26.9	146	332	53	47	46	98	30	18	3.35
	4dib	38	3	3	13	13	7.9	134	320	49	50	35	100	29	5	3.19
	5ld5	85	22	9	44	61	25.9	163	338	89	39	35	119	22	22	3.31
	5utm	235	38	20	87	109	16.2	138	339	66	32	40	95	26	17	3.27

	6fzh	430	50	15	74	118	11.6	111	331	51	28	32	76	20	15	3.35
MDH-ba	1b8p	197	47	23	102	132	23.9	150	322	59	31	60	103	29	18	3.36
	1bdm	89	35	21	81	99	39.3	156	316	64	49	43	117	22	17	3.39
	1emd	86	39	17	65	99	45.3	195	309	85	66	44	157	18	20	3.23
	1guz	188	15	6	25	35	8	51	304	26	21	4	46	4	1	3.38
	1gv1	130	9	7	30	32	6.9	90	304	40	38	12	72	4	14	3.43
	1ur5	265	62	25	112	170	23.4	172	297	106	49	17	140	25	7	3.39
	1z2i	172	39	21	84	106	22.7	182	348	59	43	80	127	40	15	3.26
	3d5t	76	25	13	55	74	32.9	173	319	75	49	49	125	30	18	3.35
	3fi9	204	34	13	56	86	16.7	117	320	54	30	33	85	13	19	3.22
	3flk	221	29	13	74	102	13.1	191	346	98	67	26	170	18	3	3.31
	3gvh	150	37	22	72	100	24.7	135	316	72	46	17	115	15	5	3.47
	3p7m	128	24	4	19	57	18.8	126	302	68	40	18	99	16	11	3.25
	3tl2	367	62	22	102	152	16.9	146	309	92	34	20	114	26	6	3.33
	4e0b	202	42	17	74	107	20.8	147	299	66	51	30	113	23	11	3.26
	4ror	304	36	18	82	111	11.8	98	316	44	37	17	81	8	9	3.36
	4tvo	237	44	22	112	157	18.6	139	323	73	39	27	122	11	6	3.32
	5ujk	282	50	24	91	125	17.7	128	312	69	43	16	105	11	12	3.35
	6aoo	163	51	29	109	139	31.3	132	313	50	47	35	99	17	16	3.33
	6bal	165	60	31	120	163	36.4	150	308	63	43	44	115	21	14	3.33
	6itk	115	29	15	74	92	25.2	134	315	72	21	41	100	24	10	3.34
	6itl	136	39	21	84	106	28.7	105	310	54	38	13	74	20	11	3.33
ADH-eu	1adb	49	16	7	45	57	32.7	167	373	25	44	98	128	26	13	3.29
	1b15	69	23	11	61	80	33.3	136	253	57	47	32	105	21	10	3.34
	1cdo	300	35	15	61	85	11.7	122	370	43	12	67	104	11	7	3.30
	1d1t	76	16	9	61	74	21.1	141	371	47	29	65	110	18	13	3.37
	1ee2	446	69	28	130	191	15.5	149	371	45	32	72	121	16	12	3.33
	1ht0	279	58	23	105	156	20.8	169	372	65	27	77	118	34	17	3.19
	1mc5	166	33	14	66	97	19.9	114	368	42	20	52	98	10	6	3.30
	1mg5	357	58	22	104	155	16.2	137	253	61	44	32	109	19	9	3.31
	1u3t	203	30	14	64	89	14.8	118	371	46	26	46	94	16	8	3.35
	3wle	222	66	35	159	202	29.7	192	333	60	72	60	131	42	19	3.29
	4jji	212	39	23	107	132	18.4	142	375	53	14	75	99	23	20	3.35
	4rqt	173	37	18	88	118	21.4	134	373	44	16	74	92	29	13	3.24
	4w6z	41	29	13	55	73	70.7	140	345	43	44	53	110	19	11	3.25
	5ilg	146	19	10	51	66	13	82	262	35	25	22	64	9	9	3.33
GDH-eu	1dss	212	38	19	55	81	17.9	148	329	49	54	44	103	17	27	3.44
	1j0x	117	23	13	53	70	19.7	153	328	59	55	39	122	22	9	3.38
	1k3t	261	30	13	64	85	11.5	164	355	76	61	27	136	19	9	3.30
	1vsu	87	31	16	68	92	35.6	178	322	65	77	36	115	35	28	3.26

	1ywg	87	18	10	56	67	20.7	152	332	67	51	34	110	22	20	3.34
	2i5p	126	29	12	50	77	23	174	300	55	67	52	136	18	20	3.31
	2vyn	206	30	15	61	82	14.6	132	325	52	47	33	96	21	15	3.33
	3cps	165	36	19	84	112	21.8	163	330	67	59	37	115	32	16	3.31
	3pym	221	36	18	92	120	16.3	187	330	79	70	38	150	20	17	3.23
	3qv1	55	21	10	47	65	38.2	137	329	54	47	36	106	21	10	3.31
	3sth	188	52	24	100	137	27.7	171	334	92	37	42	110	25	36	3.29
	4k9d	121	25	13	58	71	20.7	137	333	64	38	34	96	24	16	3.34
	5c7i	203	29	14	42	65	14.3	124	326	46	16	62	96	19	9	3.34
	5tso	264	35	12	65	99	13.3	122	328	51	39	31	86	28	7	3.36
MDH-eu	1civ	93	49	24	106	155	52.7	249	369	105	45	99	182	38	29	3.30
	1mld	123	45	20	88	123	36.6	145	311	76	41	28	120	22	3	3.25
	1sev	142	28	16	58	71	19.7	132	311	55	47	30	105	19	8	3.39
	2dfd	188	44	23	105	133	23.4	136	310	77	36	23	109	18	9	3.28
	2fn7	89	30	14	60	81	33.7	147	296	79	44	24	113	24	10	3.33
	2g76	130	14	6	26	37	10.8	86	301	37	18	31	58	18	10	3.26
	2hjr	108	18	11	48	55	16.7	89	311	40	18	31	66	11	12	3.31
	2i6t	141	22	12	41	53	15.6	113	279	37	33	43	75	20	18	3.30
	3i0p	96	42	22	101	130	43.8	150	359	69	18	63	110	32	8	3.39
	4h7p	316	52	25	105	146	16.5	117	309	53	37	25	97	7	11	3.39
	4mdh	190	43	24	104	128	22.6	199	329	97	73	29	150	26	23	3.46
	4plh	122	24	18	63	73	19.7	102	311	47	37	18	88	13	1	
	4plt	235	44	20	91	120	18.7	139	315	78	29	32	106	20	13	3.34
	4uuo	19	4	2	8	14	21.1	130	323	94	19	17	99	20	11	3.24
	5nue	294	49	23	98	131	16.7	129	324	64	37	28	106	13	10	3.36
	5zi2	103	51	25	107	145	49.5	134	336	72	37	25	101	14	19	3.29
	6um4	213	27	13	74	100	12.7	111	321	48	24	39	89	10	12	3.31
	7mdh	59	23	12	66	81	39	173	347	88	38	47	118	30	25	3.38

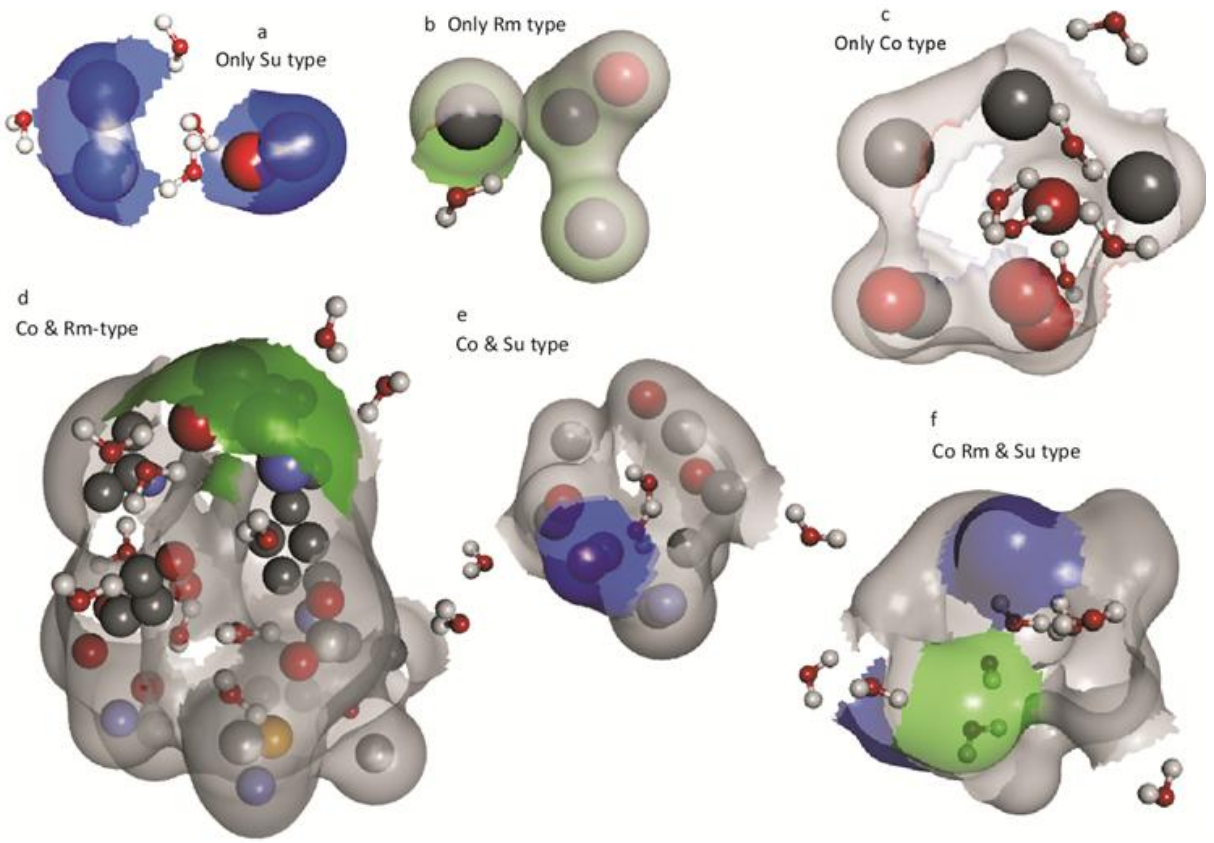


Figure 10. Shell-water and atoms in and out of cavity.

Different types of cavities formed by protein atoms from only the su type (**a**), only the rm (**b**), only the co (**c**), co and rm (**d**), co and su (**e**), and co, rm and su (**f**) regions. Atoms from su, rm, and co regions are shown by blue, green, and grey colors, respectively.

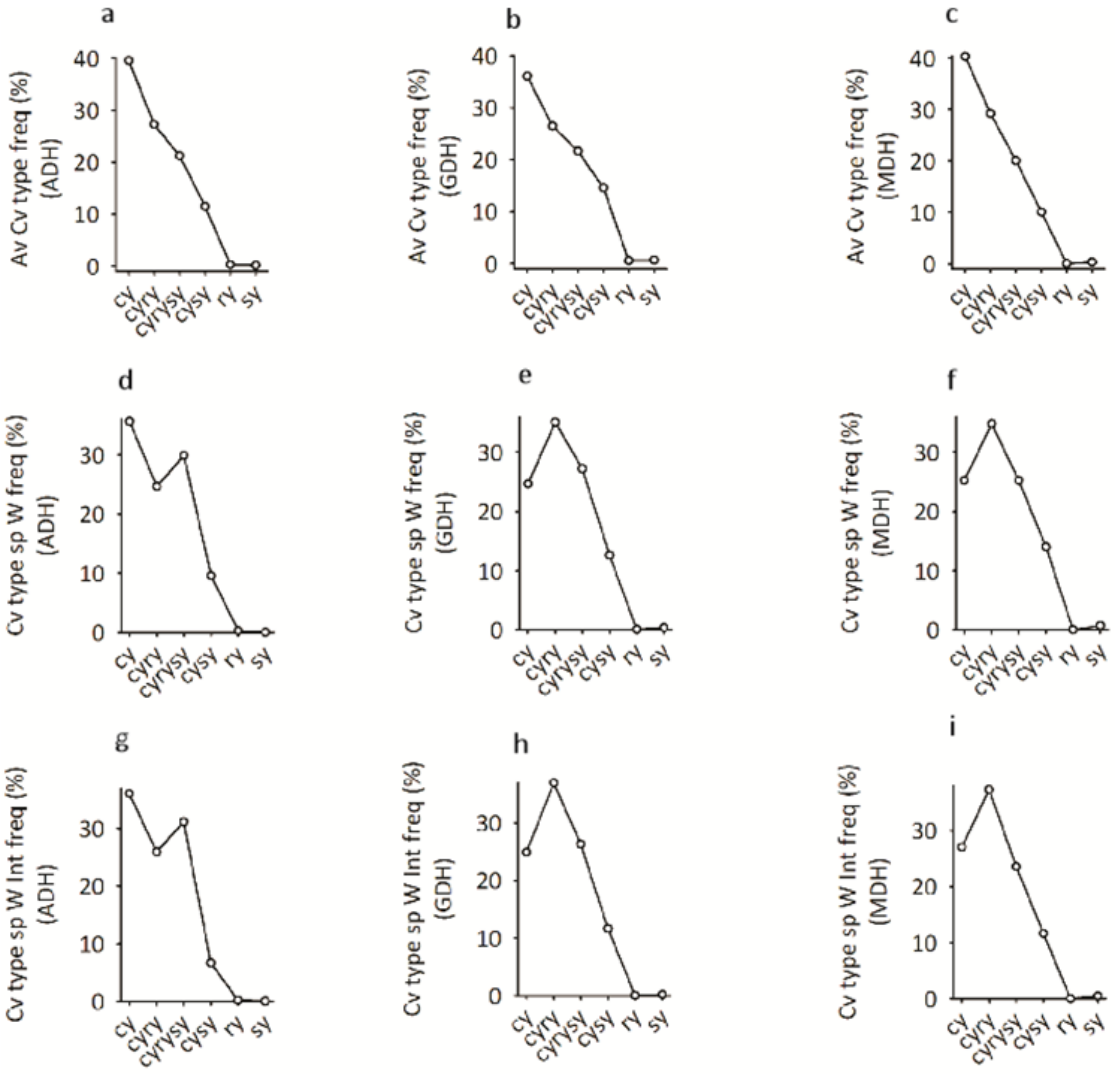


Figure 11. Enzyme-specific percent frequency of cavity

Average cavity types normalized frequency for ADH (**a**), GDH (**b**), and MDH (**c**). Similarly, cavity type specific average (%) heteroatom content for ADH (**d**), GDH (**e**) and MDH (**f**). Again, cavity type specific cavity-water and cavity-atom interaction normalized frequency.

Table 6. Domain-specific and enzyme-specific residue propensity to be part of cavity.

adh_ar	A	V	L	I	F	M	C	P	G	S	T	N	Q	Y	W	H	R	K	E	D
1h2b	0.7	0.6	1.5	0.8	1.5	1.3	1.3	1.7	1.2	1.9	0.7	1.5	0.8	1.4	5.1	1.7	0.7	0.0	0.5	0.8
1jvb	1.3	1.8	1.7	1.1	2.3	0.0	0.0	0.9	0.1	0.6	1.0	0.8	1.4	2.1	4.5	1.3	0.3	0.2	0.5	0.3
1rhc	1.3	1.1	1.5	2.1	1.3	0.6	1.2	0.4	0.2	0.8	1.1	1.2	1.4	1.3	2.2	0.9	1.0	0.3	1.1	0.3
2eer	1.0	1.1	1.5	0.7	1.1	2.1	0.0	2.0	0.8	0.6	1.4	0.9	1.8	1.5	2.7	1.8	0.3	0.5	0.4	0.7
2h6e	1.1	1.4	1.1	1.6	1.7	0.0	0.0	1.3	0.7	0.7	0.6	0.3	1.7	1.7	0.0	2.3	0.7	0.9	0.8	0.7
4jbg	0.5	1.4	1.7	1.3	1.8	0.0	0.0	0.6	0.8	1.8	1.4	1.4	2.5	1.0	2.0	0.0	0.6	0.4	0.3	1.1
6c75	0.8	1.3	1.7	1.7	1.7	1.4	0.0	0.8	0.7	1.1	1.1	1.1	0.0	1.0	1.0	0.8	0.2	0.9	0.5	0.7

av	1.0	1.2	1.5	1.3	1.6	0.8	0.4	1.1	0.6	1.1	1.0	1.0	1.4	1.4	2.5	1.3	0.5	0.5	0.6	0.7
sd	0.3	0.4	0.2	0.5	0.4	0.8	0.6	0.6	0.4	0.6	0.3	0.4	0.8	0.4	1.8	0.8	0.3	0.3	0.3	0.3
adh_ba	**A**	**V**	**L**	**I**	**F**	**M**	**C**	**P**	**G**	**S**	**T**	**N**	**Q**	**Y**	**W**	**H**	**R**	**K**	**E**	**D**
1rjw	0.7	0.9	1.7	1.0	1.4	1.1	0.6	1.0	1.3	0.0	1.9	0.0	0.0	2.1	3.8	0.6	1.4	0.6	1.1	0.4
3ox4	1.1	1.2	1.6	0.2	1.8	1.9	0.6	0.2	0.5	1.0	1.0	1.0	0.0	2.6	0.0	1.3	1.1	0.3	1.4	0.5
3uog	1.0	1.3	1.3	2.0	1.8	0.0	4.4	0.9	0.6	1.2	1.1	2.2	2.6	0.0	1.5	0.5	0.8	0.7	0.0	0.7
4cpd	0.5	0.9	0.8	2.4	3.3	2.4	2.9	1.0	1.0	0.6	0.0	0.4	0.8	2.1	0.0	1.1	1.0	0.6	0.2	0.7
4eez	0.9	1.2	1.5	1.2	1.7	1.2	1.8	1.0	0.9	0.7	1.7	0.7	0.0	0.7	2.4	1.0	1.2	0.6	0.6	0.6
4gkv	0.6	1.7	1.0	0.7	2.4	1.8	1.4	0.7	1.1	0.8	0.7	0.6	1.1	2.4	3.2	0.7	1.6	0.6	0.3	0.3
4j6f	0.7	1.3	1.6	1.4	1.8	0.0	2.4	0.2	0.7	1.6	1.0	0.5	1.4	1.7	0.0	1.6	0.9	0.4	0.2	0.8
4z6k	1.5	1.9	1.1	0.0	0.7	0.6	0.8	1.1	0.7	0.7	0.9	1.0	1.1	2.1	2.9	0.9	0.7	1.0	0.3	0.6
5yln	1.2	0.7	2.5	2.0	2.2	1.5	0.0	0.5	0.2	0.4	1.7	0.8	1.5	0.6	0.0	1.4	1.2	0.5	0.3	0.3
av	0.9	1.2	1.5	1.2	1.9	1.2	1.7	0.7	0.8	0.8	1.1	0.8	0.9	1.6	1.5	1.0	1.1	0.6	0.5	0.5
sd	0.3	0.4	0.5	0.8	0.7	0.8	1.4	0.4	0.3	0.5	0.6	0.6	0.9	0.9	1.6	0.4	0.3	0.2	0.5	0.2
adh_eu	**A**	**V**	**L**	**I**	**F**	**M**	**C**	**P**	**G**	**S**	**T**	**N**	**Q**	**Y**	**W**	**H**	**R**	**K**	**E**	**D**
1adb	0.7	0.7	1.4	1.4	2.7	1.1	0.7	0.5	0.8	1.5	1.0	1.2	1.2	0.0	2.4	0.7	0.8	0.7	0.7	0.9
1b15	0.9	1.6	1.2	2.3	0.8	4.1	2.1	0.4	0.8	0.7	0.5	1.6	0.0	2.8	0.8	0.0	0.0	0.5	0.8	0.3
1cdo	0.6	1.0	1.9	0.8	3.3	2.1	1.6	0.3	0.9	1.5	0.0	1.3	1.2	0.0	0.0	1.1	0.0	0.6	0.6	0.6
1d1t	0.5	1.0	2.2	1.0	1.2	0.6	1.1	0.5	0.7	1.1	0.7	1.1	3.1	1.0	5.2	1.5	0.9	0.8	1.0	0.6
1ee2	0.3	0.8	1.8	1.5	2.2	1.2	1.1	0.8	0.7	1.6	0.5	1.3	1.2	0.0	2.6	1.5	1.9	0.5	0.3	0.3
1ht0	0.3	0.7	1.8	1.5	2.2	1.4	0.6	1.0	0.6	1.7	1.1	1.2	0.7	1.2	2.4	0.8	1.0	0.7	0.7	0.3
1mc5	0.6	1.0	1.5	0.9	0.9	3.7	0.4	1.6	0.8	1.6	1.1	0.7	2.2	1.8	1.6	1.8	0.7	0.4	0.5	0.5
1mg5	0.6	2.0	2.0	1.3	0.8	0.0	1.8	0.0	0.6	0.5	0.8	1.8	1.0	1.2	0.7	0.0	0.7	0.6	0.4	0.0
1u3t	0.5	0.9	1.8	1.6	1.7	1.6	1.0	1.0	0.6	2.1	0.7	1.4	0.7	2.6	2.6	0.7	0.7	0.5	0.3	0.0
3wle	1.0	0.7	1.3	1.4	1.7	2.3	0.8	0.9	1.0	0.4	1.4	0.8	1.9	0.8	3.4	2.1	0.9	0.1	0.9	1.2
4jji	0.9	0.8	1.9	0.7	2.0	2.1	1.4	0.3	0.4	0.5	0.8	0.4	0.5	1.6	3.1	1.7	1.4	0.9	1.6	0.9
4rqt	1.4	0.5	1.9	1.3	1.8	0.6	2.1	0.6	0.5	0.5	1.2	1.3	0.9	0.9	0.0	1.6	2.0	0.6	0.5	0.3
4w6z	1.0	0.7	1.6	2.3	1.2	1.6	1.2	0.0	0.5	1.4	1.0	1.3	1.1	1.4	1.9	1.0	1.2	0.4	0.5	0.9
5ilg	1.0	2.2	0.6	1.6	2.1	1.5	0.0	0.0	0.6	0.5	1.2	1.0	0.0	2.0	0.0	0.0	0.6	1.0	0.0	0.4
av	0.7	1.0	1.6	1.4	1.8	1.7	1.1	0.6	0.7	1.1	0.9	1.2	1.1	1.2	1.9	1.0	0.9	0.6	0.6	0.5
sd	0.3	0.5	0.4	0.5	0.7	1.1	0.6	0.5	0.2	0.6	0.4	0.4	0.8	0.9	1.5	0.7	0.6	0.2	0.4	0.4
gdh_ar	**A**	**V**	**L**	**I**	**F**	**M**	**C**	**P**	**G**	**S**	**T**	**N**	**Q**	**Y**	**W**	**H**	**R**	**K**	**E**	**D**
1b7g	0.9	1.5	1.4	2.3	0.0	0.5	0.0	0.4	0.0	1.0	0.8	1.0	0.0	1.4	0.0	0.9	1.5	0.4	0.8	0.7
1cf2	0.9	0.9	1.2	1.5	2.2	0.6	0.0	1.0	0.4	1.1	0.8	0.5	1.1	1.4	7.6	1.1	2.7	0.6	0.6	0.8
1uxt	1.1	1.0	1.3	1.5	1.1	0.9	3.5	1.4	0.9	1.0	0.5	2.3	1.4	1.8	0.9	3.5	0.6	0.5	0.4	0.5
2czc	0.8	1.4	2.0	1.6	1.0	0.0	0.0	0.0	0.3	0.6	1.3	1.2	0.9	0.0	4.4	2.2	1.9	0.5	0.5	1.0
2yyy	1.0	1.0	2.1	2.2	1.0	0.9	0.0	0.7	0.4	0.0	0.7	1.1	0.6	2.7	0.0	2.0	1.2	0.5	0.5	0.4
av	0.9	1.2	1.6	1.8	1.1	0.6	0.7	0.7	0.4	0.7	0.8	1.2	0.8	1.5	2.6	1.9	1.6	0.5	0.6	0.7
sd	0.1	0.3	0.4	0.4	0.8	0.4	1.6	0.5	0.3	0.5	0.3	0.7	0.5	1.0	3.3	1.0	0.8	0.1	0.2	0.2
gdh_ba	**A**	**V**	**L**	**I**	**F**	**M**	**C**	**P**	**G**	**S**	**T**	**N**	**Q**	**Y**	**W**	**H**	**R**	**K**	**E**	**D**

1euh	0.4	1.2	1.6	1.3	1.7	1.9	3.4	0.9	1.0	1.3	0.8	1.2	0.8	0.9	0.0	0.0	1.3	0.7	0.6	0.7
1gad	0.7	0.9	1.9	1.9	1.1	1.2	1.4	0.9	0.6	0.0	0.6	1.4	3.3	1.0	2.8	0.0	1.7	0.8	0.6	0.8
1gd1	0.6	1.2	1.5	1.4	0.7	1.5	1.8	1.1	0.3	0.6	0.4	1.4	1.2	2.2	3.6	0.8	1.7	0.5	0.9	0.7
1obf	0.4	0.9	1.5	1.9	1.8	0.7	2.2	0.7	0.3	1.3	0.6	0.8	1.5	2.4	1.5	1.0	2.2	0.2	0.4	1.3
2d2i	0.7	1.3	1.7	1.3	0.6	0.0	2.6	0.5	0.5	0.7	0.0	2.0	0.6	3.9	1.0	0.9	1.4	0.6	0.3	0.9
2g82	0.5	1.3	2.0	1.4	1.0	1.0	0.0	0.6	0.3	1.0	0.0	0.6	1.7	2.0	2.2	0.7	1.7	1.2	0.7	0.9
3gnq	0.5	1.0	1.4	0.8	2.0	2.0	2.7	0.7	0.7	1.1	1.2	1.3	1.3	0.8	1.3	0.8	1.4	0.9	0.5	0.7
4dib	1.1	1.1	1.8	1.9	1.8	0.7	4.4	0.0	0.6	1.3	0.2	0.5	0.0	2.6	1.5	1.5	1.6	0.2	0.2	0.7
5ld5	1.0	1.2	1.7	1.6	2.3	1.3	1.4	0.3	0.7	0.8	0.4	0.6	0.8	1.5	0.0	0.0	1.7	1.0	0.6	0.6
5utm	1.1	1.0	2.1	0.7	1.2	1.3	2.3	0.8	1.5	1.2	0.8	0.5	0.4	2.9	2.3	0.0	1.1	0.5	0.2	0.4
6fzh	0.7	0.8	1.8	1.2	1.2	0.7	2.4	1.0	1.5	1.3	1.0	0.5	1.2	3.5	2.4	0.7	0.8	0.2	0.5	0.4
av	0.7	1.1	1.7	1.4	1.4	1.1	2.2	0.7	0.7	1.0	0.5	1.0	1.2	2.2	1.7	0.6	1.5	0.6	0.5	0.7
sd	0.3	0.2	0.2	0.4	0.6	0.6	1.1	0.3	0.4	0.4	0.4	0.5	0.9	1.0	1.1	0.5	0.4	0.3	0.2	0.3
gdh_eu	**A**	**V**	**L**	**I**	**F**	**M**	**C**	**P**	**G**	**S**	**T**	**N**	**Q**	**Y**	**W**	**H**	**R**	**K**	**E**	**D**
1dss	0.9	1.4	1.7	2.0	0.9	0.8	2.1	0.3	0.7	0.8	0.9	1.0	2.4	0.5	0.0	0.0	2.3	0.3	0.2	1.5
1j0x	1.3	1.2	1.9	0.9	1.1	2.5	0.0	0.7	0.2	1.3	1.3	1.4	0.6	1.3	1.3	0.3	1.9	0.3	0.3	0.7
1k3t	0.7	1.4	2.1	2.0	2.6	1.2	2.0	0.3	0.4	0.6	0.4	1.4	1.0	1.7	2.0	0.4	1.1	0.2	0.3	0.6
1vsu	1.0	1.4	1.9	1.1	0.3	1.1	1.5	0.0	0.7	0.8	1.2	0.8	0.0	0.9	0.0	2.3	2.3	0.9	0.3	0.9
1ywg	1.0	1.0	1.4	1.8	1.4	0.0	1.4	0.3	0.7	0.9	0.7	0.9	1.4	1.9	0.0	1.3	1.8	0.6	0.5	0.7
2i5p	1.0	1.4	1.1	1.4	1.9	1.2	1.5	0.3	0.9	0.5	0.6	0.8	2.3	0.9	1.5	1.5	1.0	0.8	0.6	0.6
2vyn	1.2	1.2	1.6	1.8	2.6	1.2	0.0	0.0	0.7	0.0	0.6	0.9	2.4	1.0	1.4	0.0	1.7	0.6	0.6	0.7
3cps	1.3	0.8	1.8	1.8	0.4	1.2	1.6	0.2	0.9	0.8	0.8	0.5	0.9	2.0	0.0	0.0	2.0	0.7	0.3	0.5
3pym	1.0	1.6	2.3	1.6	1.0	1.5	3.4	0.0	0.7	0.5	0.4	1.0	1.4	1.2	1.1	0.4	1.5	0.7	0.2	0.7
3qv1	0.7	1.4	1.5	1.4	2.0	0.0	3.2	0.6	0.5	0.4	0.8	0.8	2.0	2.0	2.0	1.6	1.3	0.5	0.7	0.4
3sth	0.7	0.8	2.1	1.3	1.5	1.1	2.5	0.5	0.9	0.5	1.2	1.2	0.7	1.3	0.9	0.8	2.5	0.4	0.3	0.7
4k9d	0.7	1.4	1.8	1.5	1.1	1.9	1.9	0.8	0.6	0.6	0.7	1.1	1.0	1.5	0.0	1.3	1.9	0.5	0.4	0.8
4lsm	0.6	1.3	1.9	1.8	0.6	2.3	1.7	0.0	0.3	0.7	0.6	0.9	1.4	1.0	1.7	0.0	1.3	0.8	0.9	1.1
5c7i	1.2	0.8	1.5	1.8	1.5	1.6	1.0	0.5	0.0	0.2	0.0	1.5	1.9	2.6	1.6	1.4	2.1	0.3	0.3	1.2
5tso	0.7	0.8	2.6	1.5	1.1	2.3	2.6	0.0	0.7	0.3	0.7	0.7	1.7	1.7	1.7	0.9	3.1	0.4	0.0	1.0
av	0.9	1.2	1.8	1.6	1.3	1.3	1.8	0.3	0.6	0.6	0.7	1.0	1.4	1.4	1.0	0.8	1.9	0.5	0.4	0.8
sd	0.2	0.3	0.4	0.3	0.7	0.7	1.0	0.3	0.3	0.3	0.3	0.3	0.7	0.6	0.8	0.7	0.6	0.2	0.2	0.3
mdh_ar	**A**	**V**	**L**	**I**	**F**	**M**	**C**	**P**	**G**	**S**	**T**	**N**	**Q**	**Y**	**W**	**H**	**R**	**K**	**E**	**D**
1o6z	0.6	0.7	2.5	2.6	1.4	2.7	0.0	1.1	0.9	0.6	0.4	0.6	0.5	1.6	0.0	0.8	2.1	0.7	0.2	0.7
1v9n	0.5	1.9	1.1	0.8	0.4	0.0	0.0	1.3	0.8	1.0	1.4	1.8	0.0	1.3	1.5	0.6	1.1	0.5	0.9	1.1
2d4a	0.8	1.5	2.2	1.7	1.0	1.5	0.0	1.2	0.2	0.8	0.7	0.6	0.0	1.4	0.0	1.7	1.1	0.0	0.4	0.7
2x0i	1.3	1.8	1.5	1.5	1.9	1.8	4.5	0.6	0.5	0.8	1.0	0.4	0.0	0.0	3.0	0.0	0.6	0.6	0.4	0.3
4bgu	0.7	1.0	1.8	2.5	1.9	1.7	4.2	1.2	1.3	1.2	0.2	0.8	1.4	0.0	2.1	1.4	1.3	0.0	0.2	0.3
4bgv	0.7	1.4	0.0	3.0	0.8	3.1	4.7	0.0	1.8	0.4	0.0	0.0	0.0	1.4	0.0	0.0	0.7	0.4	0.0	0.9
6ihd	1.1	1.8	1.6	2.0	1.8	0.8	0.0	0.3	0.7	0.4	0.0	0.9	1.3	2.5	1.3	0.8	0.5	0.8	0.4	0.0
av	0.8	1.4	1.5	2.0	1.3	1.7	1.9	0.8	0.9	0.7	0.5	0.7	0.5	1.2	1.1	0.8	1.1	0.4	0.4	0.6
sd	0.3	0.5	0.8	0.8	0.6	1.1	2.4	0.5	0.5	0.3	0.5	0.6	0.6	0.9	1.2	0.6	0.5	0.3	0.3	0.4

mdh-ba	A	V	L	I	F	M	C	P	G	S	T	N	Q	Y	W	H	R	K	E	D
1b8p	0.7	1.3	1.6	1.4	1.2	1.0	0.0	0.3	0.8	0.8	1.2	0.7	0.6	2.9	3.0	2.0	0.9	0.9	0.6	0.4
1bdm	0.7	1.1	1.0	1.7	0.8	1.5	0.0	0.0	0.4	1.7	1.4	2.1	0.9	0.5	0.9	3.8	1.2	0.7	0.8	1.1
1emd	0.7	1.1	2.2	1.6	1.8	1.5	2.0	0.7	0.5	0.7	0.8	0.3	0.6	2.2	0.0	1.5	0.3	0.6	0.9	0.8
1guz	1.2	1.2	1.5	2.6	0.0	0.9	3.5	0.8	0.4	0.7	2.1	0.0	0.0	0.0	0.0	2.6	0.8	0.0	1.2	0.0
1gv1	0.4	1.5	1.0	3.0	0.9	0.9	1.8	0.4	0.8	0.4	2.4	1.0	0.0	0.0	0.0	1.4	0.4	0.3	0.7	0.3
1ur5	1.1	1.3	0.8	1.4	0.6	1.5	2.2	1.3	0.8	1.3	1.6	0.0	0.0	2.5	0.0	2.2	0.0	0.3	0.9	0.6
1z2i	0.9	1.0	1.2	0.5	1.4	1.9	2.5	1.2	0.7	0.6	1.0	1.1	0.7	2.3	3.7	2.1	0.9	0.5	0.2	0.5
3d5t	0.3	1.5	1.3	1.4	1.1	1.9	1.6	0.2	0.3	0.8	1.1	2.0	1.4	1.6	3.3	2.4	1.6	0.5	0.4	0.9
3fi9	0.9	1.5	1.5	2.2	2.2	0.9	0.0	0.5	0.2	0.8	1.2	0.4	0.5	1.7	1.7	0.7	2.1	0.0	0.7	0.4
3flk	1.0	1.5	1.4	1.4	1.7	0.0	0.9	1.2	0.9	0.7	0.9	0.9	0.0	1.5	1.1	0.4	1.1	0.3	0.5	0.7
3gvh	0.7	1.5	1.6	1.8	1.6	1.0	0.0	0.6	1.1	0.5	1.3	0.0	1.0	1.1	2.2	1.5	0.0	0.4	0.6	0.8
3nep	0.9	1.3	1.2	3.5	4.3	2.3	2.1	1.4	0.0	0.8	1.3	0.0	0.0	0.0	0.0	2.1	0.6	0.0	0.4	0.3
3p7m	1.2	1.6	1.8	1.7	1.1	0.0	1.1	0.4	1.0	0.0	1.7	0.0	0.0	0.9	0.0	2.9	1.2	0.5	0.3	0.4
3tl2	0.9	1.5	0.6	1.3	1.5	2.3	0.0	0.9	1.0	1.7	1.3	0.0	1.3	0.4	0.0	2.0	1.1	0.4	0.4	0.8
4e0b	0.9	0.9	1.7	1.8	2.1	0.0	2.5	0.8	0.8	0.5	0.8	0.8	0.5	1.8	0.0	0.0	0.0	1.2	0.5	0.8
4ror	0.7	0.9	1.1	2.2	1.0	1.3	1.8	1.8	1.2	0.3	0.8	0.0	1.1	2.0	2.7	0.0	0.5	0.8	0.3	1.1
4tvo	0.3	0.9	1.7	1.4	1.8	1.7	0.0	0.6	1.2	1.0	1.4	1.1	1.2	2.8	1.7	1.4	1.7	0.4	0.2	0.4
5ujk	1.2	1.0	1.2	2.2	1.3	0.6	1.6	0.8	0.9	1.2	1.2	0.0	0.9	1.7	2.3	0.0	0.4	0.5	0.5	0.8
6aoo	1.1	1.9	1.4	0.8	1.5	2.3	3.1	0.3	0.9	0.5	0.5	0.0	1.4	1.8	0.0	0.0	0.8	0.5	0.5	1.5
6bal	0.8	1.2	1.0	1.8	1.6	0.9	2.4	0.7	0.9	0.7	0.6	0.3	1.2	2.9	0.0	1.8	1.5	0.6	0.6	1.2
6itk	0.8	1.3	1.3	2.1	1.5	1.0	0.0	1.0	0.6	0.2	1.6	0.9	0.5	3.1	2.1	1.0	1.0	0.9	0.4	0.4
6itl	0.6	1.6	1.4	1.1	2.4	0.0	1.8	1.3	0.2	0.7	0.9	0.8	1.1	2.1	0.0	1.8	0.5	0.8	1.0	1.3
av	0.8	1.3	1.3	1.8	1.5	1.2	1.4	0.8	0.7	0.8	1.2	0.6	0.7	1.6	1.1	1.5	0.8	0.5	0.6	0.7
sd	0.3	0.3	0.4	0.7	0.8	0.7	1.1	0.4	0.3	0.4	0.5	0.6	0.5	1.0	1.3	1.0	0.6	0.3	0.3	0.4
mdh-eu	**A**	**V**	**L**	**I**	**F**	**M**	**C**	**P**	**G**	**S**	**T**	**N**	**Q**	**Y**	**W**	**H**	**R**	**K**	**E**	**D**
1civ	0.2	0.8	1.6	0.8	1.1	1.3	0.0	1.0	0.7	0.8	1.2	0.9	1.5	1.3	1.1	1.4	1.8	1.3	1.1	0.9
1mld	0.7	1.7	0.9	2.6	1.9	0.7	1.1	0.8	0.6	0.5	1.2	1.1	0.5	1.7	0.0	0.8	1.1	0.4	0.6	0.3
1sev	1.1	1.3	1.5	1.0	1.3	0.6	2.8	0.8	0.7	0.9	0.6	0.9	1.2	0.8	0.0	1.6	1.0	0.6	0.3	0.7
2dfd	0.6	1.2	1.1	2.5	2.4	1.5	0.9	1.0	0.6	0.9	0.6	0.9	0.9	1.5	0.0	0.7	1.6	0.3	0.7	0.6
2fn7	0.6	1.9	0.8	1.7	1.1	1.6	1.9	0.9	0.6	1.1	0.6	0.3	0.0	1.1	3.7	2.1	1.6	0.2	0.0	1.0
2g76	0.7	1.9	1.7	1.0	3.4	2.2	1.6	0.0	1.1	0.4	0.7	0.5	1.2	0.0	0.0	2.0	0.7	0.0	0.6	0.4
2hjr	0.5	1.6	1.3	1.5	1.8	0.8	1.0	1.1	1.0	0.2	1.5	0.7	0.0	1.8	0.0	2.0	2.0	0.0	0.9	1.1
2i6t	0.3	1.3	2.1	1.3	1.8	1.8	1.5	0.0	1.1	0.3	1.4	0.4	0.9	1.5	1.1	0.0	0.0	0.9	0.4	1.1
3i0p	1.3	0.9	1.4	1.9	1.2	1.1	0.0	0.6	0.8	1.1	1.2	1.3	0.0	1.3	0.0	1.6	0.8	0.4	0.8	0.4
4h7p	0.4	0.9	1.7	2.0	0.0	0.8	2.0	1.2	0.6	0.9	1.0	2.7	0.0	2.5	3.3	2.1	1.4	0.0	0.3	0.7
4mdh	0.2	1.3	1.5	1.7	1.8	0.4	0.7	0.8	0.5	0.6	1.1	1.1	1.2	1.4	2.5	2.5	0.7	1.3	0.8	0.4
4plh	1.0	1.6	2.1	1.6	1.9	1.1	2.5	0.8	0.7	0.3	1.1	1.0	0.0	1.5	5.0	1.0	0.0	0.0	0.2	0.6
4plt	1.2	0.9	1.5	1.7	1.4	1.2	0.9	0.3	0.8	0.8	1.1	0.5	0.0	1.9	1.9	0.7	1.9	0.6	0.6	0.9
4uuo	1.0	0.6	2.2	1.6	0.7	1.1	1.9	0.6	0.2	0.4	1.3	1.3	0.6	1.8	2.2	0.0	2.0	0.9	0.5	0.6

5nue	0.6	1.3	1.5	1.7	1.4	1.0	1.4	1.0	0.6	0.3	0.7	1.4	1.4	1.2	3.2	2.2	1.3	0.4	0.0	1.0
5zi2	0.5	2.1	1.1	1.5	0.9	0.8	0.0	1.1	0.4	1.3	0.7	1.4	0.0	0.8	0.0	0.9	1.0	0.6	1.1	0.4
6um4	0.7	1.5	1.3	1.2	1.2	2.8	1.4	1.4	0.4	1.0	0.9	1.3	0.5	0.0	2.8	2.8	0.8	0.3	0.3	0.7
7mdh	0.8	0.8	1.6	2.3	1.9	2.3	0.0	0.5	0.7	0.8	0.8	0.9	1.7	1.1	1.6	1.9	1.0	0.6	0.9	0.2
av	0.7	1.3	1.5	1.6	1.5	1.3	1.2	0.8	0.7	0.7	1.0	1.0	0.6	1.3	1.6	1.5	1.2	0.5	0.6	0.7
sd	0.3	0.4	0.4	0.5	0.7	0.6	0.9	0.4	0.2	0.3	0.3	0.5	0.6	0.6	1.6	0.8	0.6	0.4	0.3	0.3

Details of residue propensity to be part of cavity for representative proteins of three enzyme classes of the domains of life. Adh _ar, archaeal ADH; adh_ba, bacterial ADH; adh_eu, eukaryote ADH; gdh_ar, archaeal GDH; gdh_ba, bacterial GDH; gdh_eu, eukaryotic GDH; mdh_ar, archaeal MDH; mdh_ba, bacterial MDH, and MDH-eu, eukaryotic MDH.

Table 7: Protein-specific frequency of types of cavity

Proteins	cy	cyry	cyrysy	cysy	ry	sy
1h2b	43.8	12.5	18.8	18.8	6.3	0.0
1jvb	46.7	13.3	33.3	6.7	0.0	0.0
1rhc	18.2	45.5	27.3	9.1	0.0	0.0
2eer	38.5	30.8	23.1	7.7	0.0	0.0
2h6e	43.8	25.0	31.3	0.0	0.0	0.0
4jbg	40.0	33.3	13.3	13.3	0.0	0.0
6c75	50.0	18.8	9.4	18.8	0.0	3.1
AD_ar_av	40.1	25.6	22.3	10.6	0.9	0.4
1rjw	53.3	20.0	26.7	0.0	0.0	0.0
3ox4	21.7	39.1	17.4	21.7	0.0	0.0
3uog	62.5	12.5	18.8	6.3	0.0	0.0
4cpd	52.2	17.4	13.0	17.4	0.0	0.0
4eez	46.7	40.0	0.0	13.3	0.0	0.0
4gkv	40.0	30.0	20.0	10.0	0.0	0.0
4j6f	38.9	33.3	22.2	5.6	0.0	0.0
4z6k	15.4	23.1	23.1	38.5	0.0	0.0
5yln	18.2	36.4	45.5	0.0	0.0	0.0
AD_ba_av	38.8	28.0	20.7	12.5	0.0	0.0
1adb	41.2	29.4	23.5	5.9	0.0	0.0
1b15	35.3	41.2	0.0	23.5	0.0	0.0
1cdo	50.0	22.2	11.1	16.7	0.0	0.0
1d1t	37.5	25.0	31.3	6.3	0.0	0.0
1ee2	41.2	29.4	17.6	11.8	0.0	0.0
1ht0	38.1	28.6	28.6	4.8	0.0	0.0
1mc5	57.1	21.4	14.3	7.1	0.0	0.0
1mg5	29.4	29.4	23.5	17.6	0.0	0.0
1u3t	35.7	28.6	28.6	7.1	0.0	0.0
3wle	30.8	30.8	30.8	7.7	0.0	0.0

4jji	26.3	21.1	26.3	26.3	0.0	0.0
4rqt	43.8	37.5	18.8	0.0	0.0	0.0
4w6z	40.0	35.0	10.0	15.0	0.0	0.0
5ilg	50.0	16.7	25.0	8.3	0.0	0.0
AD_eu_Av	39.7	28.3	20.7	11.3	0.0	0.0
1b7g	56.3	0.0	31.3	12.5	0.0	0.0
1cf2	44.4	11.1	11.1	33.3	0.0	0.0
1uxt	41.0	35.9	15.4	7.7	0.0	0.0
2czc	33.3	16.7	33.3	16.7	0.0	0.0
2yyy	42.9	17.9	21.4	14.3	3.6	0.0
GD_ar_av	43.6	16.3	22.5	16.9	0.7	0.0
1euh	35.7	28.6	28.6	7.1	0.0	0.0
1gad	36.8	31.6	21.1	10.5	0.0	0.0
1gd1	15.8	31.6	26.3	21.1	5.3	0.0
1obf	29.4	29.4	23.5	11.8	5.9	0.0
2d2i	42.9	35.7	7.1	14.3	0.0	0.0
2g82	9.1	45.5	27.3	18.2	0.0	0.0
3gnq	26.3	42.1	15.8	15.8	0.0	0.0
4dib	35.3	47.1	11.8	5.9	0.0	0.0
5ld5	21.1	31.6	15.8	26.3	0.0	5.3
5utm	27.8	38.9	27.8	5.6	0.0	0.0
6fzh	13.3	40.0	33.3	13.3	0.0	0.0
GD_ba_Av	26.7	36.5	21.7	13.6	1.0	0.5
1dss	26.1	34.8	13.0	21.7	0.0	4.3
1j0x	52.4	19.0	19.0	9.5	0.0	0.0
1k3t	50.0	22.7	13.6	13.6	0.0	0.0
1vsu	26.9	23.1	30.8	19.2	0.0	0.0
1ywg	27.8	22.2	33.3	11.1	0.0	5.6
2i5p	50.0	9.1	27.3	13.6	0.0	0.0
2vyn	40.0	20.0	25.0	15.0	0.0	0.0
3cps	39.1	21.7	26.1	13.0	0.0	0.0
3pym	45.8	29.2	12.5	12.5	0.0	0.0
3qv1	38.9	33.3	16.7	11.1	0.0	0.0
3sth	25.0	20.8	25.0	20.8	0.0	8.3
4k9d	38.9	22.2	27.8	11.1	0.0	0.0
4lsm	35.7	35.7	21.4	7.1	0.0	0.0
5c7i	38.9	50.0	5.6	0.0	0.0	5.6
5tso	33.3	33.3	13.3	20.0	0.0	0.0
GD_eu_Av	37.9	26.5	20.7	13.3	0.0	1.6
1o6z	46.2	46.2	7.7	0.0	0.0	0.0

1v9n	20.0	15.0	30.0	35.0	0.0	0.0
2d4a	41.2	29.4	29.4	0.0	0.0	0.0
2x0i	50.0	25.0	25.0	0.0	0.0	0.0
4bgu	50.0	27.8	11.1	11.1	0.0	0.0
4bgv	87.5	12.5	0.0	0.0	0.0	0.0
6ihd	26.1	39.1	34.8	0.0	0.0	0.0
MD_ar_av	45.8	27.9	19.7	6.6	0.0	0.0
1b8p	18.2	31.8	40.9	9.1	0.0	0.0
1bdm	28.6	38.1	19.0	14.3	0.0	0.0
1emd	50.0	15.4	19.2	11.5	0.0	3.8
1guz	50.0	33.3	0.0	16.7	0.0	0.0
1gv1	53.3	13.3	13.3	20.0	0.0	0.0
1ur5	52.9	23.5	17.6	5.9	0.0	0.0
1z2i	29.2	37.5	25.0	8.3	0.0	0.0
3d5t	38.1	14.3	33.3	14.3	0.0	0.0
3fi9	31.3	18.8	31.3	18.8	0.0	0.0
3flk	44.4	44.4	5.6	5.6	0.0	0.0
3gvh	52.6	31.6	10.5	5.3	0.0	0.0
3nep	57.1	14.3	14.3	14.3	0.0	0.0
3p7m	25.0	33.3	41.7	0.0	0.0	0.0
3tl2	43.8	43.8	6.3	0.0	0.0	6.3
4e0b	42.1	26.3	21.1	10.5	0.0	0.0
4ror	46.2	15.4	30.8	7.7	0.0	0.0
4tvo	40.0	40.0	13.3	6.7	0.0	0.0
5ujk	55.0	10.0	30.0	5.0	0.0	0.0
6aoo	42.1	15.8	26.3	15.8	0.0	0.0
6bal	45.0	25.0	25.0	5.0	0.0	0.0
6itk	31.6	42.1	26.3	0.0	0.0	0.0
6itl	21.4	28.6	35.7	14.3	0.0	0.0
MD_ba_Av	40.8	27.1	22.1	9.5	0.0	0.5
1civ	36.7	20.0	23.3	20.0	0.0	0.0
1mld	55.6	33.3	11.1	0.0	0.0	0.0
1sev	44.4	27.8	16.7	11.1	0.0	0.0
2dfd	44.4	33.3	16.7	5.6	0.0	0.0
2fn7	38.1	28.6	28.6	4.8	0.0	0.0
2g76	22.2	44.4	33.3	0.0	0.0	0.0
2hjr	27.3	36.4	0.0	27.3	0.0	9.1
2i6t	23.5	17.6	41.2	17.6	0.0	0.0
3i0p	30.0	45.0	20.0	5.0	0.0	0.0
4h7p	33.3	33.3	0.0	33.3	0.0	0.0
4mdh	28.0	36.0	16.0	20.0	0.0	0.0

4plh	50.0	42.9	7.1	0.0	0.0	0.0
4plt	45.0	30.0	15.0	10.0	0.0	0.0
4uuo	15.8	47.4	15.8	21.1	0.0	0.0
5nue	42.1	26.3	26.3	5.3	0.0	0.0
5zi2	35.3	11.8	17.6	35.3	0.0	0.0
6um4	27.3	36.4	18.2	18.2	0.0	0.0
7mdh	19.0	38.1	23.8	19.0	0.0	0.0
MD_eu_AV	34.3	32.7	18.4	14.1	0.0	0.5

Domains of life and enzyme class specific representative protein's normalized frequency of different types of cavity and their average (grey row) value.

Table 8: Details on unique shell-water in different types of cavities.

	cyUw	cyryUw	cyrysyUw	cysyUw	ryUw	syUw
1h2b	41.2	9.8	25.5	17.6	5.9	0.0
1jvb	29.6	22.2	40.7	7.4	0.0	0.0
1rhc	5.4	43.2	43.2	8.1	0.0	0.0
2eer	31.6	42.1	15.8	10.5	0.0	0.0
2h6e	20.0	26.7	53.3	0.0	0.0	0.0
4jbg	48.1	29.6	11.1	11.1	0.0	0.0
6c75	40.0	36.0	8.0	16.0	0.0	0.0
AD_ar_av	30.8	30.0	28.2	10.1	0.8	0.0
1rjw	46.2	30.8	23.1	0.0	0.0	0.0
3ox4	31.0	34.5	20.7	13.8	0.0	0.0
3uog	42.4	21.2	33.3	3.0	0.0	0.0
4cpd	65.0	20.0	5.0	10.0	0.0	0.0
4eez	27.8	59.3	0.0	13.0	0.0	0.0
4gkv	29.0	35.5	29.0	6.5	0.0	0.0
4j6f	53.1	31.3	15.6	0.0	0.0	0.0
4z6k	2.9	32.4	32.4	32.4	0.0	0.0
5yln	0.0	60.0	40.0	0.0	0.0	0.0
AD_ba_av	33.1	36.1	22.1	8.7	0.0	0.0
1adb	50.0	31.3	18.8	0.0	0.0	0.0
1b15	0.0	78.3	0.0	21.7	0.0	0.0
1cdo	22.9	11.4	31.4	34.3	0.0	0.0
1d1t	62.5	18.8	0.0	18.8	0.0	0.0
1ee2	34.8	34.8	21.7	8.7	0.0	0.0
1ht0	31.0	27.6	36.2	5.2	0.0	0.0
1mc5	45.5	27.3	27.3	0.0	0.0	0.0
1mg5	8.6	37.9	29.3	24.1	0.0	0.0
1u3t	30.0	30.0	30.0	10.0	0.0	0.0
3wle	24.2	24.2	43.9	7.6	0.0	0.0

4jji	30.8	15.4	28.2	25.6	0.0	0.0
4rqt	21.6	35.1	43.2	0.0	0.0	0.0
4w6z	51.7	24.1	6.9	17.2	0.0	0.0
5ilg	42.1	10.5	26.3	21.1	0.0	0.0
AD_eu_Av	32.6	29.0	24.5	13.9	0.0	0.0
1b7g	28.0	0.0	60.0	12.0	0.0	0.0
1cf2	24.0	16.0	32.0	28.0	0.0	0.0
1uxt	22.2	48.9	16.7	12.2	0.0	0.0
2czc	10.5	15.8	68.4	5.3	0.0	0.0
2yyy	27.6	20.7	41.4	10.3	0.0	0.0
GD_ar_av	22.5	20.3	43.7	13.6	0.0	0.0
1euh	12.7	36.4	44.1	6.8	0.0	0.0
1gad	3.7	48.1	33.3	14.8	0.0	0.0
1gd1	8.1	48.6	24.3	16.2	2.7	0.0
1obf	2.6	48.7	28.2	20.5	0.0	0.0
2d2i	43.8	37.5	12.5	6.3	0.0	0.0
2g82	0.0	28.0	44.0	28.0	0.0	0.0
3gnq	2.9	60.0	22.9	14.3	0.0	0.0
4dib	0.0	33.3	33.3	33.3	0.0	0.0
5ld5	4.5	50.0	18.2	18.2	0.0	9.1
5utm	15.8	39.5	36.8	7.9	0.0	0.0
6fzh	16.0	26.0	48.0	10.0	0.0	0.0
GD_ba_Av	10.0	41.5	31.4	16.0	0.2	0.8
1dss	15.8	34.2	10.5	31.6	0.0	7.9
1j0x	21.7	47.8	26.1	4.3	0.0	0.0
1k3t	13.3	40.0	20.0	26.7	0.0	0.0
1vsu	6.5	29.0	54.8	9.7	0.0	0.0
1ywg	11.1	38.9	38.9	11.1	0.0	0.0
2i5p	20.7	24.1	34.5	20.7	0.0	0.0
2vyn	23.3	33.3	33.3	10.0	0.0	0.0
3cps	5.6	36.1	47.2	11.1	0.0	0.0
3pym	13.9	38.9	25.0	22.2	0.0	0.0
3qv1	19.0	52.4	28.6	0.0	0.0	0.0
3sth	1.9	28.8	34.6	30.8	0.0	3.8
4k9d	12.5	45.8	25.0	16.7	0.0	0.0
4lsm	13.6	54.5	27.3	4.5	0.0	0.0
5c7i	24.1	48.3	17.2	0.0	0.0	10.3
5tso	8.6	74.3	8.6	8.6	0.0	0.0
GD_eu_Av	14.1	41.8	28.8	13.9	0.0	1.5
1o6z	31.6	60.5	7.9	0.0	0.0	0.0
1v9n	13.6	31.8	31.8	22.7	0.0	0.0

2d4a	28.6	57.1	14.3	0.0	0.0	0.0
2x0i	50.0	0.0	50.0	0.0	0.0	0.0
4bgu	50.9	16.4	20.0	12.7	0.0	0.0
4bgv	100.0	0.0	0.0	0.0	0.0	0.0
6ihd	100.0	0.0	0.0	0.0	0.0	0.0
MD_ar_av	53.5	23.7	17.7	5.1	0.0	0.0
1b8p	19.1	21.3	51.1	8.5	0.0	0.0
1bdm	25.7	54.3	11.4	8.6	0.0	0.0
1emd	35.9	15.4	28.2	20.5	0.0	0.0
1guz	33.3	60.0	0.0	6.7	0.0	0.0
1gv1	11.1	11.1	33.3	44.4	0.0	0.0
1ur5	53.2	25.8	17.7	3.2	0.0	0.0
1z2i	28.2	41.0	20.5	10.3	0.0	0.0
3d5t	28.0	12.0	48.0	12.0	0.0	0.0
3fi9	11.8	5.9	55.9	26.5	0.0	0.0
3flk	24.1	72.4	0.0	3.4	0.0	0.0
3gvh	48.6	8.1	24.3	18.9	0.0	0.0
3nep	27.3	9.1	36.4	27.3	0.0	0.0
3p7m	4.2	29.2	66.7	0.0	0.0	0.0
3tl2	30.6	61.3	1.6	0.0	0.0	6.5
4e0b	26.2	26.2	19.0	28.6	0.0	0.0
4ror	44.4	5.6	38.9	11.1	0.0	0.0
4tvo	56.8	25.0	11.4	6.8	0.0	0.0
5ujk	64.0	4.0	26.0	6.0	0.0	0.0
6aoo	15.7	29.4	27.5	27.5	0.0	0.0
6bal	33.3	30.0	30.0	6.7	0.0	0.0
6itk	37.9	31.0	31.0	0.0	0.0	0.0
6itl	20.5	30.8	38.5	10.3	0.0	0.0
MD_ba_Av	30.9	27.7	28.1	13.1	0.0	0.3
1civ	38.8	24.5	28.6	8.2	0.0	0.0
1mld	46.7	28.9	24.4	0.0	0.0	0.0
1sev	14.3	53.6	28.6	3.6	0.0	0.0
2dfd	31.8	43.2	20.5	4.5	0.0	0.0
2fn7	56.7	16.7	23.3	3.3	0.0	0.0
2g76	0.0	71.4	28.6	0.0	0.0	0.0
2hjr	38.9	22.2	0.0	27.8	0.0	11.1
2i6t	9.1	18.2	45.5	27.3	0.0	0.0
3i0p	19.0	69.0	9.5	2.4	0.0	0.0
4h7p	17.3	32.7	0.0	50.0	0.0	0.0
4mdh	9.3	51.2	23.3	16.3	0.0	0.0

4plh	62.5	29.2	8.3	0.0	0.0	0.0
4plt	43.2	22.7	27.3	6.8	0.0	0.0
4uuo	25.0	50.0	0.0	25.0	0.0	0.0
5nue	36.7	8.2	53.1	2.0	0.0	0.0
5zi2	25.5	2.0	17.6	54.9	0.0	0.0
6um4	22.2	44.4	22.2	11.1	0.0	0.0
7mdh	26.1	17.4	43.5	13.0	0.0	0.0
MD_eu_AV	29.1	33.6	22.5	14.2	0.0	0.6

Domains of life and enzyme class specific representative protein's normalized frequency of unique water in different types of cavities and their average (grey row) value.

Table 9: Details of interaction frequency of shell-water in different types of cavities.

	cyTw	cyryTw	cyrysyTw	cysyTw	ryTw	syTw
1h2b	43.5	7.5	30.6	14.3	4.1	0.0
1jvb	23.2	23.2	50.0	3.6	0.0	0.0
1rhc	1.7	46.1	46.1	6.1	0.0	0.0
2eer	31.5	40.7	22.2	5.6	0.0	0.0
2h6e	12.9	23.5	63.5	0.0	0.0	0.0
4jbg	46.7	25.3	20.0	8.0	0.0	0.0
6c75	40.2	47.4	2.1	10.3	0.0	0.0
AD_ar_av	28.5	30.5	33.5	6.8	0.6	0.0
1rjw	42.2	42.2	15.6	0.0	0.0	0.0
3ox4	32.6	42.1	9.5	15.8	0.0	0.0
3uog	43.2	30.6	22.5	3.6	0.0	0.0
4cpd	72.7	14.5	1.8	10.9	0.0	0.0
4eez	23.8	66.0	0.0	10.2	0.0	0.0
4gkv	23.5	47.0	25.2	4.3	0.0	0.0
4j6f	52.7	36.3	11.0	0.0	0.0	0.0
4z6k	2.3	34.1	39.8	23.9	0.0	0.0
5yln	0.0	64.7	35.3	0.0	0.0	0.0
AD_ba_av	32.6	41.9	17.8	7.6	0.0	0.0
1adb	59.6	24.6	15.8	0.0	0.0	0.0
1b15	0.0	83.8	0.0	16.3	0.0	0.0
1cdo	29.4	14.1	35.3	21.2	0.0	0.0
1d1t	66.2	23.0	0.0	10.8	0.0	0.0
1ee2	35.6	36.6	20.9	6.8	0.0	0.0
1ht0	38.5	28.2	31.4	1.9	0.0	0.0
1mc5	47.4	29.9	22.7	0.0	0.0	0.0
1mg5	4.5	47.7	31.0	16.8	0.0	0.0
1u3t	43.8	30.3	16.9	9.0	0.0	0.0
3wle	26.2	23.3	44.1	6.4	0.0	0.0
4jji	27.3	19.7	32.6	20.5	0.0	0.0

4rqt	19.5	39.0	41.5	0.0	0.0	0.0
4w6z	56.2	24.7	6.8	12.3	0.0	0.0
5ilg	42.4	7.6	21.2	28.8	0.0	0.0
AD_eu_Av	35.5	30.9	22.9	10.8	0.0	0.0
1b7g	28.3	0.0	64.2	7.5	0.0	0.0
1cf2	21.5	24.6	35.4	18.5	0.0	0.0
1uxt	21.3	50.6	18.4	9.6	0.0	0.0
2czc	21.1	15.8	60.5	2.6	0.0	0.0
2yyy	23.9	23.9	45.1	7.0	0.0	0.0
GD_ar_av	23.2	23.0	44.7	9.1	0.0	0.0
1euh	12.6	34.5	47.3	5.6	0.0	0.0
1gad	2.5	51.9	32.1	13.6	0.0	0.0
1gd1	5.5	58.2	14.5	20.9	0.9	0.0
1obf	1.1	43.5	37.0	18.5	0.0	0.0
2d2i	39.2	37.3	13.7	9.8	0.0	0.0
2g82	0.0	31.6	46.1	22.4	0.0	0.0
3gnq	1.0	62.9	25.7	10.5	0.0	0.0
4dib	0.0	7.7	69.2	23.1	0.0	0.0
5ld5	11.5	55.7	13.1	16.4	0.0	3.3
5utm	20.2	38.5	33.9	7.3	0.0	0.0
6fzh	21.2	24.6	49.2	5.1	0.0	0.0
GD_ba_Av	10.4	40.6	34.7	13.9	0.1	0.3
1dss	14.8	33.3	11.1	35.8	0.0	4.9
1j0x	24.3	55.7	17.1	2.9	0.0	0.0
1k3t	11.8	51.8	22.4	14.1	0.0	0.0
1vsu	6.5	31.5	53.3	8.7	0.0	0.0
1ywg	16.4	40.3	32.8	10.4	0.0	0.0
2i5p	22.1	28.6	24.7	24.7	0.0	0.0
2vyn	19.5	32.9	37.8	9.8	0.0	0.0
3cps	5.4	41.1	45.5	8.0	0.0	0.0
3pym	10.8	44.2	27.5	17.5	0.0	0.0
3qv1	27.7	44.6	27.7	0.0	0.0	0.0
3sth	0.7	45.3	28.5	23.4	0.0	2.2
4k9d	10.0	55.7	28.6	5.7	0.0	0.0
4lsm	9.2	63.2	19.7	7.9	0.0	0.0
5c7i	21.5	53.8	16.9	0.0	0.0	7.7
5tso	8.1	77.8	7.1	7.1	0.0	0.0
GD_eu_Av	13.9	46.6	26.7	11.7	0.0	1.0
1o6z	33.9	59.6	6.4	0.0	0.0	0.0
1v9n	10.0	36.7	30.8	22.5	0.0	0.0
2d4a	31.6	57.9	10.5	0.0	0.0	0.0

2x0i	58.8	0.0	41.2	0.0	0.0	0.0
4bgu	60.7	15.6	17.8	5.9	0.0	0.0
4bgv	100.0	0.0	0.0	0.0	0.0	0.0
6ihd	100.0	0.0	0.0	0.0	0.0	0.0
MD_ar_av	56.4	24.2	15.2	4.1	0.0	0.0
1b8p	25.8	13.6	55.3	5.3	0.0	0.0
1bdm	27.3	50.5	10.1	12.1	0.0	0.0
1emd	33.3	16.2	30.3	20.2	0.0	0.0
1guz	25.7	71.4	0.0	2.9	0.0	0.0
1gv1	9.4	12.5	21.9	56.3	0.0	0.0
1ur5	50.0	28.2	19.4	2.4	0.0	0.0
1z2i	35.8	39.6	18.9	5.7	0.0	0.0
3d5t	28.4	21.6	40.5	9.5	0.0	0.0
3fi9	23.3	8.1	47.7	20.9	0.0	0.0
3flk	22.5	70.6	0.0	6.9	0.0	0.0
3gvh	48.0	10.0	22.0	20.0	0.0	0.0
3nep	27.3	4.5	31.8	36.4	0.0	0.0
3p7m	8.8	28.1	63.2	0.0	0.0	0.0
3tl2	29.6	63.2	0.7	0.0	0.0	6.6
4e0b	20.6	20.6	20.6	38.3	0.0	0.0
4ror	51.4	9.9	30.6	8.1	0.0	0.0
4tvo	66.9	21.7	9.6	1.9	0.0	0.0
5ujk	66.4	6.4	20.8	6.4	0.0	0.0
6aoo	12.9	26.6	34.5	25.9	0.0	0.0
6bal	31.3	27.6	36.8	4.3	0.0	0.0
6itk	30.4	29.3	40.2	0.0	0.0	0.0
6itl	22.6	42.5	24.5	10.4	0.0	0.0
MD_ba_Av	31.7	28.3	26.3	13.3	0.0	0.3
1civ	41.3	16.1	32.3	10.3	0.0	0.0
1mld	39.8	31.7	28.5	0.0	0.0	0.0
1sev	15.5	52.1	31.0	1.4	0.0	0.0
2dfd	32.3	37.6	28.6	1.5	0.0	0.0
2fn7	59.3	21.0	16.0	3.7	0.0	0.0
2g76	0.0	83.8	16.2	0.0	0.0	0.0
2hjr	52.7	21.8	0.0	18.2	0.0	7.3
2i6t	11.3	15.1	39.6	34.0	0.0	0.0
3i0p	23.1	63.8	9.2	3.8	0.0	0.0
4h7p	21.2	31.5	0.0	47.3	0.0	0.0
4mdh	9.4	56.3	25.8	8.6	0.0	0.0
4plh	54.8	42.5	2.7	0.0	0.0	0.0
4plt	41.7	24.2	30.8	3.3	0.0	0.0

4uuo	28.6	57.1	0.0	14.3	0.0	0.0
5nue	45.8	7.6	45.8	0.8	0.0	0.0
5zi2	29.7	2.1	17.2	51.0	0.0	0.0
6um4	23.0	44.0	21.0	12.0	0.0	0.0
7mdh	40.7	11.1	35.8	12.3	0.0	0.0
MD_eu_AV	31.7	34.4	21.1	12.4	0.0	0.4

Domains of life and enzyme class specific representative protein's normalized frequency of interacting water in different types of cavities and their average (grey row) value.

Table 10 shows the normalized quantitative details on cavity SW and residue for candidate proteins of the domain-specific protein family of our database. Several points are noteworthy from the table. First, a cavity is a homogenous or heterogeneous structural unit formed by region-specific protein atoms. In addition to these, SW can also be a component. Irrespective of domains of life and enzyme families, $1/4^{th}$ to $1/5^{th}$ of total residues of a protein can participate in cavity. More than half of which are of the hydrophobic type (Table 10; Table 11). Second, at the same time, on average, ~20% of the SW of protein also participates in cavity formation. Notably, in all enzyme classes, the SW of cavity of archaea is much less than that of ba and eu. However, about 70% of the total cavities of a protein are filled with SW and the rests are empty. Although lower in the case of archaea, the average interaction multiplicity of SW and protein atoms is ~3. Third, in cavity, the predominance of atoms in the helix is greater (~43%) than that of the strand (28%) and coil (28%). Similarly, the atoms in the co are much higher (~76%) in the cavity than that in the rm and su. Forth, a typical cavity is shown in figure 9a. It is formed by atoms of residues from the co, rm, and su regions (Table 12). The majority of the atoms are present in the helix. Cavity SWs that are present inside the cavity have much higher multiplicity than outside (Figure 9a and Table 5). Since more cavities (~76%) are present in the co, and since more are filled with SW (~70%), their role in the overall property and stability of the co is immense. Finally, it can be said that the characteristics of this cavity in a class of three enzymes regardless of the domains of life follow a certain pattern.

Table 10: Details of average cavity compositions along with standard deviation.

	ADH (average %)			MDH (average %)			GDH (average %)		
Items	**ar (n=7)**	**ba (n=9)**	**eu (n=14)**	**ar (n=7)**	**ba (n=22)**	**eu (n=18)**	**ar (n=5)**	**ba (n=11)**	**eu (n=15)**
R‡	20.7±5.9	20.3±4.6	20.9±3.7	20.2±4.8	20.8±5.4	23.0±4.6	20.3±6.5	22.7±3.9	25.2±3.7
R_hb	14.2±4.0	13.1±2.9	12.8±2.6	13.9±3.5	14.8±3.4	15.1±2.4	12.1+_5.2	13.0±2.8	15.2±2.7
R_po	4.9±1.4	3.9±1.3	4.8±1.3	3.1±1.3	4.3±1.7	4.3±1.4	3.9±1.0	5.3±1.4	5.3±0.9
R_cr	3.6±1.3	3.3±0.9	3.2±1.3	3.2±1.2	3.8±1.5	3.6±2.2	4.2±0.4	4.4±1.3	4.6±0.8
Fr(W)#	18.1±6.1	23.9±11.5	24.2±15.0	15.0±6.8	22.1±10.6	26.1±13.0	16.9±10.5	19.0±7.0	20.2±8.5
FillW#	70.7±18.2	69.7±20.5	77.1±17.0	54.5±30.6	73.8±12.3	68.9±20.1	61.8±20.9	71.6±20.0	64.4±10.0
Mul#	2.9±0.4	3.1±0.5	3.1±0.6	2.8±0.7	2.8±0.4	3.0±0.4	2.4±0.3	3.0±0.5	2.9±0.4
H*	43.2±16.0	39.8±16.1	34.9±7.3	48.3±11.9	47.4±7.8	48.9±8.4	42.9±8.8	40.3±9.4	40.7±5.6
S*	24.8±10.6	27.6±9.6	23.0±9.3	31.7±14.2	32.1±8.3	25.6±7.3	27.3±6.7	31.3±7.1	33.3±7.5

C*	32.0±10.6	32.7±8.7	42.1±12.0	20.0±21.4	20.5±10.1	25.5±9.4	29.8±7.2	28.3±8.2	26.1±7.5
co*	73.1±6.2	79.3±5.7	76.9±5.6	79.6±11.0	78.6±6.1	76.6±5.7	78.6±6.1	71.3±5.6	73.2±5.8
rm*	16.0±5.0	12.8±4.4	14.6±4.1	13.6±5.6	12.9±4.6	14.2±3.9	11.4±4.2	17.5±4.1	15.7±3.8
su*	10.9±2.7	7.9±3.6	8.5±2.3	6.9±7.3	8.5±3.9	9.3±4.2	10.0±4.4	11.2±4.4	11.1±4.7

The normalization is done by total cavity atoms of a protein. ‡Unique residues in cavities in reference to total residue of a protein. #Unique waters in reference to total shell-water. R, amino acid residue; W, Shell-water; fr(W), % fraction of shell-water in cavity (i.e. cavW*100/totW); Mul, Multiplicity; H, Helix; S, strand; C, Coil; co, core; rm, rim; su, surface

The average measurement of items of the cavity in the enzyme class and domains of life-specific manner. Items include residue (R) class (hydrophobic, R_hb; polar, R_po; charged, R-cr), SW (fractional frequency fr(W) in %; frequency of the SW-filled cavity, fillW; interaction multiplicity of SW, mul), secondary structure (helix, H; strand, S; coil, C), and accessibility regions (co, rm and su).

Table 11: Details of normalized residue content in the cavity.

Residue	ADHar		ADHba		ADHeu		mdhar		mdhba		mdheu		GDH-ar		GDH-ba		GDH-eu	
	av%	sd	%av	sd	av%	sd	av	sd	av	sd	av	sd	av	sd	av	sd	av	sd
A	2.3	1.0	2.0	1.0	1.3	0.7	1.6	0.9	2.2	0.8	1.4	0.6	1.8	1.1	1.7	0.6	2.3	0.8
V	2.8	0.9	2.7	1.0	2.2	1.1	2.7	1.1	2.8	0.9	2.8	1.0	2.5	0.9	2.5	0.6	3.3	1.2
L	3.2	1.2	2.5	1.0	2.6	1.2	2.5	1.2	3.0	1.5	3.1	1.1	2.2	1.0	3.0	0.4	2.9	0.7
I	1.9	1.0	1.4	0.9	2.1	1.1	3.0	1.0	2.5	0.9	2.7	1.2	3.3	1.6	2.2	0.8	2.6	0.6
F	1.1	0.8	1.1	0.5	1.5	0.7	0.8	0.6	1.1	0.6	1.1	0.6	0.5	0.3	0.9	0.5	1.2	0.6
M	0.5	0.5	0.5	0.6	0.6	0.3	0.8	0.5	0.6	0.5	0.7	0.3	0.3	0.2	0.5	0.3	0.9	0.5
C	0.1	0.2	0.5	0.3	0.7	0.4	0.1	0.2	0.3	0.2	0.5	0.3	0.0	0.1	0.4	0.2	0.6	0.4
P	1.1	0.4	0.7	0.4	0.6	0.4	0.7	0.4	0.8	0.5	0.8	0.5	0.8	0.9	0.5	0.3	0.3	0.3
G	1.3	0.8	1.7	0.7	1.4	0.6	1.6	0.8	1.6	0.9	1.3	0.4	0.7	1.0	1.4	1.0	1.3	0.6
S	1.0	0.5	0.8	0.6	1.3	0.7	0.8	0.4	0.9	0.5	1.0	0.6	0.6	0.5	1.2	0.5	1.0	0.5
T	1.0	0.4	1.1	0.5	1.1	0.5	0.5	0.4	1.3	0.5	1.2	0.5	0.8	0.3	1.0	0.7	1.2	0.5
N	0.6	0.2	0.4	0.4	1.0	0.9	0.6	0.5	0.5	0.7	1.0	0.4	1.0	0.4	1.3	0.7	1.2	0.4
Q	0.7	0.4	0.4	0.4	0.5	0.4	0.3	0.4	0.4	0.4	0.5	0.4	0.3	0.2	0.4	0.3	0.6	0.3
Y	1.1	0.4	0.9	0.6	0.5	0.4	0.7	0.6	0.8	0.5	0.7	0.4	1.1	0.9	1.2	0.4	1.1	0.5
W	0.5	0.5	0.3	0.3	0.3	0.2	0.2	0.3	0.3	0.4	0.4	0.4	0.2	0.2	0.3	0.2	0.2	0.2
H	0.5	0.4	0.5	0.2	0.5	0.3	0.3	0.2	0.4	0.4	0.5	0.3	0.5	0.3	0.3	0.3	0.4	0.4
R	0.6	0.3	0.9	0.6	0.5	0.4	0.9	0.5	0.9	0.8	0.9	0.5	1.4	0.2	1.4	0.4	1.5	0.3
K	0.8	0.8	0.6	0.2	0.9	0.4	0.7	0.7	0.7	0.4	0.9	0.9	0.7	0.2	0.9	0.5	1.0	0.5
E	1.1	0.6	0.6	0.6	0.7	0.4	0.6	0.6	0.8	0.4	0.8	0.6	0.8	0.1	0.7	0.5	0.4	0.3
D	0.6	0.2	0.7	0.4	0.6	0.6	0.7	0.4	1.0	0.5	0.8	0.4	0.9	0.1	1.1	0.4	1.3	0.3

Domains of life and enzyme class-specific normalized residue content (%) in the cavity along with standard deviation.

Table 12: Details of region (reg) specific atoms of a cavity.

	ATOM		RES	CH	ID	X	Y	Z	ASA	SS	Reg	
ATOM	1	CZ	TYR	A	24	77.200	-11.273	-16.829	23.5	C	rm	C
ATOM	2	OH	TYR	A	24	77.356	-12.108	-15.712	23.5	C	rm	O
ATOM	3	CE2	TYR	A	24	76.793	-9.948	-16.667	23.5	C	rm	C
ATOM	4	CA	LEU	A	60	78.264	-6.035	-13.663	0.0	H	co	C
ATOM	5	CD2	LEU	A	60	78.671	-7.963	-11.132	0.0	H	co	C
ATOM	6	O	LEU	A	60	77.763	-7.028	-15.778	0.0	H	co	O
ATOM	7	CA	GLY	A	63	80.567	-7.435	-18.164	31.5	H	rm	C
ATOM	8	N	TRP	A	65	82.969	-8.397	-14.317	19.3	H	rm	N
ATOM	9	CB	TRP	A	65	82.252	-9.140	-12.098	19.3	H	rm	C
ATOM	10	C	TRP	A	65	82.935	-10.842	-13.793	19.3	H	rm	C
ATOM	11	O	TRP	A	65	82.948	-11.793	-13.012	19.3	H	rm	O
ATOM	12	N	HIS	A	66	82.670	-11.033	-15.100	39.8	H	su	N
ATOM	13	CA	HIS	A	66	82.287	-12.325	-15.659	39.8	H	su	C
ATOM	14	CB	HIS	A	66	81.866	-12.148	-17.141	39.8	H	su	C
ATOM	15	CB	LEU	A	69	82.273	-13.356	-10.014	31.6	H	rm	C
ATOM	16	CD	PRO	A	71	79.096	-14.930	-13.625	11.3	C	co	C
ATOM	17	CG	PRO	A	71	77.727	-14.363	-13.269	11.3	C	co	C
ATOM	18	OH	TYR	A	75	75.691	-9.963	-12.033	1.2	S	co	O
ATOM	19	O	LEU	A	131	78.065	-11.628	-8.947	9.5	C	co	O
TER	20		LEU	A	131							
HETATM	3	O	HOH	A	2004	77.948	-10.686	-13.444	3.3	6	O	
HETATM	3	O	HOH	A	2045	79.476	-9.082	-15.125	3.4	8	O	
HETATM	3	O	HOH	A	2047	84.703	-10.446	-17.620	3.4	3	O	
HETATM	3	O	HOH	A	2048	86.349	-10.838	-14.422	3.6	3	O	
END												

A typical cavity in PDB format (auto generated by the AWK script) along with residue specific accessibility (ASA), secondary structure (SS) and region (co, rm and su) specific values. There are about 2000 such PDB for all proteins (Table S1). These files were further analyzed for cavity types, inside and outside water frequency, region specificity using fully-automated script.

Region-specific non-bonded interaction

The three-dimensional structure of proteins is formed by weak force or non-bonded interactions such as hydrogen bonds, electrostatic (salt-bridge; ion-pair; π-cation; π-anion, etc.), hydrophobic (π-π; π-amide; π-σ; π--alkyl; alkyl-alkyl, etc.), etc. [8, 18-24]. Hydrogen bonds are formed with different types of donor and acceptor atoms (main-chain, side-chain polar, carbon, π-systems, and SW).

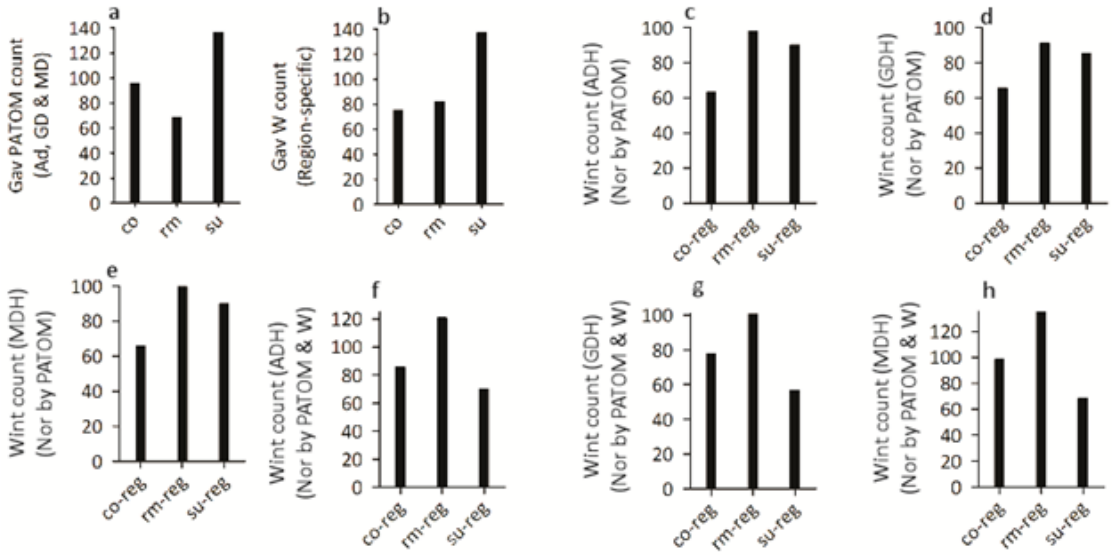


Figure 12. Protein region-specific overall frequencies of polar atoms.

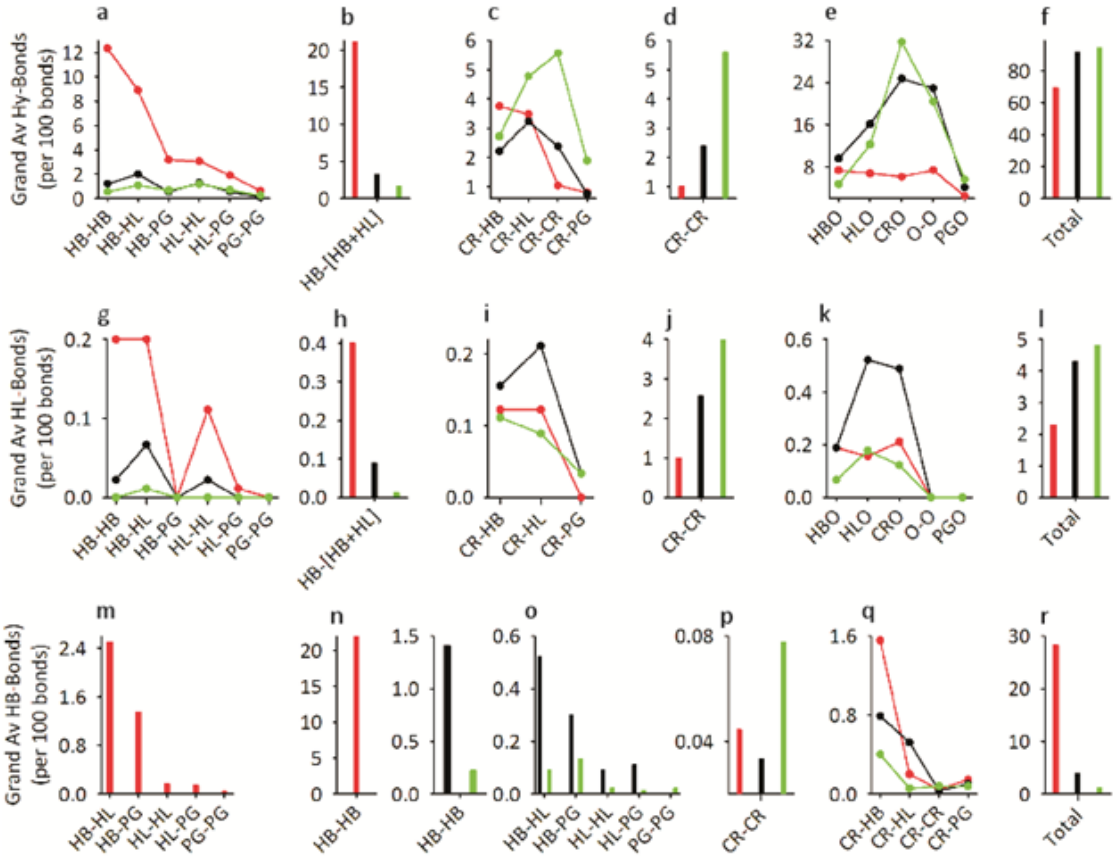


Figure 13. Normalized average nNon-bonded interactions.

Region specific grand average (irrespective of domains of life and enzyme classes) counts of protein polar atoms (PATOM) (**a**). Region specific grand average (irrespective of domains of life and enzyme classes) counts of total shell-waters (W) (**b**). PATOM normalized PATOM-

interacting Wint counts in ADH (**c**), GDH (**d**) and MDH (**e**). Both PATOM and W normalized Wint counts in ADH (**f**), GDH (**g**) and MDH (**h**).

Grand average frequency of hydrogen-bond (**a-f**), other hydrophilic (**g-l**) and hydrophobic (**m-s**) interactions regardless of enzyme classes and domains of life in different types of inter-residue connections (hydrophobic residue with hydrophobic residue i.e. HB-HB; hydrophobic residue with charged residue i.e. HB-CR, and, etc.). The interaction of each type is broadly divided into four groups namely hydrophobic and hydrophilic classes; charged class, and SW mediated, and total for each group. Some region-specific plots have been plotted together and some are plotted separately for the convenience of comparison.

Of the three regions of the protein (co, rm, and su), the grand average absolute frequency of protein atoms and SW is the highest in the su-region (Figure 12a-b). Notably, although rm is significantly less dominant in protein residues, residue-classes, and secondary structures than co and su, normalized (either by only protein atoms or both by protein atoms and total SW of protein) average SW is higher than even the su (Figure 12c-e and h-j). Here, we have compared the normalized grand average of different types of inter-residue interaction for the three regions. In the case of hydrogen bonds, in co, HB-HB, HB-HL, HB-PG, HL-HL, HL-PG, and PG-PG types dominate (Figure 13a and Table 13-14) over other regions. rm maintains an intermediate level, especially for HB-HB and HB-HL types (Figure 13b). On the other hand, except for CR-HB type, which is the highest in the co (Figure 13c), other CR types hydrogen bond (CR-HL, CR-CR, and CR-PG) are the highest in the su-region (Figure 13c, d, and Table 12-13). Here too, the rm-region is at an intermediate level (Figure 13d). Interestingly, in the case of SW-mediated hydrogen bonds, in the case of other types except for CRO, the rm again exceeds su's level so that the hydrogen bond levels of these two regions are equivalent (Figure 3e, f). In this case, it is clear that while co and su dominate in HB- and CR-mediated hydrogen bonds respectively, rm, in turn, acts as an intermediate. As far as electrostatic interactions (salt-bridge, ion-pair, cation-π, anion-π pi-sulfur, π-HB, etc.) are concerned, except for su dominating CR-CR type (Figure 3j), while HB-HB, HB-HL, and HL-HL types dominate in co (Figure 13g, h), CR-HB, CR-HL types are higher in the rm region (Figure 13i). In the case of SW-mediated electrostatic interactions, rm, as before, is dominant (Figure 13k). In this case, too, although rm is of intermediate level, the total interaction is almost equivalent to that of su (Figure 13l). Similarly, in hydrophobic interactions (π-π, π-σ, π-amide, π-alkyl, and alkyl-alkyl) especially for the HB-HB, HB-HL, HB-CR types co is highest, which is followed by the rm-region (Figure 13m-p). Notably, su and rm regions maintain the highest level of interaction for CR-CR and CR-HB types, respectively (Figure 13q, r). All in all, this rm region maintains an intermediate level here as well (Figure 13s). It should be noted here that the types that have low frequencies show a much lower level due to normalization based on the total interactions. We see that rm dominates in many cases in SW-mediated and mixed types (HB-CR and HB-HL etc) inter-residue interactions (Figure 13 and Table 13-14).

Table 13: ASA-region-specific interactions.

Total Int→	**2744**			**883**			**1631**		
ADH_ar_reg	**co**	**rm**	**su**	**co**	**rm**	**su**	**co**	**rm**	**su**
Types	Hydrogen bond category			Hydrophobic category			Hydrophilic Category		
HB-HB	354	7	6	589	8	5	6	0	0
PG-PG	17	2	1	2	0	1	0	0	0
HL-HL	68	16	9	10	1	1	2	0	0
CR-CR	22	21	100	3	0	1	24	21	77
HB-PG	100	2	6	40	4	2	0	0	0
HB-HL	204	11	21	86	6	4	11	2	0
HB-CR	109	23	32	35	5	5	6	2	2
HL-PG	60	10	8	8	2	0	0	0	0
CR-PG	33	7	30	8	2	3	0	2	2
CR-HL	104	23	63	6	10	2	5	2	1
HBO	189	107	76	0	0	0	3	1	1
PGO	73	37	102	0	0	0	0	0	0
HLO	168	133	179	0	0	0	7	12	4
CRO	175	206	552	0	0	0	7	2	4
O-O	210	196	331	0	0	0	0	0	0
Total Int→	**1886**	**801**	**1516**	**787**	**38**	**24**	**71**	**44**	**91**
Total Int	**2152**			**697**			**1488**		
GDH_ar_reg	co	rm	su	co	rm	su	co	rm	su
Types	Hydrogen bond category			Hydrophobic category			Hydrophilic Category		
HB-HB	216	11	6	424	19	5	1	0	0
PG-PG	2	1	1	5	0	1	0	0	0
HL-HL	81	6	11	0	2	1	1	0	0
CR-CR	62	24	105	2	0	2	65	19	77
HB-PG	54	4	8	29	2	6	0	0	0
HB-HL	154	30	11	62	2	3	1	1	0
HB-CR	101	19	42	50	6	9	3	1	2
HL-PG	26	4	6	4	0	0	0	0	0
CR-PG	22	4	32	0	0	0	0	0	1
CR-HL	129	14	53	8	1	0	4	0	1
HBO	149	71	60	0	0	0	2	2	1
PGO	41	33	72	0	0	0	0	0	0
HLO	138	107	149	0	0	0	1	0	5
CRO	168	176	551	0	0	0	1	1	1
O-O	146	137	266	0	0	0	0	0	0
Total Int→	**1489**	**641**	**1373**	**584**	**32**	**27**	**79**	**24**	**88**

Total Int	**1875**			**703**			**1900**		
MDH_ar_reg	co	rm	su	co	rm	su	co	rm	su
Types	Hydrogen bond category			Hydrophobic category			Hydrophilic Category		
HB-HB	267	12	12	462	4	5	4	0	0
PG-PG	30	0	7	0	0	0	0	0	0
HL-HL	47	6	19	2	0	0	1	0	0
CR-CR	10	11	126	0	0	2	7	18	61
HB-PG	69	6	10	32	1	0	0	0	0
HB-HL	194	14	12	29	6	0	4	0	0
HB-CR	51	19	50	19	7	7	3	2	0
HL-PG	42	2	16	3	1	0	0	0	0
CR-PG	12	3	46	0	0	2	0	0	0
CR-HL	33	30	109	2	5	4	2	5	7
HBO	126	40	63	0	0	0	1	0	2
PGO	66	22	73	0	0	0	0	0	0
HLO	104	99	185	0	0	0	1	3	0
CRO	97	205	637	0	0	0	3	8	0
O-O	152	174	445	0	0	0	0	0	0
Total Int→	**1300**	**643**	**1810**	**549**	**24**	**20**	**26**	**36**	**70**
Total Int	**3472**			**994**			**1379**		
ADH_ba_reg	co	rm	su	co	rm	su	co	rm	su
Types	Hydrogen bond category			Hydrophobic category			Hydrophilic Category		
HB-HB	455	9	3	866	11	3	7	0	0
PG-PG	27	3	4	0	0	0	0	0	0
HL-HL	84	10	18	10	0	0	4	0	0
CR-CR	30	30	103	1	2	2	33	33	77
HB-PG	144	3	14	51	10	3	0	0	0
HB-HL	282	14	7	80	5	0	11	0	0
HB-CR	134	27	30	40	12	6	10	1	2
HL-PG	64	4	10	3	3	1	3	0	0
CR-PG	40	15	32	7	4	2	0	0	0
CR-HL	120	25	77	5	2	0	3	1	0
HBO	211	89	86	0	0	0	8	1	1
PGO	93	53	65	0	0	0	0	0	0
HLO	249	106	162	0	0	0	12	0	5
CRO	175	285	400	0	0	0	10	2	3
O-O	200	234	263	0	0	0	0	0	0
Total	**2308**	**907**	**1274**	**1063**	**49**	**17**	**101**	**38**	**88**

Total Int→	**4008**			**1456**			**3032**		
GDH_ba_reg	co	rm	su	co	rm	su	co	rm	su
Types	Hydrogen bond category			Hydrophobic category			Hydrophilic Category		
HB-HB	440	17	23	795	20	2	8	0	0
PG-PG	7	0	3	0	0	0	0	0	0
HL-HL	196	13	30	3	0	0	7	1	1
CR-CR	50	45	133	3	5	2	23	41	112
HB-PG	94	10	19	33	4	0	0	0	0
HB-HL	379	19	31	114	6	2	8	0	0
HB-CR	163	23	76	81	12	11	6	3	0
HL-PG	55	4	15	1	1	0	0	0	0
CR-PG	19	11	42	10	0	0	0	0	2
CR-HL	202	65	143	13	7	8	8	1	3
HBO	286	122	168	0	0	0	9	3	4
PGO	72	52	209	0	0	0	0	0	0
HLO	350	266	405	0	0	0	6	9	1
CRO	238	339	928	0	0	0	12	14	3
O-O	317	343	656	0	0	0	0	0	0
Total	**2868**	**1329**	**2881**	**1053**	**55**	**25**	**87**	**72**	**126**
Total Int→Int	**7218**			**2594**			**6138**		
MDH_ba_reg	co	rm	su	co	rm	su	co	rm	su
Types	Hydrogen bond category			Hydrophobic category			Hydrophilic Category		
HB-HB	983	26	51	1669	33	24	9	0	0
PG-PG	64	0	29	1	0	0	0	0	0
HL-HL	131	14	63	10	2	0	2	0	0
CR-CR	25	53	230	1	0	7	48	62	183
HB-PG	234	9	81	136	9	11	0	0	0
HB-HL	686	38	121	124	8	1	14	0	0
HB-CR	220	39	249	89	25	11	6	8	5
HL-PG	148	6	58	17	3	1	0	0	0
CR-PG	33	21	113	6	1	2	0	0	1
CR-HL	123	73	278	12	6	0	2	6	3
HBO	603	243	297	0	0	0	12	9	7
PGO	224	109	335	0	0	0	0	0	0
HLO	451	407	768	0	0	0	17	19	7
CRO	466	698	1847	0	0	0	6	4	3
O-O	646	663	1352	0	0	0	0	0	0
Total	**5037**	**2399**	**5872**	**2065**	**87**	**57**	**116**	**108**	**209**

Total Int→	**7559**			**2629**			**4300**		
ADH_eu_reg	co	rm	su	co	rm	su	co	rm	su
Types	Hydrogen bond category			Hydrophobic category			Hydrophilic Category		
HB-HB	885	31	18	1626	30	4	29	0	0
PG-PG	54	2	11	2	1	1	0	0	0
HL-HL	225	52	27	15	0	1	6	1	0
CR-CR	56	45	208	8	0	3	75	52	173
HB-PG	285	19	12	75	3	2	0	0	0
HB-HL	642	38	33	182	10	6	16	2	0
HB-CR	273	49	73	121	18	10	16	8	3
HL-PG	168	17	23	13	5	1	0	0	0
CR-PG	105	22	62	16	8	7	1	0	2
CR-HL	189	79	169	10	6	2	5	4	2
HBO	618	312	198	0	0	0	23	4	1
PGO	236	114	243	0	0	0	0	0	0
HLO	513	436	502	0	0	0	13	17	4
CRO	441	569	1495	0	0	0	29	5	8
O-O	588	670	996	0	0	0	0	0	0
Total	**5278**	**2455**	**4070**	**2068**	**81**	**37**	**213**	**93**	**193**
Total Int→	**5211**			**1842**			**3480**		
GDH_eu_reg	co	rm	su	co	rm	su	co	rm	su
Types	Hydrogen bond category			Hydrophobic category			Hydrophilic Category		
HB-HB	610	26	7	1085	37	4	12	3	0
PG-PG	8	0	5	0	0	0	0	0	0
HL-HL	222	28	71	3	0	1	12	1	0
CR-CR	75	42	147	0	2	0	36	55	133
HB-PG	113	8	15	36	14	4	0	0	1
HB-HL	500	44	29	186	4	1	8	2	1
HB-CR	231	34	87	104	12	16	3	1	2
HL-PG	67	12	27	0	2	0	1	0	0
CR-PG	27	10	54	15	0	2	0	0	1
CR-HL	233	86	151	18	4	3	12	2	0
HBO	423	203	187	0	0	0	16	5	0
PGO	73	67	260	0	0	0	0	0	0
HLO	375	366	550	0	0	0	4	15	12
CRO	347	354	1060	0	0	0	8	14	6
O-O	348	389	643	0	0	0	0	0	0
Total	**3652**	**1669**	**3293**	**1447**	**75**	**31**	**112**	**98**	**156**

Total Int→	6160			1798			3801		
MDH_eu_reg	co	rm	su	co	rm	su	co	rm	su
Types	Hydrogen bond category			Hydrophobic category			Hydrophilic Category		
HB-HB	799	20	43	1463	30	9	15	0	0
PG-PG	38	3	16	0	0	0	0	0	0
HL-HL	139	37	96	6	0	1	11	0	0
CR-CR	46	35	215	2	1	2	57	36	160
HB-PG	197	13	37	114	4	8	0	0	0
HB-HL	618	45	76	113	9	2	17	1	2
HB-CR	210	44	155	94	17	23	9	2	8
HL-PG	171	12	45	21	1	0	0	0	0
CR-PG	21	11	81	4	4	1	0	1	0
CR-HL	156	48	260	12	8	0	7	3	4
HBO	467	155	144	0	0	0	12	4	0
PGO	127	74	189	0	0	0	0	0	0
HLO	418	344	546	0	0	0	8	4	8
CRO	372	452	988	0	0	0	8	11	5
O-O	408	369	677	0	0	0	0	0	0
Total	**4187**	**1662**	**3568**	**1829**	**74**	**46**	**144**	**62**	**187**

Non-bonded interactions for the co (core), rm (rim) and su (surface) accessibility regions of ADH, GDH and MDH of ar, ba and eu. These interactions are subcategorized as hydrogen bond (HyB), hydrophobic (hb) such as π-sigma (PS), π-π (PP), amide-π (AP), alkyl-alkyl (AL), π-alkyl (PA), and hydrophilic (hl) such as π-sulfur, π-cation (PC), π-anion (PA), π-hydrogen bond (PH), salt-bridge (SB), ion-pair (IP). Each of these interaction category is classified based on the protein's residue group such as hydrophobic-hydrophobic (HB-HB), Pro and Gly (PG-PG), Hydrophilic-Hydrophilic (HL-HL), Charged-charged (CR-CR), and other inter class residues interactions (HB-PG, HB-HL, HB-CR, HL-PG, CR-PG, CR-H). Hydrophobic-shell-water (HBO), PG-shell-water (PGO), Hydrophilic-shell-water (HLO), Charged-Shell-water (CRO), and shell-water-shell-water (O-O) types of interactions are also assessed. For each accessibility region specific PDB file (co-PDB, rm-PDB and su-PDB), all these distance dependent interactions were initially determined in the Biovia Discovery Studio Visualizer v20.1.0.19295. The interaction-detailed file thus obtained was formatted to analyze automatically using home-built AWK-script. The absolute interaction values (frequency of interaction) are placed in the table.

Table 14: domain-specific ASA-specific weak interactions.

HyB	arCOad	arCOgd	arCOmd	raRMad	arRMgd	arRMmd	arSUad	arSUgd	arSUmd
HB-HB	12.9	10	14.2	0.8	1.6	1.7	0.4	0.4	0.6
PG-PG	0.6	0.1	1.6	0.2	0.1	0	0.1	0.1	0.4
HL-HL	2.5	3.8	2.5	1.8	0.9	0.9	0.6	0.7	1
CR-CR	0.8	2.9	0.5	2.4	3.4	1.6	6.1	7.1	6.6

HB-PG	3.6	2.5	3.7	0.2	0.6	0.9	0.4	0.5	0.5
HB-HL	7.4	7.2	10.3	1.2	4.3	2	1.3	0.7	0.6
HB-CR	4	4.7	2.7	2.6	2.7	2.7	2	2.8	2.6
HL-PG	2.2	1.2	2.2	1.1	0.6	0.3	0.5	0.4	0.8
CR-PG	1.2	1	0.6	0.8	0.6	0.4	1.8	2.2	2.4
CR-HL	3.8	6	1.8	2.6	2	4.3	3.9	3.6	5.7
HBO	6.9	6.9	6.7	12.1	10.2	5.7	4.7	4	3.3
PGO	2.7	1.9	3.5	4.2	4.7	3.1	6.3	4.8	3.8
HLO	6.1	6.4	5.5	15.1	15.4	14.1	11	10	9.7
CRO	6.4	7.8	5.2	23.3	25.3	29.2	33.8	37	33.5
O-O	7.7	6.8	8.1	22.2	19.7	24.8	20.3	17.9	23.4

HyB	baCOad	baCOgd	baCOmd	baRMad	baRMgd	baRMmd	baSUad	baSUgd	baSUmd
HB-HB	13.1	11	13.6	0.9	1.2	1	0.2	0.8	0.8
PG-PG	0.8	0.2	0.9	0.3	0	0	0.3	0.1	0.5
HL-HL	2.4	4.9	1.8	1	0.9	0.5	1.3	1	1
CR-CR	0.9	1.2	0.3	3	3.1	2	7.5	4.4	3.7
HB-PG	4.1	2.3	3.2	0.3	0.7	0.3	1	0.6	1.3
HB-HL	8.1	9.5	9.5	1.4	1.3	1.5	0.5	1	2
HB-CR	3.9	4.1	3	2.7	1.6	1.5	2.2	2.5	4.1
HL-PG	1.8	1.4	2.1	0.4	0.3	0.2	0.7	0.5	0.9
CR-PG	1.2	0.5	0.5	1.5	0.8	0.8	2.3	1.4	1.8
CR-HL	3.5	5	1.7	2.5	4.5	2.8	5.6	4.7	4.5
HBO	6.1	7.1	8.4	9	8.4	9.4	6.2	5.5	4.8
PGO	2.7	1.8	3.1	5.3	3.6	4.2	4.7	6.9	5.5
HLO	7.2	8.7	6.2	10.7	18.3	15.7	11.7	13.3	12.5
CRO	5	5.9	6.5	28.7	23.3	26.9	29	30.6	30.1
O-O	5.8	7.9	8.9	23.5	23.6	25.6	19.1	21.6	22

HyB	euCOad	euCOgd	euCOmd	euRMad	euRMgd	euRMmd	euSUad	euSUgd	euSUmd
HB-HB	11.7	11.7	13	1.2	1.4	1.1	0.4	0.2	1.1
PG-PG	0.7	0.2	0.6	0.1	0	0.2	0.3	0.1	0.4
HL-HL	3	4.3	2.3	2	1.5	2.1	0.6	2	2.5
CR-CR	0.7	1.4	0.7	1.7	2.3	1.9	4.8	4.2	5.7
HB-PG	3.8	2.2	3.2	0.7	0.4	0.7	0.3	0.4	1
HB-HL	8.5	9.6	10	1.4	2.4	2.5	0.8	0.8	2
HB-CR	3.6	4.4	3.4	1.9	1.8	2.4	1.7	2.5	4.1
HL-PG	2.2	1.3	2.8	0.6	0.7	0.7	0.5	0.8	1.2
CR-PG	1.4	0.5	0.3	0.8	0.5	0.6	1.4	1.6	2.1

CR-HL	2.5	4.5	2.5	3	4.7	2.7	3.9	4.3	6.8
HBO	8.2	8.1	7.6	11.9	11	8.6	4.6	5.4	3.8
PGO	3.1	1.4	2.1	4.3	3.6	4.1	5.7	7.5	5
HLO	6.8	7.2	6.8	16.6	19.9	19.1	11.7	15.8	14.4
CRO	5.8	6.7	6	21.6	19.2	25.1	34.8	30.5	26
O-O	7.8	6.7	6.6	25.5	21.1	20.5	23.2	18.5	17.8

HL	arCOad	arCOgd	arCOmd	arRMad	arRMgd	arRMmd	arSUad	arSUgd	arSUmd
HB-HB	0.2	0	0.2	0	0	0	0	0	0
PG-PG	0	0	0	0	0	0	0	0	0
HL-HL	0.1	0	0.1	0	0	0	0	0	0
CR-CR	0.9	3	0.4	2.4	2.7	2.5	4.8	5.2	3.2
HB-PG	0	0	0	0	0	0	0	0	0
HB-HL	0.4	0	0.2	0.2	0.1	0	0	0	0
HB-CR	0.1	0.1	0.3	0.2	0.1	0.2	0.2	0.2	0
HL-PG	0	0	0	0	0	0	0	0	0
CR-PG	0	0	0	0.2	0	0	0.1	0.1	0
CR-HL	0.1	0.1	0.2	0.2	0	0.7	0.1	0.1	0.4
HBO	0.1	0.1	0.1	0.1	0.3	0	0.1	0.1	0.1
PGO	0	0	0	0	0	0	0	0	0
HLO	0.3	0	0.1	1.4	0	0.4	0.2	0.3	0
CRO	0.3	0	0.2	0.2	0.1	1.1	0.2	0.1	0
O-O	0	0	0	0	0	0	0	0	0

HL	baCOad	baCogd	baCOmd	baRMad	baRMgd	baRMmd	baSUad	baSUgd	baSUmd
HB-HB	0.2	0.2	0.1	0	0	0	0	0	0
PG-PG	0	0	0	0	0	0	0	0	0
HL-HL	0.1	0.2	0	0	0.1	0	0	0	0
CR-CR	1	0.6	0.6	3.3	2.8	2.3	5.6	3.6	2.9
HB-PG	0	0	0	0	0	0	0	0	0
HB-HL	0.3	0.1	0.2	0	0	0	0	0	0
HB-CR	0.3	0.1	0	0.1	0.2	0.3	0.2	0	0.1
HL-PG	0.1	0	0	0	0	0	0	0	0
CR-PG	0	0	0	0	0	0	0	0.1	0
CR-HL	0.1	0.2	0	0.1	0.1	0.3	0	0.1	0
HBO	0.2	0.2	0.2	0.1	0.2	0.3	0.1	0.1	0.1
PGO	0	0	0	0	0	0	0	0	0
HLO	0.3	0.1	0.2	0	0.6	0.7	0.4	0	0.1
CRO	0.3	0.3	0.1	0.2	1	0.2	0.2	0.1	0
O-O	0	0	0	0	0	0	0	0	0

HL	euCOad	euCOgd	euCOmd	euRMad	euRMgd	euRMmd	euSUad	euSUgd	euSUmd
HB-HB	0.4	0.3	0.2	0	0.2	0	0	0	0
PG-PG	0	0	0	0	0	0	0	0	0
HL-HL	0.1	0.2	0.2	0	0.1	0	0	0	0
CR-CR	1	0.6	0.9	2	3.1	2	4.1	3.8	4.2
HB-PG	0	0	0	0	0	0	0	0	0
HB-HL	0.2	0.2	0.2	0	0.2	0.1	0	0	0.1
HB-CR	0.1	0	0.1	0.1	0.1	0.1	0	0	0.3
HL-PG	0	0	0	0	0	0	0	0	0
CR-PG	0	0	0	0	0	0.1	0	0	0
CR-HL	0.1	0.2	0.1	0.1	0.2	0.2	0	0	0.1
HBO	0.3	0.3	0.2	0.2	0.3	0.2	0	0	0
PGO	0	0	0	0	0	0	0	0	0
HLO	0.2	0.1	0.1	0.6	0.8	0.2	0.1	0.3	0.2
CRO	0.4	0.2	0.1	0.2	0.8	0.6	0.2	0.2	0.1
O-O	0	0	0	0	0	0	0	0	0

HB	arCOad	arCOgd	arCOmd	arRMad	arRMgd	arRMmd	arSUad	arSUgd	arSUmd
HB-HB	21.5	19.6	24.7	0.9	2.6	0.5	0.4	0.4	0.3
PG-PG	0.1	0.2	0	0	0	0	0.1	0.1	0
HL-HL	0.4	0	0.2	0.1	0.3	0	0.1	0.1	0
CR-CR	0.1	0.1	0	0	0	0	0.1	0.1	0.1
HB-PG	1.4	1.3	1.7	0.4	0.3	0.1	0.1	0.4	0
HB-HL	3.1	2.9	1.7	0.7	0.3	0.8	0.3	0.2	0
HB-CR	1.3	2.2	1	0.5	0.9	1	0.3	0.6	0.4
HL-PG	0.3	0.1	0.2	0.2	0	0.1	0	0	0
CR-PG	0.3	0	0	0.2	0	0	0.2	0	0.2
CR-HL	0.2	0.3	0.2	1.1	0.1	0.7	0.1	0	0.2
HBO	0	0	0	0	0	0	0	0	0
PGO	0	0	0	0	0	0	0	0	0
HLO	0	0	0	0	0	0	0	0	0
CRO	0	0	0	0	0	0	0	0	0
O-O	0	0	0	0	0	0	0	0	0

HB	baCOad	baCOgd	baCOmd	baRMad	baRMgd	baRMmd	baSUad	baSUgd	baSUmd
HB-HB	25	19.7	23.2	1.1	1.4	1.3	0.2	0.1	0.3
PG-PG	0	0	0	0	0	0	0	0	0
HL-HL	0.3	0.1	0.2	0	0	0.1	0	0	0

CR-CR	0	0.1	0	0.2	0.4	0	0.2	0.1	0.1
HB-PG	1.5	0.8	1.9	1	0.3	0.3	0.2	0	0.2
HB-HL	2.3	2.9	1.7	0.5	0.4	0.3	0	0.1	0
HB-CR	1.2	1.9	1.2	1.2	0.9	0.9	0.5	0.4	0.2
HL-PG	0.1	0	0.2	0.3	0.1	0.1	0.1	0	0
CR-PG	0.2	0.2	0	0.4	0	0	0.2	0	0
CR-HL	0.1	0.3	0.1	0.2	0.5	0.2	0	0.2	0
HBO	0	0	0	0	0	0	0	0	0
PGO	0	0	0	0	0	0	0	0	0
HLO	0	0	0	0	0	0	0	0	0
CRO	0	0	0	0	0	0	0	0	0
O-O	0	0	0	0	0	0	0	0	0

HB	**euCOad**	**euCOgd**	**euCOmd**	**euRMad**	**euRMgd**	**euRMmd**	**euSUad**	**euSUgd**	**euSUmd**
HB-HB	21.6	20.8	23.8	1.1	2	1.6	0	0.1	0.2
PG-PG	0	0	0	0	0	0	0	0	0
HL-HL	0.2	0.1	0	0	0	0	0	0	0
CR-CR	0.1	0	0	0	0.2	0.1	0	0	0
HB-PG	1	0.7	1.8	0.1	0.8	0.2	0	0.1	0.2
HB-HL	2.4	3.6	1.9	0.4	0.2	0.5	0.1	0	0.1
HB-CR	1.7	2	1.5	0.7	0.6	1	0.2	0.4	0.6
HL-PG	0.1	0	0.3	0.2	0.1	0.1	0	0	0
CR-PG	0.2	0.3	0.1	0.3	0	0.2	0.1	0	0
CR-HL	0.1	0.3	0.2	0.2	0.2	0.5	0	0	0
HBO	0	0	0	0	0	0	0	0	0
PGO	0	0	0	0	0	0	0	0	0
HLO	0	0	0	0	0	0	0	0	0
CRO	0	0	0	0	0	0	0	0	0
O-O	0	0		0	0	0	0	0	0

Non-bonded interactions for the co (core), rm (rim) and su (surface) accessibility regions of ADH (ad), GDH (gd) and MDH (md) of ar, ba and eu. These interactions are subcategorized as hydrogen bond (HyB), hydrophobic (hb) such as π-sigma (PS), π-π (PP), amide-π (AP), alkyl-alkyl (AL), π-alkyl (PA), and hydrophilic (hl) such as π-sulfur, π-cation (PC), π-anion (PA), π-hydrogen bond (PH), salt-bridge (SB), ion-pair (IP). Each of these interaction categories is classified based on the protein's residue group such as hydrophobic-hydrophobic (HB-HB), Pro and Gly (PG-PG), Hydrophilic-Hydrophilic (HL-HL), Charged-charged (CR-CR), and other inter class residues interactions (HB-PG, HB-HL, HB-CR, HL-PG, CR-PG, CR-H). Hydrophobic-shell-water (HBO), PG-shell-water (PGO), Hydrophilic-shell-water (HLO), Charged-Shell-water (CRO), and shell-water-shell-water (O-O) types of interactions are also assessed. For each accessibility region specific PDB file (co-PDB, rm-PDB and su-PDB), all

these distance dependent interactions were initially determined in the Biovia Discovery Studio Visualizer v20.1.0.19295. The interaction-detailed file thus obtained was formatted to analyze automatically using home-built AWK-script. The absolute interaction values (frequency of interaction) are normalized using the total frequency of interaction. For different domains of life (ar, ba, and eu) the number of PDB file are different and thus, the Ti value is different. Such a variation is normalized using Ti value for comparison purpose.

Discussion

Rim has a distinct composition from the core and surface regions of the protein structure

To gain insights into the pattern of globular protein structures and interactions (Rose et al., 1985; Gutheil et al. 1992; Lins et al., 2003; Sen et al., 2017; Bandyopadhyay et al., 2019; Islam et al., 2019; Bandyopadhyay et al., 2020; Biswas et al., 2020; Banerjee et al., 2021; Roy et al., 2023), we have done the current study of three protein families of three domains of life. Our observation implies that the core and surface of the globular proteins have strong preferences towards hydrophobic and charged classes, respectively. Moreover, the polar class appears to be equally likely for these two regions. We know that the polar, like hydrophobic residues, are equally important in the core's organization (Bolon & Mayo, 2001). At the same time, some polar residues (e.g., Ser) play an important role in protein solubility, like the charged class (Trevino et al., 2007). However, the rim area's KD-neutral (Fleming et al., 2006) nature may indicate that it is a non-preferential region with an equal number of hydrophobic and hydrophilic residues. The existence of such a region was speculated earlier (Lins et al., 2003; Sen Gupta et al., 2017). Although compared to the rim, these differences for the core and the surface are significant in all enzyme cases, the observed variation in its level seems to be originating from the divergence in domain-specific orthologous sequences. Taken together, the existence of the rim between the core and the surface of these three enzyme structures appears to be a general characteristic of globular proteins.

Rim is topologically and conformationally distinct from the core and the surface

The topology of the tertiary structure of the protein is largely determined by the secondary structure (Islam et al., 2018; Mitra et al., 2019). Furthermore, because the strand is extended and the helix is compact (Bolon & Mayo, 2001), these two types of secondary structure may influence protein residue packing. The significant preference for the strand and helix in the core may indicate that they contribute to the core's balance of rigidity and flexibility. At the same time, the preference for coil and helix on the surface could play a comparable role. The preference for secondary structures in the core and on the surface appears to be general phenomena. In the context of protein function, it appears that ar, ba, and eu enzymes preserve the topological pattern of their structures for these accessible regions (core and surface) via creating combinational variations in secondary structures. Notably, the predominance of such distinctive secondary structural preference strategies is less apparent in the rim region. This

could imply that the rim zone serves as a balance between these two regions (co and su) with alternate and similar preferences (for helix) for secondary structures.

Compared to the surface, the interior cavities are more distinctive of the core and rim

The internal cavity, like residue classes and secondary structures, appears to play a crucial role in the region-specific structural organization of globular proteins. To illustrate their numerical abundance, a 100-residue protein may have six to eight cavities. Our findings suggest that the cavity is closely related to the stability of the core in particular. This is most likely why charged residue, which has a surface propensity, is significantly less concentrated in the cavity. These cavities, like the secondary structure, appear to be higher-level, confined structures in the protein folding path. These, on the one hand, limit the movement of secondary structure elements (e.g., helixes) and SW by incorporating them into their structure; while on the other hand, these stability units appear to operate as a limiting step in spontaneous residue packing during the folding process (Eisenberg, 2003; Sadqi et al., 2003). In other words, without these structural units, the protein's core, in particular, could have been significantly more compact, as it contains two-thirds of the total cavities. Most cavities (~70%) trap shell-water to reinforce their structure from inside and yet retain space, particularly in compact cores. This approach is most likely used by Archaea to compact their core structure, as the water-filled cavity is substantially lower than in other domains. Archaea's protein folding environment (such as high temperature) appears to improve the mobility of water and protein components, as opposed to a shell-water trap in the cavity. Because the rim region contains the most shell-water, the water content of the cavity increases when an atom from this region is present. This propensity most likely causes heterogeneity in the cavity at the residue, residue-class, and secondary structure levels. The helix specifies the compact structure, while the cavity includes helix residues in it. Again, hydrophobic residue, which plays an important role in hydrophobic collapse, has been the primary constituent for cavity. According to these events, just as an extended beta-structure with a strong tendency for core provides flexibility at the core, the cavity, with space within itself and the inclusion of compact structure units i.e., helix and hydrophobic residues into its structure, brings a balance of flexibility and rigidity to the core. Regardless of domains of life and enzyme classes, the correlations of shell-water and its interactions may imply that the interior cavity is a generalized and pattern-wise structural feature in globular proteins.

The rim is an interaction bridge between the core and the surface

Hydrogen bonding creates a complex interaction network while preserving the distinct properties of these three accessible zones. While the core and surface have peaks in hydrophobic (also hydrophilic) and charged residue-mediated hydrogen bonding, respectively, the rim has an intermediate level in both situations, indicating that the latter region contains both types of residue. In the case of other hydrophilic and hydrophobic interactions (Nayek et al., 2015), the presence of identical events in these three zones lends credence to the theory that

the rim exists as a distinct mixed entity between the core and surface. Furthermore, in the case of shell-water interactions, the fact that the rim is often superior to the core and surface demonstrates the abundance of SW here. Taken together, it seems that the position of the residue light but mixed and SW abundant rim between the strong residue-class trend core and the surface, on the one hand, supports their structures and, on the other hand, makes the protein dynamic.

Conclusion

The core and surface of globular proteins, which are KD-positive and KD-negative, respectively, have a strong tendency towards residue types, its classes, and secondary structure elements. The rim, on the other hand, which lies between these two regions, is KD-neutral; residue-wise light, mixed, and abundance in SW. SW filled interior cavities, which are abundant in the core and rim regions in particular, act as a sub-structural entities to limit residue compaction. The predominance of cavities with space within itself and extended β-structures in the core, and the predominance of the flexible coil on the surface seem to lower the rigidity in these regions. In non-bonded interactions, the core and surface are superior to hydrophobic and charged-mediated interactions, respectively, but the rim is at the intermediate level in both cases. The non-bonded interactions between these two types of residues are highest in their frequency in the rim region. This distinct entity of the rim seems to help maintain the structure of the core and the surface, on the one hand, and the dynamism of the protein, on the other. Our study finds applications in protein folding and protein bioinformatics.

Although our study demonstrates the existence of RIM, we believe that it is necessary to analyze it for many more proteins in the database.

Acknowledgment

We are grateful for the computational facility laboratory of the Department of Biotechnology of the University of Burdwan. We like to thank Rifat Nowaz UL Islam for his technical helps in this work.

Competing interests:

The authors declare no competing interests.

Abbreviation: hb, hydrophobic residue; cr, charged residue; po, polar residue; ar, Archaea; ba, Bacteria; eu, Eukaryote; SW, shell-water; co, core; rm, rim; su, surface; ss, secondary structure

Data availability:

While the coordinate files are available in the Protein Data Bank, these structure files were minimized for their use in the study. Minimized files are available upon request. There are also approximately 2000 cavity files in PDB format along with residue atoms' accessibility, secondary structure, and interaction multiplicity details. These PDB files can also be obtained upon request.

Code availability:

The binary version of the fully automated AWK scripts was used for the analysis of accessibility region-specific residue composition, residue class, and secondary structure and, KD value, cavity properties. Scripts are also used for the analysis of cavity PDB format files for their composition, residue propensity, types, and shell-water properties. While we have tested the program to be fully functional, error-free, and user-friendly for the proteins used here, we are presently continuing our verification works. The binary version of these could be obtained by experts working in the field upon request.

Reference:

Agashe, V. R., Shastry, M. C. R., & Udgaonkar, J. B. (1995). Initial hydrophobic collapse in the folding of barstar. *Nature*, *377*(6551), 754-757. https://doi.org/10.1038/377754a0

Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science*, *181*(4096), 223-230. https://doi.org/10.1126/science.181.4096.223

Baker, E. N., & Hubbard, R. E. (1984). Hydrogen bonding in globular proteins. *Progress in biophysics and Molecular Biology*, *44*(2), 97-179. https://doi.org/10.1016/0079-6107(84)90007-5

Bandyopadhyay, A. K., & Sonawat, H. M. (2000). Salt dependent stability and unfolding of [Fe2-S2] ferredoxin of *Halobacterium salinarum*: spectroscopic investigations. *Biophysical journal*, *79*(1), 501-510. https://doi.org/10.1016/S0006-3495(00)76312-0

Bandyopadhyay, A. K., Islam, R. N. U., & Hazra, N. (2020). Salt-bridges in the microenvironment of stable protein structures. *Bioinformation*, *16*(11), 900. https://doi.org/10.6026/97320630016900

Bandyopadhyay, A. K., Islam, R. N. U., Mitra, D., Banerjee, S., & Goswami, A. (2019). Structural insights from water-ferredoxin interaction in mesophilic algae and halophilic archaea. *Bioinformation*, *15*(2), 79. https://doi.org/10.6026/97320630015079

Bandyopadhyay, A. K., Islam, R. N. U., Mitra, D., Banerjee, S., & Goswami, A. (2019). Stability of buried and networked salt-bridges (BNSB) in thermophilic proteins. *Bioinformation*, *15*(1), 61. https://doi.org/10.6026/97320630015061

Bandyopadhyay, A. K., Islam, R. N. U., Mitra, D., Banerjee, S., Yasmeen, S., & Goswami, A. (2019). Insights from the salt bridge analysis of malate dehydrogenase from *H. salinarum* and *E. coli*. *Bioinformation*, *15*(2), 95. https://doi.org/10.6026/97320630015095

Bandyopadhyay, A. K., Krishnamoorthy, G., & Sonawat, H. M. (2001). Structural stabilization of [2Fe-2S] ferredoxin from *Halobacterium salinarum*. *Biochemistry*, *40*(5), 1284-1292. https://doi.org/10.1021/bi001614j

Bandyopadhyay, A. K., Krishnamoorthy, G., Padhy, L. C., & Sonawat, H. M. (2007). Kinetics of salt-dependent unfolding of [2Fe–2S] ferredoxin of *Halobacterium salinarum*. *Extremophiles*, *11*, 615-625. https://doi.org/10.1007/s00792-007-0075-0

Banerjee, S., Gupta, P. S. S., Islam, R. N. U., & Bandyopadhyay, A. K. (2021). Intrinsic basis of thermostability of prolyl oligopeptidase from *Pyrococcus furiosus*. *Scientific Reports*, *11*(1), 11553. https://doi.org/10.1038/s41598-021-90723-4

Banerjee, S., Gupta, P. S. S., Nayek, A., Das, S., Sur, V. P., Seth, P., ... & Bandyopadhyay, A. K. (2015). PHYSICO2: an UNIX based standalone procedure for computation of physicochemical, window-dependent and substitution based evolutionary properties of protein sequences along with automated block preparation tool, version 2. *Bioinformation*, *11*(7), 366. https://doi.org/10.6026/97320630011366

Banerjee, S., Mondal, B., Islam, R. N. U., Gupta, P. S. S., Mitra, D., & Bandyopadhyay, A. (2018). POWAINDv1. 0: A Program for Protein-Water Interactions Determination. *Bioinformation*, *14*(9), 530. https://doi.org/10.6026/97320630014530

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., ... & Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Research*, *28*(1), 235-242. https://doi.org/10.1093/nar/28.1.235

Betts, M. J., & Russell, R. B. (2003). Amino acid properties and consequences of substitutions. *Bioinformatics for geneticists*, 289-316. https://doi.org/10.1002/0470867302.ch14

Biswas, I., Mitra, D., Bandyopadhyay, A. K., & Mohapatra, P. K. D. (2020). Contributions of protein microenvironment in tannase industrial applicability: An in-silico comparative study of pathogenic and non-pathogenic bacterial tannase. *Heliyon*, *6*(11). https://doi.org/10.1016/j.heliyon.2020.e05359

Bolon, D. N., & Mayo, S. L. (2001). Polar residues in the protein core of Escherichia coli thioredoxin are important for fold specificity. *Biochemistry*, *40*(34), 10047-10053. https://doi.org/10.1021/bi010427y

Bottini, S., Bernini, A., De Chiara, M., Garlaschelli, D., Spiga, O., Dioguardi, M., ... & Niccolai, N. (2013). ProCoCoA: a quantitative approach for analyzing protein core composition. *Computational biology and chemistry*, *43*, 29-34. https://doi.org/10.1016/j.compbiolchem.2012.12.007

Ceccarelli, C., Liang, Z. X., Strickler, M., Prehna, G., Goldstein, B. M., Klinman, J. P., & Bahnson, B. J. (2004). Crystal structure and amide H/D exchange of binary complexes of alcohol dehydrogenase from *Bacillus stearothermophilus*: insight into thermostability and cofactor binding. *Biochemistry*, *43*(18), 5266-5277. https://doi.org/10.1021/bi049736p

Dalhus, B., Saarinen, M., Sauer, U. H., Eklund, P., Johansson, K., Karlsson, A., ... & Eklund, H. (2002). Structural basis for thermophilic protein stability: structures of thermophilic and mesophilic malate dehydrogenases. *Journal of Molecular Biology*, *318*(3), 707-721. https://doi.org/10.1016/S0022-2836(02)00050-5

Dill, K. A. (1990). Dominant forces in protein folding. *Biochemistry*, *29*(31), 7133-7155. https://doi.org/10.1021/bi00483a001

Dill, K. A., & MacCallum, J. L. (2012). The protein-folding problem, 50 years on. *Science*, *338*(6110), 1042-1046. https://doi.org/10.1126/science.1219021

Eisenberg, D. (2003). The discovery of the α-helix and β-sheet, the principal structural features of proteins. *Proceedings of the National Academy of Sciences*, *100*(20), 11207-11210. https://doi.org/10.1073/pnas.2034522100

Ernst, J. A., Clubb, R. T., Zhou, H. X., Gronenborn, A. M., & Clore, G. M. (1995). Demonstration of positionally disordered water within a protein hydrophobic cavity by NMR. *Science*, *267*(5205), 1813-1817. https://doi.org/10.1126/science.7892604

Finney, J. L. (1977). The organization and function of water in protein crystals. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *278*(959), 3-32. https://doi.org/10.1098/rstb.1977.0029

Fleming, P. J., Gong, H., & Rose, G. D. (2006). Secondary structure determines protein topology. *Protein Science*, *15*(8), 1829-1834. https://doi.org/10.1110/ps.062305106

Flower, D. R., North, A. C., & Sansom, C. E. (2000). The lipocalin protein family: structural and sequence overview. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*, *1482*(1-2), 9-24. https://doi.org/10.1016/S0167-4838(00)00148-5

Gupta, P. S. S., Banerjee, S., Islam, R. N. U., Mondal, S., Mondal, B., & Bandyopadhyay, A. K. (2014a). PHYSICO: An UNIX based Standalone Procedure for Computation of Individual and Group Properties of Protein Sequences. *Bioinformation*, *10*(2), 105. https://doi.org/10.6026/97320630010105

Gupta, P. S. S., Mondal, S., Mondal, B., Islam, R. N. U., Banerjee, S., & Bandyopadhyay, A. K. (2014b). SBION: A program for analyses of salt-bridges from multiple structure files. *Bioinformation*, *10*(3), 164. https://doi.org/10.6026/97320630010164

Gupta, P. S. S., Nayek, A., Banerjee, S., Seth, P., Das, S., Sur, V. P., ... & Bandyopadhyay, A. K. (2015). SBION2: Analyses of salt bridges from multiple structure files, version 2. *Bioinformation*, *11*(1), 39. https://doi.org/10.6026/97320630011039

Gutheil, W. G., Holmquist, B., & Vallee, B. L. (1992). Purification, characterization, and partial sequence of the glutathione-dependent formaldehyde dehydrogenase from Escherichia coli: a class III alcohol dehydrogenase. *Biochemistry*, *31*(2), 475-481. https://doi.org/10.1021/bi00117a025

Islam, R. N. U., Mitra, D., Gupta, P. S. S., Banerjee, S., Mondal, B., & Bandyopadhyay, A. K. (2018). AUTOMINv1. 0: an automation for minimization of Protein Data Bank files and its usage. *Bioinformation*, *14*(9), 525. https://doi.org/10.6026/97320630014525

Islam, R. N. U., Roy, C., Gupta, P. S. S., Banerjee, S., Mitra, D., Banerjee, S., & Bandyopadhyay, A. K. (2018). PROPAB: Computation of Propensities and Other Properties from Segments of 3D structure of Proteins. *Bioinformation*, *14*(5), 190. https://doi.org/10.6026/97320630014190

Islam, R. N. U., Roy, C., Mitra, D., Bandyopadhyay, A. K., Yasmeen, S., & Banerjee, S. (2019). Salt-bridge, aromatic-aromatic, cation-π and anion-π interactions in proteins from different domains of life. In *Biotechnology and Biological Sciences* (pp. 152-158). CRC Press. https://doi.org/10.1201/9781003001614-25

Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., & Phillips, D. C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, *181*(4610), 662-666. https://doi.org/10.1038/181662a0

Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, *157*(1), 105-132. https://doi.org/10.1016/0022-2836(82)90515-0

Lim, V. I. (1974). Structural principles of the globular organization of protein chains. A stereochemical theory of globular protein secondary structure. *Journal of Molecular Biology*, *88*(4), 857-872. https://doi.org/10.1016/0022-2836(74)90404-5

Lins, L., Thomas, A., & Brasseur, R. (2003). Analysis of accessible surface of residues in proteins. *Protein Science*, *12*(7), 1406-1417. https://doi.org/10.1110/ps.0304803

Madern, D. (2002). Molecular evolution within the L-malate and L-lactate dehydrogenase super-family. *Journal of molecular evolution*, *54*(6), 825-840. https://doi.org/10.1007/s00239-001-0088-8

Martinez, C. R., & Iverson, B. L. (2012). Rethinking the term "pi-stacking". *Chemical Science*, *3*(7), 2191-2201. https://doi.org/10.1039/c2sc20045g

Mecozzi, S., West Jr, A. P., & Dougherty, D. A. (1996). Cation-pi interactions in aromatics of biological and medicinal interest: electrostatic potential surfaces as a useful qualitative guide. *Proceedings of the National Academy of Sciences*, *93*(20), 10566-10571. https://doi.org/10.1073/pnas.93.20.10566

Mitra, D., Bandyopadhyay, A. K., Islam, R. N. U., Banerjee, S., Yasmeen, S., & Mohapatra, P. K. D. (2019). Structure, salt-bridge's energetics and microenvironments of nucleoside diphosphate kinase from halophilic, thermophilic and mesophilic microbes. In *Biotechnology and Biological Sciences* (pp. 107-113). CRC Press. https://doi.org/10.1201/9781003001614-18

Moon, J. H., Lee, H. J., Park, S. Y., Song, J. M., Park, M. Y., Park, H. M., ... & Kim, J. S. (2011). Structures of iron-dependent alcohol dehydrogenase 2 from *Zymomonas mobilis* ZM4 with and without NAD+ cofactor. *Journal of molecular biology*, *407*(3), 413-424. https://doi.org/10.1016/j.jmb.2011.01.045

Nayek, A., Banerjee, S., Sen Gupta, P. S., Sur, B. P., Seth, P., Das, S., ... & Bandyopadhyay, A. K. (2015). ADSETMEAS: Automated determination of salt-bridge energy terms and micro environment from atomic structures using APBS method, version 1.0. In *Protein Science, 24*, 216-217. 111 RIVER ST, HOBOKEN 07030-5774, NJ USA: WILEY-BLACKWELL.

Nayek, A., Gupta, P. S. S., Banerjee, S., Sur, V. P., Seth, P., Das, S., ... & Bandyopadhyay, A. K. (2015b). ADSBET2: Automated determination of salt-bridge energy-terms version 2. *Bioinformation*, *11*(8), 413. https://doi.org/10.6026/97320630011413

Nayek, A., Gupta, P. S., Banerjee, S., Das, S., Sur, V. P., Seth, P., ... & Bandyopadhyay, A. K. (2015a). ADSBET: Automated determination of salt-bridge energy terms. *Int. J. Inst. Pharm. Life Sci.*, *5*, 28-36.

Nayek, A., Sen Gupta, P. S., Banerjee, S., Mondal, B., & Bandyopadhyay, A. K. (2014). Salt-bridge energetics in halophilic proteins. *Plos One*, *9*(4), e93862. https://doi.org/10.1371/journal.pone.0093862

Nayek, A., Sen Gupta, P. S., Banerjee, S., Mondal, B., & Bandyopadhyay, A. K. (2014). Salt-bridge energetics in halophilic proteins. *Plos one*, *9*(4), e93862. https://doi.org/10.1371/journal.pone.0093862

Petitjean, C., Deschamps, P., López-García, P., & Moreira, D. (2015). Rooting the Domain Archaea by Phylogenomic Analysis Supports the Foundation of the New Kingdom Proteoarchaeota, *Genome Biology and Evolution*, *7*(1), 191–204. https://doi.org/10.1093/gbe/evu274

Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H., & Zehfus, M. H. (1985). Hydrophobicity of amino acid residues in globular proteins. *Science*, *229*(4716), 834-838. https://doi.org/10.1126/science.4023714

Rovers, E., & Schapira, M. (2022). Methods for computer-assisted PROTAC design. *Methods in Enzymology*, *690*, 311-340. https://doi.org/10.1016/bs.mie.2023.06.020

Roy, C., Islam, R. N. U., Banerjee, S., & Bandyopadhyay, A. K. (2023). Underlying features for the enhanced electrostatic strength of the extremophilic malate dehydrogenase interface salt-bridge compared to the mesophilic one. *Journal of Biomolecular Structure and Dynamics*, 1-16. https://doi.org/10.1080/07391102.2023.2295972

Sadqi, M., Lapidus, L. J., & Munoz, V. (2003). How fast is protein hydrophobic collapse? *Proceedings of the National Academy of Sciences*, *100*(21), 12117-12122. https://doi.org/10.1073/pnas.2033863100

Sen Gupta, P., Islam, R.N.U., Yasmeen, S. (2017). Cosurim: A program for automated computation of composition of core, rim and surface of protein data bank files and its applications. *International Journal of Engineering Science and Technology*, *9*(11), 993-999.

Song, S. Y., Xu, Y. B., Lin, Z. J., & Tsou, C. L. (1999). Structure of active site carboxymethylated D-glyceraldehyde-3-phosphate dehydrogenase from Palinurus versicolor. *Journal of Molecular Biology*, *287*(4), 719-725. https://doi.org/10.1006/jmbi.1999.2628

Song, S. Y., Xu, Y. B., Lin, Z. J., & Tsou, C. L. (1999). Structure of active site carboxymethylated D-glyceraldehyde-3-phosphate dehydrogenase from Palinurus versicolor. *Journal of Molecular Biology*, *287*(4), 719-725. https://doi.org/10.1006/jmbi.1999.2628

Soukri, A., Mougin, A., Corbier, C., Wonacott, A., Branlant, C., & Branlant, G. (1989). Role of the histidine 176 residue in glyceraldehyde-3-phosphate dehydrogenase as probed by site-directed mutagenesis. *Biochemistry*, *28*(6), 2586-2592. https://doi.org/10.1021/bi00432a036

Stollar, E. J., & Smith, D. P. (2020). Uncovering protein structure. *Essays in Biochemistry*, 64(4), 649–680. https://doi.org/10.1042/ebc20190042

Surti, M., Patel, M., Adnan, M., Moin, A., Ashraf, S. A., Siddiqui, A. J., ... & Reddy, M. N. (2020). Ilimaquinone (marine sponge metabolite) as a novel inhibitor of SARS-CoV-2 key target proteins in comparison with suggested COVID-19 drugs: designing, docking and molecular dynamics simulation study. *RSC Advances*, *10*(62), 37707-37720. https://doi.org/10.1039/D0RA06379G

Trevino, S. R., Scholtz, J. M., & Pace, C. N. (2007). Amino acid contribution to protein solubility: Asp, Glu, and Ser contribute more favorably than the other hydrophilic amino acids in RNase Sa. *Journal of Molecular Biology*, *366*(2), 449-460. https://doi.org/10.1016/j.jmb.2006.10.026

Tsodikov, O. V., Record Jr, M. T., & Sergeev, Y. V. (2002). Novel computer program for fast exact calculation of accessible and molecular surface areas and average surface curvature. *Journal of Computational Chemistry*, *23*(6), 600-609. https://doi.org/10.1002/jcc.10061

Williams, R. W., Chang, A., Juretić, D., & Loughran, S. (1987). Secondary structure predictions and medium range interactions. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*, *916*(2), 200-204. https://doi.org/10.1016/0167-4838(87)90109-9

Williamson, V. M., & Paquin, C. E. (1987). Homology of *Saccharomyces cerevisiae* ADH4 to an iron-activated alcohol dehydrogenase from *Zymomonas mobilis*. *Molecular and General Genetics MGG*, *209*, 374-381. https://doi.org/10.1007/BF00329668

Xiao, Y., Wu, K., Batool, S.S., Wang, Q., Chen, H., Zhai, X., Yu, Z., & Huang, J. (2023). Enzymatic properties of alcohol dehydrogenase PedE_M.s. derived from *Methylopila* sp. M107 and its broad metal selectivity. *Front. Microbiol.*, *14*, 1191436. https://doi.org/10.3389/fmicb.2023.1191436

Zielenkiewicz, P., & Saenger, W. (1992). Residue solvent accessibilities in the unfolded polypeptide chain. *Biophysical Journal*, *63*(6), 1483-1486. https://doi.org/10.1016/S0006-3495(92)81746-0

HOW TO CITE

Amal Kumar Bandyopadhyay, Sahini Banerjee and Somnath Das (2024). Discovery of rim region between core and surface of proteins. © International Academic Publishing House (IAPH), Dr. Somnath Das, Dr. Jayanta Kumar Das, Dr. Mayur Doke and Dr. Vincent Avecilla (eds.), *Life as Basic Science: An Overview and Prospects for the Future Volume: 3*,pp. 41-96. ISBN: 978-81-978955-7-9 doi:https://doi.org/10.52756/lbsopf.2024.e03.003