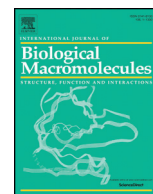




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Papain-like cysteine proteinase zone (PCP-zone) and PCP structural catalytic core (PCP-SCC) of enzymes with cysteine proteinase fold

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

^a Structural Bioinformatics Laboratory, Faculty of Science and Engineering, Biochemistry, Åbo Akademi University, Turku 20520, Finland

^b 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

^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

Article history:

Received 15 August 2020

Received in revised form 13 September 2020

Accepted 3 October 2020

Available online 12 October 2020

Keywords:

Papain

Cysteine proteinases

Catalytic triad

Fold

Zone

Structural catalytic core

COVID-19

SARC-CoV-2

ABSTRACT

There are several families of cysteine proteinases with different folds – for example the (chymo)trypsin fold family and papain-like fold family – but in both families the hydrolase activity of cysteine proteinases requires a cysteine residue as the catalytic nucleophile. In this work, we have analyzed the topology of the active site regions in 146 three-dimensional structures of proteins belonging to the Papain-like Cysteine Proteinase (PCP) superfamily, which includes papain as a typical representative of this protein superfamily. All analyzed enzymes contain a unique structurally closed conformation – a “PCP-Zone” – which can be divided into two groups, Class A and Class B. Eight structurally conserved amino acids of the PCP-Zone form a common Structural Core. The Structural Core, catalytic nucleophile, catalytic base and residue Xaa – which stabilizes the side-chain conformation of the catalytic base – make up a PCP Structural Catalytic Core (PCP-SCC). The PCP-SCC of Class A and Class B are divided into 5 and 2 types, respectively. Seven variants of the mutual arrangement of the amino-acid side chains of the catalytic triad – nucleophile, base and residue Xaa – within the same fold clearly demonstrate how enzymes with the papain-like fold adapt to the need to perform diverse functions in spite of their limited structural diversity. The roles of both the PCP-Zone of SARS-CoV-2-PLpro described in this study and the NBCZone of SARS-CoV-2-3CLpro presented in our earlier article (Denesyuk AI, Johnson MS, Salo-Ahen OMH, Uversky VN, Denessiouk K. *Int J Biol Macromol.* 2020;153:399–411) that are in contacts with inhibitors are discussed.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

The catalytic triad (Acid-Base-Nucleophile) is the widely known structural motif found in the active site of many enzymes [1,2]. Serine and cysteine are the two most common nucleophilic residues in these enzymes. In our previous work, we analyzed the structural environment of the catalytic triad (Structural Catalytic Core (SCC)) in the α/β hydrolase [3] and (chymo)trypsin-like [4] fold enzymes. In proteins from these two superfamilies, serine is the most frequently occurring amino acid used as a nucleophile in the active sites.

The superfamily of papain-like cysteine proteinases (Fold d.3 in Structural Classification of Proteins – extended (SCOPe) <https://scop.berkeley.edu/> [5]) is a classic example of enzymes that use a cysteine residue as a nucleophile. Papain-like cysteine proteases are found in

most organisms, including animals, plants, protozoa, yeast, bacteria, and virus [6]. Papain-like family enzymes are involved in numerous physiological and pathological processes, such as antigen presentation, extracellular matrix remodeling, immune invasion, hormone processing, parasite invasion, processing surface proteins, cardiovascular diseases and cancer [7–10]. This year, as the pandemic of COVID-19 began, it became particularly important to understand the molecular mechanisms behind the function of the cysteine proteases because the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has only two functionally important proteinases: papain-like (PLpro) and (chymo)trypsin-like (3CLpro) (<https://swissmodel.expasy.org/repository/species/2697049>), both belonging to the family of cysteine proteinases [11,12].

The comparison of Nucleophile-Base Catalytic Zone (NBCZone) among 169 eukaryotic, prokaryotic, and viral (chymo)trypsin-like proteases showed that: 1) the majority of eukaryotic and prokaryotic proteases have a Cys42_T-Cys58_T disulfide bond near the catalytic triad, while viral proteases do not have it, and 2) the vast majority of eukaryotic proteases have glycine residues at position 43_T. In almost all

* Corresponding authors at: 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.

E-mail address: vversky@usf.edu (V.N. Uversky).

prokaryotic and viral proteases, the amino acid at position 43_T is not glycine. Only one group of viral cysteine proteases (mainly coronavirus) has the amino acid asparagine in position 43_T [4].

The importance of the results obtained on the fundamental structural differences in the active sites of viral (chymo)trypsin-like proteases from eukaryotic and bacterial ones prompted us to carry out a similar study on papain-like cysteine proteases. In addition, in the annotation to papain-like cysteine proteases superfamily, SCOPe notes that the families included in the superfamily are distinguished “by insertion into and circular permutation of the common catalytic core made of one α -helix and 3-strands of β -sheet” [5]. Nothing of the kind is observed in (chymo)trypsin-like proteases. The structural analysis of the active site of the proteins of the papain-like cysteine protease superfamily carried out below is a next important step in the systematic study of enzymes with various folds using the catalytic triad in the enzymatic mechanism.

2. Results and discussion

2.1. PCP-Zone, Class A

2.1.1. Zone and SCC of papain

We begin the presentation of our results with an analysis of the structural environment of the catalytic triad in papain (PDB ID 1PPN) [13]. The tertiary structure of papain is well known and this fundamental structure can be found in enzymes that are classified as Fold d.3: Cysteine proteinases, Superfamily d.3.1: Cysteine proteinases, Family d.3.1.1: Papain-like [5]. The catalytic triad of papain consists of three amino acids: catalytic nucleophile Cys₂₅, catalytic base His₁₅₉ and residue Xaa (Asn₁₇₅) (Table 1). Structure visualization and structural analysis of interactions between amino acids in proteins was carried out using the Discovery Studio Modeling Environment (Dassault Systèmes BIOVIA, Discovery Studio Modeling Environment, Release 2017, San Diego: Dassault Systèmes, 2016) and the Ligand-Protein Contacts (LPC) software [14].

As a result of the visual analysis of the tertiary structure of papain in the area of its active site, we found two connected zones, a 30 amino-acid “N-Zone” Val₁₃₂-Val₁₆₁ and a 13 amino-acid “C-Zone” Ala₁₆₂-Lys₁₇₄ (Fig. 1A, Table 1). We refer to the structural union of these two Zones as the “Papain-Zone”. The Papain-Zone is formed by a 43-residue-long continuous fragment (Class A). A characteristic structural feature of the N-Zone is the presence of four main-chain hydrogen bonds between the tripeptide Val₁₃₂-Val₁₃₃-Leu₁₃₄ at the N-terminal end of the zone and the tripeptide His₁₅₉-Ala₁₆₀-Val₁₆₁ at the C-terminal end of the zone (Fig. 1B). A characteristic structural feature of the C-Zone is the presence of two hydrogen bonds between Ala₁₆₂ at the N-terminal end of the zone and Lys₁₇₄ at the C-terminal end of the zone. The three-dimensional structure of papain in the region of the Papain-Zone represents a unique mutual arrangement (papain structural core) of eight amino acids: six residues, including catalytic base (histidine), from the N-Zone and two residues from the C-Zone.

The nucleophile Cys₂₅ and residue Asn₁₇₅Xaa of the catalytic triad are located outside of the papain structural core, nevertheless, these two amino acids and the structural core interact with each other. The amino acids Ala₁₆₀ and Asn₁₇₅Xaa support the spatial arrangement of the catalytic cysteine and catalytic histidine, respectively (Fig. 1B, Tables 1 and S1). Ala₁₆₀ directly follows base His₁₅₉ in the amino acid sequence (position Base+1) and, like His₁₅₉, is involved in formation of the papain structural core; and Asn₁₇₅Xaa immediately follows the Val₁₃₂-Lys₁₇₄ fragment of the Papain-Zone in the amino-acid sequence. Ten amino acids, including eight residues of the papain structural core, catalytic cysteine and residue Xaa, form the Papain Structural Catalytic Core (Papain-SCC). According to the sequential numbering of the residues of the catalytic triad of papain, we will refer to the Cys₂₅...His₁₅₉...Asn₁₇₅Xaa triad type of SCC as “C(H)X”, where H in brackets signifies that only the catalytic histidine is located within the N-Zone.

In addition to the ten residues of the SCC, papain contains Trp₁₇₇, which is located at position Xaa + 2 of the amino-acid sequence. Trp₁₇₇ is in contact with both His₁₅₉ and Asn₁₇₅Xaa [6,15]. Trp₁₇₇ takes part in substrate binding [16] as does the N-zone, the importance of which is shown in several three-dimensional structures of papain-inhibitor complexes [17–20].

2.1.2. Cysteine proteinase fold enzymes: PCP-Zone, Class A and PCP-SCC, Type C(H)X

The superfamily of cysteine proteinases contains 24 families (SCOPe) [5]. Eighteen families (88 proteins) have a structural organization similar to the arrangement of a Class A Zone and Type C(H)X SCC as found in papain (Table S1), referred to as the Papain-like Cysteine Proteinase Zone (PCP-Zone) and PCP-SCC. In addition to the 88 three-dimensional structures, Table S1 also represents the data for 22 X-ray structures extracted from RSCB PDB [21,22]. Variation is found among these structures: the length of the fragment of the amino-acid sequence that forms the N-Zone range from 8 to 73 residues and residue Xaa of the triad can not only be asparagine, but also aspartic acid, glutamine, glutamic acid, histidine, serine, cysteine and even tryptophan. In two proteins, the catalytic nucleophile is serine (Table S1, enzymes numbered 77 and 105).

In many cases, the amino acid Xaa and the residue at position Xaa + 2 of the amino-acid sequence form a widely observed Asn-turn motif [23] (Tables 1 and S1). The equivalent structural-functional principle of using the Asx-turn motif to coordinate the catalytic histidine and substrate binding was described earlier for the catalytic acid in the *unrelated* α/β hydrolase fold enzymes [3].

Often, instead of the amino acid at position Xaa + 2 of the amino acid sequence, the residue at position Xaa + 1 takes part in coordinating the conformation of the side chain of the catalytic histidine. An exception to the rule of using the residue at position Xaa + 1 or Xaa + 2 of the amino acid sequence to coordinate catalytic histidine are cysteine proteinases belonging to the Family d.3.1.23: Papain-like viral protease catalytic domain, including proteases from SARS coronavirus (Table S1, enzymes numbered 79 and 80).

Not only is the N-Zone important for the functioning of cysteine proteinases, but the C-Zone also can play some other significant roles. For example, for the major allergens Der p 1 and f 1, amino acids from the C-Zone have been proposed to take part in recognition by the human IgE antibody [24,25].

2.1.3. Type (CH)X of PCP-SCC

Enzymes with the cysteine proteinase fold demonstrate that the N-Zone can contain not only the catalytic histidine, but also the catalytic cysteine. This (CH)X type of organization of the PCP-SCC is observed in the Family d.3.1.6: Ubiquitin carboxyl-terminal hydrolase UCH-L (Tables 1 and S1). Strictly speaking, Types C(H)X and (CH)X of the PCP-SCC are not structurally different from each other. However, because type (CH)X enzymes contain a large insertion, up to 118 residues, the cysteine of the catalytic triad is positioned within the N-zone.

2.1.4. Type (H)XC of PCP-SCC

As noted above, the location of the catalytic cysteine preceding the N-Zone or within the N-Zone may not affect the principles of the PCP-SCC organization. From our structural analysis of enzymes with the Papain-like cysteine proteinase fold, we identified a third variation, in which the catalytic nucleophile is located *after* the N-Zone. This arrangement was found in 14 proteins: Family d.3.1.7: Adenain-like, Family d.3.1.21: Yix-like and 6 cysteine peptidases from the RSCB PDB (Tables 1 and S1). For these 14 proteins, we found the following regularity: if the N-Zone contains only the catalytic histidine of the triad, then the length of the N-Zone is approximately the same as that in Type C(H)X of PCP-SCC. In one of the cysteine proteinases (Table S1, enzyme numbered 128) the residue Xaa is an aromatic amino acid. In this case,

Table 1

The geometric characteristics of the contacts of the amino acids that make up two classes of the Papain-like Cysteine Proteinase Zone (PCP-Zone) and seven types of the PCP Structural Catalytic Core (PCP-SCC) in 7 representative enzymes with the cysteine proteinase fold.

N	Protein	PDB_ID Res. (Å)	PCP-Zone			Xaa-Base Nucleophile-(Base+1) Xaa/Asx-turn	Nucleophile Xaa	Ref.
			N-Zone	N-Zone	C-Zone			
SCOPE 2.07, Fold d.3: Cysteine proteinases, Superfamily d.3.1: Cysteine proteinases								
Papain-like Cysteine Proteinases Zone (PCP-Zone): Class A								
PCP Structural Catalytic Core (PCP-SCC): I. Type C(H)X								
1	Papain	1PPN_A 1.60	N/H159-O/L134 3.1 O/H159-N/L134 3.0	N/V161-O/V132 2.9 O/V161-N/V132 3.1	N/A162-O/K174 3.3 O/A162-N/K174 2.9	OD1/N175-NE2/H159 2.6 N/A160-SG/C25 3.5	Cys25 Asn175	[34]
II. Type (CH)X								
2	Ubiquitin carboxyl-terminal esterase L3	1XD3_A 1.45	N/H169-O/F56 2.9 O/H169-N/F56 2.9	N/I171-O/L54 3.0 O/I171-N/L54 2.9	N/A172-O/L183 3.1 O/A172-N/L183 2.8	OD1/D184-NE2/H169 2.8 N/F170-SG/C95 3.7 OD1/D184-N/R186 2.8	Cys95 Asp184	[35]
III. Type (H)XC								
3	Adenain	4WX4_A 1.03	N/H54-OG1/T45 3.5 O/H54-N/T45 2.9	N/M56-O/V43 2.9 O/M56-N/V43 2.8	N/A57-O/F70 2.9 O/A57-N/F70 2.8	OD1/D71-NE2/H54 2.8 N/W55-SG/C122 3.5 OD1/D71-N/F73 2.9	Cys122 Asp71	[36]
IV. Type C(X)H								
4	Microbial transglutaminase	5M6Q_A 1.98	N/D175-O/R131 3.0 O/D175-N/R131 2.9	N/G177-O/S129 2.8 O/G177-N/S129 2.8	N/W178-O/W187 2.8 O/W178-N/W187 2.9	ND1/H188-OD2/D175 3.1 N/Y176-SG/C46 3.6	Cys46 His188	[37]
V. Type CXH								
5	Cysteine protease ATG4B	2CY7_A 1.90	N/S262-O/K259 2.8 O/S262-HOH425 2.6 HOH425-N/K259 2.9	O/H264-N/G257 3.2 N/H264-O/G257 2.9	N/Y265-O/L277 3.2 O/Y265-N/L277 2.9	OD1/D278-ND1/H280 2.6 N/A263-SG/C74 3.6 OD1/D278-N/H280 2.6	Cys74 Asp278	[38]
PCP-Zone: Class B								
PCP-SCC: VI. Type XC(H)								
6	Papain-like protease 2	4IUM_A 1.45	N/H332-O/D329 3.0 O/H332-N/D329 2.9	N/R334-O/I327 2.9 O/R334-N/I327 2.8	N/V335-O/Y262 2.8 O/V335-N/Y262 2.9	OD1/N263-NE2/H332 2.9 N/W333-SG/C270 3.7	Cys270 Asn263	[39]
VII. Type C(HX)								
7	Ubiquitin thioesterase OTU1	3BY4_A 1.55	N/H222-OD1/N219 2.9 O/H222-N/N219 3.1	N/D224-O/L217 3.0 O/D224-N/L217 2.8	N/S225-O/H112 3.1 O/S225-N/H112 2.7	CA/P113-OD1/D224 3.4 (2.4) 152° N/V114-OD1/D224 3.2 OD1/D224-NE2/H222 2.7 N/Y223-SG/C120 3.4	Cys120 Asp224	[40]
Not classified Fold								
VCP-Zone: Class C								
VCP-SCC: Type C(H)X								
8	Papain-like cysteine protease	3MTV_A 2.80	N/H159-O/Q148 2.7 O/H159-N/Q148 2.6	N/K161-O/V146 3.2 O/K161-N/V146 3.0	N/R158-O/W201 2.9 O/R158-N/G203 3.2	OXT/G203-CD2/H159 2.6 (1.6) 163° OXT/G203-SG/C90 2.9 N/L160-SG/C90 3.6	Cys90 Gly203	[27]
VCP-Zone: Class D								
VCP-SCC: Type C(HX)								
9	Venezuelan equine encephalitis virus protease	5EZS_A 2.16	N/H546-O/R543 3.0 O/H546-N/R543 3.0	N/D548-O/S541 2.8 O/D548-N/S541 2.9	N/W547-SG/C477 3.2	OD2/D548-ND1/H546 3.9	Cys477 Asp548	[41]

For CH...O contacts, in addition to the distance between the C and O atoms, the distance between the H and O atoms (in brackets), and the CHO angle are also presented. The standard criteria for contact CH...O to be considered a weak hydrogen bond are as follows: C-O ≤ 4.1 Å, H-O ≤ 3.1 Å and ∠CHO ≥ 130° [42].

the structural coordination of the side chain of the catalytic histidine is achieved via the aromatic/hydrophobic interactions.

2.1.5. Type C(X)H of PCP-SCC

In the cases considered above, the location of the catalytic base and residue Xaa in a PCP-SCC did not change relative to each other, but the location of the catalytic nucleophile in the amino-acid sequence can differ. Cysteine proteinases of the Family d.3.1.8: Microbial transglutaminase illustrate one more type of PCP-SCC organization. In fact, despite sharing the common fold, in the microbial transglutaminase from *Kutzneria albida*, for example, the sequential position of catalytic His₁₈₈ and Asp₁₇₅Xaa are interchanged compared to catalytic histidine and amino acid Xaa in the cysteine proteinases

analyzed above; i.e., Type C(H)X of PCP-SCC (Fig. 1C, Tables 1 and S1). As a result, residue Xaa is located within the N-Zone, and the catalytic histidine is attached to the C-Zone at its C-terminal end.

2.1.6. Type CXH of PCP-SCC

Family d.3.1.22: the Autophagin-like cysteine proteinase represents Type CXH of PCP-SCC, in which both the catalytic histidine and residue Xaa are located outside of the N-Zone (Fig. 1D, Tables 1 and S1). However, the catalytic histidine in this case is located two residues after Xaa (position Xaa/Asp+2) in the amino acid sequence, forming the Asx-turn motif [23]. Therefore, while in the aforementioned cases the position Xaa/Asp+2 of the amino-acid sequence was auxiliary for the positions of the catalytic triad (see sections “Zone and SCC of papain”

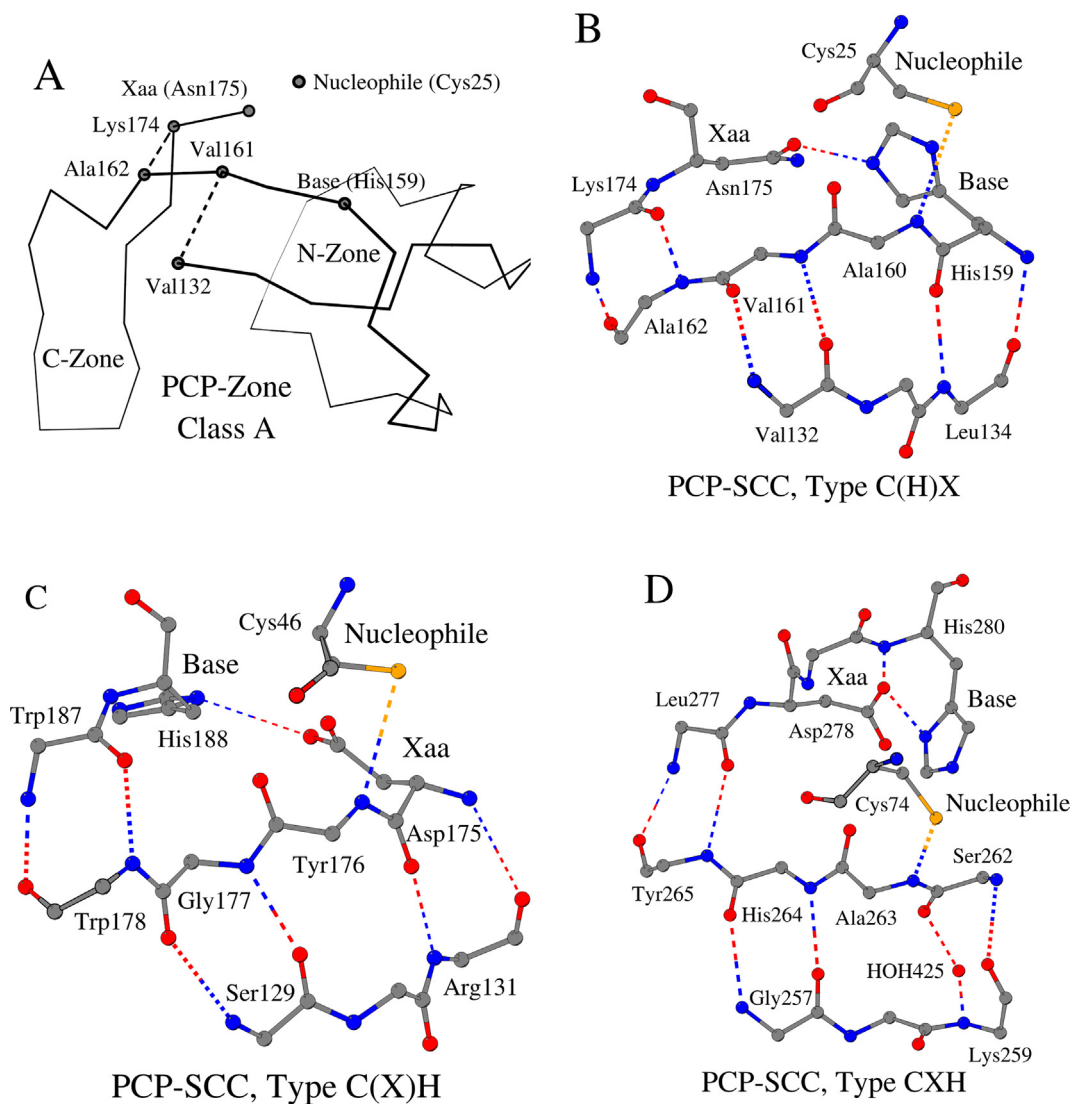


Fig. 1. PCP-Zone Class A as a structural union of the N-Zone and C-Zone. A) A schematic representation of a PCP-Zone Class A using papain three-dimensional structure as an example. Two dashed lines connect the amino acids that define the boundaries of the N- and C-Zones. Positions of the catalytic triad (Nucleophile, Base and residue Xaa) are also shown. B), C), D) Contact schemes of ten amino acids of PCP-SCC of three types: C(H)X, C(X)H and CXH, respectively.

and “Cysteine proteinase fold enzymes: PCP-Zone, Class A and PCP-SCC, Type C(H)X”), in this case it is the main one.

2.1.7. Inactive peptidases

In addition to 134 cysteine peptidases containing a catalytic triad, we found 3 inactive peptidases (Table S1). The catalytic nucleophile is absent in two inactive peptidases, and both the catalytic nucleophile and catalytic base are absent in the third inactive peptidase. Despite the lack of catalytic activity, the Zone (Class A) and SCC (Type C(H)X) characteristics of these three proteins are similar to those of active papain-like proteases.

2.2. PCP-Zone, Class B

Of the 137 analyzed structures, 23 out of 24 cysteine proteinase families (SCOPe) [5] corresponded to the types described above. The one remaining family is the Family d.3.1.11: Ubiquitin thiolesterase protein OTUB2 (Otubain-2), and one proteinase: papain-like protease 2 - have the Class B structural organization of PCP-Zone (Fig. 2A, Tables 1 and S1 (enzymes numbered 138–146)). In the PCP-Zone Class B proteinases, the C-Zone is structurally transformed (see below).

2.2.1. Type XC(H) of PCP-SCC

The papain-like protease 2 is a Class B variant of the structural organization of PCP-Zone (Tables 1 and S1). The C-Zone in this case consists of two dipeptides Gly₂₆₁-Tyr₂₆₂ and Val₃₃₅-Lys₃₃₆. Val₃₃₅ that follows the N-Zone is linked by two hydrogen bonds to Tyr₂₆₂ preceding Asn₂₆₃Xaa. The structure of this protease represents not only a new class of the Zone, but also a new Type XC(H) of the PCP-SCC (Fig. 2B).

2.2.2. Type C(HX) of PCP-SCC

In the Family d.3.1.11: Ubiquitin thiolesterase protein OTUB2 (Otubain-2) of cysteine proteinases, the C-Zone is also structurally transformed. However, unlike papain-like protease 2, the size of the C-Zone is larger in this case. For example, in ubiquitin thiolesterase OTU1, the C-Zone consists of two pentapeptides Val₁₀₈-...-His₁₁₂ and Ser₂₂₅-...-Asn₂₂₉ (Tables 1 and S1). Furthermore, residue Xaa (Asp, Asn or Ser), which coordinates the conformation of the catalytic histidine, is situated two sequence positions after the catalytic base (position Base+2) within the N-Zone and is classified as Type C(HX) of PCP-SCC (Fig. 2C). Residue Xaa is the last residue of the N-Zone. Pro₁₁₃, whose position corresponds to the position of the canonical residue Xaa, acts in this case as its structural mediator. In OTU deubiquitinase A20, a water

molecule serves as an additional mediator of the structural coordination of the side chain of the catalytic histidine (Table S1). Note that Family d.3.1.11: Ubiquitin thiolesterase protein OTUB2 (Otubain-2) is the only family of the 24 families studied that has the amino acid Asp/Asn (Xaa) at position Base+2 (Table S1).

2.3. Viral Cysteine Proteinases (VCPs) of the currently unclassified fold(s)

2.3.1. VCP-Zone, Class C and VCP-SCC, Type C(H)X

In the course of our analysis, we identified seven VCPs, the three-dimensional structures of which could not be unambiguously attributed to Fold d.3: Cysteine proteinases (Table S1, enzymes numbered 147–153). The structural analysis divided them into two groups. The first group includes three proteases (Table S1, enzymes numbered 147–149), which demonstrate the third variant (Class C) of the VCP-Zone arrangement (Tables 1 and S1 (enzymes numbered 147–149)). The structural principles of the N-Zone organization have not changed in this group. However, the C-Zone is arranged differently. Firstly, unlike the features seen in Classes A and B, in Class C, the amino acids of the C-Zone directly interact with the amino acids of the N-Zone, and significantly modify position and structure of the ligand binding site (Fig. 3A). Secondly, in connection with the previous point, instead of the antiparallel direction of the amino acid sequence fragments that form the C-Zone, their parallel direction is observed (Fig. 3B). Despite these structural differences, Class C of VCP-SCC and Class A of PCP-SCC share the same C(H)X Type. For example, in papain-like cysteine protease (Fig. 3B), the amino acid Arg₁₅₈ (as opposed to Leu₁₆₂, as it would be if the protein belonged to either Class A or B) forms the C-Zone and coordinates the three-dimensional position of amino acid Xaa with the help of two hydrogen bonds. Contact between the N-Zone and the catalytic nucleophile (N/Leu₁₆₀-SG/Cys₉₀ = 3.6 Å) still takes place. This tertiary organization of the catalytic triad was found to be suitable, for example, for the formation of a zinc-binding site, as observed in non-structural protein Nsp1 α (Table S1 (enzyme numbered 148)). The zinc ion is important for proteolytic self-release of Nsp1 α [26]. Furthermore, the N-Zone and C-Zone together (VCP-Zone) might be important for the dimerization of some viral cysteine proteinases. An example is given by the nonstructural protein Nsp1 β (which cleaves itself from the downstream Nsp2 protein via a C-terminal papain-like cysteine protease (PCP) domain) from the porcine reproductive and respiratory syndrome virus (PRRSV) that forms a functional dimer (PDB ID: 3MTV) [27]. Another example of the N-Zone and C-Zone together playing a role in homooligomerization is an octamer of the autocatalytic cysteine protease domain of potyvirus helper-component proteinase (PDB ID: 3RNV). On the other hand, the homodimer of the PRRSV Nsp1 α , which is crucial for the subgenomic mRNA synthesis, serving as a transcription factor, and which utilizes its PCP domain to self-release from the Nsp1 β , is organized differently from the aforementioned cases (PDB ID: 3IFU; [26]). In fact, Nsp1 α contains the zinc ion in the active site and is characterized by the N-Zone (Tyr₁₄₁-Ser₁₄₈), which is noticeably shorter than the N-Zones of the Nsp1 β (Val₁₄₆-Lys₁₆₁; PDB ID: 3IFU) and the autocatalytic cysteine protease domain of potyvirus helper-component proteinase (Arg₄₀₆-Val₄₁₉; PDB ID: 3RNV).

2.3.2. VCP-Zone, Class D and VCP-SCC, Type C(H)X

In the second group of proteases, which consists of four enzymes (Tables 1 and S1 (enzymes numbered 150–153)), one of the three strands of the basis β -sheet is absent in the tertiary structures. The loss of the strand results in the absence of the C-Zone in these proteases. However, these proteases have the N-Zone, including catalytic histidine, and also a conservative contact between the N-Zone and the catalytic nucleophile. The loss of the C-Zone also led to the absence of residue Xaa at its most frequently encountered position. Nevertheless, it can be assumed that the residue Asp/Ser that is situated two sequence

positions after the catalytic base can play the role of the residue Xaa, as is the case for the proteases of PCP-Zone, Class B, Type C(H)X of PCP-SCC.

2.4. Relevance of structural analysis of cysteine proteinases

The three-dimensional structures of papain-like proteins have been studied for over 50 years [28]. During that time, structural information for more than 150 proteins was obtained [21,22]. This year, as the pandemic of COVID-19 began, it became particularly important to understand the molecular mechanisms behind the function of various proteins of the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) [29]. It turned out that the SARS-CoV-2 has only two functionally important proteinases: papain-like (PLpro) and (chymo) trypsin-like (3CLpro) (<https://swissmodel.expasy.org/repository/species/2697049>), both belonging to the family of cysteine proteinases. Structural organizations of PLpro and 3CLpro complexed with the inhibitors are shown in Fig. 4A and B, respectively. Table S1 (enzyme numbered 80, PDB_ID 6WX4) shows structural data for SARS-CoV-2-PLpro [30]. The three-dimensional structures of papain and SARS-CoV-2-PLpro belong to the same Class A of PCP-Zone and Type C(H)X of PCP-SCC. In section “Zone and SCC of papain”, the importance of the N-Zone for substrates binding by papain was noted. Fig. 4C shows the interaction between N-Zone of SARS-CoV-2-PLpro (fragment of sequence Glu₂₆₃-Lys₂₇₄) and an inhibitor. The majority of protein-ligand interactions are mediated through four potential hydrogen bonds: OH/Tyr₂₆₄-N/DPP3, O/Tyr₂₆₈-NH/DPP3, N/Gly₂₇₁-O/DPP3, O/Gly₂₇₁-NH/GVE5 and van der Waals contacts between residues Tyr₂₆₄, Tyr₂₆₈, Tyr₂₇₃ and 73O2, Gly₄. Tetrapeptide Asn₂₆₇-Cys₂₇₀ of the N-Zone forms Asx-turn [23]. The C-Zone does not take part in the interaction with the inhibitor.

Residues Tyr₂₆₄ and Tyr₂₇₃ are located in the immediate vicinity of the beginning and end of the amino acid sequence fragment that defines the N-Zone. Since all papain-like cysteine proteinases have corresponding two amino acid positions, it was decided to check which pair of residues is observed in them. It turned out that in Class A proteinases, only three coronaviral proteinases (Table S1 (enzymes numbered 79–81)) have the same pair of residues. Additionally, three proteinases from the Class B Family d.3.1.11: Ubiquitin thiolesterase protein OTUB2 (Otubain-2) (Table S1 (enzymes numbered 140–142)) also have a tyrosine pair of amino acids in the analogous amino acid positions (data not shown). It is known that the SARS-CoV-2-PLpro is a deubiquitinating enzyme [12]. Amino acids Tyr₂₆₄ and Tyr₂₇₃ of PLpro structurally correspond to amino acids Val₁₃₃ and Ala₁₆₀ of papain (PDB ID 1PPN). Earlier, in section “Zone and SCC of papain”, it was noted that the amino acid Ala₁₆₀ supports the three-dimensional arrangement of the catalytic cysteine. Indeed, there are three contacts between these two amino acids: N/Ala₁₆₀-SG/Cys₂₅ = 3.5 Å, O/Ala₁₆₀-CB/Cys₂₅ = 3.4 (2.4) 172° and CB/Ala₁₆₀-O/Cys₂₅ = 3.7 (2.7) 151°. Similar contacts take place between amino acids Tyr₂₇₃ and Cys₁₁₁ in PLpro: N/Tyr₂₇₃-SG/Cys₁₁₁ = 3.5 Å, O/Tyr₂₇₃-CB/Cys₁₁₁ = 3.7 (2.7) 164° and CB/Tyr₂₇₃-O/Cys₁₁₁ = 4.0 (3.2) 128°. It can be assumed that this tyrosine pair of amino acids is a functionally specific structural marker of the deubiquitinating enzymes.

At the time of writing about the NBCZone in the superfamily of (chymo)trypsin-like folds proteases [4], the three-dimensional structure of SARS-CoV-2-3CLpro was not known. However, the situation is changed dramatically now and the three-dimensional structures of about 170 variants of this enzyme are currently known (<https://swissmodel.expasy.org/repository/species/2697049>). We selected one file (PDB_ID 7BQY [31]) from this list to analyze the structural details of the interaction between the NBCZone and the inhibitor (Fig. 4D), as in the case of SARS-CoV-2-PLpro. The NBCZone consists of eleven amino acids: Leu₂₇-Asn₂₈, Cys₃₈-Pro₃₉-Arg₄₀-His₄₁ (base)-Val₄₂, Cys₁₄₅ (nucleophile)-Gly₁₄₆-Ser₁₄₇ and His₁₆₃. The Fig. 4D additionally shows the main chain atoms of residue His₁₆₄, the carbonyl oxygen of which forms a universal weak hydrogen bond [32] with catalytic histidine:

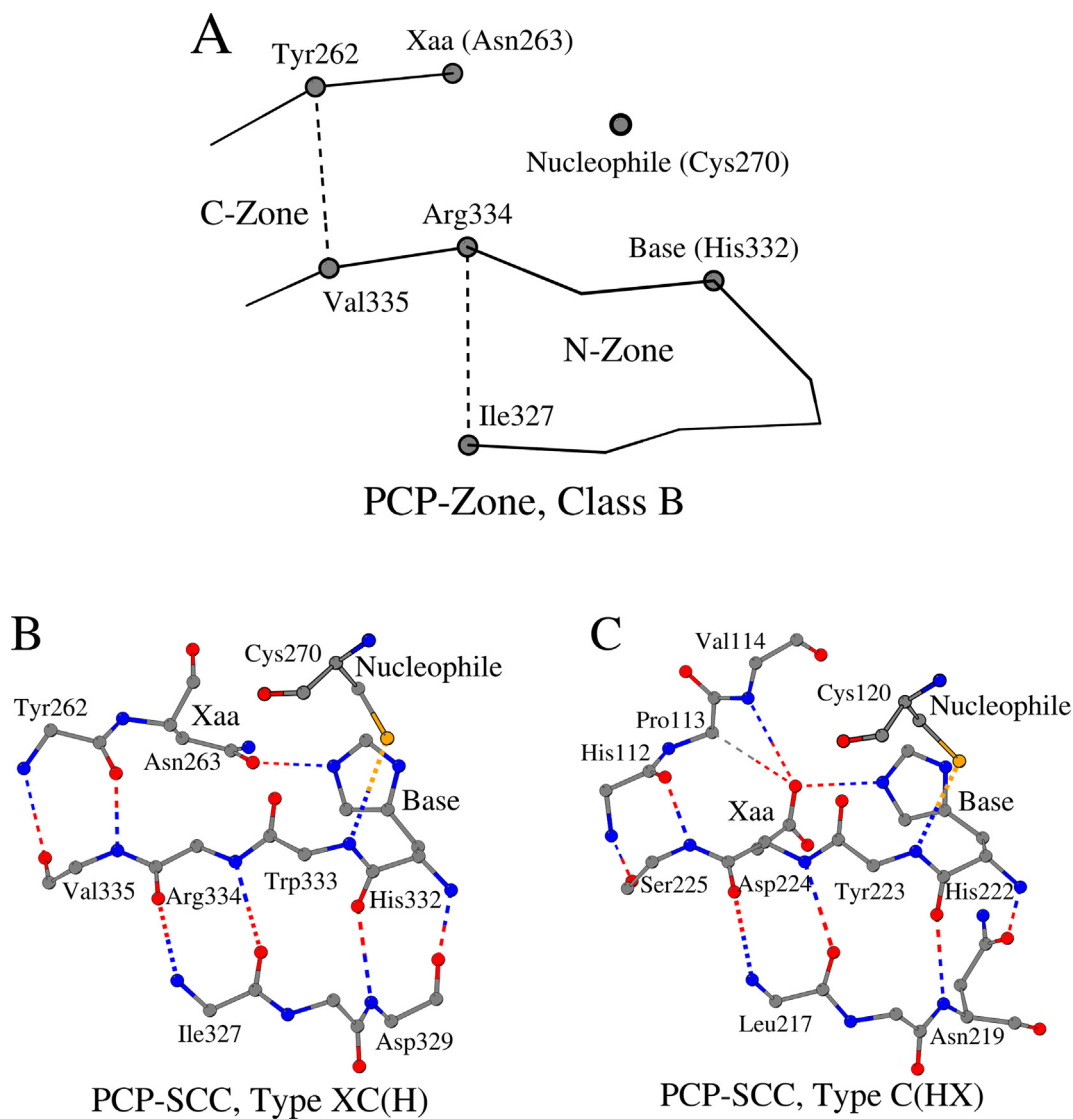


Fig. 2. PCP-Zone Class B. A. A schematic representation of a PCP-Zone Class B using three-dimensional structure of papain-like protease 2 as an example (for the designation of the details of the figure, see the legend to Fig. 1A) B. and C. Contact schemes of ten amino acids of PCP-SCC of two types: XC(H) and C(HX), respectively.

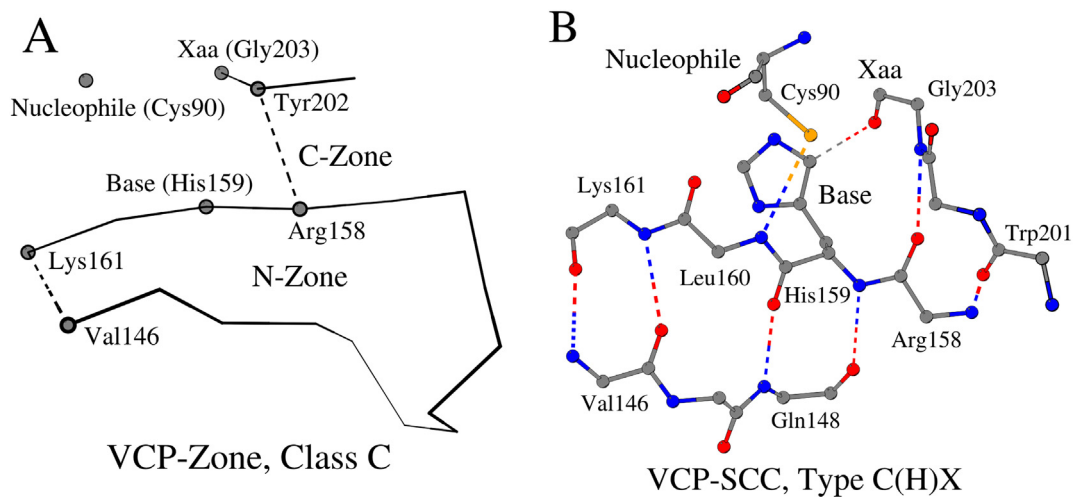


Fig. 3. VCP-Zone Class C. A. A schematic representation of a VCP-Zone Class C using three-dimensional structure of papain-like cysteine protease as an example (for the designation of the details of the figure, see the legend to Fig. 1A) B. Contact scheme of eleven amino acids of VCP-SCC, Type C(H)X.

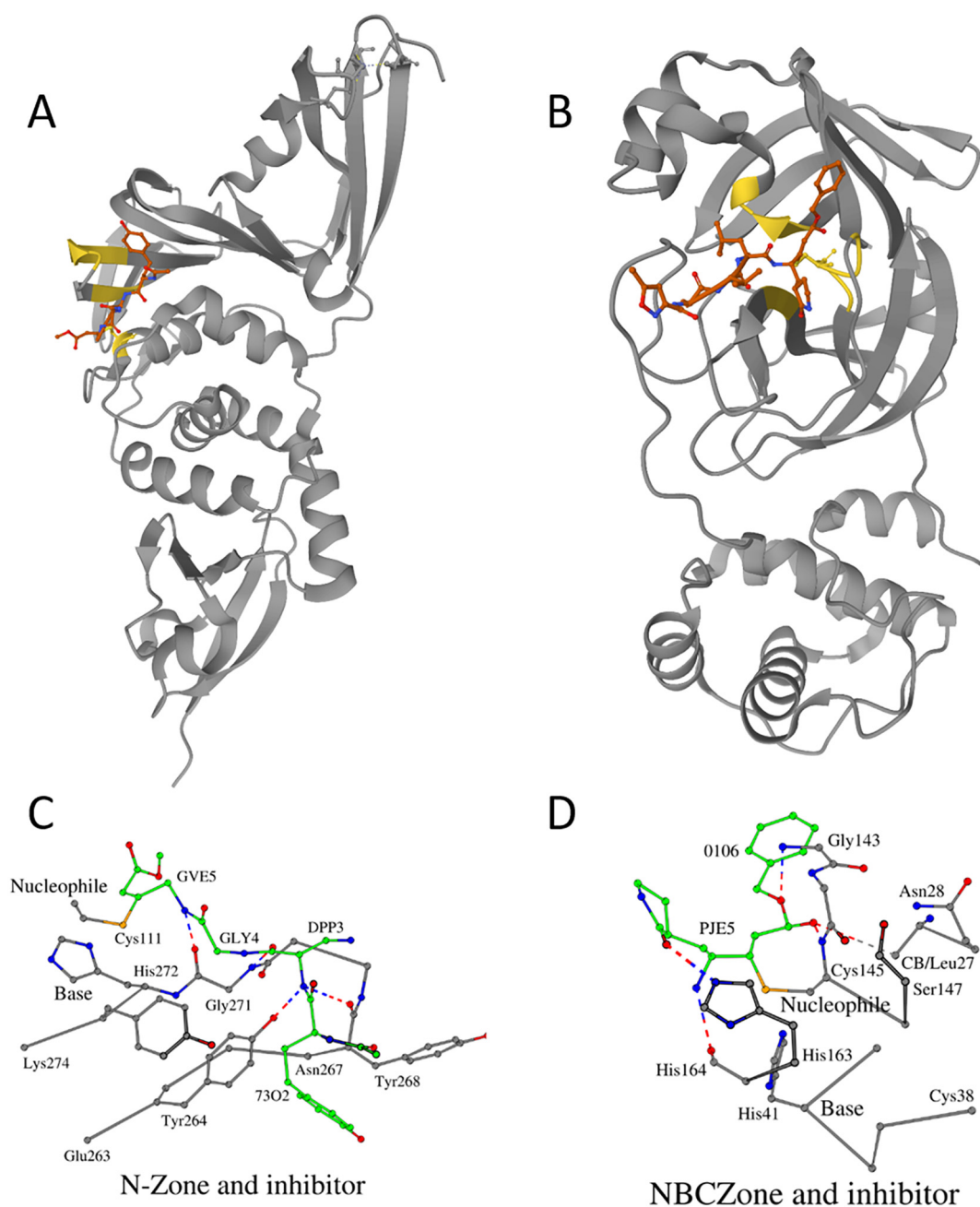


Fig. 4. Structural characterization of the SARS-CoV-2 proteases. A. Crystal structure of SARS-CoV-2-PLpro complexed with the inhibitor (PDB ID: 6WX4; [30]). B. Crystal structure of SARS-CoV-2-3CLpro complexed with the inhibitor (PDB ID: 7BQY [31]). Positions of the residues belonging to the N-Zone and the NBCZone are shown in yellow. Closer look at the complexes of the N-Zone of SARS-CoV-2-PLpro (C) and the NBCZone of SARS-CoV-2-3CLpro (D) with inhibitors. The color of carbon atoms in proteinases is gray, and in inhibitors, green. Contacts (potential hydrogen bonds) between proteinases and inhibitors are shown by dotted lines.

O/His₁₆₄-(CE1...H)/His₄₁, and a dipeptide Gly₁₄₃-Ser₁₄₄ containing a first oxyanion nitrogen atom N/Gly₁₄₃. There are four canonical (N/Gly₁₄₃-O/0106, N/Cys₁₄₅(second oxyanion atom)-O7/PJE5, NE2/His₁₆₃-O8/PJE5 and O/His₁₆₄-N5H/PJE5) and one weak ((CB...H)/Leu₂₇-O7/PJE5) hydrogen bonds between these fourteen amino acids of the SARS-CoV-2-3CLpro and two residues (PJE5 and 0106) of the inhibitor. The other four residues of the inhibitor (02J1, Ala₂, Val₃ and Leu₄) do not participate in similar bonding interactions. Two intraprotein hydrogen bonds (ND2/Asn₂₈-O/Gly₁₄₃ and OG/Ser₁₄₇-O/Ser₁₄₄) provide the necessary arrangement of the oxyanionic atoms (N/Gly₁₄₃ and N/Cys₁₄₅) for such contacts. The results presented here clearly show that the use of amino acids of both N-Zone and NBCZone for the search for effective

inhibitors of SARS-CoV-2-PLpro and SARS-CoV-2-3CLpro will accelerate the production of drugs that stop viral replication.

3. Conclusions

The 146 enzymes with the papain-like cysteine proteinase (PCP) fold have a unique, structurally closed conformation (PCP-Zone) of two Classes: A and B. The PCP-Zone, Class A consists of two Zones (N- and C-Zones) formed by a continuous fragment of the amino acid sequence. In enzymes of Class B, the C-Zone is structurally transformed. Eight structurally conserved amino acids of the PCP-Zone, in both Classes A and B, form the common Structural Core. The Structural Core,

catalytic nucleophile, catalytic base and residue Xaa, which coordinates the catalytic base, make up a PCP Structural Catalytic Core (PCP-SCC). Comparison of the PCP-SCC. Class A enzymes with each other showed that they can be divided into 5 types in accordance with the order of arrangement of the residues of the catalytic triad in the amino-acid sequence and their individual location within the N-Zone. Class B enzymes demonstrate two types of the PCP-SCC. In summary, seven variants of the mutual arrangement of the amino acids of the catalytic triad within the same fold clearly demonstrate how the enzymes of the papain-like cysteine proteinases adapt to the need to perform diverse functions in spite of the limited structural diversity.

Finding and describing local structural similarities, structural motifs, if they exist, in the active sites of proteins from different fold families is one of the fundamental areas of structural biology. Such analysis makes it possible to compare and group proteins without making a superposition of the entire tertiary structures. It serves as the basis for possible classification of proteins based on such local structural similarities. We show here that structures of active sites of different PCPs share a unique common structural organization (Structural Catalytic Core (PCP-SCC)), and can be divided into seven groups. Therefore, the members of the PCP family are characterized by one fold, two classes of supersecondary structure, and seven types of catalytic triad arrangement. The proteins that belong to each group have different tertiary structures, belong to different families and fulfill different functions, and yet their PCP-SCCs are the same. Within each group, the PCP-SCCs incorporates an identical set of conserved interactions and bonds. Furthermore, the PCP-SCCs are not only conserved in papain-like proteases with different tertiary structures and functions, but also play key role in interactions with other proteins. In particular, for PLpro of SARS-CoV-2, it was assumed that tyrosine pair of amino acids near the catalytic triad is a functionally specific structural marker of the deubiquitinating enzymes.

4. Materials and methods

Using the SCOPe classification database [5], the Sequence Similarity (<https://www.rcsb.org/pdb/explore/sequenceCluster.do?structureId=1ppn>) and Structural Similarities (<https://www.rcsb.org/pdb/explore/structureCluster.do?structureId=1ppn>) of the Protein Data Bank (PDB, <http://www.rcsb.org/> [21,22]), we have retrieved 146 different entries with the papain-like cysteine proteinases fold d.3 which represent 24 families according to SCOPe definition. Additionally, we identified 7 viral cysteine proteinases, the three-dimensional structures of which could not be unambiguously attributed to papain-like cysteine proteinases fold d.3. These 153 entries are representative structures with the highest resolution for each individual proteinase. Nine structures are used to illustrate the structural diversity of the active site observed across the papain-like cysteine proteinases fold families.

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 (Dassault Systèmes BIOVIA, Discovery Studio Modeling Environment, Release 2017, San Diego: Dassault Systèmes, 2016) and the Ligand-Protein Contacts (LPC) software [14]. Figures are drawn with MOLSCRIPT [33].

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijbiomac.2020.10.022>.

Funding

The project was supported by the Sigrid Jusélius Foundation and Joe, Pentti and Tor Borg Memorial Fund (A.I.D. and M.S.J.).

CRediT authorship contribution statement

Vladimir N. Uversky: Formal analysis, Methodology, Writing - Original Draft, Writing - Review & Editing; **Konstantin Denessiouk**: Study

design, Formal analysis, Methodology, Visualization, Investigation, 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; **Alexander I. Denesyuk**: Study design, Formal analysis, Methodology, Visualization, Writing - Original Draft, Writing - Review & Editing;

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgments

This work is supported by a grant from the Sigrid Jusélius Foundation and Joe, Pentti, and Tor Borg Memorial Fund. 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 Drug Development and Diagnostics Platform of Åbo Akademi University.

References

- [1] G. Dodson, A. Wlodawer, Catalytic triads and their relatives, *Trends Biochem. Sci.* 23 (9) (1998) 347–352.
- [2] A.R. Buller, C.A. Townsend, Intrinsic evolutionary constraints on protease structure, enzyme acylation, and the identity of the catalytic triad, *Proc. Natl. Acad. Sci. U. S. A.* 110 (8) (2013) E653–E661.
- [3] A. Denesyuk, P.S. Dimitriou, M.S. Johnson, T. Nakayama, K. Denessiouk, The acid-base-nucleophile catalytic triad in ABH-fold enzymes is coordinated by a set of structural elements, *PLoS One* 15 (2) (2020), e0229376, .
- [4] A.I. Denesyuk, M.S. Johnson, O.M.H. Salo-Ahen, V.N. Uversky, K. Denessiouk, 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 (2020) 399–411.
- [5] N.K. Fox, S.E. Brenner, J.M. Chandonia, SCOPe: structural classification of proteins—extended, integrating SCOP and ASTRAL data and classification of new structures, *Nucleic Acids Res.* 42 (Database issue) (2014) D304–D309.
- [6] M. Novinec, B. Lenarcic, Papain-like peptidases: structure, function, and evolution, *Biomol. Concepts* 4 (3) (2013) 287–308.
- [7] Y.M. Baez-Santos, S.E. St John, A.D. Mesecar, The SARS-coronavirus papain-like protease: structure, function and inhibition by designed antiviral compounds, *Antivir. Res.* 115 (2015) 21–38.
- [8] P. Gurumalles, K. Alagu, B. Ramakrishnan, S. Muthusamy, A systematic reconsideration on proteases, *Int. J. Biol. Macromol.* 128 (2019) 254–267.
- [9] H. Liu, M. Hu, Q. Wang, L. Cheng, Z. Zhang, Role of papain-like cysteine proteases in plant development, *Front. Plant Sci.* 9 (2018) 1717.
- [10] S. Verma, R. Dixit, K.C. Pandey, Cysteine proteases: modes of activation and future prospects as pharmacological targets, *Front. Pharmacol.* 7 (2016) 107.
- [11] Z. Zehra, M. Luthra, S.M. Siddiqui, A. Shamsi, N.A. Gaur, A. Islam, Corona virus versus existence of human on the earth: a computational and biophysical approach, *Int. J. Biol. Macromol.* 161 (2020) 271–281.
- [12] V. Clemente, P. D'Arcy, M. Bazzaro, Deubiquitinating enzymes in coronaviruses and possible therapeutic opportunities for COVID-19, *Int. J. Mol. Sci.* 21 (10) (2020).
- [13] G.W. Harris, R.W. Pickersgill, B. Howlin, D.S. Moss, The segmented anisotropic refinement of monoclinic papain by the application of the rigid-body TLS model and comparison to bovine ribonuclease A, *Acta Crystallogr. B* 48 (Pt 1) (1992) 67–75.
- [14] V. Sobolev, A. Sorokine, J. Prilusky, E.E. Abola, M. Edelman, Automated analysis of interatomic contacts in proteins, *Bioinformatics* 15 (4) (1999) 327–332.
- [15] S. Hussain, A. Khan, S. Gul, M. Resmini, C.S. Verma, E.W. Thomas, K. Brocklehurst, Identification of interactions involved in the generation of nucleophilic reactivity and of catalytic competence in the catalytic site Cys/His ion pair of papain, *Biochemistry* 50 (49) (2011) 10732–10742.
- [16] D. Bromme, P.R. Bonneau, E. Purisima, P. Lachance, S. Hajnik, D.Y. Thomas, A.C. Storer, Contribution to activity of histidine-aromatic, amide-aromatic, and aromatic-aromatic interactions in the extended catalytic site of cysteine proteinases, *Biochemistry* 35 (13) (1996) 3970–3979.
- [17] M.T. Stubbs, B. Laber, W. Bode, R. Huber, R. Jerala, B. Lenarcic, V. Turk, The refined 2.4 Å X-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction, *EMBO J.* 9 (6) (1990) 1939–1947.
- [18] M.S. Alphey, W.N. Hunter, High-resolution complex of papain with remnants of a cysteine protease inhibitor derived from *Trypanosoma brucei*, *Acta Crystallogr. Sec. F Struct. Biol. Cryst. Commun.* 62 (Pt 6) (2006) 504–508.
- [19] I. Redzynia, A. Ljunggren, A. Bujacz, M. Abrahamson, M. Jaskolski, G. Bujacz, Crystal structure of the parasite inhibitor chagasin in complex with papain allows identification of structural requirements for broad reactivity and specificity determinants for target proteases, *FEBS J.* 276 (3) (2009) 793–806.

- [20] M.H. Chu, K.L. Liu, H.Y. Wu, K.W. Yeh, Y.S. Cheng, Crystal structure of tarocystatin-papain complex: implications for the inhibition property of group-2 phytocystatins, *Planta* 234 (2) (2011) 243–254.
- [21] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, *Nucleic Acids Res.* 28 (1) (2000) 235–242.
- [22] H.M. Berman, T. Battistuz, T.N. Bhat, W.F. Bluhm, P.E. Bourne, K. Burkhardt, Z. Feng, G.L. Gilliland, L. Lype, S. Jain, P. Fagan, J. Marvin, D. Padilla, V. Ravichandran, B. Schneider, N. Thanki, H. Weissig, J.D. Westbrook, C. Zardecki, The protein data bank, *Acta Crystallogr. D Biol. Crystallogr.* 58 (Pt 6 1) (2002) 899–907.
- [23] W.Y. Wan, E.J. Milner-White, 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 (5) (1999) 1633–1649.
- [24] M. Chruszcz, A. Pomes, J. Glesner, L.D. Vailes, T. Osinski, P.J. Porebski, K.A. Majorek, P.W. Heymann, T.A. Platts-Mills, W. Minor, M.D. Chapman, Molecular determinants for antibody binding on group 1 house dust mite allergens, *J. Biol. Chem.* 287 (10) (2012) 7388–7398.
- [25] T. Osinski, A. Pomes, K.A. Majorek, J. Glesner, L.R. Offermann, L.D. Vailes, M.D. Chapman, W. Minor, M. Chruszcz, Structural analysis of Der p 1-antibody complexes and comparison with complexes of proteins or peptides with monoclonal antibodies, *J. Immunol.* 195 (1) (2015) 307–316.
- [26] Y. Sun, F. Xue, Y. Guo, M. Ma, N. Hao, X.C. Zhang, Z. Lou, X. Li, Z. Rao, Crystal structure of porcine reproductive and respiratory syndrome virus leader protease Nsp1alpha, *J. Virol.* 83 (21) (2009) 10931–10940.
- [27] F. Xue, Y. Sun, L. Yan, C. Zhao, J. Chen, M. Bartlam, X. Li, Z. Lou, Z. Rao, The crystal structure of porcine reproductive and respiratory syndrome virus nonstructural protein Nsp1beta reveals a novel metal-dependent nuclease, *J. Virol.* 84 (13) (2010) 6461–6471.
- [28] J. Drenth, J.N. Jansonius, R. Koekoek, H.M. Swen, B.G. Wolthers, Structure of papain, *Nature* 218 (5145) (1968) 929–932.
- [29] V. Ranga, E. Niemela, M.Z. Tamirat, J.E. Eriksson, T.T. Airene, M.S. Johnson, Immunogenic SARS-CoV-2 epitopes: in silico study towards better understanding of COVID-19 disease-paving the way for vaccine development, *Vaccines (Basel)* 8 (3) (2020).
- [30] W. Rut, Z. Lv, M. Zmudzinski, S. Patchett, D. Nayak, S.J. Snipas, F. El Oualid, T.T. Huang, M. Bekes, M. Drag, S.K. Olsen, Activity profiling and structures of inhibitor-bound SARS-CoV-2-PLpro protease provides a framework for anti-COVID-19 drug design, *bioRxiv*, 2020.
- [31] Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L.W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao, H. Yang, Structure of M(pro) from SARS-CoV-2 and discovery of its inhibitors, *Nature* 582 (7811) (2020) 289–293.
- [32] Z.S. Derewenda, U. Derewenda, P.M. Kobos, (His)C epsilon-H...O=C < hydrogen bond in the active sites of serine hydrolases, *J. Mol. Biol.* 241 (1) (1994) 83–93.
- [33] P.J. Kraulis, Molscript - a program to produce both detailed and schematic plots of protein structures, *J. Appl. Crystallogr.* 24 (1991) 946–950.
- [34] R.W. Pickersgill, G.W. Harris, E. Garman, Structure of monoclinic papain at 1.60 angstroms resolution, *Acta Crystallogr. B* 48 (1992) 59–67.
- [35] S. Misaghi, P.J. Galaray, W.J. Meester, H. Ovaas, H.L. Ploegh, R. Gaudet, Structure of the ubiquitin hydrolase UCH-L3 complexed with a suicide substrate, *J. Biol. Chem.* 280 (2) (2005) 1512–1520.
- [36] P. Grosche, F. Sirockin, A. Mac Sweeney, P. Ramage, P. Erbel, S. Melkko, A. Bernardi, N. Hughes, D. Ellis, K.D. Combrink, N. Jarousse, E. Altmann, Structure-based design and optimization of potent inhibitors of the adenoviral protease, *Bioorg. Med. Chem. Lett.* 25 (3) (2015) 438–443.
- [37] W. Steffen, F.C. Ko, J. Patel, V. Lyamichev, T.J. Albert, J. Benz, M.G. Rudolph, F. Bergmann, T. Streidl, P. Kratzsch, M. Boenitz-Dulat, T. Oelschlaegel, M. Schraeml, Discovery of a microbial transglutaminase enabling highly site-specific labeling of proteins, *J. Biol. Chem.* 292 (38) (2017) 15622–15635.
- [38] K. Sugawara, N.N. Suzuki, Y. Fujioka, N. Mizushima, Y. Ohsumi, F. Inagaki, Structural basis for the specificity and catalysis of human Atg4B responsible for mammalian autophagy, *J. Biol. Chem.* 280 (48) (2005) 40058–40065.
- [39] P.B. van Kasteren, B.A. Bailey-Elkin, T.W. James, D.K. Ninaber, C. Beugeling, M. Khajehpour, E.J. Snijder, B.L. Mark, M. Kikkert, Deubiquitinase function of arterivirus papain-like protease 2 suppresses the innate immune response in infected host cells, *Proc. Natl. Acad. Sci. U. S. A.* 110 (9) (2013) E838–E847.
- [40] T.E. Messick, N.S. Russell, A.J. Iwata, K.L. Sarachan, R. Shiekhhattar, J.R. Shanks, F.E. Reyes-Turcu, K.D. Wilkinson, R. Marmorstein, Structural basis for ubiquitin recognition by the Otu1 ovarian tumor domain protein, *J. Biol. Chem.* 283 (16) (2008) 11038–11049.
- [41] X. Hu, J.R. Compton, D.H. Leary, M.A. Olson, M.S. Lee, J. Cheung, W. Ye, M. Ferrer, N. Southall, A. Jadhav, E.M. Morazzani, P.J. Glass, J. Marugan, P.M. Legler, Kinetic, mutational, and structural studies of the Venezuelan equine encephalitis virus nonstructural protein 2 cysteine protease, *Biochemistry* 55 (21) (2016) 3007–3019.
- [42] Z.S. Derewenda, L. Lee, U. Derewenda, The occurrence of C-H...O hydrogen bonds in proteins, *J. Mol. Biol.* 252 (2) (1995) 248–262.