Chemistry & Biology, Vol. 10, 373–380, May, 2003, ©2003 Elsevier Science Ltd. All rights reserved. DOI 10.1016/S1074-5521(03)00104-2

Understanding HTLV-I Protease

Review

Suzanne Beckham Shuker,* Victoria L. Mariani, Bryan E. Herger, and Kelly J. Dennison
School of Chemistry and Biochemistry
Georgia Institute of Technology
Atlanta, Georgia 30332

Human T cell leukemia virus type I (HTLV-I) is a human retrovirus that has been clinically associated with adult T cell leukemia/lymphoma, and it has been designated as a dangerous emerging pathogen by the Centers for Disease Control. Like other retroviruses, proteolytic processing of specific sites in polyprotein precursors by a retroviral protease is an essential step in the viral life cycle. HTLV-I protease is a 28 kDa homodimeric aspartic acid protease that has only modest homology to other retroviral proteases. The enzymology of HTLV-I protease has only recently begun to be investigated, and although it shares many characteristics of other retroviral proteases, it exhibits distinct substrate specificity and different susceptibility to aspartic acid protease inhibitors. This review describes what has been reported to date on the structural characterization, specificity, and inhibition of HTLV-I protease.

Introduction

Human T cell leukemia virus type I (HTLV-I) is an oncovirus in the *Retroviridae* family [1]. HTLV-I was the first human retrovirus to be discovered and was isolated in the early 1980's from patients with adult T cell leukemia/lymphoma (ATL) [2]. It was subsequently shown to be clinically associated with ATL [3], tropical spastic paraparesis/HTLV-I-associated myelopathy (TSP/HAM) [4], and a number of chronic diseases, including uveitis, arthritis, and infective dermatitis [5, 6]. HTLV-I infections are endemic in high-risk groups in Melanesia, Japan, the Caribbean, sub-Saharan Africa, and the United States. Currently, it is estimated that 15 to 20 million individuals worldwide are infected with HTLV-I, and up to 10% of these infected individuals will develop ATL or TSP/HAM [6, 7]. The observation that Caribbean strains of HTLV-I are more easily transmitted and cause disease at a much earlier age than Japanese strains suggests that HTLV-I may be evolving into a much more dangerous virus, and the Centers for Disease Control have recently designated HTLV-I as a dangerous emerging pathogen [8, 9].

Based on its molecular organization and structure, HTLV-I has been taxonomically grouped in the *delatvirus* genus of retroviruses with HTLV-II, primate T cell leukemia virus (PTLV), and bovine leukemia virus (BLV) [10]. The mature type C virion is spherical and has a diameter of 110 to 140 nm. The host cell-derived viral membrane contains two glycoproteins, a 21 kDa transmembrane protein (TM) and a 46 kDa membrane surface glycoprotein (SU), encoded by the viral *env* gene. The center of the HTLV-I virion consists of a dense, spherical nucleocapsid containing two copies of the 9032 base-pair RNA genome [11]. It is made up of the genes *gag, pro, pol*, and *env* flanked by long terminal repeats on each end, and expression of the genome yields Gag and Gag-Pro-Pol polyprotein precursors [12, 13]. HTLV-I protease is encoded in a different reading frame than the Gag precursor and spans from the 3′ end of the *gag* region to the 5′ end of *pol* region. Synthesis of the protease occurs by ribosomal frameshifting [14], and a second ribosomal frameshift produces the 95 kDa Pol polyprotein. Upon viral maturation, the retroviral protease (PR) cleaves the 55 kDa Gag precursor into the matrix (MA), capsid (CA), and nucleocapsid (NC) proteins. The Pol polyprotein is cleaved into reverse transcriptase/RNase H (RT-RH) and integrase (IN). Because of its essential role in viral replication, HTLV-I protease is an attractive target for the development of inhibitors to treat HTLV-I infection. However, it has received little attention compared to its more famous cousin HIV protease, which has been studied extensively. This review details what is known about the activity, specificity, and inhibition of HTLV-I protease.

Retroviral Aspartic Acid Proteases

Cellular aspartic proteases, which are grouped in the A1 family of aspartic proteases, contain two aspartic acids in a highly conserved Asp-Thr/Ser-Gly motif. The three-dimensional structures for members of the A1 family reveal that they are bilobed, pseudodimeric monomers in which the lobes have similar structures. Retroviral proteases were quickly classified as a new family of aspartic proteases, the A2 family, when their primary sequences were reported [15, 16]. However, retroviral proteases are much smaller than their cellular counterparts, such as pepsin, and their sequences contain only one Asp-Thr-Gly motif. It was postulated, therefore, that retroviral proteases are symmetric homodimers and this was confirmed upon structure elucidation. Structures have been determined for the proteases from HIV-1, HIV-2, simian immunodeficiency virus (SIV), feline immunodeficiency virus (FIV), equine infectious anemia virus (EIAV), and Rous sarcoma virus (RSV). From these structures, it can be seen that the dimer is held together by a four-stranded antiparallel β sheet involving the amino and carboxyl termini of each monomer. They contain one active site that is formed at the dimer interface using residues from each monomer. The active site is located in a cleft above the dimer interface, with the catalytic aspartyl residues positioned in a loop on the floor of the cleft. The top of the cleft is formed by a β sheet containing a β-turn, referred to as "the flap," from each monomer. These features can be seen in one of the published structures of HIV-1 protease, shown in Figure 1 [17].

Most of the structural studies of retroviral proteases have been on HIV-1 protease with bound inhibitor. Peptidic inhibitors bind in a β sheet conformation, so it has been inferred that substrates bind in a similar manner. Typically, a peptide must be at least seven amino acids long to be recognized as a substrate. Following conven-

*Correspondence: suzy.shuker@chemistry.gatech.edu

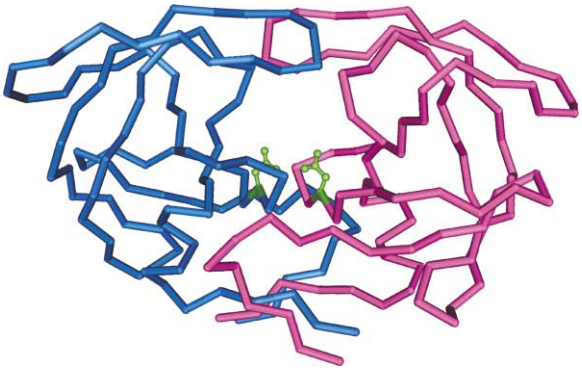


Figure 1. Structure of HIV-1 Protease

One monomer is colored blue and the other is colored magenta. The catalytic aspartic acids are highlighted in green.

tion, the amino acids toward the amino terminus of the substrate are numbered P1 to P4, with P1 being adjacent to the scissile bond, and those toward the carboxyl terminus are numbered P1′ to P4′. Retroviral proteases have eight binding pockets that accommodate the side chains of the substrate, and these subsites are numbered S1 to S4 and S1′ to S4′ to indicate which substrate residue is bound within the subsite. The strongest interactions occur between the S3 to S3′ subsites and the P3 to P3′ side chains, while the S4/P4 and S4′/P4′ interactions are much weaker. In addition to the contacts between the subsites and the substrate side chains, the protease also forms specific interactions with the backbone of the substrate. The side chain interactions confer specificity to the protease, while the interactions with the backbone provide the proper orientation of the substrate within the active site.

Sequence alignment of several retroviral proteases is shown in Figure 2. They exhibit only modest homology to one another. The regions of highest homology are found at the dimer interface (the N- and C-terminal regions) and in the residues surrounding the catalytic aspartic acid, indicated by the asterisk. There is also significant homology in the residues that form the binding pockets for the substrate, which results in similarity in substrate specificity among the retroviral proteases. The substrates are predominantly hydrophobic, and, although there are some differences observed between different proteases, examination of the substrate sequences (Table 1) reveals some trends. For example, while there is variability in the P4 position among different proteases, there is similarity among substrates for a single protease. Proline is prevalent in HTLV-I and BLV substrates, while serine and threonine are common in MuLV and HIV-1 substrates. The P4′ position is usually hydrophobic, but the identities of the side chains found in this position show no patterns. The P3 and P3′ positions show great variability, and the handful of charged residues found in the natural substrates are most often accommodated in these positions. Large hydrophobic amino acids are prevalent in the P2 and P2′ positions, with leucine, isoleucine, and valine being the most common. The P1 and P1′ positions are also predominantly hydrophobic, and no charged residues are observed at these positions in any of the natural substrates that have been identified to date. The P1 amino acid is most often a leucine, and β-branched side chains are not accommodated in the S1 subsite. On the other hand, side chains with a β-branch are found in the P1′ position in some substrates, but the most common amino acid is proline. The substrates and subsites of HTLV-I protease, in particular, will be discussed in greater detail in the next section.

HTLV-I Protease

HTLV-I protease was identified and isolated in the late 1980's [18]. Recombinant protease has been expressed in *E. coli*, and it correctly processes HTLV-I Gag polyprotein into MA, CA, and NC proteins in vitro [19–22]. HTLV-I protease is a small dimer composed of two identical subunits containing 125 amino acids each. Structurally, this protease shares the active site motif Asp-Thr-Gly with other retroviral proteases and has homology in the sequences flanking the active site. Sequence alignment of retroviral proteases indicates that HTLV-I protease is most closely related to HTLV-II protease and BLV protease (Figure 2). One notable difference between the leukemia virus proteases and the other retroviral

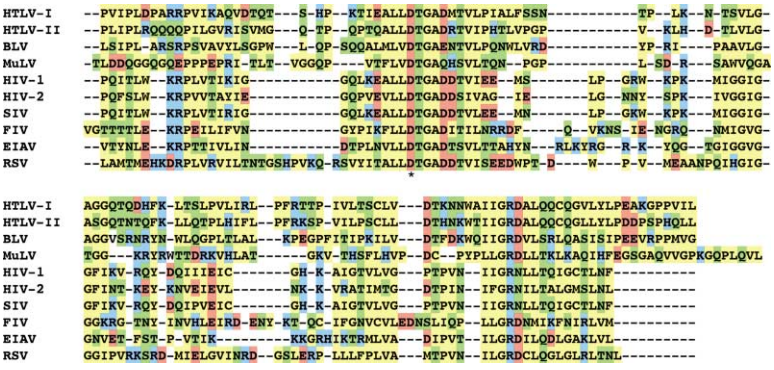


Figure 2. Sequence Alignment of the Leukemia Retrovirus Proteases with Retroviral Proteases of Known Structure

The alignment for HIV-1, HIV-2, SIV, FIV, EIAV, and RSV proteases was generated based on the reported structures (PDB IDs: HIV-1, 1A30; HIV-2, 1HSI; SIV, 1YTH; FIV, 1B11; EIAV, 1FMB; and RSV, 1BAI). These structures have the following mutations from the wild-type proteases: HIV-1: Q7K, L33I, L63I; SIV: Q7K; EAIV: I54G; RSV: S38T, I42D, I44V, M73V, A100L, V104T, R105P, G106V, and S107N. Hydrophobic residues are indicated in yellow, hydrophilic residues in green, acidic residues in red, and basic residues in blue. Abbreviations: BLV, bovine leukemia virus; MuLV, murine leukemia virus; SIV, simian immunodeficiency virus; FIV, feline immunodeficiency virus; EIAV, equine infectious anemia virus; RSV, Rous sarcoma virus.

Table 1. Retroviral Protease Cleavage Sites

Cleavage Junction	HTLV-I	BLV	MuLV	RSV	HIV-1	FIV	EIAV
MA/CA	PQVL/PVMH	PAIL/PIIS	-	VVAM/PVVI	SQNY/PIVQ	PQAY/PIQT	SEEY/PIMI
MA/pX	-	-	SSLY/PALT	-	-	-	-
X/CA	-	-	SQAF/PLRA	-	-	-	-
CA/NC	TKVL/VVQP	PAIL/VHTP	SKLL/ATVV	AAAM/SSIA	-	KMQL/LAEL	-
CA/pX	-	-	-	-	ARVL/AEAM	-	KMML/LAKA
pX/NC	-	-	-	-	ATIM/MQRG	-	AKAL/QTGL
NC/pX	-	-	-	-	RQAN/FLGK	-	-
Gag/PR	ASIL/PVIP	LECL/LSIP	TSLL/TLDD	PAVS/LAMT	SFNF/PQIT	GFVN/YNKV	QFVG/VTYN
PR/Pol	PVIL/PIQA	PMVG/VLDA	LQVL/TLNL	LTNL/IGRA	TLNF/PISP	RLVM/AIQS	KLVL/AQLS
Pro/RT	PAVL/GLEL	-	-	-	-	-	-
RT-RH/IN	VLQL/SPAD	unk	unk	FQAY/PLRE	RKIL/FLDG	CQTM/MIIE	TGVF/WVEN

The scissile bond is indicated by the slash.

proteases is the presence of additional amino acids at the C terminus. It can be seen from the sequence alignment with other retroviral proteases that the leukemia virus proteases are the only ones that contain these "extra" amino acids. There is a report that suggests the ten C-terminal residues of HTLV-I protease are required for activity, but this was based on the detection of unprocessed Gag produced during coexpression of the deletion mutant and the Gag polyprotein [23]. However, the mutant was not purified and tested further. On the other hand, a BLV protease mutant in which the ten C-terminal residues were deleted has been reported and it retains proteolytic activity [24]. Recently, a mutant of HTLV-I protease in which the ten C-terminal residues have been removed was prepared and cleavage data indicate that this deletion has no effect on activity (S. Shuker et al., submitted). The function of these "extra" residues, if any, is unclear, but it is evident that they are not necessary for the proteolytic activity of these enzymes.

Many of the natural cleavage sites for HTLV-I protease have been identified [19, 20, 25, 26], and quantitative measurements of the kinetic parameters of HTLV-I protease cleavage of peptides corresponding to the MA/CA, CA/NC, and PR/pX processing sites [19, 27–30], as well as processing sites for other retroviral polyproteins [28, 29], have been reported (Table 2). From these data, it can be seen that although only seven amino acids are typically required for substrate recognition, changes in the length of the peptide substrate have a significant effect on the kinetic parameters for cleavage. It was also recently demonstrated that activity of HTLV-I protease is not dependent on the concentration of NaCl [30]. This is in contrast to a prior report stating that a concentration of 0.5 M $(NH_4)_2SO_4$ decreases K_m by 20-fold [19]. It has also been reported that there is an increase in proteolytic activity upon an increased ionic strength for HIV-1 and BLV proteases [31–33]. On the other hand, the activity of HIV-1 protease has been reported to be independent of sodium ion concentration, but decreases upon an increase in magnesium ion concentration [31]. This is attributed to dissociation of the homodimer because a red shift in tryptophan fluorescence is observed along with an increase in quantum yield, which is indicative of increased solvent exposure for those residues.

Unlike HIV protease, no structures of HTLV-I protease have been reported to date. To elucidate the factors important for the specificity of the protease, a model (Dennison, Herger, Shuker: PDB ID 1O0J) was generated based on homology to a mutant of RSV protease, whose structure has been reported [29, 34] (PDB ID 1BAI). Although HTLV-I protease has higher homology to HIV protease, the protease from RSV was chosen because fewer insertions are required and too many insertions can result in a less accurate model. While the relatively

Table 2. Kinetic Parameters for Cleavage of HTLV-I Junctions and Junctions from Other Retroviruses by HTLV-I Protease

Cleavage Junction	Peptide Substrate	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($mM^{-1}s^{-1}$)	Ref
MA/CA	APQVL/PVMHP	8.7 ± 0.6	120 ± 18	75	27
	APQVL/PVMHP	5.9 ± 0.6	69 ± 20	85	28
	PQVL/PVMHP	0.22 ± 0.01	60 ± 7	3.7	30
	YVEPTAPQVL/PVMHP	0.085	500	0.17	19
PR/p3	KGPPVIL/PIQAP	11.8 ± 0.9	40 ± 10	290	28
	KGPPVIL/PIQAP	6.9 ± 0.4	30 ± 6	230	27
CA/NC	KTKVL/VVQPK	5.2 ± 0.1	33 ± 4	160	27
	KTKVL/VVQPK	7.7 ± 0.2	51 ± 5	150	28
	KTKVL/VVQPK	10.0	63	160	29
	TKVL/VVQPK	2.7	87	31	29
gag/PR	DPASIL/PVIP	0.7 ± 0.05	230 ± 35	3.0	27
HIV-1 MA/CA	VSQNY/PIVQ		not cleaved		29
BLV MA/CA	PPAIL/PIISE	3.3 ± 0.2	20 ± 6	165	28
BLV CA/NC	KQPAIL/VHTPG		not cleaved		28
BLV PR/pX	PPVMG/VLDAP	0.010 ± 0.001	14 ± 4	0.7	28

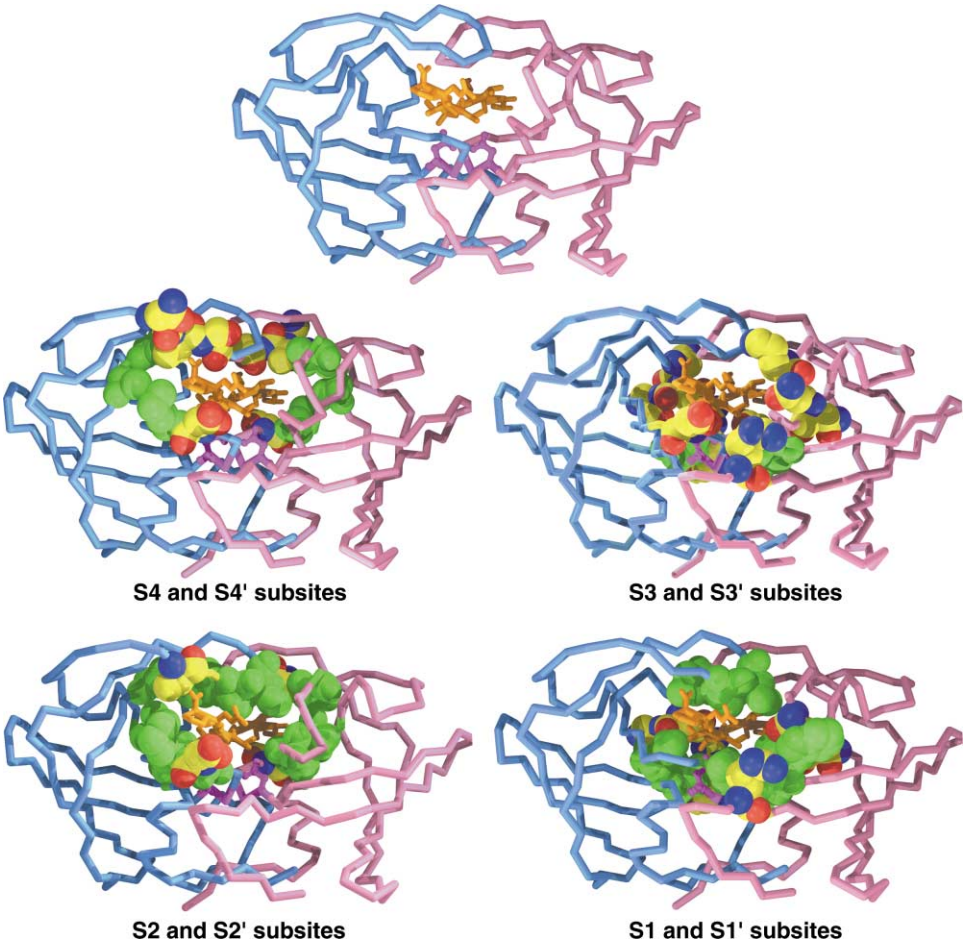


Figure 3. Homology Model of HTLV-I Protease without the Ten C-Terminal Residues

The subsites are shown in spacefilling representation (S4 and S4′, top left; S3 and S3′, top right; S2 and S2′, bottom left; S1 and S1′, bottom right) with hydrophobic residues colored in green and hydrophilic residues colored by element. The catalytic aspartic acids are shown in purple and the substrate is shown in orange. PDB ID of HTLV-I protease is 1O0J.

low homology to RSV protease (25%) calls into question the reliability of the finer details of the model, it is useful for identifying residues that may be important for specificity of the subsites, which are shown in Figure 3. In addition, Tözsér and coworkers have also prepared a model and have reported the effect on the kinetic parameters of amino acid changes in a peptide corresponding to the CA/NC cleavage site (summarized in Table 3) [29]. These data, in conjunction with the homology model and characterization of native HTLV-I and other retroviral processing sites, provide insight into the specificity for the protease subsites.

S4 and S4′ Subsites

The S4 and S4′ subsites are shallow, surface-exposed pockets formed by residues Asp36, Met37, Asn53, Thr54, Ser55, Cys90, and Val92. In the natural substrates, the P4 and P4′ positions are most often hydrophobic residues, with proline being the most prevalent. A variety of side chains are tolerated, however, including polar (serine and threonine) and charged (aspartic acid and histidine) residues. Truncation of the CA/NC peptide to remove the amino acid at P4 results in a decrease in catalytic efficiency. Replacement of the native CA/NC P4 threonine with valine or leucine produces an increase in k_{cat}/K_m due to a lower K_m value. Other uncharged residues cause a slight decrease in catalytic efficiency, while replacement with aspartic acid results in a dramatic decrease in cleavage efficiency. In all cases, the change in efficiency is due to a change in K_m. Not surprisingly, hydrophobic residues favor the association of the substrate with the protease. Therefore, although this is a surface exposed binding pocket made up of mostly polar or charged residues, peptides with a hydrophobic amino acid at P4 are cleaved with the highest efficiency. No substitutions have been reported for P4′ in a natural substrate.

S3 and S3′ Subsites

The S3 and S3′ subsites are large pockets that can accommodate a variety of side chains, as seen in the various P3 and P3′ groups found in the natural substrates (alanine, valine, leucine, glutamic acid, glutamine, serine, lysine; see Table 1). They are formed by a combination of hydrophobic side chains closer to the protein interior (Leu30) and charged and hydrophilic side chains at the protein surface (Arg10, Asp36, Lys95, Asn96, and Asn97). Substrate hydrophobic residues can

Table 3. Catalytic Efficiency of HTLV-I Protease Cleavage of Peptides Corresponding to the CA/NC Site with Single Amino Acid Substitutions

Peptide Substrate	k_{cat}/K_m ($mM^{-1}s^{-1}$)	Peptide Substrate	k_{cat}/K_m ($mM^{-1}s^{-1}$)
KTKVL/VVQPK	159	KTKV**F**/VVQPK	335
TKVL/VVQPK	31.0	KTKV**Y**/VVQPK	61.6
KVL/VVQPK	only slight cleavage	KTKV**A**/VVQPK	30.0
K**V**KVL/VVQPK	267	KTKV**M**/VVQPK	27.3
K**L**KVL/VVQPK	210	KTKV**G**/VVQPK	2.2
K**G**KVL/VVQPK	98.6	KTKV**S**/VVQPK	only slight cleavage
K**P**KVL/VVQPK	58.9	KTKV**D**/VVQPK	not cleaved
K**S**KVL/VVQPK	40.0	KTKV**K**/VVQPK	not cleaved
K**A**KVL/VVQPK	36.0	KTKVL/**L**VQPK	134
K**D**KVL/VVQPK	2.3	KTKVL/**I**VQPK	12.5
KT**F**VL/VVQPK	145	KTKVL/**F**VQPK	2.0
KT**A**VL/VVQPK	114	KTKVL/**P**VQPK	1.1
KT**L**VL/VVQPK	112	KTKVL/**S**VQPK	only slight cleavage
KT**V**VL/VVQPK	10	KTKVL/**G**VQPK	not cleaved
KT**S**VL/VVQPK	38.9	KTKVL/**D**VQPK	not cleaved
KT**G**VL/VVQPK	19.4	KTKVL/**K**VQPK	not cleaved
KT**D**VL/VVQPK	12.8		
KTK**L**L/VVQPK	118		
KTK**I**L/VVQPK	63.1		
KTK**F**L/VVQPK	11.5		
KTK**A**L/VVQPK	0.4		
KTK**G**L/VVQPK	not cleaved		
KTK**N**L/VVQPK	not cleaved		
KTK**D**L/VVQPK	not cleaved		
KTK**K**L/VVQPK	not cleaved		
KTK**S**L/VVQPK	not cleaved		

interact with the interior, hydrophobic portion of the binding site, while hydrophilic residues can interact with the more polar groups near the surface. Although the natural amino acid in the P3 position in the CA/NC peptide is a lysine, peptides with hydrophobic residues such as phenylalanine, leucine, and valine in this position are cleaved with almost the same catalytic efficiency. Substitution with an aspartic acid, serine, or glycine at the P3 position, however, causes a decrease in activity. Interestingly, an acidic residue is found at P3′ in the natural pX/RT cleavage site and the BLV PR/pX junction PPVMG/VLDAP is cleaved by HTLV-I protease. No substitutions have been reported for the P3′ position; however, inspection of the natural substrates and other polyprotein junctions that have been cleaved with HTLV-I protease reveals that a variety of side chains are accommodated. The P3 and P3′ positions appear to be the most tolerant to a variety of different side chains, and charged amino acids in the native substrates are most often found in these positions.

S2 and S2′ Subsites

The S2 and S2′ subsites are large, hydrophobic pockets that accommodate aliphatic groups in the natural substrates, particularly the β-branched residues valine and isoleucine. They are formed by predominately hydrophobic side chains (Ala35, Asp36, Met37, Val39, Thr54, Val56, Leu57, Val92, and Ile100). For the CA/NC peptide, the P2 position is much less tolerant to substitution than P3 or P4, and replacement of the native valine with leucine is the only reported substitution that does not cause a dramatic drop in k_{cat}/K_m. Substitution with charged or hydrophilic residues results in peptides that are not cleaved by the protease. Consistent with this is the observation that the HIV MA/CA peptide VSQNY/PIVQ, which has an asparagine at P2, is not cleaved by HTLV-I protease. No substitutions have been reported for the P2′ position, but it is evident from examination of the natural substrates that hydrophobic residues are preferred. The only peptide with a charged residue at P2′ whose incubation with HTLV-I protease has been reported is the BLV CA/NC substrate KQPAIL/VHTPG, and no cleavage is observed for this peptide. It is clear that hydrophobic residues are the preferred residues in the P2 and P2′ positions.

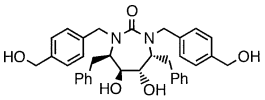
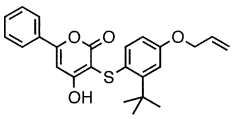
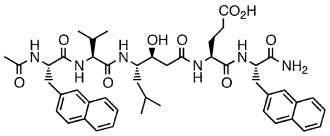
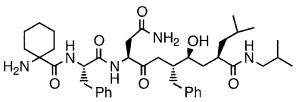
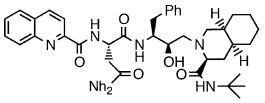
S1 Subsite

The S1 subsite is hydrophobic and is formed by residues Arg10, Leu30, Gly34, Val56, Leu57, Gln96, Gln97, Trp98, and Ile100. All of the natural substrates for HTLV-I protease have a leucine in the P1 position and amino acids with a β-branch are not accommodated [27]. The P1 position is quite sensitive to substitution, and replacement of the P1 leucine with a charged amino acid results in loss of cleavage while replacement with hydrophilic serine gives a peptide that shows only slight cleavage when incubated with the protease. Replacement with a phenylalanine gives an increase in k_{cat}/K_m, while other hydrophobic residues reduce the catalytic efficiency. The presence of a leucine at this position appears to be one of the critical factors for the specificity of HTLV-I protease.

S1′ Subsite

The S1′ subsite most often accommodates proline in natural HTLV-I substrates. Replacement of the native CA/NC P1′ valine with a leucine has little effect on k_{cat}/K_m, but all of the other reported substitutions result in less efficient cleavage. As with the P2 and P1 positions, substitution with charged or hydrophilic residues results in peptides that either are not cleaved or show only slight cleavage when incubated with the protease. This is in contrast to recent reports identifying new cleavage

Table 4. K_i Values for Compounds that Have Been Tested for Inhibition of HTLV-I Protease

Inhibitor	Structure	K_i (μM) HTLV-I	K_i (μM) HIV-1	Ref
DMP-323		>10	0.0003	27
MES13-099		0.24	0.007	35
LP-149		>10	0.0017	27
UK-88,897		>10	0.012	27
Ro-31-8959		>10	0.0004	27
N-1270	Arg-Val-Leu*-Phe-Glu-Ala-Nle-NH_2[a]	127	0.050	35
N-1465	Ac-Thr-Ile-Nle*-Nle-Gln-Arg[a]	12.7	0.78	35
N-1460	Ser-Gln-Asn-Phe*-Pro-Ile-Val-Gln[a]	7.3	1.0	35
1	Boc-Val-Val-Phe(PC)Phe-Val-Val-NH_2[b]	0.080	0.0004	27
JG-365	Ac-Ser-Leu-Asn-Phe-CH(OH)CH_2NH-Pro-Ile-Val-OMe	0.006	0.00066	35
2	Lys-Thr-Lys-Val-Sta-Val-Gln-Pro-Lys	>10	>10	27
3	Pro-Pro-Cys-Val-Phe-Sta-Ala-Met-Thr-Met	>10	ND[c]	27
4	Pro-Tyr-Val-Phe-Sta-Ala-Met-Thr	>10	12.5	27
5	Ala-Pro-Gln-Val-Sta-Val-Met-His-Pro	0.050	0.13	27
6	Ac-Pro-Gln-Ile-Thr-Leu-Trp-Gln-Arg-Pro-NH_2	1700	IC_{50} > 2000	22
7	Ac-Thr-Leu-Asn-Phe	252	IC_{50} = 1500	22
8	Ac-Thr-Val-Ser-Phe-Asn-Phe	217	6.1	22
N-1395	Ac-Leu-Val-Phe-H	93	IC_{50} = 0.90	35
9	Pro-Val-Ile-Pro-Leu-Asp-Pro-Ala-Arg-Arg-Pro-Val	87.2	17.1	22
10	Gln-Met-Gln-Gly-Val-Leu-Tyr-Leu	70.0	IC_{50} > 400	22
11	Ac-Leu-Lys-Ala-Gln-Ile-His-Phe	50.6	no inhibition	22
12	Tyr-Leu-Pro-Glu-Ala-Lys-Arg-Pro-Pro-Val-Ile-Leu	5.7	84	22

[a] Asterisk denotes a reduced peptide bond.
[b] PC denotes replacement of the amide bond with -P(O)(OH)- .
[c] ND = not determined

sites that contain a serine or glycine in this position. Interestingly, substitution of the CA/NC P1′ valine with a proline also results in less efficient cleavage, even though proline is the most prevalent P1′ amino acid in the other natural substrates. Clearly, the surrounding residues in the peptide substrate have an impact on the interactions of a particular side chain with the protease.

Identification of residues that form the substrate binding pockets in the model provides insight into the interactions that may be responsible for the specificity of the protease and suggest mutations that can be made to test the importance of each residue. However, no active site mutants have yet been reported. The only mutants that have been described to date are L40I, C90A, and C109A [27, 35]. In the protease primary sequence, leucine 40 is immediately followed by a proline, and examination of the surrounding residues, ...MTVL/PIAL..., show that this sequence is very similar to the processing sites that have been identified. Indeed, upon standing in solution, the protease will undergo self-cleavage at this position to produce inactive fragments. The mutation of leucine 40 to an isoleucine prevents autoproteolysis because β-branched amino acids are not accommodated within the S1 subsite and a more stable recombinant protease is provided [27]. Measurement of the kinetic parameters of the L40I mutant demonstrates that this mutation has no effect on the cleavage efficiency of natural substrates. The C90A and C109A mutants were also prepared to form a more stable protease [27]. These residues are not believed to be involved in the formation of structurally important disulfide bonds, and removing them was expected to yield protease that would be more amenable to structure determination due to elimination of potential unwanted disulfide bond formation. These mutations also had little effect on the kinetic parameters of cleavage.

Inhibitors of HTLV-I Protease

At the present time, there is no effective treatment for adult T cell leukemia and HTLV-I infection. For HIV-1 infection, there has been great success with the use of HIV-1 protease inhibitors. Some of these inhibitors have also been tested on HTLV-I protease; however, differences in the amino acid sequences of HTLV-I and HIV-1 protease produce differences in susceptibilities to these inhibitors. Table 4 summarizes what has been reported on the inhibition of HTLV-I protease [22, 27, 36]. The compounds MES13-099 and DMP-323 are nonpeptide inhibitors, while all of the others are peptides or modified peptides. MES13-099 has a K_i for HTLV-I protease that is submicromolar, while DMP-323 shows no measurable inhibition. However, MES13-099 is two orders of magnitude more potent against HIV-1 protease. Of the modified peptide inhibitors (LP-149, UK-88,947, Ro-31-8959, N-1270, N-1465, N-1460, entry 1, and JG-365), two (entry 1 and JG-365) have submicromolar K_i's, but again both are more potent against HIV-1 protease. The best inhibitor of HTLV-I protease reported to date is JG-365, which is the peptide-based compound Ac-SLNF(CH(OH)CH_2N-H)PIV-OMe. This predominantly hydrophobic molecule has substitution that places a phenylalanine in the S1 subsite and a proline in the S1′ subsite. As discussed in the previous section, a phenylalanine in the P1 position or a proline in the P1′ position typically results in peptides that make excellent substrates for the protease, and inhibition of the protease by JG-365 is consistent with this analysis of the active site binding pockets.

Several peptide-based inhibitors incorporating the amino acid statine, an isosteric mimetic of the tetrahedral intermediate of aspartyl protease-catalyzed peptide cleavage, have also been tested. The statine isobutyl side chain binds in the S1 subsite, analogous to the P1 leucine. Two of the peptides tested were designed to mimic HTLV-I cleavage sites: entry 5, APQV(Sta)VMHP (for the MA/CA junction) and entry 2, KTKV(Sta)VQPK (for the CA/NC junction). The APQV(Sta)VMHP peptide shows good inhibition, with a K_i of 50 nM, while the CA/NC mimic has no detectable inhibition. The other statine-containing peptides tested correspond to a Mason-Pfizer monkey virus protease cleavage site (entry 4) or an avian myeloblastosis virus protease cleavage site (entry 3) and neither inhibit HTLV-I protease, although both are potent inhibitors of their respective proteases. The final group of compounds listed in Table 4 is a set of peptides that corresponds to the N or C termini of retroviral proteases. Because the protease homodimer is formed by interactions between the N and C termini, peptides corresponding to these regions may disrupt the dimer structure and therefore inactivate the enzyme. Entries 6, 7, and 8 correspond to the N terminus of HIV protease, the C terminus of HIV protease, and the C terminus of HIV transcription factor, respectively, and all show little activity against HTLV-I protease. The N terminus of HTLV-I protease (entry 9), the C terminus of HLTV-I protease without the 10 C-terminal residues (entry 10), and the C terminus of MuLV protease without the 18 C-terminal residues (entry 11) show moderate inhibition of HTLV-I protease. The truncated C termini of HTLV-I and MuLV protease were used because these proteases contain "extra" amino acids that are not believed to be involved with formation of the dimer interface. Surprisingly, the most active peptide in this class is the one corresponding to the C terminus of full-length HTLV-I protease. Because it is unlikely that this region of the protease is important for interactions at the dimer interface, the activity of this peptide may not involve the disruption of the dimer but may result from binding of the peptide in the active site. The inhibition was not further characterized as competitive or noncompetitive, and more work will be required to determine the mode of action of these peptides.

In addition to the studies described herein, it has been demonstrated that the HIV-1 protease inhibitors approved for use in therapy fail to block HTLV-I protease Gag processing in vitro [37]. Combination chemotherapies have been tested in vivo but have had limited impact on the 7- to 18-month median survival time of individuals with the acute form of ATL, which is the most prevalent form [38]. Clearly, much remains to be done in the development of treatments for adult T cell leukemia, and more generally, for HTLV-I infection before the onset of ATL. Based on the success of HIV protease inhibitors and the low mutation rate of HTLV-I protease [39], HTLV-I protease is a promising target for chemotherapy.

Conclusions and Outlook

HTLV-I has been identified as a dangerous emerging pathogen and it is estimated that 20 million people worldwide are infected with this retrovirus. This is a significant problem because there are no effective treatments to control HTLV-I infection or to prevent or treat HTLV-I-induced ATL and TSP/HAM. The protease from HTLV-I, which produces the mature structural proteins and enzymes essential for viral replication, is an attractive target for inhibitor design. However, HTLV-I protease and its enzymology are just beginning to be investigated. Comparison of HTLV-I to other retroviral proteases reveals significant differences in amino acid sequence, substrate specificity, and response to protease inhibitors. It is clear that while certain generalizations can be made among the retroviral proteases, HTLV-I protease has features that make it distinct and impart unique characteristics to its activity. The research described in this review lays the foundation for further work that will delineate the important structural features of the enzyme, define the specificity of substrate and inhibitor binding within the active site, and ultimately lead to the development of potent and specific HTLV-I protease inhibitors.

References

1. Murphy, R.A., and Kingsbury, D.W. (1990). Virus taxonomy. In Virology, Second Edition, B.N. Fields and D.M. Knipe, eds. (New York: Raven Press), pp. 26–27.
2. Poiesz, B.J., Ruscetti, F.W., Gazdar, A.F., Bunn, P.A., Minna, J.D., and Gallo, R.C. (1980). Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. Proc. Natl. Acad. Sci. USA *77*, 7415–7419.
3. Yoshida, M., Seiki, M., Yamaguchi, K., Takatsuki, T., and Murphy, K. (1984). Monoclonal integration of human T-cell leukemia provirus in all primary tumors of adult T-cell leukemia suggests causative role of human T-cell leukemia virus in the disease. Proc. Natl. Acad. Sci. USA *81*, 2543–2637.

4. Gessain, A., Vernant, J.C., Maurs, L., Barin, F., Gout, O., Calender, A., and Dethe, G. (1985). Antibodies to human T-lymphotropic virus type-I in patients with tropical spastic paraparesis. Lancet *2*, 407–410.
5. Kaplan, J.E., Khabbaz, R.F., and Fukuda, K. (1993). Guidelines for counseling human T-lymphotropic virus type I (HTLV-I) and HTLV type II-infected persons. Transfusion *33*, 694.
6. Gessain, A. (1996). Virological aspects of tropical spastic paraparesis/HTLV-I associated myelopathy and HTLV-I infection. J. Neurovirol. *2*, 299–306.
7. Zucker-Franklin, D., and Pancake, B.A. (1998). Human T-cell lymphotropic virus type 1 tax among American blood donors. Clin. Diagn. Lab. Immunol. *5*, 831–835.
8. Ewald, P. (1996). Guarding against the most dangerous emerging pathogens: Insights from evolutionary biology. Emerg. Infect. Dis. *2*, 245–256.
9. Satcher, D. (1995). Emerging infections: getting ahead of the curve. Emerg. Infect. Dis. *1*, 1–6.
10. Coffin, J.M. (1990). Retroviridae and their replication. In Virology, Second Edition, B.N. Fields and D.M. Knipe, eds. (New York: Raven Press), pp. 1437–1500.
11. Seiki, M., Hattori, S., Hirayama, Y., and Yoshida, M. (1983). Human T-Cell Leukemia Virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. Proc. Natl. Acad. Sci. USA *80*, 8–22.
12. Carrington, C.V., Weiss, R.A., and Schulz, T.F. (1994). A truncated HTLV-I envelope protein, lacking the hydrophobic membrane anchor domain, is associated with cellular membranes and virions. Virology *202*, 61–69.
13. Koralnik, I.J., Lemp, J.F., Jr., Gallo, R.C., and Franchini, G. (1992). *In vitro* infection of human macrophages by human T-cell leukemia/lymphotropic virus type I (HTLV-I). AIDS Res. Hum. Retroviruses *8*, 1845–1849.
14. Nam, S.H., Copeland, T.D., Hatanaka, M., and Oroszlan, S. (1993). Characterization of ribosomal frameshifting for expression of Pol gene products of Human T-Cell Leukemia Virus Type-I. J. Virol. *67*, 196–203.
15. Ratner, L., Haseltine, W., Patarca, R., Livak, K., Starcich, B., Josephs, S., Doran, E., Rafalski, J., Whitehorn, E., Baumeister, K., et al. (1983). Complete nucleotide sequence of the AIDS virus, HTLV-III. Nature *313*, 277–284.
16. Toh, H., Ono, M., Saigo, K., and Miyata, T. (1985). Retroviral protease like sequence in the yeast transposon Ty1. Nature *315*, 691.
17. Louis, J.M., Dyda, F., Nashed, N.T., Kimmel, A.R., and Davies, D.R. (1998). Hydrophilic peptides derived from the transframe region of Gag-Pol inhibit the HIV-1 protease. Biochemistry *37*, 2105–2110.
18. Hatanaka, M., and Nam, S.H. (1989). Identification of HTLV-I Gag protease and its sequential processing of the Gag gene product. J. Cell. Biochem. *40*, 15–30.
19. Kobayashi, M., Ohi, Y., Asano, T., Hayakawa, T., Kato, K., Kakinuma, A., and Hatanaka, M. (1991). Purification and characterization of Human T-Cell Leukemia-Virus Type I protease produced in *Escherichia coli*. FEBS Lett. *293*, 106–110.
20. Hayakawa, T., Misumi, Y., Kobayashi, M., Ohi, Y., Fujisawa, Y., Kakinuma, A., and Hatanaka, M. (1991). Expression of Human T-Cell Leukemia Virus Type I protease in *Escherichia coli*. Biochem. Biophys. Res. Commun. *181*, 1281–1287.
21. Saiga, A., Tanaka, T., Orita, S., Sato, A., Sato, S., Hachisu, T., Abe, K., Kimura, Y., Kondo, Y., Fujiwara, T., et al. (1993). Human T-Cell Leukemia Virus Type 1 protease protein expressed in *Escherichia coli* possesses aspartic proteinase activity. Arch. Virol. *128*, 195–210.
22. Daenke, S., Schramm, H.J., and Bangham, C.R.M. (1994). Analysis of substrate cleavage by recombinant protease of Human T-Cell Leukemia-Virus Type-1 reveals preferences and specificity of binding. J. Gen. Virol. *75*, 2233–2239.
23. Hayakawa, T., Misumi, Y., Kobayashi, M., Yamamoto, Y., and Fujisawa, Y. (1992). Requirement of N-terminal and C-terminal regions for enzymatic activity of Human T-Cell Leukemia Virus Type-I protease. Eur. J. Biochem. *206*, 919–925.
24. Precigoux, G., Geoffre, S., Leonard, R., Llido, S., Dautant, A., Destaintot, B.L., Picard, P., Menard, A., Guillemain, B., and Hospital, M. (1993). Modeling, synthesis and biological activity of a BLV proteinase, made of (only) 116 amino acids. FEBS Lett. *326*, 237–240.
25. Heidecker, G., Hill, S., Lloyd, P.A., and Derse, D. (2002). A novel protease processing site in the transframe protein of human T-cell leukemia virus type 1 PR76(gag-pro) defines the N terminus of RT. J. Virol. *76*, 13101–13105.
26. Mariani, V.L., and Shuker, S. (2003). Identification of the RT-RH/IN cleavage site of HTLV-I. Biochem. Biophys. Res. Commun. *300*, 268–270.
27. Louis, J.M., Oroszlan, S., and Tözsér, J. (1999). Stabilization from autoproteolysis and kinetic characterization of the human T-cell leukemia virus type 1 proteinase. J. Biol. Chem. *274*, 6660–6666.
28. Zahuczky, G., Boross, P., Bagossi, P., Emri, G., Copeland, T.D., Oroszlan, S., Louis, J.M., and Tözsér, J. (2000). Cloning of the bovine leukemia virus proteinase in *Escherichia coli* and comparison of its specificity to that of human T-cell leukemia virus proteinase. Biochim. Biophys. Acta *1478*, 1–8.
29. Tozser, J., Zahuczky, G., Bagossi, P., Louis, J.M., Copeland, T.D., Oroszlan, S., Harrison, R.W., and Weber, I.T. (2000). Comparison of the substrate specificity of the human T-cell leukemia virus and human immunodeficiency virus proteinases. Eur. J. Biochem. *267*, 6287–6295.
30. Ha, J.J., Gaul, D.A., Mariani, V.L., Ding, Y.S., Ikeda, R.A., and Shuker, S.B. (2002). HTLV-1 protease cleavage of p19/24 substrates is not dependent on NaCl concentration. Bioorganic Chem. *30*, 138–144.
31. Tyagi, S.C., Simon, S.R., and Carter, C.A. (1994). Effect of pH and nonphysiological salt concentrations on Human Immunodeficiency Virus-1 protease dimerization. Biochem. Cell Biol. *72*, 175–181.
32. Szeltner, Z., and Polgar, L. (1996). Conformational stability and catalytic activity of HIV-1 protease are both enhanced at high salt concentration. J. Biol. Chem. *271*, 5458–5463.
33. Konvalinka, J., Heuser, A.M., Hruskovaheidingsfeldova, O., Vogt, V.M., Sedlacek, J., Strop, P., and Krausslich, H.G. (1995). Proteolytic processing of particle-associated retroviral polyproteins by homologous and heterologous viral proteinases. Eur. J. Biochem. *228*, 191–198.
34. Wu, J., Adomat, J.M., Ridky, T.W., Louis, J.M., Leis, J., Harrison, R.W., and Weber, I.T. (1998). Structural basis for specificity of retroviral proteases. Biochemistry *37*, 4518–4526.
35. Teruya, K., Kawakami, T., Akaji, K., and Aimoto, S. (2002). Total synthesis of L40I, C90A, C109A-human T-cell leukemia virus type 1 protease. Tetrahedron Lett. *43*, 1487–1490.
36. Ding, Y.S., Rich, D.H., and Ikeda, R.A. (1998). Substrates and inhibitors of human T-cell leukemia virus type I protease. Biochemistry *37*, 17514–17518.
37. Pettit, S.C., Sanchez, R., Smith, T., Wehbie, R., Derse, D., and Swanstrom, R. (1998). HIV type 1 protease inhibitors fail to inhibit HTLV-I gag processing in infected cells. AIDS Res. Hum. Retroviruses *14*, 1007–1014.
38. Matsushita, K., Matsumoto, T., Ohtsubo, H., Fujiwara, H., Imamura, N., Hidaka, S., Kukita, T., Tei, C., Matsumoto, M., and Arima, N. (1999). Long-term maintenance combination chemotherapy with OPEC/MPEC (vincristine or methotrexate, prednisolone, etoposide and cyclophosphamide) or with daily oral etoposide and prednisolone can improve survival and quality of life in adult T-cell leukemia/lymphoma. Leuk. Lymphoma *36*, 67–75.
39. Gessain, A. (1996). Epidemiology of HTLV-I and associated diseases. In Human T-Cell Lymphotropic Virus Type I, P. Hollsberg and D.A. Hafler, eds. (Chichester, UK: Wiley), pp. 31–64.