PATHOLOGY ONCOLOGY RESEARCH
Vol 3, No 2, 1997

ARÁNYI LAJOS
Foundation

SPECIAL REPORT

Specificity of Retroviral Proteinases Based on Substrates Containing Tyrosine and Proline at the Site of Cleavage+

József TÖZSÉR

Department of Biochemistry, University Medical School of Debrecen, Hungary

The retroviral proteinase (PR) plays crucial roles in the viral life cycle, therefore it is a target for chemotherapy. However, resistance rapidly develops due to frequent mutations. Studies to determine the common features of the specificity of different retroviral PRs may help to design broad spectrum inhibitors and reduce the possibility of viable mutants. We have studied the specificity of various retroviral proteinases including those the PR of HIV-1, HIV-2, equine infectious anemia virus and avian myeloblastosis virus using oligopeptide substrates. A series of oligopeptides containing substitutions in a sequence Val-Ser-Gln-Asn-Tyr*Pro-Ile-Val-Gln (asterisk indicates the site of cleavage) representing a naturally occurring cleavage site in HIV-1 was used to characterize the seven substrate binding subsites of the enzymes. The unsubstituted substrate is a typical class 1 cleavage site substrate containing an aromatic amino acid and a proline residue at the site of cleavage. The largest differences in kinetics of substrate hydrolysis were obtained with peptides containing substitutions of the Ser and Asn residues. Detailed analysis of the results by molecular modeling and comparison with previously reported data revealed the common characteristics of the specificity of the PRs as well as its strong dependence on the sequence context of the substrate. However, molecular modeling in many cases provided explanation for the sequence context dependence. Also, comparison of the specificity of the enzymes suggests that the specificity of HIV-1 and -2 PRs is rather exceptional preferring hydrophilic residues at the most discriminative positions while other PRs prefer hydrophobic residues. (Pathology Oncology Research Vol 3, No 2, 142–146, 1997)

Key words: proteinases, retroviral, HIV, cleavage, specificity

All replication competent retroviruses including HIV, code for a proteolytic enzyme which is essential for virus replication, therefore it is an attractive target for chemotherapy.[1,2] Potent inhibitors are being developed and some of them are now used in clinical practice.[3,4] However, as with reverse transcriptase inhibitors, resistance rapidly develops. Therefore better understanding of the specificity of retroviral PRs may help to design broad spectrum inhibitors.

The retroviral PRs are homodimeric aspartic proteinases with 99-126 residue long subunits.[2] Three characteristic conserved regions are the active site triplet, the so called flap region and a third conserved region close to the C-terminal end *(Fig.1)*. An interesting feature of the naturally occurring PR cleavage sites is the lack of a consensus sequence *(Table 1)*, which would suggest that the retroviral proteinases are rather unspecific enzymes; however, they process their natural substrates in a restrictive way. Another interesting feature is that Pro frequently occurs at the site of cleavage *(Table 1)*, in a position where it is not found in cleavage sites of cellular proteinases. In fact, one of the earliest indications that a retroviral PR exists was the finding of N terminal proline residues in animal retroviral proteins[5]. Initially three types of cleavage sites were proposed for HIV-1, HIV-2 and simian immunodeficiency viruses (SIV).[6] Subsequently, two major types of cleavage sites were proposed for retroviral proteinases, type 1 having -Tyr(Phe)*Pro- and type 2 having hydrophobic residues (excluding Pro) at the site of

Received: April 8, 1997; *accepted:* May 29, 1997

Correspondence: Dr. József TÖZSÉR, Department of Biochemistry, University Medical School of Debrecen, Nagyerdei krt 98, 4012 Debrecen, Hungary; Tel/fax: (36) (52) 416-432

+Research sponsored in part by the Hungarian Science and Research Fund (OTKA T 16396, T 22670) and U.S.-Hungarian Science and Technology Fund (JF.No.320).

1219-4956/97/020142+05 $ 12.00/0

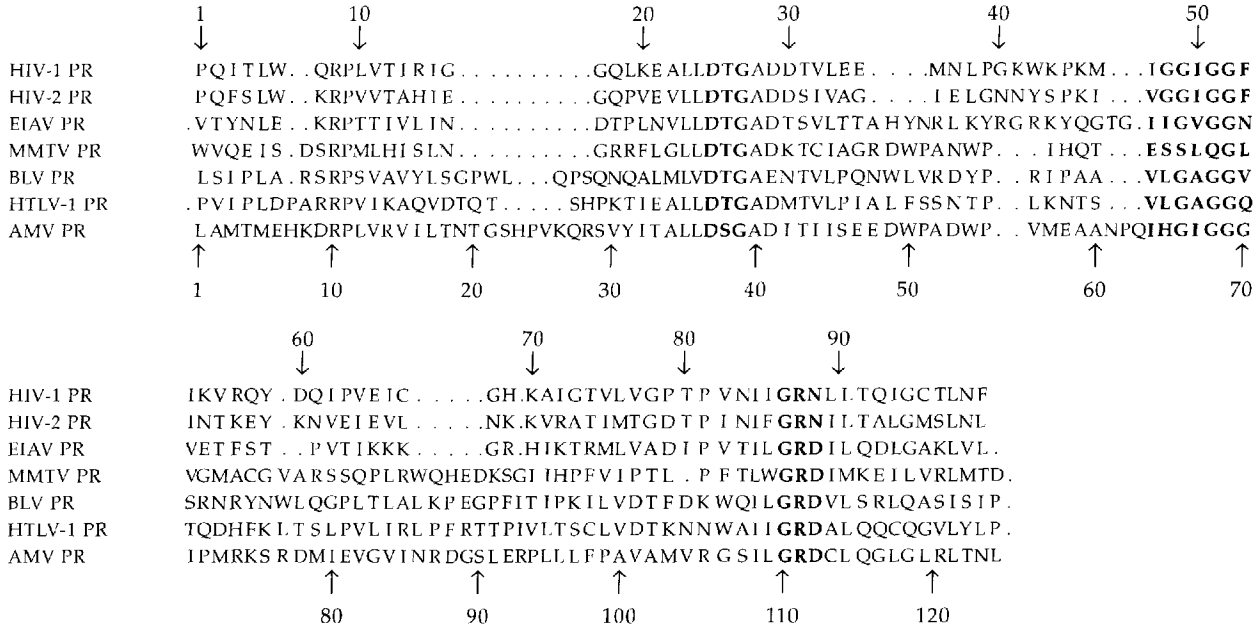

Figure 1. *Sequence comparison of retroviral proteinases. The amino acids of the conserved regions are in bold.*

cleavage.[7-9] These two types of cleavage sites seem to have different preferences for the P2 and P2' positions, where the peptide bond between P1 and P1' is cleaved (notation is according to Schechter and Berger[10]).

To better understand the specificity of retroviral proteinases we have studied and compared the specificity of HIV-1, HIV-2 PRs,[9,11] equine infectious anemia virus (EIAV) PR[12] and avian myeloblastosis virus (AMV) PR[13] using a type 1 series oligopeptide substrate. These oligopeptides contained substitutions in the sequence Val-Ser-Gln-Asn-Tyr*Pro-Ile-Val-Gln (asterisk indicates the site of cleavage) representing a naturally occurring cleavage site between the matrix and capsid proteins of HIV-1. Substrates were cleaved by the enzymes, then products were analysed by reversed-phase HPLC analysis to determine the kinetic constants. In some cases relative activities (activity on a given peptide determined at 0.4 mM concentration divided by the activity obtained for the unmodified peptide at the same substrate concentration) were determined. These values correlated well with the k_{cat}/K_m values.[13] The same substrate series was also used to characterize wild-type and mutant murine leukemia virus (MuLV) PR[14] as well as mutant HIV-1 proteinases.[15,16] To complement these studies, the specificity of human T-cell leukemia virus type I (HTLV-I), bovine leukemia virus (BLV) and mouse mammary tumor virus (MMTV) PRs have been also characterized by the same substrate set (our unpublished results).

The crystal structures of several PRs have been determined.[17-19] For those enzymes for which a crystal structure

Table 1. **Comparison of the naturally occurring cleavage site sequences of retroviral proteinases.**

HIV-1		*HIV-2*		*EIAV*		*AMV*	
site	*sequence*	*site*	*sequence*	*site*	*sequence*	*site*	*sequence*
MA/CA	SQNY*PIV	MA/CA	GGNY*PVQ	MA/CA	SEEY*PIM	MA/p2A	TSCY*HCG
CA/p2	ARVL*AEA	CA/p2	ARLM*AEA	CA/p1	KMML*LAK	p2A/p2B	CNCA*TAS
p2/NC	ATIM*MQR	p2/NC	PFAA*AQQ	p1/NC	AKAL*QTG	p2B/p10	PYVG*SGL
NC/p1	RQAN*FLG	NC/p1	RQAG*FLG	NC/p9	KQTF*PIQ	p10/CA	VVAM*PVV
p1/p6	PQNF*LQS	p1/p6	PRNF*PVA	TF/PR	QFVG*VTY	CA/p3	AAAM *SSA
TF/PR	SFNF*PQI	TF/PR	GLAA*PQF	PR/RT	KLVL*AQL	p3/NC	PLIM*AVV
PR/RT	TLNF*PIS	PR/RT	SLNL*PVA	RT/p15^{pol}	EEIM*LAY	NC/PR	PAVS*LAM
RT/IN	RKIL*FLD	RT/IN	RQVL*FLE	p15^{Pol}/IN	TGVF*WVE	PR'/RT	ATVL*TVA
						RT/IN	FQAY*PLR

The cleavage sites are indicated by asteriscs. Nomenclature of retroviral proteins is according to Leis et al.[21]

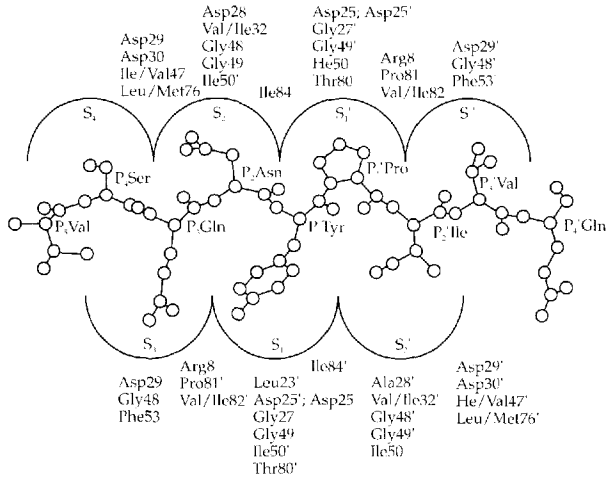

Figure 2. *A schematic representation of the HIV-1 MA/CA substrate in the S3-S3' subsites of HIV-1 and HIV-2 proteinases. The relative size of each subsite is indicated approximately as the area enclosed by the curved line around each substrate side chain. Many of the residues contribute to more than one subsite, as indicated by the position of the label. Proteinase residues forming the subsites are shown. Residues that differ between the two enzymes are given as HIV-1/HIV-2 residues.*

was not available at the time of study, as in the case of EIAV PR, we built up the structures by homologous modeling, and used these models to interpret the results of kinetic measurements.[12] Later the crystal structure of the EIAV PR revealed that substrate binding sites were properly predicted by molecular modeling, although some details of the overall structure proved to be somewhat different.[18]

Based on enzyme-inhibitor crystal structures, a substrate is also expected to bind in an extended beta conformation *(Fig.2)*. The original substrate was nine amino acid long. Sequential shortening of the substrate revealed, that 7 residues are required for efficient hydrolysis with HIV-1 and HIV-2 PRs, 4 at the N-terminal side, and 3 at the C-terminal side.[11] Some enzymes including EIAV and AMV may even interact with P5.[13] Peptides having polar or charged residues at P1 were not substrates of the studied PRs. Peptides containing small hydrophobic residues or β-branched residues were also very poor substrates or were not even hydrolyzed. The variation of relative activities of the hydrolyzed P1 substituted peptides was surprisingly small. In the given sequence context, all PRs preferred the original Tyr, and the also aromatic Phe (not shown). On the contrary, the largest variations were obtained with the P2 substituted peptides *(Fig.3)*. A common characteristic of the PRs is that they cannot hydrolyze a peptide containing Lys in this position. HIV PR prefers the original Asn, as well as the small hydrophobic Ala and Cys in this sequence context. EIAV PR follows the pattern found with HIV PRs, however, the original Asn gave a less efficient substrate than those with hydrophobic residues. AMV PR also has a similar pattern, however, the relative activities are higher for peptides having hydrophobic residues. BLV and MMTV PRs at this position seem to have a preference for larger, hydrophobic residues, getting the best values with Ile and Leu substitutions, respectively (not shown).

Asp 30, which is part of the HIV-1 S2 binding site, is a Thr in EIAV, and an uncharged residue in all the other studied PRs *(Fig.1)*. This could be one reason for the more hydrophobic nature of their S2 subsites. The other important change could be that instead of Ile 54, another residue of S2, a smaller Val is in the identical position in EIAV PR. In the case of the other PRs, except for AMV PR, which also contains Ile in this position, smaller residues can also be found in this place, which may contribute to the preference for larger hydrophobic residues at this subsite.

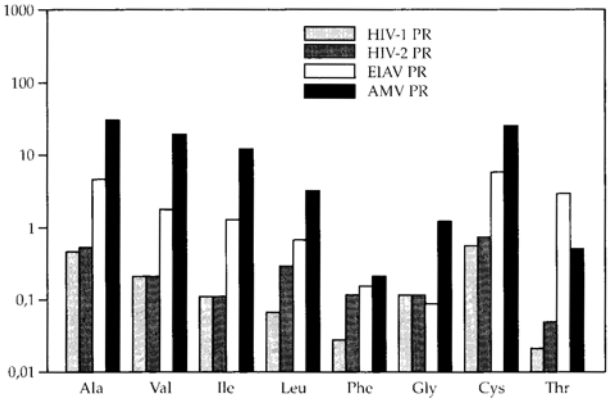

Figure 3. *Relative activities obtained for P2 Asn substituted Val-Ser-Gln-Asn-Tyr*Pro-Ile-Val-Gln substrates. The relative activity obtained with the unsubstituted peptide is 1.00 by definition.*

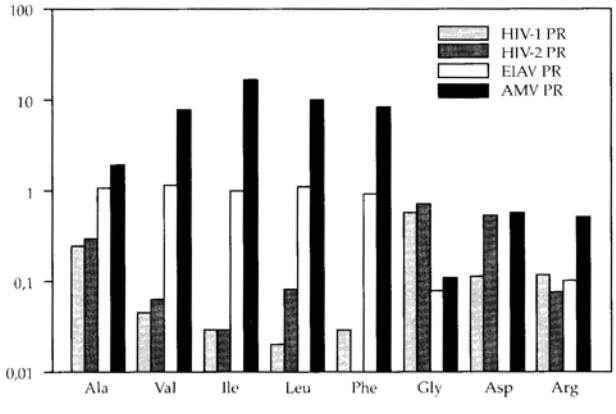

Figure 4. *Relative activities obtained for P4 Ser substituted Val-Ser-Gln-Asn-Tyr*Pro-Ile-Val-Gln substrates. The relative activity obtained with the unsubstituted peptide is 1.00 by definition.*

Crystal structures and molecular models of PRs suggested that S3 subsites are generally large. These pockets can accept various side chains, and the variation of activity is relatively small as compared to S4 and S2 (not shown). S4 is near the surface of the PRs, and the side chain may partially be exposed to the solvent. In this position HIV PRs prefer small residues. The best values were obtained with an unsubstituted substrate containing Asn and with the peptide containing a Gly substitution. In contrast, other PRs had a preference for large hydrophobic residues, suggesting that they have a more enclosed, more hydrophobic S4 pocket. The preferred size of P4 is different. For AMV, Ile substitution gave the best results while for BLV PR, Phe substitution worked most efficiently (not shown).

A common characteristic for the studied retroviral PRs is that changing the P1' Pro to any other tested amino acids gave nonhydrolyzable or very poor substrates.[9,12,13] Substitutions of P2' and P3' residues did not give as large variations in relative activities as obtained with substituting P2 and P4 residues (not shown).

In summary, it seems to be generally true, that the retroviral proteinases can recognize 6-7 amino acid residues of the substrate, generally spanning from P4 to P3'. The substrate binding sites are generally hydrophobic. Specificity of the proteinases showed the largest variations at the S2 and to a lesser degree, the S4 subsites. This site seems to be the major determinant for the differences in specificity of retroviral proteinases, and mutations harboring resistance frequently involve residues of this subsite.

Another important feature of the PR specificity is that the preference for a given subsite is strongly dependent on the residues occupying the other subsites. For example preference for P2 depends on the P1' residue,[8,9] P1 and perhaps on the P4 residue. It should be mentioned that the so called type 2 cleavage site is relatively small, but β-branched Ile is preferred at S2 of HIV PR.[20] Specificity in S3 is a function of the P1 residue: a large P1 side chain restricts the size of the P3 residue which can be accommodated.[13] Although S4 is close to the surface, it shows a preference for hydrophobic residues except for HIV PRs, although the size of the preferred residue is a function of P2. Our results suggest that the specificity of retroviral proteinases is very complex, and strongly depends on the context of the substrate sequence. The preference at a given position may depend not only on the complementarity of residues at the same side of the β-sheet (like P3 and P1, P2 and P1'), but also on those at the opposite side. However, modeling in many cases could give an explanation for the sequence context dependence and is a promising tool to "decode" the specificity of the retroviral proteinases. The strong sequence context dependence should be taken into account in the design of proteinase inhibitors, since developing resistance is one of the most serious problems in treatment of AIDS. A mutation in a substrate binding subsite of the PR indirectly could influence the specificity of the other binding sites. Conversely, changing the ligand at those other affected positions could complement the changes and regain the high potency of the enzyme-ligand interaction.

References

1. *Kräusslich, HG, Oroszlan S and Wimmer E:* Viral proteinases as targets for chemotherapy. Curr Commun Mol Biol. Cold Spring Harbor Laboratory Press, New York, 1989.
2. *Oroszlan S and Luftig RB:* Retroviral proteinases. Curr Topics Microbiol Immunol 157:153-185, 1990.
3. *Winslow DL and Otto MJ:* HIV protease inhibitors. AIDS 9 (suppl A): S183-S192, 1995.
4. *Mellors JW:* Closing in on human immunodeficiency virus-1. Nat. Medicine 2:274-275, 1996.
5. *Oroszlan S, Henderson LE, Stephenson JR et al:* Amino and carboxyl-terminal amino acid sequences of proteins coded by gag gene of murine leukemia virus. Proc Natl Acad Sci USA 75:1404-1408, 1978.
6. *Henderson LE, Benveniste RE, Sowder R et al:* Molecular characterization of gag proteins from simian immunodeficiency virus (SIV_{Mne}). J Virol 62: 2587-2595, 1988.
7. *Pettit SC, Simsic J, Loeb DD et al:* Analysis of retroviral protease cleavage sites reveals two types of cleavage sites and the structural requirements of the P1 amino acid. J Biol Chem 266:14539-14567, 1991.
8. *Griffith JT, Phylip LH, Konvalinka J et al:* Different requirements for productive interaction between the active site of HIV-1 proteinase and substrates containing -hydrophobic*hydrophobic- or -aromatic*Pro- cleavage sites. Biochemistry 31:5193-5200, 1992.
9. *Tözsér J, Weber IT, Gustchina A et al:* Kinetic and modeling studies of the S_3-S_3' subsites of HIV proteinase. Biochemistry 31:4793-4800, 1992.
10. *Schechter I and Berger A:* On the size of the active site in proteases. I. papain. Biochem Biophys Res Commun 27:157-162, 1967.
11. *Tözsér J, Gustchina A, Weber IT et al:* Studies on the role of the S_4 binding site of HIV proteinases. FEBS Lett 279:356-360, 1991.
12. *Weber IT, Tözsér J, Wu J et al:* Molecular model of equine infectious anemia virus proteinase and kinetic measurements for peptide substrates with single amino acid substitutions. Biochemistry 32:3354-3362, 1993.
13. *Tözsér J, Bagossi P, Weber IT et al:* Comparative studies on the substrate specificity of avian myeloblastosis virus proteinase and lentiviral proteinases. J Biol Chem 271:6781-6788, 1996.
14. *Menéndez-Arias L, Weber IT, Soss J et al:* Kinetic and modeling studies of subsites S_4-S_3 of Moloney murine leukemia virus protease. J Biol Chem 269:16795-16801, 1994.

15. *Bagossi P, Cheng YSE, Oroszlan S and Tözsér J:* Activity of linked HIV-1 proteinase dimers containing mutations in the active site region. Protein Eng 9:997-1003, 1996.

16. *Tözsér J, Yin FH, Cheng YSE et al:* Activity of tethered human immunodeficiency virus 1 protease containing mutations in the flap region of one subunit. Eur J Biochem 244:235-241, 1997.

17. *Wlodawer A and Erickson JW:* Structure-based inhibitors of HIV-1 protease. Ann Rev Biochem 62:543-585, 1993.

18. *Gustchina A, Kervinen J, Powell, DJ et al:* Structure of equine infectious anemia virus proteinase complexed with an inhibitor. Protein Sci 5: 1453-1465, 1996.

19. *Wlodawer A, Gustchina A, Reshetnikova L et al:* Structure of an inhibitor complex of the proteinase from feline immunodeficiency virus. Nature Stuct Biol 2:480-488, 1995.

20. *Dunn BM, Gustchina A, Wlodawer A and Kay J:* Subsite preferences of retroviral proteinases. Methods Enzymol 241, 254-278, 1994.

21. *Leis J, Baltimore D, Bishop JM et al:* Standardized and simplified nomenclature for proteins common to all retroviruses. J Virol 62:1808-1809, 1988.