ISSN 0006-2979, Biochemistry (Moscow), 2008, Vol. 73, No. 1, pp. 1-13. © Pleiades Publishing, Ltd., 2008.
Original Russian Text © G. N. Rudenskaya, D. V. Pupov, 2008, published in Biokhimiya, 2008, Vol. 73, No. 1, pp. 3-17.

REVIEW

Cysteine Proteinases of Microorganisms and Viruses

G. N. Rudenskaya[1]* and D. V. Pupov[2]

[1]Faculty of Chemistry and [2]Faculty of Biology, Lomonosov Moscow State University, 119991 Moscow, Russia; fax: (495) 939-3181; E-mail: laboratoriahps@hotmail.com

Received May 7, 2007
Revision received July 18, 2007

Abstract—This review considers properties of secreted cysteine proteinases of protozoa, bacteria, and viruses and presents information on the contemporary taxonomy of cysteine proteinases. Literature data on the structure and physicochemical and enzymatic properties of these enzymes are reviewed. High interest in cysteine proteinases is explained by the discovery of these enzymes mostly in pathogenic organisms. The role of the proteinases in pathogenesis of several severe diseases of human and animals is discussed.

DOI: 10.1134/S000629790801001X

Key words: cysteine proteinases, properties, protozoa, bacteria, viruses

Classification and Catalytic Mechanism of Cysteine Proteinases

Cysteine proteinases are peptidyl hydrolases in which the role of the nucleophilic group of the active site is performed by the sulfhydryl group of a cysteine residue. Cysteine proteinases were first discovered and investigated in tropic plants. They have now been found in all representatives of the animal world.

The catalytic mechanism of cysteine proteinases is similar to that of serine peptidases, which implies the participation of a nucleophilic group and a proton donor. The role of the proton donor in all cysteine peptidases, as in most serine peptidases, is played by a histidine residue. Some cysteine proteinases need a third catalytic residue that serves for the correct spatial orientation of the imidazole ring of the histidine residue.

The temporary classification of proteolytic enzymes MEROPS (http://merops.sanger.ac.uk) based on the catalytic mechanism of proteinases and their primary and spatial structure includes nine main clans of cysteine proteinases.

The first discovered cysteine proteinase, papain, belongs to clan CA. Analysis of the crystal structures of papain and related peptidases showed that the catalytic residues are arranged in the following order in the polypeptide chain: Cys, His, and Asn. Also, a glutamine residue preceding the catalytic cysteine is also important for catalysis. This residue is probably involved in the formation of the oxyanion cavity of the enzyme. The catalytic cysteine residue is usually followed by a residue of an aromatic hydrophobic amino acid, but sometimes this position is occupied by a glycine residue.

The representatives of clan CB contain the catalytic dyad His-Cys. Comparative analysis of tertiary structures of typical representatives of clan CB, proteinases of picornavirus 3C and hepatitis A virus, with the structure of chymotrypsin reveals a significant similarity in their spatial structures. These data suggest that these representatives of the cysteine and serine proteinases may originate from a common ancestor.

In clan CC, the cysteine catalytic residue usually follows a hydrophobic residue, while in clan CB a glycine residue follows the active site cysteine. The catalytic mechanism of the clan CC proteinases is similar to that of papain, which involves the participation of four amino acid residues, but in the polypeptide chain of clan CC proteinases these residues are arranged in a different order: His, Glu, Gln (or Asp), Cys. Endopeptidases of clan CC usually hydrolyze a single bond during the processing of the virus polyprotein, while the proteinases of clan CB are capable of cleaving several bonds.

The enzymes of clan CD contain the catalytic dyad His-Cys. Typical representatives of this clan are the

Abbreviations: DTT) dithiothreitol; FMDV) foot-and-mouth disease virus; RDRP) RNA-dependent RNA polymerase; TLCK) tosyl-L-lysine chloromethyl ketone; TPCK) tosyl-L-phenylalanine chloromethyl ketone.

* To whom correspondence should be addressed.

enzymes gingipains K and R, clostripain, caspases, and legumains. All enzymes of this clan are highly selective. Their specificity is determined by the amino acid residue in the P1 position, which is uncharacteristic for papain-like cysteine proteinases. Proteinases of clan CD, unlike other cysteine proteinases, are not inhibited (or are inhibited reversibly) by the inhibitor E-64. Moreover, they all have a similar structure of the active site, and gingipain R and caspase-1 exhibit similar spatial packing of polypeptide chains.

Currently, the following five clans of cysteine proteinases have been specified: CF, CH, CL, CM, and CN. The serine proteinases of clan CF contain in the active site the catalytic triad Glu-Cys-His. A representative of this clan is pyroglutamyl-peptidase I from the bacterium *Thermococcus litoralis*. The tertiary structure of pyroglutamyl-peptidase I differs from the structures of other known cysteine proteinases and belongs to the new clan CF. However, there is some similarity between the representatives of clan CF and clans of metalloproteinases MC, MF, and MH.

A cysteine proteinase of clan CH (hedgehog protein) was found in *Drosophila melanogaster*. The active site of this enzyme is composed of the triad Cys-Thr-His. Clan CH includes autolytic cysteine proteinases.

Clan CL includes sortase A and B from *Staphylococcus aureus*, whose active sites contain the dyad His-Cys. The spatial structures of sortases A and B were determined, and sortase A was shown to be activated by calcium ions. Clan CL includes bacterial proteinases with high transferase activity.

Clan CM includes viral endopeptidases existing as homodimers in the active state, which is of special interest. Their active site contains the triad His-Glu-Cys, but the first two residues belong to one monomer, and the cysteine residue belongs to the other monomer. The tertiary structure of proteinase NS2 from hepatitis C virus was determined; it appears to be unique for clan CM.

The active sites of cysteine proteinases of clan CN contain the dyad Cys-His. The tertiary structure for proteinase nsP2 from Venezuelan horse virus has been determined. The domain structure of clan CN cysteine proteinases is unique, but the C-terminal domain is similar to that of S-adenosine-L-methionine-dependent enzymes of the methyltransferase family.

Cysteine Proteinases of Protozoa

Various protozoans differ significantly in their nutritional metabolism. There are photosynthesizing autotrophs and microorganisms that use carbohydrate substrates instead of proteins as the source of carbon (Euglenales), so the secretion of proteolytic enzymes into the environment is not characteristic for them. However, representatives of protozoa causing numerous human and animal diseases possess a number of proteinases, including cysteine proteinases that are important factors of their pathogenicity.

Cysteine proteinases of protozoa of the genus *Plasmodium*. Malaria is one of the most severe fever-causing infectious diseases. There are one million lethal cases per year worldwide. *Plasmodium falciparum* is the most pathogenic species of parasitic protozoa that causes malaria in humans. The main proteinase secreted by this parasite is falcipain; it belongs to the papain family of cysteine proteinases (clan CA, family C1). Falcipain was first isolated from the trophozoite form of the parasite [1]. The enzyme exhibits maximal activity in an acid medium (pH 5.0-6.5), this being connected with its localization in the digestive vacuoles of trophozoites. The activity of falcipain is significantly stimulated by reducing agents such as dithiothreitol (DTT), cysteine, and glutathione. The enzyme cleaves protein substrates: azocoll, gelatin, and casein. Among synthetic substrates, falcipain hydrolyzes those containing in the P1 position Lys or Arg residues, preferring large neutral amino acid residues in the P2 position. Falcipain is inactivated by the peptide inhibitors of cysteine proteinases leupeptin, E-64, and chymostatin [2].

The translation product of the falcipain gene is a preproenzyme (569 amino acid residues, 66.8 kD) [3]. Comparison with papain and the relative proteinases suggests that the formation of the active enzyme form is provided by cleavage between Lys332 and Val333. The proteolysis yields the mature proteinase of 26.8 kD with the amino acid sequence exhibiting 37 and 33% of homology with human cathepsin L and papain, respectively.

It is supposed that the mature falcipain molecule contains four disulfide bonds. The gene encoding this protein was named the falcipain-1 gene, since another gene was found that was named the falcipain-2 gene [2, 4]. It is considered that all previously determined biochemical properties of falcipain are related to the product of the falcipain-2 gene. This gene encodes a proteinase that is a typical member of papain family containing a large prodomain. Without this prodomain, the enzyme exhibits molecular weight of 27 kD. The predicted amino acid sequences of mature falcipain-2 and falcipain-1 exhibit only 37% of homology. Falcipain-2 lacks in the characteristic signal peptide, but has a hydrophobic sequence of 20 amino acid residues that is probably a transmembrane domain. The gene of falcipain-3 has also been characterized [5]. The mature falcipain-3 exhibits 65% of homology with falcipain-2.

The most important period of the life cycle of the parasite takes place in erythrocytes. Falcipains are hemoglobinases that are can digest hemoglobin in a nutrition vacuole in an acid-reducing medium. The stability of the erythrocyte membrane is maintained by a bidimensional protein system on the internal surface of the membrane. Proteolytic cleavage of these proteins results in the lysis of

the erythrocytes. Falcipain-2 is capable of cleaving skeletal proteins of the erythrocyte membrane at neutral pH values. For example, falciapin-2 hydrolyzes ankyrin and band 4.1 protein in their C-terminal region. The activity of this enzyme allows the parasites to leave the erythrocyte.

Analogs of falcipain have been found in nine species of malarial parasites. The primary structures of the mature enzyme are rather conservative. For example, falcipain-1 exhibits 71 and 56% homology with its analogs from *P. vivax* and *P. vinckei*, respectively. All these proteinases have a non-conservative region of 30 amino acids (499-529 in falcipain), which is absent in most enzymes of the papain family. All cysteine proteinases of malarial parasites have unusually large sequences.

Cysteine proteinases of *P. vivax* (vivapain-2 and vivapain-3) have also been isolated and characterized [6]. Vivapains, like falcipains, exhibit maximal activity in acid media. The character of inhibition of vivapains is typical for proteinases of the papain family.

Cysteine proteinases of the ameba *Entamoeba histolytica*. The protozoan *E. histolytica* is a causative agent of amebiasis. Infection with this microorganism results in hemorrhagic colitis and abscess of the liver, sometimes being lethal. Based on biochemical, immunological, and genetic data, *E. histolytica* has been subdivided into two species that are identical morphologically but different genetically: the pathogenic *E. histolytica* and nonpathogenic *E. dispar*. It is important to note that the most significant difference between these two species consists in the expression level of cysteine proteinases. *Entamoeba dispar* exhibits a much lower activity of cysteine proteinases compared with *E. histolytica*. Such a difference is probably connected with the number of functioning genes encoding cysteine proteinases [7]. Cysteine proteinases with molecular weights in the range of 16-96 kD have been found in different extracts of *Entamoeba*. The first isolated and characterized proteinases were amebapain [8] and histolysin [9]. Later, two other cysteine proteinases of 27 kD differing in affinity to laminin were described: amebapains EhCP1 and EhCP2 [10].

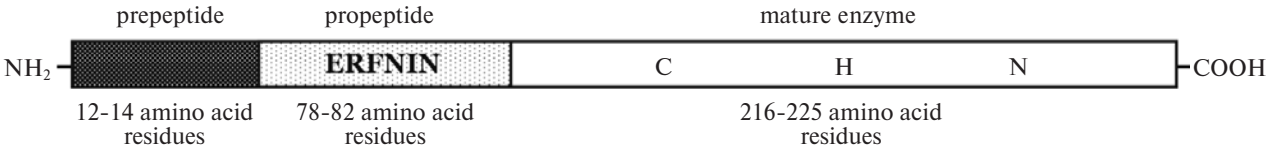
The substrate specificity of cysteine proteinases of *E. histolytica* is typical for the cathepsin B-like enzymes of the papain family [11]. Most of these cysteine proteinases exhibit the highest activity towards the substrates containing arginine residues in the P1 and P2 positions. The proteinases are activated in the presence of reducing agents (DTT, cysteine, and β-mercaptoethanol) and inactivated in the presence of inhibitors of cysteine proteinases (iodoacetate, iodoacetamide, N-ethylmaleimide, E-64), as well as tosyl-L-lysine chloromethyl ketone (TLCK), leupeptin, and α_1-antitrypsin [12, 13].

At least seven genes of *E. histolytica* encoding different preproforms of cysteine proteinases have been identified and sequenced. About 90% of the total activity of cysteine proteinases in *E. histolytica* is accounted for by EhCP1, EhCP2, and EhCP5 [14]. The product of the gene EhCP5 is considered as the most important cysteine proteinase of *E. histolytica* and, consequently, the main factor of pathogenicity of this microorganism. It was supposed that the nonpathogenic species *E. dispar* lacks this gene. However, it was demonstrated that the gene is present but inactive [15]. The amino acid sequences of the products of seven genes of cysteine proteinases of *E. histolytica* are proteins exhibiting 34-39 and 36-46% homology with the amino acid sequences of papain and cathepsin L, respectively, being absolutely conservative in the positions of the amino acid residues that are essential for the functioning of cysteine proteinases.

Cysteine proteinases of *E. histolytica* are synthesized as the inactive precursors containing a hydrophobic signal peptide of 12-14 amino acid residues, a prodomain of 78-82 amino acid residues, a catalytic domain of 216-225 amino acid residues, and a C-terminal region [16]. A surface cysteine proteinase EhCP12 containing a transmembrane region and an RGD domain for the binding to integrin, which can indicate the participation of the enzyme in the adhesion of the parasite cells to the host cells, was also found (Scheme 1).

Two main reasons for tissue damage during infection by *E. histolytica* have been determined: lysis of cells by proteins that are capable of forming pores and proteolytic activity of cysteine proteinases leading to the destruction of cells and extracellular matrix. Cysteine proteinases can adhere to host cells and destroy matrix proteins such as collagens, laminin, and fibronectin, but not elastin [12]. Thus, cysteine proteinases provide not only proteolytic cleavage of the extracellular matrix proteins, but they also take part in the adhesion to host cells and their subsequent lysis.

Cysteine proteinases of trypanosomatids. *Trypanosoma cruzi* is a causative agent of American trypanosomi-



Polypeptide chain of cysteine proteinase of *E. histolytica*. The scheme shows the cathepsin-L-like motif ERFNIN in the prodomain and the catalytic cysteine, histidine, and asparagine residues in the region corresponding to the mature enzyme

Scheme 1

asis (Chagas' disease). Humans are infected when bitten by bloodsucking insects, which are the main carriers of the disease. The main cysteine proteinase of the parasite is cruzipain belonging to the papain family (clan CA, family C1).

Cruzipain is an endopeptidase that cleaves casein, bovine albumin, and hemoglobin (optimal range of pH is 3-5) and different synthetic chromogenic and fluorogenic substrates (at pH 7-9). Cruzipain binds and hydrolyzes peptide substrates containing both arginine and phenylalanine in the P2 position, the efficiency of these processes differing 15-fold at pH 6.0 [17]. The specificity of cruzipain is higher at low pH values, this being connected with protonation of the amino acid residues of the pocket that is responsible for the substrate specificity of the enzyme [18].

The enzyme is inhibited by organomercuric reagents, E-64, TLCK, leupeptin, and some peptidyl derivatives of chloro- and fluoromethane. Cystatins, stefins, and kininogens are also strong inhibitors of the proteinase [17, 19]. The enzyme is stable within the range of pH 4.5-9.5 [20].

Cruzipain is encoded by several genes encoding a signal peptide and a proprotein. The mature enzyme (molecular weight 40 kD including two mannose polysaccharide chains) includes a catalytic part (215 amino acid residues) that is homologous to cathepsin S and a C-terminal domain (130 amino acid residues) that is unique for the cysteine proteinases of trypanosomatids. The C-terminal domain contains a core (76 amino acid residues) stabilized by disulfide bridges, an N-terminal segment (27 amino acid residues), and a highly hydrophilic C-terminal tail (27 amino acid residues). The N-terminal segment contains seven proline residues and presumably acts as a hinge between the catalytic and C-terminal domains. It should be noted that the C-terminal domain is the main immunogenic part of the cruzipain molecule. So, after infection the host antibodies bind to the C-terminal domain of the proteinase, but the catalytic part of the cruzipain in the resulting complex is not inhibited. Consequently, the resulting immune complex retains enzymatic activity [21].

The spatial structure of recombinant cruzipain in a complex with the synthetic inhibitor Z-Phe-Ala-CH_2F has been determined [22]. The catalytic part of the enzyme consists of one polypeptide chain that forms two domains. The first domain exhibits predominantly α-structure (L domain), and the second has antiparallel β-structure (R domain). The catalytic triad Cys25-His159-Asn175 is located in the active site situated in the cleft between the two domains. The general model of the structure and the arrangement of the amino acid residues in the active site are analogous to those of papain.

The expression of cruzipain is different in the four main stages of the life cycle of the parasite: its activity is 10-100-fold higher in the epimastigotes than in other forms. The enzyme is localized in lysosomes, but presumably it can be released to the surface of the cell. The role of cruzipain in the nutrition of the parasite is obvious. However, it also plays an important role in the penetration of the trypomastigotes into the mammalian cell [23, 24].

Leishmania is a protozoan that is a causative agent of different internal and skin diseases of mammals in the tropics and subtropics. The intermediate hosts of the parasite and its carriers are blood-sucking Diptera. Investigations of *L. mexicana* showed that it possesses several isoforms of cysteine proteinases exhibiting similar biochemical properties [25, 26]. Type I proteinases (lmcpb) are cathepsin L-like proteinases characterized by an unusual C-terminal domain. Type II proteinases (lmcpa) are also cathepsin L-like proteinases, but their biochemical properties are poorly characterized. Type III proteinases are cathepsin B-like enzymes. Type I proteinases includes seven groups of enzymes (A-H). Some of them exhibit differences in substrate specificity, particularly selectivity in P1 position [27].

For *L. mexicana*, it was shown that a high content of cysteine proteinases is characteristic for the amastigote form of the parasite that is located in the phagolysosome compartment of the mammalian macrophages. Thus, it can be supposed that cysteine proteinases are important for the survival of the parasite in the host organism.

Cysteine endopeptidases of trichomonads. Representatives of the protozoa *Trichomonas*, *Tritrichomonas*, and *Giardia* infest the urogenital and gastrointestinal tracts of human and other mammals. Cysteine proteinases of trichomonads do not have trivial names and are usually identified by their molecular weight determined by SDS-PAGE. The proteolytic activity of cysteine proteinases of these parasites is one of the highest among the cysteine proteinases of the studied groups of protozoa.

At least four genes of cysteine endopeptidases were found in the genome of *Trichomonas vaginalis* [28], and nine genes encoding these enzymes were revealed in the genome of *Tritrichomonas foetus* (a parasite of cattle) [29]. The molecular weights of the proteinases encoded by these genes are in the range of 18-34 kD. Proteinases of *Giardia* are less well investigated.

Proteinases of trichomonads can hydrolyze numerous proteins including possible physiological substrates such as hemoglobin, immunoglobulins [30], and proteins of extracellular matrix [31].

Undoubtedly, cysteine proteinases play an important role in the interactions of *T. vaginalis* with the host cell and contribute to the total pathogenicity of the parasite. It was shown that the activity of 43-kD cysteine proteinase located on the cell surface is necessary for the adhesion of the parasite to the host cells [32]. Due to their ability to cleave immunoglobulins [33], cysteine proteinases of trichomonads can be involved in the suppression of the immune response of the host organism.

Cysteine Proteinases of Bacteria

Not all bacteria are capable of releasing cysteine proteinases into the environment. For example, representatives of the genus *Bacillus* do not possess such proteinases, having only cysteine exopeptidases. In the genomes of the most studied representatives of *Bacillus* used in the microbiological industry (*B. subtilis*, *B. cereus*, and *B. thuringiensis*) the genes of cysteine proteinases have not been revealed.

Cysteine proteinases of *S. aureus*—staphylopain and staphopains. Staphylococci are causative agents of numerous diseases that are accompanied by extensive damage to connective tissues. Staphylopain is a cysteine proteinase, one of three proteinases secreted by staphylococcus. Staphylopain is a 13-kD (by results of SDS-PAGE) basic protein of 175 amino acid residues with p*I* of 9.4 [34]. This proteinase is characterized by a wide substrate specificity that is independent of the nature of the amino acid residues adjacent to the cleaved peptide bond. Reducing agents are required to maintain the enzyme in the active state. The maximal activity determined by the hydrolysis of elastin was observed at pH 6.5.

Staphylopain is inhibited irreversibly by E-64, heavy metal ions, and oxidants. The proteolytic activity is also inhibited by T-kininogen and α_2-macroglobulin, but not by human cystatin C and kininogen.

Two other cysteine proteinases, staphopains A and B secreted by *S. aureus*, exhibit 47% homology in their amino acid sequences [35]. The spatial structure of staphopain A has been determined. Based on this structure, it can be concluded that staphopains are long-distance members of the papain family, the closest structure being observed in cathepsin B. The spatial structure of staphopain B [36] is similar to that of staphopain A.

Like most cysteine proteinases of the papain family, the staphopain molecule consists of two domains (L- and R-domains), the active site being located in the cleft between these domains. The L-domain of staphopains is more compact than in most other papain-like proteinases. The R-domain is represented by an antiparallel pseudobarrel that differs significantly from the barrels of other cysteine proteinases including cathepsin B. It should be noted that the binding of the protein inhibitor (staphostatin) by the staphopain molecule is similar to the binding of its protein substrate. A glycine residue is located in the P1 position of the inhibitor, and the replacement of this residue transforms the inhibitor into a substrate.

The connection between the pathogenicity of *S. aureus* and the proteolytic activity of the excreted proteinases has not yet been determined. However, considering that staphylopain possesses elastinolytic activity, it may be assumed that that the enzyme takes part in the destruction of the host tissues during staphylococcal infection.

Cysteine proteinase of *Clostridium histolyticum*—clostripain. The anaerobic bacterium *C. histolyticum* is a causative agent of myonecrosis (gaseous gangrene) complicated by profuse necrosis of tissues. Clostripain is the main proteinase of *C. histolyticum* that belongs to clan CD and forms its own family C11. The family includes the enzymes hydrolyzing selectively the peptide bonds after arginine residues depending on the presence of calcium ions [37]. Analysis of the specificity of clostripain in experiments on hydrolysis of parvalbumin containing one Arg residue and 12 Lys residues demonstrated that the polypeptide chain is cleaved only after the Arg residue. The enzyme can also hydrolyze the bonds formed by the carboxyl group of Lys, but with much lower rate [38]. Clostripain is able to hydrolyze the peptide bonds with proline in the P1′ position [39].

Clostripain is a heterodimeric protein composed of heavy and light chains of 43 and 15.4 kD, respectively [38]. The N-terminal amino acid residues of the heavy and light chains are alanine and asparagine, respectively. The chains are connected by noncovalent interactions. The two polypeptide chains are encoded in a single gene. The translation product of the gene contains a signal peptide (27 amino acid residues), a propeptide (23 amino acid residues), the light chain (131 amino acid residues), a linker peptide (nine amino acid residues), and the heavy chain (336 amino acid residues) [40]. Clostripain is synthesized as the inactive preproenzyme. In the presence of calcium ions, the core protein (residues 51-526) catalyzes the removal of the linker nonapeptide (residues 182-190) independently of its amino acid composition [41]. The catalytic residues are Cys231 and His176, and the substrate specificity for the P1 position is determined by the presence of Asp229.

Clostripain is inhibited by oxidants, thiol-blocking agents, EDTA, Co^{2+}, Cu^{2+}, Cd^{2+}, and heavy metal ions. Citrate, borate, and Tris also partially inhibit the protein [42]. The strongest protein inhibitors are leupeptin and antipain (both contain Arg in the P1 position). Protein inhibitors of some serine proteinases containing Lys (soybean Kunitz inhibitor, aprotinin, lima bean trypsin inhibitor) and Arg (soybean Bowman–Birk inhibitor) residues in the P1 position are also strong inhibitors of clostripain. In the case of the enzyme inactivation by these inhibitors, no preference was observed towards Lys or Arg. In spite of the fact that the character of inhibition of clostripain is similar to that of trypsin, the binding of inhibitors by clostripain is not analogous to their binding by trypsin [43]. It is of importance that histatin 5, a low molecular weight protein (24 amino acid residues) from the family of the histidine-rich saliva proteins, is a very strong inhibitor of clostripain. Histatin 5 is not an analog of the substrate, although it competes with the substrate for the active site of the enzyme [44]. Nothing is known concerning the biological function of clostripain.

Cysteine proteinases of *Porphyromonas gingivalis*. *Porphyromonas gingivalis* is a gram-negative anaerobic bacterium that is considered as the main cause of the

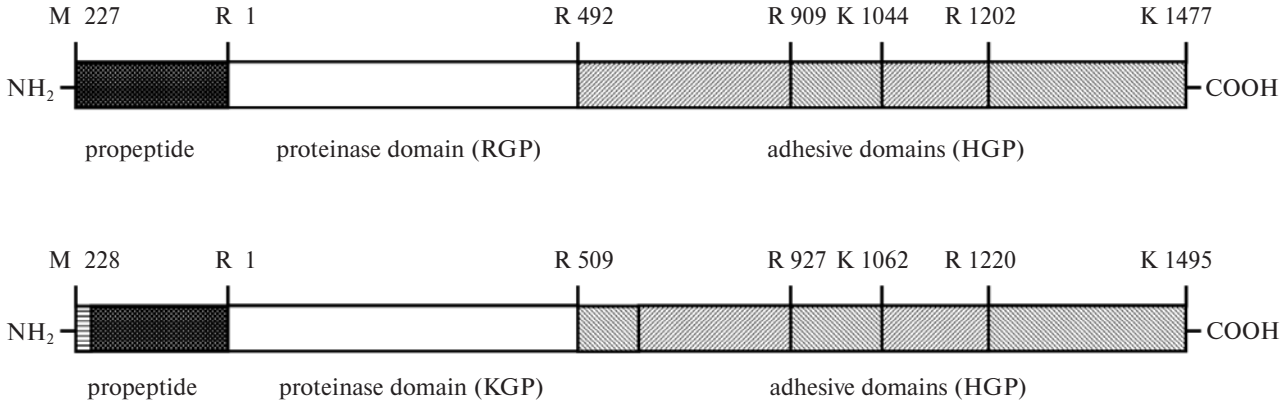
development of adult parodontitis. *Porphyromonas gingivalis* secrets numerous proteinases including aminopeptidases, carboxypeptidases, di- and tripeptidylpeptidases, oligopeptidases, and several endopeptidases [45]. All secreted proteinases (gingipains, periodontain, Prt-protease, and Tpr-protease) belong to the class of cysteine peptidases and play the main role in the pathogenesis during parodontitis. Gingipains R and K are specific towards Arg-Xaa and Lys-Xaa bonds, respectively, and constitute 85% of the total proteolytic activity of the pathogen [46].

Gingipain R is specific towards substrates containing Arg in the P1 position [47]. It has been determined that the arginine-specific activity is due to two proteinases encoded by two genes, *rgpA* and *rgpB*. The first proteinase is encoded by the single gene *rgpA*, but was isolated in two forms: a single chain of 50-kD (RGP-1) and a high molecular weight protein of 95-kD (HRGP). The latter form consists of the catalytic domain (50 kD) fused with the hemagglutinin adhesive domain [48]. The second arginine-specific proteinase (RGP-2) is encoded by the gene *rgpB* and isolated as four isoforms with different isoelectric points [49]. All isoforms have identical N-terminal sequences and similar molecular weight (48 kD), which indicates that the differences must be connected with modifications of the side chains of amino acid sequences. This does not affect the enzymatic properties of gingipains, since the isoforms do not differ in their stability, pH optima, and kinetic characteristics.

Gingipain R is inhibited by chelating agents (EDTA and EGTA), zinc ions, iodoacetic acid, iodoacetamide, N-ethylmaleimide, E-64, leupeptin, and antipain. Cystatin C does not inhibit gingipain R. TLCK and tosyl-L-phenylalanine chloromethyl ketone (TPCK) inhibit gingipain due to their ability to modify cysteine residues. It should be noted that SDS stimulates the gelatinolytic activity of this enzyme. In the absence of reducing agents, the proteolytic and amidolytic activities of purified RGP-2 and HRGP constitute 1% of their activity in the presence of 10 mM cysteine, which is the most efficient activator of gingipains. The amidolytic activity also depends on the presence of glycylglycine, which is used to reveal *P. gingivalis* in parodontitis and to differentiate between RGP-1 and RGP-2 that are stimulated by glycylglycine in different ways.

Gingipain K also exists in several forms encoded by the single gene *kgp* [50]. The product of its expression is a proteinase-adhesive polypeptide of 186.8 kD that is subjected to posttranslational modification on arginine residues, yielding numerous forms of gingipain K (Scheme 2). The activity of gingipain K towards synthetic substrates is maximal at pH 8.0. The characters of inhibition of gingipains R and K are similar, but there are some differences—gingipain K is not inhibited by zinc ions, E-64, leupeptin, antipain, and EDTA, and is not activated by glycylglycine. Similarly to gingipain R, cysteine is the most efficient activator of gingipain K. Gingipain K exhibits specificity towards the amino acid residue in the P1 position and does not hydrolyze the bonds with lysine and arginine in the P2 position.

The crystal structure of the complex of gingipain R (RgpB) with the inhibitor D-Phe-Phe-Arg-chloromethyl ketone has been determined [51]. The enzyme consists of two domains: the catalytic domain (351 amino acid residues) and the domain of the immunoglobulin family IgSF (84 amino acid residues). The catalytic domain is subdivided into subdomains A and B. The surface between these subdomains is relatively hydrophilic and contains a cluster of charged amino acid residues with coordinated calcium ion providing the correct spatial orientation of the subdomains. The B-subdomain contains the catalytic dyad Cys244-His211. The tertiary structure of the catalytic domain of gingipain R is similar to that of



Polypeptide chain of gingipains RGP and KGP: the prodomain is shown dark, the proteinase domain is clear, adhesive domains are shaded. Some amino acid residues involved in the proteolytic cleavage of the peptide are indicated

Scheme 2

caspase-1 and caspase-3. Also, caspases and gingipain contain identically situated histidine and cysteine residues, equivalent oxyanion cavities, and S_1-pockets of similar shape and location that determine the substrate specificity of the enzymes.

Gingipains R are capable of activating prothrombin [52]. Gingipains R (HRGP and RGP-2) cannot cleave hemoglobin, while gingipain K cleaves this protein yielding different low molecular weight fragments. Gingipain K is capable of hydrolyzing transferrin, while gingipain R hydrolyzes it very slowly. Gingipain R can hydrolyze the C3 component of the complement yielding C3b- and C3a-like fragments, which are significant mediators of inflammation [53].

The other cysteine proteinase isolated from *P. gingivalis* is peridontain [54]. The gene of peridontain encodes a protein of 93 kD (843 amino acid residues). The molecular weight of peridontain determined by gel filtration constitutes 75 kD, while SDS-PAGE yields two bands corresponding to 55 and 20 kD. Thus, the mature enzyme is a heterodimer with the active site in the large subunit. The p*I* value of peridontain is 5.3. The enzyme does not hydrolyze azocasein, casein, lysozyme, collagen, fibrin, plasminogen, and fibrinogen. However, when lysozyme was reduced and carboxymethylated, complete cleavage of the protein was observed after less than 10 min. Thus, it can be concluded that peridontain hydrolyses denatured proteins or easily accessible polypeptide chains, but cannot hydrolyze proteins with pronounced secondary and tertiary structures. Peridontain is active within the range of pH 6.0-9.0, exhibiting maximal activity at pH 7.5-8.0. The enzyme is absolutely inactive in the presence of reducing agents such as cysteine, β-mercaptoethanol, and DTT. Ca^{2+} does not stabilize the molecule of peridontain, and glycylglycine does not activate the enzyme. Peridontain is inhibited by leupeptin, TLCK, iodoacetamide, and E-64. These results suggest that peridontain is closer to the papain family than gingipains.

Porphyromonas gingivalis is asaccharolytic organism obtaining carbon and energy mostly by cleaving proteins. Evidently, gingipains possessing a high specificity are capable of cleaving large proteins yielding peptides, while several proteinases of wide specificity including peridontain are required for the complete cleavage of these peptides.

Cysteine proteinase of *Streptococcus pyogenes*—streptopain. *Streptococcus pyogenes* is one of the widespread bacterial pathogens of humans. Diseases caused by streptococcal infection vary from relatively mild, such as pharyngitis, to severe diseases that are dangerous for human life. Cysteine proteinase of *S. pyogenes* (SpeB) is the first cysteine proteinase isolated from prokaryotes. Streptopain, also known as exotoxin B, is present in all strains of *S. pyogenes* and constitutes 95% of all secreted proteins. Streptopain hydrolyzes the bonds in the peptides containing hydrophobic residues in the P1 position [55]. The proteinase is inhibited by iodoacetate, iodoacetamide, N-ethylmaleimide, *p*-chloromercuribenzoate, heavy metal ions (Cu^{2+}, Hg^{2+} and Ag^{+}), as well as atmospheric oxygen [56].

The spatial structure of the streptopain proenzyme has been determined [57]. The characteristic features of packing of the polypeptide chain suggest that streptopain belongs to the papain family. The molecule of this cysteine proteinase consists of two typical domains. The first domain is formed by α-helixes, and the second one by antiparallel β-structures. The enzyme molecule contains the dyad Cys-His, but not the triad Cys-His-Asn. The profragment of the immature enzyme shifts the catalytic His195 from the active site and prevents its interaction with Cys47, thus maintaining the proteinase in the inactive state.

Streptopain is one of the most important factors of pathogenicity of streptococci [58]. SpeB contains the motif Arg-Gly-Asp (RGD) that is required for the binding to integrins of the host cells, allowing the microorganism to adhere to the host cell surface. Streptopain is capable of hydrolyzing fibronectin and vitronectin, proteins of extracellular matrix, but does not cleave laminin [59]. SpeB hydrolyzes the hinge region of IgG and the heavy chains of immunoglobulins of all types [60]. Streptopain activates the 66-kD matrix metalloproteinase cleaving type IV collagen, and also activates some proteins of *S. pyogenes*: M1 protein, C5a peptidase, and IgG-binding protein. SpeB cleaves the inactive precursor of human interleukin 1β yielding biologically active interleukin 1β [61]. Thus, streptopain performs numerous functions and is an important factor of pathogenicity of *S. pyogenes*.

Cysteine Proteinases of Viruses

Viruses reproduce and function using the genetic apparatus of the host cell. The extreme extent of parasitism has led to the significant simplification of the virus structure, reducing the number of proteins required for their existence compared to other organisms. For example, the grippe virus does not produce proteolytic enzymes, and its genome lacks in the structures corresponding to proteolytic enzymes. Nevertheless, some virus families are capable of producing proteinases including cysteine proteinases that are important factors of their pathogenicity.

Proteinase L of foot-and-mouth disease virus. Foot-and-mouth disease virus (FMDV) belongs to the genus of autoviruses from the picornavirus family. The genome of picornaviruses is represented by a plus-strand RNA. After penetrating into the host cell, the RNA is translated yielding a long polyprotein. Then the polyprotein is processed yielding separate viral proteins. Viruses of two genera of this family (auto- and cardioviruses), as well as

some viruses of indefinite systematic position, for example *Equine rhinovirus* (ERV), encode in their genome a special leader (L) protein [62].

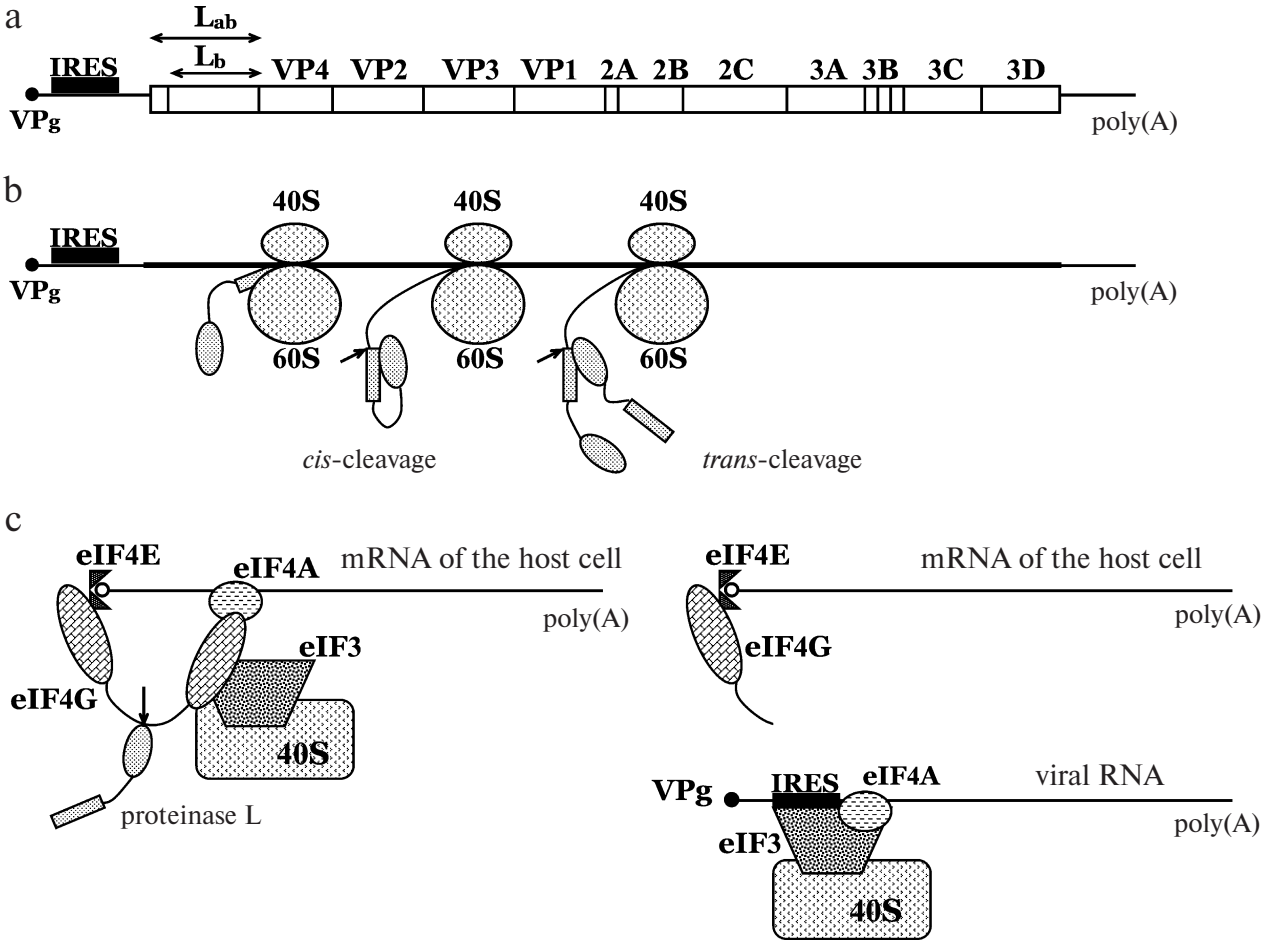
Protein L was named because of its position in the polyprotein and due to the early emergence in the life cycle of the virus: it corresponds to the 5′-region that precedes the region of structural genes encoding the virus capsid proteins (Scheme 3a).

Protein L of FMDV is a proteolytic enzyme that belongs to clan CA of the papain-like cysteine proteinases [63]. FMDV is characterized by the presence of two start codons AUG within the single reading frame, so two different L proteins can be formed, these being designated as La_b and L_b. Protein L_b, the smaller of these two proteins, is the main synthesized proteinase L.

Peptidase L_b is a protein of 21 kD (173 amino acid residues) exhibiting p*I* of 4.8. X-Ray analysis revealed in the active site of the enzyme two amino acid residues (Cys51 and His148) that are essential for its functioning [64].

Protein L is cleaved autocatalytically from the translated viral polyprotein (Scheme 3b). Protein L is also involved in the cleavage of translation initiation factor eIF-4G, a cell protein also known as p220 or eIF-4g that is required for the initiation of cap-dependent protein synthesis [65]. The L_b form of peptidase L cleaves two different bonds in the eIF-4G molecule: first, Gly479–Arg480, and then Lys318–Arg319.

The main function of peptidase L_b in cells infected by the virus is the cleavage of eIF-4G protein that is



Genome of FMDV and the main functions of the encoded proteinase L. a) Genome plus-strand RNA of FMDV. Names of the genes are designated. The 5′-terminus of the genome RNA contains the covalently bound protein VPg, and the 3′-terminus is polyadenylated. The IRES element responsible for the cap-independent translation of the virus RNA is located in the non-translated region at the 5′-terminus. b) Translation of mRNA of FMDV: autocatalytic cleavage of proteinase L can be catalyzed by the intramolecular (*cis*) or intermolecular (*trans*) mechanisms. The ribosomal subunits (40S and 60S) are designated as ovals. Proteinase L is shown by dashed oval and rectangle; the site of the cleavage is shown by the arrow. c) Role of eIF-4G factor in the cap-dependent initiation of protein synthesis and its proteolytic cleavage by proteinase L. The cell mRNA cannot be involved in the formation of the initiation complex during the translation. However, the viral RNA can interact directly with eIF3 without cap-binding eIF4E and eIF4G proteins yielding the initiation complex

Scheme 3

required for the efficient translation of the cap-dependent mRNA of the host cell (Scheme 3c). The cleavage of this protein ceases the synthesis of most host proteins. However, the genome mRNA of the virus is translated by the cap-independent mechanism and is not suppressed after the cleavage of eIF-4G. The changed variant of FMDV with a deletion in the region encoding peptidase L_b [66] is less productive compared to the wild type, suggesting that peptidase L_b is necessary for the normal production of the virus and for the development of the dramatic cytopathic effects connected with FMDV infection [67].

Papain-like endopeptidases of coronaviruses. Coronaviruses comprise a family of viruses containing plus-strand genome RNA of large size (28-32 kb) encoding their own RNA-dependent RNA polymerase (RDRP). During the translation of the viral genome, RDRP is translated as a protein precursor, a polyprotein that must be subjected to the proteolytic cleavage for the correct replication of the viral genome. In the well-studied representative of coronaviruses mouse hepatitis virus (MHV), RDRP encoded inside the genome region (22 kb) results in the translation of a protein precursor of 750 kD [68]. The presence of different functional regions in this polyprotein was predicted. There are three regions corresponding to proteinases: two of them are papain-like proteinases, and the third is picornain-like proteinase. The amino acid sequence of the first papain-like cysteine proteinase (PCP-1) exhibits a limited similarity with the sequence of papain. The similarity mainly consists in the presence of the conservative dipeptide Cys-Trp containing the catalytic cysteine residue and in the interval between the catalytic cysteine and the histidine residue [69].

The second papain-like region (PCP-2) exhibits little similarity with papain, retaining only catalytic residues Cys and His and the interval between them. Therefore, proteinases PCP-2 of hepatitis C are now included into a separate family of clan CM.

It has been demonstrated that the PCP-1 region is necessary for the emergence of p28 protein. *In vitro*, p28 splits off from the polyprotein cotranslationally, this supporting the *cis*-mechanism of the cleavage. The proteolytic activity of proteinase PCP-1 is connected with the liberation from the polyprotein of another protein (p65) that is adjacent to p28 [70].

Commercial inhibitors of proteinase PCP-1 suppressing its activity have been found. For example, leupeptin and zinc chloride block the proteolytic activity of PCP-1 towards the cleavage of p28 *in vitro*, but none of them showed an efficient blocking of the viral protease *in vivo* due to their high toxicity. The membrane-penetrating derivative of irreversible inhibitor of cysteine proteinases E-64 (E-64d) is capable of blocking efficiently the cleavage of p65 from the polyprotein *in vivo*, exhibiting no effect on the proteolytic cleavage of protein p28 [71].

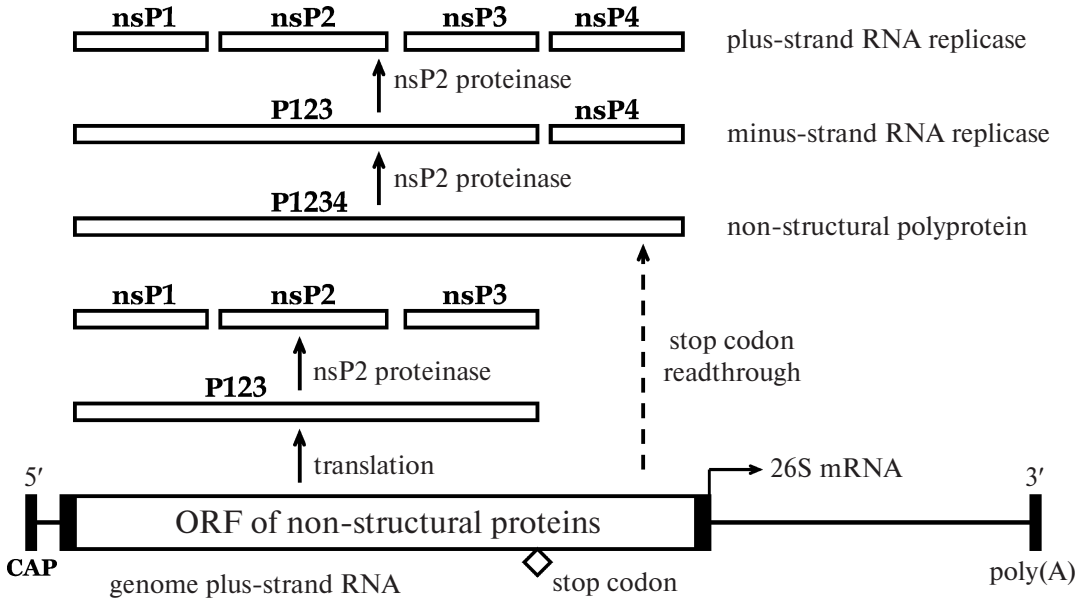
Endopeptidase PCP-1 is supposed to play an important role in the processing of the polyprotein through both *cis*- and *trans*-mechanisms yielding non-structural proteins connected with the replication of the viral genome. The existence of the regions of proteinase PCP-1 was predicted, but was not functionally demonstrated for three other representatives of coronaviruses with the completely sequenced gene of RDRP: avian infectious bronchitis virus, human coronavirus 229E, and swine transmissible gastroenteritis virus. Other members of the superfamily that includes arteriviruses and coronaviruses, as previously described, also encode regions that are similar to those of PCP-1 [72].

Papain-like endopeptidase of Sindbis virus. Sindbis virus (Togaviridae family, *Alphavirus* genus) is a virus of animals containing in the genome a single plus-strand RNA. Sindbis virus has a wide range of hosts and is widespread in different tissues, also infecting numerous cell types in culture including cells of mammals, birds, and mosquitoes. The translation of the genome RNA yields four non-structural proteins that are formed in the cytoplasm of the infected cells as two precursors: polyproteins P123 (1896 amino acid residues) and P1234 (2513 amino acid residues, formed due to readthrough of the stop codon in the 1897 position). Further, the polyprotein P1234 is processed by the viral nsP2 proteinase yielding (through a number of intermediate products) four products: nsP1, nsP2, nsP3, and nsP4 [73] (Scheme 4).

Sindbis virus proteinase nsP2 is a polypeptide of 807 amino acid residues encoded by nucleotides 1680-4100 of the viral genome. The N-terminal part of the protein is supposed to be helicase that is required for replication of RNA, and the C-terminal region between the 460 and 807 residues corresponds to the proteinase [74]. The supposed catalytic pair Cys481 and His558 is accompanied by the Trp559 residue that is required to stabilize the conformation of the active site of the enzyme [75]. Currently, the proteinase is referred to clan CN. All known alphaviruses encode nsP2 proteinases exhibiting more than 50% similarity in the amino acid sequences.

The substrates for proteinase nsP2 of Sindbis virus are only the non-structural polyproteins P123 and P1234, and the products of their processing (P12, P23, and P34). The site of the cleavage has a consensus sequence (Ala/Ile)-Gly-(Ala/Cys/Gly)(Ala/Tyr).

The nsP2 proteinase and protein precursors containing nsP2 are active proteolytic enzymes differing in their substrate specificity. For example, P1234 can cleave itself autoproteolytically by the *cis*-mechanism yielding two separate proteins, P123 and nsP4, while P123 can be processed only by the *trans*-mechanism. The cleavage of a P123 molecule by another molecule of P123 or P12 results in the formation of two proteins, nsP1 and P23, while the cleavage of P123 by P23 yields P12 and nsP3. The proteins that lost the nsP3 region during processing, namely nsP2 or P12, are incapable of cleaving the site nsP3/nsP4. Such



Organization and processing of non-structural proteins formed during translation of the genome RNA of Sindbis virus

Scheme 4

changes in the substrate specificity of proteinase nsP2 that depend on the original polyprotein allow fine regulation of RNA synthesis during the infection cycle.

During the early stages of infection, proteins P123 and nsP4 together with some proteins of the host organism form a replicase complex synthesizing the minus-strand RNA of the viral genome. Proteolytic decomposition of P123 within the replicase complex changes its specificity. As a result, the complex acquires the ability to synthesize efficiently the plus-strand RNA, but cannot take part in the replication of the minus-strand RNA [76, 77]. Since during infection the concentration of the *trans*-acting proteinases grows, the non-structural polyprotein is cleaved at the nsP2/nsP3 site during its synthesis (co-translationally) yielding the processing products nsP1, nsP2, nsP3, and nsP34. At this stage, the synthesis of the minus-strand RNA is blocked, since the processing products cannot form the replicase that is capable of synthesizing the minus-strand RNA. The use of proteinase nsP2 for regulation of genome RNA synthesis is an interesting evolutionary adaptation of viruses that have to encode in their genome an additional protein providing the correct cleavage of the polyprotein in the cytoplasm of the infected cells.

Thus, besides the processing and liberation of proteins required for the reproduction of the virus, proteinase nsP2 from Sindbis virus is important for the fine regulation of the genome RNA replication, thus providing high infectivity.

Cysteine proteinase of picornaviruses—picornain 3C. Picornaviruses are the smallest of the known RNA-containing viruses (*pico*, very little, and *rna* stands for RNA). They comprise one of the most numerous and important families of causative agents of diseases of humans and farm animals. FMDV, poliovirus, and human hepatitis A virus are picornaviruses. The main causative agents of cold-related diseases (rhinoviruses) are also representatives of this family.

The main strategy of protein synthesis used by picornaviruses is the following: the genome plus-strand RNA contains a single extended reading frame encoding a polyprotein that is cleaved during translation, so the full-length protein is not formed. The cleavage cascade results in 11 or 12 products.

The first viral protein with proteolytic activity was 3C protein of poliovirus [78]. Later, this proteolytic activity was identified in all genera of the picornavirus family, and the corresponding protein was named picornain 3C. Picornain 3C is a protein containing from 180 (human rhinoviruses) to 220 (hepatitis A virus) amino acid residues. The p*I* value of the protein is in the range of 8.0-10.0. The active site includes the residues Cys, His, and Asp. There are extended regions of homology in picornain 3C and chymotrypsin-like serine proteinase (family C1). The structure of the active site of picornain 3C and the packing of the polypeptide chain that is similar to that of chymotrypsin suggest that picornain 3C is an example of evolutionary relationship of cysteine and serine proteinases [79].

Picornain 3C, as a typical representative of cysteine proteinases, is inhibited by N-ethylmaleimide, iodoacetamide, and Hg^{2+}, but is not inhibited by E-64.

Picornain 3C cleaves sites containing Gln and Gly residues in the P1 and P1′ positions, respectively. However, these residues and their environment can differ in different genera of viruses. For the successful processing, not only the primary sequence of the site of cleavage is important, but also the character of the second structure. It was demonstrated that the Gln-Gly pairs situated between two β-barrels are hydrolyzed more efficiently than pairs situated inside ordered secondary structures [80].

Picornain 3C is the main enzyme involved in the processing of the polyprotein in picornaviruses, providing from 8 to 10 disruptions depending on the genus of the virus. Some differences between the enzymes from different genera of picornaviruses concern the substrate specificity and number of recognized sites of cleavage.

Besides the viral polyprotein, picornain 3C can cleave different proteins of the host cell. For example, proteinase 3C of poliovirus blocks the transcription of the host proteins by specific proteolysis of the transcription factors TFIIIC [81] and TFIID [82]. One of the proteins associated with microtubules (MAP-4) is also a substrate of the poliovirus picornain 3C [83]. The cleavage of this protein results in destruction of the cytoskeletal microtubules of the infected cell. The picornain of FMDV is involved in proteolysis of histone H3 [84]. Thus, picornain 3C is responsible for many severe cytopathic effects during viral infection.

Picornain of polioviruses binds specifically to hairpin structures on the 5′ ends of the genome RNA [85]. The picornain precursor 3CD protein binds to the viral RNA much more efficiently than the mature 3C protein. Presumably, such binding of the picornain precursor 3CD to the RNA provides correct spatial orientation of the viral 3D polymerase for the initiation of genome RNA synthesis. Autoproteolysis of 3CD protein directly after its binding to the genome RNA releases the polymerase 3D in the required site. The supposed RNA-binding site of 3C protein is on the opposite side of the active site side and is enriched with basic amino acid residues.

Thus, picornain 3C is one of the important factors of pathogenicity of picornaviruses that determines not only processing and maturation of the viral polyprotein, but also severe cytopathic effects such as inhibition of the transcription and breaking of the microtubular cytoskeleton. Besides, this proteinase recognizes specifically and binds to certain hairpin structures of the viral RNA in the form of the protein precursor 3CD, this providing the correct spatial orientation of the viral polymerase 3D on the molecule of genome RNA.

Picornain 2A of poliomyelitis virus 1 (PV1). The assumption that 2A protein of PV1 possesses proteolytic activity was made based on the revelation of specific picornain 3C motifs in the sequence encoding this protein. Later it was confirmed that the disruption of the sequence encoding 2A protein inhibited the proteolytic cleavage in the PV1/2A site of the viral polyprotein. The cleavage is an intramolecular process, since the 2A protein is cut out before the translation of the viral polyprotein is finished. It should be noted that 2A proteinase is also responsible for the cleavage of the eukaryotic translation initiation factor eIF-4G (also known as p220 or eIF-4g) [86]. Picornain 2A catalyzes both intramolecular (*cis*) and intermolecular (*trans*) cleavage of the viral polyprotein.

Picornain 2A is a protein with a single polypeptide chain (149 amino acid residues) exhibiting homology with the amino acid sequence of serine proteinases, this suggesting the similarity of its spatial structure with that of representatives of the bacterial α-lytic-like proteinase family [87].

Similarly to picornain 3C, picornain 2A is inhibited by N-ethylmaleimide, iodoacetamide, and Hg^{2+}, but not by E-64. The inhibitors of serine proteinases chymostatin and elastatin also inhibit efficiently the activity of picornain 2A [88].

Picornain 2A performs at least two main functions in the replication cycle of polioviruses. It catalyzes the fist stage of proteolytic processing of the viral polyprotein, cleaving the site located between the C-terminus of PV1 and its own N-terminus. Besides, picornain 2A takes part in the cleavage of 3CD protein yielding the molecules 3C′ and 3D′ [89]. Picornain 2A from enteroviruses and rhinoviruses is also responsible for the cleavage of eIF-4G, a cell protein involved in the initiation of protein syntheses.

Picornain 2A from enteroviruses and rhinoviruses differs from picornain 3C by smaller size and by substrate specificity. No structural similarity was observed between picornain 2A from enteroviruses and rhinoviruses and proteinase L of FMDV or 2A protein of encephalomyocarditis virus.

The most widespread and investigated proteinases are cysteine proteinases from plants. As seen from this review, these proteinases are also found in pathogenic protozoa and microorganisms, where these enzymes are important factors of pathogenicity. It is of interest that the closely related pathogenic and non-pathogenic species differ mostly in the expression of cysteine proteinases, as demonstrated for *E. histolytica* and *E. dispar*, and also for *T. cruzi* and *T. rangeli*.

Based on the homology of the primary and spatial structures, most cysteine proteinases of protozoa can be referred to clan CA of papain. The bacterial cysteine proteinases belong to clan CD of cysteine proteinases including caspases and legumins. The common feature of the enzymes of this clan is a high selectivity that is determined by the amino acid residue in the P1 position: arginine for clostripain, arginine and lysine for gingipains R and K, respectively, aspartate for caspases, and asparagine for legumains. Besides, they all are not inhibited by E-64, an irreversible inhibitor of most cysteine proteinases. The

reversible inhibition of clostripain and gingipain R by this reagent can be explained by the fact that this compound is an analog of arginine.

This work was supported by the Russian Foundation for Basic Research (05-04-49087).

REFERENCES

1. Rosenthal, P. J., McKerrow, J. H., Aikawa, M., Nagasawa, H., and Leech, J. H. (1988) *J. Clin. Invest.*, **82**, 1560-1566.
2. Shenai, B. R., Sijwali, P. S., Singh, A., and Rosenthal, P. J. (2000) *J. Biol. Chem.*, **275**, 29000-29010.
3. Rosenthal, P. J., and Nelson, R. G. (1992) *Mol. Biochem. Parasitol.*, **51**, 143-152.
4. Hanspal, M. (2000) *Biochim. Biophys. Acta*, **1493**, 242-245.
5. Sijwali, P. S., Shenai, B. R., Gut, J., Singh, A., and Rosenthal, P. J. (2001) *Biochem. J.*, **360**, 481-489.
6. Na, B. K., Shenai, B. R., Sijwali, P. S., Choe, Y., Pandey, K. C., Singh, A., Craik, C. S., and Rosenthal, P. J. (2004) *Biochem. J.*, **378**, 529-538.
7. Tannich, E., Scholze, H., Nickel, R., and Horstmann, R. D. (1991) *J. Biol. Chem.*, **266**, 4798-4803.
8. Scholze, H., and Schulte, W. (1988) *Biomed. Biochim. Acta*, **47**, 115-123.
9. Luaces, A. L., and Barrett, A. J. (1988) *Biochem. J.*, **250**, 903-909.
10. Li, E., Yang, W. G., Zhang, T., and Stanley, S. L., Jr. (1995) *Infect. Immun.*, **63**, 4150-4153.
11. Jacobs, T., Bruchhaus, I., Dandekar, T., Tannich, E., and Leippe, M. (1998) *Mol. Microbiol.*, **27**, 269-276.
12. Keene, W. E., Petitt, M. G., Allen, S., and McKerrow, J. H. (1986) *J. Exp. Med.*, **163**, 536-549.
13. Spinella, S., Levavasseur, E., Petek, F., and Rigothier, M. C. (1999) *Eur. J. Biochem.*, **266**, 170-180.
14. Bruchhaus, I., Jacobs, T., Leippe, M., and Tannich, E. (1996) *Mol. Microbiol.*, **22**, 255-263.
15. Willhoeft, U., Hamann, L., and Tannich, E. (1999) *Infect. Immun.*, **67**, 5925-5929.
16. Que, X., and Reed, S. L. (2000) *Clin. Microbiol. Rev.*, **13**, 196-206.
17. Serveau, C., Lalmanach, G., Juliano, M. A., Scharfstein, J., Juliano, L., and Gauthier, F. (1996) *Biochem. J.*, **313**, 951-956.
18. Gillmor, S. A., Craik, C. S., and Fletterick, R. J. (1997) *Protein Sci.*, **6**, 1603-1611.
19. Turk, B., Stoka, V., Turk, V., Johansson, G., Cazzulo, J. J., and Bjork, I. (1996) *FEBS Lett.*, **391**, 109-112.
20. Stoka, V., Nycander, M., Lenarcic, B., Labriola, C., Cazzulo, J. J., Bjork, I., and Turk, V. (1995) *FEBS Lett.*, **370**, 101-104.
21. Martinez, J., Campetella, O., Frasch, A. C., and Cazzulo, J. J. (1993) *Immunol. Lett.*, **35**, 191-196.
22. McGrath, M. E., Eakin, A. E., Engel, J. C., McKerrow, J. H., Craik, C. S., and Fletterick, R. J. (1995) *J. Mol. Biol.*, **247**, 251-259.
23. Yokoyama-Yasunaka, J. K., Pral, E. M., Oliveira, O. C., Jr., Alfieri, S. C., and Stolf, A. M. (1994) *Acta Trop.*, **57**, 307-315.
24. Cazzulo, J. J., Stoka, V., and Turk, V. (1997) *Biol. Chem.*, **378**, 1-10.
25. Robertson, C. D., and Coombs, G. H. (1990) *Mol. Biochem. Parasitol.*, **42**, 269-276.
26. Souza, A. E., Waugh, S., Coombs, G. H., and Mottram, J. C. (1992) *FEBS Lett.*, **311**, 124-127.
27. Mottram, J. C., Souza, A. E., Hutchison, J. E., Carter, R., Frame, M. J., and Coombs, G. H. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 6008-6013.
28. Mallinson, D. J., Lockwood, B. C., Coombs, G. H., and North, M. J. (1994) *Microbiology*, **140**, 2725-2735.
29. Mallinson, D. J., Livingstone, J., Appleton, K. M., Lees, S. J., Coombs, G. H., and North, M. J. (1995) *Microbiology*, **141**, 3077-3085.
30. Parenti, D. M. (1989) *J. Infect. Dis.*, **160**, 1076-1080.
31. Williams, A. G., and Coombs, G. H. (1995) *Int. J. Parasitol.*, **25**, 771-778.
32. Arroyo, R., and Alderete, J. F. (1989) *Infect. Immun.*, **57**, 2991-2997.
33. Provenzano, D., and Alderete, J. F. (1995) *Infect. Immun.*, **63**, 3388-3395.
34. Potempa, J., Dubin, A., Korzus, G., and Travis, J. (1988) *J. Biol. Chem.*, **263**, 2664-2667.
35. Rzychon, M., Sabat, A., Kosowska, K., Potempa, J., and Dubin, A. (2003) *Mol. Microbiol.*, **49**, 1051-1066.
36. Filipek, R., Rzychon, M., Oleksy, A., Gruca, M., Dubin, A., Potempa, J., and Bochtler, M. (2003) *J. Biol. Chem.*, **278**, 40959-40966.
37. Kembhavi, A. A., Buttle, D. J., Rauber, P., and Barrett, A. J. (1991) *FEBS Lett.*, **283**, 277-280.
38. Gilles, A. M., Imhoff, J. M., and Keil, B. (1979) *J. Biol. Chem.*, **254**, 1462-1468.
39. Ullmann, D., and Jakubke, H. D. (1994) *Eur. J. Biochem.*, **223**, 865-872.
40. Dargatz, H., Diefenthal, T., Witte, V., Reipen, G., and von Wettstein, D. (1993) *Mol. Gen. Genet.*, **240**, 140-145.
41. Witte, V., Wolf, N., and Dargatz, H. (1996) *Curr. Microbiol.*, **33**, 281-286.
42. Porter, W. H., Cunningham, L. W., and Mitchell, W. M. (1971) *J. Biol. Chem.*, **246**, 7675-7682.
43. Siffert, O., Emod, I., and Keil, B. (1976) *FEBS Lett.*, **66**, 114-119.
44. Gusman, H., Grogan, J., Kagan, H. M., Troxler, R. F., and Oppenheim, F. G. (2001) *FEBS Lett.*, **489**, 97-100.
45. Potempa, J., Sroka, A., Imamura, T., and Travis, J. (2003) *Curr. Protein Pept. Sci.*, **4**, 397-407.
46. Potempa, J., and Travis, J. (1996) *Acta Biochim. Pol.*, **43**, 455-465.
47. Kadowaki, T., Yoneda, M., Okamoto, K., Maeda, K., and Yamamoto, K. (1994) *J. Biol. Chem.*, **269**, 21371-21378.
48. Pike, R. N., Potempa, J., McGraw, W., Coetzer, T. H., and Travis, J. (1996) *J. Bacteriol.*, **178**, 2876-2882.
49. Potempa, J., Mikolajczyk-Pawlinska, J., Brassell, D., Nelson, D., Thogersen, I. B., Enghild, J. J., and Travis, J. (1998) *J. Biol. Chem.*, **273**, 21648-21657.
50. Pavloff, N., Pemberton, P. A., Potempa, J., Chen, W. C., Pike, R. N., Prochazka, V., Kiefer, M. C., Travis, J., and Barr, P. J. (1997) *J. Biol. Chem.*, **272**, 1595-1600.
51. Eichinger, A., Beisel, H. G., Jacob, U., Huber, R., Medrano, F. J., Banbula, A., Potempa, J., Travis, J., and Bode, W. (1999) *EMBO J.*, **18**, 5453-5462.

52. Imamura, T., Banbula, A., Pereira, P. J., Travis, J., and Potempa, J. (2001) *J. Biol. Chem.*, **276**, 18984-18991.
53. Wingrove, J. A., DiScipio, R. G., Chen, Z., Potempa, J., Travis, J., and Hugli, T. E. (1992) *J. Biol. Chem.*, **267**, 18902-18907.
54. Nelson, D., Potempa, J., Kordula, T., and Travis, J. (1999) *J. Biol. Chem.*, **274**, 12245-12251.
55. Gerwin, B. I., Stein, W. H., and Moore, S. (1966) *J. Biol. Chem.*, **241**, 3331-3339.
56. Liu, T. Y., and Elliott, S. D. (1965) *Nature*, **206**, 33-34.
57. Kagawa, T. F., Cooney, J. C., Baker, H. M., McSweeney, S., Liu, M., Gubba, S., Musser, J. M., and Baker, E. N. (2000) *Proc. Natl. Acad. Sci. USA*, **97**, 2235-2240.
58. Lukomski, S., Montgomery, C. A., Rurangirwa, J., Geske, R. S., Barrish, J. P., Adams, G. J., and Musser, J. M. (1999) *Infect. Immun.*, **67**, 1779-1788.
59. Kapur, V., Topouzis, S., Majesky, M. W., Li, L. L., Hamrick, M. R., Hamill, R. J., Patti, J. M., and Musser, J. M. (1993) *Microb. Pathog.*, **15**, 327-346.
60. Collin, M., and Olsen, A. (2001) *Infect. Immun.*, **69**, 7187-7189.
61. Kapur, V., Majesky, M. W., Li, L. L., Black, R. A., and Musser, J. M. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 7676-7680.
62. Li, F., Browning, G. F., Studdert, M. J., and Crabb, B. S. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 990-995.
63. Roberts, P. J., and Belsham, G. J. (1995) *Virology*, **213**, 140-146.
64. Guarne, A., Kirchweger, R., Verdaguer, N., Liebig, H. D., Blaas, D., Skern, T., and Fita, I. (1996) *Protein Sci.*, **5**, 1931-1933.
65. Medina, M., Domingo, E., Brangwyn, J. K., and Belsham, G. J. (1993) *Virology*, **194**, 355-359.
66. Piccone, M. E., Rieder, E., Mason, P. W., and Grubman, M. J. (1995) *J. Virol.*, **69**, 5376-5382.
67. Brown, C. C., Piccone, M. E., Mason, P. W., McKenna, T. S., and Grubman, M. J. (1996) *J. Virol.*, **70**, 5638-5641.
68. Bredenbeek, P. J., Pachuk, C. J., Noten, A. F., Charite, J., Luytjes, W., Weiss, S. R., and Spaan, W. J. (1990) *Nucleic Acids Res.*, **18**, 1825-1832.
69. Gorbalenya, A. E., Koonin, E. V., and Lai, M. M. (1991) *FEBS Lett.*, **288**, 201-205.
70. Bonilla, P. J., Hughes, S. A., Pinon, J. D., and Weiss, S. R. (1995) *Virology*, **209**, 489-497.
71. Kim, J. C., Spence, R. A., Currier, P. F., Lu, X., and Denison, M. R. (1995) *Virology*, **208**, 1-8.
72. Den Boon, J. A., Faaberg, K. S., Meulenberg, J. J., Wassenaar, A. L., Plagemann, P. G., Gorbalenya, A. E., and Snijder, E. J. (1995) *J. Virol.*, **69**, 4500-4505.
73. De Groot, R. J., Hardy, W. R., Shirako, Y., and Strauss, J. H. (1990) *EMBO J.*, **9**, 2631-2638.
74. Hardy, W. R., and Strauss, J. H. (1989) *J. Virol.*, **63**, 4653-4664.
75. Strauss, E. G., de Groot, R. J., Levinson, R., and Strauss, J. H. (1992) *Virology*, **191**, 932-940.
76. Lemm, J. A., Rumenapf, T., Strauss, E. G., Strauss, J. H., and Rice, C. M. (1994) *EMBO J.*, **13**, 2925-2934.
77. Shirako, Y., and Strauss, J. H. (1994) *J. Virol.*, **68**, 1874-1885.
78. Hanecak, R., Semler, B. L., Ariga, H., Anderson, C. W., and Wimmer, E. (1984) *Cell*, **37**, 1063-1073.
79. Brenner, S. (1988) *Nature*, **334**, 528-530.
80. Ypma-Wong, M. F., Filman, D. J., Hogle, J. M., and Semler, B. L. (1988) *J. Biol. Chem.*, **263**, 17846-17856.
81. Clark, M. E., Hammerle, T., Wimmer, E., and Dasgupta, A. (1991) *EMBO J.*, **10**, 2941-2947.
82. Clark, M. E., Lieberman, P. M., Berk, A. J., and Dasgupta, A. (1993) *Mol. Cell. Biol.*, **13**, 1232-1237.
83. Joachims, M., Harris, K. S., and Etchison, D. (1995) *Virology*, **211**, 451-461.
84. Tesar, M., and Marquardt, O. (1990) *Virology*, **174**, 364-374.
85. Andino, R., Rieckhof, G. E., Achacoso, P. L., and Baltimore, D. (1993) *EMBO J.*, **12**, 3587-3598.
86. Krausslich, H. G., Nicklin, M. J., Toyoda, H., Etchison, D., and Wimmer, E. (1987) *J. Virol.*, **61**, 2711-2718.
87. Bazan, J. F., and Fletterick, R. J. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 7872-7876.
88. Sommergruber, W., Ahorn, H., Zophel, A., Maurer-Fogy, I., Fessl, F., Schnorrenberg, G., Liebig, H. D., Blaas, D., Kuechler, E., and Skern, T. (1992) *J. Biol. Chem.*, **267**, 22639-22644.
89. Sommergruber, W., Ahorn, H., Klump, H., Seipelt, J., Zoephel, A., Fessl, F., Krystek, E., Blaas, D., Kuechler, E., Liebig, H. D., et al. (1994) *Virology*, **198**, 741-745.