Amino acid preferences of retroviral proteases for amino-terminal positions in a type-1 cleavage site

Helga Eizert[1], Péter Bagossi[1], Tamás Sperka[1], Gabriella Miklóssy[1], Pálma Bander[1], Péter Boross[1], Irene T. Weber[2] and József Tözsér[1*]

[1]Department of Biochemistry and Molecular Biology, Research Center for Molecular Medicine, Medical and Health Science Center, University of Debrecen, Debrecen, Egyetem tér 1, Life Science Building, Hungary, and [2]Department of Biology, Molecular Basis of Disease Program, Georgia State University, Atlanta, GA, USA

[*]Corresponding author; Tel: +36-52-416432; fax: +36-52-314989; e-mail: tozser@dote.hu

Running title: Substrate specificity of retroviral proteases.

Word count for the abstract: 248

Word count for the text: 2761

Keywords: HIV protease; retroviral protease; oligopeptide substrates; substrate specificity; enzyme kinetics; molecular modeling.

2008.02.22.

Abstract

The specificity of the proteases of eleven retroviruses was studied using a series of oligopeptides having amino acid substitutions in P1, P3 and P4 positions of a naturally occurring type-1 cleavage site (Val-Ser-Gln-Asn-Tyr↓Pro-Ile-Val-Gln) in human immunodeficiency virus type-1 (HIV-1). Previously, substrate specificity for P2 site was studied of the same representative set of retroviral proteases, which included at least one member from each of the seven *genera* of the family *Retroviridae* (Bagossi, P., T. Sperka, A. Feher, J. Kadas, G. Zahuczky, G. Miklossy, P. Boross, and J. Tozser. 2005. J. Virol. 79:4213-4218). Our enzyme set composed the proteases of HIV-1, HIV-2, equine infectious anemia virus, avian myeloblastosis virus, Mason-Pfizer monkey virus (MPMV), mouse mammary tumor virus (MMTV), Moloney murine leukemia virus, human T-lymphotropic virus type 1, bovine leukemia virus, human foamy virus and walleye dermal sarcoma virus. Molecular models were used to interpret the specificity similarities and differences between these retroviral proteases. The results showed that retroviral proteases had similar preferences (Phe, Tyr) for the P1 position in this sequence context, but more differences were found for the P3 and P4 positions. Importantly, the size of the P3 and P4 residues appear to be a major contributor for specificity. The substrate specificities correlated well with the phylogenetic tree of retroviruses. Furthermore, while the specificity of some enzymes belonging to different genera appeared to be very similar (e.g. that of AMV and MMTV), the specificity of primate lentiviral proteases substantially differed from that observed for a nonprimate lentiviral protease.

Introduction

Infection by retroviruses, most notably human immunodeficiency virus (HIV), causes severe immunodeficiency, cancer, and anemia in humans and in many economically important vertebrates. With the exception of HIV/AIDS, relatively few studies have focused on human diseases arising from retroviral infections.

The total number of people living with HIV has continued to increase and it is estimated to be about 33 million people in the world (36). There were more than 2 million new infections, and about 2 million people died of acquired immunodeficiency syndrome (AIDS) related illnesses in 2007 (36). Antiretroviral therapies have proved to be effective since the introduction of the combination of drugs including inhibitors of the retroviral protease (PR). The PR of HIV is an excellent target for chemotherapy (for reviews, see references (16, 33)), since the PR activity is essential for the maturation and infectivity of the virus (as reviewed in references (14, 21)). Specificity studies of wild type and mutant HIV PRs have provided a basis for the rational design of potent, selective inhibitors (4, 30, 38) and also may help to circumvent the problems caused by the rapidly developing resistance against the compounds used in therapy (28).

Human T-lymphotropic virus (HTLV) infection is also a global epidemic: 10–20 million individuals are estimated to be carriers of the virus, and the risk of developing disease is estimated to be 5% in infected patients (15). HTLV-1 has been etiologically associated with a number of diseases including adult T-cell leukemia and HTLV-1 associated myelopathy (10). Studies indicate that viral replication is critical for the development of HTLV-1 associated myelopathy, and initial studies suggested that blocking replication with AZT had a therapeutic effect (26). HTLV PR inhibitors may also have therapeutic value in the HTLV-associated diseases, similar to the successful application of HIV PR inhibitors to treat AIDS.

Another retrovirus, which can infect people, is the human foamy virus (HFV) or prototype foamy virus (PFV). Retroviruses of the foamy virus subgroup have several unusual features (8, 23), but the PR of the HFV is also essential for viral infectivity (11). HFV holds great promise in gene therapy (20), because of its inability to cause pathogenic disease (12). However, such an application requires detailed knowledge of the replication strategy as well as the enzymes of this virus.

Retroviral PR cleavage sites are currently classified into two major groups (reviewed in (33)). Type-1 cleavage sites have an aromatic residue and Pro, while type-2 cleavage sites have hydrophobic residues (excluding Pro) at the site of cleavage. Other positions, especially P2 and P2' (nomenclature is according to reference (25)), also substantially contribute to the specificity and showed significant differences in terms of preference in the two types of cleavage sites.

Comparative study of divergent members of the retroviral PR family is a promising approach not only to recognize general and specific features of the PR, but also to discover the mutational capacity of the PR. Several of the mutations causing drug resistance of HIV-1 PR introduce residues into the substrate binding sites that are found in the equivalent position of other retroviruses (13, 24). Retroviral PRs can recognize at least seven residues of the substrates, which bind to the enzyme in an extended conformation, as schematically shown in Fig. 1. Previously, a large set of peptides containing single amino acid substitutions in the P4-P3' region of the Val-Ser-Gln-Asn-Tyr↓Pro-Ile-Val-Gln oligopeptide, a typical type-1 substrate (where the arrow denotes the site of cleavage) was used to characterize the specificity of the PR of various retroviruses including HIV-1, HIV-2 (32, 34), equine infectious anemia virus (EIAV) (37), Moloney murine leukemia virus (MMLV) (18) and avian myeloblastosis virus (AMV) (31). Based on these studies, the P2 position was found to be critical for determining the substrate specificity

differences of retroviral proteases. Therefore, the P2 position was characterized in the first part of our study (1). Here, we have continued with studies of the specificity of P1, P3 and P4 positions using the same PR set previously utilized for the mapping of the P2 position, comprising at least one member of each *genera* of the family *Retroviridae* including HIV-1, HIV-2, EIAV, MMLV, AMV, Mason-Pfizer monkey virus (MPMV), mouse mammary tumor virus (MMTV), HTLV-1, bovine leukemia virus (BLV), HFV and walleye dermal sarcoma virus (WDSV) proteases. Our previous (1) and present study has the advantage that the different retroviral proteases were mapped with the same peptide series in the same reaction conditions in the same laboratory, therefore the results are directly comparable.

Materials and methods

Retroviral proteases

Chemically synthesized HIV-2 PR was purified and refolded as described (34). Purified MPMV PR (the shortest, 12 kDa form) was prepared as described previously (39). MMTV protease was expressed as a GST-fusion protein, processed with Factor Xa and purified following the published procedure (19). MMLV protease was cloned with MBP and hexahistidine tags and purified as described (6). HTLV-1 protease (13), BLV protease (40) and WDSV protease (9) were purified from inclusion bodies after expression as described previously. HFV protease was cloned in fusion with MBP and it was used in its fusion form (7). The activity of MBP-fusion form of HFV protease was similar to that of the processed form (3).

Oligopeptides

Oligopeptides synthesized by solid-phase peptide synthesis were described previously (17, 19, 34). Stock solutions and dilutions were made 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 one hour or 24 hours in case of HFV and WDSV proteases, and were stopped by the addition of 9 volumes 1 % trifluoroacetic acid (TFA) then 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 acetonitrile gradient (0-100%) in water, in the presence of 0.05% TFA. 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 (34). 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 less than 20% substrate turnover, by dividing the activity on a given peptide by the activity on the reference substrate, as described (1, 2). Measurements were performed in duplicate and the average values were calculated. The standard error was less than 10%. The relative activities for the HIV-1 PR (34), EIAV PR (37) and AMV PR (31) have been reported previously.

Molecular modeling

The sequence alignment of the proteases used in this study and the construction of molecular models were described previously (1). A model of VSQNY↓PIVQ oligopeptide was docked into the substrate binding site of each retroviral protease and the minimization and analysis procedure were applied as described (1) with the Sybyl program package (Tripos Inc., St. Loius, MO, USA) run on the Silicon Graphics Fuel computer graphics system. Sequence identity and similarity percentage values for the full protease sequences as well as for the set of residues involved in the substrate binding (18-20 residues of the proteases) are provided in the Supplemental Table 1. Cavities were calculated from the minimized structures containing Gly at P1, P3 or P4 position using the SiteID modul of Sybyl. It should be noted that the calculated cavity volumes are comparable only for a given subsite and not suitable for comparison of different subsites, as a consequence of the cavity-finding algorithm of SiteID program and the diverse solvent accessibility of the binding sites (e.g. S1 is fully buried while S4 is opened to the surface). The volume of the amino acid residues was retrieved from the literature (41).

Results

Substrate specificity of the S1 subsite of retroviral proteases

Results obtained with representative P1-modified substrates are visualized in Fig. 2, and the complete set of relative activity values is provided in the Table 2 of the Supplemental material. None of the substrates were cleavable by HTLV-1 and HFV proteases. For the other enzymes, the activities were calculated relative to those obtained with the original, P1 Tyr-containing substrate. The S1 binding site of the PRs appears to be hydrophobic. The HIV-1, HIV-2, EIAV, AMV,

MMLV and WDSV proteases shared similar specificity, showing the preference order of Phe > Tyr > Leu ≅ Met > Ala, although the individual values for relative activity showed some variation. MPMV and MMTV PRs were more selective for the large aromatic residues. On the other hand, the PR of BLV showed the highest preference for Leu at P1, followed by Phe or Met. Therefore, the S1 site is large, hydrophobic and well conserved among retroviral proteases. These results agree with predictions from the molecular models that a bulky hydrophobic P1 side chain will fill the S1 subsite to obtain efficient cleavage. However, there are some fine specificity distinctions, in terms of whether the enzymes would also favor (or tolerate) smaller residues at this position. HIV-1, HIV-2, EIAV, AMV, MMLV and WDSV proteases tolerated the smaller Leu and Met residues substantially better than did MPMV and MMTV proteases (Fig. 2, Table 2 of Supplemental material). Interestingly, in the case of HIV-1 PR a similar preference order of Phe >> Met ≅ Tyr ≅ Leu was observed using another substrate set (Lys-Thr-Lys-Val-Xaa↓Val-Val-Gln-Pro-Lys) based on the HTLV-1 CA/NC (typical type-2) cleavage site (35), while HTLV-1 and BLV proteases preferred Phe > Leu > Tyr > Ala and Tyr ≅ Phe > Leu ≅ Met, respectively (27). The strong sequence context dependence of the preference at a given site was recognized earlier (reviewed in (33)).

Substrate specificity of the S3 subsite of retroviral proteases

Unlike the S1 subsite, various residues were preferred when the S3 binding sites were mapped for different retroviral PRs (Fig. 3). The complete set of relative activity values is provided in the Table 3 of the Supplemental material. The HIV and EIAV PRs preferred the original polar Gln at P3 (Fig. 3), but the second best residue was a hydrophobic one (Val for HIV-1 and HIV-2 PRs, Phe for EIAV PR). Smaller residues (Gly and Ala) were also well tolerated. A

pronounced preference for the large hydrophobic residues, Phe and Leu, characterized AMV, MPMV and MMTV proteases. In fact, MMTV PR did not hydrolyze the substrates with Gly or Ala at P3. Also, MPMV PR was unusual in recognizing Asp at P3 in this substrate. On the other hand, the small Ala, as well as the polar Lys were preferred by the BLV PR, and Gly by the WDSV PR (Fig. 3). The size of the residue appears to be the main specificity determinant at this position, since, with the exception of the WDSV PR, the size of the average volume of the two most preferred P3 residues correlated well with the mean cavity of the S3 subsite of the enzymes (Fig. 4), as previously established for the P2 position (1). In the exceptional case of WDSV protease, the S3 subsite binding pocket that was predicted to be relatively large and less suitable for small residues at P3. However, the uncertainty of the model of WDSV protease may contribute to this discrepancy.

Substrate specificity of the S4 subsite of retroviral proteases

Various residues were found to fit preferably to the S4 sites of the PRs, similar to the results for S3, as indicated by Fig. 5. The activities were calculated relative to that of the P4 Val-containing substrate instead of the original Ser-containing one, since the former was also a substrate of the HTLV-1 PR. The complete data are provided in the Table 4 of the Supplemental material. In the case of primate lentiviral proteases, like HIV-1 and HIV-2 proteases, the polar Ser as well as the smaller Gly or Ala residues are strongly preferred. EIAV PR prefers hydrophobic residues Val, Leu, Phe and Pro. AMV, MPMV and MMTV proteases preferred medium-sized or large hydrophobic residues Ile, Leu and Phe (Fig. 5). In contrast, the BLV and WDSV proteases preferred small hydrophobic residues (Ala, Pro) at P4, while MMLV protease preferred medium-sized polar (Ser, Thr, Asp) residues (Fig. 5). However, this subsite lies at the PR surface, and

lacks the well-defined pocket of the internal subsites. For many of these PRs a correlation was observed between the size of the average volume of the best two residues and the mean cavity of the S4 subsite of the enzymes, similar to the results for S3. However, HIV-1 and MMLV PRs did not fit this correlation for S4 (Fig. 6), that could be due to the less well defined S4 subsites of these proteases suggested by the molecular models.

Discussion

These results, together with the previously published analysis of the S2 subsite, provide a basis for characterization of the specificity of the amino-terminal subsites (S4-S1) of retroviral proteases using a type-1 cleavage site substrate, having Tyr and Pro at the site of cleavage. A schematic representation of the specificities is provided in Fig. 7. The alpharetroviral AMV and betaretroviral MMTV proteases appear to share very similar specificity. All their substrate binding sites are hydrophobic and large, except for the small S2 pocket (Fig. 7). The deltaretroviral BLV PR has large hydrophobic S1 and S2 pockets, and smaller S3 and S4 subsites, and only the S3 subsite has some polar character. The gammaretroviral MMLV PR has a similar specificity, except the S4 site appears to be small and hydrophilic. On the other hand, the epsilonretroviral WDSW PR has hydrophobic S4, S1, and somewhat hydrophilic S2 and S3 subsites. Interestingly, the specificity of lentiviral proteases can be subdivided into two groups. The primate lentiviral HIV-1 and HIV-2 PRs have a substantially different specificity as compared to that of the EIAV PR, both in the hydrophobicity of the S2 and S4 subsites as well as the size of S4. The important conclusion is that the specificity pattern of the subsites agrees with the evolutionary relationship among the proteases as represented by the phylogenetic tree (Fig 7). For example, EIAV, AMV and MMTV share very similar specificity at S1, S2 (1) and S4

subsites. It is of interest to note that while the degree of sequence identity and similarity of retroviral protease sequences is fairly low, the residues forming the substrate binding sites are substantially more conserved (Supplemental Material, Table 1), a likely consequence of the selective evolutionary pressure to maintain these residues as compared to those that are not critical for the structure and activity of the enzyme.

HIV-1 PR clinical inhibitors typically only weakly inhibit, if at all, the other retroviral proteases, with the exception of MMLV PR, which was inhibited by an HIV-1 PR inhibitor (5). So there is an apparent contradiction between the relatively conserved substrate specificity among retroviral PRs and an almost complete lack of inhibition by HIV-1 PR inhibitors. This contradiction might be due to the fact that the clinical inhibitors are typically rigid molecules, while substrates are more flexible, and capable of adapting to altered substrate binding sites in different retroviral PRs. The same phenomenon is observed in drug resistant HIV-1 PR variants (22, 29). The development of resistance towards the drugs designed against HIV-1 PR is one of the main problems in the protease inhibitor therapy of AIDS. Many of the mutations occurring in drug resistance introduce amino acids that can be found at the equivalent position in other retroviral proteases. Therefore, characterization of the specificity similarities and differences of these enzymes may help to design broad-spectrum inhibitors against HIV-1 PR.

Acknowledgements

This research was supported by the Hungarian Science and Research Fund (OTKA K68288, F35191), the Ministry of Public Health and Welfare (Egészségügyi Tudományos Tanács, ETT 88/2003) and United States National Institute of Health grants AIDS FIRCA TW01001 and GM062920.

We thank Dr. Stephen Oroszlan for providing the MMTV protease clone, Dr. John M. Louis for the HIV-1 protease clone, Dr. Iva Pichova for the MPMV protease clone, and Dr. Volker M. Vogt for the WDSV protease clone. The synthesis of the peptides is acknowledged to Dr. Ivo Bláha. We thank Gabriella Emri, Zsolt Oláh and Gábor Zahuczky for help in some of the measurements and Szilvia Petö for technical assistance.

References

1. **Bagossi, P., T. Sperka, A. Feher, J. Kadas, G. Zahuczky, G. Miklossy, P. Boross, and J. Tozser.** 2005. Amino acid preferences for a critical substrate binding subsite of retroviral proteases in type 1 cleavage sites. J. Virol. **79:**4213-4218.
2. **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.
3. **Boross, P., J. Tozser, and P. Bagossi.** 2006. Improved purification protocol for wild-type and mutant human foamy virus proteases. Protein Expr Purif **46:**343-7.
4. **Dunn, B. M., A. Gustchina, A. Wlodawer, and J. Kay.** 1994. Subsite preferences of retroviral proteinases. Methods Enzymol. **241:**254-278.
5. **Feher, A., P. Boross, T. Sperka, G. Miklossy, J. Kadas, P. Bagossi, S. Oroszlan, I. T. Weber, and J. Tozser.** 2006. Characterization of the murine leukemia virus protease and its comparison with the human immunodeficiency virus type 1 protease. J Gen Virol **87:**1321-30.
6. **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.
7. **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.
8. **Flügel, R. M., and K. I. Pfrepper.** 2003. Proteolytic processing of foamy virus Gag and Pol proteins. Curr. Top. Microbiol. Immunol. **277:**63-88.
9. **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.

10. **Gallo, R. C.** 2002. Human retroviruses after 20 years: a perspective from the past and prospects for their future control. Immunol Rev **185:**236-65.

11. **Konvalinka, J., M. Löchelt, H. Zentgraf, R. M. Flügel, and H. G. Krausslich.** 1995. Active foamy virus proteinase is essential for virus infectivity but not for formation of a Pol polyprotein. J. Virol. **69:**7264-7268.

12. **Linial, M.** 2000. Why aren't foamy viruses pathogenic? Trends Microbiol. **8:**284-289.

13. **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.

14. **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.

15. **Macchi, B., E. Balestrieri, and A. Mastino.** 2003. Effects of nucleoside-based antiretroviral chemotherapy on human T cell leukaemia/lymphotropic virus type 1 (HTLV-1) infection in vitro. J Antimicrob Chemother **51:**1327-30.

16. **Mellors, J. W.** 1996. Closing in on human immunodeficiency virus-1. Nat. Med. **2:**274-275.

17. **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.

18. **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.

19. **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.

20. **Mergia, A., and M. Heinkelein.** 2003. Foamy virus vectors. Curr Top Microbiol Immunol **277:**131-59.

21. **Oroszlan, S., and R. B. Luftig.** 1990. Retroviral proteinases. Curr. Top. Microbiol. Immunol. **157:**153-185.

22. **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.

23. **Rethwilm, A.** 2003. The replication strategy of foamy viruses. Curr. Top. Microbiol. Immunol. **277:**1-26.

24. **Ridky, T., and J. Leis.** 1995. Development of drug resistance to HIV-1 protease inhibitors. J. Biol. Chem. **270:**29621-29623.

25. **Schechter, I., and A. Berger.** 1967. On the size of the active site in proteases. I. Papain. Biochem. Biophys. Res. Commun. **27:**157-162.

26. **Sheremata, W. A., D. Benedict, D. C. Squilacote, A. Sazant, and E. DeFreitas.** 1993. High-dose zidovudine induction in HTLV-I-associated myelopathy: safety and possible efficacy. Neurology **43:**2125-9.

27. **Sperka, T., G. Miklossy, Y. Tie, P. Bagossi, G. Zahuczky, P. Boross, K. Matuz, R. W. Harrison, I. T. Weber, and J. Tozser.** 2007. Bovine leukemia virus protease: comparison with human T-lymphotropic virus and human immunodeficiency virus proteases. J Gen Virol **88:**2052-63.

28. **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.

29. **Tie, Y., P. I. Boross, Y. F. Wang, L. Gaddis, F. Liu, X. Chen, J. Tozser, R. W. Harrison, and I. T. Weber.** 2005. Molecular basis for substrate recognition and drug resistance from 1.1 to 1.6 angstroms resolution crystal structures of HIV-1 protease mutants with substrate analogs. Febs J **272:**5265-77.

30. **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.

31. **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.

32. **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.

33. **Tözsér, J., and S. Oroszlan.** 2003. Proteolytic events of HIV-1 replication as targets for therapeutic intervention. Curr. Pharm. Des. **9:**1803-1815.

34. **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.

35. **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.

36. **UNAIDS.** 2007. AIDS epidemic update: December 2007. http://www.unaids.org.

37. **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.

38. **Wlodawer, A.** 2002. Rational approach to AIDS drug design through structural biology. Annu. Rev. Med. **53:**595-614.

39. **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.

40. **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.

41. **Zamyatin, A. A.** 1972. Protein volume in solution. Prog. Biophys. Mol. Biol. **24:**107-123.

Figure legends

Figure 1. A schematic representation of the HIV-1 matrix↓capsid substrate in the S4-S3' subsites of HIV-1 PR. The relative size of each subsite is indicated approximately by the area enclosed by the curved line around each substrate side chain.

Figure 2. Comparison of relative activities obtained on selected Val-Ser-Gln-Asn-Xaa↓Pro-Ile-Val-Gln peptides for retroviral proteases.

Figure 3. Comparison of relative activities obtained on selected Val-Ser-Xaa-Asn-Tyr↓Pro-Ile-Val-Gln peptides for retroviral proteases.

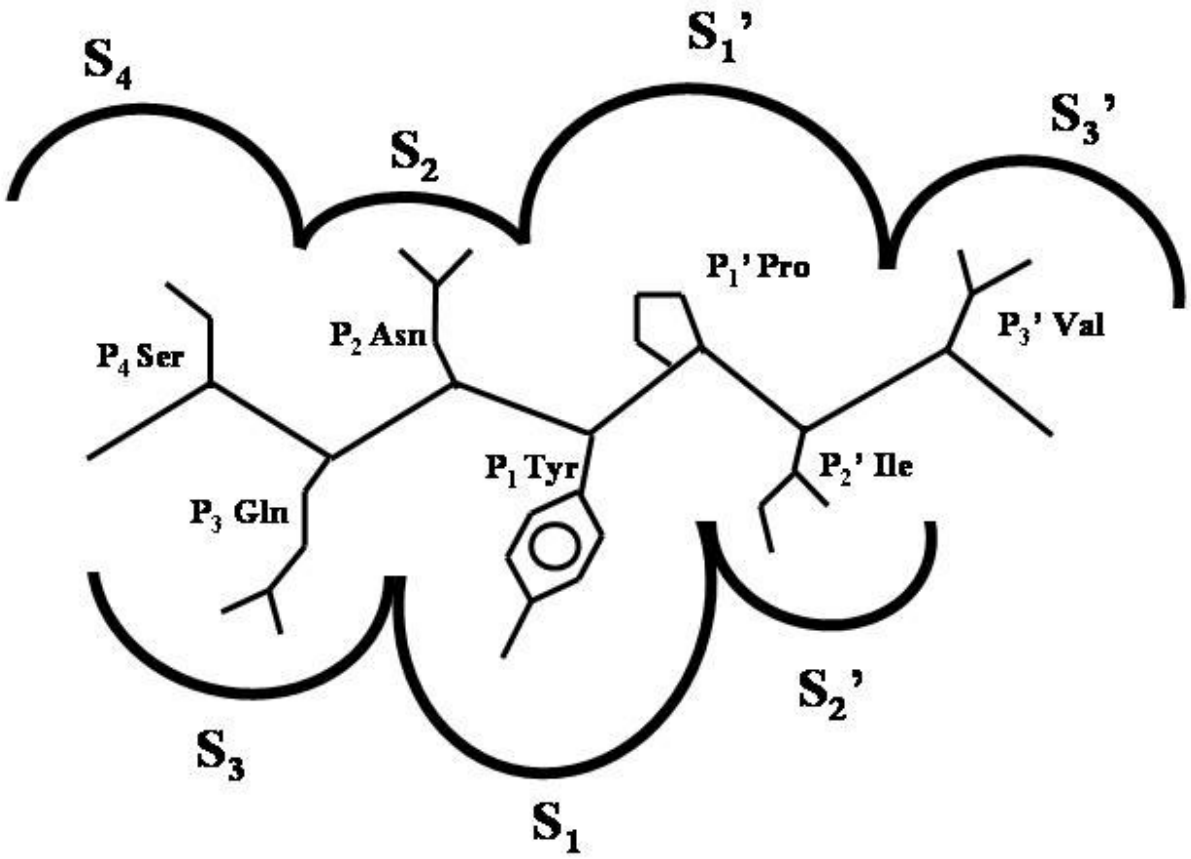
Figure 4. Mean cavity volume of the S3 subsite of various retroviral proteases *versus* averaged volume of the two P3 residues for which the measured relative activity was the highest.

Figure 5. Comparison of relative activities obtained on selected Val-Xaa-Gln-Asn-Tyr↓Pro-Ile-Val-Gln peptides for retroviral proteases.

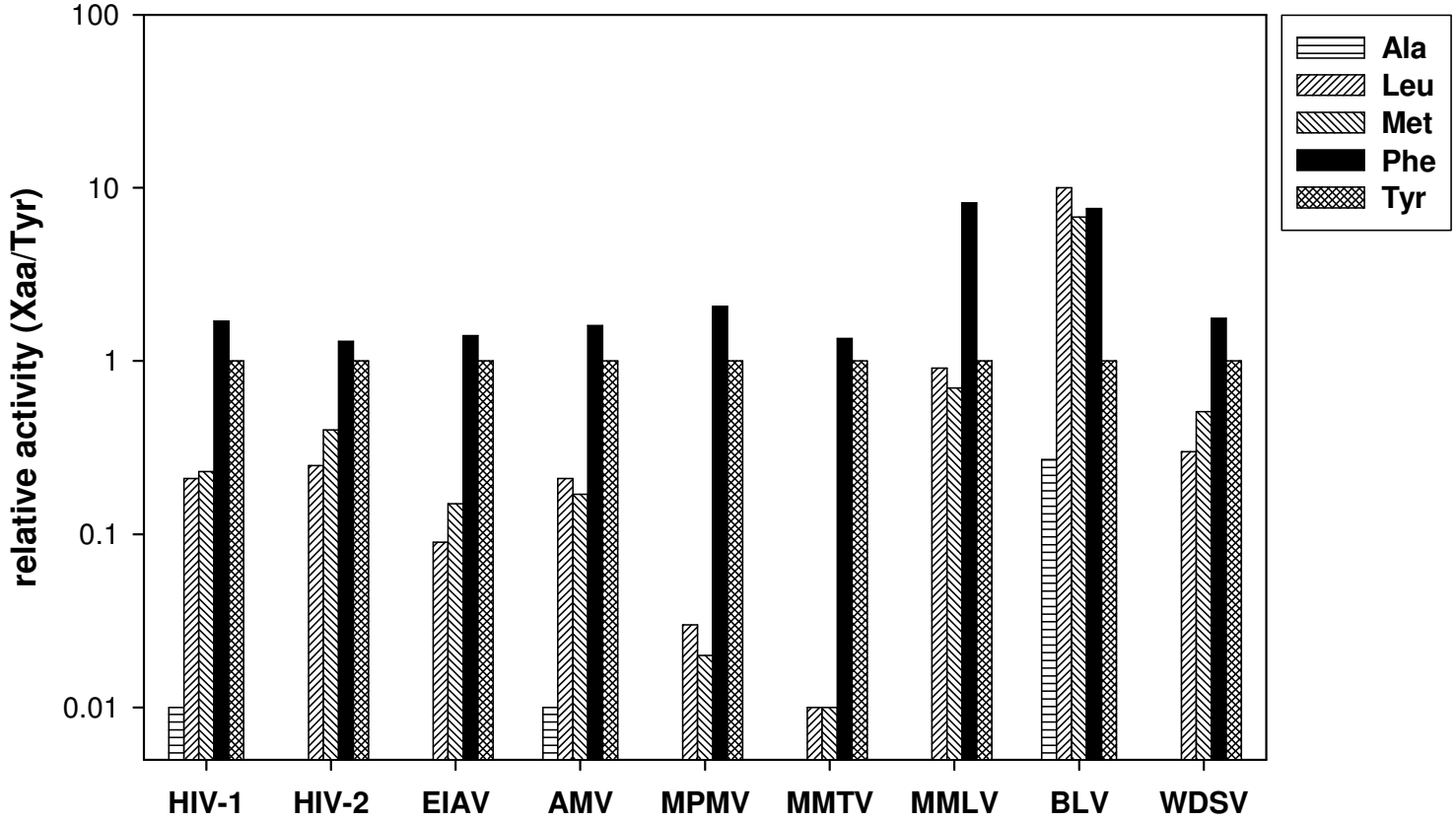
Figure 6. Mean cavity volume of S4 subsite of various retroviral proteases *versus* averaged volume of the two P4 residues for which the measured relative activity was the highest.

Figure 7. Phylogenetic tree and schematic representation of the preferred residues of the S4-S1 subsites of representative retroviral proteases. The phylogenetic tree contains Rous sarcoma virus (RSV) PR while measurements were performed using AMV PR. These two PRs differ only in two residues that do not influence the specificity. The size of the oval objects representing the substrate amino acid side chains approximates the size of the most preferred residues. Gray objects represent sites that prefer predominantly hydrophobic residues, while white objects represent sites that do not discriminate based on hydrophobicity. Dashed lines for the S4 subsites indicate that these pockets are less defined than the other ones, due to the proximity to the protein surface.

Eizert et al., Fig. 1.



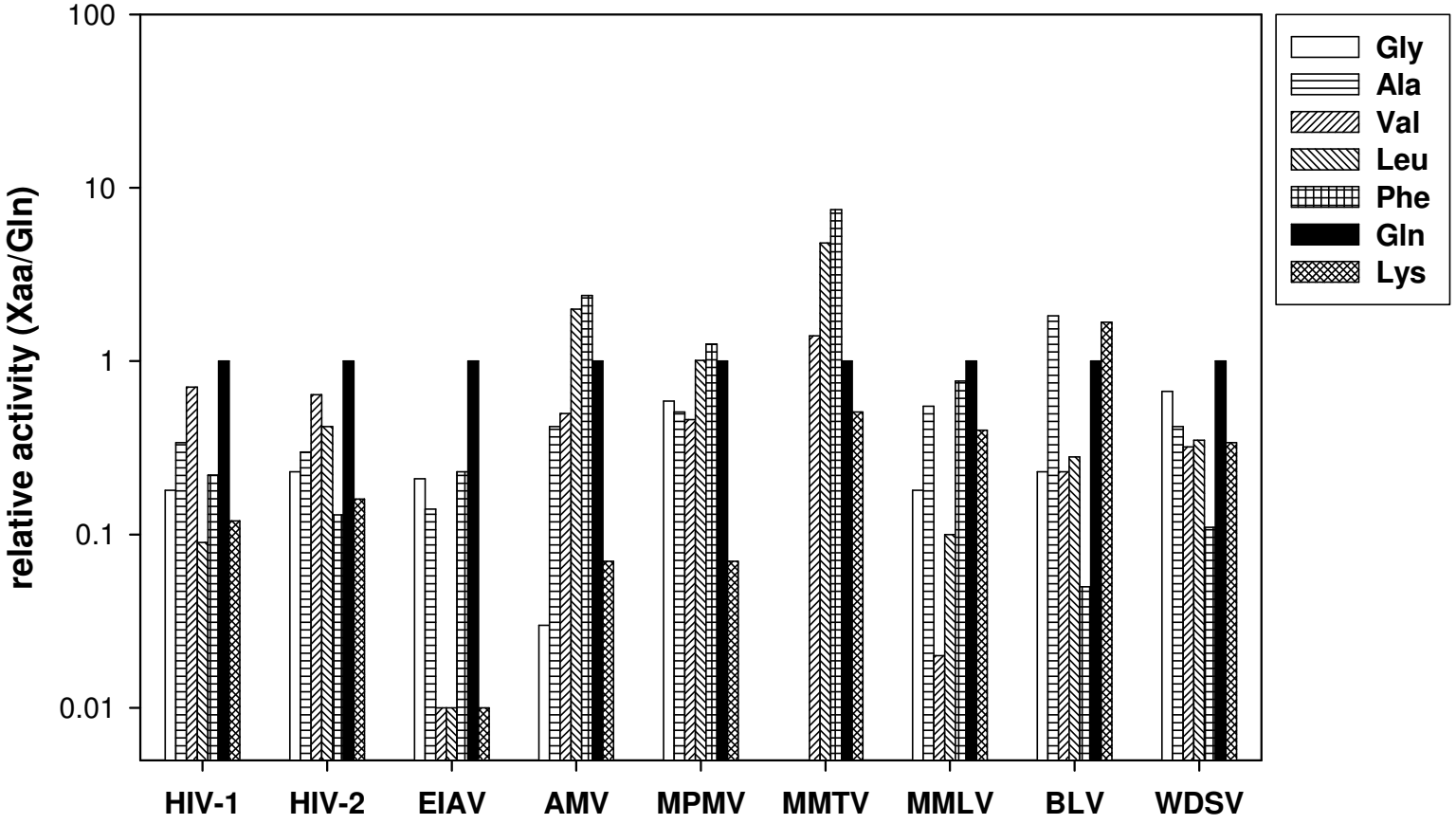
1 **Eizert et al., Fig. 2.**



2

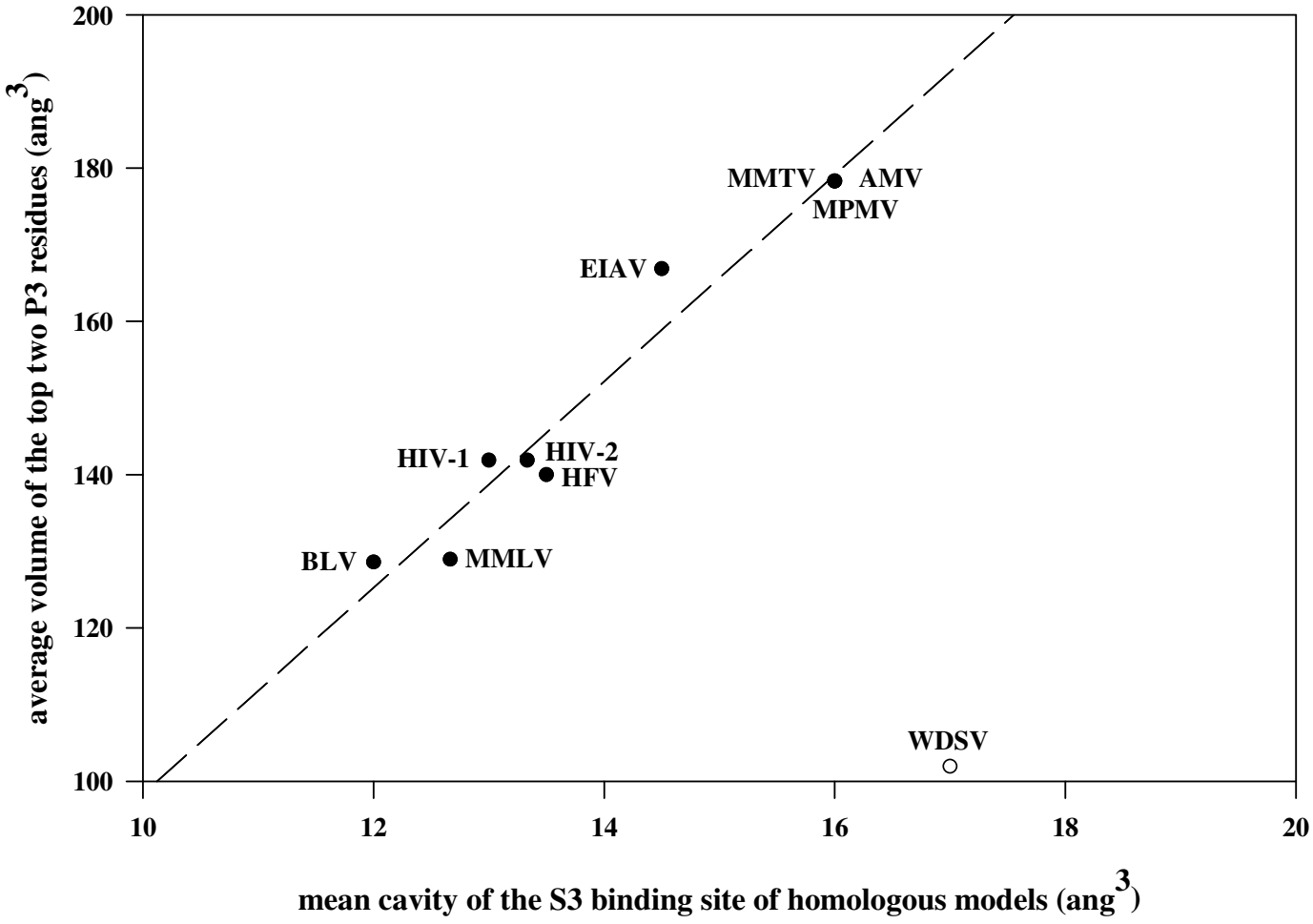
3

1 **Eizert et al., Fig. 3.**



2

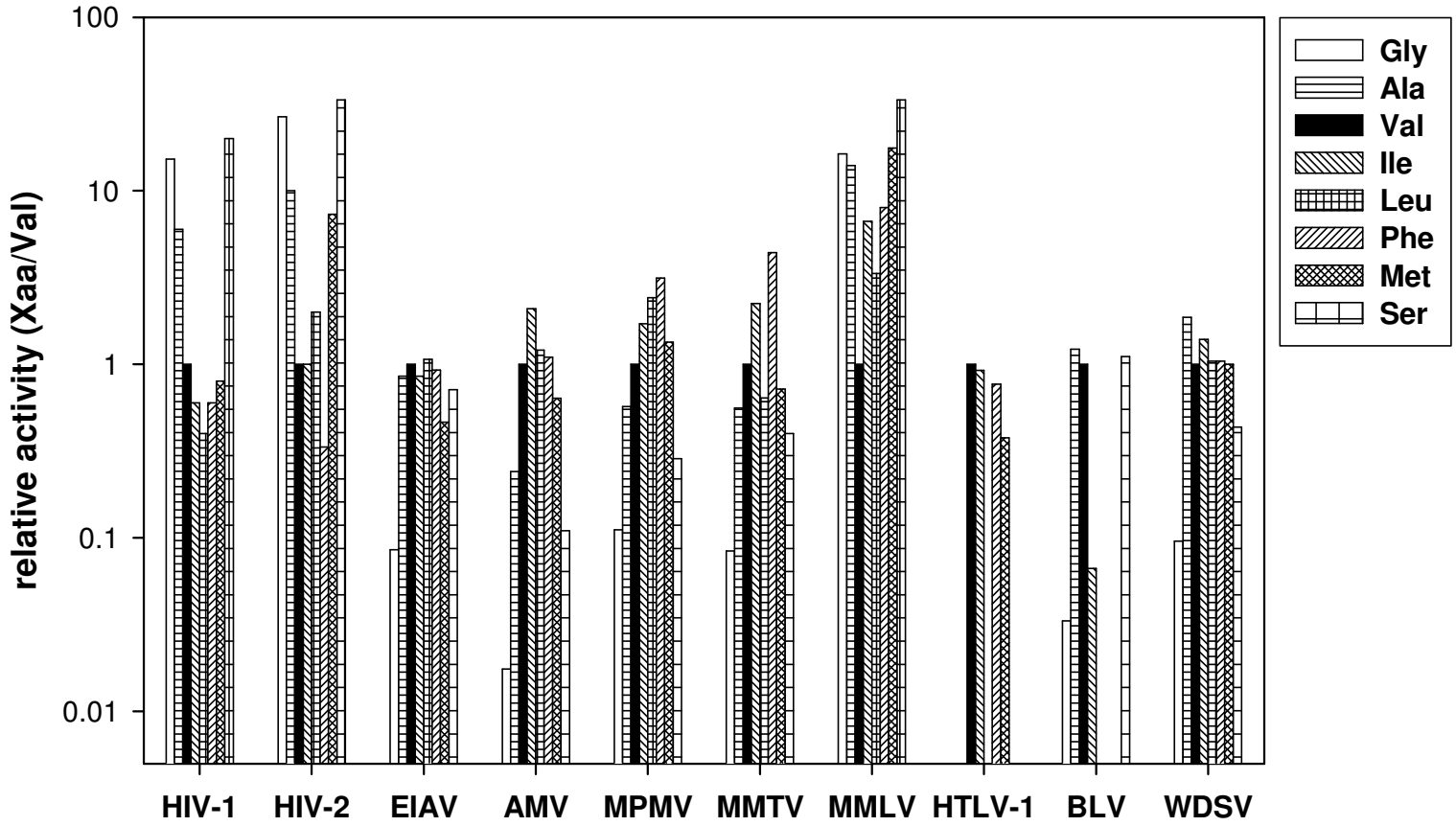
1 **Eizert et al., Fig. 4**



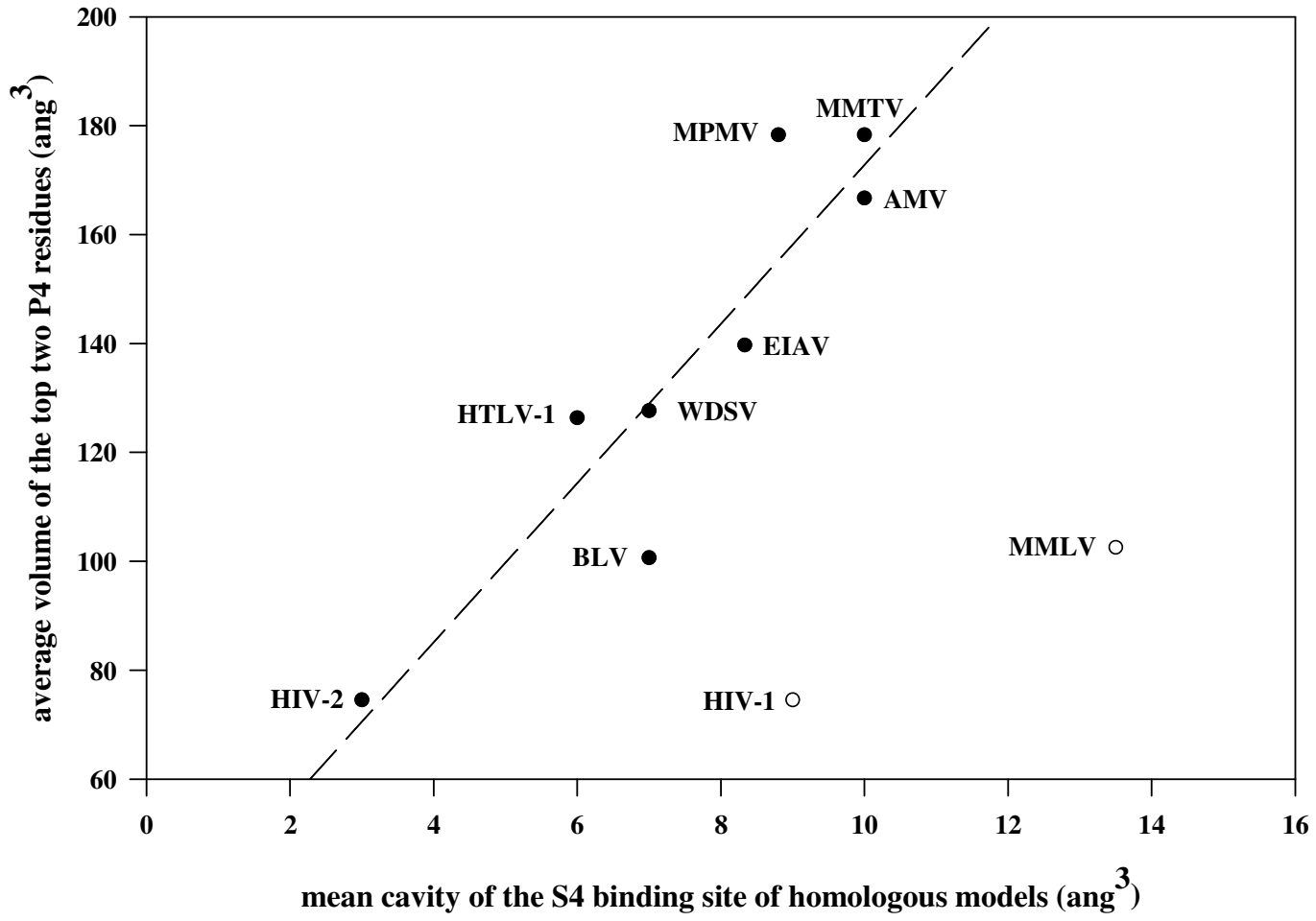
2

3

Eizert et al., Fig. 5.

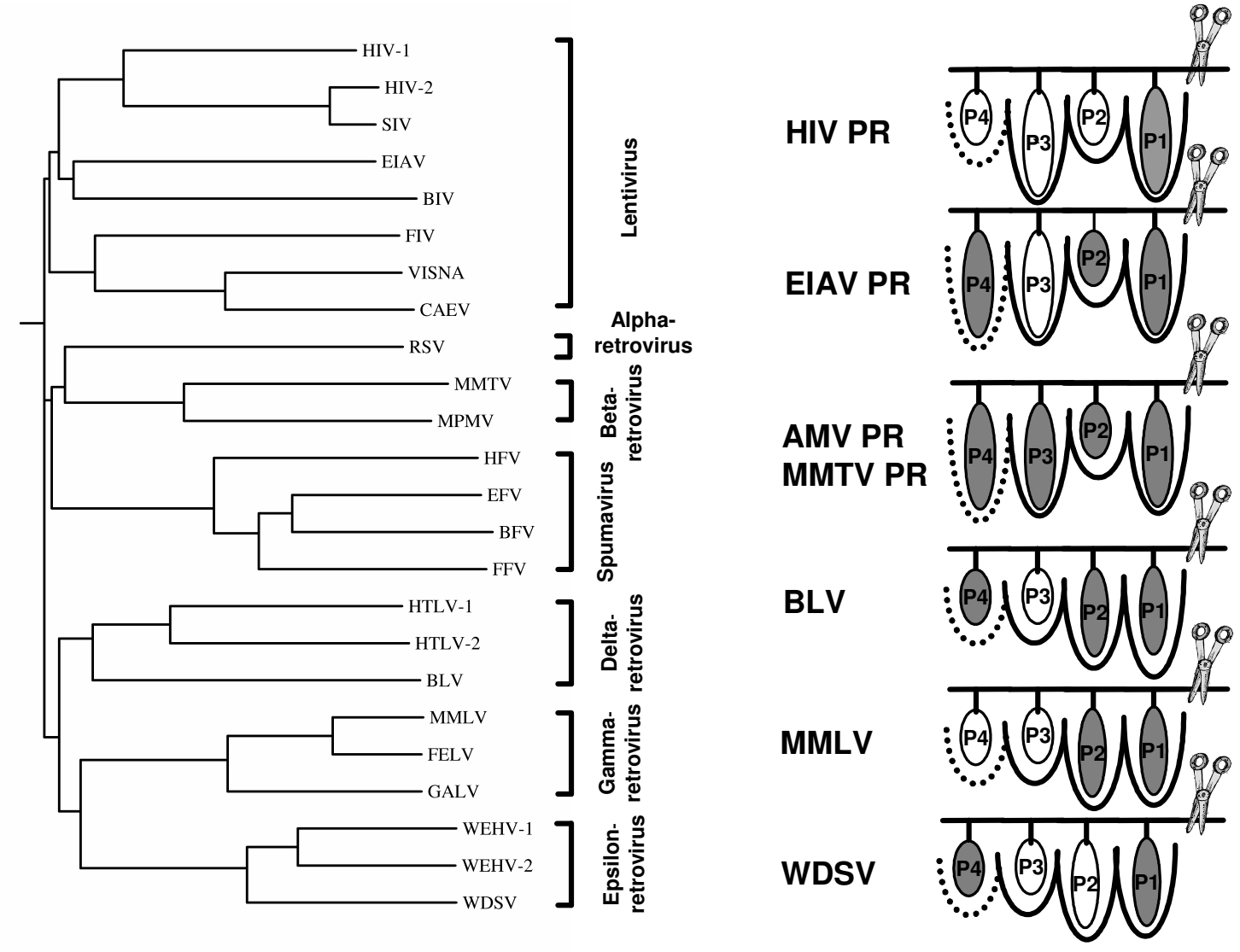


1 **Eizert et al., Fig. 6.**



2

1 **Eizert et al., Fig. 7.**



2

Supplemental Material, Table 1. Sequence identity and similarity of retroviral proteases.

PR	HIV-1	HIV-2	EIAV	AMV	MPMV	MMTV	MMLV	HTLV-1	BLV	WDSV	HFV
HIV-1	--	--	--	--	--	--	--	--	--	--	--
HIV-2	47/63 **65/80**	--	--	--	--	--	--	--	--	--	--
EIAV	31/40 **70/75**	36/46 **55/75**	--	--	--	--	--	--	--	--	--
AMV	31/41 **50/55**	26/40 **50/55**	27/45 **45/60**	--	--	--	--	--	--	--	--
MPMV	24/39 **50/75**	22/36 **45/65**	20/37 **40/75**	25/41 **40/70**	--	--	--	--	--	--	--
MMTV	22/32 **45/65**	20/31 **40/55**	22/37 **40/65**	24/38 **35/60**	46/57 **55/65**	--	--	--	--	--	--
MMLV	24/41 **40/50**	24/36 **35/45**	24/35 **40/50**	16/33 **25/40**	21/32 **35/50**	16/30 **40/55**	--	--	--	--	--
HTLV-1	34/42 **50/55**	26/40 **45/50**	29/42 **50/55**	24/37 **40/50**	20/33 **30/55**	18/27 **40/55**	22/29 **45/55**	--	--	--	--
BLV	28/45 **45/55**	22/39 **40/50**	25/38 **45/55**	22/34 **35/50**	17/30 **25/55**	15/27 **30/50**	27/36 **55/60**	32/43 **75/85**	--	--	--
WDSV	15/26 **35/35**	19/32 **40/40**	22/29 **30/35**	17/28 **35/40**	19/28 **35/45**	18/32 **30/40**	22/34 **25/35**	19/29 **30/40**	13/21 **30/35**	--	--
HFV	23/30 **35/35**	16/30 **25/40**	15/25 **25/30**	22/28 **30/30**	12/20 **25/40**	10/19 **20/30**	15/21 **15/20**	17/26 **15/25**	18/27 **15/20**	14/22 **25/40**	--

Sequence identity and similarity of full retroviral protease sequences as well as of residues involved in substrate binding (numbers in bold) are given as identity/similarity percentage values.

Supplemental material, Table 2. Relative activities obtained on Val-Ser-Gln-Asn-Xaa↓Pro-Ile-Val-Gln peptides for retroviral proteases.

	Lentivirus[1]			Alpha-retrovirus	Betaretrovirus		Gamma-retrovirus	Deltaretrovirus		Epsilon-retrovirus	Spuma-virus
	Protease of										
P1 residue	HIV-1	HIV-2	EIAV	AMV	MPMV	MMTV	MMLV	HTLV-1	BLV	WDSV	HFV
Gly	0	0	0	0	0	0	0	0	nd[2]	0	0
Ala	0.01	0	0	0.01	0	0	0	0	0.27	0	0
Val	0	0	0	0.02	0	0	0	0	0.13	0	0
Leu	0.21	0.25	0.09	0.21	0.03	0.01	0.91	nd	**10**	0.30	0
Phe	**1.7**	**1.3**	**1.4**	**1.6**	**2.1**	**1.3**	**8.2**	0	**7.6**	**1.8**	0
Tyr	**1**	**1**	**1**	**1**	**1**	**1**	**1**	0	1	**1**	0
Trp	0.01	nd	0	0.02	0	0	0	0	0.18	0	0
Cys	<0.01	nd	0.69	<0.01	0	nd	0	0	0.86	nd	0
Met	0.23	0.40	0.15	0.17	0.02	0.01	0.70	0	6.8	0.51	0
Ser	0	0	0	0	0	0	0	0	0	0	0
Asp	0	0	0	0	0.02	0	0	nd	0	0	0
Lys	0	0	0	0	0	0	0	0	0	0	0

[1]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). Activities were expressed relative to that obtained with the Val-Ser-Gln-Asn-Tyr↓Pro-Ile-Val-Gln substrate, except for HTLV-1 and HFV PRs. [2]nd, not determined. The best two values are highlighted in bold.

Supplemental Material, Table 3. Relative activities obtained on Val-Ser-Xaa-Asn-Tyr↓Pro-Ile-Val-Gln peptides for retroviral proteases.

	Lentivirus[1]			Alpha-retrovirus	Betaretrovirus		Gamma-retrovirus	Deltaretrovirus		Epsilon-retrovirus	Spuma-virus
	Protease of										
P3 residue	HIV-1	HIV-2	EIAV	AMV	MPMV	MMTV	MMLV	HTLV-1	BLV	WDSV	HFV
Gly	0.18	0.23	0.21	0.03	0.59	0	0.18	0	0.23	**0.67**	0
Ala	0.34	0.30	0.14	0.42	0.51	0	0.55	0	**1.8**	0.42	0
Val	**0.71**	**0.64**	0.01	0.50	0.46	1.4	0.02	0	0.23	0.32	**nc**[2]
Leu	0.09	0.42	0.01	**2.0**	**1.0**	**4.8**	0.10	0	0.28	0.35	0
Phe	0.22	0.13	**0.23**	**2.4**	**1.3**	**7.5**	0.77	0	0.05	0.11	0
Asn	0.50	0.32	0.08	0.05	0.43	1.1	**0.81**	0	0.26	0.11	0
Gln	**1**	**1**	**1**	1	1	1	**1**	0	1	**1**	0
Asp	0.01	0.10	0	0.01	0.81	0.06	0.01	0	0.22	0	0
Lys	0.12	0.16	0.01	0.07	0.07	0.51	0.40	0	**1.7**	0.34	0

[1]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). Activities were expressed relative to that obtained with the Val-Ser-Gln-Asn-Tyr↓Pro-Ile-Val-Gln substrate, except for HFV PR for which only one peptide was cleaved. The best two values are highlighted in bold. [2]nc, not calculated. It should be noted, that in case of HFV, only the P3 Val-containing peptide was hydrolyzed, therefore it was not possible to calculate the relative activity value.

Supplemental Material, Table 4. Relative activities obtained on Val-Xaa-Gln-Asn-Tyr↓Pro-Ile-Val-Gln peptides for retroviral proteases.

	Lentivirus[1]			Alpha-retrovirus	Betaretrovirus		Gamma-retrovirus	Deltaretrovirus		Epsilon-retrovirus	Spuma-virus
	Protease of										
P4 residue	HIV-1	HIV-2	EIAV	AMV	MPMV	MMTV	MMLV	HTLV-1	BLV	WDSV	HFV
Gly	**15**	**27**	0.09	0.02	0.11	0.08	16	0	0.03	0.10	0
Ala	6.0	10	0.86	0.24	0.57	0.56	14	0	**1.2**	**1.9**	0
Val	1	1	1	1	1	1	1	**1**	1	1	0
Ile	0.60	1.0	0.86	**2.1**	1.7	**2.2**	6.7	0.92	0.07	**1.4**	0
Leu	0.40	2.0	**1.1**	**1.2**	**2.4**	0.64	3.3	nd	0	1.0	0
Phe	0.60	0.33	0.93	1.1	**3.1**	**4.4**	8.0	0.77	0	1.0	0
Met	0.80	7.3	0.46	0.64	1.3	0.72	18	0.38	0	1.0	0
Ser	**20**	**33**	0.71	0.11	0.29	0.40	**33**	0	1.1	0.43	0
Thr	5.4	19	0.24	0.53	0.43	nd	**26**	0	1.0	0.48	0
Pro	2.8	4.0	**1.4**	0.68	0.34	nd	15	**1.7**	**1.6**	1.3	0
Asn	3.4	nd	0.49	0.15	0.34	0.23	10	nd	0.03	0.43	0
Asp	3.4	21	0.49	0.09	0.13	0.14	23	0	0	0.03	0
Lys	1.2	1.0	0.18	0.32	2.1	0	7.3	0	0.01	1.0	0

[1]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). Relative activities were expressed as activity relative to that obtained with the Val-Val-Gln-Asn-Tyr↓Pro-Ile-Val-Gln substrate. The best two values are highlighted in bold.