JOURNAL OF VIROLOGY, Apr. 2005, p. 4213–4218 Vol. 79, No. 7
0022-538X/05/$08.00+0 doi:10.1128/JVI.79.7.4213–4218.2005

Amino Acid Preferences for a Critical Substrate Binding Subsite of Retroviral Proteases in Type 1 Cleavage Sites

Péter Bagossi, Tamás Sperka, Anita Fehér, János Kádas, Gábor Zahuczky, Gabriella Miklóssy, Péter Boross, and József Tözsér*

Department of Biochemistry and Molecular Biology, Research Center for Molecular Medicine, Medical and Health Science Center, University of Debrecen, Debrecen, Hungary

Received 15 July 2004/Accepted 31 October 2004

The specificities of the proteases of 11 retroviruses representing each of the seven genera of the family *Retroviridae* were studied using a series of oligopeptides with amino acid substitutions in the P2 position of a naturally occurring type 1 cleavage site (Val-Ser-Gln-Asn-Tyr ↓ Pro-Ile-Val-Gln; the arrow indicates the site of cleavage) in human immunodeficiency virus type 1 (HIV-1). This position was previously found to be one of the most critical in determining the substrate specificity differences of retroviral proteases. Specificities at this position were compared for HIV-1, HIV-2, equine infectious anemia virus, avian myeloblastosis virus, Mason-Pfizer monkey virus, mouse mammary tumor virus, Moloney murine leukemia virus, human T-cell leukemia virus type 1, bovine leukemia virus, human foamy virus, and walleye dermal sarcoma virus proteases. Three types of P2 preferences were observed: a subgroup of proteases preferred small hydrophobic side chains (Ala and Cys), and another subgroup preferred large hydrophobic residues (Ile and Leu), while the protease of HIV-1 preferred an Asn residue. The specificity distinctions among the proteases correlated well with the phylogenetic tree of retroviruses prepared solely based on the protease sequences. Molecular models for all of the proteases studied were built, and they were used to interpret the results. While size complementarities appear to be the main specificity-determining features of the S2 subsite of retroviral proteases, electrostatic contributions may play a role only in the case of HIV proteases. In most cases the P2 residues of naturally occurring type 1 cleavage site sequences of the studied proteases agreed well with the observed P2 preferences.

The retroviral proteases (PR), especially those of human immunodeficiency viruses (HIV), have received a great deal of attention as a target for chemotherapy (for reviews, see references 20 and 39), since the protease activity is required for the maturation and infectivity of the viruses (reviewed in references 16 and 24). Specificity studies of wild-type and mutant HIV PRs have provided a basis for the rational design of potent, selective inhibitors (6, 32, 46) and also may help to circumvent the problems caused by the rapidly developing resistance against the compounds used therapeutically (30).

Another experimental approach to better understand PR specificity is to characterize the PRs of divergent retroviruses, since several of the mutations causing drug resistance in HIV type 1 (HIV-1) PR introduce residues into the substrate binding sites found in equivalent positions in other retroviruses (15, 27). Previously, a large series of peptides containing single-amino-acid substitutions in the P4-P3′ (nomenclature according to reference 29) region of the Val-Ser-Gln-Asn-Tyr ↓ Pro-Ile-Val-Gln oligopeptide (the arrow indicates the site of cleavage) were used to characterize the specificities of the proteases of various retroviruses, including those of HIV-1, HIV-2 (38, 40), equine infectious anemia virus (EIAV) (44), Moloney murine leukemia virus (MMLV) (22), and avian myeloblastosis virus (AMV) (34). Activities relative to the unmodified peptide were determined for substituted peptides (except for MMLV PR) to compare the specificities of the enzymes. Based on these studies, the P2 position was found to be one of the most critical in determining the substrate specificity differences of retroviral proteases. Here, we report an extension of these studies of the P2 position with a protease set which now contains at least one member of each genus of the family *Retroviridae* (Table 1), including HIV-1, HIV-2, EIAV, MMLV, AMV, Mason-Pfizer monkey virus (MPMV), mouse mammary tumor virus (MMTV), human T-cell leukemia virus type 1 (HTLV-1), bovine leukemia virus (BLV), human foamy virus (HFV), and walleye dermal sarcoma virus (WDSV) proteases. Our present study has the advantage that the different retroviral proteases were mapped with the same peptide series under the same reaction conditions in the same laboratory, and therefore, the results are directly comparable. Similarities and differences in the P2 specificities of retroviral proteases were interpreted by molecular modeling using the crystal structures and homologous models of retroviral proteases.

MATERIALS AND METHODS

Retroviral proteases. Chemically synthesized HIV-2 protease was purified and refolded as described previously (40). Purified MPMV protease (the shortest, 12-kDa, form) was prepared as described previously (48). MMTV protease was expressed as a glutathione *S*-transferase fusion protein, processed with factor Xa, and purified following the published procedure (23). MMLV protease was cloned with maltose binding protein (MBP) and hexahistidine tags and purified as described previously (7). HTLV-1 protease (15), BLV protease (49), and WDSV protease (11) were purified from inclusion bodies after expression as described previously. HFV protease was cloned in fusion with MBP and used in its fusion form (10).

Oligopeptides. Oligopeptides synthesized by solid-phase peptide synthesis were described previously (21, 23, 40). Stock solutions and dilutions were made

* Corresponding author. Mailing address: Department of Biochemistry and Molecular Biology, Research Center for Molecular Medicine, Medical and Health Science Center, University of Debrecen, Nagyerdei krt. 98, P.O. Box 6, H-4012 Debrecen, Hungary. Phone: 36-52-416-432. Fax: 36-52-314-989. E-mail: tozser@indi.biochem.dote.hu.

TABLE 1. Comparison of relative activities obtained on Val-Ser-Gln-Xaa-Tyr ↓ Pro-Ile-Val-Gln peptides for retroviral proteases

P2 residue	Relative activity on protease of:										
	Lentivirus[a]			Alpharetrovirus	Betaretrovirus		Gammaretrovirus	Deltaretrovirus		Epsilonretrovirus	Spumavirus
	HIV-1	HIV-2	EIAV	AMV	MPMV	MMTV	MMLV	HTLV-1	BLV	WDSV	HFV
Ala	2.5	1.5	2.0	1.9	1.02	0.14	0.76	0	0.098	0.95	1.05
Val[b]	1	1	1	1	1	1	1	1	1	1	1
Ile	0.50	0.25	0.52	0.59	0.70	1.9	1.2	3.3	1.3	0.59	0.038
Leu	0.30	0.75	0.28	0.20	0.53	0.35	3.6	3.0	3.9	2.0	0.18
Phe	0.15	0.25	0.080	0.014	0.018	0.22	0.12	0	0.023	0.23	0.066
Gly	0.50	0.25	0.040	0.073	0.088	0	0.007	0	0.0006	0	0
Cys	4.5	2.8	2.9	1.6	0.90	0.59	1.9	0.67	0.094	2.8	1.14
Thr	0.10	0.075	1.3	0.041	0.18	0.11	0.43	0	0.042	0.18	0.063
Asn	5.0	2.5	0.40	0.045	0.18	0.015	0.24	0	0.021	2.6	0

[a] Classification is based on the latest report of the International Committee on Taxonomy of Viruses (ICTV), as detailed at the ICTV database (http://www.ncbi.nlm.nih.gov/ICTVdb).

[b] Relative activities are expressed as activity relative to that obtained with the Val-Ser-Gln-Val-Tyr ↓ Pro-Ile-Val-Gln substrate.

in distilled water (or in 5 mM dithiothreitol for the Cys-containing peptide), and the proper peptide concentrations were determined by amino acid analysis.

Protease assay. Protease assays were performed at 37°C using purified retroviral proteases and chemically synthesized oligopeptides (0.4 mM) in 0.25 M potassium phosphate buffer, pH 5.6, containing 7.5% glycerol, 5 mM dithiothreitol, 1 mM EDTA, 0.2% Nonidet P-40, and 2 M NaCl. The reaction mixtures were incubated at 37°C for 1 h, or 24 h in the cases of HFV and WDSV proteases, and the reactions were stopped by the addition of 9 volumes of 1% trifluoroacetic acid. The reaction mixtures were then injected onto a Nova-Pak C_{18} reversed-phase chromatography column (3.9 × 150 mm; Waters Associates, Inc.) using an automatic injector. The substrates and the cleavage products were separated using an acetonitrile gradient (0 to 100%) in water in the presence of 0.05% trifluoroacetic acid. The cleavage of peptides was monitored at 206 nm, and the peak areas were integrated. Amino acid analysis of the collected peaks was used to confirm the site of cleavage with HIV-1 PR (40). For the other retroviral proteases, the cleavage products were identified by the retention time, which was found to be identical to that obtained with HIV-1 PR. Relative activities were calculated from the molar amounts of peptides cleaved per unit time at <20% substrate turnover by dividing the activity on a given peptide by the activity on the Val-Ser-Gln-Val-Tyr ↓ Pro-Ile-Val-Gln substrate, which has the smallest hydrocarbon side chain at the P2 site among the peptides that were hydrolyzable by all of the studied proteases, as described in reference 1. Measurements were performed in duplicate, and the average values were calculated. The standard error was <10%. The relative activities for the HIV-1 PR (40), EIAV PR (44), and AMV PR (34) have been reported previously. The published values were converted to values relative to our reference substrate. Previous studies also indicated a strong correlation between the relative activities and the specificity constants (34); therefore, the determined activity values can be considered a measure of the k_{cat}/K_m values.

For peptides representing naturally occurring cleavage sites, at least six substrate concentrations were tested, selected based on the approximate K_m values. Amino acid analysis of the collected product peaks was used to confirm the sites of cleavages and to establish the relationship between integration values and product quantities. Kinetic parameters were determined by fitting the data obtained at <20% substrate hydrolysis to the Michaelis-Menten equation by using the Fig. P program (Fig. P Software Corp.). The standard errors of the kinetic parameters were <20%.

Molecular modeling. Structure-based alignment of the HIV-1, HIV-2, simian immunodeficiency virus, EIAV, feline immunodeficiency virus (FIV), and Rous sarcoma virus (RSV) PRs was used as a template for the alignment of BLV, HTLV-1, MPMV, MMTV, MMLV, HFV, and WDSV protease sequences, as shown in Fig. 1. Structural alignment was done by WHAT IF (43), and the initial multiple sequence alignment was made by ClustalW (31) and was corrected by hand based on the structural alignment. The phylogenetic tree was made by ClustalW and Phylip (9).

The program Modeller (28) was used to build the initial models of all studied retroviral proteases. The program allows the use of multiple template crystal structures as an input and can create multiple homologous models as an output. We used the highest-resolution (1.2- to 2.4-Å) crystal structures of HIV-1 (18), HIV-2 (33), EIAV (13), FIV (14), and RSV (47) proteases (either as wild-type or mutant forms) available at the time of modeling as templates (Protein Data Bank accession codes [and resolutions]: 1K1T [1.2 Å], 1IDA [1.7 Å], 1FMB [1.8 Å], 4FIV [1.8 Å], and 1BAI [2.4 Å], respectively) and generated three simultaneous models for each protease. RSV protease structure was used to interpret the AMV protease results: AMV and RSV proteases differ in only two residues, which are not expected to be involved in the enzyme-substrate interactions (34). A model of the QVY ↓ PIV hexapeptide was docked into the substrate binding site of each retroviral protease model in forward and reverse directions, and a water molecule was also placed in a conserved position between the flaps of the protease and the substrate. This sequence has the smallest hydrocarbon side chain at the P2 site among the peptides that were hydrolyzable by all of the studied proteases, and the appropriate substrate was also used as a reference substrate for activity measurements.

Unfavorable van der Waals interactions were removed by short minimizations using Sybyl (Tripos Inc., St. Louis, Mo.) with the following parameters: Kollman all-atom force field (45) as implemented in Sybyl, 8 Å cutoff; 20 Simplex and 100 Powell iterations with distance constraints between the carboxyl groups of the two catalytic aspartates and between the H bond donor and acceptor atoms of the enzyme and the substrate in the P3-P3′ region to maintain the conserved H bond network of the substrate binding site. Another 100 Powell iterations were applied with distance constraints only between the catalytic aspartates.

The structures were examined on a Silicon Graphics Indigo2 or O2 computer graphics system using the program Sybyl or WHAT IF. Root mean square (RMS) deviations were calculated by the structural superposition module of WHAT IF. Cavities were calculated on the minimized structures containing P2 Gly substrate using the SiteID module of Sybyl, and they were averaged on three models and two directions for each enzyme. At least three values were included in the averaging procedure for each enzyme. The volume of the amino acid residues was retrieved from the literature (50). The residues forming the S2 subsites (Fig. 1) were predicted previously for HIV-1 and HIV-2 PRs (40); EIAV (44), AMV (34), MMLV (2), and HTLV-1 (41) proteases and the corresponding residues in BLV, MMTV, MPMV, HFV, and WDSV proteases were obtained from the sequence alignment in Fig. 1 and verified using the crystal structures and homologous models.

RESULTS AND DISCUSSION

It is a special feature of the retroviral proteases that it is not possible to give consensus substrate sequences for them, even though they are fairly specific. Retroviral protease cleavage sites are presently classified into two groups. Type 1 cleavage sites have the aromatic residue Pro, and type 2 sites have hydrophobic residues (excluding Pro) at the site of cleavage (12, 25, 40). The P2 position is also critical in determining the type of cleavage site (12, 40). In type 1 cleavage sites of primate lentiviruses, like HIV-1, there is a preference for Asn at P2, while in type 2 cleavage sites, the P2 position is typically β branched. The type 1 cleavage site is very important for several reasons. No protease except pepsin is known to act at the imino side of a Pro residue. Proline residues, especially after Tyr or

```
                    1         10                  20        30              40          50
                    |         |                   |         |               |           |
HIV-1     ------------PQITLW--QRPLVTIRIG----------GQLKEALLDTGADDTVLEE--M--N--LPGKWK--PKMIGGIGGFIKVRQY
HIV-2     ------------PQFSLW--KRPVVTAHIE----------GQPVEVLLDTGADDSIVAG--I--E--LGNNYS--PKIVGGIGGFINTKEY
SIV       ------------PQFSLW--RRPVVTAHIE----------GQPVEVLLDTGADDSIVTG--I--E--LGPHYT--PKIVGGIGGFINTKEY
EIAV      ------------VTYNLE--KRPTTIVLIN----------DTPLNVLLDTGADTSVLTTAHYNRLKYRGRKYQ--GTGIGGVGGNVETFST
FIV       -------YNKVGTTTTLE--KRPEILIFVN----------GYPIKFLLDTGADITILNRRDFQ-VKN-SIENG--RQNMIGVGGGKRGTNY
RSV       ------------LAMTMEHKDRPLVRVILTNTGSHPVKQRSVYITALLDSGADITIISEEDWP---TDWPVMEAANPQIHGIGGGIPMRKS
HTLV-1    ------------PVIPLDPARRPVIKAQVDTQTSHP-----KTIEALLDTGADMTVLPIALFS----SNTPLK--NTSVLGAGGQTQDHFK
HTLV-2    ------------PLIPLRQQQQPILGVRISVMGQTP-----QPTQALLDTGADLTVIPQTLVP----GPVKLH--DTLILGASGQTNTQFK
BLV       ------------LSIPLA-RSRPSVAVYLSGPWLQP---SQNQALMLVDTGAENTVLPQNWLV----RDYPRI--PAAVLGAGGVSPNRYN
MMLV      -----TLDDQGGQGQDPP--PEPRITLKVG----------GQPVTFLVDTGAQHSVLTQ--N--P--GPLSDK--SAWVQGATGGKRYRWT
WEHV-1    ----PSAQSVGCQGAQVCTNREAEILVKVN----------NTIIPMLVDTGACLTAIGGSAPIVP-KLKLTDT--TAYAVGISSDPVEHIK
WEHV-2    ----PQEQDVDCQGAQVYANQDAEVMVTLN----------GTQIPMLVDTGACLTAIGGAATKVP-ELTLTHK--TAYAMGISAEPVEHIK
WDSV      PIDCPYEKSGTKTTQDVITTKNAEIMVTVN----------HTKIPMLVDTGACLTAIGGAATVVP-DLKLTNT--EIIAVGISAEPVPHVL
MMTV      ------------WVQEISD-SRPMLHISL----------NGRRFLGLLDTGADKTCIAGRDWP---ANWPIHQTESSLQGLVGMACGVARS
MPMV      ------------WVQPITC-QKPSLTLWL----------DDKMFTGLIDTGADVTIIKLEDWP---PNWPITDTLTNLRG-IGQSNNPKQS
HFV       ---------MNPLQLL-----QPLPAEIK-----------GTKLLAHWDSGATITCIPESFLEDEQP--I--K--KTLIKTIHGEKQQNVY
EFV       ------------MQAL-----QPLQVQIK-----------GNSLKGFYDTGAEITCVPAIFLIEEEP--I--G--ERTIQTIHGITKEKVY
FFV       ------------MDLL-----KPLTVERK-----------GVKIKGYWDSQADITCVPKDLLQGEEP--V--R--QQNVTTIHGTQEGDVY
BFV       ------------MPAL-----RPLQVEIK-----------GNHLKGYWDSGAEITCVPAIYIIEEQP--V--G--KKLITTIHNEKEHDVY

                    60              70              80          90
                    |               |               |           |
HIV-1     --DQIPVEICG------HKAIGTVLVG----PTPVNIIGRNLLTQIGCTLNF------------------
HIV-2     --KNVEIEVLN------KKVRATIMTG----DTPINIFGRNILTALGMSLNL------------------
SIV       --KNVEIEVLG------KRIKGTIMTG----DTPINIFGRNILTALGMSLNL------------------
EIAV      --P-VTIKKKG------RHIKTRMLVA----DIPVTILGRDILQDLGAKLVL------------------
FIV       --INVHLEIRDENYKT-QCIFGNVCVL-EDNSLIQPLLGRDNMIKFNIRLVMAQ----------------
RSV       -RDMIELGVINRDGSLERPLLLFPAVA----MVRGSILGRDCLQGLGLRLTNL------------------
HTLV-1    -LTSLPVLIRLPFRT-TPIVLTSCLV---DTKNNWAIIGRDALQQCQGVLYLPEAKGPPVIL--------
HTLV-2    -LLQTPLHIFLPFRR-SPVILSSCLL---DTHNKWTIIGRDALQQCQGLLYLPDDPSPHQLL--------
BLV       -WLQGPLTLALKPEG-PFITIPKILV---DTFDKWQILGRDVLSRLQASISIPEEVRPPMVG--------
MMLV      --TDRKVHLAT------GKVTHSFLHV---PDCPYPLLGRDLLTKLKAQIHFEGSGAQVMGPMGQPLQVL
WEHV-1    -TEQVKVKVKD------ASCNIEPWFN---KDQTFHILGRDTLTKMRAMIAFLPCGRMEVVFPARYHQN-
WEHV-2    -ANKIRLQIDD------ASCSIQPWYN---KNQTFHILGRDTLSKMRAIIHFHPDGEMEVVFPAKYHQL-
WDSV      -AKPTKIQIEN------TNIDISPWYN---PDQTFHILGRDTLSKMRAIVSFEKNGEMTVLLPPTYHKQL
MMTV      -SQPLRWQH-EDK-SGIIHPFVIPTL-------PFTLWGRDIMKEILVRLMTDSPDDSQDL---------
MPMV      -SKYLTWRDKENN-SGLIKPFVIPNL-------PVNLWGRDLLSQMKIMMCS------------------
HFV       -YVTFKVKGRKVE----AEVIASPYEYILLSPTDVPWLTQQPLQLTILVPLQEYQEKILSKTALPEDQKQ
EFV       -YLTFKIQGRKLA----AEVIGTQLDYVIIAPSDIPWYKKYELELTIKIDIQKQQEQLLHTTNLSSEGKK
FFV       -YVNLKIDGRRIN----TEVIGTTLDYAIITPGDVPWILKKPLELTIKLDLEEQQGTLLNNSILSKKGKE
BFV       -YVEMKIEKRKVQ----CEVIATALDYVLVAPVDIPWYKPGPLELTIKIDVESQKHTLITESTLSPQG--
```

FIG. 1. Sequence alignment of the retroviral proteases. Residues that are proven or predicted to form S2 subsites are shown in boldface.

Phe in the sequence (as is the case in type 1 cleavage sites), have a relatively high probability of forming the *cis* isomer rather than the *trans* isomer of the preceding peptide bond (17). Conformational selectivity of the HIV-1 PR toward the *trans* isomer of the cleaved peptide bond was demonstrated by nuclear magnetic resonance and kinetic studies (19, 42).

Previously, we characterized the specificities of HIV-1, HIV-2, EIAV, and AMV proteases using an oligopeptide substrate series based on the naturally occurring type 1 cleavage site (Val-Ser-Gln-Asn-Tyr ↓ Pro-Ile-Val-Gln) between the matrix (MA) and capsid (CA) proteins of HIV-1 (34, 40, 44). In this study, we extended these studies to the proteases of MPMV, MMTV, MMLV, HTLV-1, BLV, HFV, and WDSV to have at least one member of each genus of the retroviruses (Table 1).

Replacement of the P2 Asn in the original peptide with medium-size hydrophobic residues resulted in peptides that were well hydrolyzed by most proteases (Table 1). The largest effects were observed with HTLV-1 and HFV proteases, in which the original nonhydrolyzable peptide was converted to hydrolyzable ones, and with BLV PR, in which the best substrate was the P2 Leu-containing peptide, which was an almost 200-fold better substrate than the unmodified peptide. P2 Leu-containing peptide was also the best substrate for MMLV PR, as described in a previous study (22). The peptide with an Ile substitution was the best substrate for MMTV and HTLV-1 proteases. The small hydrophobic amino acid Ala was preferred by AMV and MPMV proteases. EIAV PR preferred the slightly polar Cys residue, while HIV-1 PR preferred the more polar Asn residue, giving the best value with the original peptide (Table 1). Although in the cases of HIV-2 and WDSV proteases, Cys was the best residue in this position, a similar value was obtained for Asn (Table 1). HFV PR showed equally high activities, within the experimental error of the measurement, on substrates containing Ala, Val, or Cys at the P2 site.

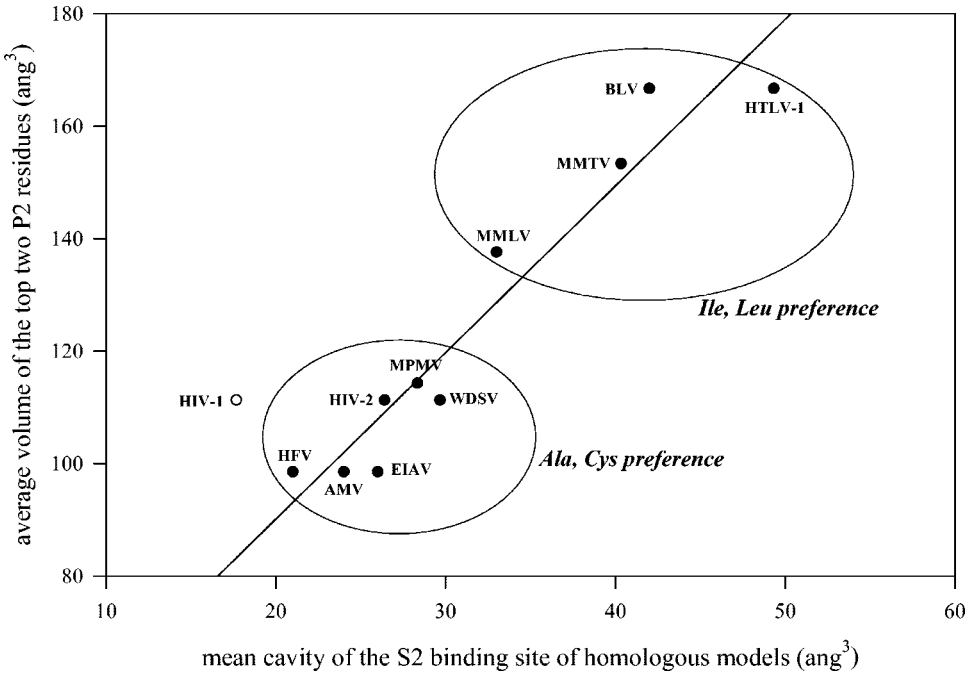

FIG. 2. Mean cavity volumes of S2 subsites of homologous models of various retroviral proteases versus averaged volume of the best two P2 residues, in which the measured relative activity was the highest. ang, angstrom.

Therefore, in this sequence context, the preference for Asn by the HIV-1 PR is rather exceptional, since the other retroviral proteases studied showed higher preferences for hydrophobic or less polar residues. It seems to be a common result for retroviral proteases that Phe or Gly at P2 of this peptide series forms a poor substrate: molecular modeling suggested that the Phe side chain is too bulky, while Gly is too small, for this subsite.

To understand the similarities and differences among the specificities of retroviral proteases, molecular models were built for all of the studied enzymes using the same procedure. The goodness of the models was estimated by calculating their RMS deviations from the five high-resolution crystal structures used as templates. Naturally, the smallest values (0.3 to 0.5 Å) were calculated for HIV-1, HIV-2, EIAV, and RSV models from their own crystal structure templates, and their range showed similar values, as can be seen in comparison to different crystallographic determinations of the same enzyme (4). Other models showed similar RMS deviations (0.9 to 2.9 Å) from the values calculated between the templates themselves (1.0 to 2.7 Å), and they also showed similar relationships between the RMS deviation and the sequence homology, as can be seen in a comparison of crystal structures (4).

Based on the molecular models, the S2 binding sites of retroviral proteases are usually small hydrophobic pockets. S2 is sterically more restricted than the other subsites, especially S4 and S3. About half of the subsite-forming residues of S2 are conserved (Fig. 1), however, in place of Asp 30 of HIV-1 PR and HIV-2 PR, which appears to be the reason why these enzymes could accommodate more polar residues in this subsite (3, 40); other enzymes contain hydrophobic or hydrophilic residues, but pointing outward from the pocket. At the tip of the flap, most enzymes contain branched side chains (Ile, Val, or Leu), but MMLV, BLV, and HTLV-1 PRs contain smaller Ala residues, which may be at least partly responsible for the greater Ile and Leu preference at the S2 sites of these enzymes. Previously, we observed that the two residues that seemed to be crucial in determining the preference for Val over Leu at P2 in our substrate series were residues of the retroviral proteases corresponding to Val 32 and Ile 47 of HIV-1 (34): in both positions, Ile favors Val at P2 in the substrate, while Val favors Leu. BLV, HTLV-1, and MMLV PRs have Val at both positions, and the relative preference for Leu over Val is in good agreement with the previously proposed rule (34). This suggests that size complementarities are the main specificity-determining features of the S2 subsites of retroviral PRs, and electrostatic contributions play a role only in the cases of proteases of primate lentiviruses (HIVs and simian immunodeficiency viruses). To illustrate this suggestion, cavity volumes were calculated in the cases of P2 Gly substrate complexes and plotted against the averaged volume of the best two P2 residues measured experimentally (Fig. 2). The data points (except that of HIV-1 PR) showed linear correlations ($r = 0.96$) and were clustered into two regions of the plot according to the P2 preferences of the enzymes. This correlation was further supported by phylogenetic analysis of the sequences of retroviral proteases (Fig. 3).

Naturally occurring type 1 cleavage site sequences of the studied retroviruses (Table 2) have P2 residues that are usually in good agreement with the findings of the S2 mapping study (Table 1). Asn can be found at the P2 position in five out of six cleavage sites of HIV-1 and HIV-2, and the peptide with P2 Leu showed a very low cleavage efficiency. The S2 specificity of EIAV PR appeared to show a transition between the specificity of primate lentiviruses (preference for polar residues) and the alpha-, beta-, and gammaretroviruses (preference for small hydrophobic residues): the preferred residues are Cys and Ala

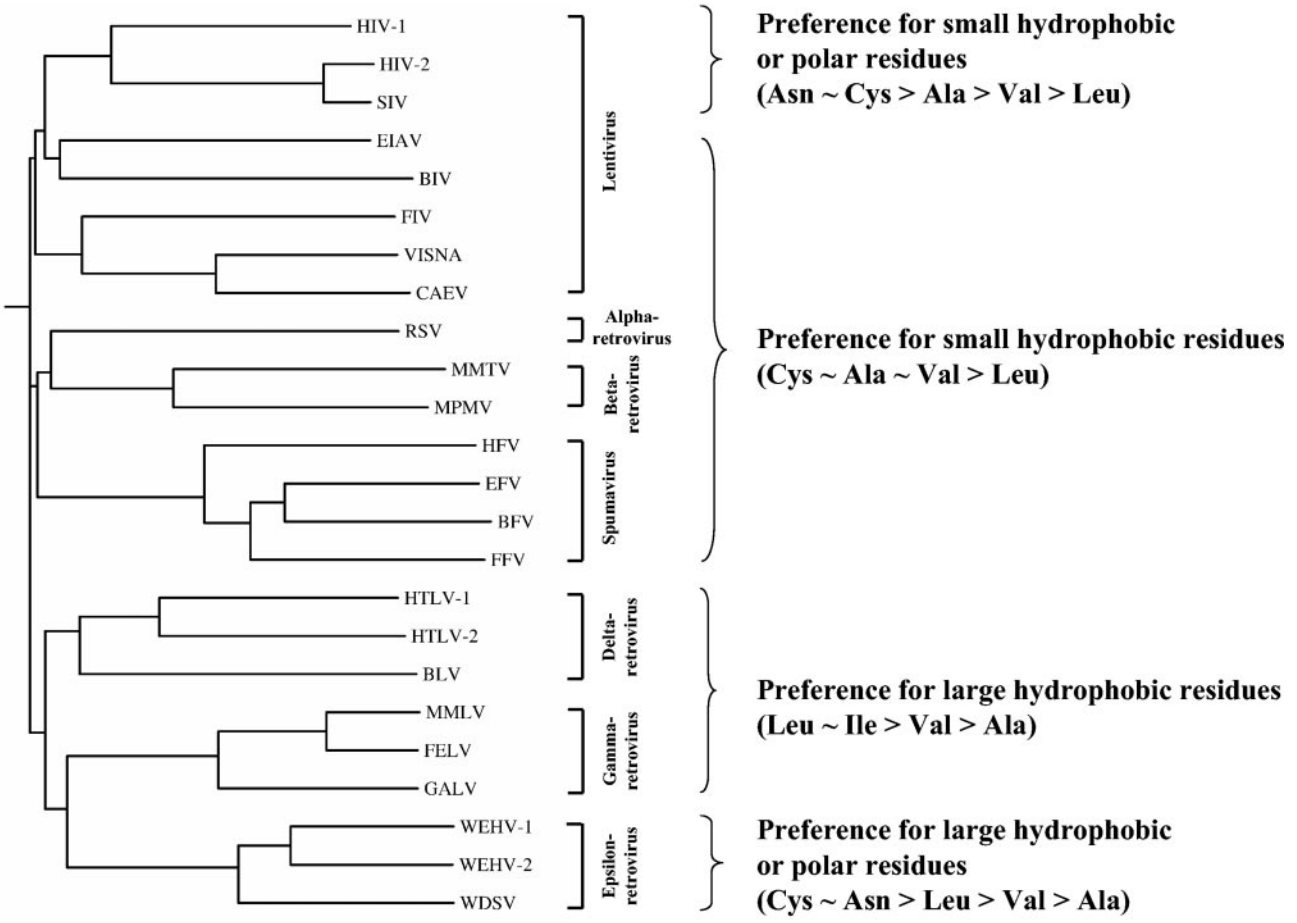

FIG. 3. Phylogenetic tree of the retroviral proteases. The distinct specificity subgroups are also indicated.

in the context of the VSQXY ↓ PIVQ sequence, while Glu and Thr are found in the natural cleavage sites of EIAV. The best P2 residue (Ala) observed for AMV protease in the mapping study appears in the P2 position of its naturally occurring type 1 cleavage site sequence; however, the MPMV and MMTV sites contain the less favorable Ile and Thr, respectively, at this position (Table 2). In one of the MMLV cleavage sites, the most preferred Leu appears in P2, while the other contains a less preferred Ala. Discrepancies between the most preferred P2 residues and those observed in the naturally occurring cleavage sites might be due to the different sequence contexts, which have been shown to have a profound effect on the subsite preference of HIV-1 PR, including S2 (35), as well as to the possibility that not all retroviral cleavage sites are optimized evolutionarily for rapid processing (8).

The importance of the proper van der Waals interactions and filling in the S2 binding site is also demonstrated by mutations appearing in drug resistance. One of the slowest HIV-1 cleavage sites, the NC/p1 cleavage site (Arg-Gln-Ala-Asn ↓ Phe-Leu-Gly-Lys) contains a small Ala residue at P2, which does not fit optimally into this pocket, as demonstrated by kinetic, modeling (8), and crystallographic studies (26); this residue is frequently mutated to Val in drug resistance (5) to provide a substantially better fit.

One of the main problems in the protease inhibitor therapy of AIDS is the development of resistance to the drugs designed against HIV-1 PR. Many of the mutations occurring in drug resistance produce residues that can be found in other retroviral proteases. Therefore, understanding the specificity simi-

TABLE 2. Type 1 (−Phe/Tyr ↓ Pro−) natural cleavage sites of the studied retroviral proteases

Virus[a]	Cleavage site[b]	Oligopeptide sequence	k_{cat}/K_m (mM^{-1} s^{-1})	Reference
HIV-1	MA/CA	VSQNY ↓ PIVQ	45.3	36
	in p6	DKELY ↓ PLTSL	0.02	
	TF/PR	VSFNF ↓ PQITL	6.9	
	PR/RT	CTLNF ↓ PISP	24.1	
HIV-2	MA/CA	EKGGNY ↓ PVQHV	3.7	36
	NC/p6	KPRNF ↓ PVAQV	1.0	
EIAV	MA/CA	PSEEY ↓ PIMID	76.3	37
	NC/p9	QKQTF ↓ PIQQK	38.5	
AMV	RT/IN	TFQAY ↓ PLREA	0.015	34
MPMV	p12/CA	PKDIF ↓ PVTET	30.6	This study
MMTV	n/CA	LTFTF ↓ PVVFMRR	6.3	This study
MMLV	MA/p12	PRSSLY ↓ PALTP	2.0	This study
	p12/CA	TSQAF ↓ PLRAG	1.7	

[a] There is no known type 1 cleavage site for HTLV-1, BLV, WDSV, or HFV. All measurements were performed under identical assay conditions.

[b] Abbreviations of viral proteins: MA, matrix; CA, capsid; NC, nucleocapsid; PR, protease; RT, reverse transcriptase; IN, integrase. Other proteins are shown as p followed by the molecular mass of that protein in kilodaltons. In MMTV the small peptide (n) presumably located upstream of CA has not been identified.

larities and differences of these enzymes may help in the design of broad-spectrum inhibitors against HIV-1 PR.

ACKNOWLEDGMENTS

This research was supported by the Hungarian Science and Research Fund (OTKA T22670, T43482, and F34479) and by the János Bolyai Fellowship of the Hungarian Academy of Sciences (to Péter Bagossi).

We thank Stephen Oroszlan for providing the MMTV protease clone, John M. Louis for the HIV-1 protease clone, Iva Pichova for the MPMV protease clone, and Volker M. Vogt for the WDSV protease clone. Ivo Bláha is acknowledged for the synthesis of the peptides. We thank Szilvia Petö for technical assistance.

REFERENCES

1. **Blaha, I., J. Tözsér, Y. Kim, T. D. Copeland, and S. Oroszlan.** 1992. Solid phase synthesis of the proteinase of bovine leukemia virus. Comparison of its specificity to that of HIV-2 proteinase. FEBS Lett. **309:**389–393.
2. **Boross, P., P. Bagossi, T. D. Copeland, S. Oroszlan, J. M. Louis, and J. Tözsér.** 1999. Effect of substrate residues on the P2′ preference of retroviral proteinases. Eur. J. Biochem. **264:**921–929.
3. **Cameron, C. E., B. Grinde, P. Jacques, J. Jentoft, J. Leis, A. Wlodawer, and I. T. Weber.** 1993. Comparison of the substrate-binding pockets of the Rous sarcoma virus and human immunodeficiency virus type 1 proteases. J. Biol. Chem. **268:**11711–11720.
4. **Chothia, C., and A. M. Lesk.** 1986. The relation between the divergence of sequence and structure in proteins. EMBO J. **5:**823–826.
5. **Doyon, L., G. Croteau, D. Thibeault, F. Poulin, L. Pilote, and D. Lamarre.** 1996. Second locus involved in human immunodeficiency virus type 1 resistance to protease inhibitors. J. Virol. **70:**3763–3769.
6. **Dunn, B. M., A. Gustchina, A. Wlodawer, and J. Kay.** 1994. Subsite preferences of retroviral proteinases. Methods Enzymol. **241:**254–278.
7. **Fehér, A., P. Boross, T. Sperka, S. Oroszlan, and J. Tözsér.** 2004. Expression of the murine leukemia virus protease in fusion with maltose-binding protein in *Escherichia coli*. Protein Expr. Purif. **35:**62–68.
8. **Fehér, A., I. T. Weber, P. Bagossi, P. Boross, B. Mahalingam, J. M. Louis, T. D. Copeland, I. Y. Torshin, R. W. Harrison, and J. Tözsér.** 2002. Effect of sequence polymorphism and drug resistance on two HIV-1 Gag processing sites. Eur. J. Biochem. **269:**4114–4120.
9. **Felsenstein, J.** 1989. PHYLIP—Phylogeny Inference Package. Cladistics **5:** 164–166.
10. **Fenyöfalvi, G., P. Bagossi, T. D. Copeland, S. Oroszlan, P. Boross, and J. Tözsér.** 1999. Expression and characterization of human foamy virus proteinase. FEBS Lett. **462:**397–401.
11. **Fodor, S. K., and V. M. Vogt.** 2002. Characterization of the protease of a fish retrovirus, walleye dermal sarcoma virus. J. Virol. **76:**4341–4349.
12. **Griffiths, J. T., L. H. Phylip, J. Konvalinka, P. Strop, A. Gustchina, A. Wlodawer, R. J. Davenport, R. Briggs, B. M. Dunn, and J. Kay.** 1992. 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.
13. **Gustchina, A., J. Kervinen, D. J. Powell, A. Zdanov, J. Kay, and A. Wlodawer.** 1996. Structure of equine infectious anemia virus proteinase complexed with an inhibitor. Protein Sci. **5:**1453–1465.
14. **Kervinen, J., J. Lubkowski, A. Zdanov, D. Bhatt, B. M. Dunn, K. Y. Hui, D. J. Powell, J. Kay, A. Wlodawer, and A. Gustchina.** 1998. Toward a universal inhibitor of retroviral proteases: comparative analysis of the interactions of LP-130 complexed with proteases from HIV-1, FIV, and EIAV. Protein Sci. **7:**2314–2323.
15. **Louis, J. M., S. Oroszlan, and J. Tözsér.** 1999. Stabilization from autoproteolysis and kinetic characterization of the human T-cell leukemia virus type 1 proteinase. J. Biol. Chem. **274:**6660–6666.
16. **Louis, J. M., I. T. Weber, J. Tözsér, G. M. Clore, and A. M. Gronenborn.** 2000. HIV-1 protease: maturation, enzyme specificity, and drug resistance. Adv. Pharmacol. **49:**111–146.
17. **MacArthur, M. W., and J. M. Thornton.** 1991. Influence of proline residues on protein conformation. J. Mol. Biol. **218:**397–412.
18. **Mahalingam, B., P. Boross, Y. F. Wang, J. M. Louis, C. C. Fischer, J. Tözsér, R. W. Harrison, and I. T. Weber.** 2002. Combining mutations in HIV-1 protease to understand mechanisms of resistance. Proteins **48:**107–116.
19. **McCornack, M. A., L. T. Kakalis, C. Caserta, R. E. Handschumacher, and I. M. Armitage.** 1997. HIV protease substrate conformation: modulation by cyclophilin A. FEBS Lett. **414:**84–88.
20. **Mellors, J. W.** 1996. Closing in on human immunodeficiency virus-1. Nat. Med. **2:**274–275.
21. **Menendez-Arias, L., D. Gotte, and S. Oroszlan.** 1993. Moloney murine leukemia virus protease: bacterial expression and characterization of the purified enzyme. Virology **196:**557–563.
22. **Menendez-Arias, L., I. T. Weber, J. Soss, R. W. Harrison, D. Gotte, and S. Oroszlan.** 1994. Kinetic and modeling studies of subsites S4–S3′ of Moloney murine leukemia virus protease. J. Biol. Chem. **269:**16795–16801.
23. **Menendez-Arias, L., M. Young, and S. Oroszlan.** 1992. Purification and characterization of the mouse mammary tumor virus protease expressed in *Escherichia coli*. J. Biol. Chem. **267:**24134–24139.
24. **Oroszlan, S., and R. B. Luftig.** 1990. Retroviral proteinases. Curr. Top. Microbiol. Immunol. **157:**153–185.
25. **Pettit, S. C., J. Simsic, D. D. Loeb, L. Everitt, C. A. Hutchison III, and R. Swanstrom.** 1991. 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**:14539–14547.
26. **Prabu-Jeyabalan, M., E. Nalivaika, and C. A. Schiffer.** 2002. Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes. Structure **10:**369–381.
27. **Ridky, T., and J. Leis.** 1995. Development of drug resistance to HIV-1 protease inhibitors. J. Biol. Chem. **270:**29621–29623.
28. **Sali, A., and T. L. Blundell.** 1993. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. **234:**779–815.
29. **Schechter, I., and A. Berger.** 1967. On the size of the active site in proteases. I. Papain. Biochem. Biophys. Res. Commun. **27:**157–162.
30. **Swanstrom, R., and J. Eron.** 2000. Human immunodeficiency virus type-1 protease inhibitors: therapeutic successes and failures, suppression and resistance. Pharmacol. Ther. **86:**145–170.
31. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.
32. **Tomasselli, A. G., and R. L. Heinrikson.** 1994. Specificity of retroviral proteases: an analysis of viral and nonviral protein substrates. Methods Enzymol. **241:**279–301.
33. **Tong, L., S. Pav, S. Mui, D. Lamarre, C. Yoakim, P. Beaulieu, and P. C. Anderson.** 1995. Crystal structures of HIV-2 protease in complex with inhibitors containing the hydroxyethylamine dipeptide isostere. Structure **3:**33–40.
34. **Tözsér, J., P. Bagossi, I. T. Weber, T. D. Copeland, and S. Oroszlan.** 1996. Comparative studies on the substrate specificity of avian myeloblastosis virus proteinase and lentiviral proteinases. J. Biol. Chem. **271:**6781–6788.
35. **Tözsér, J., P. Bagossi, I. T. Weber, J. M. Louis, T. D. Copeland, and S. Oroszlan.** 1997. Studies on the symmetry and sequence context dependence of the HIV-1 proteinase specificity. J. Biol. Chem. **272:**16807–16814.
36. **Tözsér, J., I. Blaha, T. D. Copeland, E. M. Wondrak, and S. Oroszlan.** 1991. Comparison of the HIV-1 and HIV-2 proteinases using oligopeptide substrates representing cleavage sites in Gag and Gag-Pol polyproteins. FEBS Lett. **281:**77–80.
37. **Tözsér, J., D. Friedman, I. T. Weber, I. Blaha, and S. Oroszlan.** 1993. Studies on the substrate specificity of the proteinase of equine infectious anemia virus using oligopeptide substrates. Biochemistry **32:**3347–3353.
38. **Tözsér, J., A. Gustchina, I. T. Weber, I. Blaha, E. M. Wondrak, and S. Oroszlan.** 1991. Studies on the role of the S4 substrate binding site of HIV proteinases. FEBS Lett. **279:**356–360.
39. **Tözsér, J., and S. Oroszlan.** 2003. Proteolytic events of HIV-1 replication as targets for therapeutic intervention. Curr. Pharm. Des. **9:**1803–1815.
40. **Tözsér, J., I. T. Weber, A. Gustchina, I. Blaha, T. D. Copeland, J. M. Louis, and S. Oroszlan.** 1992. Kinetic and modeling studies of S3–S3′ subsites of HIV proteinases. Biochemistry **31:**4793–4800.
41. **Tözsér, J., G. Zahuczky, P. Bagossi, J. M. Louis, T. D. Copeland, S. Oroszlan, R. W. Harrison, and I. T. Weber.** 2000. Comparison of the substrate specificity of the human T-cell leukemia virus and human immunodeficiency virus proteinases. Eur. J. Biochem. **267:**6287–6295.
42. **Vance, J. E., D. A. LeBlanc, P. Wingfield, and R. E. London.** 1997. Conformational selectivity of HIV-1 protease cleavage of X-Pro peptide bonds and its implications. J. Biol. Chem. **272:**15603–15606.
43. **Vriend, G.** 1990. WHAT IF: a molecular modeling and drug design program. J. Mol. Graph. **8:**52–56.
44. **Weber, I. T., J. Tözsér, J. Wu, D. Friedman, and S. Oroszlan.** 1993. Molecular model of equine infectious anemia virus proteinase and kinetic measurements for peptide substrates with single amino acid substitutions. Biochemistry **32:**3354–3362.
45. **Weiner, S. J., P. A. Kollman, D. T. Nguyen, and D. A. Case.** 1986. An all atom force field for simulations of proteins and nucleic acids. J. Comput. Chem. **7:** 230–252.
46. **Wlodawer, A.** 2002. Rational approach to AIDS drug design through structural biology. Annu. Rev. Med. **53:**595–614.
47. **Wu, J., J. M. Adomat, T. W. Ridky, J. M. Louis, J. Leis, R. W. Harrison, and I. T. Weber.** 1998. Structural basis for specificity of retroviral proteases. Biochemistry **37:**4518–4526.
48. **Zabransky, A., M. Andreansky, O. Hruskova-Heidingsfeldova, V. Havlicek, E. Hunter, T. Ruml, and I. Pichova.** 1998. Three active forms of aspartic proteinase from Mason-Pfizer monkey virus. Virology **245:**250–256.
49. **Zahuczky, G., P. Boross, P. Bagossi, G. Emri, T. D. Copeland, S. Oroszlan, J. M. Louis, and J. Tözsér.** 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:**1–8.
50. **Zamyatin, A. A.** 1972. Protein volume in solution. Prog. Biophys. Mol. Biol. **24:**107–123.