THESIS OF PHD DISSERTATION

Studies on the substrate specificity of retroviral proteinases and caspases with the methods of enzimology, molecular biology and molecular modelling

Gábor Zahuczky

Medical and Health Science Center
University of Debrecen
Debrecen, 2003

1. INTRODUCTION

Proteases are essential to physiologic processes such as inflammation, infection, fertilization, allergic reactions, cell growth and death, blood clotting, tumor growth and bone remodeling. Proteases, as target proteins are also therapeutically important, since a large number of molecules are able to inhibit or attenuate the undesirable action of proteases.

Various proteinase inhibitors are now used in antiretroviral therapy of AIDS. However, as in the case of therapeutical use of reverse transcriptase inhibitors, resistance rapidly develops against these inhibitors, both *in vitro* and *in vivo*. Most of the HIV PR inhibitors including those in clinical use were specifically designed against the wild-type HIV-1 PR. Residues which are conferring resistance of HIV-1 PR against inhibitors frequently can be found in equivalent positions of other wild-type PRs, as was the case with HTLV-I and BLV PR. Comparative studies of various PRs are expected to reveal the common features of their specificity. Design and use of inhibitors which are efficient against different retroviral proteinases may reduce the possibility of selection for viable mutants. Specificity studies can help us in this design.

Phosphorylation is one of the most fundamental regulatory processes in the cells. The level of phosphorylation of proteinase substrates can substantially influence their cleavability. In order to determine whether the phosphorylation of a given protein on a determined residue have a direct regulatory function, specificity studies are also useful tools.

Retroviral proteinases (PR*)

The research of retroviruses excited a great interest due to the appearance and rapid spread of the acquired immunodeficiency syndrome (AIDS). The existence of retroviruses was already known even in the beginning of the last century, however until the discovery of human T-cell leukemia virus (HTLV) and human immunodeficiency virus (HIV) their capability to infect human was not obvious. There are examples for infections of all vertebrates now and they can have many different outcomes: viremia without illness, tumor formation, alterations in nervous system, anemia and immunodeficiency.

The first role of retroviral proteinases in the viral life cycle discovered was the cleavage of Gag and GagPol precursor proteins into functional structural proteins and enzymes. Additionally, cleavage of nucleocapsid protein was observed within the viral

* The nomenclature of retroviral proteins is according to Leis at al. (1988). Proteins with unknown function are designated with p following a number expressing its molecular weight.

capsids in the early phase of virus infection. Later it turned out that many cellular proteins are also substrates of HIV proteinases (vimentin, MAP2, NF-κB, etc.) which can contribute to the pathogenicity of the virus.

Retroviral proteinases consists of 99-126 residues. Their molecular weight is 11-15 kDa and they are aspartic proteases, active in homodimer form. There is a close homology in the first and secondary structure between retroviral proteinases and one domain of cellular aspartic proteinases. They contain many β-sheets and one or two short α-helixes depending on the enzyme. The N- and C-terminal regions of the two monomers form a four layer β-sheet. There are three specific regions in the enzyme: region of the catalytic triad (-Asp-Thr/Ser-Gly-) which is characteristic of aspartic proteinases and found close to the N-terminal end; flap region, which is rather flexible bending to the substrate during the formation of the enzyme-substrate complex; dimerisation region, for which the N- and C-terminal sequences are responsible.

The residues found in the natural cleavage sites of retroviral proteinases are usually hydrophobic, however, a general consensus sequence cannot be given. The classification of cleavage sites is possible in the case of HIV-1 PR, where type 1 cleavage sites contain Asn-Tyr/Phe↓Pro residues, while type 2 cleavage sites are hydrophobic in P2-P2'[†] positions. However, this generalization does not seem to be appropriate for other retroviral proteinases.

Human T-cell leukemia virus (HTLV) and bovine leukemia virus (BLV) belong to the family of retroviruses, and together with simian T-cell lymphotropic viruses (STLV) to the subfamily of HTLV-BLV group. The characteristics of this group is substantially different from other retroviruses, however the members of the group are closely related, they share a common genome organization, presence of regulatory proteins Tax and Rex, and nucleotide sequence similarity. Thus the BLV enzyme may be considered as a relevant model for its human counterpart. So far crystal structure about the proteinases of this group is not available, but according to previous studies there may be a significant difference in the substrate specificity of proteinases between the representatives of HTLV-BLV group and lentiviruses (HIV-1, HIV-2, etc.).

[†] Nomenclature of substrate residues and substrate binding sites is according to Schechter and Berger (1967). Substrate residues are designated from the cleavage site to the N-terminus with P1, P2, P3, etc., and to the C-terminus with P1', P2', P3', etc. The appropriate substrate binding sites are designated with S1, S2, S3, etc. or S1', S2', S3', etc.

Caspases

Programmed cell death (apoptosis) is an essential mechanism to eliminate unwanted cells during the development and homeostasis of multicellular organisms. This extremely well organized process involves DNA fragmentation, membrane blebbing, cell shrinkage and disassembly into membrane-enclosed vesicles. These are eliminated by phagocytosis, therefore preventing an inflammatory response to the intracellular components. In this evolutionarily highly conserved mechanism caspases play a crucial role by taking part as initiators or executors in a cascade which occurs through proapoptotic signals and culminate in the cleavage of many cellular proteins. Two canonical pathways of caspase activation have been described so far. One begins with ligation of specialized cell-surface receptors termed death receptors, the other is caused by cellular stress and is carried out by the assistance of citochrome c release from mitochondrion.

Caspases carry their major characteristics in their name, namely caspase stands for cysteine-dependent aspartate specific protease. Since the identification of the first representative of caspases (caspase-1 or interleukin 1β converting enzyme (ICE)), 14 mammalian and 5 *Drosophila* caspases have been cloned. They are synthesized as latent precursor zymogene procaspases and they are activated by a highly regulated mechanism.

Structurally, all procaspases contain a highly homologous protease domain, the signature motif of this family of proteases. This domain can be further divided into two subunits, a large subunit of approximately 20 kDa (p20) and a small subunit of approximately 10 kDa (p10). For the activity of the enzyme both subunits are necessary. Each procaspase also contains a prodomain or NH_2-terminal peptide of variable length. These domains mediate homophylic interaction between procaspases and their adapters and play important roles in procaspase activation. So far the crystal structure of 5 caspases have been solved with their tetrapeptide aldehide inhibitors. These structures reveal that a mature caspase is a tetramer (homodimer of the p20 and p10 heterodimers arranged in twofold rotational symmetry), with the two adjacent small subunits surrounded by two large subunits. Each p20-p10 heterodimer forms a single globular domain, and the core of the globular domain is a six-stranded beta-sheet flanked on either side by alpha-helices. These two heterodimers associate with each other primarily through the interaction between the p10 subunits.

Caspases are highly specific endopeptidases, since their enzymatic properties are governed by a dominant specificity for substrates containing Asp in P1 position. However, there are known substrates now, which contain Glu in their cleavage site P1 position, such as tumor necrosis factor receptor-I (p60) protein cleaved by caspase-7 at a GELE motif, Max

protein containing IEVE cleavage site and DRONC *Drosophila* caspase is able to selfprocess itself at TQTE sequence. Caspases can be subdivided on the basis of their substrate specificity, where the major specificity determinant is the S4 subsite. Group 1 caspases (1, 4, 5, 13) are tolerant of liberal substitions in P4 but prefer bulky hydrophobic amino acids such as Tyr or Trp. The group 2 caspases (2, 3, 7) are substantially more stringent in S4, requiring a P4 Asp. Group 3 caspases (6, 8, 9, 10), on the other hand prefer branched chain aliphatic amino acids in P4.

To date, more than 60 proteins have been shown to be substrates of one or more caspases in mammalian cells, and the list is still growing. These substrate proteins contain one or very few caspase cleavage sites in interdomain linker sequences. Substrate proteins are thus not degraded by caspase processing; instead, caspase cleavage may activate or inactivate the substrate protein's functions. Among them there are executors of apoptosis, cell cycle regulator proteins, enzymes of the DNA metabolism, cytosceletal scaffold proteins, repair and housekeeping enzymes, signaling molecules, neurodegenerative disease proteins, transcription factors and many others.

Several caspase cleavage sites contain potential or verified Ser/Thr phosphorylation site. In the latter ones these residues have been proved to be phosphorylation sites in many cases. A recent study demonstrated that both Ser 327 and Ser 330 of presenilin 2 (PS-2) at the caspase cleavage sites are targets of phosphorylation; phosphorylation inhibited caspase mediated cleavage, and enhanced the antiapoptotic properties of the protein in cell culture. It has been reported that phosphorylation of IκB-α inhibited its cleavage by a caspase *in vitro* and the P1' Ser of the caspase cleavage site was reported to be a phosphorylation site by MEKK1, MST1 and Max, also suggesting that P1’ Ser phosphorylation may regulate caspase-mediated proteolysis. An important part of the $G_0 \rightarrow G_1$ transition of the cell cycle is the downregulation of the expression of a set of *gas* genes and the hyperphosphorylation of the Gas protein. Gas2 is a protein of the microfilament system and is also involved in the regulation of apoptosis. Gas2 was found to be a substrate of caspase-3. Furthermore, its function is regulated by phosphorylation at serine residues. The cleavage site in Gas2 contains a P4 Ser instead of Asp, in a potential phosphorylation site. A serine residue in position P4 is also present in the caspase-3 substrate SREBP-1, and PAK-2. Phosphorylation at other sites like P2 and P2' appears to be less frequent, but may also protect from caspase-mediated cleavages, as found for the cleavage of Bid by caspase-8.

2. OBJECTIVES

HTLV-I and BLV proteinases are highly similar in their sequence and specificity, therefore BLV PR appeared to be a promising model for its human counterpart. Since crystal structure of the proteinases of the HTLV-BLV group is not available and the specificity of BLV PR was only compared to that of HIV-1 and HIV-2 proteinases we aimed at cloning and purification of BLV PR than comparing its specificity to that of HTLV-I PR using oligopeptides representing the natural cleavage sites of HTLV-I and BLV proteinases.

The HTLV-I PR has been partially characterized. However, its detailed specificity comparison to that of HIV-1 PR has not been described yet, therefore we tried to make such a specificity comparison, by using a series of oligopeptide substrates based on the HIV-1 PR MA/CA type 1 natural cleavage site, that has already been used to characterize representative PRs of various retrovirus subgroups. Since most of the members of this series proved to be not cleavable by HTLV-I PR, we performed a comparison using another series of oligopeptides based on the HTLV-I PR CA/NC natural cleavage site. Furthermore, we intended to build the model of HTLV-I PR for interpreting our results with the help of molecular modeling tools.

The proteolysis and protein phosphorylation as posttranslational modifications are usually studied separately: only a few studies have demonstrated to date, that these two crucial processes may act in a concerted way. For the identification of residues that can be phosphorylated and for the elucidating of the effect of phosphorylation usually site-specific mutagenesis is used, mutating the candidate Ser to Ala or Gly in the first case and to Asp in the second case. Recently our study proved that phosphorylated oligopeptides might serve as excellent tools to study the effect of phosphorylation on proteolysis, therefore we planned to study the effect of P4, P3 and P1' Ser phosphorylation of caspase substrates on proteolysis. We used PS-2 and Gas-2 caspase substrate cleavage site mimicking oligopeptides and their phosphorylated counterparts in our studies. Furthermore, we intended to characterize the role of P1 Asp in the cleavability using a good substrate (PARP) analogue, and P1 Ser or phospho-Ser mutated counterpart peptides.

3. MATERIALS AND METHODS

ENZYMES: The proteinase coding region of a cDNA BLV clone was amplified by polymerase chain reaction (PCR) by using a 5' primer which contains the nucleotide sequence located upstream of the start of the proteinase coding region flanked with an EcoRI site, and two kinds of 3' primers, containing the C-terminal sequence of the PR or the pro reading frame, two stop codons and a SalI restriction site. The PCR yielded DNA fragments having the expected size. The amplified DNA fragments were cloned into pMal–c2 vector (New England Biolabs), after the maltose binding protein (MBP) gene. Ligation and transformation of Subcloning Efficiency DH5 cells (Gibco-BRL) were performed by using standard protocols. Both constructs contained an eight-residue N-terminal viral flanking sequence. DH5-α cells bearing the constructs for expression were grown in Luria–Bertani medium in the presence of 100 g/ml ampicillin. When the cell suspension reached an absorbance of 0.6–0.8 at 600 nm, protein expression was induced by the addition of 1 mM IPTG for 2 h. To verify expression, cells were collected by centrifugation, and disrupted directly in sodium dodecyl sulfate (SDS)–polyacrylamide gel sample loading buffer. For large-scale purification cells were harvested by centrifugation and suspended in lysis buffer (50 mM Tris–HCl (pH 8.0), 1 mM EDTA and 100 mM NaCl, 1 mM PMSF). Cells were lysed with the lysozyme-DOC method. The inclusion bodies were washed with lysis buffer containing 0.5% Triton X-100, and the pellet was dissolved in 100 mM Tris–HCl (pH 8.0) containing 6 M guanidine–HCl on ice. The sample was dialyzed stepwise against 100 mM Tris–HCl (pH 8.0) containing 2 M guanidine–HCl, then against buffer A (20 mM Tris (pH 7.2) containing 100 mM NaCl, 1 mM EDTA, 5% glycerol, 0.1% Triton X-100) at 4°C. Finally, the solution was clarified by centrifugation and filtered through a 0.22-m membrane. The clarified solution was applied to an Econo-Pak High S column (Bio-Rad) and proteins were eluted with a linear gradient of buffer A and buffer A+1 M NaCl. Fractions with the highest activity were concentrated with Amicon-10, and applied to a Superdex G-75 gel filtration column (Pharmacia), equilibrated with buffer A. Fractions with the highest activity were pooled and used for the kinetic measurements. SDS–polyacrylamide gel electrophoresis was performed according to Laemmli, and immunoblotting was performed according to Towbin et al. Rainbow molecular mass markers (Amersham) were used for comparison.

The coding region of recombinant HTLV-I PR was cloned into pET-23b plasmid. The expression and purification of the enzyme was according to a recently publicated method.

The human recombinant caspase-3, -7 and -8 enzymes were purchased from Biomol Research Laboratories (Plymouth Meeting, PA, USA) or expressed and purified according to a recently published method. The clones were a kind gift from Dr. Guy S. Salvesen (Burnham Institute, La Jolla, California). Enzyme preparations were active site titrated by using acetyl-Asp-Glu-Val-Asp-CHO (Calbiochem), a potent inhibitor of these caspases, as described.

Mutant caspase-3 enzymes were produced by using QuickChangeTM site-specific mutagenesis kit (Stratagene) and pET-23b template plasmid containing caspase-3 proteinase nucleotide sequence. Expression and purification were carried out as described above.

OLIGOPEPTIDES: Oligopeptides were synthesized by solid-phase peptide synthesis on a Model 430A automated peptide synthesizer (Applied Biosystems) or on the Vega Coupler 250C using Boc chemistry, and were purified by reversed-phase high-performance liquid chromatography. Stock solutions and dilutions were made in distilled water (or in 5 mM dithiothreitol for the Cys-containing peptide), and the proper peptide concentration was determined by amino acid analysis with a Beckman 6300 amino acid analyzer. Peptides were obtained from Dr. Stephen Oroszlan or from Dr. Terry D. Copeland (Molecular Virology and Carcinogenesis Laboratory, NCI-FCRDC, Frederick, MD, USA). Synthesis of phosphorylated peptides was carried out as described previously.

PROTEINASE ASSAYS: The retroviral PR assays were performed in the following solutions: 0.01-4 mM substrate 1-40 nM retroviral PR preparation, 0.25 M potassium phosphate buffer, pH 5.6, containing 5% glycerol, 1 mM EDTA, 5 mM dithiothreitol, 2 M NaCl. The caspase assays were performed in 50 mM HEPES, pH 7.4, containing 100 mM NaCl, 0.1 % CHAPS, 1 mM EDTA, 10 mM DTT and 10 % glycerol at 0.01 – 6.0 mM final substrate concentration. The range of substrate concentration was selected depending on the approximate K_M values. The 20 μl final volume reaction mixture was incubated at 37°C for 1 h and terminated by the addition of 180 μl 1% trifluoroacetic acid (TFA), and an aliquot was injected onto a Nova-Pak C_{18} reversed-phase chromatography column (3.9×150 mm, Waters Associates) using an automatic injector. Substrates and the cleavage products were separated using an increasing water–acetonitrile gradient (0–100%) in the presence of 0.05% TFA. Cleavage products were identified by amino acid analysis and/or peptide sequencing. Kinetic parameters were determined by fitting the data obtained at less than 20% substrate hydrolysis to the Michaelis–Menten equation by using the Fig. P program (Fig. P Software Corp.). The error of the k_{cat}/K_M values determined in this way was less than 20%. For some peptides the

k_{cat}/K_M values were determined from the linear part of the rate *versus* concentration profile. For peptides, where it was not possible to determine the K_M value from the Michaelis–Menten curve, due to enzyme saturation even et very low substrate concentration, k_{cat}/K_M values were determined using competition assays and calculated from the values of substrates having known k_{cat}/K_M values.

MOLECULAR MODELING: The crystal structure of the Rous sarcoma virus (RSV) S9 mutant with the HIV-1 CA/p2 analog reduced peptide inhibitor Arg-Val-Leu-r-Phe-Glu-Ala-Nle was used as the starting model, as it showed all the flap residues. The model was built with two deletions: a five residue deletion between RSV PR residues 21 and 22 in the turn between beta strands b and c, and a three residue deletion at the end of beta strand a' in the flap. Because of conformational differences in the two flaps of the RSV S9 PR, the two subunits of HTLV-1 PR were modeled with a slightly different position of the deletion in the flaps. The dimer was modeled with the HTLV-1 PR peptide substrate Lys-Val-Leu*Val-Val-Gln-Pro and the conserved water molecule between the flaps and the inhibitor. The amino-acid residues of RSV S9 PR were replaced by those of HTLV-1 PR. The positions of all new atoms were generated and minimized with the program AMMP. The sp4 potential set was used. The peptide bond included additional distance restraints. The positions were generated for all new atoms using 10 cycles of the Kohonen algorithm, which is based on self-organizing or Kohonen networks as implemented for proteins, followed by 20 steps of Gauss–Seidel distance geometry minimization, as described in Harrison *et al*. The geometrical terms for the new atoms were minimized with 200 steps of conjugate gradients, and then the nonbonded and geometrical terms were minimized for a further 400 steps, including a short run of molecular dynamics. Then, the complex was minimized using 200 steps of conjugate gradients. Finally, the P4 Thr was added using the above procedure and adjusted manually to resemble the P4' residue present in the crystal structure, and the whole complex was minimized further by using 200 steps of conjugate gradients. All computations were performed on a 233-MHz DEC AlphaStation. The new crystal structure of HIV-1 PR K45I mutant with the HIV CA/p2 analog was used for comparison, as it was refined at the high resolution of 1.55 Å.

Caspase-3 with a model of the PS-2 substrate (DSYDS) was built from a human caspase-3-inhibitor crystal structure by changing residues of the inhibitor to residues of substrate using the Sybyl 6.7 software package (Tripos Inc., St. Louis, MO, USA). Then each of the side chain torsion angles for substituted residues in the peptide substrate was

rotated through 360 ° in steps of 15 ° to search for alternate conformations using Kollman all atom force field and 8-angstrom cutoff. Twenty Powell iterations were applied for the whole enzyme-substrate complex after finding the lowest energy conformers of the modified residues. The RMS deviation between the Cα atoms of the final enzyme model and the Cα atoms of the initial crystal structure was 0.13 angstrom.

The structures were built, minimized and examined on Silicon Graphics Indigo2 or O2 computer graphics systems. For structural comparison a caspase-7-inhibitor structure and a caspase-8-inhibitor structure were used.

4. RESULTS AND DISCUSSION

Cloning of BLV PR and comparison of its specificity to that of human T-cell leukemia virus proteinase

The proteinase coding region of BLV was cloned into pMal-c2 vector in fusion with the *malE* gene of *E. coli*. Two constructs were made, both having an eight-residue N-terminal flanking region; however, pMal–Blvpr clones contained stop codons at the end of the PR, while pMal–Blvprpro clones contained the entire coding sequence of the post-proteinase region of the *pro* gene. After expression we performed immunoblot analysis, using polyclonal rabbit anti-BLV PR antibody, of total cell lysates from duplicate clones before and after IPTG induction. Instead of the appearance of the fusion proteins with the expected size of 58 and 60 kDa, a strong 14 kDa protein appeared with both types of clones. The size of the appearing band is in good agreement with the size of the PR purified from virion, suggesting that not only the N-terminal processing, but also the C-terminal processing was very efficient in vivo, during the IPTG expression.

The majority of the BLV proteinase appeared in inclusion bodies after induction, as verified by SDS–polyacrylamide gel electrophoresis and immunoblotting. To purify the enzyme, cation-exchange chromatography followed by gel filtration was applied. By using these steps, apparently pure enzyme was obtained. The final concentration of the purified enzyme was 2 μM. Further concentration of the proteinase as well as longer storage at -20°C lead to partial self processing resulting in a closely spaced doublet in SDS–polyacrylamide gel electrophoresis, predicted to be at the C-terminus between residues 116 and 117 as previously described for another clone of BLV PR.

Oligopeptide substrates representing naturally occurring cleavage sites in BLV as well as HTLV-1 were used to compare the specificity of BLV and HTLV-1 PRs. Except one BLV peptide which was hydrolyzed only by the BLV PR, all other studied substrates were cleaved by both enzymes, although in some cases with substantially different rates. A common characteristics of the BLV PR is that the k_{cat} values appear to be generally substantially lower than those obtained for the HTLV-1 enzyme; however, due to the very low K_M values, the specificity constant k_{cat}/K_M values are in a similar range for the two enzymes. Due to the very low K_M for the BLV PR, it was not possible to determine the k_{cat}/K_M values for most of the substrates from the Michaelis–Menten curve. For these peptides competition assays were performed with another substrate of known k_{cat}/K_M value, as previously done for other low K_M substrates of HIV proteinases. Several inhibitors of HIV-1 PR were assayed as inhibitors of the BLV PR, as previously was performed for the HTLV-1 PR. None of these inhibited the BLV PR, while Compound 3 has previously been shown to be a fairly good inhibitor of HTLV-1 PR. Inhibitors of MPMV and AMV proteinases also did not inhibit the BLV PR. However, two statine-based inhibitors, which were based on HTLV-1 cleavage sites inhibited the BLV PR, even though one of them was inactive against the HTLV-1 PR.

Comparison of the substrate specificity of human T-cell leukemia virus and human immunodeficiency virus proteinases

First, we attempted to characterize the specificity of the HTLV-I PR with a series of oligopeptides based on the MA/CA cleavage site of HIV-1 PR (H-Val-Ser-Gln-Asn-Tyr↓Pro-Ile-Val-Gln-NH2). However, most of these peptides were not cleavable by HTLV-I PR with the exception of peptides containing hydrophobic P2 residues or Leu, Val, Ile, Met, Phe, and Thr in P4 position.

As most of the peptides of the HIV-1 MA/CA series were not cleaved, or were hydrolyzed only inefficiently by the HTLV-1 PR, a new peptide series based on the CA/NC cleavage site of HTLV-1 PR (Lys-Thr-Lys-Val-Leu↓Val-Val-Gln-Pro-Lys) was used to compare the specificity of the two enzymes. Peptides with N-terminally shortened sequences as well as single amino-acid substitutions in the P4-P1' positions of this starting sequence were assayed as substrates of the HTLV-1 and HIV-1 PRs. Generally, most of these peptides were hydrolized by HTLV-I PR with higher catalytic rate then with HIV-1 PR, but in the case of K_M values we did not observe a similar trend.

While the decapeptide substrate was well hydrolyzed by both enzymes, the N-terminally shortened nonapeptide was hydrolyzed with smaller catalytic constants by the HTLV-1 PR. The octapeptide with P3-P5' residues was cleaved only poorly by the HTLV-1 PR, but it was still an efficient substrate of the HIV-1 PR. These results suggested that HTLV-I PR has a more extended substrate binding site than that of the HIV-1 PR.

All P4-substituted peptides formed fairly good substrates of the HTLV-1 PR, except for the peptide with P4 Asp. The highest k_{cat}/K_M values were obtained with the unsubstituted peptide having Thr in P4 and the peptides with Val or Leu at P4. For the HIV-1 PR the best substrates had Ser and Thr at this position, whereas only the peptide with Gly at P4 was a relatively inefficient substrate.

Similar to the P4 substituted peptides, all P3 substituted substrates were hydrolyzed by both PRs, but while the original P3 Lys containing peptide was the best substrate of HTLV-1 PR, HIV-1 PR preferred Phe and Leu residues in this position. Interestingly, peptides with negatively charged Asp and without a P3 side-chain (Gly) provided the least efficient substrates for both PRs.

The substitution of the original P2 Val to Lys provided for both enzymes uncleavable substrate. For HTLV-I PR Asn, Asp, Ser and Gly were also not tolerated at P2 and P2 Gly analogue was also hydrolyzed very inefficiently. For HIV-1 PR P2 Gly, Leu, Ser and Asp substitution provided a similar effect.

The best substrate for both enzymes had Phe at P1. Another common feature of the two enzymes is that exchange of Phe to Tyr or Met provided also a good substrate. The most remarkable difference in the specificity of the two enzymes was found in the P1 Ala substituted peptide, which was hardly cleaved by HIV-1 PR, meanwhile it was a fairly good substrate of HTLV-I PR.

P1’ Gly, Ser, Asp, Lys and Pro substituted peptides were rather inefficient substrates for both enzymes. Interestingly, the S1' subsite of HIV-1 PR appears to be more flexible than that of HTLV-I PR: substitution of Val to Ile or Phe caused more substantial changes in the kinetic constants for HTLV-I than for HIV-1 PR.

In the absence of a crystal structure, a model was built of the HTLV-I PR dimer with its substrate, in order to understand the molecular basis for the specificity. The amino-acid sequence of HTLV-1 PR was aligned with the sequences of other retroviral PRs of known structure, to determine the best starting structure for building the model. The length and sequence of the three PRs are different: HTLV-1 PR is the largest (125 residues), RSV S9 PR has 124 residues, and HIV-1 PR has 99 residues. Deletions and insertions in HTLV-1 PR were

positioned in surface loops in the PR structures. Models built by homology are most accurate for regions of very similar amino-acid sequence, and least reliable for those with insertions or deletions in the sequence alignment.

Like the peptidic inhibitors, the peptide substrate is expected to bind to the enzyme in an extended β-like conformation (Fig. 1.). The HTLV-1 PR is predicted to be similar to the other retroviral PRs in having at least seven substrate binding subsites. The respective primed and nonprimed subsites are more or less symmetric in the two subunits of the PR dimer, although the symmetry is distorted due to the asymmetric hydrogen bonding pattern.

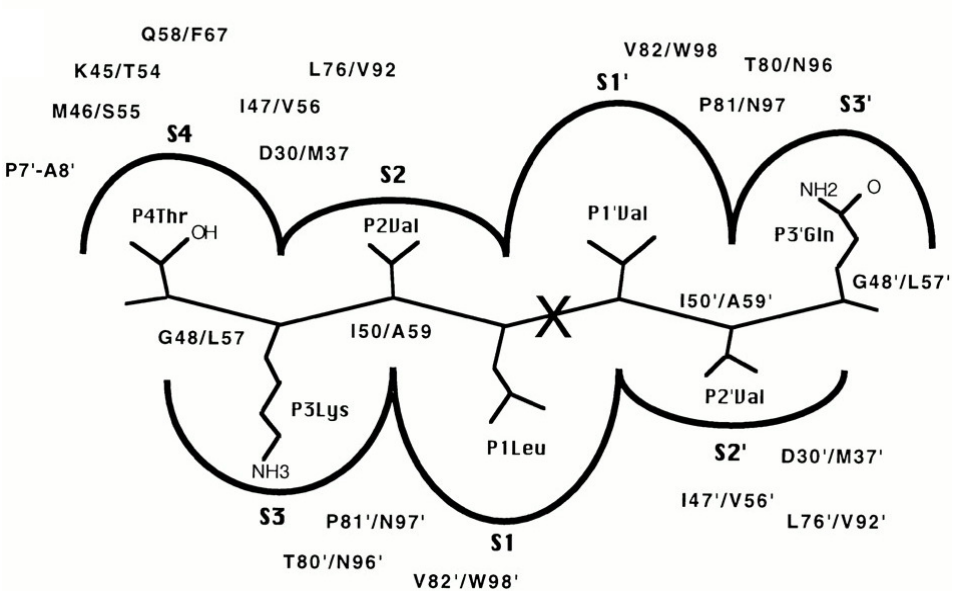

Fig. 1. Schematic representation of the HTLV-1 CA/NC substrate (KTKVL↓VVQPK) in the S4-S3' subsites of PR (A). The relative size of each subsite is indicated approximately by the area enclosed by the curved line around each substrate side-chain. PR residues forming the subsites are shown only for those that differ between the two PRs as HIV-1/HTLV-1 residues.

The S4 subsite is close to the protein surface and partly exposed to solvent for all retroviral PRs. Generally, it would be expected that this subsite would be formed of more polar amino acids compared to the less exposed internal subsites, and therefore, polar residues would be preferred at P4. In contrast, the preference for hydrophobic P4 residues obtained for the HTLV-1 PR with both substrate sets suggested that the S4 subsite is more hydrophobic. The residues at subsite S4 that are predicted to contribute to specificity differences between HIV-1 and HTLV-1 PRs are Asp30/Met37, Ile47/Val56, and Leu76/Val92. In HIV-1 the residue corresponding to Met37 is Asp30. Asp30 is a unique residue in this position, and can be found only in primate immunodeficiency virus proteinases. The presence of an insertion and hydrophobic instead of polar residues at S4 is consistent with the observed preference of

HTLV-1 PR for extended substrates with more hydrophobic residues such as Leu or Val in P4, compared to the preference of HIV-1 PR for P4 Ser or Thr.

The PR crystal structures show that the S3 subsite is usually large, and it is similar to S4 in being partly exposed to solvent at the surface of the enzyme. Consequently, the P3 side-chain may be positioned either to interact with the more polar residues of the PR surface, or to interact with the hydrophobic internal residues of the enzyme, as seen in many HIV-1-PR inhibitor crystal structures. Based on the molecular model, most of the residues participating in the S3-P3 interactions are identical in both in HIV-1 and HTLV-1 enzyme-substrate complexes. The most significant difference is Gly48/Leu57 substitution which may account for the less affinity of HTLV-I PR for larger hydrophobic residues.

The common characteristics of all the S2 subsites of the retroviral proteinases is that they are relatively small and hydrophobic, compared with the S4 and S3 subsites. However, it is evident from the results obtained with the set of substrates based on the HIV-1 MA/CA cleavage site that HIV-1 PR prefers small, polar residues like Asn and Cys, whereas the HTLV-1 PR prefers larger, hydrophobic ones. The major differences in residues, Ile47/Val56, Ile50/Ala59 result in a somewhat larger, more hydrophobic S2 subsite in HTLV-1 PR, as compared to HIV-1 PR, which is consistent with the observed selection for Val, Ile, Leu or Phe at P2 in HTLV-1 PR, as compared to a preference for Ile, Val or Ala in HIV-1 PR in the HTLV-1 CA/NC peptide series.

The S1 and S1’ subsites appear to be conserved (like the S3 subsite) of the retroviral PRs, and several of the amino-acid residues participating in this subsite are conserved in HIV-1 and HTLV-1 PRs (Gly34, Arg10, Ala59, Gly58, Leu30 HTLV-I numbering). They compose a relatively large and hydrophobic, internal pocket in the enzyme. The catalytic Asp residues also contribute to these subsites. Phe at P1 was better then P1 Tyr for both enzymes. Phe was found to be optimal for HIV-1 also in the HIV-1 MA/CA peptide series. However, P1 Met and P1 Ala substitutions result in much less decrease in the specificity constant for HTLV-I PR then for HIV-1 PR. These results suggest a higher flexibility of S1 subsite for HTLV-I then for HIV-1 PR. According to the model in the place of Pro81 in HIV-1 PR there is a more flexible Asn in HTLV-I PR and due to another substitution (Val82/Trp98) there is a Trp in the region which as a surface residue seems to be rotating since it points inward in S1 subsite and outward in S1’.

S1' subsite appears to be somewhat smaller than the S1 subsite due to the asymmetry introduced by the peptide bond. Interestingly, the S1' subsite of HIV-1 PR appears to be more flexible than that of HTLV-1 PR: substitution of Val to Ile or Phe caused more substantial

changes in the kinetic constants for HTLV-1 than for HIV-1. Similar to the findings for the S1 pocket, peptides with hydrophilic side-chains did not provide efficient substrates, despite that in HTLV-I PR there are two asparagines in the subsite which appears to be not enough for forming more specific hydrophilic interactions.

Effect of caspase cleavage site phosphorylation on proteolysis

To provide a direct evidence for the effect of phosphorylation at the PS-2 sites on the susceptibility towards proteolysis, oligopeptides with or without phosphorylated Ser residues, representing these cleavage sites in PS-2, were assayed as substrates of caspases. Although the nonphosphorylated peptide was properly hydrolyzed by the caspase 3 and 7 enzymes, this substrate was not cleaved at the other Asp residue, which is the minor site determined in the protein. The only two residues at the N-terminal part of this peptide may not provide a sufficient length for cleavage and the inherent cleavage rate is also substantially lower at this site of the protein. It should be noted, that the predicted K_M value for this substrate was substantially higher than the K_M for the peptide representing the PARP cleavage site.

We have also tested P3 Ser- and P1' Ser- phosphorylated versions of this peptide, but while the P3 Ser-phosphorylated peptide was cleaved with a very similar specificity constant by both enzymes, the P1' Ser-phosphorylated was not cleaved by the proteinases. Interestingly, in both cases increase of the concentration of the P3-phosphorylated peptide resulted in a decrease of the cleavage rate, that may be due to an alternative, nonproductive binding of the phosphorylated peptide.

Caspase-8 did not cleave any of these peptides. Since caspase 8 prefers branched chain aliphatic amino acids, especially Leu at P4 instead of Asp, we have also tested the P4 Leu substituted analogs of these peptides to test the sensitivity of caspase 8 for cleavage site phosphorylation at P3 and P1' positions. Although the specificity constants for the hydrolyzed peptides were very low, the results were in good agreement with the findings for the two other caspases: the enzyme was sensitive to P1' phosphorylation but not for P3 phosphorylation.

It should be noted that the specificity constant for the unmodified PS-2 substrate was much higher for the caspase 3 than for caspase 7, which may be at least partly due to its much better tolerance for P2 Tyr. Furthermore, caspase 3 is a much more efficient enzyme on small synthetic substrates, as compared to caspase 7, which may also substantially contribute to the observed effects.

The molecular model of the PS-2 substrate binding to the caspase 3 was built on the basis of an enzyme-inhibitor crystal structure. Side chains of P3 and P2 residues point more or less outward, explaining the rather tolerant nature of the respective subsites, although preference for Glu at P3, Val, Leu, Pro and Thr at P2 was demonstrated for caspases. Less efficient hydrolysis of substrates with large P1' side chains could be the result of steric hindrance between P1' side chain and the side chains of Thr 166, Tyr 204 of caspase-3 and the usually hydrophobic P2 side chains. Negatively charged amino acids were also found to be very unfavorable at this position, due the electrostatic repulsion of Glu 123 in caspase-3. Caspase-7 also contains Glu while caspase-8 contains Asp in the equivalent position. Phosphorylation of P1' Ser increased the size and added negative charges, and the accumulation of these unfavorable changes could lead to an uncleavable peptide.

Effect of P4 Ser phosphorylation of a caspase substrate on its susceptibility toward cleavage was studied with oligopeptides representing Gas-2 caspase substrate cleavage site. Although the consensus sequence for caspase 3 cleavage sites is Asp-Xaa-Yaa-Asp-↓-, the cleavage site in Gas2 contains a P4 Ser instead of Asp, in a potential phosphorylation site. While the nonphosphorylated peptide was an acceptable substrate of the caspase 3 and 7, phosphorylation of the P4 Ser residue decreased the specificity constant, suggesting that unlike Asp, the also negatively charged phospho-Ser is very unfavorable in this position, even more unfavorable than the noncharged Ser residue. When the P4 residue was substituted with Asp, much more favorable kinetics were obtained. The values of catalytic constant (k_{cat}) and specificity constant (k_{cat}/K_m) for this peptide altered with a similar trend for both enzyme, but this change was more expressed in the case of caspase 3. None of these peptides were cleaved by caspase-8.

Based on the enzyme-inhibitor structures, the P4 Asp forms strong H-bond interactions with the side chain of Trp 214 and the amide N of Phe 250 and weaker interaction is also possible with the side chain of Asn 208. These interactions appear to anchor the substrate and they cannot be formed with smaller (Ser) or bigger (pSer) residues. In a detailed specificity study of caspase-3, the P4 Asp preference was found to be almost absolute, peptides containing charge conserving (Glu) and isosteric (Asn) substitutions were only poor substrates [9]. Our results suggest that a phospho-Ser residue is also very unfavorable by caspase-3 at this position.

Since all caspases cleave substrates at the negatively charged Asp residue, we tested, whether substitution of P1 Asp in a consensus caspase 3 cleavage site sequence representing the cleavage site in PARP could provide a substrate for the enzyme. The unmodified substrate

was efficiently hydrolyzed by caspase 3 and 7, while it was a poor substrate for caspase 8, although P4 Leu substitution in this sequence somewhat improved the specificity constant. The peptides with P1 phospho-Ser residue were not hydrolyzed by the caspases, similar to the peptides with P1 Ser. The P1 Asp residue of the substrate is anchored by interactions with the side chain guanidino groups of Arg 64 and Arg 207, as well as with the side chain of Gln 161 in caspase 3. These residues are conserved in other caspases, including caspases 7 and 8. Due to this complex H-bond/salt bridge network, neither smaller nor larger charged residues than Asp can form similarly strong interactions. Our mutagenesis studies also confirm the essentiality of these residues, since in caspase-3 enzyme mutating any of them to a small and uncharged residue, resulted in mutant procaspase-3 enzymes which were unable for self-processing during expression.

5. SUMMARY

In our study we managed to produce the main representatives of the HTLV-BLV group of retroviruses and compare their aspartic proteinases' specificity with each other or the specificity of HTLV-I PR with that of HIV-1 PR. We also studied caspases, a very important family of cysteinyl proteinases, in the relation of phosphorylation and cleavability. We attempted to elucidate the interactions between the representatives of these enzyme-families and their substrates, in order to realize universal relations concerning these interactions.

In order to produce BLV proteinase the proteinase cDNA was cloned into pMal–c2 vector, from that we expressed and purified it. During purification, it turned out that the enzyme is able to self-process itself both from N- and C-terminal direction. The majority of the proteinase formed inclusion bodies, therefore it was selected as a source of purification of the enzyme by a cation-exchange chromatography step, followed by gel filtration. Finally, it was possible to purify the enzyme to homogeneity. Longer storage even at -20°C led to the appearance of a doublet band on SDS–polyacrylamide gels, most likely due to a C-terminal processing, however, this slow self-processing was only partial, and did not result in any loss of the activity on oligopeptide substrates. Based on the results of specificity measurements, the specificity of BLV proteinase appears to be broader than that of HTLV-I, but the degree of differences in specificity of BLV and HTLV-I proteinases on substrates representing naturally occurring cleavage sites is more substantial than the differences between the HIV-1 and HIV-2 proteinases. This is in good agreement with the degree of homology between the core region of these enzymes

We attempted to make the specificity studies of the HTLV-I PR using HIV-1 MA/CA peptides that has been used to characterize representative PRs of various retrovirus subgroups, however we found that most of these peptides were not cleavable by HTLV-I PR due to the significantly different specificity of the HTLV-I PR compared to that of other retroviral PRs. Finally, a set of substrates based on the cleavage site of HTLV-1 CA/NC was used. Detailed analysis of the substrate binding pockets with the aid of a molecular model of the HTLV-1 PR provided a molecular interpretation of these specificity differences. To summarize our results, we can say that unlike in the case of HIV-1 PR, the S4 and S2 subsites of HTLV-I PR are more hydrophobic interactions, while there is a higher degree of similarity in the case of S3 and S1 and S1 subsites.

Our caspase specificity studies were focused on the effect of phosphorylation, which is one of the most fundamental regulatory mechanisms. In several studies, introduction of Asp residues is used to mimic phosphorylated serine residues for the functional analysis of phosphoproteins. In terms of caspase specificity, our studies demonstrate that the Asp and phospho-Ser residues are nonequivalent: while Asp is exclusively required at P1 position of the substrates for caspases 3, 7 and 8 and is also important at P4 position of the substrates for caspases 3 and 7, phospho-Ser at the same positions diminished or completely prevented the susceptibility towards cleavage. Therefore, phosphorylation at these sites could provide a regulatory mechanism to protect substrates from caspase-mediated degradation and direct cleavage site phosphorylation at proper positions may generally provide a regulatory system to control caspase-mediated proteolysis.

REFERENCES

References relating to the dissertation:

Zahuczky, G., Boross, P., Bagossi, P., Emri, G., Copeland, T.D., Oroszlan, S., Louis, J.M. and Tőzsér, J. (2000) Cloning of the bovine leukemia virus proteinase in Escherichia coli and comparison of its specificity to that of human T-cell leukemia virus proteinase. Biochim. Biophys. Acta. 1478 (1): 1-8. I.F.: 2.112

Tőzsér, J., **Zahuczky, G.**, Bagossi, P., Louis, J.M., Copeland, T.D., Oroszlan, S., Harrison, R.W. and Weber, I.T. (2000) Comparison of the substrate specificity of the human T-cell leukemia virus and human immunodeficiency virus proteinases. Eur. J. Biochem. 267 (20): 6287-6295. I.F.: 2.849

Tőzsér, J., Bagossi, P., **Zahuczky, G.**, Specht, S.I., Majerova, E. and Copeland, T.D. (2003) Effect of caspase cleavage site phosphorylation on proteolysis. Biochem. J. [epub ahead of print] I.F.: 4.326

Lectures and posters relating to the dissertation:

Zahuczky, G., Bagossi, P., Oroszlan, S., Tőzsér, J.. A HIV-1 és HTLV-I proteinázok szubsztrátspecificitásának és szubsztrátkötő helyeinek összehasonlítása. MBKE Molekuláris Biológiai Szakosztálya 3. Munkaértekezlete, Sárospatak, Május 11-14, 1998. (AP-16)

Zahuczky, G., Oroszlan, S., Bagossi, P., Tőzsér, J. (1998) Characterization of the specificity of HTLV-1 proteinase. 6th Cell- and Developmental Biology Meeting, Szeged, Hungary

Sperka, T., **Zahuczky, G.**, Boross, P., Emri, G., Oroszlan, S., Tőzsér, J. A humán T-sejtes vírus és a marha leukémia vírus proteolitikus enzimjeinek jellemzése. MBKE Molekuláris Biológiai Szakosztálya 4. Munkaértekezlete, Eger, Május 10-13, 1999. (AP-13)

Tőzsér, J., Sperka, T., Bagossi, P., **Zahuczky, G.**, Copeland, T.D., Oroszlan, S. and Louis, J.M. Characterization of the HTLV-1 proteinase. VIIIth International Aspartic Proteinase Conference, Funchal, Madeira, Portugal, Sep. 7-12, 1999.

Tőzsér, J, Bagossi P, Sperka T, **Zahuczky G**, Louis, J.M., Copeland, T.D., Oroszlan, S., Harrison, R.W. and Weber, I.T. Characterization of the specificity of HTLV-1 and HIV-1 proteinases. Retroviruses, Cold Spring Harbor Laboratory, New York, USA, May 23-28, 2000.

Other references:

Bagossi, P.**, Zahuczky, G.**, Tőzsér, J., Weber, I.T. and Harrison, R.W. (1999) Improved Parameters for Generating Partial Charges: Correlation with Observed Dipole Moments. J. Mol. Model., 5, 143 – 152. I.F.: 1.011