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.

International Journal of Biological Macromolecules 193 (2021) 2113–2120

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

journal homepage: www.elsevier.com/locate/ijbiomac

Structural and functional significance of the amino acid differences $Val_{35}Thr$, $Ser_{46}Ala$, $Asn_{65}Ser$, and $Ala_{94}Ser$ in 3C-like proteinases from SARS-CoV-2 and SARS-CoV

Alexander I. Denesyuk [a,b,*], Eugene A. Permyakov [a], Mark S. Johnson [b], Sergei E. Permyakov [a], Konstantin Denessiouk [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 Moscow Region 142290, Russia
[b] Structural Bioinformatics Laboratory, Biochemistry, InFLAMES Research Flagship Center Faculty of Science and Engineering, 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

Keywords:
(Chymo)trypsin-like proteinases
Catalytic tetrad
Structural catalytic core
Interdomain loop
Autolysis
COVID-19
SARS-CoV-2

ABSTRACT

Three dimensional structures of (chymo)trypsin-like proteinase ($3CL^{pro}$) from SARS-CoV-2 and SARS-CoV differ at 8 positions. We previously found that the $Val_{86}Leu$, $Lys_{88}Arg$, $Phe_{134}His$, and $Asn_{180}Lys$ mutations in these enzymes can change the orientation of the N- and C-terminal domains of $3CL^{pro}$ relative to each other, which leads to a change in catalytic activity. This conclusion was derived from the comparison of the structural catalytic core in 169 (chymo)trypsin-like proteinases with the serine/cysteine fold. $Val_{35}Thr$, $Ser_{46}Ala$, $Asn_{65}Ser$, $Ala_{94}Ser$ mutations were not included in that analysis, since they are located far from the catalytic tetrad. In the present work, the structural and functional roles of these variable amino acids at positions 35, 46, 65, and 94 in the $3CL^{pro}$ sequences of SARS-CoV-2 and SARS-CoV have been established using a comparison of the same set of proteinases leading to the identification of new conservative elements. Comparative analysis showed that, in addition to interdomain mobility, which could modulate catalytic activity, the $3CL^{pro}$(s) can use for functional regulation an autolytic loop and the unique Asp_{33}-Asn_{95} region (the Asp_{33}-Asn_{95} Zone) in the N-terminal domain. Therefore, all 4 analyzed mutation sites are associated with the unique structure-functional features of the $3CL^{pro}$ from SARS-CoV-2 and SARS-CoV. Strictly speaking, the presented structural results are hypothetical, since at present there is not a single experimental work on the identification and characterization of autolysis sites in these proteases.

1. Introduction

The coronavirus disease 2019 (COVID-19) pandemic, due to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has raised many important issues for the international scientific community especially regarding the molecular mechanisms involved in the viral infection process and SARS-CoV-2 replication. In coronaviruses, there are two functionally important proteinases papain-like (PL^{pro}) and the (chymo)trypsin-like cysteine proteinase ($3CL^{pro}$, also known as viral main proteinase, M^{pro}), both belonging to the family of cysteine proteinases (https://swissmodel.expasy.org/repository/species/2697049) [1,2]. The main protease $3CL^{pro}$, which corresponds to the coronavirus nonstructural protein 5 (NSP5), splits the central and C-terminal regions of the polyprotein at 11 conserved sites generating 11 mature viral NSPs required for viral replication and infection [3,4].

The coordinates of the three-dimensional (3D) structures of the $3CL^{pro}$(s) from SARS-CoV (PDB ID 1UJ1) [5] and SARS-CoV-2 (PDB ID 6LU7) [6] first appeared in the Protein Data Bank (PDB [7,8]) in 2003

* Correspondence to: A.I. Denesyuk, Structural Bioinformatics Laboratory, Biochemistry, Faculty of Science and Engineering, Åbo Akademi University, Tykistökatu 6, BioCity3A, 20520 Turku, Finland.
** Correspondence to: V.N. Uversky, Department of Molecular Medicine and USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, 12901 Bruce B. Downs Blvd., MDC 07, Tampa, FL 33612, USA.
E-mail addresses: adenesyu@abo.fi (A.I. Denesyuk), vuversky@usf.edu (V.N. Uversky).

https://doi.org/10.1016/j.ijbiomac.2021.11.043
Received 4 September 2021; Received in revised form 7 October 2021; Accepted 5 November 2021
Available online 11 November 2021

and 2020, respectively. These two viral proteases differ in their amino acid sequence at 12 positions: $Thr_{35}Val$, $Ala_{46}Ser$, $Ser_{65}Asn$, $Leu_{86}Val$, $Arg_{88}Lys$, $Ser_{94}Ala$, $His_{134}Phe$, $Lys_{180}Asn$, $Leu_{202}Val$, $Ala_{267}Ser$, $Thr_{285}Ala$ and $Ile_{286}Leu$ [9,10]. Only the first 8 out of the 12 variable amino acids are resolved in the (chymo)trypsin-like 3D structures, whereas the last 4 amino acid positions – 202, 267, 285, and 286 – belong to the an additional C-terminal extension that forms domain, but which lies outside the solved 3D structure of the $3CL^{pro}$. The sequence differences at these 12 amino acid positions are not expected to significantly affect the polarity and hydrophobicity of SARS-CoV-2 $3CL^{pro}$ compared to $3CL^{pro}$ from SARS-CoV [9]. It is of importance that all 12 variable amino acids are located outside of the catalytic and substrate binding regions of the enzyme.

During 2020–2021, structural information from X-ray structures for more than 300 SARS-CoV-2 $3CL^{pro}$ complexes with various inhibitor molecules were reported (https://swissmodel.expasy.org/repository/species/2697049). In addition, the 3D structures of complexes of SARS-CoV-2 $3CL^{pro}$ with 33 known and potential inhibitor molecules have been studied using computational methods in order to discover potential inhibitors that can be used as antiviral therapeutic agents targeting the (chymo)trypsin-like cysteine proteinase (see recent review [11]). As a result, numerous ligand-binding amino acids of SARS-CoV-2 $3CL^{pro}$ have been identified. However, only 2 (Ser_{46} and Leu_{286}) amino acids of the above mentioned 12 variable residue positions are mentioned in the review [11].

Motivated by the lack of insights on any structural and functional consequences of these amino acid differences at the aforementioned 12 positions in the sequences of the $3CL^{pro}$ of SARS-CoV and SARS-CoV-2, we first identified the Structural Catalytic Core (SCC) in 169 (chymo) trypsin-like proteinases with serine/cysteine fold [12,13]. Next, we compared the NBCZone(s) of the $3CL^{pro}$ of SARS-CoV [13] and SARS-CoV-2 [12], and found that these NBCZones of both viral proteinases form compact structures around the catalytic nucleophile and base that consist of 11 conserved amino acids: Leu_{27}, Asn_{28}, Cys_{38}, Pro_{39}, Arg_{40}, His_{41} (catalytic base), Val_{42}, Cys_{145} (catalytic nucleophile), Gly_{146}, Ser_{147} and His_{163}. Furthermore, it turned out that the NBCZones of $3CL^{pro}$ of the SARS-CoV-2 (PDB ID 7BQY) [6] and SARS-CoV (PDB ID 6XHN) [14] are identical to each other [12].

The NBCZone is only a part of the SCC. Therefore, the structural regions around the third (Cys_{85}) and fourth (His_{164}) members of the structural catalytic tetrad have been analyzed as well [12]. The compact complex of four amino acids around the catalytic acid analogue Cys_{85} is referred to as 102_T-Core. "T" indicates that the canonical residue numbering based on the trypsin sequence is used; in this case referring to the third member of the structural catalytic tetrad, the catalytic Asp_{102} in trypsin (PDB ID 4I8H) [15]). In our work, for each protease we used both the original numbering of the amino acid sequence and the canonical numbering based on trypsin. Consequently, the 102_T-Core and 85-Core of $3CL^{pro}$ of SARS-CoV-2 consists of Gln_{83}, Cys_{85}, Val_{86} and Leu_{87}. In SARS-CoV $3CL^{pro}$, leucine replaces valine at position 86. The inclusion of the four amino acids of the 102_T-Core in the SCC composition made it possible to reveal one more important difference between the $3CL^{pro}$ of SARS-CoV-2 and SARS-CoV: $Lys_{88}Arg$. This amino acid position – 88 (105_T) – is located at the conserved position of the β-sheet of the N-terminal β-barrel [16].

A set of six amino acids at positions 134, 135, 136, 180, 181, and 182, located spatially next to the fourth member of the structural catalytic tetrad, His_{164}, is called the S-Core [12]. The S-Core from the $3CL^{pro}$ of SARS-CoV-2 and SARS-CoV are characterized by two amino acid differences: $Phe_{134}His$ and $Asn_{180}Lys$. These amino acids form a part of the SCC, but they are located on its periphery. The tertiary structures of (chymo)trypsin-like proteinases with the serine/cysteine fold are separable into groups on the basis of the super-secondary structure differences within this region [12].

Amino acids at positions 86 and 180 are involved in the contacts between the N- and C-terminal β-barrels of the $3CL^{pro}$. The sequence differences between the SARS-CoV-2 and SARS-CoV $3CL^{pro}$s at positions 86 and 180 seem to affect the nature of the interaction between N- and C-terminal β-barrels, which ultimately leads to the modulation of enzymatic activity [12].

These results made it possible to explain the structural and functional significance of 4 (positions 86, 88, 134 and 180) out of 8 observed amino acid differences in the SARS-CoV-2 and SARS-CoV $3CL^{pro}$ sequences [12]. In the present work, we compared distinct 3D structures of 170 (chymo)trypsin-like proteinases with the serine/cysteine fold, identified new conserved elements, and established the structural and functional roles of the remaining four variable amino acids at positions 35, 46, 65 and 94 in the amino acid sequences of SARS-CoV-2 and SARS-CoV $3CL^{pro}$s, for which the structural context has been reported. It has been suggested that these 4 positions are associated with the autolysis process in two loops of the $3CL^{pro}$ of SARS-CoV-2 as shown for trypsin and chymotrypsin proteinases [17–19]. Currently, a strategy for inhibiting serine/cysteine proteases by targeting its autolysis loops is actively developing [20].

Therefore, the goal of this study was to find some important regularities in the 3D structures of the family of (chymo)trypsin-like serine/cysteine proteases (including $3CL^{pro}$ (chymo)trypsin-like proteases from SARS-CoV-2 and SARS-CoV), which are currently missing in the structural description of these proteins and which can be used to answer some functional questions. To this end, we utilized a structural biology approach based on the multiple structural comparison and subsequent analysis. Earlier, application of this analysis revealed the presence in the alpha/beta-hydrolases of unique structural motifs termed the structural catalytic zones (SCZs) and the SCCs that serve to properly position the catalytic machinery and coordinate function. The advantage of the use of the SCZs and SCCs for the comparative analysis is in the capability of this approach to compare and group proteins without making superposition of their entire tertiary structures. Therefore, this approach provides useful means to classify proteins into various groups on the basis of such local structural similarities. Earlier, this analysis revealed that all proteases with the (chymo)trypsin-like serine/cysteine fold contain a universal 3D structural motif in their structural catalytic cores, the Nucleophile-Base Catalytic Zone (NBCZone), that includes eleven amino acids near the catalytic nucleophile and base [12]. We also analyzed in detail the peculiarities of the amino acid content of the SCCs in 169 proteinases with the (chymo)trypsin-like serine/cysteine fold [12,13]. This analysis revealed that based on the differences in their SCCs, these proteinases can be divided into two classes and four groups, with the proteinases belonging to different classes and groups differing from each other by the nature of the interaction between their N- and C-terminal β-barrels. The utility of this approach for gaining important information of the functional peculiarities of proteins was proven by the comparative analysis of the $3CL^{pro}$(s) from SARS-CoV-2 and SARS-CoV, which showed that amino acids at positions 103_T and 179_T affect the nature of the interaction of the "catalytic acid" core (102_T-Core, N-terminal β-barrel) with the "supplementary" core (S-Core, C-terminal β-barrel), which ultimately results in the modulation of an enzymatic activity. It was also found that the $Val_{86}Leu$, $Lys_{88}Arg$, $Phe_{134}His$, and $Asn_{180}Lys$ mutations in these enzymes can change the orientation of the N- and C-terminal domains of $3CL^{pro}$ relative to each other, which leads to a change in catalytic activity [13]. However, $Val_{35}Thr$, $Ser_{46}Ala$, $Asn_{65}Ser$, $Ala_{94}Ser$ mutations were not included in the previous analysis, since they are located far from the catalytic tetrad. In the present work, we are filling this gap and are using the aforementioned structural biology approach to establish the structural and functional roles of these variable amino acids at positions 35, 46, 65, and 94 in the $3CL^{pro}$ sequences of SARS-CoV-2 and SARS-CoV. A comparison of the same set of 169 proteinases with the (chymo)trypsin-like serine/cysteine fold allowed us to identify new conservative elements. We found that, in addition to interdomain mobility, which could modulate catalytic activity, the $3CL^{pro}$(s) can use an autolytic loop and the unique Asp_{33}-Asn_{95} region (the Asp_{33}-Asn_{95} Zone) in the N-terminal domain.

Therefore, all 4 analyzed mutation sites are associated with the unique structure-functional features of the 3CL^{pro} from SARS-CoV-2 and SARS-CoV.

2. Results and discussion

2.1. Selection of residue Asn$_{28}$ of the 3CLpro SARS-CoV-2 as the starting amino acid for structural analysis

Earlier, a comparative structural analysis of 169 (chymo)trypsin-like proteases with the serine/cysteine fold was carried out. The analysis was based on the identification in each protein of an SCC near the catalytic tetrad and their subsequent comparison with each other [12,13]. However, in those studies we found only 4 amino acid sequence positions, in which 3CL^{pro}s from SARS-CoV-2 and SARS-CoV sequences differed from each other. Later, it became clear that coronavirus proteases modulate their enzymatic activity using amino acids that also are *not* located in structural proximity of the catalytic tetrad (this work).

The NBCZone of the SARS-CoV-2 3CL^{pro} (PDB ID 7BQY) [6], which is a part of the SCC around the catalytic nucleophile Cys$_{145}$ and the catalytic base His$_{41}$, includes the amino acids Asn$_{28}$ and His$_{163}$ (Fig. 1A). Structural analysis of the region around His$_{163}$ and adjacent fourth member of the structural catalytic tetrad, His$_{164}$, made it possible to elucidate the functional role of the amino acid differences at positions 86 and 180 [13].

The tertiary structure in the vicinity of Asn$_{28}$ has previously been subjected to rigorous structural analysis as well [12]. It has been shown that due to the presence of the Asn$_{28}$ side chain (position 43$_T$), prokaryotic and viral proteases cannot undergo a structural transition from zymogen to zyme type, which is observed in eukaryotic proteases [12]. Asn$_{28}$ is not directly involved in interactions with inhibitors [6]. In spite of this, mutation of Asn$_{28}$ to alanine disrupts dimerization (active form of enzyme) and completely inactivates the 3CLpro SARS-CoV [21]. Although Asn$_{28}$ is not directly involved in interactions with inhibitors [6], it has been shown that 8 out of 33 promising and potential 3CL^{pro} SARS-CoV-2 inhibitor molecules are in contact with the adjacent Leu$_{27}$ [11]. A hydrophobic contact between Val$_{42}$, which follows the catalytic base His$_{41}$, and Leu$_{27}$ can maintains the conformation of the polypeptide chain near the active site (Fig. 1A). Let us clarify that this statement is a hypothesis.

2.2. Val$_{42}$-Leu$_{27}$ zone of 3CLpro SARS-CoV-2

In view of the importance of the Leu$_{27}$-Asn$_{28}$ dipeptide and the Val$_{42}$ for the catalytic activity of the SARS-CoV-2 3CL^{pro}, this region of the tertiary structure; i.e., located "above" the NBCZone in Fig. 1A, has been analyzed 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 [22].

Visual analysis of the tertiary structure of the SARS-CoV-2 3CL^{pro} in the vicinity of Leu$_{27}$, Asn$_{28}$ and Val$_{42}$, allowed identification of what we refer to as the "Val$_{42}$-Leu$_{27}$ Zone" (Fig. 1A). A characteristic structural feature of the Val$_{42}$-Leu$_{27}$ Zone is the presence of four main-chain hydrogen bonds between the two pairs of amino acids: Leu$_{27}$ and Val$_{20,}$ as well as Thr$_{21}$ and Leu$_{67}$ (Table S1, row numbered 1, columns 6 and 7). These four amino acids are located at the ends of three antiparallel β-strands and form the first well-structured part of the V42-L27 Zone. The second part of the Val$_{42}$-Leu$_{27}$ Zone is the Val$_{42}$-Leu$_{67}$ loop, which is also called loop B [23] or 60-loop [24]. If Leu$_{27}$ belongs to the "first" β-strand in Fig. 1A, then Leu$_{67}$ belongs to the third β-strand, which is parallel to the first β-strand in the β-sheet. As noted above, the hydrophobic contact between Val$_{42}$ and Leu$_{27}$ unites two structurally dissimilar parts of the Val$_{42}$-Leu$_{27}$ Zone into a single compact structure. Note that Val$_{42}$ and Leu$_{27}$ of the Val$_{42}$-Leu$_{27}$ Zone are also components of the NBCZone.

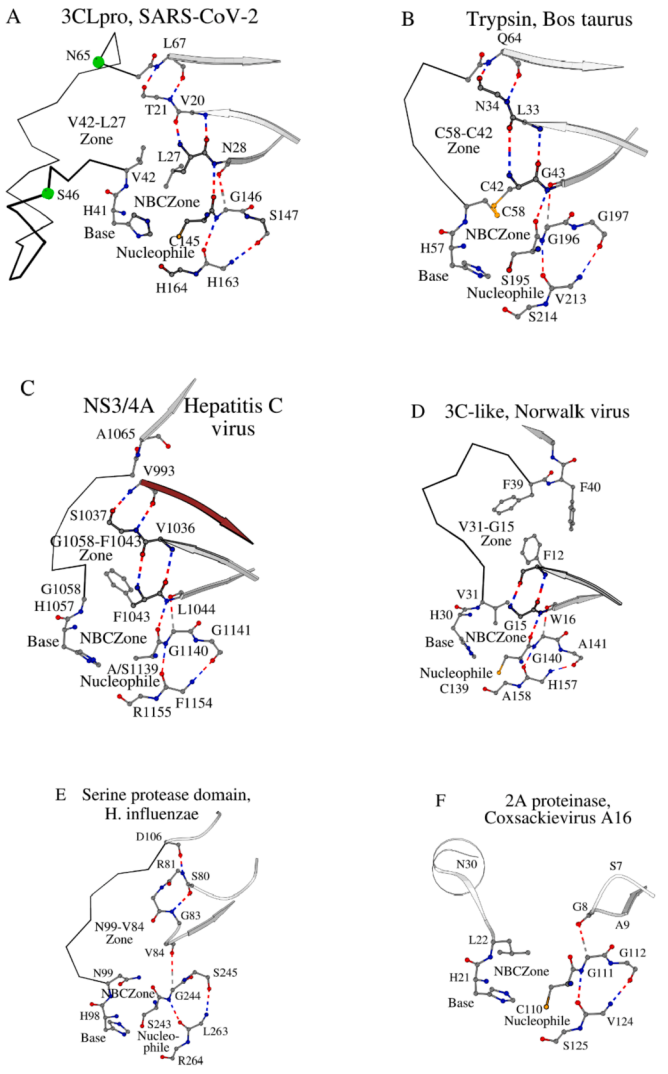

Fig. 1. (A) and (B) show the "Val$_{42}$-Leu$_{27}$ Zone" and "Cys$_{58}$-Cys$_{42}$ Zone" of the SARS-CoV-2 3CL^{pro} and Trypsin *Bos Taurus*, respectively. 58$_T$–42$_T$ Zone is the representative zone for 148 (chymo)trypsin-like proteinases with serine/cysteine fold. 58$_T$–42$_T$ Zone consists of the two part: well-structured (four hydrogen bonds) part, which consists of the residues in positions 42$_T$, 33$_T$, 34$_T$ and 64$_T$, and the 58$_T$–64$_T$ loop, variable in length. The positions of the C$_α$-atoms of the amino acids Ser$_{46}$ and Asn$_{65}$ of the SARS-CoV-2 3CL^{pro} (A), which have changed in comparison with the SARS-CoV 3CL^{pro}, are marked with large green circles. The location of the 58$_T$–42$_T$ Zone in relation to the NBCZone formed by amino acids: 42$_T$, 43$_T$, 57$_T$ (catalytic base), 58$_T$, 195$_T$ (catalytic nucleophile), 196$_T$, 197$_T$ and 213$_T$ is also shown. Some variations in the organization of the well-structured part of the 58$_T$–42$_T$ Zone: (C) shows 3D complex of the NS3 protease and NS4A cofactor (brown); (D) 3C-like viral cysteine protease shows another variant of structural organization of the well-structured part of the 58$_T$–42$_T$ Zone. The second and third β-strands are connected by aromatic residues; (E) Instead of four hydrogen bonds, only two remained in this variant of the 58$_T$–42$_T$ Zone; and (F) The 2A proteinase has no structural analogue of the 58$_T$–42$_T$ Zone at all. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.3. 58$_T$–42$_T$ zone of (chymo)trypsin-like proteinases with the serine/cysteine fold

The SARS-CoV-2 3CL^{pro} structure is one of 170 structures studied by us that have the same b.47 fold classification within the Structural Classification of Proteins–extended (SCOPe) (https://scop.berkeley.edu/ [25]). The tertiary structure of trypsin (PDB ID 4I8H) [15] is a representative example of this fold. Val$_{42}$ and Leu$_{27}$ of the SARS-CoV-2

3CL^{pro} structurally correspond to Cys_{58} and Cys_{42} of trypsin, respectively (Fig. 1B). Visual analysis of the tertiary structure of trypsin in the vicinity of Cys_{42}, Gly_{43}, and Cys_{58}, allowed identification of the "Cys_{58}-Cys_{42} Zone", whose characteristic structural feature is the presence of four main-chain hydrogen bonds between two pairs of amino acids: Cys_{42} and Leu_{33}, as well as Asn_{34} and Gln_{64} (Table S1, row numbered 13, columns 6 and 7).

At first glance, the Val_{42}-Leu_{27} and Cys_{58}-Cys_{42} Zones are structurally similar (see Fig. 1A and B). However, one can see two fundamental differences. The hydrophobic contact between Val_{42} and Leu_{27} that closes the Val_{42}-Leu_{27} Zone is replaced by the Cys_{58}-Cys_{42} disulfide bond, which is conserved in eukaryotes [12]. The second significant difference between the two zones is the different lengths of the Val_{42}-Leu_{67} and Cys_{58}-Gln_{64} loops: 26 and 7 residues, respectively (Table S1, rows numbered 1 and 13, column 9).

The Val_{42}-Leu_{27} Zone for the 3CL^{pro} SARS-CoV-2 can be also written in a universal form as the 58_T–42_T Zone. Unlike prokaryotic and viral proteases, most eukaryotic proteases have disulfide-linked cysteines at positions 42_T and 58_T [12]. The need for a strong side-chain interaction forming the single ring-shaped structure (Zone) is apparently directly related to the functioning of eukaryotic proteases. In fact, it has been shown that the presence or absence of the Cys_{42T}-Cys_{58T} disulfide bond affects the overall thermal stability of trypsin [26].

The results of the structural analysis of the 58_T–42_T Zone for all 170 structures are presented in Table S1. 148 structures (Table S1, rows numbered 1–148) have an identical structural organization of the first well-structured part of the 58_T–42_T Zone. This large group of proteins includes all eukaryotic and prokaryotic proteases, TA and [KR]P groups of the viral serine proteases, [TA]N and [ΨC][PQ] groups of the viral cysteine proteases and inactive proteases. The names of the groups and the list of the corresponding proteases are taken from the study on the characterization of the NBCZones in (chymo)trypsin-like proteinases with the serine/cysteine fold [12]. The [ΨC][PQ] group includes twelve viral cysteine proteases, which lack the catalytic acid [13]; i.e., instead of the catalytic triad, they have a catalytic dyad in the active site (Table S1, rows numbered 1–12). Eleven out of twelve proteins are coronavirus proteases. It is important to note that, despite the structural identity of the well-structured part of the 58_T–42_T Zone of this group of 148 proteases, the 58_T–64_T loop varies significantly in length from 4 to 37 residues (Table S1, column 9).

Some variation in the organization of the well-structured part of the 58_T–42_T Zone is demonstrated by the proteases belonging to the viral serine proteases, [ST]Ψ group (Table S1, rows numbered 149–156). Unlike the 148 proteases considered earlier, the PDB files of 8 proteases belonging to this group contain non-covalent, heterodimer complexes formed by two proteins, the N-terminal serine protease domain of NS3 (catalytic subunit) and the NS4A cofactor (activation subunit). Fig. 1C illustrates the 58_T–42_T (Gly_{1058}-Phe_{1043}) Zone for the complex of the NS3 protease and NS4A protein from the hepatitis C virus (PDB ID 3SU6) [27]. The formation of a heterodimeric complex results in a structure, where a β-strand from NS4A (brown in Fig. 1C) is located in the Gly_{1058}-Phe_{1043} Zone structure, replacing the third β-strand of NS3 that begins with Ala_{1065}. The inclusion of the NS4A cofactor in the NS3 protease structure increases the well-structured part of the 58_T–42_T Zone by one amino acid: Val_{993}. Therefore, in this case, the length of the Gly_{1058}-Ala_{1065} loop is 8 residues (Table S1, row numbered 154, column 9).

The viral cysteine proteases, T[TSA] group (Table S1, rows numbered 157–161), show another variant of the organization of the well-structured part of the 58_T–42_T Zone (Fig. 1D). Instead of two hydrogen bonds formed between the main-chain atoms that connect the second and third β-strands, the role of such an interchain clamp in the 3C-like protease from the Norwalk virus (PDB ID 5E0G) [28] is performed by aromatic contacts among Phe_{12}, Phe_{39}, and Phe_{40}. The length of the Val_{31}-Phe_{39} loop is 9 residues (Table S1, row numbered 159, column 9).

The distinctive structural characteristics of the adhesion and penetration protein autotransporter (Fig. 1E, PDB ID 3SYJ, [29]), which belongs to the SPATE family [30,31], were previously studied [12]. Six compared proteins belonging to this family (Table S1, rows numbered 162–167) have a main-chain conformation at position 43_T (position 84 in Fig. 1E) that is different from that found in all other proteins analyzed so far. As a result, instead of four hydrogen bonds, only two remained in the Asn_{99}-Val_{84} Zone of these proteins. Furthermore, two fragments of the polypeptide chain have a loop-like conformation, instead of the β-structural one. The length of Asn_{99}-Asp_{106} loop is 8 residues (Table S1, row numbered 162, column 9).

Three 2A proteinases (Table S1, rows numbered 168–170) have the same structural characteristics at position 43 as the proteases of the SPATE family. However, the N-terminal domain in these proteinases is not a β-barrel, but a four-stranded antiparallel β-sheet [32]. As a result, the 2A proteinase has no structural analogy to the 58_T–42_T Zone (Fig. 1F, PDB ID 4MG3, [33]).

Summarizing everything said in this section, we can conclude that amino acid residues at positions 33_T, 34_T, and 43_T (Val_{20}, Thr_{21}, and Asn_{28} in the 3CL^{pro} SARS-CoV-2 (Fig. 1A)) should also be included in the conserved structural core of the (chymo)trypsin-like proteinases with the serine/cysteine fold. This conclusion is fully consistent with the results of the structural analysis of 13 widely divergent serine proteinases (see Fig. 10 in [16]). The most frequently observed changes in the local secondary structure are found near position 64_T. These structural rearrangements change the nature of the proteolytic activity of the corresponding proteins, but do not eliminate it.

2.4. Val_{42}-Leu_{67} loop of 3CLpro SARS-CoV-2

The Val_{42}-Leu_{67} loop of the SARS-CoV-2 3CL^{pro} contains 26 amino acids (Table S1, row numbered 1, columns 8 and 9) and the residues at positions 46 and 65 (Figs. 1A and 2A) were the subject of our structural analysis. This loop is long and contains 2 of 8 positions, in which amino acid residues are different in the 3CL^{pro}s from SARS-CoV-2 and SARS-CoV. Therefore, we studied the structural and functional role of this loop. Since the 58_T–64_T loop length in 170 proteases varies significantly from 4 to 37 residues, it was impossible to compare them structurally. For this reason, we analyzed structural and functional role of the Val_{42}-Leu_{67} (58_T–64_T) loop of the SARS-CoV-2 3CL^{pro}.

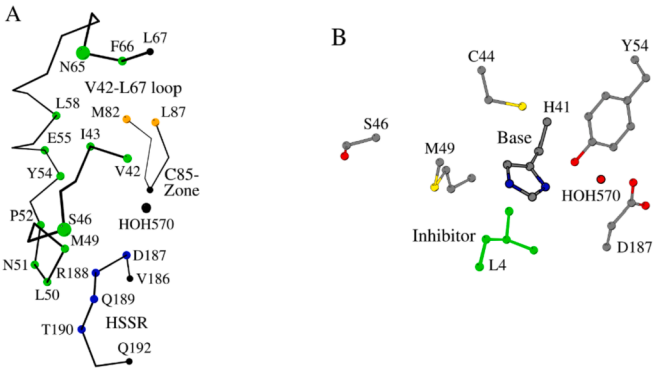

Fig. 2. (A) Hydrophobic interactions (small green and orange circles) between Val_{42}-Leu_{67} loop and Cys_{85}-Zone in the SARS-CoV-2 3CL^{pro}. The positions of the C_α-atoms of the amino acids Ser_{46} and Asn_{65} of the SARS-CoV-2 3CL^{pro} (A), which have changed in comparison with the SARS-CoV 3CL^{pro}, are marked with large green circles. Polar contacts (small blue and green circles) between the conserved parts of the interdomain loop (IDL): Val_{186}-Gln_{192}, Horse Shoe-Shaped Region (HSSR) and Val_{42}-Leu_{67} loop. (B) In the complex between ligand and SARS-CoV-2 3CL^{pro} hydrophobic amino acids Cys_{44}, Met_{49} and Tyr_{54} of the Val_{42}-Leu_{67} loop and the Leu_4 residue of the ligand interact with each other and maintain the 3D position of the catalytic histidine in position 41. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.4.1. Hydrophobic interactions between the Val_{42}-Leu_{67} loop and Cys_{85}-zone

The tertiary conformation of the Val_{42}-Leu_{67} loop is stabilized by numerous hydrophobic contacts of the amino acids Val_{42}, Ile_{43}, Tyr_{54}, Leu_{58}, and Phe_{66} with the N- and C-terminal hydrophobic amino acids Met_{82} and Leu_{87} of the Cys_{85}-Zone (Fig. 2A). The Cys_{85}-Zone is actually a structural analogue of the catalytic acid (position 102_T) zone [13]. With a few exceptions, a similar type of interactions between the 58_T–64_T loop and 102_T-Zone is typical for all (chymo)trypsin-like proteinases with the serine/cysteine fold (data not shown).

The mutation $Asn_{65}Ser$ of the SARS-CoV-2 $3CL^{pro}$ is located near the C-terminal end of the Val_{42}-Leu_{67} loop. Positions 65 and 66 are adjacent in the amino acid sequence, therefore, it can be reasonably assumed that $Asn_{65}Ser$ sequence difference would affect the contact between the Val_{42}-Leu_{67} loop and the Cys_{85}-Zone through the amino acid Phe_{66}. The change in the nature of this contact compared to homologous contact in the $3CL^{pro}$ of the SARS-CoV is caused both by a change in the size of the side chain group of the amino acid at position 65 and by the appearance of an additional positively charged group of NH_2 atoms.

2.4.2. The N-end half of the Val_{42}-Leu_{67} loop interacts with ligand

An analysis of contacts in the complex (PDB ID 7BQY) between the PRD_002214 ligand and SARS-CoV-2 $3CL^{pro}$ performed by means of the LPC software [22] showed that Cys_{44}, Met_{49}, and Tyr_{54} of the Val_{42}-Leu_{67} loop and residue Leu_4 of the ligand interact with each other and maintain the 3D positioning of the catalytic histidine at the sequence position 41 mainly via hydrophobic interactions (Fig. 2B).

These results are consistent with the data presented in the recent review of application of various computational methods (see Table 1 in [11]) for the identification of amino acids that are involved in contacts between 33 antiviral ligands and the SARS-CoV-2 $3CL^{pro}$. Compared to other residues of the Val_{42}-Leu_{67} loop, Met_{49} (17 of 33 cases) and Tyr_{54} (8 of 33 cases) are the most frequently observed as residues participating in contacts. Therefore, the N-terminal half of the Val_{42}-Leu_{67} loop is responsible for modulating the catalytic activity of the SARS-CoV-2 $3CL^{pro}$. Let us clarify that this statement is a hypothesis. It is possible that the observed replacement of the hydrophobic amino acid alanine of in the SARS-CoV-2 $3CL^{pro}$ by the polar serine at position 46 will affect this modulation. Ser_{46} and Met_{49} are located on a short, one-turn α-helix Thr_{45}-Met_{49}. The side-chains of Ser_{46} and Met_{49} groups are in tight contact with each other. Apparently, Met_{49} is a key intermediate amino acid, through which a change in the residue at position 46 affects the catalytic base.

Fig. 3 shows the structural alignment of the Val_{42}-Leu_{67} loop from the SARS-CoV-2 $3CL^{pro}$, with loops from ten coronavirus proteases and the loop of one 3Cl protease from the Cavally virus (Table S1, rows numbered 1–12). The alignment was built using the Protein Structure Comparison Server Dali [34]. The length of loops in coronavirus proteases corresponding to the Val_{42}-Leu_{67} loop of the SARS-CoV-2 $3CL^{pro}$ is fairly uniform and varies from 23 to 26 amino acids (Table S1, columns 8 and 9). It is important to note that the alignment contains deletions at the N-terminus within a narrow range of positions 45–48 (SARS-CoV-2 $3CL^{pro}$ numbering). Therefore, the use of the position 46 to modulate the catalytic activity of coronavirus proteases is quite possible.

2.4.3. The N-end half of the Val_{42}-Leu_{67} loop interacts with the N-terminus of the interdomain loop (IDL)

The SARS-CoV-2 $3CL^{pro}$ structure is comprised of two β-barrels (domains I and II), bringing the catalytic residues together at their interface (canonical (chymo)trypsin-like structure, residues Ser_1-Asp_{176}), and an additional C-terminal extension [6]. Asp_{176} is the last amino acid of the His_{164}-Core [13]. The C-terminal extension that starts with Leu_{177} contains interdomain loop (IDL, residues 184–199) (Fig. 2A). The conserved Val_{186}-Gln_{192} Horse-Shoe-Shaped Region (HSSR) is a part of the IDL [35]. Asp_{187} and the water molecule HOH_{570} compensate for the absence of catalytic acid in the SARS-CoV-2 $3CL^{pro}$ (Fig. 2A). The residues Ser_{46}, Met_{49}, Leu_{50}, Asn_{51}, Pro_{52} and Tyr_{54} of the Val_{42}-Leu_{67} loop are in contact with Asp_{187}, Arg_{188}, Gln_{189} and Thr_{190} of the IDL. It has been suggested that the IDL may represent a regulatory site, since it does not contain the amino acids of the catalytic triad [35].

Therefore, the $Ser_{46}Ala$ sequence difference between the SARS-CoV-2 and SARS-COV $3CL^{pro}$s, located near position 58_T, can affect the catalytic activity of the SARS-CoV-2 $3CL^{pro}$ not only by itself, but also indirectly through the IDL C-terminal extension. As noted above, the residue Ser_{46} is an essential functional element affecting the binding process of the ligand [11,36]. It was also suggested that the mutation $Ser_{46}Ala$ may increase the contribution of other hydrophilic amino acids to the structure of the active site [37].

2.4.4. $Ser_{46}Ala$ and $Asn_{65}Ser$, and probable autolysis regulation in the SARS-CoV-2 $3CL^{pro}$

There is another possibility related to the functional consequences of the $Ser_{46}Ala$ and $Asn_{65}Ser$ sequence differences between the SARS-COV-2 and SARS-COV $3CL^{pro}$ proteins. Dissolved trypsin is known to undergo autolytic degradation [17]. Three important autolysis sites have been reported for bovine trypsin: Lys_{60}-Ser_{61}, Arg_{117}-Val_{118} and Lys_{145}-Ser_{146}. In rat (chymo)trypsin the autolytic site Phe_{114} (Leu_{114} in trypsin) was identified [19]. Mutant $Phe_{114}Ile$ has the same enzymatic activity and molecular stability as the wild-type enzyme, but exhibites significantly slower autolytic inactivation. In rat trypsin, Lys_{61} (Lys_{60} in bovine trypsin) and Arg_{117} are both replaced by Asn [18]. Kinetic parameters of the mutants did not change, but the autolysis rate slowed down significantly. Lys_{60} of bovine trypsin belongs to the 58_T–64_T loop. Since the 58_T–64_T loop of the SARS-CoV-2 $3CL^{pro}$ differs significantly in amino acid sequence from 58_T–64_T loop of bovine trypsin, it is not known

```
SARS-2   7BQY_A  42 VICTSEDMLNPNYEDLLIRKS-NHNFL 67
SARS-1   6XHN_A  42 VICTAEDMLNPNYEDLLIRKS-NHSFL 67
MERS     5WKK_A  42 VMCPADQLSDPNYDALLISMT-NHSFS 67
HKU4     2YNA_A  42 IMCPADQLTDPNYDALLISKT-NHSFI 67
229E     2ZU2_A  42 VIASN-TTSAIDYDHEYSIMR-LHNFS 66
NL63     3TLO_A  42 VIAPS-TTVLIDYDHAYSTMR-LHNFS 66
IBV      2Q6F_A  42 VLGK---FSGDQWGDVLNLAN-NHEFE 64
MHV      6JIJ_A  42 VICS--DMTDPDYPNLLCRVT-SSDFC 65
PEDV     4XFQ_A  42 VIASS-TTSTIDYDYALSVLR-LHNFS 66
FIPV     4ZRO_A  42 VIASD-TSRVINYENELSSVR-LHNFS 66
TGEV     1LVO_A  42 VIASD-TTRVINYENEMSSVR-LHNFS 66
CAVV     5LAC_B  49 LFG----SKKQEFEACYNNGKGLLNCK 71
```

Fig. 3. The structure-based multiple sequence alignment of Val_{42}-Leu_{67} loop of SARS-CoV-2 $3CL^{pro}$, ten corresponding coronavirus proteases and one 3Cl protease from Cavally virus loops.

whether there are autolytic sites in the 58_T–64_T loop of the coronavirus proteases. If there are any, then the $Ser_{46}Ala$ and/or $Asn_{65}Ser$ sequence differences could alter the functional life-time of the SARS-CoV-2 $3CL^{pro}$ without changing its activity. This structural assumption is hypothetical and should be verified by appropriate experiments.

2.5. Structural relationship of amino acids Val35 and Ala94 of the SARS-CoV-2 $3CL^{pro}$

The two remaining positions 35 and 94 of the SARS-CoV-2 $3CL^{pro}$ (PDB ID 7BQY), in which amino acid changes are observed compared to the SARS-CoV $3CL^{pro}$ (PDB ID 6XHN), are located respectively at the N- and C-terminal ends of the functionally important β-strands Val_{35}-Pro_{39} and Val_{86}-Lys_{90} (Fig. 4A). The first of these two β-strands contains catalytic histidine His_{41} near its C-terminal end, and the second β-strand contains at its N-terminal Cys_{85} that replaces the catalytic acid. The last β-strand also contains the positions 86 and 88 where SARS-CoV-2 and SARS-CoV sequences have different amino acids. In the SARS-CoV-2 $3CL^{pro}$ 3D structure, these two β-strands form an antiparallel β-hairpin, held together by six interchain hydrogen bonds, starting from the bond N/Arg_{40}-O/Cys_{85} (3.0 Å) and ending with the O/Asp_{34}-N/Val_{91} bond (3.0 Å) (Fig. 4A). The 3D contacts between these fragments, being separated within the amino acid sequence, take place within the 3D structure, since two more hydrogen bonds are observed: O/Asp_{33}-CA/Ala_{94} and O/Asp_{33}-N/Asn_{95}. In addition, the tetrapeptide Leu_{32}-Val_{35} forms a β-turn. As a result, two fragments Asp_{33}-Asp_{34} and Val_{91}-Asn_{95} form a Asp_{33}-Asn_{95} Zone. This zone directly contains the position of interest to us: Ala_{94}. Val_{35} lies at the border of the Asp_{33}-Asn_{95} Zone and an antiparallel β-hairpin.

Due to the fact that positions 35 and 94 are spatially close to each other in the structure of the Asp_{33}-Asn_{95} Zone from the SARS-CoV-2 $3CL^{pro}$, it was of interest to analyze the equivalent region from all the 3D structures included in the superfamily of (chymo)trypsin-like proteinases with the serine/cysteine fold. The results of the analysis of 128 proteinases (75% of the total set) are collected in Table S2. In the structure of trypsin, the amino acids Ser_{49} and Ala_{112} respectively correspond to the Asp_{33} and Asn_{95} of the SARS-CoV-2 $3CL^{pro}$ (Fig. 4B). Fig. 4B shows the 3D organization of the Ser_{49}-Ala_{112} Zone in trypsin. The first 108_T–112_T fragment consists of 5 residues, and the second 49_T–50_T fragment is formed by two amino acids. Therefore, the SARS-CoV-2 $3CL^{pro}$ and trypsin have a structurally identical Asp_{33}-Asn_{95} and Ser_{49}-Ala_{112} Zones, which are located far from the catalytic tetrad. In the course of the structural analysis, we identified 104 examples of such 3D sub-structures (Table S2, rows numbered 1–11, 13–79, 88–91, 104, 141–147 and 157–170). In addition to the identity of the 49_T–112_T Zones, 104 proteases have an identical 3D arrangement of the 49_T–112_T Zone in relation to the catalytic base at position 57_T and the catalytic acid at position 102_T (or its structural analog). This group of proteases was given the code name "(5 + 2)".

Eight other proteases form a (8 + 2) group (Table S2, rows numbered 133–140). Fig. 4C shows an example of the 49_T–112_T Zone for the nuclear inclusion protein A from the Tobacco vein mottling virus from a (8 + 2) group. Apparently, the presence of 8 amino acids in the first fragment of the 49_T–112_T Zone is the maximum possible number. This conclusion is derived from the structural observations, where the extention of the first fragment only by one amino acid leads to the impossibility of the formation of contact between the C_α-atom of the residue at position 111_T and the main-chain oxygen atom of the residue in position 49_T. However, four proteases (Table S2, rows numbered 27, 53, 67, and 76) overcome this restriction as their 49_T–112_T Zone is formed through the use of a disulfide bond Cys_{111}-Cys_{50}, thereby defining the (9 + 2) group (Fig. 4D). Further lengthening of the first fragment of the 49_T–112_T Zone (Fig. 4E) to a (11 + 2) group results in a situation, where the disulfide bond Cys_{111}-Cys_{50} is no longer used for mutual structural stabilization of the ends of the 49_T–112_T Zone (Table S2, rows numbered 117–121).

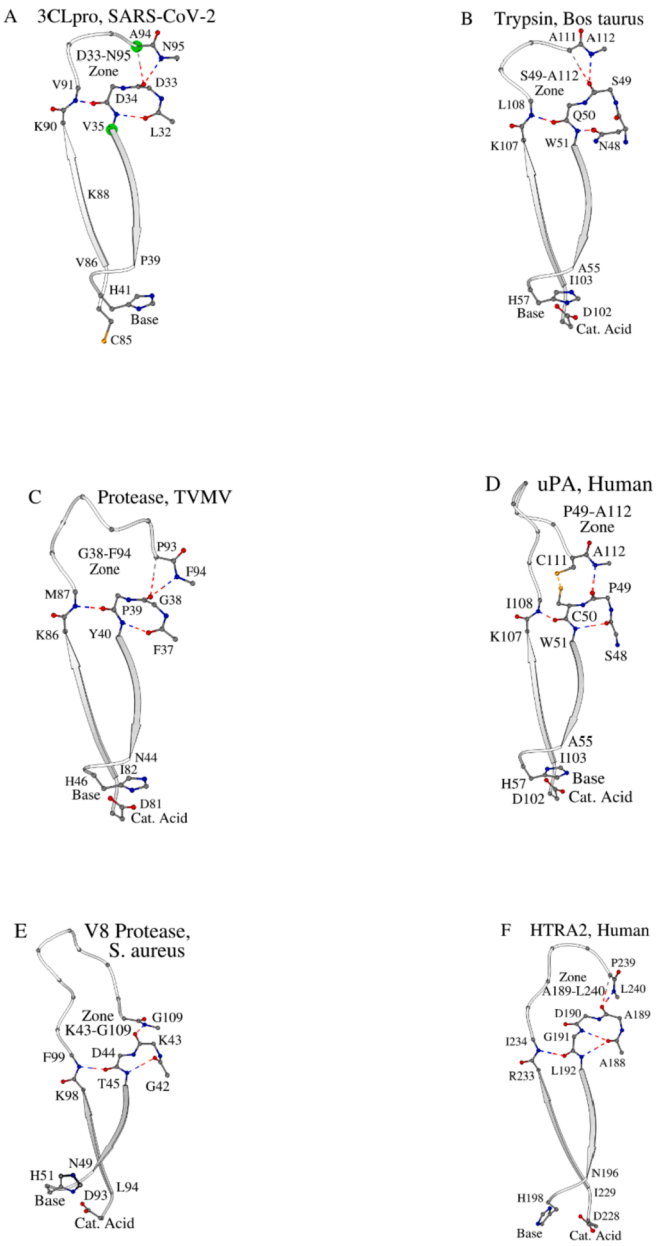

Fig. 4. (A) and (B) show the "Asp_{33}-Asn_{95} Zone" and "Ser_{49}-Ala_{112} Zone" of the SARS-CoV-2 $3CL^{pro}$ and Trypsin *Bos Taurus*, respectively. 49_T–112_T Zone is the representative zone for 104 (chymo)trypsin-like proteinases with serine/cysteine fold: the code name of this zone is "(5 + 2)". The positions of the C_α-atoms of the amino acids Val_{35} and Ala_{94} of the SARS-CoV-2 $3CL^{pro}$ (A), which have changed in comparison with the SARS-CoV $3CL^{pro}$, are marked with large green circles. The location of the 49_T–112_T Zone in relation to the catalytic base and catalytic acid is also shown. (C) a zone with the code name is "(8 + 2)"; (D) a zone with the code name "(9 + 2)" and disulfide bond Cys_{111}-Cys_{50}; (E) a zone with the code name: "(11 + 2)", and the lack of 49_T–111_T contact or disulfide bond Cys_{50}-Cys_{111}; and (F) a zone with the code name: "(21 + 3)", a special version of the zone in which the second fragment of the zone has 3 amino acids residues instead of two. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The last seven examples of proteases in Table S2 show that the 49_T–112_T Zone can be modified not only by changing the amino acid length of the first fragment but also by extending the second fragment by one residue. Fig. 4F shows serine protease HTRA2 from the *Homo sapiens* as an example of the Ala_{189}-Leu_{240} Zone from such a (7 + 3) group. Four proteases have been found with a similar zone arrangement (Table S2,

rows numbered 80–82, and 148). The last three prokaryotic proteases form a (21 + 3) group (see Table S2, rows numbered 101–103) and demonstrate the variant of the zone, in which the second fragment has 3 amino acids allowing for significantly wider variations in the length of the first fragment compared to the zone, in which the second fragment has 2 amino acids.

In the Section 2.4.4. we cited the works in which Arg_{117}-Val_{118} dipeptide of trypsin was mentioned as an autolysis site [18]. There appears to be a close relationship between the 49_T–112_T Zone and the autolysis process. At positions 35 and 94, the 3CLpro SARS-CoV-2 has two small hydrophobic amino acids Val_{35} and Ala_{94} instead of two small polar residues Thr_{35} and Ser_{94} in the 3CLpro SARS-CoV. It is possible that such changes in amino acids give the 3CLpro SARS-CoV-2 structure additional hydrophobic resistance to autolysis that is amino acid replacements at positions 35 and 94 of the 3CLpro SARS-CoV-2 can change the autolysis rate. This structural assumption is hypothetical and should be verified by appropriate experiments.

3. Conclusions

The 3D structures of (Chymo)trypsin-like $3CL^{pro}$ from SARS-CoV-2 and SARS-CoV have different amino acid residues at 8 positions of their amino acid sequences: $Val_{35}Thr$, $Ser_{46}Ala$, $Asn_{65}Ser$, $Val_{86}Leu$, $Lys_{88}Arg$, $Ala_{94}Ser$, $Phe_{134}His$, and $Asn_{180}Lys$ (Fig. 5). These residues can be divided into two structural groups. The first group includes 4 amino acids at positions 86, 88, 134, and 180 that form the structural catalytic core. The second group includes 3 amino acids at positions 46, 65, and 94 located in the loop regions. This group is also adjoined by the amino acid at position 35. The first group of residues modulates the catalytic activity of the $3CL^{pro}$ by changing the nature of the interaction between the N- and C-terminal β-barrels. The second group includes 3 amino acids at positions 46, 65, and 94 located in the loop regions (Fig. 5A and B). This group is also adjoined by the amino acid at position 35 (Fig. 5A and B). The first group of residues modulates the catalytic activity of the $3CL^{pro}$ by changing the nature of the interaction between the N- and C-terminal β-barrels. The second group of residues can be involved in modulation of the activity using such unique features of the tertiary structure as the existence of the C-terminal extension (IDL loop). In addition, the amino acids of the second group and the sites of possible autolysis of the $3CL^{pro}$ intersect in the amino acid sequence, which suggests that the process of autolysis of proteases plays an essential role in modulating catalytic activity of this important viral protease. This result opens up a new field of scientific research for those researchers involved in protein characterization and inhibition of the $3CL^{pro}$ from SARS-CoV-2.

4. Materials and methods

4.1. The choice of structures to be analyzed

In this work, as in the previous publications [12,13], the same dataset of 170 3D structures of the (chymo)trypsin-like proteinases with serine/cysteine fold was used. In these two publications, all the procedural details of the compilation of the required set of 3D structures are described in detail.

4.2. Modeling software for structural analysis

Structure visualization and structural analysis of interactions between amino acids in proteins (hydrogen bonds, hydrophobic, other types of weak interactions) were 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 [22]. Figures are drawn with MOLSCRIPT [38].

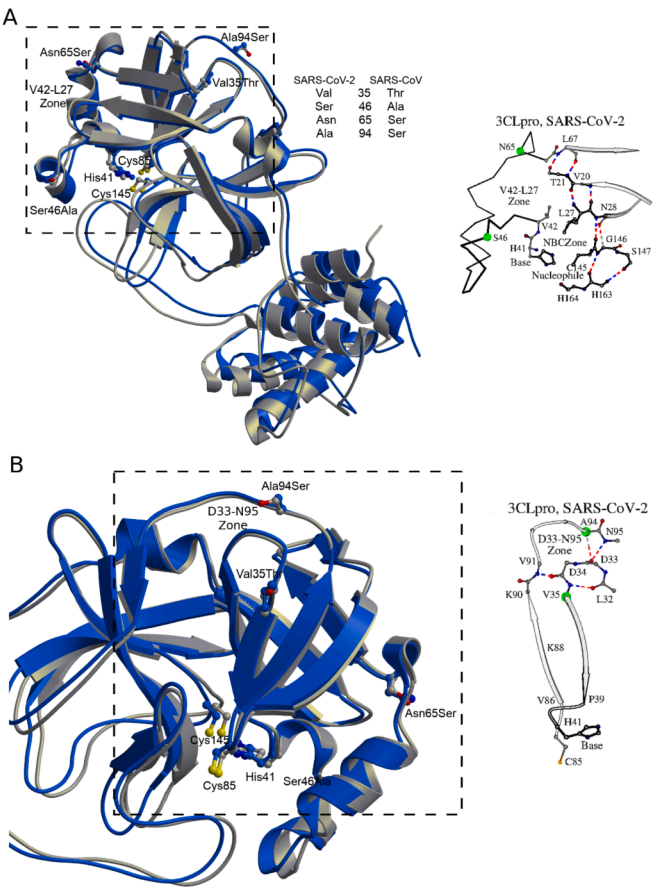

Fig. 5. Superposition of three-dimensional (3D) structures of the 3CLpro(s) from SARS-CoV (PDB ID 1UJ1; shown in gray) and SARS-CoV-2 (PDB ID 6LU7; shown in blue). (A) Shows the location within the entire structure of the V42-L27 Zone (for reference, see Fig. 1A), the catalytic triad (H41, C85 and C145 in SARS-CoV-2), and the variable amino acids at the positions 35, 46, 65 and 94. (B) Shows the location of the D33-N95 Zone (for reference, see Fig. 4A), the catalytic triad (H41, C85 and C145 in SARS-CoV-2), and the variable amino acids at the positions 35, 46, 65 and 94. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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

Alexander I. Denesyuk: Study design, Formal analysis, Methodology, Visualization, Writing – Original Draft, Writing – Review & Editing; **Eugene A. Permyakov**: Formal analysis, Writing – Review & Editing; **Mark S. Johnson**: Formal analysis, Methodology, Writing – Original Draft; **Sergei E. Permyakov**: Formal analysis, Writing – Review & Editing; **Konstantin Denessiouk**: Formal analysis, Methodology, Visualization, Writing – Original Draft, Writing – Review & Editing; **Vladimir N. Uversky**: Study design, Formal analysis, Methodology, Visualization, Investigation, Writing – Original Draft, Writing – Review & Editing.

Declaration of competing interest

The authors declare no conflict of interest.

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 stategic 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.ijbiomac.2021.11.043.

References

[1] 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).
[2] 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.
[3] A. Hegyi, J. Ziebuhr, Conservation of substrate specificities among coronavirus main proteases, J. Gen. Virol. 83 (Pt 3) (2002) 595–599.
[4] V. Thiel, K.A. Ivanov, A. Putics, T. Hertzig, B. Schelle, S. Bayer, B. Weissbrich, E. J. Snijder, H. Rabenau, H.W. Doerr, A.E. Gorbalenya, J. Ziebuhr, Mechanisms and enzymes involved in SARS coronavirus genome expression, J. Gen. Virol. 84 (Pt 9) (2003) 2305–2315.
[5] H. Yang, M. Yang, Y. Ding, Y. Liu, Z. Lou, Z. Zhou, L. Sun, L. Mo, S. Ye, H. Pang, G. F. Gao, K. Anand, M. Bartlam, R. Hilgenfeld, Z. Rao, The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor, Proc. Natl. Acad. Sci. U. S. A. 100 (23) (2003) 13190–13195.
[6] 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.
[7] H.M. Berman, T. Battistuz, T.N. Bhat, W.F. Bluhm, P.E. Bourne, K. Burkhardt, Z. Feng, G.L. Gilliland, L. Iype, 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 No 1) (2002) 899–907.
[8] 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.
[9] N. Chitranshi, V.K. Gupta, R. Rajput, A. Godinez, K. Pushpitha, T. Shen, M. Mirzaei, Y. You, D. Basavarajappa, V. Gupta, S.L. Graham, Evolving geographic diversity in SARS-CoV2 and in silico analysis of replicating enzyme 3CL(pro) targeting repurposed drug candidates, J. Transl. Med. 18 (1) (2020) 278.
[10] Y.W. Chen, C.B. Yiu, K.Y. Wong, Prediction of the SARS-CoV-2 (2019-nCoV) 3C-like protease (3CL (pro)) structure: virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates, F1000Res 9 (2020) 129.
[11] E. Singh, R.J. Khan, R.K. Jha, G.M. Amera, M. Jain, R.P. Singh, J. Muthukumaran, A.K. Singh, A comprehensive review on promising anti-viral therapeutic candidates identified against main protease from SARS-CoV-2 through various computational methods, J. Genet. Eng. Biotechnol. 18 (1) (2020) 69.
[12] 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.
[13] A.I. Denesyuk, S.E. Permyakov, M.S. Johnson, E.A. Permyakov, V.N. Uversky, K. Denessiouk, Structural leitmotif and functional variations of the structural catalytic core in (chymo)trypsin-like serine/cysteine fold proteinases, Int. J. Biol. Macromol. 179 (2021) 601–609.
[14] R.L. Hoffman, R.S. Kania, M.A. Brothers, J.F. Davies, R.A. Ferre, K.S. Gajiwala, M. He, R.J. Hogan, K. Kozminski, L.Y. Li, J.W. Lockner, J. Lou, M.T. Marra, L. J. Mitchell Jr., B.W. Murray, J.A. Nieman, S. Noell, S.P. Planken, T. Rowe, K. Ryan, G.J. Smith 3rd, J.E. Solowiej, C.M. Steppan, B. Taggart, Discovery of ketone-based covalent inhibitors of coronavirus 3CL proteases for the potential therapeutic treatment of COVID-19, J. Med. Chem. 63 (21) (2020) 12725–12747.
[15] D. Liebschner, M. Dauter, A. Brzuszkiewicz, Z. Dauter, On the reproducibility of protein crystal structures: five atomic resolution structures of trypsin, Acta Crystallogr. D Biol. Crystallogr. 69 (Pt 8) (2013) 1447–1462.
[16] A.M. Lesk, W.D. Fordham, Conservation and variability in the structures of serine proteinases of the chymotrypsin family, J. Mol. Biol. 258 (3) (1996) 501–537.
[17] S. Maroux, P. Desnuelle, On some autolyzed derivatives of bovine trypsin, Biochim. Biophys. Acta 181 (1) (1969) 59–72.
[18] E. Varallyay, G. Pal, A. Patthy, L. Szilagyi, L. Graf, Two mutations in rat trypsin confer resistance against autolysis, Biochem. Biophys. Res. Commun. 243 (1) (1998) 56–60.
[19] A. Bodi, G. Kaslik, I. Venekei, L. Graf, Structural determinants of the half-life and cleavage site preference in the autolytic inactivation of chymotrypsin, Eur. J. Biochem. 268 (23) (2001) 6238–6246.
[20] L. Jiang, C. Yuan, M. Huang, A general strategy to inhibit serine protease by targeting its autolysis loop, FASEB J. 35 (2) (2021), e21259.
[21] J. Barrila, S.B. Gabelli, U. Bacha, L.M. Amzel, E. Freire, Mutation of Asn28 disrupts the dimerization and enzymatic activity of SARS 3CL(pro), Biochemistry 49 (20) (2010) 4308–4317.
[22] V. Sobolev, A. Sorokine, J. Prilusky, E.E. Abola, M. Edelman, Automated analysis of interatomic contacts in proteins, Bioinformatics 15 (4) (1999) 327–332.
[23] J.J. Perona, C.S. Craik, Evolutionary divergence of substrate specificity within the chymotrypsin-like serine protease fold, J. Biol. Chem. 272 (48) (1997) 29987–29990.
[24] P. Goettig, H. Brandstetter, V. Magdolen, Surface loops of trypsin-like serine proteases as determinants of function, Biochimie 166 (2019) 52–76.
[25] 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), D304-9.
[26] T.T. Baird Jr., W.D. Wright, C.S. Craik, Conversion of trypsin to a functional threonine protease, Protein Sci. 15 (6) (2006) 1229–1238.
[27] K.P. Romano, A. Ali, C. Aydin, D. Soumana, A. Ozen, L.M. Deveau, C. Silver, H. Cao, A. Newton, C.J. Petropoulos, W. Huang, C.A. Schiffer, The molecular basis of drug resistance against hepatitis C virus NS3/4A protease inhibitors, PLoS Pathog. 8 (7) (2012), e1002832.
[28] P.M. Weerawarna, Y. Kim, A.C. Galasiti Kankanamalage, V.C. Damalanka, G. H. Lushington, K.R. Alliston, N. Mehzabeen, K.P. Battaile, S. Lovell, K.O. Chang, W. C. Groutas, Structure-based design and synthesis of triazole-based macrocyclic inhibitors of norovirus protease: structural, biochemical, spectroscopic, and antiviral studies, Eur. J. Med. Chem. 119 (2016) 300–318.
[29] G. Meng, N. Spahich, R. Kenjale, G. Waksman, J.W.St Geme III, Crystal structure of the Haemophilus influenzae Hap adhesin reveals an intercellular oligomerization mechanism for bacterial aggregation, EMBO J. 30 (18) (2011) 3864–3874.
[30] T.A. Johnson, J. Qiu, A.G. Plaut, T. Holyoak, Active-site gating regulates substrate selectivity in a chymotrypsin-like serine protease the structure of haemophilus influenzae immunoglobulin A1 protease, J. Mol. Biol. 389 (3) (2009) 559–574.
[31] F. Ruiz-Perez, J.P. Nataro, Bacterial serine proteases secreted by the autotransporter pathway: classification, specificity, and role in virulence, Cell. Mol. Life Sci. 71 (5) (2014) 745–770.
[32] J.F. Petersen, M.M. Cherney, H.D. Liebig, T. Skern, E. Kuechler, M.N. James, The structure of the 2A proteinase from a common cold virus: a proteinase responsible for the shut-off of host-cell protein synthesis, EMBO J. 18 (20) (1999) 5463–5475.
[33] Y. Sun, X. Wang, S. Yuan, M. Dang, X. Li, X.C. Zhang, Z. Rao, An open conformation determined by a structural switch for 2A protease from coxsackievirus A16, Protein Cell 4 (10) (2013) 782–792.
[34] L. Holm, C. Sander, Dali: a network tool for protein structure comparison, Trends Biochem. Sci. 20 (11) (1995) 478–480.
[35] B.C. Nick, M.C. Pandya, X. Lu, M.E. Franke, S.M. Callahan, E.F. Hasik, S. T. Berthrong, M.R. Denison, C.C. Stobart, Identification of a Critical Horseshoe-shaped Region in the nsp5 (Mpro, 3CLpro) Protease Interdomain Loop (IDL) of Coronavirus Mouse Hepatitis Virus (MHV), bioRxiv, 2020, 2020.06.18.160671.
[36] S.T. Ngo, N.M. Tam, M.Q. Pham, T.H. Nguyen, Benchmark of popular free energy approaches revealing the inhibitors binding to SARS-CoV-2 mpro, J. Chem. Inf. Model. 61 (5) (2021) 2302–2312.
[37] A. Gahlawat, N. Kumar, R. Kumar, H. Sandhu, I.P. Singh, S. Singh, A. Sjostedt, P. Garg, Structure-based virtual screening to discover potential Lead molecules for the SARS-CoV-2 Main protease, J. Chem. Inf. Model. 60 (12) (2020) 5781–5793.
[38] P.J. Kraulis, MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures, J. Appl. Crystallogr. 24 (1991) 946–950.