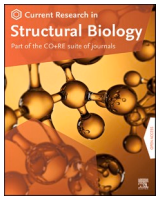
Current Research in Structural Biology 7 (2024) 100123

Contents lists available at ScienceDirect

Current Research in Structural Biology

journal homepage: www.journals.elsevier.com/current-research-in-structural-biology



Research Article

The active site of the SGNH hydrolase-like fold proteins: Nucleophile–oxyanion (Nuc-Oxy) and Acid–Base zones

Konstantin Denessiouk [a,b], Alexander I. Denesyuk [a,b,**], Sergei E. Permyakov [a], Eugene A. Permyakov [a], Mark S. Johnson [b], Vladimir N. Uversky [a,c,*]

[a] Institute for Biological Instrumentation of the Russian Academy of Sciences, Federal Research Center "Pushchino Scientific Center for Biological Research of the Russian Academy of Sciences", Pushchino, 142290, Russia
[b] Structural Bioinformatics Laboratory, Biochemistry, InFLAMES Research Flagship Center, Faculty of Science and Engineering, Biochemistry, Åbo Akademi University, Turku, 20520, Finland
[c] Department of Molecular Medicine and USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL, 33612, USA

ARTICLE INFO

Handling Editor: Dr A Wlodawer

Keywords:
SGNH-Hydrolases
Catalytic triad
Oxyanion hole
Nuc–Oxy and Acid–Base zones
SHLink

ABSTRACT

SGNH hydrolase-like fold proteins are serine proteases with the default Asp-His-Ser catalytic triad. Here, we show that these proteins share two unique conserved structural organizations around the active site: (1) the Nuc-Oxy Zone around the catalytic nucleophile and the oxyanion hole, and (2) the Acid-Base Zone around the catalytic acid and base. The Nuc-Oxy Zone consists of 14 amino acids cross-linked with eight conserved intra- and inter-block hydrogen bonds. The Acid–Base Zone is constructed from a single fragment of the polypeptide chain, which incorporates both the catalytic acid and base, and whose N- and C-terminal residues are linked together by a conserved hydrogen bond. The Nuc-Oxy and Acid-Base Zones are connected by an SHLink, a two-bond conserved interaction from amino acids, adjacent to the catalytic nucleophile and base.

1. Introduction

Earlier, we described a unique structural organization, the "catalytic core", in proteins with the alpha/beta Hydrolases (ABH) fold (Denesyuk et al., 2020a; Dimitriou et al., 2017a; Dimitriou et al., 2019). The "catalytic core" was a combination of three Zones, or characteristic areas, each around a respective residue of the catalytic triad: the catalytic Nucleophile Zone, the catalytic Base Zone, and the catalytic Acid Zone. According to SCOP (SCOP2; https://scop.mrc-lmb.cam.ac.uk/), fold alpha/beta-Hydrolases (the ABH fold; SCOP ID: 2000076) contain one superfamily of alpha/beta-Hydrolases (the ABH superfamily; SCOP ID 3000102), which includes 44 families of 248 domains (https://scop.mrc-lmb.cam.ac.uk/term/3000102) (Andreeva et al., 2020). 15 of these 44 families are various families of lipases and esterases, which all have a characteristic acid-base-nucleophile catalytic triad with serine as the nucleophile residue (Denesyuk et al., 2020a; Dimitriou et al., 2019). From the latest update of the SCOP database (released June 29, 2022), it can be shown that such esterases with serine as the catalytic nucleophile may belong not only to the ABH protein superfamily, but also to a one other superfamily of "SGNH hydrolase-like" proteins (14 protein families; SCOP ID: 3001315) with the Flavodoxin-like fold (SCOP ID: 2000016).

The hydrolases of these two superfamilies, ABH and SGNH, share structural and functional homology. Structurally, they both belong to the α/β protein class, and their folds share a 3-layer α/β/α core structure and only differ in the topology of the central β-sheet, with the mixed (seven parallel plus one anti-parallel) β-sheet in the alpha/beta-Hydrolases (the ABH fold) and the five-stranded parallel β-sheet in the SGNH hydrolase-like proteins (the Flavodoxin-like fold). Functionally, both the ABH and SGNH hydrolases are serine proteases mainly utilizing the Asp-His-Ser catalytic triad.

Historically, the SGNH hydrolases were called GDSL esterases and lipases, because of the GDSL consensus sequence segment around the catalytic serine nucleophile in those enzymes (Dalrymple et al., 1997;

* Corresponding author. Department of Molecular Medicine and USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL, 33612, USA..
** Corresponding author. Structural Bioinformatics Laboratory, Biochemistry, InFLAMES Research Flagship Center, Faculty of Science and Engineering, Biochemistry, Åbo Akademi University, Turku, 20520, Finland.
E-mail addresses: alexandre.denesyuk@abo.fi (A.I. Denesyuk), vuversky@usf.edu (V.N. Uversky).

https://doi.org/10.1016/j.crstbi.2023.100123
Received 26 October 2023; Received in revised form 25 December 2023; Accepted 27 December 2023
Available online 29 December 2023

Upton and Buckley, 1995). Later, however, after the new structures of the same protein superfamily were identified, the name was changed to the SGNH hydrolase-like superfamily, where the first serine and the fourth histidine are the two residues of catalytic triad, and the glycine and asparagine between them are the invariant catalytic residues in all the enzymes of the superfamily (Lo et al., 2003; Molgaard et al., 2000). However, the four residues S, G, N, and H are not located within a consecutive sequence segment, but belong to four consensus sequence blocks, named Blocks I, II, III, and V, respectively, that constituted the catalytic center of SGNH hydrolases through a conserved hydrogen-bonding network connecting the catalytic triad, the residues of the catalytic center, and the surrounding water molecules (Akoh et al., 2004; Anderson et al., 2022; Lo et al., 2003). For reference, the historical GDSL consensus sequence is simply Block I, and thus the "S" residue in GDSL and SGNH is the same catalytic serine of the catalytic triad. Besides the catalytic histidine, consensus Block V also contains the catalytic acid of the catalytic triad, an aspartate in most enzymes of the SGNH hydrolase-like superfamily.

Currently, the SGNH hydrolase-like superfamily consists of 14 families with more than 30 representative domains (Andreeva et al., 2020). Here, we aimed to carry out a comparative study of active sites of these representative domains to show how these proteins are similar or different with respect to the presence or absence of core catalytic elements, how conserved or varied are their structural surrounding of the catalytic amino acids, and what kind of grouping can be deduced based on the structural attributes of their active sites rather than the overall folds or sequences.

2. Results and discussion

2.1. Creating a dataset of the SGNH hydrolase-like superfamily fold proteins

Renaming the GDSL superfamily to the SGNH hydrolase-like superfamily, described above, indicated a change in core amino acids of the structural motif, when new structures of the superfamily were identified and described. In the case of the GDSL/SGNH hydrolases, the transition went from the characteristic sequence segment to a set of specific amino acids forming the conserved hydrogen-bonding network connecting the catalytic triad, the residues of the catalytic center, and the surrounding water molecules. Therefore, as the first step of analyzing the structure of the active site of the SGNH hydrolase-like fold proteins, we assessed the conservation of the four key amino acids, S, G, N, and H, which gave the name to the SGNH superfamily. One family out of 14, the esterase domain of haemagglutinin-esterase-fusion glycoprotein HEF1 (SCOP ID: 4003705), was removed from the analysis because its best representative structure, PDB ID: 1FLC, had a low resolution of 3.20 Å. That left for analysis the SGNH hydrolase-like superfamily with 13 protein families (Table 1).

Table 1 lists 30 structures representing the SGNH hydrolase-like superfamily. The set of 30 analyzed structures listed in Table 1 was constructed as follows: (1) first, 25 out of the 30 structures (marked as SCOP) were direct members of the 13 SGNH families from SCOP (SCOP ID: 3001315) (Andreeva et al., 2020), where, if available, at least one ligand-bound and/or one non ligand-bound structures were taken for each of the 13 SGNH families; and (2) the remaining 5 out of 30 structures (marked as Additional members) are also members of the SGNH hydrolase-like superfamily; however, they were not present and analyzed in SCOP, but were independently found from the Protein Data Bank (PDB) (Berman et al., 2000). The choice of structures for the TAP-like family (SCOP ID: 4000470) is an exception. In Table 1, this family is represented by four structures, PDB IDs: 1IVN, 5TIC, 1YZF, and 7C82. Structures 1IVN (row number 1) and 5TIC (row number 2) were included for historical reasons, representing well-studied reference protein, after which the entire SGNH hydrolase-like superfamily was defined. Structure 1YZF (row number 3) was the current representative

Table 1
Analyzed members of the SGNH hydrolase-like superfamily fold proteins.

N	PDB ID	R (Å)	Protein, Ligand, EC number	Ref.
Thirteen families				
Family: TAP-like				
1	1IVN_A	1.90	Thioesterase I, **GOL$_{301}$**, EC: 3.1.2.2	Lo et al. (2003)
2	5TIC_A	1.65	Acyl-CoA thioesterase I, EC: 3.1.2.2	Grisewood et al. (2017)
3	1YZF_A	1.90	Lipase/acylhydrolase, N/A	https://doi.org/10.2210/pdb1YZF/pdb
Family: Rhamnogalacturonan acetylesterase				
4	1K7C_A	1.12	Rhamnogalacturonan acetylesterase, **SO4$_{303}$**, EC: 3.1.1.86	Molgaard and Larsen (2002)
5	1DEX_A	1.90	Rhamnogalacturonan acetylesterase, EC: 3.1.1.86	Molgaard et al. (2000)
Family: Acetylhydrolase				
6	1ES9_A	1.30	Platelet-activating factor acetylhydrolase IB subunit alpha1, EC: 3.1.1.47	McMullen et al. (2000)
7	1WAB_A	1.70	Platelet-activating factor acetylhydrolase, **ACT$_{300}$**, EC: 3.1.1.47	Ho et al. (1997)
Family: YxiM C-terminal domain-like				
8	2O14_A	2.10	Hypothetical protein yxiM, EC: 3.1.-.-	https://doi.org/10.2210/pdb2O14/pdb
Family: BT2961-like				
9	3BZW_A	1.87	*Bacteroides thetaiotaomicron* putative lipase, **ACT$_{301}$**, N/A	https://doi.org/10.2210/pdb3BZW/pdb
10	3DC7_A	2.12	Putative uncharacterized protein lp_3323, **SO4$_{233}$**, N/A	https://doi.org/10.2210/pdb3DC7/pdb
Family: Esterase				
11	1ESC_A	2.10	*Streptomyces scabies* esterase, EC: 3.1.1.-	Wei et al. (1995)
12	1ESD_A	2.30	*Streptomyces scabies* esterase, **VXA$_{400}$**, EC: 3.1.1.-	Wei et al. (1995)
Family: BACUNI_00748-like				
13	4M8K_A	1.90	Hypothetical protein, GDSL-like lipase/acylhydrolase family protein, **ACT$_{301}$**, N/A	https://doi.org/10.2210/pdb4M8K/pdb
14	4NRD_A	2.10	GDSL-like lipase BACOVA_04955, N/A	https://doi.org/10.2210/pdb4NRD/pdb
Family: Putative acetylxylan esterase-like				
15	2APJ_A	1.60	At4g34215 putative esterase, **SEB$_{31}$**, EC: 3.1.-.-	Bitto et al. (2005)
16	1ZMB_A	2.61	Acetylxylan esterase related enzyme, N/A	https://doi.org/10.2210/pdb1ZMB/pdb
Family: BACUNI_01406-like				
17	4I8I_A	1.50	Hypothetical protein BACUNI_01406, **ACT$_{300}$**, N/A	https://doi.org/10.2210/pdb4I8I/pdb
Family: DltD-like				
18	6PFX_A	1.50	D-alanyl transferase DltD, N/A	https://doi.org/10.2210/pdb6PFX/pdb
19	6O93_A	2.18	D-alanyl transferase DltD, **PO4$_{505}$**, **NA$_{510}$**, N/A	https://doi.org/10.2210/pdb6O93/pdb
Family: Hypothetical protein alr1529				
20	1VJG_A	2.01	Putative lipase from the G-D-S-L family, N/A	https://doi.org/10.2210/pdb1VJG/pdb
21	1Z8H_A	2.02	Putative lipase from the G-D-S-L family, **UNL$_{301}$**, N/A	https://doi.org/10.2210/pdb1Z8H/pdb
Family: OSK domain-like				
22	5A4A_A	1.70	RNA-binding Oskar domain, **SO4$_{1607}$**, **SO4$_{1608}$**, **SO4$_{1609}$**, N/A	Jeske et al. (2015)

(*continued on next page*)

Table 1 (*continued*)

N	PDB ID	R (Å)	Protein, Ligand, EC number	Ref.
Family: AlgX acetyltransferase domain-like				
23	4O8V_A	1.81	Alginate biosynthesis protein AlgJ, EC: 2.3.1,-	Baker et al. (2014)
24	7ULA_A	2.46	Alginate biosynthesis protein AlgX, **NI_{504}**, N/A	Gheorghita et al. (2022)
Family: TAP-like				
25	7C82_A, B	1.18	SGNH-hydrolase family esterase AlinE4, **ACT_{205}**, **ACT_{206}**, **CD_{203}**, N/A	13
Additional members				
26	3PT5_A	1.60	NANS (YJHS), A 9-O-acetyl N-acetylneuraminic acid esterase, N/A	Rangarajan et al. (2011)
27	4H08_A	1.80	*Bacteroides thetaiotaomicron* putative hydrolase, **UNL_{301}**, N/A	https://doi.org/10.2210/pdb4H08/pdb
28	3SKV_A	2.49	Salicylyl-acyltransferase SsfX3, EC: 2.3,-,-	Pickens et al. (2011)
29	5V8E_A	2.20	O-acetyltransferase PatB1, **CIT_{401}**, **CIT_{402}**, N/A	Sychantha et al. (2018)
30	4C1B_A, B	2.50	ORF1-encoded protein esterase, EC: 3.1,-,-	Schneider et al. (2013)

N/A – Not Available.

structure of the TAP-like family (SCOP ID: 4000470), while the 7C82 structure (row number 25) stood alone as a dimer, and could be even considered as forming its own family, which we describe below.

Using only references to SCOP was not an oversight for the purposes of this particular manuscript. We have processed the other protein databases, however, for the purpose of the current study, we decided to leave references only to SCOP for the following reasons:

The ESTHER database incorporates proteins with the Alpha-Beta Hydrolase fold (the ABH fold). We have extensively gone through this database in our previous studies and publications. Despite the similarity in names, the ABH fold and the SGNH fold (the matter of this manuscript) are not the same, and sets of proteins with these folds do not overlap. Actually, none of the 30 SGNH structures from Table 1 of this manuscript are listed in ESTHER (https://bioweb.supagro.inrae.fr/ESTHER/allstructure (Lenfant et al., 2013);). We have started discussion from the ABH fold, and consequently from the ESTHER database, because it was logical from the historical point of view, but we moved to discussing the SGNH fold here. A reference to ESTHER is only introductory for the purpose of this manuscript.

The CATH database (https://www.cathdb.info (Sillitoe et al., 2021);) is probably the closest to SCOP in terms of analysis of tertiary structures. Usually, we would incorporate references to CATH, and we did go through it entirely, but specifically for the SGNH fold proteins, CATH is incomplete and describes only a portion of what SCOP does. For example, CATH does not contain 4 proteins of two unique DHSGN and $(SGND)_AH_B$ classes (Table 2, rows 23, 24, 29 and 25). In addition, for the $(SGN)_A(DH)_B$ class (Table 2, row 30), CATH does not specify the SGNH superfamily, which means that CATH contains only tertiary structures of the SGNDH class in terms of this study. There are other issues as well. Thus, bringing CATH for the SGNH fold is not exactly overly beneficial. CATH is great for many other folds and families, though.

The InterPro (Pfam and CDD; https://www.ebi.ac.uk/interpro (Paysan-Lafosse et al., 2023);) database serves somewhat a different purpose and lacks classification. It does list functional domains and motifs based mostly on the functional significance of protein sequence, but it does not classify structures into superfamilies, families and representative structures for each. Instead, in this manuscript we spelled out the presence of the well-known structural motifs, such as the Asx-turn motif and the ST-motif, where appropriate.

All in all, using SCOP would be most logical for the purposes of the current study. Bringing other databases would require additional lengthy descriptions of how data is connected, and filling exceptions and

Table 2
Sequence-structure alignment of amino acid sequence Blocks I, II, III and V in the SGNH hydrolase-like superfamily fold proteins.

N	PDB ID	Block I	Block II	Block III	Block V
SGNDH class					
Thirteen families					
[+++++] group					
Family: TAP-like					
1	1IVN_A	8 GD**S**LS 12	42 ISGDTS 47	69 ELGG**N**D 74	150 W[3]**D**[1]I**H**PN 159
2	5TIC_A	8 GD**S**LS 12	42 ISGDTS 47	69 ELGG**N**D 74	150 W[3]**D**[1]I**H**PN 159
3	1YZF_A	8 GD**S**IT 12	47 MPGDTT 52	74 FFGA**N**D 79	162 F[3]**D**[1]L**H**FS 171
Family: Rhamnogalacturonan acetylesterase					
4	1K7C_A	7 GD**S**TM 11	40 VAGRSA 45	70 EFGH**N**D 75	188 Y[3]**D**[1]T**H**TS 197
5	1DEX_A	7 GD**S**TM 11	40 VAGRSA 45	70 EFGH**N**D 75	188 Y[3]**D**[1]T**H**TS 197
Family: Acetylhydrolase					
6	1ES9_A	45 GD**S**LV 49	72 IG**G**DST 77	100 WVGT**N**N 105	189 D[2]**D**[1]L**H**LS 197
7	1WAB_A	45 GD**S**LV 49	72 IG**G**DST 77	100 WVGT**N**N 105	189 D[2]**D**[1]L**H**LS 197
Family: YxiM C-terminal domain-like					
8	2O14_A	169 GD**S**TV 173	207 SG**G**QIA 212	237 QLGI**N**D 242	334 L[4]**D**[1]L**H**PN 344
Family: BT2961-like					
9	3BZW_A	52 GD**S**IT 56	85 IS**G**RQW 90	114 FMGT**N**D 119	251 Y[6]**D**[1]L**H**PD 263
10	3DC7_A	28 GDSIT 32	60 IS**G**STI 65	89 FGGV**N**D 94	197 Y[2]**D**[1]L**H**PN 205
[+++-+] group					
Family: Esterase					
11	1ESC_A	12 GD**S**YT 16	64 CG**G**ALI 69	102 SLGG**N**T 107	261 G[18]**W**[1]A**H**PN 285
12	1ESD_A	12 GD**S**YT 16	64 CG**G**ALI 69	102 SLGG**N**T 107	261 G[18]**W**[1]A**H**PN 285
Family: BACUNI_00748-like					
13	4M8K_A	30 GD**S**YS 34	85 YS**G**STV 90	121 FGGT**N**D 126	203 D[2]**W**….G**H**PS 210
14	4NRD_A	30 GD**S**YS 34	84 FS**G**ATI 89	120 FGAT**N**D 125	202 D[2]**S**….G**H**PS 209
[++-++] group					
Family: Putative acetylxylan esterase-like					
15	2APJ_A	29 GQ**S**NM 33	120 SG**G**TAI 125	158 YQGE**S**D 163	231 P[3]**D**[1]L**H**LT 240
16	1ZMB_A	9 GQ**S**NM 13	80 EG**G**SSI 85	117 HQGE**S**D 122	200 T[3]**D**[1]I**H**ID 209
Family: BACUNI_01406-like					
17	4I8I_A	38 GN**S**FS 42	68 IG**G**CSL 73	121 QQAS**P**L 126	220 M[2]**D**[1]Y**H**LD 228
Family: DltD-like					
18	6PFX_A	69 GS**S**EL 73	99 AP**G**TQS 104	126 IISP**Q**W 131	367 F[2]**D**[1]I**H**LG 375
19	6O93_A	71 GS**S**EL 75	101 EA**G**TQS 106	128 ILSP**Q**W 133	369 F[2]**D**[1]I**H**LG 377
[+-+++] group					
Family: Hypothetical protein alr1529					
20	1VJG_A	15 GD**S**FV 19	52 IR**R**DTS 57	83 SFGL**N**D 88	174 E[4]**D**[1]V**H**PQ 184
21	1Z8H_A	15 GD**S**FV 19	52 IR**R**DTS 57	83 SFGL**N**D 88	174 E[4]**D**[1]V**H**PQ 184
[- + —] group					
Family: OSK domain-like					

(*continued on next page*)

Table 2 (*continued*)

N	PDB ID	Block I	Block II	Block III	Block V
22	5A4A_A	427 GD**D**FM 431	456 VS**G**LTI 461	482 NIGS**V**D 487	570 C[3]**S** [11]L**F**WN 589
SGNDH class					
Additional members					
[++−+] group					
26	3PT5_A	17 GQ**S**NA 21	120 RG**G**SAF 125	178 MQGE**F**D 183	267 L[30]**R**[1]**S**H**F**S 303
[+−++] group					
27	4H08_A	50 GN**S**IT 54	76 N.**S**KSV 80	104 NNGL**H**G 109	192 Y[4]**D**[1]T**H**PI 202
[+—+] group					
28	3SKV_A	172 GD**S**ICH177	204S**F**A**A**DGS 210	231 RVGT**S**N 236	327 L[2]**E**[6]T**H**PN 340
DHSGN class					
Thirteen families					
[++++-] group					
Family: AlgX acetyltransferase domain-like					
23	4O8V_A	286 GT**S**YS 290	314 ED**G**HGP 319	343 EFPE**R**Y 348	185 V[4]**D** ….T**H**WT 194
24	7ULA_A	269 GT**S**HS 273	296 FP**G**GGL 301	325 EFSP**L**Y 330	171 F[4]**D** ….**Q**HWT 180
Additional members					
[+++−] group					
29	5V8E_A	335 KD**S**FA 339	357 DL**R**H.. 360	381 VYSD**S**N 386	195 M[4]**D** **….**H**H**WN 204
$(SGND)_AH_B$ class					
Thirteen families					
[+++++] group					
Family: TAP-like					
25	7C82_ A, B	11 GD**S**LF 15	48 VS**G**DTT 53	77 ELGG**N**D 82	158 L[3]**D**[1]V**H**PT 167 A: 158–162 B: 164-167
$(SGN)_A(DH)_B$ class					
Additional members					
[+++++] group					
30	4C1B_ A, B	141 GD**S**IV 145	163 FP**G**ARV 168	191 HVGT**N**D 196	275 L[3]**D**[1]L**H**PS 284 B chain

gaps.

2.2. Classes and groups of the SGNH hydrolase-like superfamily fold proteins

As described in the Introduction section above, the SGNH superfamily is named after the four amino acids, S, G, N, and H, which lie in four consensus sequence blocks, named Blocks I, II, III, and V (Table 2). The columns 3–6 in Table 2 show the sequence-structure alignment of these four blocks in the 30 SGNH hydrolase-like superfamily fold proteins. Thus, Table 2 details the consensus block structure for these 30 representative structures. As mentioned above, based on the relationship between the overall tertiary protein structure and function, the SCOP database further separates the SGNH superfamily into 15 structural families, which incorporate most, but not all SGNH proteins (Table 1). Because our aim is to study the catalytic site of the SGNH hydrolase-like fold proteins and arrangement of the catalytic triad and the surrounding scaffold zones, we will introduce two more classification parameters that lie in between the SCOP "superfamily" and SCOP "family" classifications.

One parameter is the relative arrangement of the catalytic residues with respect to each other, and specifically the location of the catalytic acid and the catalytic base with respect to the catalytic nucleophile in the amino acid sequence. The second parameter is the actual conservation of the key amino acids, S, G, N, and H, and the catalytic acid (D). As the result, we will bring "substance" and "form" of the catalytic triad to the study of the active sites of the SGNH hydrolase-like fold proteins. Based on the location of the catalytic acid and base with respect to the catalytic nucleophile in the amino acid sequence, the 30 structures from Table 1 can be divided into four classes: SGNDH, DHSGN, $(SGND)_AH_B$, and $(SGN)_A(DH)_B$ (Table 2). Based on the conservation of five functionally important amino acids, the catalytic S (nucleophile)-G-N-H(base) and the catalytic acid (D), the same 30 structures from Table 1 can be also divided into groups from "+++++" to "−+−−−" depending on whether the catalytic residues are conserved (+) or not (−). In this study, the spatial structure of thioesterase I (Table 2, row number 1, PDB ID: 1IVN) was used as the reference structure for making pairwise superpositions with the other 29 structures using the Dali server (Holm, 2022).

The SGNDH class represents the canonical positioning of the catalytic residues, S(catalytic nucleophile)/G/N/D(catalytic acid)/H(catalytic base), within one chain. The DHSGN class proteins show the amino acid rearrangement within the same chain, where the D (catalytic acid) and H (catalytic base) come before the S/G/N segment in sequence. The $(SGND)_AH_B$ class (example: the AlinE4 dimer structure; PDB ID: 7C82 in Table 2 (Li et al., 2020);) and the $(SGN)_A(DH)_B$ class (example: the esterase domain of the ZfL2-1 ORF1 protein from the zebrafish ZfL2-1 retrotransposon; PDB ID: 4C1B in Table 2 (Schneider et al., 2013);) create their "chimeric" catalytic triads from chains A and B. These "chimeric triads" by definition were not classified by the domain fold-based SCOP database. The four classes, SGNDH, DHSGN, $(SGND)_AH_B$, and $(SGN)_A(DH)_B$, included structures of eight, two, one, and one different "+++++/−+−−−" groups, respectively, totaling 12 different class/group combinations (Table 2). This classification includes all known SGNH hydrolase-like fold proteins, where about half (16 out of 30) of the representative structures strictly followed the SGNH naming, and the rest exhibited varying degree of variability, with the highest variability observed with the RNA-binding OSK domain (PDB ID: 5A4A), in which the catalytic triad was completely absent (the SGNDH class; [−+−−−] group) (Table 2). Still, the overall fold/function properties allowed SCOP to classify 5A4A as the SGNH hydrolase-like superfamily protein, even though it had no respective conserved catalytic amino acids in the active site at all.

2.3. Structural core around catalytic residues in SGNH hydrolases, the catalytic zones

Earlier, we have described catalytic cores in many catalytic triad-based proteins with the ABH, (chymo)trypsin-like, and papain-like folds, and showed the presence of unique structure/functional environments, or "zones", around the catalytic sites in these proteins (Denessiouk et al., 2020a; Denesyuk et al., 2020a; Denesyuk et al., 2020b). Each zone incorporated a segment of the catalytic core, connected to their respective element of protein functional machinery. The "nucleophile zone", for example, would incorporate structural environment around the catalytic nucleophile, and connected to it through a network of conserved hydrogen bonds and other interactions. Here, we will use a similar approach to describe the catalytic core in the SGNH hydrolase-like fold proteins, and define the catalytic zones specific to these proteins. As shown above, the four residues, S, G, N, and H belong to four consensus sequence blocks, named Blocks I, II, III, and V, respectively (Akoh et al., 2004; Anderson et al., 2022; Lo et al., 2003). These four blocks are not equal to the catalytic cores, but historically include amino acids that are connected to the catalytic triad and the surrounding structural environment through a conserved hydrogen-bonding network. Fig. 1 illustrates the five "central" functionally important residues of the SGNH hydrolase-like proteins, "S" (Catalytic Nucleophile and Oxy I; Block I), "G" (OxyII; Block II), "N" (OxyIII; Block III), "H" (Catalytic Base; Block V) and "D" (Catalytic Acid; Block V), and functional connections among them with the example of

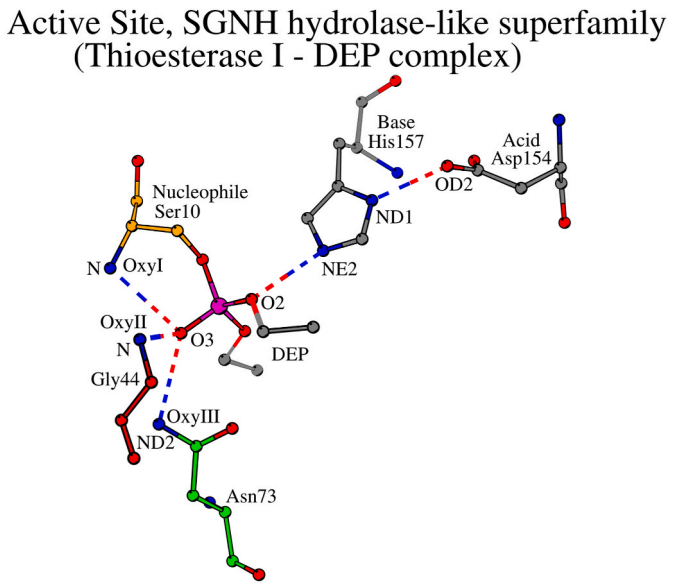


Fig. 1. Functional connection among the five "central" functionally important residues of the SGNH hydrolase-like proteins that give the hydrolases their name, "S" (Catalytic Nucleophile and Oxy I; Block I), "G" (OxyII; Block II), "N" (OxyIII; Block III), "H" (Catalytic Base; Block V) and "D" (Catalytic Acid; Block V). Amino acid numbers are taken from the complex of Thioesterase I with the inhibitor (PDB ID: 1J00). The chemical drawing showing the general SGNH hydrolase reaction scheme among these residues is shown in the recent review on the SGNH hydrolase family (Fig. 2 in (Anderson et al., 2022)).

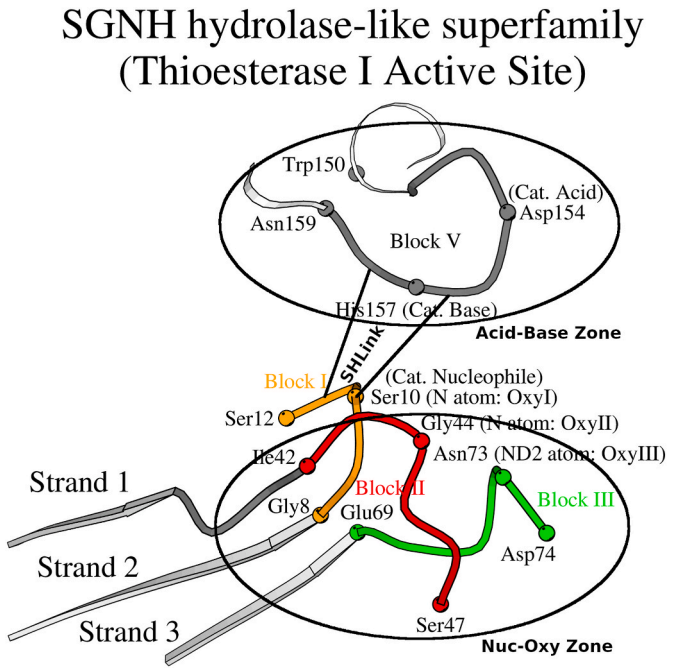


Fig. 2. Three-dimensional structure of the active site in SGNH hydrolase-like superfamily fold proteins. Amino acid numbers are taken as in thioesterase I (PDB ID: 1IVN). The catalytic triad includes Ser_{10} (the catalytic nucleophile), Asp_{154} (the catalytic acid) and His_{157} (the catalytic base). Two main-chain nitrogen atoms, N/Ser_{10} (OxyI) and N/Gly_{44} (OxyII), and side-chain nitrogen $ND2/Asn_{73}$ (OxyIII) form the oxyanion hole. Blocks (I (orange), II (red), III (green) and V (gray)) are a conserved amino acid sequence blocks in the SGNH hydrolase-like superfamily. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

structure of Thioesterase I. Fig. 1 should be taken together with the chemical drawing showing the general SGNH hydrolase reaction scheme from the recent review on the SGNH hydrolase family (Fig. 2 in (Anderson et al., 2022)). Here, we will use Blocks I, II, III, and V and the "central" functionally important residues of the SGNH hydrolase-like proteins from each block (Fig. 1) as a starting point for investigating the catalytic core in the SGNH hydrolases.

2.3.1. Structural core around the catalytic nucleophile in SGNH hydrolases. The Nuc-Oxy zone in thioesterase I

Let us consider thioesterase I (PDB ID: 1IVN; row number 1 in Table 2), which belongs to the TAP-like family (SCOP ID: 4000470), and whose conserved hydrogen bonding network around the catalytic core was well identified and shown in (Lo et al., 2003). The catalytic triad in thioesterase I is Ser_{10} – Asp_{154} – His_{157}, and Blocks I, II, and III form the basis of the active site around the Ser_{10} nucleophile (Table 2; Fig. 2). Two main-chain nitrogen atoms, N/Ser_{10} (OxyI) and N/Gly_{44} (OxyII), and the side-chain nitrogen $ND2/Asn_{73}$ (OxyIII) form the oxyanion hole in this enzyme. Taken together, in thioesterase I in the area surrounding the catalytic nucleophile, there are 14 amino acids that are connected by a network of conserved interactions and contain the oxyanion hole. This 14 amino acid structure is the "Nucleophile-Oxyanion (Nuc-Oxy) Zone" in thioesterase I, and it consists of segments Gly_8-Asp_9 from Block I, Ile_{42}-Ser_{47} (Block II) and Glu_{69}-Asp_{74} (Block III) (Fig. 2; row number 1 in Table 2). The three blocks of the Nuc-Oxy Zone in thioesterase I are connected by a network of five interactions: (1, 2, and 3) one canonical N–H···O hydrogen bond and two weak C–H···O hydrogen bonds (Manikandan and Ramakumar, 2004) between Gly_8-Asp_9 (Block I), Ile_{42} (Block II), and Glu_{69} (Block III); and (4 and 5) two canonical hydrogen bonds between Ser_{47} and Asp_{74} (Fig. 3A, Table 3). Geometrical data for hydrogen bond OG/Ser_{47} … $OD1/Asp_{74}$ = 2.4 Å is not included in Table 3.

In addition to these five inter-block contacts, there are also four functionally important intra-block hydrogen bond contacts: 1–2) O/Asp_9 … N/Ser_{12} and $OD1/Asp_9$ … N/Leu_{11} in Block I coordinate the catalytic nucleophile Ser_{10} in general, N/Ser_{10} atom (aka the OxyI atom of the oxyanion hole) in particular, and amino acid Leu_{11}; 3) O/Ile_{42} … N/Asp_{45} in Block II coordinates the N/Gly_{44} atom (aka OxyII); and 4) O/Gly_{71} … N/Asp_{74} in block III coordinates the $ND2/Asn_{73}$ atom (aka OxyIII) (Fig. 3B, Table 3). As mentioned above, atoms N/Ser_{10} (OxyI), N/Gly_{44} (OxyII), and $ND2/Asn_{73}$ (OxyIII) together constitute the oxyanion hole in thioesterase I (Fig. 2

) (Lo et al., 2003), and the contacts O/Asp_9 … N/Ser_{12} and $OD1/Asp_9$ … N/Leu_{11} in Block I form the well-known Asx-motif (Wan and Milner-White, 1999a).

2.3.2. Nuc-Oxy Zones of the SGNH hydrolase-like superfamily fold proteins

Using the extensive structural data known for thioesterase I, we have described the nucleophile-oxyanion (Nuc-Oxy) zone, which binds and coordinates the catalytic nucleophile and the oxyanion hole in this enzyme. The Nuc-Oxy Zone structure is present in all SGNH hydrolase-like superfamily fold proteins. Table S1 shows structural comparison of the Nuc-Oxy Zones in 30 representative SGNH hydrolase-like superfamily fold proteins. The proteins exhibit a high degree of similarity in contacts between Blocks II and III, with two Block I amino acids as a structural mediator at their N-terminal ends. Only one protein out of 30 (salicylyl-acyltransferase SsfX3; PDB ID: 3SKV) has some modification in the contacts between Block I and Block II at their N-terminal ends ("Blocks I and II, N-ends" in Table S1; columns 4 and 5). Different atom types in contacts in this structure can also be artificial, and due to the fact that the 3SKV structure has resolution of 2.49 Å with three amino acid differences of catalytically important residues.

Interactions between the C-terminal ends of Blocks II and III in 30 SGNH hydrolase-like proteins are less conserved compared to their N-terminal ends (last column "Blocks II and III, C-ends" in Table S1). Only 18 out of 30 representative structures have C-terminal contacts between

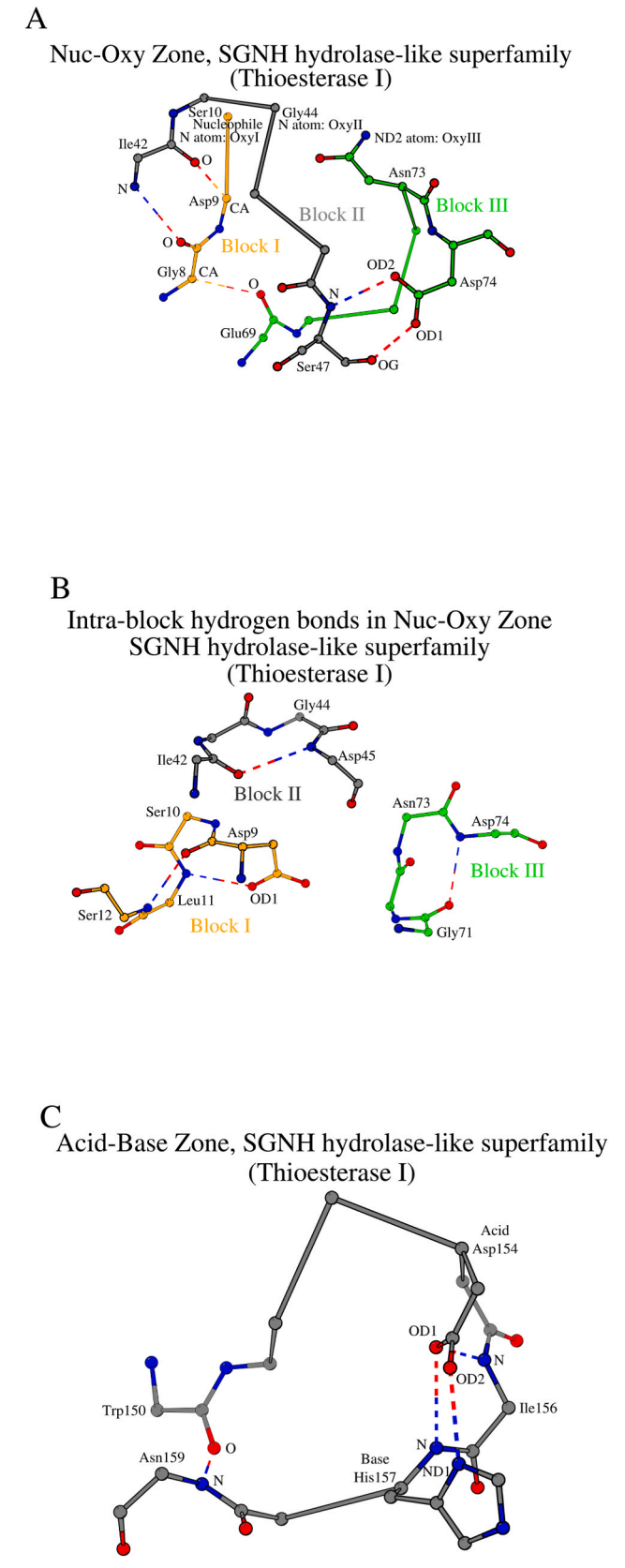


Fig. 3. Contact schemes of the Nuc-Oxy Zone with the inter-block hydrogen bonds shown in 3 A and intra-block hydrogen bonds shown in 3 B; and contact scheme of the Acid-Base Zone shown in 3C in the SGNH hydrolase-like superfamily fold proteins. For the image of the schemes, the tertiary structure of the thioesterase I is taken as an example. Dashed lines represent hydrogen bonds as well as weak hydrogen bonds.

Table 3
The hydrogen bonds in the active site of the Thioesterase I.

Inter-block hydrogen bonds in Nuc-Oxy Zone				
PDB ID Ligand	Blocks I and II, N-ends		Blocks I and III, N-ends	Blocks II and III, C-ends
1IVN_A **GOL_{301}**	O/Gly_8-N/ Ile_{42} 3.2 Å	CA/Asp_9-O/ Ile_{42} 3.2 Å (2.2 Å) 144^{O}	CA/Gly_8-O/ Glu_{69} 3.0 Å (2.2 Å) 131^{O}	N/Ser_{47}-OD2/Asp_{74} 3.0 Å
5TIC_A	O/Gly_8-N/ Ile_{42} 3.1 Å	CA/Asp_9-O/ Ile_{42} 3.3 Å (2.3 Å) 146^{O}	CA/Gly_8-O/ Glu_{69} 3.1 Å (2.1 Å) 157^{O}	N/Ser_{47}-OD2/Asp_{74} 2.9 Å
Intra-block hydrogen bonds in Nuc-Oxy Zone				
PDB ID **Ligand**	Block I		Block II	Block III
1IVN_A **GOL_{301}**	O/Asp_9-N/ Ser_{12} 3.0 Å	OD1/Asp_9-N/ Leu_{11} 3.0 Å	O/Ile_{42}-N/ Asp_{45} 3.3 Å	O/Gly_{71}-N/ Asp_{74} 2.8 Å
5TIC_A	O/Asp_9-N/ Ser_{12} 3.0 Å	OD1/Asp_9-N/ Leu_{11} 3.1 Å	O/Ile_{42}-N/ Asp_{45} 3.2 Å	O/Gly_{71}-N/ Asp_{74} 3.1 Å
Intra-block hydrogen bonds in Acid-Base Zone and a contacts (SHLink) between Acid-Base and Nuc-Oxy Zones				
PDB ID **Ligand**	First-last residues contact	Functional Acid-Base contact	Additional Acid-Base contact	SHLink
1IVN_A **GOL_{301}**	O/Trp_{150}-N/ Asn_{159} 2.9 Å	OD2/Asp_{154}-ND1/His_{157} 2.6 Å	OD1/Asp_{154}-N/ Ile_{156} 2.7 Å OD1/Asp_{154}-N/ His_{157} 3.0 Å	CD2/Leu_{11}-O/Ile_{156} 4.3 Å (3.5 Å) 133^{O} Leu_{11}-Pro_{158} 3.7 Å, 34.3 $Å^2$
5TIC_A	O/Trp_{150}-N/ Asn_{159} 2.9 Å	OD2/Asp_{154}-CD2/His_{157} 2.7 Å (1.7 Å) 163^{O}	OD1/Asp_{154}-N/ Ile_{156} 2.9 Å OD1/Asp_{154}-N/ His_{157} 3.1	CD2/Leu_{11}-O/Ile_{156} 4.3 Å (3.5 Å) 130^{O} Leu_{11}-Pro_{158} 3.6 Å, 35.1 $Å^2$

Blocks II and III identical to thioesterase I (Table S1). Where two proteins did not have any C-terminal hydrogen bonds between Blocks II and III, we tabulated the two following parameters for the non-polar and van der Waals forces: 1) the distance (in Å) between nearest atoms, and 2) the contact surface area (in $Å^2$) between two respective residues (Sobolev et al., 1999). One protein, O-acetyltransferase PatB1 (PDB ID: 5V8E) did not have any type of C-terminal contacts between Blocks II and III. The absence of any contact between the C-terminal amino acids of Blocks II and III in this protein is caused by the specific shortened structure of Block II (row number 29 in Table 2).

2.3.3. Structural organization of the Nuc-Oxy Zone in the SGNH hydrolase-like proteins

In the SGNH hydrolase-like proteins, Block I contains serine as the catalytic nucleophile with the exception of only the RNA-binding Oskar domain (PDB ID: 5A4A), which contains aspartic acid instead of serine at the same position (Table 2). The proper spatial geometry of the catalytic nucleophile is usually governed by the preceding residue, Asp, Asn, or Gln, which is the central amino acid of the well-known structural Asx-motif (Wan and Milner-White, 1999a). Alternatively, the position preceding the catalytic serine may contain threonine. In this case, the Asx-motif is changed into the ST-motif (Wan and Milner-White, 1999b), as in the alginate biosynthesis proteins AlgJ and AlgX (PDB IDs: 4O8V and 7ULA). Block II in the SGNH hydrolase-like proteins also has a conserved secondary structure, which is observed for 27 of the 30 representative structures listed in Table S2. Block III amino acids show conservation in all 30 representative structures. Thus, only 2 out of 30

representative proteins, salicylyl-acyltransferase SsfX3 and O-acetyl-transferase PatB1, show some structural variation of the Nuc-Oxy Zone. Because 16 out of 30 analyzed structures contained a small ligand in the region of the active site, it can be concluded that small ligands do not cause any noticeable structural distortion of the Nuc-Oxy Zone.

2.4. Structural core around catalytic acid and base in SGNH hydrolases. The Acid-Base Zone

The Nuc-Oxy Zone includes the structural core around catalytic nucleophile and the oxyanion hole in sequence Blocks I, II, and III that are defined by the S, G, and N in the SGNH hydrolase name (Table 2). Both the catalytic acid and catalytic base reside in Block V, which is designated by H in the SGNH hydrolase name (Table 2). In this section, we will describe the structural surrounding around the catalytic acid and base, the Acid-Base Zone, for the SGNH hydrolase-like superfamily fold proteins. Similar to the approach above, we will take as the template the well-described structure of thioesterase I (PDB ID: 1IVN).

2.4.1. Acid-Base Zone in thioesterase I

The Block V amino acid sequence fragment Trp_{150}-Asn_{159} of the thioesterase I (Lo et al., 2003) contains both the catalytic acid Asp_{154} and the catalytic base His_{157} (Table 2). The N-terminal Trp_{150} and C-terminal Asn_{159} amino acids of Block V are linked by a conventional hydrogen bond O/Trp_{150} … N/Asn_{159}, forming a closed structure of the Acid-Base Zone as the result (Fig. 3C; see the "First-last residues contact" column in the Acid-Base Zone section in Table 3). The mutual spatial arrangement of the catalytic acid and the catalytic base is coordinated by three hydrogen bonds: (1) the first-last residues contact, (2) the functional acid-base contact, and (3) the additional acid-base contact (Table 3). The "Additional Acid-Base contact" in the Acid-Base Zone section in Table 3 represents the Asx-turn motif (Duddy et al., 2004).

2.4.2. Acid-Base Zones of the SGNH hydrolase-like superfamily fold proteins

The last column in Table 2 shows the structural alignment of Block V in the 30 SGNH hydrolase-like superfamily fold proteins. Among all the superfamily members, only the SGNH-hydrolase family esterase AlinE4 (PDB ID: 7C82) and ORF1-encoded esterase (PDB ID: 4C1B) differ from other proteins in structural organization of Block V by including amino acids from two chains and not one. The crystal structures of these two proteins show the presence of a symmetric dimer through the swapped C-terminal domains that contain Block V residues (Li et al., 2020; Schneider et al., 2013). As a result, in the SGNH-hydrolase family esterase AlinE4, Block V consists of residues Leu_{158}-Asp_{162} from the A-chain and residues Val_{164}-Thr_{167} of the B-chain (Table 2). Residues His_{163} (A chain) and His_{163} (B chain) are in close contact with each other through π-π stacking interactions (Clementel et al., 2022), which provides a structural transition from fragment A:Leu_{158}-Asp_{162} to fragment B:Val_{164}-Thr_{167}. In the ORF1-encoded esterase both the catalytic acid Asp_{279} and catalytic base His_{282} belong to the B chain, while the catalytic nucleophile Ser_{143} resides in the A chain.

2.4.3. Structural organization of the Acid-Base Zone in the SGNH hydrolase-like proteins

Table S3 shows structural comparison of Acid-Base Zones in 30 SGNH hydrolase-like superfamily fold proteins. The lengths of the Block V fragments in the SGNH hydrolase-like superfamily fold proteins vary from 8 to 37 amino acids (Table S3, column L1). The hydrogen bond located between the first and last amino acids in Block V can be found in all 30 analyzed structures (Table S3). As seen with the Nuc-Oxy Zone, binding of small ligands does not cause noticeable structural distortion of around the catalytic acid and the catalytic base. Aspartate is found as the catalytic acid in 23 out of the 30 structures (Table 2). All 23 structures have an identical functional spatial arrangement of the catalytic acid and the catalytic base (see "Functional Acid-Base Contact" column in Table S3). The number of residues between the functional acid and the catalytic base varies from 1 to 12 amino acids (Table S3, column L2). In 22 structures out of 30, the contact between the catalytic aspartate and the catalytic base forms an Asx-turn (Duddy et al., 2004) (see "Additional Acid-Base contact" column in Table S3). Esterase AlinE4 has no such Acid-Base contact because the catalytic acid and the catalytic base are within different amino acid chains (Table 2). Salicylyl-acyltransferase SsfX3 (PDB ID: 3SKV) has glutamate as the catalytic acid. In the remaining structures (PDB IDs: 1ESC, 1ESD, 4M8K, 4NRD, and 3PT5), the main-chain oxygen atom of the catalytic residue, which acts as the catalytic acid, forms a hydrogen bond with the side-chain ND1 atom of the catalytic base (see "Functional Acid-Base contact" column in Table S3). Finally, the RNA-binding Oskar domain is not an enzyme, as it lacks a catalytic triad and most of the key contacts.

2.5. Interactions between Nuc-Oxy and Acid-base zones: the SHLink

We described two conserved structural arrangements, the Nuc-Oxy Zone and the Acid-Base Zone, in the SGNH hydrolase-like superfamily fold proteins. The two conserved zones provide a framework for the (catalytic nucleophile)/(oxyanion hole) and the (catalytic acid)/(catalytic base) structural pairs, respectively. Structural analysis of thioesterase I shows that in this protein, in addition to the commonly known functional hydrogen bond between the catalytic nucleophile and base, OG/Ser_{10} … NE2/His_{157}, there are also other contacts joining their surroundings. One such interaction is conserved throughout the entire SGNH hydrolase-like superfamily fold. It is the contact between the residue following the catalytic nucleophile (Leu_{11} in thioesterase I) and the two residues surrounding the catalytic base (Ile_{156} and Pro_{158} in thioesterase I) (Fig. 4, Table 3). The interaction between the Nuc-Oxy and Acid-Base Zones described above are found in all 30 representative structures with the SGNH hydrolase-like superfamily fold and can be logically named as the SHLink (see last column in Table S3). Where non-polar amino acids are present in interacting segments of Nuc-Oxy and Acid-Base Zones, we calculated: 1) the distance (in Å) between two nearest atoms, and 2) the contact surface area (in $Å^2$) between the two residues (Sobolev et al., 1999).

3. Conclusions

We analyzed structural conservation and domain organization of the

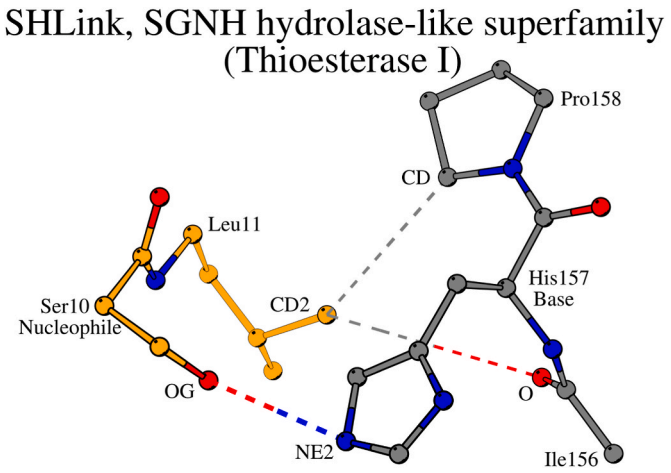


Fig. 4. Contact scheme of the SHLink connection between the Nuc-Oxy and Acid-Base Zones in the SGNH hydrolase-like superfamily fold proteins. The figure shows the contacts between the dipeptide Ser_{10}-Leu_{11} near the catalytic nucleophile Ser_{10} and a tripeptide Ile_{156}-His_{157}-Pro_{158} near the catalytic base His_{157} in thioesterase I.

catalytic triad active sites in all SGNH hydrolase-like superfamily fold proteins that were both classified by SCOP database as the SGNH hydrolase-like proteins (SCOP ID: 3001315) and those that were found elsewhere, totaling to 30 representative structures. The SGNH hydrolase-like proteins share the same fold, and have similar set of single key functional amino acids, which gave the name to the protein superfamily. These key amino acids are spread differently along protein sequences and chains in relative order and distance, forming four classes, SGNDH, DHSGN, $(SGND)_AH_B$, and $(SGN)_A(DH)_B$, and twelve groups, and yet in protein structure they all come together in a similar way to ensure protein function. It would be logical to assume that such similar local functional assemblies should be organized and maintained by larger conserved or similar sub-structures, which extend from the key amino acids and interact with them by networks of similar interactions. We call such unique conserved sub-structures around key catalytic amino acids as Zones.

As the result of our study, we describe two such structurally conserved assemblies, the "nucleophile - oxyanion" (Nuc - Oxy) Zone, which governs structural arrangement of the substrate handling catalytic nucleophile/oxyanion hole unit, and the "catalytic acid – base" (Acid - Base) Zone, which governs structural arrangement of the charge modulating catalytic acid/base unit. The two zones are connected by a structurally conserved link, the SHLink. As a result, the conformation of the active site of SGNH hydrolase-like superfamily fold proteins is uniformly formed to accommodate the catalytic triad in all the proteins of this study.

Comparative analysis of small ligand-bound and ligand-free structures of the same proteins show that the conformation of all zones stay intact upon ligand binding. Connected to that, we observed small structural motifs inside zones, such as the well-known Asx-turn motif and ST-motif (Wan and Milner-White, 1999a, 1999b), which would ensure zone rigidity.

Besides the SHLink, the Nuc-Oxy Zone and the Acid-Base Zone are stitched by a network of conserved intra-zone interactions, including hydrophobic and polar interactions and hydrogen bonds. Thus, it is not surprising that in a handful of SGNH hydrolase-like proteins one or several functionally important amino acids, S(nucleophile)-G-N-H(base) or the catalytic acid (D), may be absent or modified giving rise to several groups ranging from "+++++" to "−+−−−" depending on whether the important residue is conserved (+) or not (−), while at the same time keeping the fold and "membership" in the structural family intact.

Describing the zones provides basis for comparison, grouping, and choosing proteins based on the relevant local structural similarities, and making right choices whether the structure is incomplete, mutated or otherwise modified. We can also conclude that evolution modulates catalytic activity not only by changes in chemistry of catalytic groups, but also by construction, conservation or variation of specific features of the fold. We earlier observed similar structure/functional approach in ABH (Dimitriou et al., 2017b) and cysteine proteinase (Denessiouk et al., 2020b) fold enzymes, whose proteins do also rely on catalytic triads to carry out their function.

4. Materials and methods

The SCOP classification database (Andreeva et al., 2020) and Protein Data Bank (PDB, http://www.rcsb.org/(Berman et al., 2000)) were used to identify and retrieve all representative structures of proteins with the SGNH hydrolase-like fold (SCOP ID: 3001315). According to SCOP, the SGNH hydrolase-like superfamily consisted of 14 families with 30 representative domains. One family out of 14, the esterase domain of haemagglutinin-esterase-fusion glycoprotein HEF1 (SCOP ID: 4003705), had been removed from analysis because of poor resolution (3.20 Å) of its best representative structure (PDB ID: 1FLC). Out of the 30 domains, the first 25 were direct members of the 13 SGNH families from SCOP. The remaining domains were also members of the SGNH hydrolase-like superfamily, but were identified separately. If a representative structure had a bound ligand in the region of the active site, then we tried to locate another structure without a bound ligand to remove possible structural disturbance. The TAP-like family (SCOP ID: 4000470) was represented by four structures: PDB IDs: 1IVN, 5TIC, 1YZF (monomers), and 7C82 (dimer) to cover all possible variations.

Structure visualization and structural analysis of interactions between amino acids in proteins (hydrogen bonds, hydrophobic, other types of weak interactions) was carried out using the Discovery Studio Modeling Environment (Discovery Studio Modeling Environment (Dassault Systèmes BIOVIA, Discovery Studio Modeling Environment, Release, 2017, San Diego: Dassault Systèmes, 2016)) (https://www.3ds.com/products/biovia/discovery-studio), Maestro (Schrödinger Release, 2023–1: Schrödinger, LLC, New York, NY, 2021) (https://www.schrodinger.com/user-announcement/announcing-schrodinger-software-release-2023-4) and the Ligand-Protein Contacts (LPC) software (Sobolev et al., 1999).

The spatial structure of the thioesterase I (PDB ID: 1IVN) was used as the reference structure when performing a pairwise superposition with 29 other representative structures using the Dali server (http://ekhidna2.biocenter.helsinki.fi/dali/) (Holm, 2022). The π-π stacking and similar contacts were analyzed using the Residue Interaction Network Generator (RING, https://ring.biocomputingup.it/submit) (Clementel et al., 2022). The dimers were built using the "Protein interfaces, surfaces and assemblies" service PISA at the European Bioinformatics Institute (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html) (Krissinel and Henrick, 2007). Figures were drawn with MOLSCRIPT (Kraulis, 1991).

Funding

The project was supported by the Sigrid Jusélius Foundation (A.I.D. and M.S.J).

CRediT authorship contribution statement

Konstantin Denessiouk: Study design, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Alexander I. Denesyuk:** Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Sergei E. Permyakov:** Formal analysis, Writing – review & editing. **Eugene A. Permyakov:** Formal analysis, Writing – review & editing. **Mark S. Johnson:** Formal analysis, Methodology, Writing – original draft. **Vladimir N. Uversky:** Study design, Formal analysis, Methodology, Visualization, Investigation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

We thank the Biocenter Finland Bioinformatics Network (Dr. Jukka Lehtonen) and CSC IT Center for Science for computational support for the project. The Structural Bioinformatics Laboratory is part of the Solution for Health strategic area of Åbo Akademi University and within the InFLAMES Flagship program on inflammation and infection, Åbo Akademi University and the University of Turku, funded by the Academy of Finland.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crstbi.2023.100123.

References

Akoh, C.C., Lee, G.C., Liaw, Y.C., Huang, T.H., Shaw, J.F., 2004. GDSL family of serine esterases/lipases. Prog. Lipid Res. 43, 534–552.

Anderson, A.C., Stangherlin, S., Pimentel, K.N., Weadge, J.T., Clarke, A.J., 2022. The SGNH hydrolase family: a template for carbohydrate diversity. Glycobiology 32 (10), 826–848. https://doi.org/10.1093/glycob/cwac045.

Andreeva, A., Kulesha, E., Gough, J., Murzin, A.G., 2020. The SCOP database in 2020: expanded classification of representative family and superfamily domains of known protein structures. Nucleic Acids Res. 48, D376–D382.

Baker, P., Ricer, T., Moynihan, P.J., Kitova, E.N., Walvoort, M.T., Little, D.J., Whitney, J. C., Dawson, K., Weadge, J.T., Robinson, H., Ohman, D.E., Codee, J.D., Klassen, J.S., Clarke, A.J., Howell, P.L., 2014. P. aeruginosa SGNH hydrolase-like proteins AlgJ and AlgX have similar topology but separate and distinct roles in alginate acetylation. PLoS Pathog. 10, e1004334.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The protein Data Bank. Nucleic Acids Res. 28, 235–242.

Bitto, E., Bingman, C.A., McCoy, J.G., Allard, S.T., Wesenberg, G.E., Phillips Jr., G.N., 2005. The structure at 1.6 Angstroms resolution of the protein product of the At4g34215 gene from Arabidopsis thaliana. Acta Crystallogr D Biol Crystallogr 61, 1655–1661.

Clementel, D., Del Conte, A., Monzon, A.M., Camagni, G.F., Minervini, G., Piovesan, D., Tosatto, S.C.E., 2022. Ring 3.0: fast generation of probabilistic residue interaction networks from structural ensembles. Nucleic Acids Res. 50, W651–W656.

Dalrymple, B.P., Cybinski, D.H., Layton, I., McSweeney, C.S., Xue, G.P., Swadling, Y.J., Lowry, J.B., 1997. Three Neocallimastix patriciarum esterases associated with the degradation of complex polysaccharides are members of a new family of hydrolases. Microbiology (Read.) 143 (Pt 8), 2605–2614.

Denessiouk, K., Uversky, V.N., Permyakov, S.E., Permyakov, E.A., Johnson, M.S., Denesyuk, A.I., 2020a. Papain-like cysteine proteinase zone (PCP-zone) and PCP structural catalytic core (PCP-SCC) of enzymes with cysteine proteinase fold. Int. J. Biol. Macromol. 165, 1438–1446.

Denessiouk, K., Uversky, V.N., Permyakov, S.E., Permyakov, E.A., Johnson, M.S., Denesyuk, A.I., 2020b. Papain-like cysteine proteinase zone (PCP-zone) and PCP structural catalytic core (PCP-SCC) of enzymes with cysteine proteinase fold. Int. J. Biol. Macromol. 165 (Pt A), 1438–1446. https://doi.org/10.1016/j.ijbiomac.2020.10.022.

Denesyuk, A., Dimitriou, P.S., Johnson, M.S., Nakayama, T., Denessiouk, K., 2020a. The acid-base-nucleophile catalytic triad in ABH-fold enzymes is coordinated by a set of structural elements. PLoS One 15, e0229376.

Denesyuk, A.I., Johnson, M.S., Salo-Ahen, O.M.H., Uversky, V.N., Denessiouk, K., 2020b. NBCZone: universal three-dimensional construction of eleven amino acids near the catalytic nucleophile and base in the superfamily of (chymo)trypsin-like serine fold proteases. Int. J. Biol. Macromol. 153, 399–411.

Dimitriou, P.S., Denesyuk, A., Takahashi, S., Yamashita, S., Johnson, M.S., Nakayama, T., Denessiouk, K., 2017a. Alpha/beta-hydrolases: a unique structural motif coordinates catalytic acid residue in 40 protein fold families. Proteins 85, 1845–1855.

Dimitriou, P.S., Denesyuk, A., Takahashi, S., Yamashita, S., Johnson, M.S., Nakayama, T., Denessiouk, K., 2017b. Alpha/beta-hydrolases: a unique structural motif coordinates catalytic acid residue in 40 protein fold families. Proteins 85 (10), 1845–1855. https://doi.org/10.1002/prot.25338.

Dimitriou, P.S., Denesyuk, A.I., Nakayama, T., Johnson, M.S., Denessiouk, K., 2019. Distinctive structural motifs co-ordinate the catalytic nucleophile and the residues of the oxyanion hole in the alpha/beta-hydrolase fold enzymes. Protein Sci. 28, 344–364.

Duddy, W.J., Nissink, J.W., Allen, F.H., Milner-White, E.J., 2004. Mimicry by asx- and ST-turns of the four main types of beta-turn in proteins. Protein Sci. 13, 3051–3055.

Gheorghita, A.A., Li, Y.E., Kitova, E.N., Bui, D.T., Pfoh, R., Low, K.E., Whitfield, G.B., Walvoort, M.T.C., Zhang, Q., Codee, J.D.C., Klassen, J.S., Howell, P.L., 2022. Structure of the AlgKX modification and secretion complex required for alginate production and biofilm attachment in Pseudomonas aeruginosa. Nat. Commun. 13, 7631.

Grisewood, M.J., Hernandez Lozada, N.J., Thoden, J.B., Gifford, N.P., Mendez-Perez, D., Schoenberger, H.A., Allan, M.F., Floy, M.E., Lai, R.Y., Holden, H.M., Pfleger, B.F., Maranas, C.D., 2017. Computational redesign of acyl-ACP thioesterase with improved selectivity toward medium-chain-length fatty acids. ACS Catal. 7, 3837–3849.

Ho, Y.S., Swenson, L., Derewenda, U., Serre, L., Wei, Y., Dauter, Z., Hattori, M., Adachi, T., Aoki, J., Arai, H., Inoue, K., Derewenda, Z.S., 1997. Brain acetylhydrolase that inactivates platelet-activating factor is a G-protein-like trimer. Nature 385, 89–93.

Holm, L., 2022. Dali server: structural unification of protein families. Nucleic Acids Res. 50, W210–W215.

Jeske, M., Bordi, M., Glatt, S., Muller, S., Rybin, V., Muller, C.W., Ephrussi, A., 2015. The crystal structure of the Drosophila germline inducer oskar identifies two domains with distinct vasa helicase- and RNA-binding activities. Cell Rep. 12, 587–598.

Kraulis, P.J., 1991. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. 24, 946–950.

Krissinel, E., Henrick, K., 2007. Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 372, 774–797.

Lenfant, N., Hotelier, T., Velluet, E., Bourne, Y., Marchot, P., Chatonnet, A., 2013. ESTHER, the database of the alpha/beta-hydrolase fold superfamily of proteins: tools to explore diversity of functions. Nucleic Acids Res. 41 (Database issue), D423–D429. https://doi.org/10.1093/nar/gks1154.

Li, Z., Li, L., Huo, Y., Chen, Z., Zhao, Y., Huang, J., Jian, S., Rong, Z., Wu, D., Gan, J., Hu, X., Li, J., Xu, X.W., 2020. Structure-guided protein engineering increases enzymatic activities of the SGNH family esterases. Biotechnol. Biofuels 13, 107.

Lo, Y.C., Lin, S.C., Shaw, J.F., Liaw, Y.C., 2003. Crystal structure of Escherichia coli thioesterase I/protease I/lysophospholipase L1: consensus sequence blocks constitute the catalytic center of SGNH-hydrolases through a conserved hydrogen bond network. J. Mol. Biol. 330, 539–551.

Manikandan, K., Ramakumar, S., 2004. The occurrence of C–H...O hydrogen bonds in alpha-helices and helix termini in globular proteins. Proteins 56, 768–781.

McMullen, T.W., Li, J., Sheffield, P.J., Aoki, J., Martin, T.W., Arai, H., Inoue, K., Derewenda, Z.S., 2000. The functional implications of the dimerization of the catalytic subunits of the mammalian brain platelet-activating factor acetylhydrolase (Ib). Protein Eng. 13, 865–871.

Molgaard, A., Larsen, S., 2002. A branched N-linked glycan at atomic resolution in the 1.12 A structure of rhamnogalacturonan acetylesterase. Acta Crystallogr D Biol Crystallogr 58, 111–119.

Molgaard, A., Kauppinen, S., Larsen, S., 2000. Rhamnogalacturonan acetylesterase elucidates the structure and function of a new family of hydrolases. Structure 8, 373–383.

Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B.L., Salazar, G.A., Bileschi, M.L., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D.H., Letunic, I., Marchler-Bauer, A., Mi, H., Natale, D.A., Orengo, C.A., Pandurangan, A.P., Rivoire, C., Sigrist, C.J.A., Sillitoe, I., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Wu, C.H., Bateman, A., 2023. InterPro in 2022. Nucleic Acids Res. 51 (D1), D418–D427. https://doi.org/10.1093/nar/gkac993.

Pickens, L.B., Sawaya, M.R., Rasool, H., Pashkov, I., Yeates, T.O., Tang, Y., 2011. Structural and biochemical characterization of the salicylyl-acyltranferase SsfX3 from a tetracycline biosynthetic pathway. J. Biol. Chem. 286, 41539–41551.

Rangarajan, E.S., Ruane, K.M., Proteau, A., Schrag, J.D., Valladares, R., Gonzalez, C.F., Gilbert, M., Yakunin, A.F., Cygler, M., 2011. Structural and enzymatic characterization of NanS (YjhS), a 9-O-Acetyl N-acetylneuraminic acid esterase from Escherichia coli O157:H7. Protein Sci. 20, 1208–1219.

Schneider, A.M., Schmidt, S., Jonas, S., Vollmer, B., Khazina, E., Weichenrieder, O., 2013. Structure and properties of the esterase from non-LTR retrotransposons suggest a role for lipids in retrotransposition. Nucleic Acids Res. 41, 10563–10572.

Sillitoe, I., Bordin, N., Dawson, N., Waman, V.P., Ashford, P., Scholes, H.M., Pang, C.S. M., Woodridge, L., Rauer, C., Sen, N., Abbasian, M., Le Cornu, S., Lam, S.D., Berka, K., Varekova, I.H., Svobodova, R., Lees, J., Orengo, C.A., 2021. CATH: increased structural coverage of functional space. Nucleic Acids Res. 49 (D1), D266–D273. https://doi.org/10.1093/nar/gkaa1079.

Sobolev, V., Sorokine, A., Prilusky, J., Abola, E.E., Edelman, M., 1999. Automated analysis of interatomic contacts in proteins. Bioinformatics 15, 327–332.

Sychantha, D., Little, D.J., Chapman, R.N., Boons, G.J., Robinson, H., Howell, P.L., Clarke, A.J., 2018. PatB1 is an O-acetyltransferase that decorates secondary cell wall polysaccharides. Nat. Chem. Biol. 14, 79–85.

Upton, C., Buckley, J.T., 1995. A new family of lipolytic enzymes? Trends Biochem. Sci. 20, 178–179.

Wan, W.Y., Milner-White, E.J., 1999a. A natural grouping of motifs with an aspartate or asparagine residue forming two hydrogen bonds to residues ahead in sequence: their occurrence at alpha-helical N termini and in other situations. J. Mol. Biol. 286, 1633–1649.

Wan, W.Y., Milner-White, E.J., 1999b. A recurring two-hydrogen-bond motif incorporating a serine or threonine residue is found both at alpha-helical N termini and in other situations. J. Mol. Biol. 286, 1651–1662.

Wei, Y., Schottel, J.L., Derewenda, U., Swenson, L., Patkar, S., Derewenda, Z.S., 1995. A novel variant of the catalytic triad in the Streptomyces scabies esterase. Nat. Struct. Biol. 2, 218–223.

Discovery Studio Modeling Environment (Dassault Systèmes BIOVIA, Discovery Studio Modeling Environment, Release, 2017, San Diego: Dassault Systèmes, 2016). Available from https://www.3ds.com/products/biovia/discovery-studio. Maestro (Schrödinger Release, 2023–1: Schrödinger, LLC, New York, NY, 2021). Available from https://www.schrodinger.com/user-announcement/announcing-schrodinger-software-release-2023-4.