THESIS OF PHD DISSERTATION

STUDIES ON THE PROTEOLYTIC ENZYMES OF HUMAN RETROVIRUSES

János Kádas

Thesis advisor:
Prof. Dr. József Tőzsér

University of Debrecen
Medical and Health Science Center
Department of Biochemistry and Molecular Biology
Debrecen
2004

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.

These homodimeric, active proteases (PR) from retroviruses have been extensively characterized, both structurally and biochemically. The importance of such knowledge to the development of new drugs against acquired immunodeficiency syndrome (AIDS) has been the driving force behind this progress.

The function of retroviral proteinases is critical for virion replication. The first identified role of retroviral proteinases in the viral life cycle was the cleavage of Gag and Gag-Pol precursor proteins into functional structural proteins and enzymes. Additionally, cleavage of nucleocapsid protein was observed within the viral capsids in the early phase of virus infection. Later it turned out that many cellular proteins are also substrates of HIV proteinases which can contribute to the pathogenicity of the virus.

1.1. The retrovíruses:

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. The research of retroviruses excited a great interest due to the appearance and rapid spread of the acquired immunodeficiency syndrome (AIDS). 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.

Retroviruses comprise a large and diverse family of enveloped RNA viruses defined by common taxonomic denominators that include structure, composition, and replicative properties. The virions are 80–100 nm in diameter, and their outer lipid envelope incorporates and displays the viral glycoproteins. The shape and location of the internal protein core are characteristic for various genera of the family. The virion RNA is 7–12 kb in size, and it is linear, single-stranded, nonsegmented, and of positive polarity. The hallmark of the family is its replicative strategy which includes as essential steps reverse transcription of the virion RNA into linear double-stranded DNA and the subsequent integration of this DNA into the genome of the cell.

Retroviruses are broadly divided into two categories—simple and complex—distinguishable by the organization of their genomes. All retroviruses contain three major coding domains with information for virion proteins: *gag*, which directs the synthesis of internal virion proteins that form the matrix, the capsid, and the nucleoprotein structures; *pol*, which contains the information for the reverse transcriptase and integrase enzymes; and *env*, from which are derived the surface and transmembrane components of the viral envelope protein. An additional, smaller, coding domain present in all retroviruses is *pro*, which encodes the virion protease. Simple retroviruses usually carry only this elementary information, whereas complex retroviruses code for additional regulatory nonvirion proteins derived from multiply spliced messages.

Retroviruses are further subdivided into seven groups defined by evolutionary relatedness, each with the taxonomic rank of *genus*. Five of these groups represent retroviruses with oncogenic potential, and one of the other two groups is the *Lentiviridae* genus, with the HIV-1.

1.2. The human T-cell leukemia vírus (HTLV)

The 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.

1.3. The retroviral life cycle

Viruses enter the cell with receptor mediated endocytosis or direkt fusion. The genome RNA positive and regular like the eukaryotic mRNA, although it is not used for protein synthesis. The viral RNA transcripted to DNA by the reverse transcriptase (RT), maybe in the citoplasm inside the core structure. During *in vitro* incubation of EIAV capsids the NC was further processed into smaller fragments by the incorporated PR. Later the HIV-1 NC was found to be processed in a similar manner by HIV-1 PR. It is possible, that active PR which enter the cell within the „core” has a crucial role in reverse transcription and integration.

The new DNA-nucleoprotein complex get into nucleus, when the viral DNA integrate to the host cell’s genome (provirus formation).

The late phase of the retroviral life cycle starts with the transcripion of viral DNA into various RNA forms by the host RNA polymerase II. A significant fraction of newly sythesized mRNA reaches the cytoplasm unspliced, to serve as template for the Gag and Gag-Pro-Pol polyproteins, and to be packaged into virions as genomic RNA. The *gag* gen encodes for the structural proteins of the virus (MA: matrix protein, CA: capsid protein, NC: nucleocapsid protein), and the *pol* gene for the replication enzymes (RT: reverse transcriptase, IN: integrase). The *pro* gene can be expressed in different ways, such as by the supression of the *gag* termination code (MuLV) or by frameshifting (HIV-1, HTLV, BLV). The relatively low frequency of these events assures that the amount of replication enzymes in the virion is only about 5-10 % of those of the structural proteins encoded by the *gag* gene, while the PR of avian myeloblastoma virus (AMV) is encoded in the *gag* gene and therefore it is synthesized equimolarly with the structural Gag proteins. In the last stage of the late phase of the viral replication, the Gag and Gag-Pro-Pol polyproteins are assembled together with the envelope proteins and the viral genomic RNA at the plasma membrane of the infected cell, where the Env proteins concentrated, then budding yields „immature" virions with a "doughnut-shaped" capsid structure. The PR cleaves the viral polyproteins at a limited number of sites, producing a condensed, cone-shaped core of „mature" infectious particle. This function of PR is essential for virus replication.

In the early phase of viral replication, the viral capsid structure enters the cell. Based on compositional studies of EIAV and HIV-1 capsids they contain the viral RNA covered by nucleocapsid protein, capsid protein, and at least the replication enzymes (RT, RNaseH, IN, and PR). Reverse transcription occurs in this highly organized structure, then a preintegration complex (PIC) is formed, containing the viral DNA as well as integrase, although other proteins such as MA and Vpr were also detected in PIC. So far the HIV-1 RT and RNase H, the NC protein and the PR itself were described as substrate of the HIV-1 proteinase.

While the late-phase function of the PR is well established , it is function int he early phase is controversial and much debated. It was demonstrated for EIAV and later for HIV that the PR is part of the core structure, which enters the infected cell. In both the receptor – mediated andocytotic and macropinocytotic routes, the incoming core is surrounded by an acidic environment, which has an optimal pH for the retroviral protease. It has been observed, that the HIV PR is capable of cleaving cytoskeletal and sarcomeric proteins including vimentin, desmin, actin, myosin and tropomyosin. It has been proposed that the processing of vimentin is important int he early phase. The journey of the core towards the nucleus requires

interaction with the actin microfilaments, therefore proteolysis of the components of the actin filament may be an important step in regulation of this process.

Cellular proteins were also detected in virions. For example the peptidyl prolyl isomerase cyclophilin A (Cyp A) that is incorporated into the virion by binding to CA and it is presence enhances viral infectivity.

Actin, and various actin-binding proteins have also been detected. The fate of these proteins after entering the cells is mostly unknown, but many were already fragmented in the virions, and the involvement of the PR at least in degrading elongation factor-1 alpha (EF1α) was demonstrated. During *in vitro* incubation of EIAV cores in the presence of EDTA the NC was further processed *in situ* into smaller fragments by the incorporated PR. Among the proteins of the core, the HIV-1 RT, RNaseH and Nef have been demonstrated to be substrates of the HIV-1 PR. Furthermore, the PR itself undergoes self-degradation, while Vpr a nucleocytoplasmic shuttling protein of the PIC, remains intact in the capsid.The proteas-mediated clavage of viral and cellular proteins of the core may be important for the formation of proper preintegration complex.

1.4. The retroviral protease

Retroviral proteinases consists of 99-138 residues. Their molecular weight is 11-15 kDa and they are aspartic proteinases, active in homodimeric 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-helices 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.

Both oligopeptide and polyprotein substrates have been extensively used to probe the activity and specificity of retroviral proteinases. The measurements are routinely conducted at rather high salt concentration in case of oligopeptide (2-3 M NaCl), but low ionic strength in case of polyprotein substrates. Modifying the oligopeptides makes their measurements easier by allowing the use of spectrophotometric and fluorimetric techniques.

The residues found in the natural cleavage sites of retroviral proteinases are usually hydrophobic, however, a general consensus sequence cannot be given.

2. OBJECTIVES

Human T-cell leukemia virus type-1 (HTLV-1) is associated with a number of human diseases; therefore its protease - like the human immunodeficiency virus type-1 (HIV-1) protease (PR) - is a potential target for chemotherapy. Residues that confer inhibitor resistance to HIV-1 PR are frequently seen in equivalent positions of other retroviral PRs, as demonstrated in case of HTLV-1 PR. Therefore, understanding the specificity differences of PRs may help the design of inhibitors effective against the mutant HIV-1 PR forms appearing in resistance.

While HTLV-1 PR shares 28 % sequence identity with HIV-1 PR, based on a molecular model of the enzyme, the substrate binding region is more conserved showing 45 % sequence identity. Nevertheless, both the substrate specificity and inhibition profile of the two enzymes are substantially different. Several comparative studies have been performed on wild type retroviral proteases, however, most of them were done on oligopeptides substrates representing naturally occurring cleavage sites in the same viruses or selected peptides having various amino acid substitutions. In a few studies oligopeptides representing cleavage sites in other retroviruses were also probed. However, no comprehensive studies have been performed between two proteases using a library of oligopeptide substrates representing cleavage sites in several different retroviruses. Such studies may provide important knowledge about the common characteristics and differences of the PR specificity. We have performed comprehensive studies between HTLV-1 and HIV-1 proteases using a library of oligopeptide substrates representing cleavage sites in several different retroviruses.. Amino acid residues of HTLV-1 protease substrate biding sites were replaced by equivalent ones of HIV-1 protease.

The retroviral proteinase plays a crucial role in the late phase of the retroviral life-cycle by processing of the Gag and Gag-Pol polyproteins at the boundaries of the functional domains, during viral assembly and maturation. But we dont know much about the function of protease in early phase. In the early phase of viral replication, the viral capsid structure enters the cell. In this phase several research groups including ours found an protease inhibitory effect on replication, while others did not observe such an effect. Our research group found an PR inhibitory effect on EIAV replication in early phase. Reverse transcription occurs in this

highly organized structure, then a preintegration complex (PIC) is formed, containing the viral DNA as well as integrase, although other proteins such as MA and Vpr were also detected in PIC. Since the virus entering the cells contains the active protease, and protease-mediated cleavage of viral proteins, may be important for the preparation of proper preintegration complex. The presence of cyclophilin A decreased the degree of capsid protein processing. Further objective of this study was the examination of retroviral proteolitic clavage of the proteins traveling in virion included capsid and integrase.

3. MATERIALS AND METHODS

3.1. Mutagenesis of the HTLV-1 protease

The clone coding for the stabilized HTLV-1 PR in a pET expression vector was used as template for the mutagenesis. Mutants were generated by the Quick-Change mutagenesis protocol with the appropriate oligonucleotide pairs obtained from Genosys Sigma. Mutations were verified by DNA sequencing performed with ABI-Prism dye terminator cycle sequencing kit and an Applied Biosystems Model 373A sequencer.

3.2.Enzyme purification

Stabilized wild type HIV-1 and HTLV-1 PRs were prepared from inclusion bodies as described previously. Mutant HTLV-1 PRs were purified by using the same procedure. The reversed-phase HPLC-purified enzymes were homogenous based on SDS-polyacrylamide gel electrophoresis. Folding of the enzymes was performed by dialysis of the HPLC fractions in large excess of 25 mM formic acid, pH 2.8, and subsequently into 50 mM sodium acetate buffer, pH 5.0, 1 mM DTT, 1 mM EDTA. Finally the enzymes were dialyzed against 20 mM Pipes, pH 7.0, containing 1 mM EDTA, 100 mM NaCl, 10% glycerol, 5% ethylene glycol, 0.5% Nonidet P-40 and 10 mM DTT, to increase the stability of the PR preparation. Amino acid composition and protein amount of the enzymes were determined by amino acid analyses with a Beckman 6300 amino acid analyzer. Active site titration for the HIV-1 PR was performed with Compound 3, a phosphinate group-containing tight binding inhibitor of the enzyme . Active site titration of the wild type and some mutant HTLV-1 PRs were performed using peptide KTKVL-r-VVQPK (IB268) where r represents a reduced peptide bond. Reduced peptide bond-containing inhibitors used in this study (IB268 and IB269: APQVL-r-PVMHP) were synthesized by Dr. Ivo Blaha (Ferring Leciva).

3.3. Expression of the wild-type and mutant HTLV proteases as MBP-fusion proteins

HTLV-1 PR coding region and an 8-residue long N-terminal flanking sequence of an infectious HTLV-1 clone, pCS-HTLV-1 (35) was amplified by PCR and was cloned into the *Eco*RI./*Bam*HI. restriction sites of pMal-c2 in frame with the maltose binding protein (MBP)-coding sequence. The internal HTLV protease sequence (residues 12-116 of the mature PR, see Fig. 1A) was exchanged to the sequence of the HTLV-1 PR coding pET expression vector containing the stabilizing using *Pac*I. and *Eco*NI. restriction sites. Ligations, transformation of DH5α cells were performed by using standard protocols (36). Individual mutations were introduced into the expression clone using the same oligonucleotide pairs and protocol as described for the pET clones, and all vectors used for protein expression were verified by DNA sequencing. Protein expression was induced by the addition of 1 mM IPTG, for 3 h. After expression, cells were collected by centrifugation, lysed in 50 mM Tris-HCl, pH 8.2 containing 1 mM EDTA, 1 mM DTT, 1 % Triton X-100 with sonication. Protein samples separated by SDS-PAGE and transferred to PVDF membrane according to Towbin et al. (37). Immunoblots were developed using a mixture of antisera of rabbits immunized against peptides having the N-terminal and C-terminal sequences of the HTLV-1 protease a peroxidase-conjugated anti-rabbit antibody and ECL detection kit (Pierce, Rockford, IL).

3.4. Oligopeptides

Oligopeptides were synthesized by solid-phase peptide synthesis on a Model 430A automated peptide synthesizer 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 10 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).

3.5. Assays with oligopeptide substrates

The PR assays were initiated by the mixing of 5 μl (8 - 8500 nM) purified HTLV-1 or HIV-1 PR with 10 μl 2x incubation buffer (0.5 M potassium phosphate buffer, pH 5.6, containing 10% glycerol, 2 mM EDTA, 10 mM dithiothreitol, 4 M NaCl) and 5 μl 0.01 – 3.0

mM substrate. The synthesis and characterization of the oligopeptide substrates were described earlier (15,16, 30-33). The substrate concentration range was selected depending on the approximate K_m values. The reaction mixture was incubated at 37 °C for one hour and the reaction was stopped 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 x 150 mm, Waters Associates, Inc.) 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. Amino acid analysis of the collected peaks was used to confirm the site of cleavage with at least one PR, (typically with the PR of the same retrovirus) and to quantitate the amount of substrate cleaved, for other proteases the site of cleavage was assumed to be identical if cleavage products eluted with the same retention time and gave the same relative integration values as those identified by analysis. Kinetic parameters were determined by fitting the data obtained at less than 20% substrate hydrolysis to the Michaelis-Menten equation (or by linear fitting for k_{cat}/K_m values determined under pseudo first order conditions) by using the Fig. P program (Fig. P Software Corp). The standard errors of the kinetic parameters were below 20 %.

3.6. Inhibitor profiling

For the inhibitor assays, a microtiter plate reader assay using fluorescent Dabcyl/Edans tagged analog of the capsid↓nucleocapsid substrate (RE(Edans)TKVL↓VVQPK(Dabcyl)R where the arrow represents the cleavable bond) was used. Briefly, enzyme, substrate and inhibitor were incubated in 250 mM phosphate buffer, pH 5.6 containing 5 % glycerol, 1 mM EDTA, 5 mM DTT, 500 mM NaCl, 1% DMSO in 96-well microtiter plates. The increase of fluorescence was detected at 460 nm, using 355 nm excitation wavelengths in a Victor Wallace fluorimeter-luminometer and corrected with the inner filter effect. K_i values were calculated according to Williams and Morrison.

3.7. Molecular modeling

The crystal structure of the Rous sarcoma virus (RSV) S9 mutant used for the HTLV-1 PR starting model, as it showed all the flap residues. The amino-acid residues of RSV S9 PR were replaced by those of HTLV-1 PR.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. The structure with Thr-Lys-Val-Leu↓Val-Val-Gln-Pro oligopeptid were generated and minimized with the program AMMP. The sp4 potential set was used

The HTLV PR model was compared with crystal structure of HIV-1, HIV-2, SIV, EIAV, FIV proteases. The structures were built, minimized and examined on Silicon Graphics Indigo2 or O2 computer graphics systems.

3.8.Purification of the HIV-1 capsid protein

Purification of viral $p24^{CA}$ protein from cells infected with HIV-1 was performed as described previously. For the expression of recombinant capsid protein (r-$p24^{CA}$), the plasmid bearing the CA protein from HIV_{IIIB} isolate with an N-terminal 6-his tag was obtained from Dr. Carol Carter (Department of Molecular Genetics and Microbiology, S.U.N.Y. Stony Brook, USA). Expression of the recombinant protein in BL21(DE3) *E. coli* cells was induced by the addition of 0.4 mM IPTG to the cultures and incubation at 37 °C for 3 h. Bacterial cells were suspended in 10 mM Tris-HCl, pH 8.0 containing 150 mM NaCl, and 1 mM PMSF, then disrupted by sonication. The insoluble fraction was collected by centrifugation and resuspended in 20 mM phosphate buffer, pH 7.5, containing 500 mM NaCl, 0.1 % Triton X-100 and 8 M urea. After sonication, the solution was clarified by filtration then applied to the Ni-NTA Superflow affinity resin equilibrated with the same buffer. After extensive consecutive washing with a buffer having the same composition as the loading buffer, but with pH 6.0 and containing 500 mM LiCl, the recombinant protein bound to the resin was eluted with 50 mM Na-acetate, pH 5.0, containing 8 M urea, 500 mM NaCl, 10 % glycerol, 1 % Triton X-100 and 5 mM β-mercaptoethanol. The purified protein was dialyzed overnight at 4°C, against 25 mM formic acid, lyophilized and reconstituted in assay buffer. Protein concentrations were determined by the Bradford spectrophotometric method.

3.9. Purification of HIV-1 integrase

HIV-1 pINSD.His.Sol plasmid coding for a soluble integrase mutant was obtained through the AIDS Research and Reference Program. Integrase was expressed in BL21(DE3) *E. coli* cells by addition of 0.4 mM IPTG to the culture and incubation at 37 °C for 3 h. Bacterial cells were suspended in 10 mM Tris-HCl, pH 8.0 containing 150 mM NaCl, 0.1 % Triton-X-100, and treated with 100 μg/ml DNAse I for 30 min at 37 °C in the presence of 10 mM $MgCl_2$. The insoluble fraction was collected by centrifugation and solubilized in 20 mM phosphate buffer, pH 7.5, containing 500 mM NaCl, 0.1 % Triton X-100 and 8 M urea. The solution was clarified by filtration and applied to nickel-charged ProBond resin equilibrated with the same buffer. After extensive consecutive washing with the loading buffer, protein was eluted with gradually decreasing the pH to 5.0. Fractions containing the recombinant

protein (as determined by SDS-PAGE) were combined and dialyzed overnight at 4°C, against 25 mM formic acid, lyophilized and reconstituted in assay buffer.

3.10 Incubation of viral capsid and integrase proteins by HIV-1 Proteinase

The capsid prepared from infected cells was a kind gift of Dr. Mangalasseril G. Sarngadharan (Advanced BioScience Laboratories Inc., MD, USA).

To determine the effect of pH on capsid protein degradation, $p24^{CA}$ isolated from virus infected cells (4.3 μM final concentration) was incubated in 100 mM potassium phosphate, 2.5 mM DTT, pH 4.0 with recombinant HIV-1 PR (0.7 μM final concentration). After 16 h incubation reactions were stopped by the addition of equal volume of 2X tricine-SDS sample buffer. The samples were analyzed using 16% or 10-20 % gradient SDS-polyacrylamide tricine buffered gel. Molecular weight of the fragments was estimated using Rainbow molecular weight markers.

For the cleavage of r-$p24^{CA}$ as well as r-$p32^{IN}$, recombinant protein (20 μM and 8 μM final concentration, respectively) was incubated with recombinant HIV-1 PR (1 μM final concentration) in 100 mM sodium acetate buffer, pH 5.5, containing 1 mM DTT, 1 mM EDTA and 150 mM NaCl. To test the effect of cyclophilin A on the capsid cleavage, 43 μM Cyp A was preincubated with the capsid protein for 1 hour at room temperature. As nonspecific controls, bovin serumalbumin and lysozyme were used in 43 μM concentration, in a same way as Cyp A. After incubation for various times at 37°C, reactions were stopped by the addition of equal volume of 2X tricine-SDS sample buffer, and the samples were analyzed using 16% or 10-20% gradient SDS-polyacrylamide tricine buffered gels.

3.11. Sequencing of fragments of capsid protein

Fragments of the CA protein cleaved by PR were separated by SDS-PAGE and transferred to PVDF membrane according to Towbin *et al*. N-terminal amino acid sequence analysis was carried out by stepwise Edman degradation in a Applied 470A gas-phase sequencer or in a Knauer 910 protein sequencer.

4. RESULTS AND DISCUSSION

4.1. Narrow substrate specificity and sensitivity towards ligand binding site mutations of human T-cell leukemia virus type-1 protease

4.1.1 Comparison of the substrate specificity of HTLV-1 and HIV-1 proteases using oligopeptides representing naturally occurring cleavage sites in retroviruses

To compare the specificity of the two proteases, a large set of 50 oligopeptides representing naturally occurring cleavage sites in HIV-1, HIV-2, EIAV, RSV, MMTV, MMLV, BLV and HTLV-1 were used. The selected peptides were previously characterized and were found to be hydrolyzed by the respective protease coded within the same virus. Only a few peptides were cleaved by appreciable kinetics with both HIV and HTLV-1 proteases. While most PR cleavage sites contain hydrophobic residues at P1 and P1', with one exception a common characteristics of those peptides which were substrates for both enzymes is having hydrophobic beta branched residue (Val or Ile) at P2, and also a hydrophobic beta branched residue or Leu at P2' positions. These are typical arrangements for the naturally occurring HTLV-1 cleavage sites and to a lesser extent for the BLV cleavage sites. Three peptides were hydrolyzed only by the HTLV-1 PR but not by HIV-1 PR. On the other hand, HIV-1 PR was able to cleave many peptides, which were not substrates of the HTLV-1 PR, indicating that it has a broader specificity.

4.1.2 Mutations introduced to the HTLV-1 protease

The sequence alignment of HTLV-1 and HIV-1 PRs based on the multiple alignments of PRs with known structures. The kinetic data have been analyzed using a molecular model of HTLV-1 PR and a crystal structure of HIV-1 PR-inhibitor complex. The substrate binding site of PR is comprised mainly by residues located at three regions: the active site region, the flap region which closes down on the bound ligand, and the C-terminal region at the shoulder of the substrate binding site. The active site region is very conserved in retroviral PRs, it contains only one amino acid substitution between HTLV-1 and HIV-1 PRs; the residue corresponding to HTLV-1 PR Met 37 is Asp in HIV-1 PR. The same residue differs in other retroviral PRs. Unlike the active site region; most of the residues of the flap and C-terminal regions are different in the two PR. In this study, fifteen HTLV-1 PR mutants were generated by replacing key residues of the substrate binding pockets with the structurally equivalent residues of HIV-1 PR. These mutations included M37D, V56I, L57G, A59I, F67Q, N96T, N97P, and W98V.

Since the residue corresponding to Met37 is an important determinant of specificity in PRs hydrophobic residues of various sizes (Ala, Val, Ile) as well as the drug resistant mutation Asn of HIV-1 PR were introduced at this position. Three mutants with combined substitutions at the flap and C-terminal regions were also tested. The mutant PRs were purified to homogeneity from inclusion bodies and folded as previously described for wild-type PRs

4.1.3. Activities of the mutant HTLV-1 proteases against a peptide substrate and their autoprocessing capabilities from MBP fusion protein

The peptide representing the capsid↓nucleocapsid cleavage site of HTLV-1 was shown to be a good substrate of both wild-type HTLV-1 and HIV PRs. Eight out of the 15 HTLV-1 PR mutants were able to hydrolyze this substrate, although, mostly with dramatically reduced kinetic efficiency as compared to the two wild-type enzymes. However, the K_m values obtained with some mutants were similar to those obtained with the wild-type enzymes, while some other mutants showed moderately increased values. Since many of the mutants showed very low activity and loss of inhibitor potency, active site titration was possible only for M37V, M37I and the mutant having HIV-1 PR-like flap. Comparison of the active protein content with the total amount of protein suggested that the folding efficiency of these mutants is much lower as compared to the wild-type enzymes. These results also suggest that the lack of activity of some mutants may be the consequence of the inability to refold properly.

To further verify the low activity/improper refolding capability of the mutants, they were also expressed as MBP fusion proteins, from which the wild-type enzyme is capable to self-process itself out. An advantage of expressing a protein in fusion with MBP is that it is very effective to promote the solubility of polypeptides to which it is fused, compared to other commonly used proteins, like GST and thioredoxin. All of the mutants that appeared to be inactive after folding from inclusion bodies using the pET expression system were also expressed in fusion with MBP. Unlike the wild-type enzyme as well as the mutant having HIV-1 PR-like flap, none of those mutants were capable of self processing during expression.

4.1.4. Mutations affecting mainly the S4/S2 binding sites of HTLV-1 protease

The S4 subsite is close to the surface of the enzyme, and partly exposed to solvent for all retroviral PRs. In contrast, the S2 subsites are relatively small and formed by mostly hydrophobic residues, and consequently would be expected to accommodate smaller

hydrophobic P2 residues. However, while HIV-1 PR preferred more polar amino acids at P2 (and P2'), all other PRs including HTLV-1 PR preferred hydrophobic residues.

Residues forming the S4 binding sites of the HTLV-1 and HIV-1 PRs include Met37/Asp30, Ser55/Met46, Val56/Ile47, Leu57/Gly48 and Val92/Leu76 with very distal contributions from Phe67/Gln58 and the Pro7-Ala8 insertion in HTLV-1 PR. Asp30 of HIV-1 PR, and structurally equivalent residues of other PRs, were shown to be important determinants for the different specificities of the PRs at S4 and S2. In another example, HIV PR mutant Asn30, harboring a mutation appearing in nelfinavir resistance, showed specificity changes compared with wild-type enzyme accompanied by small structural changes in PR-substrate interactions. Therefore, HTLV-1 PR Met37 was substituted by Asp, Asn, Ala, Val and Ile. Interestingly, the mutants containing Asp37 or Asn37 were inactive on all substrates tested (not published). They were also not able to self-process themselves from MBP fusion proteins. Substituting the smaller Ala in this position also rendered the enzyme inactive on most of the substrates (not published), except for the substrate with Phe at P1, which was the best substrate for the wild-type enzyme, suggesting that proper hydrophobic interactions at this position of HTLV-1 PR are required for efficient catalysis. The other mutants with conserved substitutions (Val or Ile) at residue 37 still showed substantially reduced k_{cat} values without an apparent change of K_m values as shown for the wild-type substrate. The specificity of these mutants was also substantially altered. Both preferred P4 Ile as compared to the P4 Val preference for the wild-type HTLV-1 PR, similar to the HIV-1 PR. However, they also preferred Leu over Val at this position, which is not seen with either of the wild-type PRs. Val56 of HTLV-1 is part of the flap region, and Ile is the equivalent residue in HIV-1 PR. The V56I mutant was a very inefficient enzyme, but it hydrolyzed the P4 Leu substituted peptide with a substantially increased rate, suggesting that the larger side chain of Leu at P4 was able to make better hydrophobic interactions with V56I than the other smaller side chains of Val, Thr and Ile at P4. The order of P4 side chain preference of the mutant having HIV-1 PR-like flap resembled to that of wild-type HIV-1 PR, except it did not hydrolyze well the substrate having Ala at this position.

The S2 subsite shows differences in residues Met37/Asp30, Val56/Ile47, Ala59/Ile50 and Val92/Leu76 for HTLV-1 and HIV-1 PRs. These differences result in a somewhat larger, more hydrophobic S2 subsite in HTLV-1 PR as compared to HIV-1 PR. Interestingly, both M37V and M37I mutations changed the Val > Leu >Ile P2 preference to Ile >Val > Leu, the same as observed for HIV-1 PR, however, they were not able to hydrolyze the substrate containing bulky aromatic Phe at this position. Introduction of beta-branched residues into this position may

restrict the size of side chain, which can fit into this subsite. Similar to the findings obtained with the P4 modified substrate, the mutant having HIV-1 PR-like flap showed a P2 specificity more similar to that of HIV-1 PR than HTLV-1 PR, but this mutant also did not tolerate Phe at this position, likely due the two beta-branched residues introduced into the sequence. These results suggested that the flap region is important not only for the enzyme activity, but also for the specificity differences of the two wild-type PRs.

4.1.5 Mutations affecting the S3/S1 binding sites of HTLV-1 protease

The S3 subsite is usually large, and 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. Several P3 substituted substrates were hydrolyzed by both wild-type PRs, indicating that this site is not restrictive, but while HTLV-1 PR preferred the original Lys-containing substrate, HIV-1 PR preferred larger, hydrophobic residues. Analysis of the HIV-1 PR crystal structure and the model structure for HTLV-1 PR suggested that about half of the residues participating in the S3-P3 interactions are identical in both enzymes. The S3 subsites differ in residues Leu57/Gly48 and Ala59/Ile50 from the flap, while Asn96'/Thr80', Asn97'/Pro81' and Trp98'/Val82' residues from the other subunit form one side of the subsite. Wild-type HTLV-1 PR was less sensitive to P3 changes: while Leu 57 and Trp 98' were predicted to interact favorably with hydrophobic residues, Asn 96' and Asn 97' can potentially interact with polar P3 residues, including Lys of the unmodified substrate. Both L57G and A59I flap mutants were very inefficient enzymes on the unsubstituted peptide. Leu57 is predicted to interact favorably with larger hydrophobic P3 residues in the wild-type HTLV-1 PR, and this interaction appears to be important for efficient hydrolysis. While P3 substitutions did not substantially improve the hydrolysis for the L57G mutant, surprisingly the P3 Ala substituted peptide was the best substrate for this enzyme (not published). Unlike the L57G mutant, introduction of larger hydrophobic residues into P3 dramatically improved the specificity constants for the A59I mutant: Ile side chain in this position was predicted to make more hydrophobic contacts with hydrophobic P3 side chains as compared to the original Ala residue. The A59I mutant showed a preference for Phe at P3, consistent with additional favorable hydrophobic interactions. Substitution of Asn96 to Thr and Asn97 to Pro increased the relative preference toward hydrophobic P3 residues over Lys at P3. Unexpectedly, W98V mutant did not have a preference toward larger hydrophobic residues, it rather preferred smaller hydrophobic ones.

The S1 subsite shows differences in Ala59/Ile50, Asn96'/Thr80', Asn97'/Pro81' and Trp98'/Val82' in HIV-1 and HTLV-1 PRs. Despite these sequence differences, the structure and the specificity at the S1 subsite of the two PRs appears to be similar. The S1 subsite is a relatively large and hydrophobic, internal pocket in the enzyme. The best substrate for both enzymes had Phe at P1. Unexpectedly, A59I, N97P and W98V mutants preferred P1 Met instead of the P1 Phe preferred by both wild-type PRs. These substitutions introduced beta branched side chains, or conformational restraints in case of Pro, which may restrict the fit of the large bulky Phe P1 side chain within the subsite.

The specificity of the mutant having HIV-1 PR-like flap resembled the specificity of HIV-1 PR on both P3 and P1, however, it also had some unique features, for example the best P3 substitution was Leu and the mutant did not favor Phe at P1 or P1'.

4.1.6. Inhibition profile of the wild-type and mutant HTLV-1 proteases

Several potent inhibitors that target HIV-1 PR are used in clinical practice. These include indinavir, saquinavir, nelfinavir, ritonavir and amprenavir. We have tested these inhibitors and two reduced peptide-containing HTLV-1 inhibitors on the wild-type and the mutant HTLV-1 PRs. The assays were performed in a high-throughput microtiter plate fluorescent assay system. Except for indinavir, the HIV-1 PR inhibitors were not effective against the wild-type HTLV-1 PR. We have previously tested several inhibitors of HIV-1 PR including saquinavir, as inhibitors of the HTLV-1 PR by using the very labor-intensive and time-consuming HPLC assay. Only Compound 3 inhibited the PR up to 10 μM. Four of the HIV-1 PR inhibitors used in clinical practice were also tested on HTLV-1 Gag processing *in vitro* and did not show any effect. These results demonstrate the large differences in specificity of the two enzymes. The most potent inhibitor of the HTLV-1 PR in our previous studies was a statine-containing inhibitor, based on the matrix↓capsid cleavage site sequence of HTLV-1. In contrast, the statine-containing peptide based on the capsid↓nucleocapsid cleavage site sequence did not inhibit the enzyme. Here, we have tested two reduced peptide bond-containing inhibitors of the HTLV-1 PR, IB268 and IB269, with the sequences of the HTLV-1 capsid↓nucleocapsid and matrix↓capsid cleavage sites, respectively. Our previous studies on HIV-1 and HIV-2 PRs indicated that reduced peptide bond-containing substrate analogs act as potent inhibitors of the retroviral proteinases in the presence of high salt. IB268 and IB269 were potent inhibitors of the wild-type HTLV-1 PR in the HPLC assay, with K_i values below 50 nM.

They were also the best inhibitors of the wild-type HTLV-1 PR in the fluorometric assay, which was performed in a substantially lower ionic strength.

Most of the HTLV-1 PR mutants were not inhibited by the HIV-1 PR inhibitors: only the enzyme having an HIV-1 PR-like flap showed some degree of sensitivity towards some of the HIV-1 PR inhibitors while it showed decreased affinity towards the HTLV-1 PR inhibitors. The large effect of the V56I mutation on inhibitor susceptibility is in good agreement with previous findings that the corresponding I47V mutation appears in drug resistance *in vitro*. This residue seems to be important for defining the differences in inhibitor susceptibility of the HIV-1 and HTLV-1 PRs.

4.2. Human Immunodeficiency Virus Type 1 Capsid Protein is a Substrate of the Retroviral Proteinase while Integrase is Resistant toward Proteolysis

4.2.1. Capsid protein is a pH-dependent substrate of HIV-1 PR, in vitro

Capsid protein ($p24^{CA}$) purified from eukaryotic cells infected with HIV-1 was incubated overnight with purified recombinant HIV-1 PR in buffers at various pH. While no cleavage was observed at pH 7.0, decreasing the pH resulted in substantial processing of the $p24^{CA}$ protein. At pH 5.0 the largest, 22 kDa fragment appeared with very low molecular weight fragments, while at pH 4.0, three intermediate-sized bands also appeared, implying that the cleavage yielding these fragments may have a lower pH optimum than the cleavage producing the 22 kDa and the very short fragments.

4.2.2. Time-dependent processing of recombinant capsid protein by HIV-1 PR

To determine the cleavage sites in the capsid protein, recombinant $p24^{CA}$ protein expressed in *E. coli* (r-$p24^{CA}$) was subjected to proteolysis by HIV-1 PR. At 1:20 PR:capsid ratio and pH 5.5 (which is optimal for the PR) a time-dependent processing was observed, which was prevented by the addition of 0.2 μM ritonavir, a specific PR inhibitor.

To identify all of the fragments obtained from $p24^{CA}$ degradation, the cleavage products were separated by SDS-polyacrylamide gel electrophoresis, transferred to a PVDF membrane, then Coomassie stained bands were cut out and subjected to N-terminal analysis. Bands were identified by their N-terminal sequence and by their apparent molecular weight. Based on these data, two cleavage sites were identified in the $p24^{CA}$ protein, one between Ala

77 and Ala 78 of the N-terminal domain of CA, and the other between Leu 189 and Leu 190 of the C-terminal domain of CA.

Based on the crystal structures of $p24^{CA}$, the cleavage site of the C-terminal domain is in helix 9, where Leu 189 is a part of the dimer interface while the cleavage site of the N-terminal domain is in helix 4, which leads to the Cyp A binding loop. These structural elements are incompatible with the structural requirement for the PR-mediated hydrolysis, since the substrate should bind into the enzyme in an extended conformation. Similar discrepancies were also found with other protein substrates of HIV-1, in which the cleavage site regions were found to be part of alpha helices. However, the crystals for determination for the $p24^{CA}$ structure were grown at pH 7.5, at which the protein appeared to be resistant toward proteolysis. The structure and higher organization of $p24^{CA}$ is very sensitive to the pH, below 6.6 the multimeric capsid forms disassemble into monomers, and its structure assumes a "molten globule" state in which the local structure around the cleavage sites may be altered to allow productive binding to the PR.

4.2.3. Effect of cyclophilin A on the processing of capsid protein by HIV-1 PR

To study the effect of cyclophilin on the $p24^{CA}$ degradation, r-$p24^{CA}$ was preincubated with cyclophilin A (Cyp A) at pH 4.0 (which is optimal for the N-terminal site cleavage) before adding of the PR into the reaction mixture. Binding of Cyp A to $p24^{CA}$ substantially reduced the rate of cleavage of capsid protein, indicating that this chaperone was able to delay the proteolytic processing of CA by HIV-1 PR. Serumalbumin or lysozyme used in the same concentration as Cyp A did not inhibit the cleavage of $p24^{CA}$ indicating that the effect of Cyp A was due to specific protein-protein interaction.

4.2.4. Cleavage of oligopeptides representing the processing sites in the capsid protein by HIV-1 PR

Oligopeptides were synthesized representing the cleavage sites at the identified sites in CA of $HIV\text{-}1_{HXB2}$ and assayed as substrates of HIV-1 PR. While at assay conditions we previously found to be optimal for HIV-1 PR (pH 5.6, 2 M NaCl) the peptide representing the C-terminal cleavage site appeared to be a somewhat better substrate. When the kinetic parameters were determined under conditions similar to that of the *in vitro* protein cleavages at pH 4.0, the other substrate was much better hydrolyzed.

HIV-1 PR cleaves the Gag and Gag-Pol proteins at various positions during viral maturation. The amino acid residues in the P4-P3' region of these sites, which are recognized

by the PR, are listed in Table 3. Most of these sites contain hydrophobic residues at the P2-P2' region. Although only one viral HIV-1 PR maturation cleavage site, the CA↓p2 site contains charged residue at this region, a P2' glutamate, this residue is highly conserved in HIV-1. The cleavage at this site is accelerated by lower pH, and this was suggested to play a regulatory role in the viral protein processing. While the sequence of the cleavage site at the C-terminal domain of p24 fits well into the other Gag and Gag-Pol sequences, the cleavage site in the N-terminal domain of p24 contains Glu residue at both the P2 and P2' positions. Crystallographic structures of HIV-1 PR have been reported with inhibitors or products having P2' Glu residues. In these structures the P2' Glu showed two weak hydrogen bond interactions with the amides of Asp 29 and Asp 30 of the enzyme. The proximity of the Glu side chain to the side chain of Asp 30 suggested that they share a proton. The presence of glutamate at P2' position of various cleavage sites was found to be responsible for the pH effect. The presence of two glutamates in this p24 cleavage site may enhance such a pH effect explaining why lowering the pH enhanced substantially the cleavage in the N-terminal domain of CA. Unlike the P2 and P2' glutamates, Glu at P3 and P3' in the peptide representing the cleavage site in the C-terminal domain did not result in such an effect.

Since the virus entering the cells contains the active protease, its activity could be boosted by the acidic pH of the endosomes, and protease-mediated cleavage of viral proteins, including CA may be important for the preparation of proper preintegration complex.

4.2.5. Integrase is not a substrate of HIV-1 PR

Recombinant IN of HIV-1 was incubated with HIV-1 PR under the same conditions as used for the capsid cleavage. During the 4 h incubation the integrase protein remained intact and active in good agreement with its presence in the preintegration complex.

5. SUMMARY

During my Ph.D work I have had the opportunity to study the proteinases of human retroviruses, and possible substrates of these enzymes in the early phase of retoviral life cycle.I have compared the specificty and inhibitor susceptibility of wild-type HTLV-1, HIV-1 and mutant HTLV-1 proteases.

The specificity of HTLV-1 and HIV-1 proteases was compared using oligopeptides representing maturation cleavage sites in various retroviruses. Based on the results, the specificity of HTLV-1 PR appears to be substantially narrower as compared to the specificity of HIV-1 PR. To further characterize the specificity of HTLV-1 PR, mutations were introduced into its ligand binding sites. Many of the mutants we have generated were either inactive or possessed scarcely detectable activity and loss of capability to self-process from an MBP fusion protein. Based on active site titration, the folding efficiency of the active mutants is much lower as compared to the wild-type enzymes. Based on our results, the specificity of HTLV-1 PR is narrower than that of HIV-1 PR, and both the folding pathway and the catalytic efficiency of HTLV-1 PR is apparently much more sensitive towards mutations than those of other retroviral PRs, especially HIV-1 PR. Our results presented here suggest that the flap region has an important role in determining the differences in both substrate specificity and inhibitor susceptibility of the HTLV-1 and HIV-l proteases. In this aspect it is important to note that after infection HTLV-1 typically replicates by cell division and not by producing exogenous virions, omitting the error-prone reverse transcription step. Therefore, HTLV-1 is much more conserved than HIV-1. As a consequence, the HTLV-1 PR has not undergone the rapid evolution that was able to optimize the HIV-1 PR for both high catalytic efficiency as well as flexibility in tolerating mutations.

In the early phase of retroviral life the viral capsid structure enters the cell. Among the proteins of the core other proteins -wich are travelling in this structure- have been demonstrated to be substrates of the retroviral protease. Our studies indicated that the CA protein is also a substrate of HIV-1 PR. Its processing is strongly dependent from a pH-induced conformational change. The capsid protein was found to be cleaved at two sites. The presence of cyclophilin A decreased the degree of capsid protein processing and it may help to stabilize the „core" structure. Unlike the capsid protein, integrase was found to be resistant towards proteolysis in good agreement with its presence in the preintegration complex.

REFERENCES

References relating to the dissertation:

Kádas, J., Weber, I.T., Bagossi, P., Miklóssy, G., Boross, P., Oroszlan, S., and Tözsér, J. (2004) Narrow substrate specificity and sensitivity toward ligand bidnding site mutations of human T-cell leukemia vírus type-1 protease. J. Biol. Chem. Apr 20 (Epub ahead of print). I.F.: 6.696

Tözsér, J., Shulenin, S., Kádas, J., Boross, P., Bagossi, P., Copeland, T.D., Nair, B.C., Sarngadharan, M.G., and Oroszlan, S. (2003) Immunodeficiency Virus Type 1 Capsid Protein is a Substrate of the Retroviral Proteinase while Integrase is Resistant toward Proteolysis. Virology. 310: 16-23. I.F.: 3.363

Other references:

Bagossi, P., Kádas, J., Miklóssy, G., Boross, P., Weber, I.T., and Tözsér, J. (2004) Development of a microtiter plate fluorescent assay for inhibition studies ont he HTLV-1 and HIV-1 proteinases. J.Virol. Methods. In press. I.F.: 1.938

Lectures and posters relating to the dissertation:

Kádas J., Sperka T., Bagossi P., Boross P. és Tőzsér J. Mutáns retrovirális proteázok kinetikai jellemzése: Mutációk hatása a specificitásra. A Magyar Biokémiai Egyesület molekuláris Biológiai Szakosztályának 5. Munkaértekezlete, Sopron, 2000. Május 08-11.

Kádas, J., Weber, I.T., Bagossi, P., Boross, P., Miklóssy, G., Louis, J.M., Copeland, T.D., and Tözsér, J. Alteration of the substrate specificity of Human T-cell Leukemia Virus Protease. Meeting on Retroviruses 2003., Cold Spring Harbor, USA. May 20-25, 2003.

Tözsér, J., Kádas, J., Bagossi, P., Boross, P., Miklóssy, G., Oroszlan, S., and Weber, I.T. Alteration of the substrate specificity of Human T-cell Leukemia Virus Protease. Fourth HIV DRP Symposium on Antiviral Drug Resistance, Chantilly, Virginia, USA, December 7-10, 2003.

Other posters:

Kádas J., Varga S., Bagossi P. és Tőzsér J. Kísérletek zöld fluoreszcens fehérjén alapuló, lentivírus fertőzés monitorozására alkalmas *in vitro* rendszer kifejlesztésére. A Magyar

Biokémiai Egyesület Molekuláris Biológiai Szakosztályának 4. Munkaértekezlete, Eger, 1999. Május 04-06.

Miklóssy G., Fehér A., Kádas J., Sperka T. és Tőzsér J. Az erbB2 receptor ”shedding”-jének vizsgálata emlőtumor sejtekben. A Magyar Biokémiai Egyesület Molekuláris Biológiai Szakosztályának 7. Munkaértekezlete, Keszthely, 2002. Május 14-17.

Kádas, J., Tözsér, J. *In vitro* monitoring Equine Infectious Anemia Virus (EIAV) infection and its inhibition by antiretroviral drugs in early phase. VectEuroTrain - 2nd European Conference and Practical Course Towards clinical gene therapy: pre-clinical gene transfer assessment, Bellaterra, Barcelona, Spain, February 1-14, 2004.