Biochemistry (Moscow), Vol. 70, No. 6, 2005, pp. 672-684. Translated from Biokhimiya, Vol. 70, No. 6, 2005, pp. 814-828.

Originally published in Biochemistry (Moscow) On-Line Papers in Press, as Manuscript BM04-127, September 12, 2004.

Cloning and Molecular Modeling of Duodenase with Respect to Evolution of Substrate Specificity within Mammalian Serine Proteases That Have Lost a Conserved Active-Site Disulfide Bond

T. S. Zamolodchikova[1]*, E. V. Smirnova[1], A. N. Andrianov[1], I. V. Kashparov[1], O. D. Kotsareva[1], E. A. Sokolova[1], K. B. Ignatov[1], and A. D. Pemberton[2]

[1]*Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, ul. Miklukho-Maklaya 16/10, 117997 Moscow, Russia; fax: (7-095) 335-7103; E-mail: tatyana@enzyme.siobc.ras.ru*

[2]*Department of Veterinary Clinical Studies, Welcome Trust Centre for Research in Comparative Respiratory Medicine, University of Edinburgh, Easter-Bush Veterinary Centre, Roslin, Midlothian EH25 9RG, UK*

Received May 19, 2004
Revision received July 2, 2004

Abstract—Mammalian serine proteases such as the chromosome 14 (*Homo sapiens*, *Mus musculus*) located granzymes, chymases, cathepsin G, and related enzymes including duodenase collectively represent a special group within the chymotrypsin family which we refer to here as "granases". Enzymes of this group have lost the ancient active-site disulfide bond Cys191–Cys220 (bovine chymotrypsinogen A numbering) which is strongly conserved in classic serine proteases such as pancreatic, blood coagulation, and fibrinolysis proteases and others (granzymes A, M, K and leukocyte elastases). We sequenced the cDNA encoding bovine (*Bos taurus*) duodenase, a granase with unusual dual trypsin-like and chymotrypsin-like specificity. The sequence revealed a 17-residue signal peptide and two-residue (GlyLys) activation peptide typical for granases. Production of the mature enzyme is apparently accompanied by further proteolytic processing of the C-terminal pentapeptide extension of duodenase. Similar C-terminal processing is known for another dual-specific granase, human cathepsin G. Using phylogenetic analysis based on 39 granases we retraced the evolution of residues 189 and 226 crucial for serine protease primary specificity. The analysis revealed that while there is no obvious link between mutability of residue 189 and the appearance of novel catalytic properties in granases, the mutability of residue 226 evidently gives rise to different specificity subgroups within this enzyme group. The architecture of the extended substrate-binding site of granases and structural basis of duodenase dual specificity based on molecular dynamic method are discussed. We conclude that the marked selectivity of granases that is crucial to their role as regulatory proteases has evolved through the fine-tuning of specificity at three levels—primary, secondary, and conformational.

Key words: duodenase cDNA, chymotrypsin family, phylogeny, substrate specificity, granase, molecular dynamic

The chymotrypsin family (family S1 of clan SA according to the Rawlings and Barrett classification [1]) is the largest evolutionary family of serine proteases, presently including more than 150 members. Enzymes belonging to the family are proteases of different activities participating in important physiological processes such as food digestion, blood coagulation and fibrinolysis, immune response, complement system activation, apoptosis and others [2]. A complex of genes residing on mouse (*Mus musculus*) and human chromosome 14 encodes a number of serine proteases that are stored in the secretory granules of cytotoxic T-lymphocytes, natural killer cells, mast cells, and some other hemopoietic cells [3]. Comparative structural–functional analysis of bovine (*Bos taurus*) duodenase and related serine proteases from the hemopoietic cells (some cytotoxic T-cell granzymes, mast cell chymases, leukocyte cathepsin G, and others) revealed a special group of serine proteases within the chymotrypsin family which we earlier named "graspases" [4] but in the present work the name is modified to

Abbreviations: MCP) mast cell protease; GLP) granzyme-like protein; NKP) natural killer protease; HLE) human leukocyte elastase (medullasin); PRN) human leukocyte protease 3 (myeloblastin); ChlP) chymase-like protease.

* To whom correspondence should be addressed.

0006-2979/05/7006-0672

"granases" (granzyme-like proteases). The main structural characteristic of granases is the absence of the "active site" disulfide bond (Cys191–Cys220, according to standard bovine chymotrypsinogen A numbering), which is highly conserved in the structure of well-known serine proteases such as trypsin, chymotrypsin, and elastase, blood clotting and fibrinolysis factors, etc. (we designate them as the "trypsin group"). Besides, it was postulated that the crucial role in granase P1-specificity is apparently played by the residue at position 226, whereas the primary specificity of trypsin group enzymes generally depends on residue 189. Analyzing all known structures of granases [4], we observed a strict correlation between the nature of residue 226 and substrate specificity of an enzyme. Based on the nature of residue 226, and the corresponding substrate specificity of the enzyme, we divided granases into three specificity subgroups. Firstly, "chymozymes" are defined as granases with an uncharged residue at 226 and displaying chymotrypsin-like or elastase-like specificity. "Duozymes" are granases having a negative charged residue at 226 and possessing dual, trypsin-like and chymotrypsin-like specificity. Finally, "aspozymes" are Asp-specific granases with Arg at 226. Some main aspects of our previous work were experimentally confirmed by the mutagenesis technique in a later published study [5].

Ruminant proteases such as bovine duodenase (localized to epithelial cells of Brunner's (duodenal) glands and mast cells within the intestinal mucosa [6-8]) and sheep (*Ovis aries*) mast cell proteases 1-3 (sMCP) [9] structurally belong to granases. Duodenase was proposed to be the primary activator of the enteropeptidase zymogen, a key enzyme of the digestive protease activation cascade [10]. In evidence to a wider physiological role, it was demonstrated that duodenase can induce pulmonary artery fibroblast DNA synthesis, apparently via proteolytic activation of a G-protein-coupled receptor [8].

Duodenase as well as sMCP 1 possesses dual trypsin- and chymotrypsin-like activity [7]. Crystal data and computer modeling reveal that the S1 subsite of duodenase (nomenclature of Shechter and Berger [11]) has structural features compatible with effective accommodation by P1 residues typical of trypsin (Arg/Lys) and chymotrypsin (Tyr/Phe) substrates [12].

In this paper we report duodenase cDNA sequence revealed duodenase precursor is similar to that of other granases, showing the closest relation to human cathepsin G. Besides, the structural basis of duodenase dual specificity based on molecular dynamic data is proposed and the peculiarities of architecture of the substrate-binding site of granases providing their pronounced secondary and tertiary (conformational) specificity are discussed. Retracing the evolution of residues 189 and 226 crucial for serine protease primary specificity within the granase group, we confirmed the defining role of residue 226 in the evolution of specificity of granases.

MATERIALS AND METHODS

Duodenase cDNA cloning and sequencing. A sample of bovine duodenum was taken at slaughter and quickly frozen in liquid nitrogen. The frozen tissue was homogenized under continuous cooling using liquid nitrogen, and RNA was isolated using acid guanidinium thiocyanate–phenol–chloroform extraction [13]. cDNA was synthesized using RNA from bovine duodenum as template, oligo dT_{20} primer, and SuperScript II reverse transcriptase (Life Technology, Inc., USA) as recommended. The resulting cDNA was subjected to PCR with the primers P1 and P2 designed from the 5′ and 3′ UTR of sheep mast cell protease (SMCP-1) gene (the closest analog of duodenase [9]). Sequences of primers used were as follows:

P1: 5′-TCTGGGCAGTCTTCCTGGAGAG,

P2: 5′-AGATGAATGTTTATTAAGTTCCAGGC.

PCR reaction mixture (30 μl) contained 10 mM Tris-HCl, pH 9.0, 50 mM KCl, 1.5 mM $MgCl_2$, 0.1% Triton X-100, 0.25 mM of each dNTP, 10 pmol of each primer, and 50 ng cDNA as template. After denaturation (95°C, 5 min), 2 units of Taq DNA polymerase were added at 72°C and the amplification program (94°C for 30 sec, 65°C for 30 sec, 72°C for 30 sec) was run for 30 cycles. The reaction product was analyzed by electrophoresis on a 2% agarose gel. The PCR product of the expected size was cut from the gel, purified by Gene Clean (Promega, USA), and after re-amplification under the same PCR conditions, was cloned into TA cloning vector pGEM-T (Promega). The nucleotide sequence was determined by the Sanger method using Sequenase Version 2.0 (Amersham, UK).

Calculation details. Structural models of duodenase complexes with oligopeptide substrates were built using the atomic coordinates of the free duodenase molecule (PDB code: 1EUF [12]). Molecular docking of the substrate molecules into the enzyme active site was done using programs DeepView [14] and eFold [15]. Both substrates had identical P4-P2 and P1′-P3′ residues and P1 residues were either Lys or Phe, referred to as P1 Lys-substrate and P1 Phe-substrate, respectively. Substrate sequences listed below (ACE means acetyl and NME means N-methyl) were taken from the substrates with high experimentally determined affinity to duodenase:

P4 P3 P2 P1 P1' P2' P3'
ACE – Ser-Ile-Pro-*Lys*-Ser-Leu-Phe-NME
Lys-substrate

P4 P3 P2 P1 P1' P2' P3'
ACE – Ser-Ile-Pro-*Phe*-Ser-Leu-Phe-NME
Phe-substrate.

The position for the center of the solvation (simulation) sphere was chosen at the centroid point of the C_α atoms of all enzyme amino acids within 6 Å from the P1 substrate residue. The molecular dynamics (MD) paths and the conformational energy of the flap were calculated using the Q program [16] and the semiempirical potential field CHARMM22 [17]. The simulation was performed in a sphere of 25 Å radius from the center chosen as above. All protein atoms within the simulation sphere were considered as mobile, while beyond it the position of atoms was strictly fixed to their experimental coordinates by imposing penalty functions as standard in the Q program. The examined enzyme region was filled with water molecules pre-equilibrated at 300 K and normal pressure. The CHARMM function was calculated using a 10 Å cutoff and taking into account the long-range electrostatic interactions computed with the LRF algorithm [18]. At the same time, no distance restrictions were imposed when calculating the non-bonded interactions of the substrate molecule with the surrounding groups. The charge neutralization model was used for ionogenic groups farther than 8 Å away from the simulation center. The water molecules closer than 2.4 Å to protein non-hydrogen atoms and to the three fixed water molecules were removed from the coordinate set. Prior to MD registration, the three model systems were heated with equilibration consecutively at 1, 50, 150, and 300 K. During heating, the displacement of protein heavy atoms from the initial X-ray coordinates was restricted by imposing a harmonic penalty function with a force constant of 5.0 kcal/(mol·Å^2). At constant particle number, pressure, and temperature of 300 K, the data were registered with a 1-fsed step for 300 psec of the productive phase following equilibration over 300 psec (monitored with the dependence of overall energy on simulation time). Use was made of the Berendsen coupling to external bath [19]. Hydrogen mobility of solvent molecules was limited with the SHAKE algorithm [17]. For an adequate scanning of the conformational space available for the bound substrate molecules, each model was subjected to six independent MD runs with different initial Maxwellian distributions of atom velocities. In every run, the mean paths were evaluated for the thermodynamically equilibrated system. The programs VMD 1.7.1 [20] and DeepView [14] were used for the visual inspection of the molecular structures and visualization of the MD trajectories.

RESULTS AND DISCUSSION

Analysis of duodenase cDNA sequence. Duodenase cDNA was obtained by RT-PCR using RNA isolated from bovine duodenum and primers which were designed from the 5′ and 3′ UTR of the sheep mast cell protease 1 (sMCP-1) [9], the closest analog of duodenase. The product of RT-PCR (850 bp) was obtained and cloned into the pGEM vector. Four individual clones containing cDNA of duodenase were sequenced. The duodenase nucleotide sequence (Gene Bank No. AF198965) is presented in Fig. 1. The open reading frame of 753 bp encodes a 251-amino acid polypeptide. The first 22 and last 26 nucleotides correspond to P1 and P2 primers, respectively. The deduced amino acid sequence was compared with that of sMCP-1 and other related proteases. This revealed the existence of a duodenase preproenzyme composed of the signal peptide (the first 17 residues), the propeptide (Gly-Lys), and the catalytic domain, beginning with Ile-Ile-Gly-Gly and consisting of 232 residues including a C-terminal pentapeptide extension (Arg-Gln-Gly-Ser-Val) which was absent in the mature protein. The nucleotide and amino acid sequences of duodenase have 89 and 85% identity with sMCP-1, respectively. Some differences between the deduced and the earlier reported sequences [6], possibly reflecting a polymorphism, were found mainly in the C-terminal region of the enzyme molecule. They include insertion of Ser49, mismatching residues at positions 175 (Lys instead of Arg), 200 (Arg instead of Lys), and 217 (Ser instead of Pro) (duodenase numbering). Besides, there are differences in residues at positions 220-222 (His-Ser-Thr instead of Lys-Arg-Val) and 224-226 (Arg-Arg-Tyr instead of Tyr-Leu-Phe). According to the cDNA data presented here (Fig. 1), there is a C-terminal five-residue extension (residues 228-232) in the duodenase molecule. Previously, the C-terminal residue of duodenase was reported as Lys226 [6] (Lys227 in the present sequence). We suppose that aside from the well known N-terminal proteolytic processing for serine proteases, duodenase undergoes C-terminal truncation. This supposition is in accordance with the duodenase X-ray data. As was described, electron density after Arg224 in the enzyme molecule was absent [12]. Analogous posttranslational processing was described for human leukocyte cathepsin G, which possesses structural and enzymatic similarity with duodenase [21].

Comparison of prepropeptide sequences of known granases with those of trypsin group enzymes revealed significant structural differences between them. In the table, the most typical preprosequences of granases including duodenase and of some trypsin group enzymes are presented. Granases generally possess similar leader peptide structures, consisting of a hydrophobic sequence of 10-11 residues flanked by conserved proline residues. The activation dipeptide is one of three main types (Gly-Lys, Gly-Glu, or Glu-Glu), and generally is preceded by an alanine residue. Prepropeptides of trypsin group enzymes are significantly longer (about 30 residues) compared with granases. Activation peptides in trypsin group enzymes differ in their structure and length.

Duodenase, like human cathepsin G, is a protease with unusual dual specificity, and as such represents a promising subject for mutagenesis investigation. Despite

the existence of solved spatial structures for duodenase [12] and cathepsin G [22], the peculiarities of their substrate-binding pocket are still far from understood. In particular, how can substrates so different as hydrophobic and positively charged ones be equally suitable for binding in the enzyme substrate-binding pocket? To answer this question, we have investigated duodenase using molecular modeling (see below), and further studies of the enzyme by mutagenesis of crucial residues are planned.

Duodenase molecular modeling. In investigating duodenase using molecular dynamic modeling methods, we aimed to elucidate the structural determinants of duozyme specificity. Two peptide substrates (seven residues) differing only in P1 residue, corresponding to the trypsin-like (P1 = Lys) and chymotrypsin-like (P1 = Phe) specificities of duodenase (see "Materials and Methods"), were used for modeling of the ground state complex with duodenase (Fig. 2). P4-P3′ residues of the substrates were selected according to the experimentally determined subsite preferences of duodenase [7]. The first important peculiarity of the duodenase substrate-binding pocket to be revealed by means of the modeling experiment was formation of the ionic contact between residues Arg217 and Asp219 (standard chymotrypsinogen A numbering is used from this point). This contact was not observed in the duodenase crystal structure, which differs from our molecular dynamic simulation of solution structure due to intermolecular salt bridge formation between Arg217 and residue Glu165 of the neighboring duodenase molecule in the crystallographic cell. The native conformation of the Gly216-Thr224 loop was revealed in the model complexes of duodenase with both substrates.

The ion-pair bond Arg217–Asp219 obviously plays a stabilizing role in the substrate-binding pocket of the enzyme. As it is clear from Fig. 2, the conformation of the substrate-binding site slightly differs when it binds a charged or aromatic substrate. However, residues Asp226 and Ser214 that form part of the S1 binding pocket may significantly change their conformation depending on the nature of the P1 residue of the bound substrate. The negatively charged Asp226 side chain is directly involved in the substrate binding in the enzyme S1 pocket when forming a salt bridge with the ε-amino group of the P1 Lys substrate residue. In the case of the aromatic substrate, direct interaction between P1 Phe residue and Asp226 is

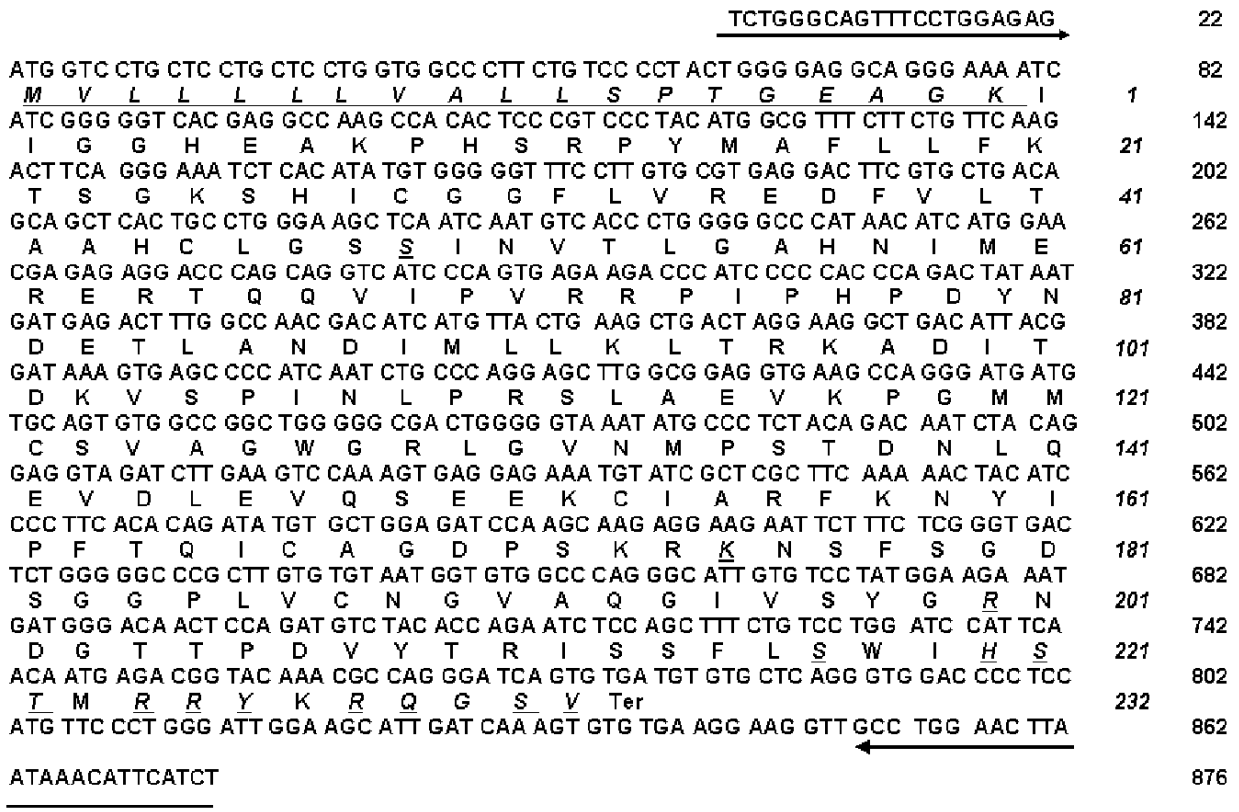

Fig. 1. Nucleotide sequence and deduced amino acid sequence of bovine duodenase. Nucleotides and deduced amino acids are numbered to the right of the figure. The numbering of amino acids (in italic) begins at the isoleucine that corresponds to the amino terminus of the mature enzyme. Differences in sequence between earlier reported [6] and present data are indicated by italic underlined letters. The locations and directions of primers used in the PCR experiments are indicated by arrows.

Comparison of preprosequences of some granases and trypsin group enzymes (granase preprosequences are aligned to maximize homology; hydrophobic sequences of the granase leader peptides are underlined)

Enzyme	Signal peptide	Activation peptide	GenBank accession No.
	Granases		
Duodenase	MV-LLLLLVALLSPTG--EA	GK	AF198965
Cathepsin G human	MQPLLLLLAFL--LPTGAEA	GE	NM_001911
MCP-8 rat	MF-LFLFFLVAILPVNTEG	GE	NM_101598
Granzyme B human	MQPILLLLAFLLLP--RADA	GE	NM_004131
Granzyme C mouse	MPPV-LILLTLLLPLRAGA	EE	NM_010371
MCP-1 mouse	MQAL-LFLMALLLPSGAGA	EE	NM_008570
Chymase human	MLLLPLPLLLFLLCSRAEA	GE	XM_007317
	Trypsin group		
Medullasin (HLE)	MTLGRRLACLFLACVLPALLLGGTAA	SE	NM_001972
Tryptase dog	MPSPLVLALALLGSLVPVS	PAPGQALQRBG	M24664
Granzyme A human	MRNSYRFLASSLSVVVSLLLIPEDVC	EK	XM_003652
Granzyme M mouse	MEVCWSLLLLLLALKTLWAAG	NRFETQ	AB015728
Chymotrypsin human	MAFLWLLSCWALLGTTFG	CGVPAIHPVLSGLSR	BC005385
Trypsin bovine	MHPLLILAFVGAAVAFP	VDDDDK	X54703
Elastase rat	MLRFLVFASLVLYGHS	TQDFPETNAR	NM012552

not observed. Moreover, the Asp226 side chain changes its conformation, noticeably deflecting from the aromatic ring of the substrate P1 residue (Fig. 2). In this case, negative charge of Asp226 distributes between the side chain nitrogen of Asn189 and well conserved in all simulations water molecule, which is bonded with the main chain oxygen of Thr223 and main chain nitrogen of Arg217 (Fig. 2). The benzyl side chain of the P1 Phe residue extends into the S1 specificity pocket and is sandwiched between the backbones of the enzyme fragments 190-192 and 215-216, similarly to the P1 Phe binding observed in human chymase [23]. Additional stabilization of the P1 Phe residue in the duodenase S1 subsite is achieved due to the hydrogen bond formation between the Phe residue amide nitrogen and Ser214 carbonyl oxygen (Fig. 2). In the case of P1 Lys residue binding, Ser214 changes its conformation and does not form the abovementioned bond and thus does not contribute to productive positioning of the scissile bond. Thus, the unique conformational behavior of duodenase S1 residues Asp226 and Ser214 provides inducible conformity between the different substrate P1 residues and the duodenase S1 subsite. The P2 Pro residue, in common with many serine proteases, is the most favorable for duodenase due to its specific conformation conferring a kink to the polypeptide chain, as necessary for the following P1 residue to insert into the S1 subsite. Gly216 in duodenase participates in positioning both substrates (charged or hydrophobic) in the enzyme active center through hydrogen bond formation with the P3 residue. Gly216 is located at the rim of the S1 pocket in duodenase and its main-chain conformation observed in this investigation is similar to that of chymotrypsin and chymases [24]. The unique properties of the duodenase substrate-binding pocket impart to residues Gly216, Ser214, and Asp226 specific conformational behavior that makes both positively charged and hydrophobic substrates equally cognate for the enzyme.

In common with other granases, duodenase shows marked selectivity towards the P1′-P4′ sequences of the substrates [7]. Molecular dynamic studies demonstrate that in contrast to trypsin and chymotrypsin, peptide groups of duodenase residues Ile41 and Ser39 are able to form two additional hydrogen bonds with the P2′ and P4′ substrate residues (see below).

Thus, the backbone of Lys-containing as well as Phe-containing substrate juxtaposes the duodenase substrate-binding site (S3-S1 and S2′-S4′ fragments) in an antiparallel β-sheet manner.

The P1 residue in the case of Lys (Arg [12]) directly interacts with the primary specificity determinant (Asp226). However, the charged substrate hydrolysis efficiency also depends on the conformity between the P3-P2 and P1′-P4′ substrate sequences and duodenase sec-

ondary specificity determinants (Gly216, Ile41, Ser39) although the P1 side chain (Lys, Arg)/S1 (Asp226) ion interaction plays the dominant role in the substrate binding. When a Phe-containing substrate binds in the enzyme binding site the P1 residue side chain hydrophobic interactions, hydrogen bond formation with Ser214, and secondary binding sites apparently play the determining role in the productive positioning of the substrate.

Evolution of residues 189 and 226 crucial for serine protease substrate specificity. Because it was supposed that the primary specificity of granases mainly depends on the nature of residue 226 [4, 25], we assumed the mutability of this crucial residue is the factor determining the origin of the granase specificity subgroups. In our previous work [4], we constructed a phylogenetic tree for the enzymes including almost all known granases and some typical representatives of the trypsin group (Fig. 3). The tree demonstrates that proteases of the chymotrypsin family are divided along two main branches, leading to granases (proteins *1-39*) and trypsin group enzymes (proteins *42-53*).

Based on the phylogeny of the enzymes, we constructed phylogenic trees for residues 189 and 226, which are the primary specificity determinants in chymotrypsin family proteases (Fig. 4). When we constructed the trees, we aimed to discover how nucleotide triplets encoding the residues have evolved from one specificity group of granases to another. Additionally, we tried to reconstruct the coding base triplets for hypothetical ancestors using post-order traversal and Farris optimization methods [26]. By these methods, we were able to determine the coding base triplets for most of the internal nodes of the trees. However, the determination of the triplet coding for residue 226 of the common ancestor of all proteases examined here was ambiguous because there were two possible combinations of the nucleotides: GG(A, C) or GC(A, C) that encode glycine and alanine residues, respectively. Taking into account the absolute conservation of Gly at position 226 in polypeptide structures of trypsins from primitive organisms (bacteria, yeast, fungi) and that these trypsins are believed to be the most ancient

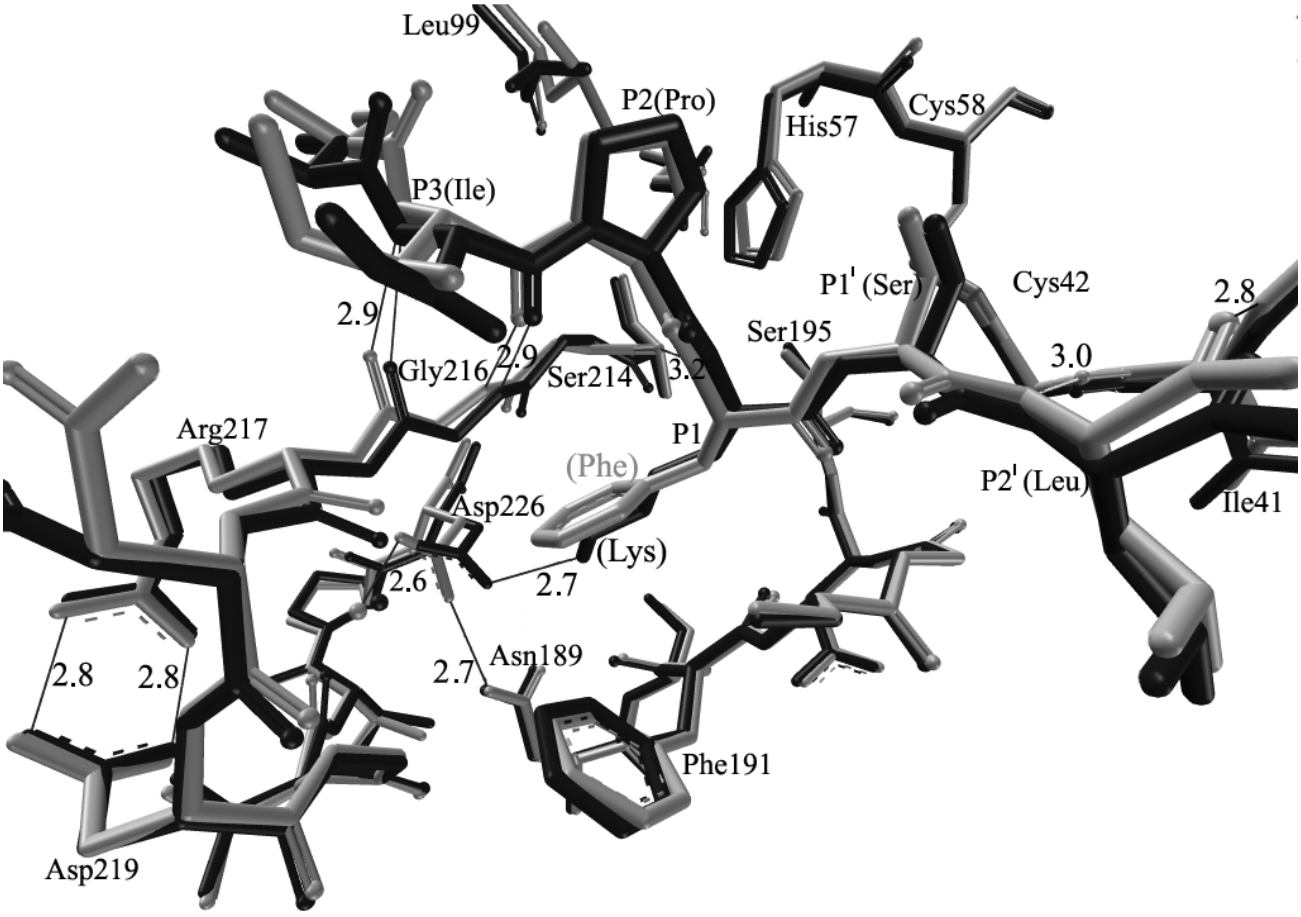

Fig. 2. Structural basis of duodenase dual specificity. The proteinase substrate-binding site is represented in complex with the P1 Lys substrate (dark) and P1 Phe substrate (light). The salt bridge between the enzyme residues Arg217 and Asp219, as well as the ion pair interaction between the enzyme S1 (Asp 226) and substrate P1 (Lys) residues, and hydrogen bonds stabilizing the substrates in the duodenase substrate-binding pocket are indicated.


MCP-1 sheep 1
MCP-3 sheep* 2
duodenase cow 3
cathepsin G mouse 4
cathepsin G human 5
MCP-8 rat 6
GLP II rat 7
granzyme J rat 8
MCP-8 mouse 9
granzyme H human 10
granzyme B human 11
GLP III rat 12
GLP I rat 13
NKP 1 rat 14
granzyme B mouse 15
granzyme C mouse 16
NKP 4 rat 17
granzyme D mouse 18
granzyme E mouse 19
granzyme G mouse 20
NKP 7 rat 21
granzyme F mouse 22
MCP-4 mouse 23
myonase mouse 24
MCP-1 rat 25
MCP-1 gerbil 26
MCP-1 hamster 27
MCP-2 mouse 28
MCP-1 mouse 29
MCP-2 rat 30
MCP-9 mouse 31
chymase human 32
chymase hamadryas 33
chymase dog 34
MCP-2 sheep 35
MCP-3 rat 36
MCP-5 mouse 37
MCP-2 gerbil 38
MCP-2 hamster 39
medullasin human (HLE) 40
myeloblastin human (PRN) 41
granzyme K mouse 42
granzyme K rat 43
granzyme K human 44
granzyme A human 45
granzyme A mouse 46
granzyme M mouse 47
granzyme M rat 48
granzyme M human 49
tryptase cow 50
chymotrypsinogen A cow 51
trypsinogen cow 52
elastase 1 cow 53
1448.0
1000
500
200
100
50
20
10
5
2
1
0

serine proteases [27], we assumed residue 226 of the common ancestor to be glycine and the corresponding coding nucleotide triplet to be GG(A, C) (indicated by asterisk in Fig. 4b).

The obtained trees illustrate that granase specificity groups (chymozymes, aspozymes, duozymes [4]) are derived from the granase ancestor that presumably possesses Ala189 and Gly226, which are encoded by GC(T, A) and GGA nucleotide triplets, respectively (indicated by open arrows in Fig. 4). Recently, a hypothetical ancestor that was recreated using phylogenetic inference, gene synthesis, and protein expression was published [5]. It has broad specificity encompassing the entire repertoire of primary specificities found in its descendants. The ancient triplets coding Ala189 and Gly226 were inherited by mouse granzyme G and rat natural killer protease 7 (proteins *20* and *21*) which structurally belong to the chymozymes. Single base substitutions in the triplets encoding residues 189 and 226 led to the appearance of a chymase ancestor with Ser189 and Ala226 (indicated by black arrow in Fig. 4), from which the α- and β-chymases (proteins *23-39*) are derived. As for residue 189, further mutations of Ser coded by the TCA triplet in the chymase ancestor led to the appearance of three additional variants of residue 189 (Thr, Ala, and Asn) (Fig. 4a).

Another group of granases also possessing chymotrypsin-like specificity and comprising proteins *18-22* (murine granzymes D-G and rat natural killer protease 7 (NKP 7)) inherited the same nucleotide triplet (GGA) as encodes Gly226 of the common ancestor (Fig. 4b). In the case of these enzymes, residue 189, being represented by three variants of uncharged residues (Thr, Ala, and Ser) (Fig. 4a), is more variable than residue 226, as in the chymases. Thus, residue 226 in chymases and the mentioned granzymes is significantly more stable, as regards evolutional mutability, than residue 189 and may be combined with the different (uncharged) residues observed at position 189 in the enzymes retaining the chymotrypsin-like specificity.

Substitution of a single base in the ancestral coding triplet (GGA→CGA) apparently resulted in a change in primary specificity of the hypothetical ancestor due to the appearance of a positively charged Arg residue at position 226 (Fig. 4b). Granases containing Arg226 (granzymes B) are known to be Asp-specific proteases [28], i.e., according to our classification are aspozymes. Proteins *16* and *17* (murine granzymes D-G and rat natural killer protease 4 (NKP 4)), although descendants of the same ancestor as aspozymes *13-15*, have Gln at position 226 due to CGA→CA(A, G) substitution. These Gln226-containing enzymes are predicted to be chymozymes [28, 29].

In the case of rat granzyme-like protein III (protein 12), which is positioned on the phylogenic tree among the aspozymes, there are substitutions of the two first nucleotides encoding Arg226 in aspozymes (CGG→GAG), leading to the appearance of a negative charge at position 226 (Fig. 4b). Consequently, the enzyme was predicted to be a duozyme [30]. A reverse synonymous base substitution (CGG→CGA) in the triplet, encoding residue 226, takes place in aspozymes *7*, *8* (rat GLPII and rat granzyme J), and *11* (human granzyme B) (Fig. 4b).

← **Fig. 3.** Phylogenetic tree for 53 aligned mammalian serine proteases of trypsin and granase groups. The tree was originally published in [4]. Granase specificity subgroups are marked (proteins *1-39*): duozymes with white letters in black boxes, aspozymes with white letters in black ovals, chymozymes with bold letters. Protein *2* (marked with an asterisk) cannot be attributed to any specific group (see text). Proteases of trypsin group (proteins *40-53*) are shown in Italic letters. Amino acid sequences of 53 proteins (numbers correspond to those in Fig. 4) were extracted from the GenBank nucleotide sequence data bank, their proposed names and accession numbers are given in parenthesis: *1*) sheep mast cell protease 1 (MCP-1 sheep, H14654); *2*) sheep mast cell protease 3 (MCP-3 sheep, Y13462); *3*) bovine duodenase (duodenase cow, AF198965); *4*) mouse cathepsin G (cathepsin G mouse, NM_007800); *5*) human cathepsin G (cathepsin G human, NM_001911); *6*) rat mast cell protease 8 (MCP-8 rat, NM_101598); *7*) rat granzyme-like protein II (GLP II rat, X68657); *8*) rat granzyme J (granzyme G rat, U72143); *9*) mouse mast cell protease (MCP-8 mouse, X78545); *10*) human granzyme H (granzyme H human, M57888); *11*) human granzyme B (granzyme B human, NM_004131); *12*) rat granzyme-like protein III (GLP III rat, X76996); *13*) rat granzyme-like protein I (GLP I rat, X66693); *14*) rat natural killer protease (NKP 1 rat, M34097); *15*) mouse granzyme B (granzyme B mouse, NM_013542); *16*) mouse granzyme C (granzyme C mouse, NM_010371); *17*) rat natural killer protease 4 (NKP 4 rat, U57062); *18*) mouse granzyme D (granzyme D mouse, J03255); *19*) mouse granzyme E (granzyme E mouse, NM_010373); *20*) mouse granzyme G (granzyme G mouse, NM_010375); *21*) rat natural killer protease 7 (NKP 7 rat, U57063); *22*) mouse granzyme F (granzyme F mouse, NM_010374); *23*) mouse mast cell protease 4 (MCP-4 mouse, AY007569); *24*) mouse myonase (myonase mouse, AB051900); *25*) rat mast cell protease 1 (MCP-1 rat, U67915); *26*) gerbil mast cell protease 1 (MCP-1 gerbil, D45173); *27*) hamster mast cell protease 1 (MCP-1 hamster, D85517); *28*) mouse mast cell protease 2 (MCP-2 mouse, NM_008571); *29*) mouse mast cell protease 1 (MCP-1 mouse, NM_008570); *30*) rat mast cell protease 2 (MCP-2 rat, J02712); *31*) mouse mast cell protease 9 (MCP-9 mouse, AY007568); *32*) human chymase (chymase human, XM_007317); *33*) hamadryas chymase (chymase hamadryas, U38521); *34*) dog chymase (chymase dog, U89607); *35*) sheep mast cell protease 2 (MCP-2 sheep, Y08133); *36*) rat mast cell protease 3 (MCP-3 rat, D38495); *37*) mouse mast cell protease 5 (MCP-5 mouse, M73759); *38*) gerbil mast cell protease 2 (MCP-2 gerbil, D45174); *39*) hamster mast cell protease 2 (MCP-2 hamster, AB007622); *40*) human leukocyte elastase (medullasin human (HLE), NM_001972); *41*) human proteinase 3 (myeloblastin human (PRN 3), NM_002777); *42*) mouse granzyme K (granzyme K mouse, NM_008196); *43*) rat granzyme K (granzyme K rat, NM_017119); *44*) human granzyme K (granzyme K human, XM_003782); *45*) human granzyme A (granzyme A human, XM_003652); *46*) mouse granzyme A (granzyme A mouse, NM_010370); *47*) mouse granzyme M (granzyme M mouse, AB015728); *48*) rat granzyme M (granzyme M rat, L05175); *49*) human granzyme M (granzyme M human, NM_005317); *50*) bovine tryptase (tryptase cow, X94982); *51*) bovine chymotrypsinogen A (chymotrypsinogen A cow, P00776); *52*) bovine trypsinogen (trypsinogen cow, X54703); *53*) bovine elastase 1 (elastase 1 cow, M80838).

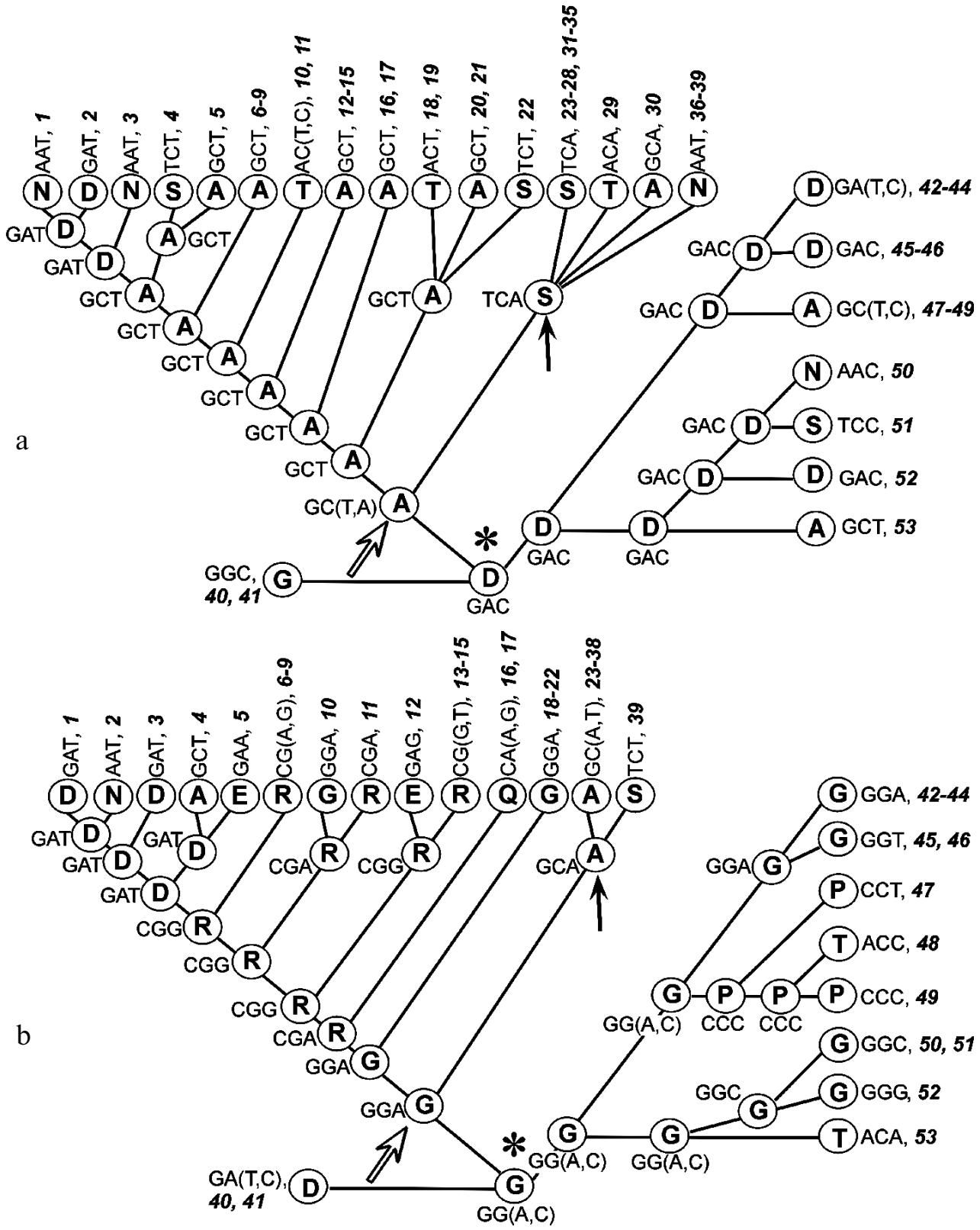

Fig. 4. Comparative phylogeny of S1 residues 189 (a) and 226 (b). Protein numbers are the same as in Fig. 3 (protein *51* is human chymotrypsin). Nucleotide triplets encoding the residues are shown near circles, and the corresponding amino acid is indicated in the circle. The residues of some hypothetical ancestors are marked (common ancestor by asterisk, granase and chymase ancestors by white and black arrows, respectively).

Further mutation (CGA→GGA) leads to the Arg226→Gly226 replacement in human granzyme H (protein *10*). This is the cause of the enzyme chymotrypsin-like activity, although it is the nearest structural relative of the classic aspozyme, human granzyme B [31]. This case may be considered as a reversion to the ancestral condition where residue Gly226, which was present in the granase ancestor, appears again in the descendant after several mutations [32]. In this subgroup of enzymes (proteins *6-17*), residue 189 is a conserved Ala (Fig. 4a), with the exception of human granzymes B and H (proteins *10*, *11*), which contain Thr189. Thus, the nature of residue 189 is more restricted in aspozymes than chymozymes, reflecting a relatively narrow specificity.

Finally, the duozyme group (proteins *1-5*) apparently branched from aspozymes due to the replacement of all three nucleotides (CGG→GAT) encoding residue 226 (Fig. 4b). As a result, Asp226 in the hypothetical duozyme ancestor was substituted for Arg226. This characteristic change was inherited by sMCP-1 (protein *1*) and duodenase (protein *3*). A further single base substitution (GAT→GAA), which occurred in the gene of human cathepsin G (protein *5*), led to the appearance of a Glu residue at position 226 of the polypeptide chain. Thus human cathepsin G, having an acidic residue at the position 226, remains a duozyme [33].

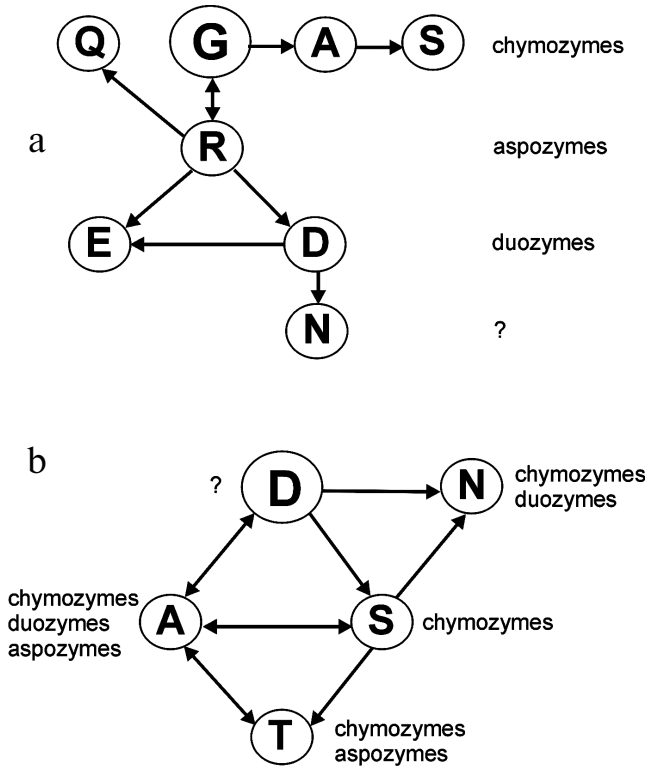

Fig. 5. Modulation of granase specificity depending on nature of the amino acid residues at positions 226 (a) and 189 (b). Residues in the big circle, glycine (a) and aspartate (b), are predicted to be ancestral (see text). Arrows indicate possible mutations of the ancestral residues leading to appearance of the novel residues at the crucial positions. Specificity of granase containing Asn226 and Asp189 (sMCP-3, see text) is ambiguous (indicated by "?").

Among the duozyme branch, there are two enzymes (sMCP-3, protein *2*, and murine cathepsin G, protein *4*), that apparently have lost dual trypsin- and chymotrypsin-like specificity. The first of them has a single nucleotide replacement (GAT→AAT), leading to the Asp226→Asn226 mutation. It is difficult to predict the substrate specificity of sMCP-3, for which only the nucleotide sequence is known [9]. In the second one, a GAT→GCT substitution at residue 226 has led to the appearance of an Ala residue in this position instead of the Glu226 present in human cathepsin G (Fig. 4b). Murine cathepsin G would therefore be expected to possess chymotrypsin-like specificity, due to the combination of uncharged residues Ala226 and Ser189 in its S1-subsite.

The majority of trypsin group proteases considered here possess a conserved Gly residue at position 226, which was inherited from the common ancestor (Fig. 4b; proteins *42-46*, *50-52*). The presence of the small Gly residue in this position apparently is a necessary condition for catalytic activity of proteases with trypsin-like specificity [34]. As was discussed above, the main specificity-determining residue in the trypsin group proteases is residue 189, positioned at the bottom of the substrate-binding pocket. The role of the residue at position 226 is associated rather with the general architecture of the enzyme substrate-binding pocket [30]. In some cases (granzymes M, proteins *47-49*), pancreatic elastase (protein *53*), leukocyte elastase and myeloblastin (proteins *40* and *41*) (Fig. 4b) position 226 is occupied by residues other than glycine. However, the side chains of the residues at position 226 as was demonstrated for human granzyme M and leukocyte elastase do not influence the substrate binding in the S1 subsite of the enzymes [35, 36]. The presence of Thr226 as well as Val216[1] reduces the substrate-binding pocket volume in pancreatic elastase and thus restricts the enzyme P1 specificity to small aliphatic amino acids. On the whole, it appears that mutations of residue 226 in proteases of the trypsin group are less effective for producing novel enzymatic properties compared with granases.

Thus, based on the above results, we conclude that the mutability of specificity-determining residue 226 is an important factor in the phylogeny of granases. Because the enzymatic properties of granases are highly dependent on the nature of this residue, then mutations leading to significant alterations in its properties (change or loss of charge) have formed the origin of the observed specificity subgroups (Fig. 5a). Residue 189, which is important for the specificity of the trypsin group enzymes, is

[1] Analogous role is played by Val216 in some rodent chymases [37].

uncharged in all granases (with the exception of sMCP-3), and there is no obvious link between its nature and the catalytic properties of granases (Fig. 5b).

Differences in structural elements crucial for substrate binding in granase and trypsin groups. When the spatial structures of the five granases for which crystal data are available (human granzyme B, chymase, and cathepsin G; rat mast cell protease 2 (rMCP-2) and bovine duodenase) were superimposed with those of rat anionic trypsin and bovine chymotrypsin A, the principal differences between granase and "classic" serine protease substrate-binding sites were observed (Fig. 6). The main differences are associated with the loops 214-227 (Fig. 6a) and 34-42 (Fig. 6b) that are crucial for the enzyme–substrate interaction. As mentioned above, granases have lost the "active site" disulfide bond Cys191–Cys220. In addition, compared with the trypsin group enzymes, residues at positions 222-224 are deleted in granases. As a result, the loop in granases is significantly shorter and has a different conformation, which is similar in all granases.

The conformational rigidity of the substrate-binding pocket in granases can be achieved by several means. As was mentioned, the absence of the "active site" disulfide bond in duodenase is compensated by salt bridge formation between Arg217 and Asp219. Analogous pairs of charged residues are present in some other granases (rMCP-1, rat granzyme-like protein I, and human chymase), apparently stabilizing their substrate-binding pocket in a similar way to that observed in duodenase. Besides, the fragment 217-219 in granases is generally formed by residues with bulky side chains, and this may also decrease conformational mobility of the loop 216-224. The side chain of Phe191, which is highly conserved in granases, spatially corresponds to the lost "active-site" disulfide bond and participates in the S1 pocket wall formation.

Primary specificity determinants, which are different for granases (Asp226, duodenase) and "classic" serine proteases (Asp189, trypsin), are indicated in Fig. 6a. Compared with trypsin and chymotrypsin, the polypeptide chain fragment 224-227 in granases is shifted in such a way as to broaden the corresponding part of the substrate-binding pocket. Relocation of the specificity-determining residue at position 226 in granases does not reduce the pocket volume as much as it does in trypsin mutant Asp189Gly/Gly226Asp [38]. This peculiarity of granases seems to explain why the alternative location of the P1 specificity determinant in the mutant trypsin led to a significantly less effective protease. Only position 189 is effective for the primary determinant negative charge location in trypsin, and this explains the conservation of Asp189 in trypsin-like serine proteases. In granases, position 226 may be occupied by rather bulky residues such as Asp, or even Glu (duozymes) or Arg (aspozymes). These residues serve as the primary specificity determinant in this enzyme group, being available for the substrate P1 residue due to the specific conformation of the substrate-binding pocket in granases.

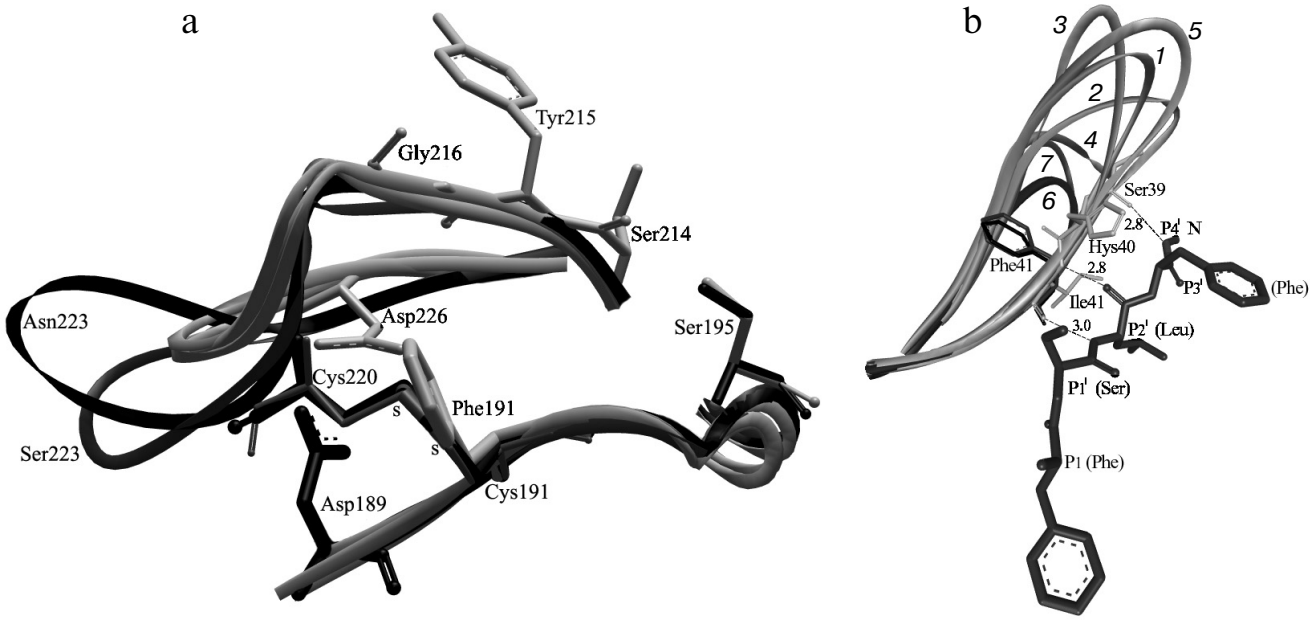

Fig. 6. Differences in structural elements crucial for substrate binding in granases and trypsin group enzymes. The models shown represent (a) the vicinity of the S1 substrate-binding site, and (b) the S2′-S4′ binding-site (loop 34-42). The structures of trypsin (dark) and chymotrypsin (grey) are superimposed with the five available granase structures (light), namely: *1*) human granzyme B [25]; *2*) human chymase [23]; *3*) rat mast cell protease 2 [39]; *4*) human cathepsin G [22]; *5*) bovine duodenase [12]; *6*) rat anionic trypsin [40]; *7*) bovine chymotrypsin [41]. Duodenase residues participating in binding the substrate P2′-P4′ residues are indicated as stick models.

Another important structural difference between granases and trypsin group enzymes is the special conformation of their loop 34-42. Compared with chymotrypsin, all known granases possess an insertion of three residues (36A–C), which causes a significant deviation of the polypeptide backbone in the loop (Fig. 6b). Although all C_α-atoms of the residues at position 41 (Phe in trypsin and chymotrypsin and Ile in duodenase) in the superimposed structures are less than 1 Å apart from one another, in granases this residue has significantly different conformation. The amide group associated with the 40–41 peptide bond and peptide carbonyl of residue 39 are quite differently located in granases compared with that in trypsin and chymotrypsin. Unlike the classic proteases, the amide and carbonyl groups in granases point towards the active-site cleft and provide additional hydrogen bond formation with a substrate P2′ and P4′ residues. Structural studies of granases particularly emphasize the importance of the interactions on the S2′-S4′ side of the enzyme extended substrate-binding site and implicate residues at positions 40 and 41 as residues with a significant role in substrate selection [25]. Because residues 39-41 in granases are involved in additional potential antiparallel β sheet forming between them and the main chain of the substrate P2′-P4′, the unique extended specificity of granases may result from interactions at the S4-S2 and S1′-S4′ sites. Moreover, special conformation of loop 34-42 in granases (due to specific insertion) allows the enzymes to prevent non-cognate substrate binding because of spatial restrictions. Thus, loop 34-42 provides the marked secondary and conformational specificity of the proteases, which evolved mainly as regulatory rather than destructive enzymes. On the other hand, the amino acid sequence within the loop varies significantly from one granase to another, and this is one of the factors determining the different secondary and conformational specificities among the enzymes having similar P1 (primary) specificity. Having together the presented results, we conclude that the marked selectivity of granases that is crucial to their role as regulatory proteases has evolved through the fine-tuning of specificity at three levels—primary, secondary, and conformational.

This work was supported by a grant from the Russian Foundation for Basic Research (99-04-48135).

REFERENCES

1. Rawlings, N. D., and Barrett, A. J. (1993) *Biochem. J.*, **290**, 205-218.
2. Barrett, A. J., and Rawlings, N. D. (1998) in *Handbook of Enzymes* (Woessner, J. F., ed.) Academic Press, London, pp. 2-16.
3. Trapani, A. G. (2001) *Genome Biol.*, **2**, 3014.1-3014.7.
4. Zamolodchikova, T. S., Sokolova, E. A., and Smirnova, E. V. (2003) *Biochemistry (Moscow)*, **68**, 309-316.
5. Wouters, M. A., Liu, K., Riek, P., and Husain, A. (2003) *Mol. Cell*, **12**, 343-354.
6. Zamolodchikova, T. S., Vorotyntseva, T. I., Nazimov, I. V., and Grishina, G. A. (1995) *Eur. J. Biochem.*, **227**, 873-879.
7. Zamolodchikova, T. S., Sokolova, E. A., Alexandrov, S. L., Mikhaleva, I. I., Prudchenko, I. A., Morozov, I. A., Kononenko, N. V., Mirgorodskaya, O. A., Da, U., Larionova, N. I., Pozdnev, V. F., Ghosh, D., Duax, W. L., and Vorotyntseva, T. I. (1997) *Eur. J. Biochem.*, **249**, 612-621.
8. Pemberton, A. D., Zamolodchikova, T. S., Scudamore, C. L., Chilvers, E. R., Miller, H. R., and Walker, T. R. (2002) *Eur. J. Biochem.*, **269**, 1171-1180.
9. McAleese, S. M., Pemberton, A. D., McGrath, M. E., Huntley, J. F., and Miller, H. R. P. (1998) *Biochem. J.*, **333**, 801-809.
10. Zamolodchikova, T. S., Sokolova, E. A., Lu, D., and Sadler, E. (2000) *FEBS Lett.*, **466**, 295-299.
11. Shechter, I. V., and Berger, A. C. (1967) *Biochem. Biophys. Res. Commun.*, **27**, 157-162.
12. Pletnev, V. Z., Zamolodchikova, T. S., Pangborn, W. A., and Duax, W. L. (2000) *Prot. Struct. Funct. Genet.*, **41**, 8-16.
13. Chomezynski, P., and Sacchi, N. (1987) *Analyt. Biochem.*, **162**, 156-159.
14. Guex, N., and Peitsch, M. C. (1997) *Electrophoresis*, **18**, 2714-2723.
15. Ruzheinikov, S. N., Popov, M. E., and Kashparov, I. V. (2002) *Bioorg. Khim.*, **28**, 28-37.
16. Marelius, J., Kolmodin, K., Feierberg, I., and Åqvist, J. Q. (1999) *J. Mol. Graph. Model.*, **16**, 213-225.
17. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) *J. Comp. Phyl.*, **23**, 327-341.
18. Lee, F. S., and Warshel, A. (1992) *J. Chem. Phys.*, **97**, 3100-3107.
19. Berendsen, H. J. C., Postma, J. P. M., Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984) *J. Chem. Phys.*, **81**, 3684-3690.
20. Humphrey, W., Dalke, A., and Schulten, K. (1996) *J. Mol. Graph.*, **14**, 33-38.
21. Salvesen, G., and Enghild, J. J. (1990) *Biochemistry*, **29**, 5304-5308.
22. Hof, P., Mayr, I., Huber, R., Korzus, E., Potemba, J., Travis, J., Powers, J. C., and Bode, W. (1996) *EMBO J.*, **15**, 5481-5491.
23. Pereira, P. J. B., Wang, Z. M., Rubin, H., Huber, R., Bode, W., Schechter, N. M., and Strobl, S. (1999) *J. Mol. Biol.*, **286**, 163-173.
24. Perona, J. J., Hedstrom, L., Rutter, W. J., and Fletterick, R. J. (1995) *Biochemistry*, **34**, 1489-1499.
25. Waugh, S. M., Harris, J. L. Fletterick, R., and Craik, C. S. (2000) *Nat. Struct. Biol.*, **7**, 762-765.
26. Page, R. D. M., and Holmes, E. C. (2000) in *Molecular Evolution: a Phylogenetic Approach*, Blackwell Science, London, pp. 135-171.
27. Rypniewski, W. R., Perrakis, A., Vorgias, C. E., and Wilson, K. S. (1994) *Protein. Eng.*, **7**, 57-64.
28. Kam, C.-M., Hudig, D., and Powers, J. C. (2000) *Biochim. Biophys. Acta*, **1477**, 307-323.
29. Ewoldt, G. R., Smyth, M. J., Darcy, P. K., Harris, H. L., Craik, C. S., Horowitz, B., Woodard, S. L., Powers, J. C., and Hudig, D. (1997) *J. Immunol.*, **158**, 4574-4583.

30. Grigorenko, V. G., Yarovoi, S. V., Paulauskaite, R., and Amerik, A. Yu. (1994) *FEBS Lett.*, **342**, 278-280.
31. Edwards, K. M., Kam, Ch.-M., Powers, J. C., and Trapani, J. A. (1999) *J. Biol. Chem.*, **274**, 30468-30473.
32. Page, R. D. M., and Holmes, E. C. (2000) in *Molecular Evolution: a Phylogenetic Approach*, Blackwell Science, London, pp. 11-36.
33. Polanovska, J., Krokoszynska, I., Czapinska, H., Watorek, W., Dadlez, M., and Otlewski, J. (1998) *Biochim. Biophys. Acta*, **1386**, 189-198.
34. Craik, Ch. S., Roczniak, S., Sprang, S., Fletterick, R., and Rutter, W. (1987) *J. Cell. Biol.*, **33**, 199-211.
35. Smith, M. J., O'Connor, M. D., Trapani, J. A., Kershaw, M. H., and Brinkworth, R. I. (1996) *J. Immunol.*, **156**, 4174-4181.
36. Bode, W., Meyer, E., and Powers, J. C. (1989) *Biochemistry*, **28**, 1951-1963.
37. Kunori, Y., Koizumi, M., Masegi, T., Kasai, H., Kawabata, H., Yamazaki, Y., and Fukamizu, A. (2002) *Eur. J. Biochem.*, **269**, 5921-5930.
38. Perona, J. J., Tsu, C. A., McGrath, M. E., Craik, C. S., and Fletterick, R. J. (1993) *J. Mol. Biol.*, **230**, 934-939.
39. Remington, S. J., Woodbury, R. G., Reynolds, R. A., Matthews, B. W., and Neurath, H. (1988) *Biochemistry*, **27**, 8097-8105.
40. Pasternak, A., Ringe, D., and Hedstrom, L. (1999) *Protein Sci.*, **8**, 253-258.
41. Frigerio, F., Coda, A., Pugliese, L., Lionetti, C., Menagatti, E., Amiconi, G., Schnebli, H. P., Ascenzi, P., and Bolognesi, M. (1992) *J. Mol. Biol.*, **225**, 107-123.