Research Article

Retroviral proteases: correlating substrate recognition with both selected and native inhibitor resistance

Gary S Laco

Roskamp Institute, Sarasota, Florida, USA

Received on September 26, 2017; Accepted on November 11, 2017; Published on December 10, 2017

Correspondence should be addressed to Gary S Laco; E-mail: gary.laco@gmail.com

Abstract

A diverse group of retroviral proteases were analyzed to correlate mechanisms of substrate recognition with resistance to HIV-1 protease active-site inhibitors. Here it was shown that HIV-1 protease utilized a pathway common to many retroviral proteases, for recognition of mutated Gag/Pol cleavage sites, in order to become resistant to active-site inhibitors. While HIV-1 and HIV-2 resulted from independent cross-species transmissions of simian immunodeficiency virus into humans, HIV-2 has native primary resistance to many HIV-1 protease inhibitors as do many other retroviral proteases. The native multi-drug resistance of those proteases contributed to the lack of treatments for the respective life-long infections. Analysis of interactions between retroviral proteases and Gag/Pol substrates revealed that protease interactions weighted towards cleavage site residues P4-P4' resulted in inhibitor sensitivity, while interactions weighted towards residues P12-P5/P5'-P12' gave inhibitor resistance. In addition, a mechanism was identified for human T-cell leukemia virus type-1 protease that allowed re-weighting of the protease interactions with substrate residues P4-P4' and P12-P5/P5'-P12' using anti-parallel beta-sheets that connected the protease flaps to the substrate-grooves. Those anti-parallel beta-sheets are common to all studied retroviral proteases. The critical role of the retroviral protease substrate-grooves in substrate recognition and inhibitor resistance makes them a potential target.

Introduction

Retroviruses infect a remarkably diverse range of vertebrate species that span fish to humans (Barre-Sinoussi *et al.* 1983, Fodor & Vogt 2002, Gallo *et al.* 1983). All retroviruses encode an aspartic acid protease (PR) in which two identical monomers each contribute a catalytic aspartic acid residue to the active-site of the symmetrical PR dimer (Kohl *et al.* 1988, Wlodawer *et al.* 1989). The retroviral PR is typically expressed as either a Gag-Pro polyprotein, or as a Gag-Pro-Pol polyprotein, due to translational frame shifting near the C-terminus of Gag (Jacks *et al.* 1988). As the virus buds from the host cell, the PR first undergoes autocatalytic processing out of the respective polyprotein, and then cleaves Gag to release the structural proteins that are needed for virion maturation: matrix; capsid; and nucleocapsid (Strickler *et al.* 1989). The PR also cleaves Pol to release the enzymes needed for completion of the viral replication cycle in the next host cell: reverse transcriptase and integrase (Strickler *et al.* 1989). The essential role of the HIV-1 PR in virion maturation and activation of the viral enzymes have made the HIV-1 PR an important target in the treatment of human immunodeficiency virus type-1 (HIV-1) infections. Clinical HIV-1 PR inhibitors are typically peptidomimetics based in part on native cleavage sites; the only non-peptidomimetic inhibitor in current clinical use is tipranavir, which perhaps as a consequence is associated with more severe adverse effects. The HIV-1 protease inhibitors are comparable in length to a four residue peptide substrate and are typically designed to bind to the HIV-1 PR active-site as a transition state mimetic with a water tetrahedrally coordinated between the inhibitor P1/P1' backbone carbonyl oxygens and the backbone nitrogens of the HIV-1 PR flaps, Figure 1. Novel side groups on the inhibitors increase both van der Waals and H-bond interactions with the HIV-1 PR active-site and that contributes to active-site inhibitors being able to outcompete Gag/Pol substrates for binding to HIV-1 PR (Erickson *et al.* 1990, Miller *et al.* 1989).

HIV-1 PR inhibitor resistance typically begins with the selection of primary mutations in the active-site that can either reduce van der Waals contacts due to shorter side-chains (i.e., Ile54Val), or change the electrostatic interactions (i.e., Asp30Asn), between the HIV-1 PR and inhibitor, Figure 1 (Wensing *et al.* 2015). Those active-site mutations can also decrease the interactions between HIV-1 PR and substrate resi-

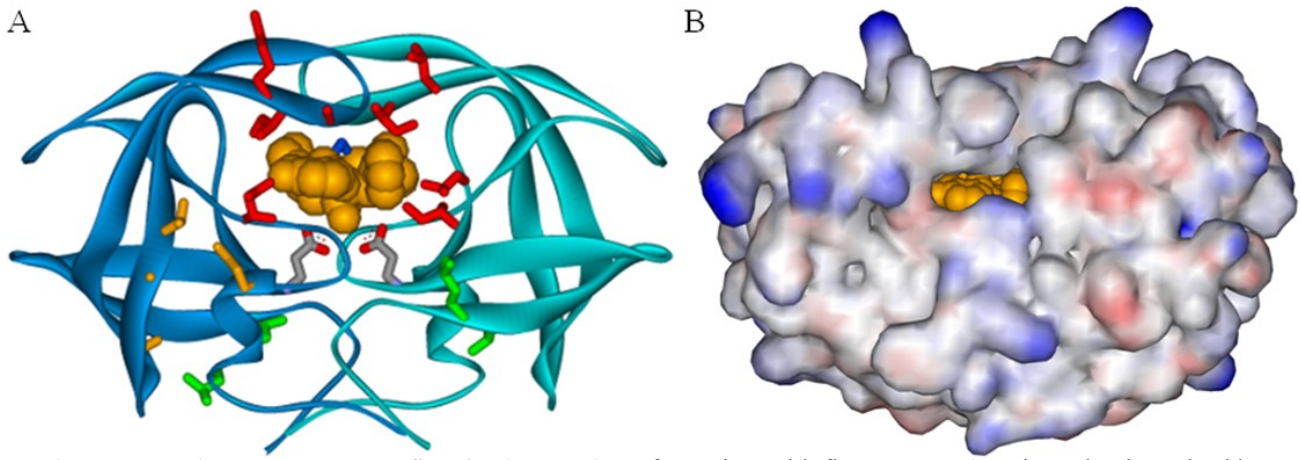

Figure 1. WT HIV-1 PR bound to nelfinavir. A, HIV-1 PR front view with flaps on top, A and B subunits as backbone ribbons, active-site Asp25 (A and B subunits, center) with side chains atoms in: oxygen, red; carbon, grey. Nelfinavir (bronze, CPK rendering) bound in the active-site, hydrogens not shown. Tetrahedrally coordinated water (center, above nelfinavir) with all atoms in blue. Solid colored PR residues that when mutated contributed to either primary, or secondary, inhibitor resistance are shown only once on either the A or B subunits. Primary active-site residues in red, clockwise from center of left subunit: Asp30; Ile47; Met46; Gly48; Ile54; Ile50; Val82; Ile84. Secondary S-groove residues in bronze, clockwise from bottom of left subunit: Ala71; Gly73; Thr74; Asn88. Secondary cleavage-site residues with residue and cleavage site indicted in green, clockwise from bottom of left subunit: Ile93 (P7 PR/RT); Leu90 (P10 PR/RT); Leu10 (P10' p6*/PR); Val11, P11' (p6*/PR). B, HIV-1 PR front view (flaps on top) with electrostatic surface potential (red negative, blue positive) and bound nelfinavir (bronze, CPK rendering).

dues P4-P4' resulting in reduced viral replicative capacity, Figure 2 (Chang & Torbett 2011, Gulnik *et al.* 1995, Kaplan *et al.* 1994, Martinez-Picado *et al.* 1999, Nijhuis *et al.* 1999, Pazhanisamy *et al.* 1996, Prabu-Jeyabalan *et al.* 2004, Schock *et al.* 1996). Next, selection of secondary resistance mutations distant from the active-site (i.e., Ala71Leu, Gly73Thr) restored both HIV-1 PR binding to Gag/Pol cleavage sites and viral replicative capacity while maintaining inhibitor resistance, Figure 1 and 2 (Chang & Torbett 2011, Gulnik *et al.* 1995, Kaplan *et al.* 1994, Martinez-Picado *et al.* 1999, Nijhuis *et al.* 1999, Pazhanisamy *et al.* 1996, Prabu-Jeyabalan *et al.* 2004, Schock *et al.* 1996)

In order to gain insight into the evolution of multi-drug resistant HIV-1 PR (MDR HIV-1 PR) researchers initially focused on understanding HIV-1 PR interaction with the ten primary Gag/Pol polyprotein cleavage sites by using short peptide substrates containing Gag/Pol cleavage site residues P4-P4'. However, a consistent cleavage order for the short peptides could not be determined in part because in order to increase peptide solubility researchers used different assay conditions and peptides of different lengths, including non-native N-and C-terminal residues (Billich

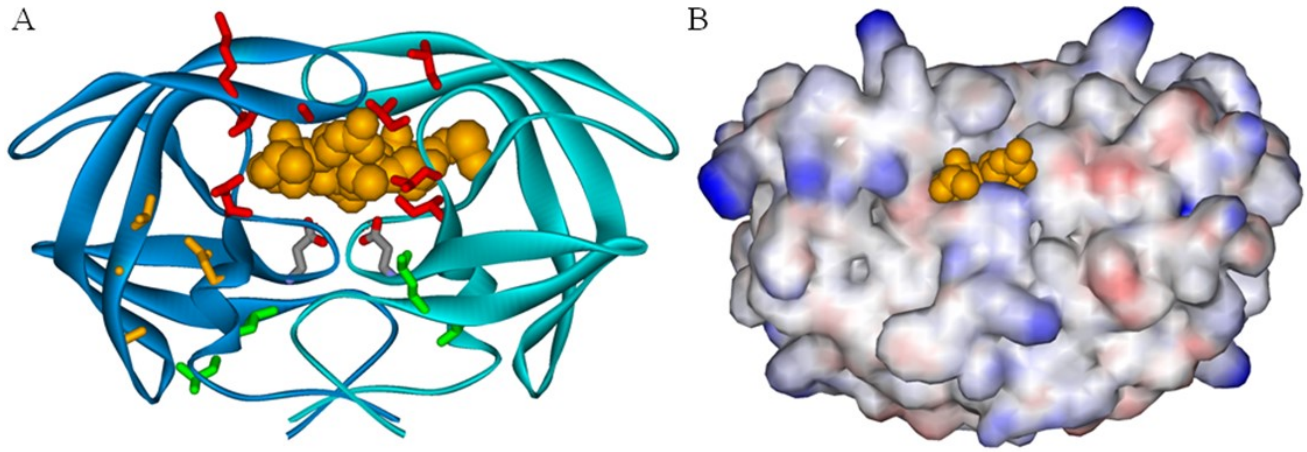

Figure 2. WT HIV-1 PR bound to Gag SP1/NC cleavage site 8-mer. A, HIV-1 PR front view with flaps on top, A and B subunits as backbone ribbons, active-site mutation D25N (A and B subunits, center) side chains atoms: oxygen, red; nitrogen blue; carbon, grey. SP1/NC cleavage site residues P4-P4' (bronze, CPK rendering) bound in the active-site, hydrogens not shown. Tetrahedrally coordinated water not visible (see Fig. 1 A). Solid colored PR residues that when mutated contributed to either primary, or secondary, inhibitor resistance are shown only once on either the A or B subunits. Primary active-site residues in red, clockwise from center of left subunit: Asp30; Ile47; Met46; Gly48; Ile54; Ile50; Val82; Ile84. Secondary S-groove residues in bronze, clockwise from bottom of left subunit: Ala71; Gly73; Thr74; Asn88. Secondary cleavage-site residues with residue and cleavage site indicted in green, clockwise from bottom of left subunit: Ile93 (P7 PR/RT); Leu90 (P10 PR/RT); Leu10 (P10' p6*/PR); Val11, P11' (p6*/PR). B, HIV-1 PR front view (flaps on top) with electrostatic surface potential (red negative, blue positive) and bound SP1/NC cleavage site residues P4-P4' (bronze, CPK rendering).

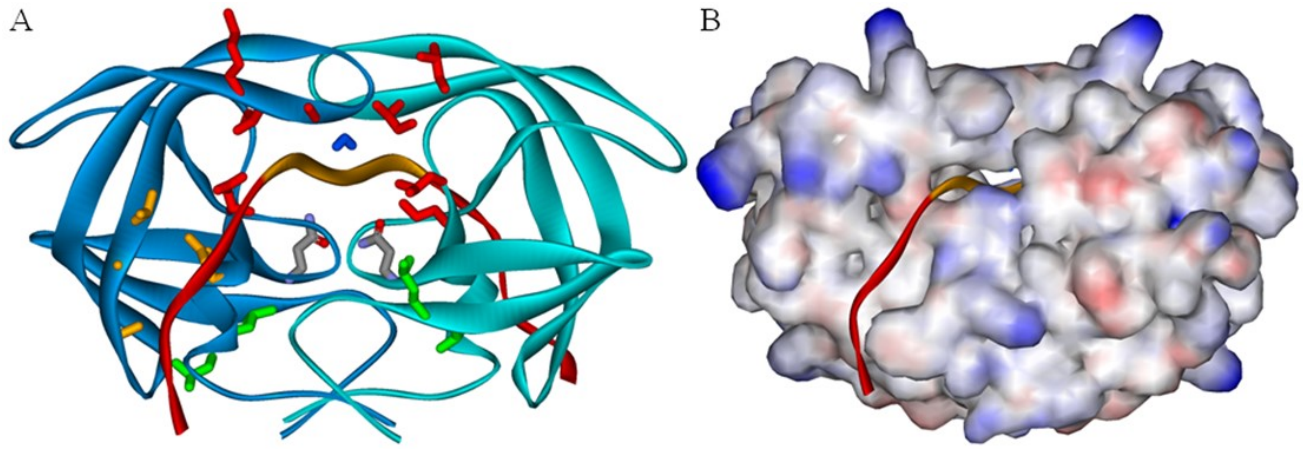

Figure 3. WT HIV-1 PR bound to Gag SP1/NC cleavage site 24-mer. A, HIV-1 PR front view with flaps on top, A and B subunits as backbone ribbons, active-site mutation D25N (center, A and B subunits) side chains atoms: oxygen, red; nitrogen, blue, carbon, grey. SP1/NC cleavage site residues P12-P12' (bronze ribbon |P12-P5|, red ribbon P4-P4') bound in the active-site and S-grooves, hydrogens not shown. Tetrahedrally coordinated water (center, above substrate) with all atoms blue. Solid colored PR residues that when mutated contributed to either primary, or secondary, inhibitor resistance are shown only once on either the A or B subunits. Primary active-site residues in red, clockwise from center of left subunit: Asp30; Ile47; Met46; Gly48; Ile54; Ile50; Val82; Ile84. Secondary S-groove residues in bronze, clockwise from bottom of left subunit: Ala71; Gly73; Thr74; Asn88. Secondary cleavage-site residues with residue and cleavage site indicted in green, clockwise from bottom of left subunit: Ile93 (P7 PR/RT); Leu90 (P10 PR/RT); Leu10 (P10' p6*/PR); Val11, P11' (p6*/PR). B, HIV-1 PR front view with flaps on top and electrostatic surface potential (red negative, blue positive) and bound SP1/NC cleavage site residues P12-P12' (bronze ribbon |P12-P5|, red ribbon P4-P4').

et al. 1988, Darke *et al.* 1988, Kotler *et al.* 1988, Krausslich *et al.* 1989, Tozser *et al.* 1991). It was only when *in vitro* transcribed/translated full length Gag and Gag-Pro-Pol polyproteins and purified full-length Gag polyprotein were used as substrates for HIV-1 PR that a consistent cleavage order was determined (Erickson-Viitanen *et al.* 1989, Pettit *et al.* 2005). These results indicated that cleavage site residues outside of P4-P4' were important for HIV-1 PR recognition of substrates and provided a clue as to how the MDR HIV-1 PR evolved. More recently, it was shown that the HIV-1 PR can also bind the Gag MA/CA cleavage site residues P12-P5/P5'-P12' (|P12-P5|) in the substrate-grooves (S-grooves), one on each face of the symmetric HIV-1 PR dimer, Figure 3 (Laco 2015). In addition, recent NMR studies on HIV-1 PR interaction with Gag polyproteins revealed that residues in the HIV-1 PR S-grooves, as previously defined (Laco 2015), interacted with Gag cleavage site residues outside of P4-P4' (Deshmukh *et al.* 2017). Here those findings were extended *in silico* to include the interaction between WT and MDR HIV-1 PRs and nine Gag/Pol cleavage sites represented as either 8-mers (P4-P4'), or 24-mers (P12-P12'), to evaluate the importance of the S-groove interaction with substrates, while eliminating the inherent variables of *in vitro* peptide cleavage assays. By understanding HIV-1 PR substrate interactions, a strategy can be developed to target HIV-1 that express MDR PR.

In contrast to HIV-1 PR where inhibitor resistance can be selected for *in vitro* and *in vivo*, many other retroviruses express PRs with native resistance to HIV-1 PR active-site inhibitors. For example, whereas HIV-1 and HIV-2 originated from independent transmissions of simian immunodeficiency virus (SIV) into humans (Hirsch *et al.* 1989, Huet *et al.* 1990, Marx *et al.* 1991, Peeters *et al.* 1989), HIV-2 PR has native primary resistance to many HIV-1 PR clinical inhibitors (Brower *et al.* 2008, Masse *et al.* 2007, Rodes *et al.* 2006, Witvrouw *et al.* 2004). Similarly, while human T-cell leukemia virus type-1 (HTLV-1) resulted from the cross species transmission of simian T-cell leukemia virus (STLV) into humans (Koralnik *et al.* 1994, Voevodin *et al.* 1997), the HTLV-1 PR has native multi-drug resistance to HIV-1 PR inhibitors (Ding *et al.* 1998, Pettit *et al.* 1998). In addition, both equine infectious anemia virus (EIAV) and feline immunodeficiency virus (FIV) PRs have been reported to have native multi-drug resistance to HIV-1 PR inhibitors (Kervinen *et al.* 1998). The native multi-drug resistance of HTLV-1, EIAV, and FIV PRs is a key reason why there are no effective treatments for the respective life-long retroviral infections. In the case of HTLV-1, long-term untreated infections can result in several debilitating neurological diseases, as well as a rapidly progressing and terminal T-cell leukemia (Goncalves *et al.* 2010, Proietti *et al.* 2005). The published structures of the above PRs allowed for the *in silico* analysis of the interactions between the PRs and both substrates and inhibitors, Results (Gustchina *et al.* 1996, Kovalevsky *et al.* 2008, Laco *et al.* 1997, Li *et al.* 2005, Rose *et al.* 1996a). In addition, the generation of three-dimensional (3D) models of the HIV-1 Gag and Pol polyproteins provided insight into the role of cleavage site accessibility on PR recognition of substrates. The analysis revealed the native resistance

mechanisms for these PRs, and will be important in the development of effective inhibitors for retroviruses that express PRs with either selected, or native, resistance to active-site inhibitors.

Materials and Methods

Computational chemistry

PR/ligand models were energy minimized prior to calculation of interaction energy scores using Accelrys Discovery Studio (Dassault Systèmes, San Diego, CA) with parameters set to approximate *in vitro* conditions for direct interactions between proteins and ligands as previously described (Laco 2011, 2015). The only explicit water in the PR substrate/inhibitor models was a water tetrahedrally coordinated between either the HIV -1 PR flaps Ile50 (A and B-subunits) backbone nitrogens, or structurally equivalent residues in the HIV-2/ SIV-cpz/SIV-sm/HTLV-1/EIAV/FIV PRs, and either the respective substrate residue P1 and P1' backbone carbonyl oxygens or inhibitor oxygens, harmonic restraints were placed on those H-bonded heavy atoms as previously described (Laco 2015). The indicated structure coordinate files were used to build the respective PR substrate/inhibitor models: WT HIV-1 PR bound to the Gag SP1/NC cleavage site was based on 1KJ7.pdb [19] with the addition of the HIV-1 PR HXB2 residues Val3 and Ser37 (Ratner *et al.* 1985), the 24-mer SP1/NC substrate backbone orientation served as the starting point for all PR/substrate models for consistency (Laco 2015). The MDR 3761 HIV-1 PR mutations relative to WT HIV-1 PR [46] were built into 1KJ4.pdb [19] as previously described (Laco 2015). HIV-1 substrates used in the *in silico* studies were based on the HIV-1 HXB2 strain (Ratner *et al.* 1985). The EIAV PR (1FMB.pdb), FIV PR (2FIV.pdb), HTLV-1 PR (2B7F.pdb) were treated the same as HIV-1 PR (Gustchina *et al.* 1996, Kovalevsky *et al.* 2008, Laco *et al.* 1997, Li *et al.* 2005). HIV-2 PR was generated by mutating the SIV-sm PR, as described herein, using the HIV-2 subtype A, isolate BEN, sequence (UniProtKB-P18096), uniprot.org (Consortium 2015). The SIV-cpz and SIV-sm PR models were generated by mutating the published SIV PR structure 1YTJ.pdb (Rose *et al.* 1996a) in Discovery Studio to the amino acid sequences of the SIV-cpz MB66 isolate (UniProtKB-Q1A267) and the SIV-sm S4 F236 isolate (UnitProtKB-P12502) (D'Arc *et al.* 2015), respectively, uniprot.org (Consortium 2015). The HTLV-1 PR was reverse engineered by either mutating two S-groove residues (I85A/T88G) to give HTLV-1 2X PR, or the two S-groove residues (I85A/ T88G) and two active-site residues (V56I/A59I) together to give HTLV-1 4X PR. The default Discovery Studio orientations for the mutated HTLV-1 2X and 4X PR side chains were used for the minimizations. Before starting the respective minimizations, the orientations of HTLV-1 2X and 4X PRs native residues, and bound 24-mer substrate residues, were identical to the HTLV-1 PR and bound 24-mer substrate.

HIV-1 Gag and Pol structure-based 3D models

The HIV-1 full-length Gag polyprotein (MA/CA/SP1/ NC/SP2/p6) and Pol polyprotein (p6*/PR/RT/RH/IN) 3D models were generated using published structures of Gag and Pol proteins (protein, PDB ID): MA-CA, 1L6N.pdb (Tang *et al.* 2002); CA-SP1, 4XFX.pdb (Gres *et al.* 2015); NC, 1MFS.pdb (Lee *et al.* 1998); p6, 2C55.pdb (Fossen *et al.* 2005); PR monomer, 1Q9P.pdb (Ishima *et al.* 2003); RT, 3T19.pdb (Gomez *et al.* 2011); IN, 4NYF.pdb (Wang *et al.* 2001), while the transframe encoded p6* structure was generated using the Robetta full-chain protein prediction server, http://robetta.bakerlab.org (Ovchinnikov *et al.* 2016). The proteins were bonded together using short random coil linkers containing missing residues, and then the structures were typed with the consistent force field (CFF) and energy minimized using implicit solvent and the same parameters in Discovery Studio as used above for HIV-1 PR (Laco 2015). Note: the relative

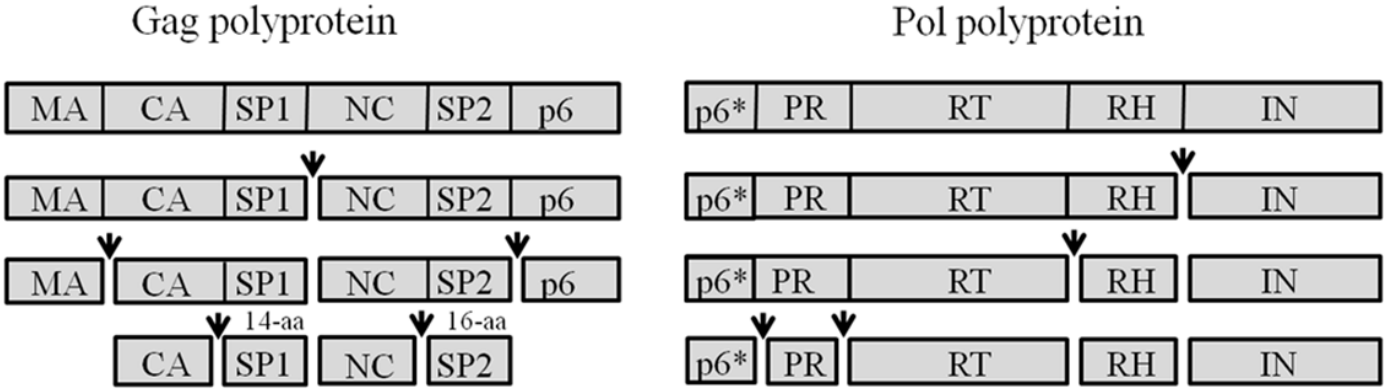

Figure 4. Schematic of HIV-1 Gag and Pol polyproteins and cleavage sites. Top row: interior vertical lines indicate cleavage sites; arrows indicate relative cleavage order from top to bottom for each polyprotein when cleaved in trans by HIV-1 PR (Pettit et al., 2005; Tritch *et al.*, 1991). The length of the CA/SP1 (14-aa) and NC/SP2 (16-aa) cleavage sites C-terminal tails are indicated. The Gag-Pro-Pol polyprotein is composed of Gag with a TF region starting after NC that changes the reading frame to express p6* and the Pol polyprotein (not shown).

orientations of the subdomains within the Gag and Pol polyproteins are in part for illustrative purposes as compared to the more extended orientations found within an immature HIV-1 virion (Bharat *et al.* 2014).

Protein sequence alignments

Protease 3D structures were aligned using PROMALS3D (Pei *et al.* 2008). The following pairs of viral Gag-Pro-Pol polyproteins were aligned using MUSCLE (Multiple Sequence Comparison by Log-Expectation, (Edgar 2004, McWilliam *et al.* 2015): HIV-1 (H2, group M, subtype B, isolate HXB2, UniProtKB-P04585) and SIV-cpz (isolate MB66, UniProtKB-Q1A267); and HIV-2 (subtype A, isolate BEN, UniProtKB-P18096) and SIV-sm (S4, isolate F236, UnitProtKB-P12502), uniprot.org (Consortium 2015).

Statistical analysis

The GraphPad QuickCalcs *t* test calculator and paired *t* test were used for all statistical analyses (http://www.graphpad.com/quickcalcs/ttest1.cfm).

Results

HIV-1 Gag/Pol cleavage site alignments, residues P12-P12'

The HIV-1 PR is essential for virion maturation in the infected cell and for activation of the enzymes required to complete the viral replication cycle in the next host cell (Kohl *et al.* 1988, Le Grice *et al.* 1988). Both processes require HIV-1 PR to cleave the ten primary Gag/Pol polyprotein cleavage sites in the correct order and rate, Figure 4.

How the HIV-1 PR accomplishes this is complicated by the fact that the HIV-1 PR cleavage sites do not have a consensus sequence for residues P4-P4' (Pearl & Taylor 1987). As a result, substrate residues bound by the HIV-1 PR S-grooves (|P12-P5|) were included in an alignment and revealed that even for this expanded cleavage site alignment (P12-P12') there were still no residues at any position conserved in all ten cleavage sites, and for P10' there were no conserved residues, Figure 5.

However, there were two positions where there was conservation at four cleavage sites: P2, Asn (4X); and P1, Phe (4X), Figure 5. The HIV-1 PR like other retroviral PRs, including the FIV PR, can bind peptide substrates in either the N-to C-terminal, or C-to N-terminal, orientation due to the structural symmetry of the PR dimer (Laco *et al.* 1997, Prabu-Jeyabalan *et al.* 2002). When the Gag/Pol cleavage site residues P12-P1 were aligned C-to N-terminal (i.e., P1-P12) with the respective P1'-P12' residues Phe was found at |P1| in three additional cleavage sites for a total of 7X, and His was found at |P12| in two additional cleavage sites for a total of 4X, followed by thirty-four residues/positions where the conservation increased by an additional cleavage site including at |P10| where six residues were conserved, Figure 5 yellow highlights. It was proposed that the heterogeneity of the Gag/Pol cleavage sites was due to the dual roles of the P12-P12' residues: first as recognition sequences that regulated Gag/Pol polyprotein cleavage order; and second as structural components in the N-and C-termini of the respective Gag/Pol cleavage products (Laco 2015), with the exception of the Gag derived SP1 and SP2 peptides, which are not part of any mature viral protein, Figure 4.

WT HIV-1 PR *in silico* interaction with Gag/Pol cleavage sites

The WT HIV-1 PR interactions with Gag/Pol cleavage sites were analyzed *in silico* using 8-mer and 24-mer peptide substrates in order to explore the role of the WT HIV-1 PR S-grooves in substrate recognition, Table 1 (Laco 2015). This approach avoided the issues associated with *in vitro* peptide assays, which while having good kinetic resolution suffered from variable results (Billich *et al.* 1988, Kotler *et al.* 1988, Krausslich *et al.* 1989, Tozser *et al.* 1991). The WT HIV-1 PR *in silico* interaction energy scores with substrates were compared to the *in vitro* cleavage of the Gag/Pol polyproteins. This was done to determine *in silico* which substrate length (8-mer/24-mer) best matched the published *in vitro* Gag/Pol cleavage order (Pettit *et al.* 2002, 2005). Previous results for WT HIV-1 PR bound to the Gag MA/CA and SP1/NC sites (Laco 2015) were based on structures of HIV-1 PR with substrates from both HIV-1 NL4-3 and HXB2 strains (King *et al.* 2012, Kozisek *et al.* 2007, Prabu-

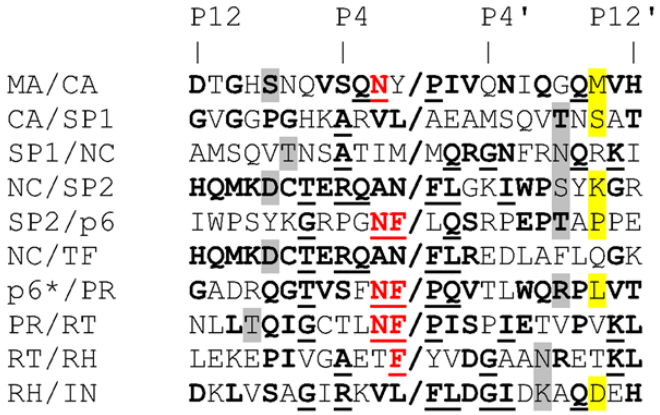

Figure 5. HIV-1 Gag/Pol cleavage site alignments P12-P12'. HXB2 sequence, / indicates scissile bond. Residue conservation across all sites indicated as follows: bold, 2X; underlined, 3X; red, 4X. The Gag NC/SP2 and NC/TF cleavage sites have identical residues P12-P2'. The translational frame shift used to generate the Gag-Pro-Pol polyprotein occurs between the NC/TF cleavage site residues P2' and P3'. The transframe region contains p6* at the C-terminus. Gray highlighted residues made electrostatic/H-bond interactions with the MDR HIV-1 PR S-groove resistance residue Thr73 (A/B subunits). P10' yellow highlighted residues were conserved when the P12-P1 residues were aligned C-to N-terminal (P1-P12) with the P1'-P12' residues (see Results, not shown).

Table 1. WT HIV-1 PR interaction energy scores for Gag/Pol substrates.

	HIV-1 PR in vitro Cleavage rate (Order)	HIV-1 PR in vitro (Cleavage Order)	WT HIV-1 PR in silico Interaction energy		24-mer Interaction Breakdown	
Cleavage Site	**Gag[1]**	**Gag-Pro-Pol[2]**	**8-mer**	**24-mer**	**|P12-P5|**	**P4-P4'**
MA/CA	10X↓(3)	(2)	-136	-248	-123	-125
CA/SP1	400X↓(5)	ND	-125	-205	-90	-114
SP1/NC	1 (1)	(1)	-122	-257	-141	-116
NC/SP2	350X↓(4)	ND	-121	-245	-137	-108
SP2/p6	9X↓(2)	ND	-120	-236	-123	-112
P6*/PR		(4)	-126	-226	-106	-120
PR/RT		(5)	-123	-224	-112	-112
RT/RH		(3)	-139*	-255*	-125	-130
RH/IN		(2)	-129*	-254*	-133	-121

WT HIV-1 PR interaction energy in kcal/mol for Gag and Pol cleavage sites as 8-mers and 24-mers, with a breakdown of the 24-mer substrates scores for |P12-P5| and P4-P4' (strongest interactions underlined). [1,2] The *in silico* interaction energy scores were compared to the published *in vitro* cleavage orders for the same sites in the context of Gag and Gag-Pro-Pol polyproteins where Gag-Pro-Pol contained an inactive PR with active HIV-1 PR supplied in trans (Pettit et al., 2002; Pettit et al., 2005). *The 3D context of RT/RH and RH/IN sites were examined in Results.

Jeyabalan *et al.* 2002, Wang *et al.* 2012). Here all substrates used for *in silico* interaction energy scores were based on the HIV-1 HXB2 strain for consistency, Figure 5 (Ratner *et al.* 1985). The WT HIV-1 PR interaction energy scores with nine Gag/Pol cleavage sites were reported alongside the published *in vitro* cleavage order for the same sites when part of Gag/Pol polyproteins, Table 1 (Pettit *et al.* 2002, 2005).

Notably, the scores for the MA/CA and SP1/NC site 8-mers were out of order with the published *in vitro* cleaved Gag/Pol polyproteins, while the 24-mers were in order (excluding RT/RH and RH/IN sites, Table 1). The RT/RH and RH/IN sites 8-mer and 24-mer scores were out of order due to being among the strongest interactions with WT HIV-1 PR (Table 1). Interestingly, the WT HIV-1 PR scores for the SP1/NC site went from one of the weaker 8-mer scores (incorrect order) to the strongest 24-mer score (correct order), Table 1. These results support the importance of the WT HIV-1 PR S-groove contacts with residues |P12-P5| in substrate recognition, Table 1 (Laco 2015). When the interaction energy scores between WT HIV-1 PR and substrates were broken down for substrate residues |P12-P1| versus P4-P4' it was found that four out of the nine cleavage site scores were weighted towards the active-site/P4-P4'. The PR/RT site scores for substrate residues |P12-P1| and P4-P4' were evenly distributed between the S-grooves and active-site, Table 1.

MDR HIV-1 PR *in silico* interaction with Gag/Pol cleavage sites

The evolution of HIV-1 PR multi-drug resistance typically begins with primary resistance mutations selected for in the active-site/flaps (i.e., Ile54Val and Asp30Asn), which reduces PR binding to inhibitors as well as Gag/Pol substrates (Laco 2015), resulting in inhibitor resistance along with reduced viral replicative capacity, Figure 1 Introduction. Next secondary resistance mutations are selected for in the PR S-grooves (i.e., A71I and G73T) that increased interactions with substrate residues |P12-P5| (Laco 2015), and that restored viral replicative capacity while maintaining inhibitor resistance, Figure 3 Introduction. The combination of HIV-1 PR primary and secondary resistance mutations resulted in multi-drug resistance and treatment failure (Chen *et al.* 1995, Gulnik *et al.* 1995, Schock *et al.* 1996).

Here the interactions between a MDR HIV-1 PR, HIV-1 PR 3761 (Wang *et al.* 2012), which contains seven resistance mutations including the S-groove A71I and G73T mutations (Laco 2015), and nine HIV-1 Gag/Pol cleavage sites was analyzed *in silico* (Table 2). Like the WT HIV-1 PR (Table 1), the MDR HIV-1 PR interactions with the MA/CA and SP1/NC site 8-mers were out of order with the *in vitro* cleaved Gag/Pol polyproteins, while the respective 24-mer scores were in order, Table 2. Similar to the WT HIV-1 PR, the MDR HIV-1 PR interactions with the RT/RH and RH/IN sites were the among the strongest making them out of order with the *in vitro* cleaved Gag/Pol polyproteins, Table 2. When the interactions between the MDR HIV-1 PR and substrates were broken down for substrate residues |P12-P5| and P4-P4', it was found that for eight out of the nine cleavage sites the interaction between the MDR HIV-1 PR and substrate was weighted towards the S-grooves and |P12-P5|. Only the CA/SP1 site was weighted towards the active-site and P4-P4', Table 2. The MDR HIV-1 PR S-groove secondary resistance residue Thr73 made direct electrostatic/H-bond interactions with residues in all nine of the Gag/Pol cleavage sites, see Figure 5

Table 2. MDR HIV-1 PR interaction energy scores for Gag/Pol substrates.

	HIV-1 PR in vitro Cleavage rate (Order)	HIV-1 PR in vitro Cleavage (Order)	MDR HIV-1 PR in silico Interaction energy		24-mer Interaction Breakdown	
Cleavage Site	**Gag[1]**	**Gag-Pro-Pol[2]**	**8-mer**	**24-mer**	**|P12-P5|**	**P4-P4'**
MA/CA	10X↓(3)	(2)	-126	-248	-132	-116
CA/SP1	400X↓(5)	ND	-115	-205	-101	-104
SP1/NC	1(1)	(1)	-117	-254	-147	-107
NC/SP2	350X↓(4)	ND	-110	-246	-144	-102
SP2/p6	9X↓(2)	ND	-116	-237	-127	-110
P6*/PR		(4)	-117	-244	-131	-113
PR/RT		(5)	-117	-224	-127	-97
RT/RH		(3)	-129*	-259*	-136	-123
RH/IN		(2)	-120*	-256*	-146	-110

MDR HIV-1 PR interaction energy in kcal/mol for Gag and Pol cleavage sites as 8-mers and 24-mers, with a breakdown of the 24-mer substrates scores for |P12-P5| and P4-P4' (strongest interactions underlined). [1,2] The *in silico* interaction energy scores were compared to the published *in vitro* cleavage orders for the same sites in the context of Gag and Gag-Pro-Pol polyproteins where Gag-Pro-Pol contained an inactive PR with active HIV-1 PR supplied in trans (Pettit et al., 2002; Pettit et al., 2005). *The 3D context of the RT/RH and RH/IN sites were examined in Results.

gray highlighted residues.

Interestingly, two incongruous results stand out in Tables 1 and 2. The first is the magnitude of difference in the *in vitro* cleavage rates reported for the CA/SP1 site (400-fold lower) and NC/SP2 site (350-fold lower) relative to the SP1/NC site when part of a Gag polyprotein (Pettit *et al.* 2002), and the 0.04 to 0.20-fold difference for the WT and MDR HIV-1 PRs *in silico* interaction energy scores for the CA/SP1 and NC/SP2 sites 24-mers relative to the SP1/NC site 24-mer (Tables 1 and 2). One explanation is that due to earlier cleavages at the SP1/NC and SP2/p6 sites, the CA/SP1 and NC/SP2 sites only had 14 and 16 residues, respectively, that extended past the C-terminal side of the scissile bond, Figure 4. This is in contrast to all other Gag/Pol cleavage sites that were stabilized by structured domains on both sides of the scissile bond, Figure 4. The orientations of the CA/SP1 and NC/SP2 site C-terminal tails could hinder HIV-1 PR binding to the respective cleavage site residues P1'-P12' *in vitro*, while *in silico* those potential issues were not taken into account. Complicating the issue is that the context of those sites within the Gag polyprotein could not be analyzed due to the lack of a full-length Gag structure. The second incongruous result from Tables 1 and 2 is that while the RT/RH and RH/IN sites were cleaved slower *in vitro* than the SP1/NC site in the respective Gag/Pol polyproteins, the RT/RH and RH/IN 24-mer substrates had some of the strongest *in silico* interactions with both the WT and MDR HIV-1 PRs (Tables 1 and 2). However, no obvious explanation could be proposed based on either the position of the cleavage sites in the Pol polyprotein, or the primary sequence of the cleavage sites, Figure 4 and 5. As a result, attention was turned to the 3D context of both the Gag CA/SP1 and NC/SP2 sites, and the Pol RT/RH and RH/IN sites. However, while the structures for nearly all of the mature proteins from the HIV-1 Gag/Pol polyproteins have been solved using crystallography and NMR, to our knowledge no full-length structure-based energy minimized 3D models of the HIV-1 Gag and Pol polyproteins have been published.

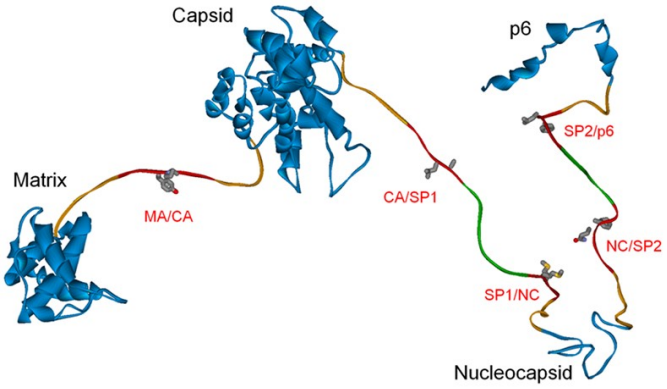

Figure 6. HIV-1 Gag polyprotein 3D model. Protein backbone as a ribbon, individual subdomains in blue with black labels, cleavage site backbone for residues |P12-P5| in bronze, P4-P4' in red. Cleavage site labels in red with residues P1/P1' shown: carbon, grey; oxygen, red; nitrogen, blue; sulfur, yellow. The overlapping portions of the CA/SP1 and SP1/NC, and NC/SP2 and SP2/p6, cleavage sites are green, Fig. 4 and 5. The overall orientation was influenced by illustrative considerations and in not meant to represent the more extended orientation within an immature virion.

HIV-1 Gag and Pol polyprotein 3D models

HIV-1 Gag and Pol polyprotein 3D models were built *in silico* using published structures of Gag and Pol derived proteins, while the p6* protein structure was generated using the Robetta full-chain protein prediction server, Materials and methods (Ovchinnikov *et al.* 2016). The individual proteins were connected as required with random coil linkers containing missing

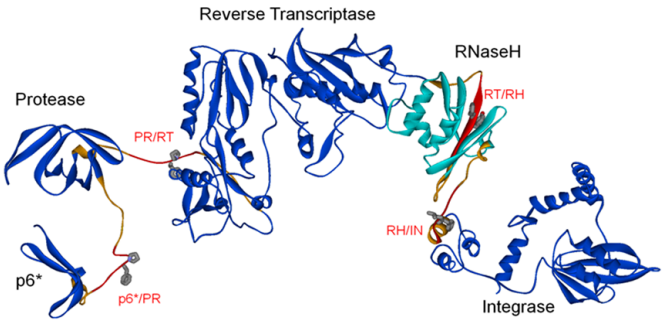

Figure 7. HIV-1 Pol polyprotein 3D model. Protein backbone as a ribbon, individual subdomains in blue with black labels, except for the RNaseH domain in teal. Cleavage site backbone for residues |P12-P5| in bronze, P4-P4' in red. Cleavage site labels in red with residues P1/P1' shown: carbon, grey; oxygen, red; nitrogen, blue. The overall orientation was influenced by illustrative considerations and is not meant to represent the more extended orientation within an immature virion.

residues to generate a full-length Gag polyprotein and Pol polyprotein with an N-terminal p6*, Figure 6 and 7. The Gag polyprotein 3D model revealed that all five cleavage site residues P12-P12' were solvent exposed in extended random coil configurations, Figure 6. As a result, the dramatically reduced *in vitro* cleavage rates for the CA/SP1 and NC/SP2 sites were likely due solely to the short C-terminal tails, Figure 4. Those C-terminal tails could be either disordered random coils, or form α-helices (Bharat *et al.* 2014), and in either case interfere with HIV-1 PR binding and cleavage (Table 1). This may explain why the Gag NC/SP2 24-mer scores for both the WT and MDR PRs were stronger than expected based on the *in vitro* Gag cleavage order, in that the high affinity primary sequence may be used to off-set the weaker *in vitro* binding of the PRs to the short C-terminal tail of the NC/SP2 cleavage site (Tables 1 and 2, Figure 4).

Compared to the Gag cleavage sites, the Pol cleavage sites were not as solvent accessible: the p6*/PR site residues P12-P8, and PR/RT site residues P12-P9, were part of α-helices, however, those structured residues likely contributed only a minor steric hindrance to HIV-1 PR binding since the remaining cleavage site residues were solvent exposed, Figure 7. In contrast, the Pol RT/RH site residues P12-P12' made extensive β-sheet contacts with the RNase H domain, while the RH/IN site residues P12-P6 and P3'-P12' were both part of α-helices that made interactions with RNase H β-sheets and integrase α-helices, respectively, Fig 7. Those finding explain why the Pol RT/RH and RH/IN site 24-mers (P12-P12') had some of the strongest *in silico* interactions with WT and MDR HIV -1 PRs (Tables 1 and 2), while *in vitro* they were reported to be the third and second fastest cleaved sites respectively, in that the high affinity primary sequence of the cleavage sites likely enhanced PR binding to the sites in the context of the Pol polyprotein, Figure 7. The results in Tables 1 and 2 indicated that the *in silico* interaction energy scores between both WT and MDR HIV-1 PRs and the 24-mer Gag/Pol cleavage sites (P12-P12') in general reflected the published *in vitro* cleavage order for sites that were in predominantly solvent exposed regions of the polyproteins, Tables 1 and 2. In contrast, the WT and MDR HIV-1 PRs *in silico* scores with the Gag/Pol cleavage site residues P4-P4' (8-mers) were more out of order with the reported *in vitro* cleavage orders and had lower resolution between the scores for the fast SP1/NC site and the slower CA/SP1, p6*/PR, and PR/

```
                    10/11       20    25   30 33          46 47
HIV-1          PQVTLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMI
SIV-cpz        PQITLWQRPILTVKIGGEIKEALLDTGADDTVIEEIQLEGKWKPKMI

               48 50  54                7173/74     82 84  88 90 93
HIV-1           GGIGGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
SIV-cpz        GGIGGFIKVKQYDNVIIEIQGKKAVGTVLVGPTPVNIIGRNFLTQIGCTLNF


                   10/11       20    25   30 33          46 47
HIV-1           PQVTLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMI
HIV-2          PQFSLWKRPVVTAYIEDQPVEVLLDTGADDSIVAGIELGDNYTPKIV
SIV-sm         PQFSLWRRPIVTAYIEEQPVEVLLDTGADDSIVAGIELGPNYTPKIV

               48 50  54                7173/74     82 84  88 90 93
HIV-1           GGIGGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
HIV-2          GGIGGFINTKEYKNVEIKVLNKRVRATIMTGDTPINIFGRNILTALGMSLNL
SIV-sm         GGIGGFINTKEYKDVKIKVLGKVIKGTIMTGDTPINIFGRNLLTAMGMSLNL
```

Figure 8. Structure-based HIV and SIV protease alignments. Top, HIV-1 PR/SIV-cpz PR; bottom, HIV-2 PR/SIV-sm PR with HIV-1 PR included as a resistance residue reference. Gray highlighted residues indicate amino acid identity between aligned PRs. HIV-1 PR residues that when mutated contributed to inhibitor resistance in bold, primary resistance residues indicated by underlined numbers, HIV-2 PR and SIV PR native resistance residues in bold. Active-site Asp25 (D25) italicized. Potential resistance residues in HIV-2 and SIV-sm at position 20 underlined.

RT sites, Tables 1 and 2.

Evolution of HIV-1 and HIV-2 from SIV: Structure -based PR alignments

The evolution of HIV-1 PR multi-drug resistance to active-site inhibitors results in treatment failure (Brown *et al.* 2003, Mocroft *et al.* 2003, Richman *et al.* 2004, Rosenbloom *et al.* 2012). This could represent a novel evolutionary pathway for HIV-1 PR inhibitor resistance. Alternatively, it could be a common pathway used by all retroviral proteases to adapt, during cross-species transmissions (Wolfe *et al.* 2005), to mutations in Gag/Pol cleavage sites selected for in the non-native host. Mutation of the Gag/Pol cleavage sites also occurs during HIV-1 adaptation to a new native host due to cytotoxic T lymphocyte (CTL) selective pressure on viral proteins including Gag/Pol and the respective cleavage sites (Phillips *et al.* 1991, Prince *et al.* 2012, Seibert *et al.* 1995). It is important to note that the HIV-1 ancestral SIV was from chimpanzees/SIV-cpz (Gao *et al.* 1999, Santiago *et al.* 2002), while the HIV- 2 ancestral SIV was from sooty mangabeys/SIV-sm (Lemey *et al.* 2003), since the amino acid sequences differ for both PRs, Figure 8. The SIV-cpz and SIV-sm PRs share 57% identity, Figure 8. It has been reported that both SIV-sm PR and HIV-2 PR had native resistance to clinical HIV-1 PR inhibitors: SIV-sm, amprenavir; HIV-2, amprenavir, atazanavir, nelfinavir, ritonavir, and tipranavir (Brower *et al.* 2008, Desbois *et al.* 2008, Rodes *et al.* 2006, Witvrouw *et al.* 2004). The 3D structures for HIV-1 and SIV-cpz PRs, and HIV-2 and SIV-sm PRs, were aligned to determine whether the PRs contained known HIV-1 PR resistance residues, Figure 8. HIV-1 PR had 83% identity with the SIV-cpz PR that had one native resistance residue (Val10), Figure 8. HIV-2 PR had 85% identity with the SIV-sm PR, with HIV-2 PR having seven native resistance residues while SIV-sm PR had six native resistance residues, Figure 8. The HIV-2, SIV-cpz, and SIV-sm PRs native resistance residues when found in HIV-1 PR at the same 3D position contributed to primary and secondary inhibitor resistance, http://hivdb.stanford.edu (Rhee *et al.* 2003, Wensing *et al.* 2015). When the HIV-2 PR and both SIV PRs were analyzed for the presence of S-grooves, all three PRs were found to have S-grooves similar to HIV-1 PR and HTLV-1 PR, Figure 3 and 10 (Laco 2015).

Evolution of HIV-1 and HIV-2 from SIV: Gag/Pol cleavage site alignments

While HIV-2 and SIV PRs were shown to contain native resistance residues, neither of those viruses had been exposed to HIV-1 PR active-site inhibitors during evolution, Figure 8. As a result, we explored how the HIV-1 and HIV-2 Gag/Pol cleavage sites evolved from the respective SIV Gag/Pol cleavage sites, during adaptation to the human host, to see how that may have contributed to the native resistance residues in the PRs.

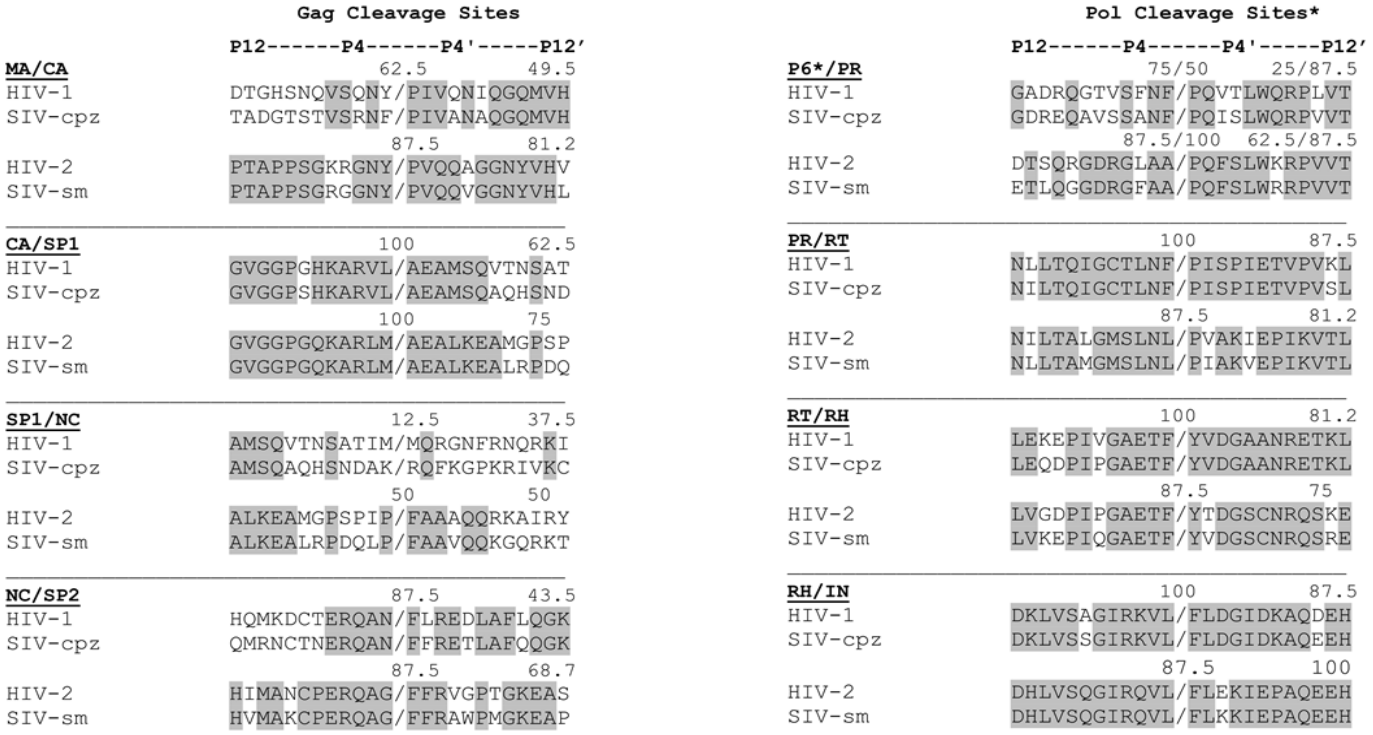

Figure 9. HIV and SIV Gag/Pol cleavage site alignments. Cleavage site residues P12-P12' shown, gray highlighted residues indicate identity between aligned sequences. For each cleavage site: top, HIV-1/SIV-cpz; bottom, HIV-2/SIV-sm. Percent identity for P4-P4' shown above cleavage site scissile bond (/); combined percent identity for |P12-P5| shown on the right side above P12'. *The p6*/PR cleave site percent identities were calculated separately for Gag (P12-P4 and P4-P1, left value) and Pol (P1'-P4' and P4'-P12', right value).

Table 3. Protease interaction energy scores with MA/CA 24-mer residues |P12-P5| and P4-P4'.

HIV-1	MDR HIV-1	HIV-2	SIV-cpz	SIV-sm	HTLV-1	EIAV	FIV
-123/-125	-132/-116	-120/-115	-135/-111	-127/-114	-128/-107	-152/-120	-130/-110

HIV-1, MDR HIV-1, HIV-2, SIV-cpz, SIV-sm, HTLV-1, EIAV, FIV PRs interaction energy scores (kcal/mol) with the corresponding MA/CA substrate residues |P12-P5| and P4-P4'.

The HIV-1 and HIV-2 Gag/Pol cleavage site residues (P12-P12') were aligned with the respective cleavage sites from the nearest ancestral SIV based on gag gene sequences, Figure 9 (D'Arc *et al.* 2015). The HIV-1 and HIV-2 cleavage site residues were divided into two groups (|P12-P5| and P4-P4') with the percent identity to the respective SIV cleavage sites indicated on the right side for |P12-P5|, and above the scissile bond for P4-P4', Figure 9. Note that the p6*/PR site residues P12-P1 were grouped with the Gag cleavage sites, while the residues P1'-P12' were grouped with the Pol cleave sites. HIV-1 had similar identity as HIV -2 with all of the respective SIV Gag cleavage site residues P4-P4' and p6*/PR residues P4-P1 (two-tailed P value = 0.1087), Figure 9. In contrast, HIV-2 had significantly more identity with the SIV-sm Gag cleavage site residues |P12-P5|, than did HIV-1 with the respective SIV-cpz cleavage site residues |P12-P5| (two-tailed P value = 0.0090), Figure 9. The HIV-1 and HIV -2 Pol cleavage site residues P12-P12' were overall highly conserved with the respective SIV Pol cleavage sites, Figure 9. These results indicated that HIV-2 had greater identity with SIV-sm Gag cleavage sites than did HIV-1 with SIV-cpz Gag sites, and the HIV-2 Gag cleavage site identities were more evenly distributed across cleavage site residues P12-P12', Figure 9. Mutation of the solvent accessible Gag cleavage sites, during viral adaptation to the human host, may have been one of the factors that drove the evolution of HIV-1 PR to weight substrate interactions towards the active-site and more conserved Gag cleavage site residues P4-P4', versus the S-grooves and divergent Gag cleavage site residues |P12-P5|, Figure 9.

Distribution of PR interactions with 24-mer substrates *in silico*

Based on the analysis of the HIV-1 and HIV-2 Gag/Pol cleavage site evolution from the respective SIVs the following hypothesis was proposed: due to CTL selective pressure in the human host (Phillips *et al.* 1991, Prince *et al.* 2012, Seibert *et al.* 1995), HIV-1 cleavage sites had more residues conserved P4-P4' versus |P12-P5| from SIV-cpz and this drove selection of an HIV-1 PR that had stronger active-site interactions with substrate residues P4-P4'. At the same time, HIV-1 PR S-groove interactions with substrate residues |P12-P5| were weakened in order to: 1) balance HIV-1 PR affinity for substrates to maintain the overall Gag/Pol polyprotein cleavage order required for virion maturation, Table 1; and 2) prevent premature activation of HIV-1 PR during formation of the immature virion (Strickler *et al.* 1989). One consequence was that HIV-1 PR became sensitive to active-site inhibitors. In contrast, during HIV-2 adaptation to the human host the Gag/Pol cleavage sites had a more balanced and higher conservation of SIV-sm cleavage site residues |P12-P5|

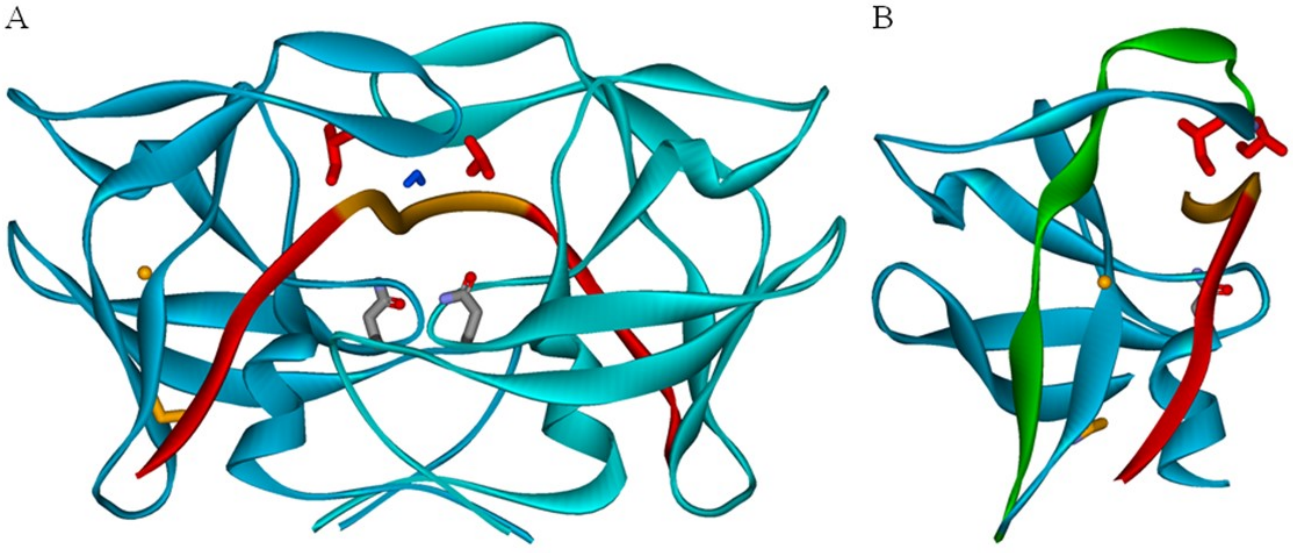

Figure 10. HTLV-1 PR with bound MA/CA 24-mer substrate. A) Front view with flaps on top, A and B subunits as backbone ribbons, active-site mutation D32N center (A and B subunits). Atom colors: oxygen, red; nitrogen, blue; carbon, grey. Tetrahedrally coordinated water with all atoms blue (center, above substrate). Back mutated residues shown starting clockwise bottom left: S-groove I85A and T88G carbons in bronze; active-site V56I and A59I carbons in red. Gag MA/CA substrate backbone |P12-P5| in red; P4-P4' in bronze. B) Side view of A (left subunit) showing the anti-parallel beta sheet with one ribbon colored green starting at the tip of the flap and ending at the bottom of the S-groove, with the other strand of the anti-parallel beta sheet below as a blue ribbon.

Table 4. HTLV-1 PRs interaction energy scores with MA/CA 24-mer and indinavir.

	HTLV-1 PR	HTLV-1 2X PR (S-groove)	HTLV-1 4X PR (S-groove/AC)	
MA/CA 24-mer	-235.9 (128.6/-107.2)	-231.0 (-123.7/-107.3)	-235.6 (-126.4/-109.2)	
Indinavir	-84.1	-84.8	-86.7	

HTLV-1 PR and mutant HTLV-1 PRs interaction energy scores (kcal/mol) with MA/CA 24-mer substrate residues: P12-P12'; |P12-P5|; P4-P4'. Bottom, scores for the same PRs and the active-site inhibitor indinavir.

Figure 9, while HIV-2 PR also acquired the secondary resistance residue Ala73, Figure 8. That mutation could help stabilize HIV-2 PR S-groove interactions with cleavage site residues |P12-P5| and along with the inherited S-groove secondary resistance residue Val71 may have contributed to HIV-2 PR weighting interactions with Gag/Pol cleavage sites more towards the S-grooves and cleavage site residues |P12-P5|. This in turn would allow HIV-2 Gag/Pol substrates to out-compete the binding of active-site inhibitors to HIV-2 PR. In order to test this hypothesis, the interaction energy scores between HIV-2, SIV-cpz, and SIV-sm PRs and the respective MA/CA 24-mer substrates were calculated for |P12-P5| and P4-P4'. WT HIV-1 PR interaction with the MA/CA substrate was weighted towards the active-site and residues P4-P4', while the HIV-2, SIV-cpz, and SIV-sm PRs interactions with the respective MA/CA substrates were all weighted towards the S-grooves and MA/CA substrate residues |P12-P5|, Table 3. The weighting of the HIV-2, SIV-cpz, and SIV-sm PRs interactions towards the S-grooves and substrate residues |P12-P5| correlated with resistance to active-site inhibitors. These findings support the hypothesis that PR/substrate interactions dominated by S-groove binding to cleavage site residues |P12-P5| allows Gag/Pol polyprotein cleavage sites to out-compete the binding of active-site inhibitors to PR (Laco 2015).

In silico analysis of HTLV-1 PR native multi-drug resistance

In contrast to the native primary resistance of HIV-2 and SIV-sm PRs (Results), the HTLV-1 PR has native multi-drug resistance to HIV-1 PR clinical inhibitors (Louis *et al.* 1999, Pettit *et al.* 1998), and contains a total of ten native primary/secondary resistance mutations (Laco 2015, Li *et al.* 2005). HTLV-1 PR has been reported to have an S-groove residue that *in silico* made direct H-bond interactions with substrate residues |P12-P5| (Laco 2015). That HTLV-1 PR S-groove residue (Thr88) aligned in 3D with the WT HIV-1 PR Gly73 that when mutated contributed to HIV-1 PR inhibitor resistance (Laco 2015). Here we tested whether HTLV-1 PR could be reverse engineered *in silico*, in a way opposite to how HIV-1 PR evolved into a MDR HIV-1 PR, to make the HTLV-1 PR inhibitor sensitive while maintaining interactions with substrates. HTLV-1 PR S-groove residues Ile85 and Thr88, equivalent to the HIV-1 PR secondary resistance mutations Ala70Ile and Gly73Thr (Rhee *et al.* 2003, Wensing *et al.* 2015), contributed to interactions with substrate residues |P12-P5|, Table 4 (Laco 2015). Here those S-groove residues were back mutated to equivalent residues found in WT HIV-1 PR to give the HTLV-1 2X PR (Ile85Ala and Thr88Gly), Figure 10.

The interaction between HTLV-1 2X PR and the MA/CA 24-mer substrate decreased by 4.9 kcal/mol consistent with a loss of S-groove contacts to substrate residues |P12-P5|, Table 4. Next, the HTLV-1 PR active-site residues Val56 and Ala59, equivalent to HIV-1 PR Ile47Val and Ile50Ala primary resistance mutations, were back mutated to Val56Ile and Ala59Ile and along with the S-groove mutations Ile85Ala and Thr88Gly resulted in the HTLV-1 4X PR, Figure 10. The HTLV-1 4X PR had a 2.6 kcal/mol stronger interaction with indinavir while restoring interactions with the MA/CA 24-mer substrate to a level similar to that for the native HTLV-1 PR, Table 4. Interestingly, when the interaction energy scores were broken down for the HTLV-1 4X PR and the MA/CA 24-mer substrate, the interactions were still found to be weighted towards the S-grooves and |P12-P5|, Table 4. Since the HTLV-1 4X PR S-groove interactions with substrate residues |P12-P5| were still the dominant interaction with the MA-CA substrate the potential exists *in vitro* for the MA/CA cleavage site, in the context of the Gag polyprotein, to out-compete active-site inhibitor binding to the HTLV-1 4X PR. These results indicate the potential complexity of reverse engineering the HTLV-1 PR into an inhibitor sensitive PR. It is interesting to note that by acquiring the bulky active-site mutations Val56Ile/Ala59Ile, the HTLV-1 4X PR flaps had to move away from the substrate residues P4-P4' in order to accommodate them. That flap movement was then translated to the S-grooves via the anti-parallel beta-sheets that connect the flaps and S-grooves and increased interactions with substrate residues |P12-P5|, Figure 10 B and 11. The HIV-1 PR and all other retroviral PRs studied here have similar anti-parallel beta-sheets connecting the flaps and S-grooves, Figure 3 B.

Non-primate retroviral PRs with native multi-drug resistance

In order to extend the native resistance findings for the HIV-2 and HTLV-1 PRs, non-primate retroviruses were next examined. EIAV and FIV have both been reported to encode native MDR PRs (Kervinen *et al.* 1998), with structures published for both (Kervinen *et al.* 1998, Laco *et al.* 1997). Structure based alignment of EIAV and FIV PRs with HIV-1 PR revealed that they had 28% and 24% identity with HIV-1 PR, re-

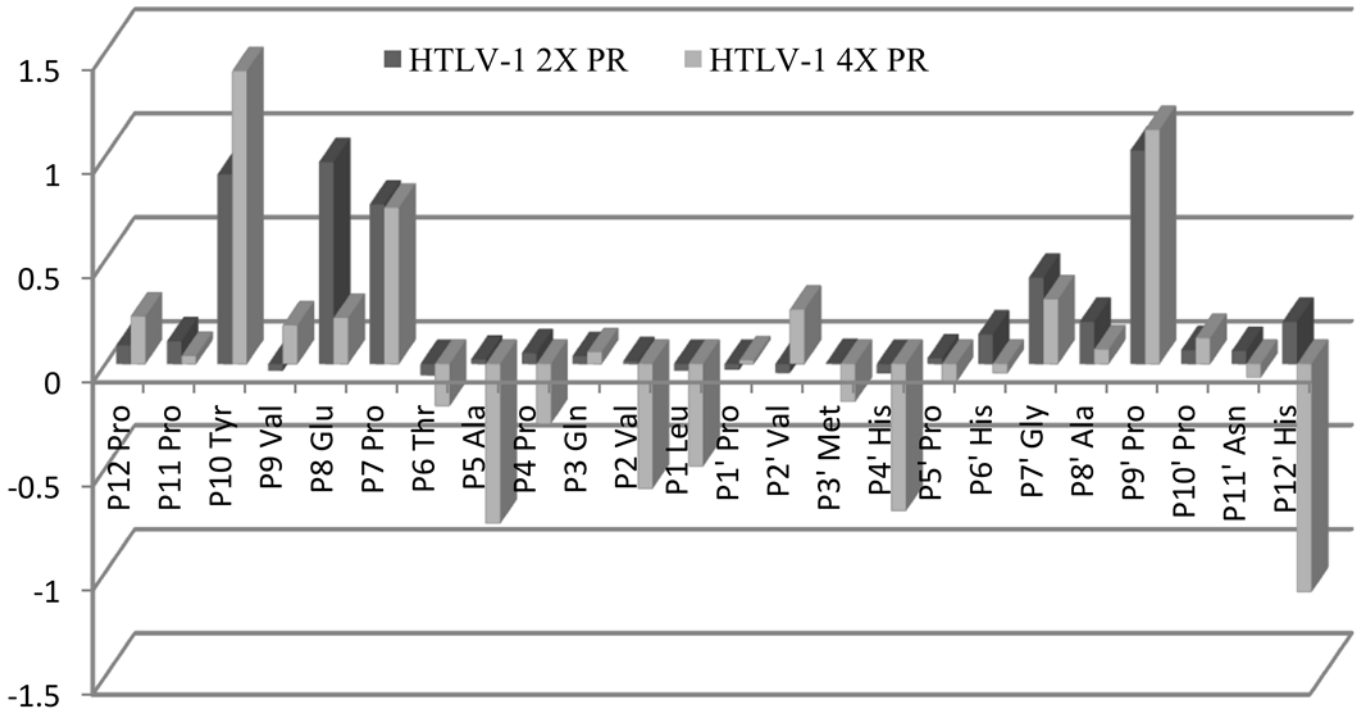

Figure 11. Reverse Engineered HTLV-1 PRs interaction with MA/CA 24-mer on a per residue basis. The reverse engineered HTLV-1 2X PR and 4X PR interaction energy scores, on a per residue basis, with the MA/CA substrate residues P12-P12' (kcal/mol). Scores relative to the native HTLV-1 PR residue interaction energy scores that were set to zero.

spectively, Figure 12. When both PRs were analyzed for native resistance residues, EIAV and FIV PRs were found to each have six residues that when found at the same 3D position in HIV-1 PR contributed to primary and secondary inhibitor resistance, and included S-groove residues, Figure 12 (Rhee *et al.* 2003, Wensing *et al.* 2015).

In the structure of an unliganded MDR HIV-1 PR the flaps were in an open conformation (Yedidi *et al.* 2014), as was the case for two MDR HIV-1 PRs with bound inhibitor (Liu *et al.* 2013a, b). In contrast, the structure of an unliganded WT HIV-1 PR had the flaps in a closed conformation (Wlodawer *et al.* 1989). These published structures demonstrated that the orientation of the flaps correlated with the sensitivity of the respective PRs to active-site inhibitors. Here inhibitor-resistant PRs were shown to have weaker interactions between the flaps/active-site and substrate residues P4-P4', which was compensated for by stronger interactions between the S-grooves and substrate residues |P12-P5|, including the MDR HIV-1, HIV-2, SIV-sm, HTLV-1, EIAV, and FIV PRs (Table 3).

In addition, like HTLV-1 PR, EIAV and FIV PRs had a number of residues that were found at resistance positions in HIV-1 PR, though had not been reported as resistance residues for HIV-1 PR (Rhee *et al.* 2003, Wensing *et al.* 2015). For example, HIV-1 PR Asp30 H-bonds to nelfinavir stabilizing the interaction between WT HIV-1 PR and nelfinavir, while the Asp30Asn mutation results in the loss of that H-bond making the mutant HIV-1 PR resistant to nelfinavir (Kolli *et al.* 2014). The EIAV PR Thr30 aligns with HIV-1 PR Asp30; it is likely that the shorter Thr30 side chain also does not H-bond to nelfinavir and so explains EIAV PR resistance to nelfinavir, Figure 12 (Kervinen *et al.* 1998). Likewise, the HIV-1 PR Gly48 when mutated to Val contributed to saquinavir resistance due to the loss of a stabilizing H-bond between the HIV-1 PR Gly48Val backbone carbonyl oxygen and saquinavir (Liu *et al.* 2008). Interestingly, the FIV PR Ile57 was found at the equivalent 3D position to HIV-1 PR Gly48 and so may play a similar role as the HIV-1 PR Gly48Val mutation in FIV PR resistance to saquinavir (i.e., RO31-8959), Figure 12 (Lin *et al.* 2000). Next the interaction energy scores were calculated for EIAV PR and FIV PR bound to the respective MA/CA 24-mer substrates. The EIAV and FIV PRs interaction with the respective MA/CA substrates were also weighted towards the S-grooves and substrate residues |P12-P5|, versus the active-site and residues P4-P4' (Table 3). Given that EIAV and FIV have been estimated to be in the respective hosts for millions of years (Cook *et al.* 2013, Pecon-Slattery *et al.* 2008), these findings suggest that there could be an evolutionary trend for retroviral PRs to weight substrate interactions towards the S-grooves and substrate residues |P12-P5| with one consequence being native resistance to active-site inhibitors.

Discussion

We have shown that the WT HIV-1 PR and MDR HIV-1 PR *in silico* interactions with solvent exposed Gag/Pol 24-mer cleavage sites correlated better with the

reported *in vitro* cleavage order than did the respective 8-mers, Tables 1 and 2. This demonstrated the importance of the HIV-1 PRs S-groove interactions with substrate residues |P12-P5| in determining the Gag/Pol cleavage order. At the same time, WT HIV-1 PR weighted the interaction with the 24-mer substrates towards the active-site and residues P4-P4' for four cleavage sites and the S-groove and |P12-P5| for four cleavage sites, while the the PR/RT site was equally weighted between the S-grooves and active-site (Table 1). The Gag/Pol cleavage sites in which the WT HIV-1 PR interactions with substrates were weighted towards the active-site/P4-P4' (MA/CA, CA/SP1, p6*/PR, RT/RH) may be responsible for the inhibitor sensitivity of the WT HIV-1 PR, due to active-site inhibitors being able to outcompete those substrates for binding to WT HIV-1 PR (Table 1). In contrast, the MDR HIV-1 PR interactions with 24-mer substrates were weighted towards the S-grooves and residues |P12-P5| for eight out of the nine Gag/Pol cleavage sites, which may be responsible for allowing the Gag/Pol substrates to outcompete inhibitor binding to the MDR HIV-1 PR active-site (Table 2), and correlates with the high level resistance of the MDR HIV-1 PR to active-site inhibitors (Wang *et al.* 2012). The only 24-mer cleavage site where the MDR HIV-1 PR interactions were weighted towards P4-P4' was the Gag CA/SP1 site, the slowest cleaved site *in vitro*, Table 2. Inhibition of cleavage at the CA/SP1 site would still allow virion maturation to proceed under optimal inhibitor levels and in the process release more active PR shifting the inhibitor/PR ratio towards PR. These results are supported by a published *in vitro* MDR HIV-1 PR cleavage assay that used a peptide substrate, that does not bind to the PR S-grooves, and resulted in a relative 8.4-fold weaker *in vitro* activity for the MDR HIV-1 PR as compared to cleavage at the SP1/NC site in full-length Gag (Laco 2015). This suggests that MDR HIV-1 PR inhibitor resistance may be significantly underestimated *in vitro* when short peptide substrates are used due to the weak binding of the MDR HIV-1 PR to substrate residues P4-P4' (Table 2). In contrast, the published *in vitro* results for the WT HIV-1 PR using the same peptide substrate and full-length Gag demonstrated that the WT HIV-1 PR had only a relative 2-fold weaker *in vitro* activity for the peptide versus Gag because the WT HIV-1 PR was more focused on flap/active-site interactions with substrate residues P4-P4', Table 1 (Laco 2015). This suggests that the sensitivity of WT HIV-1 PR to active-site inhibitors may be modestly overestimated *in vitro* when short peptide substrates are used.

Next we examined whether the inhibitor resistance residues selected for in the MDR HIV-1 PR, including primary active-site mutations and secondary S-groove mutations, were found in closely related retroviruses that had not been exposed to protease inhibitors (Results). This was done in order to determine if the MDR HIV-1 PR resistance evolution was either a novel pathway used solely for inhibitor resistance, or a pathway common to all retroviral proteases in order to adapt to mutated Gag/Pol substrates selected to escape *in vivo* CTL detection (Phillips *et al.* 1991, Prince *et al.* 2012, Seibert *et al.* 1995). When the HIV-1 and HIV-2 PRs were aligned with the evolutionary closest SIV PRs, HIV-1 PR was found to have lost the one native resistance residue found in SIV-cpz PR, while HIV-2 PR gained one native resistance residue in addi-

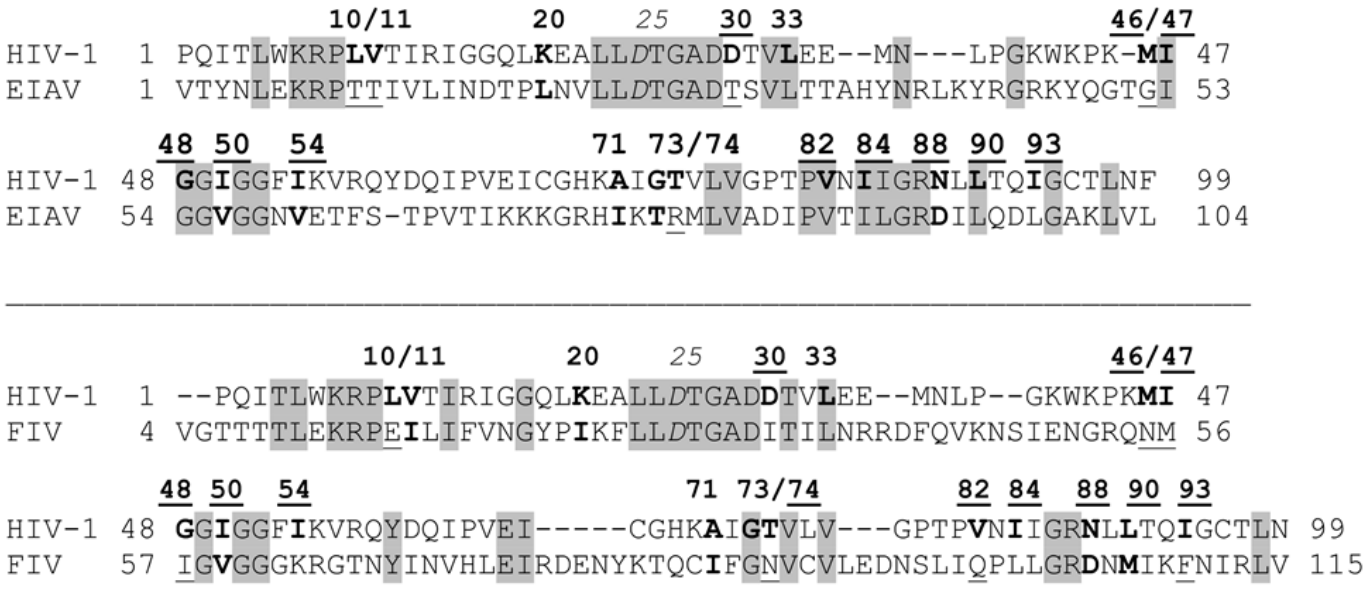

Figure 12. Structure-based alignment of HIV-1 PR with EIAV and FIV PRs. Gray highlighted residues indicate identity between aligned sequences. HIV-1 PR residues that when mutated contributed to inhibitor resistance in bold, positions of primary resistance residue indicated by underlined numbers, EIAV and FIV PRs native resistance residues in bold, underlined EIAV and FIV residues indicate potential native resistance residues. HIV-1 PR active-site Asp25 (D25) italicized, as are the active-site residues for EIAV (D25) and FIV (D30).

tion to inheriting the six native resistance residues present in SIV-sm PR, Figure 8. Since neither HIV-1, nor HIV-2, evolved from the respective SIV in the presence of protease inhibitors, the evolution of the corresponding Gag/Pol substrates were examined to see how they could explain the resistance residue inheritance pattern of HIV-1 and HIV-2 PRs, Figure 9. The HIV-1 Gag/Pol cleavage sites had a significant loss in identity with the respective SIV-cpz Gag/Pol cleavage site residues |P12-P5|, and this may have contributed to the WT HIV-1 PR weighting interactions towards the flaps/active-site and the conserved cleavage site residues P4-P4', Figure 9 and Table 1. The HIV-1 PR focus on substrate residues P4-P4' correlated with HIV-1 PR losing the SIV-cpz PR Val10 native resistance residue since Val10 is also the P10' residue in the p6*/PR cleavage site and would be bound in the S-grooves, Figure 3 and 5. In contrast, the HIV-2 Gag/Pol cleavage sites overall retained significantly more identity with the SIV-sm cleavage sites, and non-conserved residues in the HIV-2 Gag/Pol cleavage sites were more evenly distributed across cleavage site residues P12-P12', Figure 9. The net result was the evolution of an HIV-2 PR that weighted S-groove interactions towards substrate residues |P12-P5| due in part to the acquisition of the native S-groove secondary resistance residue Ala73. And so during adaptation to the human host, the HIV-1 PR focused more on substrate residues P4-P4' and became acutely sensitive to active-site inhibitors (Rhee *et al.* 2003, Wensing *et al.* 2015). In contrast, the HIV-2 PR weighted S-groove interactions towards substrate residues |P12-P5| as shown for the MA/CA site (Table 3) resulting in native resistance to many HIV-1 PR active-site inhibitors, Results. These findings support the hypothesis that HIV-1 PR used a common retroviral PR pathway to acquire multi-drug resistance. As a result, HIV-1 PR active-site inhibitors can drive the evolution of HIV-1 PR resistance as well as select for adaptive mutations in the Gag/Pol cleavage sites (McKinnon *et al.* 2011). Transmission of the resulting MDR HIV-1 eliminates PR active-site inhibitors as a treatment option. By targeting the HIV-1 PR S-grooves, along with the active-site, the potential loss of PR active-site inhibitors could be avoided. This approach may be critical given that in 2012 of the estimated 1.5 million HIV-1 infected individuals in the United States and Puerto Rico only 30% had the virus suppressed to < 200 copies/mL (Frieden *et al.* 2015). Likewise, the HIV-2, HTLV-1, EIAV, and FIV PRs, which all contain S-groove resistance residues, could be targeted with both S-groove inhibitors and active-site inhibitors in order to prevent the evolution of PRs that re-weighted substrate interactions towards P4-P4' in order to escape S-groove inhibitors.

The HIV-1 Gag and Pol polyprotein 3D models revealed the solvent accessibility of the Gag cleavage site residues versus several solvent inaccessible Pol cleavage sites (i.e., RT/RH and RH/IN), Figure 6 and 7. In the cell, soluble proteases degrade cytosolic proteins at low levels to release approximately ten residue long peptides that are then bound by MHC class I receptors for presentation on the cell surface for CTL surveillance (Kourjian *et al.* 2014, Lazaro *et al.* 2015, Yewdell *et al.* 1999, Zervoudi *et al.* 2013). Due to the solvent accessibility of the Gag cleavage sites, they would be prime targets for those cellular proteases, in contrast to the less accessible Pol cleavage sites, Figure 6 and 7. The net result being that during SIV-cpz cross-species transmission into humans CTL selective pressure likely contributed to the mutation of Gag cleavage sites, while Pol cleavage site sequences were highly conserved, Figure 9. The accessibility of the HIV-1 Gag cleavage sites could be a strategy to direct the cellular immune response towards Gag to minimize mutation of the more structurally sensitive Pol enzymes that typically have ordered N-and C-termini, Figure 7. The remarkable ability of HIV-1 to evade the CTL response may be due in part to the Gag cleavage sites being used as bait for cellular proteases since HIV -1 PR recognizes 24-residues of a cleavage site, with the ability to differentially weight substrate interactions between |P12-P5| and P4-P4' in order to accommodate Gag/Pol cleavage site CTL escape mutations. This substrate recognition strategy may allow HIV-1 PR and other retroviral PRs to minimize the impact of cleavage site CTL escape mutations on PR/substrate interactions and consequently virus maturation (Phillips *et al.* 1991, Prince *et al.* 2012, Seibert *et al.* 1995).

A complicating factor in the proposed HIV-1 PR adaptation to mutations in Gag cleavage sites is the timeline; the CTL response selects for escape mutations in Gag cleavage sites during a round of replication, while the selection of adaptive mutations in HIV-1 PR for the mutated cleavage sites would take place in subsequent rounds of replication. Perhaps HIV-1 PR uses a strategy similar to that shown *in silico* for the HTLV-1 4X PR, which adapted to bulky cleavage site residues P4-P4' by moving the flaps away, with a concomitant strengthening of the S-groove contacts with residues |P12-P5| via the anti-parallel beta sheets, Tables 4 and 5 and Figure 10. This approach to substrates with mutated residues could allow the HIV-1 PR to get by CTL escape mutations in Gag cleavage sites during the same replication cycle. Then in subsequent rounds of replication adaptive mutations in HIV-1 PR could be selected for. It is interesting to note that HIV-1 PR secondary resistance polymorphisms were found in treatment naive patients, including the S-groove A71V/T and G73C/R mutations (Birk & Sonnerborg 1998, Bossi *et al.* 1999, Kearney *et al.* 2008, Kozal *et al.* 1996, Rose *et al.* 1996b). The HIV-1 PR secondary resistance polymorphisms could represent an archive of adaptive HIV-1 PR mutations selected in response to CTL escape mutations in Gag cleavage sites over the course of an infection (Birk & Sonnerborg 1998, Bossi *et al.* 1999, Kearney *et al.* 2008, Kozal *et al.* 1996, Servais *et al.* 2001). Those PR S-groove muta-

tions could then be selected for during subsequent inhibitor treatment and contribute to PR resistance. In untreated chronically infected patients >98% of the latent HIV-1 reservoir contained CTL escape mutations (Deng *et al.* 2015).

The HIV-1 PR may respond to both mutations in substrate residues P4-P4' and active-site inhibitors by reweighting substrate interactions towards the S-grooves, since active-site inhibitors essentially mimic mutated substrate from the perspective of the PR in that both disrupt the processing of the Gag/Pol polyproteins. At the same time, the S-grooves may allow for non-covalent tethering of the HIV-1 PR to the solvent exposed Gag/Pol cleavage sites, even with inhibitor bound in the active-site (Figure 1), and allow for virion maturation at a rate concordant with the inhibitor dissociation rate. Together these findings are directed towards a strategy to combat HIV-1 infections for which current treatments are not effective (Frieden *et al.* 2015), and HTLV-1 infections for which there are no treatments.

Conclusions

Based on the finding presented here and elsewhere (Deshmukh *et al.* 2017, Laco 2015), retroviral PRs interact with the Gag/Pol cleavage sites using both the active-site and S-grooves. The PRs can weight those interactions towards either the active-site and cleavage site residues P4-P4, or S-grooves and residues |P5-P12|, to accommodate cleavage site CTL escape mutations while maintaining Gag/Pol cleavage order. Many retroviral PRs evolved S-groove dominated interactions with substrates, and as a consequence native resistance to active site inhibitors. In contrast, HIV-1 PR evolved a more active-site dominated interaction with substrates that resulted in sensitivity to active-site inhibitors, while retaining the ability to reweight interactions towards the S-grooves and substrate residues |P5-P12| to outcompete active-site inhibitors.

Conflict of interest

The author declares no conflict of interest.

Acknowledgements

This work was supported by the Roskamp Foundation and the Laco Science Institute.

References

Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vezinet-Brun F, Rouzioux C, Rozenbaum W & Montagnier L 1983 Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **220** 868-871

Bharat TA, Castillo Menendez LR, Hagen WJ, Lux V, Igonet S, Schorb M, Schur FK, Krausslich HG & Briggs JA 2014 Cryo-electron microscopy of tubular arrays of HIV-1 Gag resolves structures essential for immature virus assembly. *Proc Natl Acad Sci U S A* **111** 8233-8238

Billich S, Knoop MT, Hansen J, Strop P, Sedlacek J, Mertz R & Moelling K 1988 Synthetic peptides as substrates and inhibitors of human immune deficiency virus-1 protease. *J Biol Chem* **263** 17905-17908

Birk M & Sonnerborg A 1998 Variations in HIV-1 pol gene associated with reduced sensitivity to antiretroviral drugs in treatment-naive patients. *Aids* **12** 2369-2375

Bossi P, Mouroux M, Yvon A, Bricaire F, Agut H, Huraux JM, Katlama C & Calvez V 1999 Polymorphism of the human immunodeficiency virus type 1 (HIV-1) protease gene and response of HIV-1-infected patients to a protease inhibitor. *J Clin Microbiol* **37** 2910-2912

Brower ET, Bacha U M, Kawasaki Y & Freire E 2008 Inhibition of HIV-2 protease by HIV-1 protease inhibitors in clinical use. *Chem Biol Drug Des* **71** 298-305

Chang MW & Torbett BE 2011 Accessory mutations maintain stability in drug-resistant HIV-1 protease. *J Mol Biol* **410** 756-760

Chen Z, Li Y, Schock HB, Hall D, Chen E & Kuo LC 1995 Three-dimensional structure of a mutant HIV-1 protease displaying cross-resistance to all protease inhibitors in clinical trials. *J Biol Chem* **270** 21433-21436

Consortium TU 2015 UniProt: a hub for protein information. *Nucleic Acids Res* **43** D204-212

Cook RF, Leroux C & Issel CJ 2013 Equine infectious anemia and equine infectious anemia virus in 2013: a review. *Vet Microbiol* **167** 181-204

D'Arc M, Ayouba A, Esteban A, Learn GH, Boue V, Liegeois F, Etienne L, Tagg N, Leendertz F H, Boesch C, Madinda NF, Robbins MM, Gray M, Cournil A, Ooms M, Letko M, Simon VA, Sharp PM, Hahn BH, Delaporte E, Mpoudi Ngole E & Peeters M 2015 Origin of the HIV-1 group O epidemic in western lowland gorillas. *Proc Natl Acad Sci U S A* **112** E1343-1352

Darke PL, Nutt RF, Brady SF, Garsky VM, Ciccarone TM, Leu CT, Lumma PK, Freidinger RM, Veber DF & Sigal IS 1988 HIV-1 protease specificity of peptide cleavage is sufficient for processing of gag and pol polyproteins. *Biochem Biophys Res Commun* **156** 297-303

Deng K, Pertea M, Rongvaux A, Wang L, Durand C M, Ghiaur G, Lai J, McHugh HL, Hao H, Zhang H, Margolick JB, Gurer C, Murphy A J, Valenzuela DM, Yancopoulos GD, Deeks SG, Strowig T, Kumar P, Siliciano JD, Salzberg SL, Flavell RA, Shan L & Siliciano RF 2015 Broad CTL response is required to clear latent HIV-1 due to dominance of escape mutations. *Nature* **517** 381-385

Desbois D, Roquebert B, Peytavin G, Damond F, Collin G, Benard A, Campa P, Matheron S, Chene G,

Brun-Vezinet F & Descamps D 2008 In vitro phenotypic susceptibility of human immunodeficiency virus type 2 clinical isolates to protease inhibitors. *Antimicrob Agents Chemother* **52** 1545-1548

Deshmukh L, Tugarinov V, Louis JM & Clore GM 2017. Binding kinetics and substrate selectivity in HIV -1 protease-Gag interactions probed at atomic resolution by chemical exchange NMR. *Proc Natl Acad Sci U S A* **114** E9855-E9862

Ding YS, Rich DH & Ikeda RA 1998 Substrates and inhibitors of human T-cell leukemia virus type I protease. *Biochemistry* **37** 17514-17518

Edgar RC 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32** 1792-1797

Erickson-Viitanen S, Manfredi J, Viitanen P, Tribe D E, Tritch R, Hutchison CA, 3rd, Loeb DD & Swanstrom R 1989 Cleavage of HIV-1 gag polyprotein synthesized in vitro: sequential cleavage by the viral protease. *AIDS Res Hum Retroviruses* **5** 577-591

Erickson J, Neidhart DJ, VanDrie J, Kempf DJ, Wang XC, Norbeck DW, Plattner JJ, Rittenhouse JW, Turon M, Wideburg N & et al. 1990 Design, activity, and 2.8 A crystal structure of a C2 symmetric inhibitor complexed to HIV-1 protease. *Science* **249** 527-533

Fodor SK & Vogt VM 2002 Characterization of the protease of a fish retrovirus, walleye dermal sarcoma virus. *J Virol* **76** 4341-4349

Fossen T, Wray V, Bruns K, Rachmat J, Henklein P, Tessmer U, Maczurek A, Klinger P & Schubert U 2005 Solution structure of the human immunodeficiency virus type 1 p6 protein. *J Biol Chem* **280** 42515-42527

Frieden TR, Foti KE & Mermin J 2015 Applying Public Health Principles to the HIV Epidemic--How Are We Doing? *N Engl J Med* **373** 2281-2287

Gallo RC, Sarin PS, Gelmann EP, Robert-Guroff M, Richardson E, Kalyanaraman VS, Mann D, Sidhu GD, Stahl RE, Zolla-Pazner S, Leibowitch J & Popovic M 1983 Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science* **220** 865-867

Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, Cummins LB, Arthur LO, Peeters M, Shaw GM, Sharp PM & Hahn BH 1999 Origin of HIV -1 in the chimpanzee Pan troglodytes troglodytes. *Nature* **397** 436-441

Gomez R, Jolly SJ, Williams T, Vacca JP, Torrent M, McGaughey G, Lai MT, Felock P, Munshi V, Distefano D, Flynn J, Miller M, Yan Y, Reid J, Sanchez R, Liang Y, Paton B, Wan BL & Anthony N 2011 Design and synthesis of conformationally constrained inhibitors of non-nucleoside reverse transcriptase. *J Med Chem* **54** 7920-7933

Goncalves DU, Proietti FA, Ribas JG, Araujo MG, Pinheiro SR, Guedes AC & Carneiro-Proietti AB 2010 Epidemiology, treatment, and prevention of human T-cell leukemia virus type 1-associated diseases. *Clin Microbiol Rev* **23** 577-589

Gres AT, Kirby KA, KewalRamani VN, Tanner JJ, Pornillos O & Sarafianos SG 2015 X-ray crystal structures of native HIV-1 capsid protein reveal conformational variability. *Science* **349** 99-103

Gulnik SV, Suvorov LI, Liu B, Yu B, Anderson B, Mitsuya H & Erickson JW 1995 Kinetic characterization and cross-resistance patterns of HIV-1 protease mutants selected under drug pressure. *Biochemistry* **34** 9282-9287

Gustchina A, Kervinen J, Powell DJ, Zdanov A, Kay J & Wlodawer A 1996 Structure of equine infectious anemia virus proteinase complexed with an inhibitor. *Protein Sci* **5** 1453-1465

Hirsch VM, Olmsted RA, Murphey-Corb M, Purcell RH & Johnson PR 1989 An African primate lentivirus (SIVsm) closely related to HIV-2. *Nature* **339** 389-392

Huet T, Cheynier R, Meyerhans A, Roelants G & Wain-Hobson S 1990 Genetic organization of a chimpanzee lentivirus related to HIV-1. *Nature* **345** 356-359

Ishima R, Torchia DA, Lynch SM, Gronenborn AM & Louis JM 2003 Solution structure of the mature HIV-1 protease monomer: insight into the tertiary fold and stability of a precursor. *J Biol Chem* **278** 43311-43319

Jacks T, Power MD, Masiarz FR, Luciw PA, Barr PJ & Varmus HE 1988 Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature* **331** 280-283

Kaplan AH, Michael SF, Wehbie RS, Knigge MF, Paul DA, Everitt L, Kempf DJ, Norbeck DW, Erickson JW & Swanstrom R 1994 Selection of multiple human immunodeficiency virus type 1 variants that encode viral proteases with decreased sensitivity to an inhibitor of the viral protease. *Proc Natl Acad Sci U S A* **91** 5597-5601

Kearney M, Palmer S, Maldarelli F, Shao W, Polis MA, Mican J, Rock-Kress D, Margolick JB, Coffin JM & Mellors JW 2008 Frequent polymorphism at drug resistance sites in HIV-1 protease and reverse transcriptase. *Aids* **22** 497-501

Kervinen J, Lubkowski J, Zdanov A, Bhatt D, Dunn BM, Hui KY, Powell DJ, Kay J, Wlodawer A & Gustchina A 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

King NM, Prabu-Jeyabalan M, Bandaranayake RM, Nalam MN, Nalivaika EA, Ozen A, Haliloglu T, Yilmaz NK & Schiffer CA 2012 Extreme entropy-enthalpy compensation in a drug-resistant variant of HIV-1 protease. *ACS Chem Biol* **7** 1536-1546

Kohl NE, Emini EA, Schleif WA, Davis LJ, Heimbach JC, Dixon RA, Scolnick EM & Sigal IS 1988 Active human immunodeficiency virus protease is required for viral infectivity. *Proc Natl Acad Sci U.S.A.* **85** 4686-4690

Kolli M, Ozen A, Kurt-Yilmaz N & Schiffer CA 2014 HIV-1 protease-substrate coevolution in nelfinavir resistance. *J Virol* **88** 7145-7154

Koralnik IJ, Boeri E, Saxinger WC, Monico AL, Fullen J, Gessain A, Guo HG, Gallo RC, Markham P, Kalyanaraman V & et al. 1994 Phylogenetic associations of human and simian T-cell leukemia/ lymphotropic virus type I strains: evidence for interspecies transmission. *J Virol* **68** 2693-2707

Kotler M, Katz RA, Danho W, Leis J & Skalka AM 1988 Synthetic peptides as substrates and inhibitors of a retroviral protease. *Proc Natl Acad Sci U S A* **85** 4185-4189

Kourjian G, Xu Y, Mondesire-Crump I, Shimada M, Gourdain P & Le Gall S 2014 Sequence-specific alterations of epitope production by HIV protease inhibitors. *J Immunol* **192** 3496-3506

Kovalevsky AY, Louis JM, Aniana A, Ghosh AK & Weber IT 2008 Structural evidence for effectiveness of darunavir and two related antiviral inhibitors against HIV-2 protease. *J Mol Biol* **384** 178-192

Kozal MJ, Shah N, Shen N, Yang R, Fucini R, Merigan TC, Richman DD, Morris D, Hubbell E, Chee M & Gingeras TR 1996 Extensive polymorphisms observed in HIV-1 clade B protease gene using high-density oligonucleotide arrays. *Nat Med* **2** 753-759

Kozisek M, Bray J, Rezacova P, Saskova K, Brynda J, Pokorna J, Mammano F, Rulisek L & Konvalinka J 2007 Molecular analysis of the HIV-1 resistance development: enzymatic activities, crystal structures, and thermodynamics of nelfinavir-resistant HIV protease mutants. *J Mol Biol* **374** 1005-1016

Krausslich HG, Ingraham RH, Skoog MT, Wimmer E, Pallai PV & Carter CA 1989 Activity of purified biosynthetic proteinase of human immunodeficiency virus on natural substrates and synthetic peptides. *Proc Natl Acad Sci U.S.A.* **86** 807-811

Laco GS 2011 Evaluation of two models for human topoisomerase I interaction with dsDNA and camptothecin derivatives. *PLoS One* **6** e24314

Laco GS 2015 HIV-1 protease substrate-groove: Role in substrate recognition and inhibitor resistance. *Biochimie* **118** 90-103

Laco GS, Schalk-Hihi C, Lubkowski J, Morris G, Zdanov A, Olson A, Elder J H, Wlodawer A & Gustchina A 1997 Crystal structures of the inactive D30N mutant of feline immunodeficiency virus protease complexed with a substrate and an inhibitor. *Biochemistry* **36** 10696-10708

Lazaro S, Gamarra D & Del Val M 2015 Proteolytic enzymes involved in MHC class I antigen processing: A guerrilla army that partners with the proteasome. *Mol Immunol* **68** 72-76

Le Grice SF, Mills J & Mous J 1988 Active site mutagenesis of the AIDS virus protease and its alleviation by trans complementation. *EMBO J* **7** 2547-2553

Lee BM, De Guzman RN, Turner BG, Tjandra N & Summers MF 1998 Dynamical behavior of the HIV-1 nucleocapsid protein. *J Mol Biol* **279** 633-649

Leigh Brown AJ, Frost SD, Mathews WC, Dawson K, Hellmann NS, Daar ES, Richman DD & Little SJ 2003 Transmission fitness of drug-resistant human immunodeficiency virus and the prevalence of resistance in the antiretroviral-treated population. *J Infect Dis* **187** 683-686

Lemey P, Pybus OG, Wang B, Saksena NK, Salemi M & Vandamme AM 2003 Tracing the origin and history of the HIV-2 epidemic. *Proc Natl Acad Sci U S A* **100** 6588-6592

Li M, Laco GS, Jaskolski M, Rozycki J, Alexandratos J, Wlodawer A & Gustchina A 2005 Crystal structure of human T cell leukemia virus protease, a novel target for anticancer drug design. *Proc Natl Acad Sci U S A* **102** 18332-18337

Lin YC, Beck Z, Lee T, Le VD, Morris GM, Olson AJ, Wong CH & Elder JH 2000 Alteration of substrate and inhibitor specificity of feline immunodeficiency virus protease. *J Virol* **74** 4710-4720

Liu F, Kovalevsky AY, Tie Y, Ghosh AK, Harrison RW & Weber IT 2008 Effect of flap mutations on structure of HIV-1 protease and inhibition by saquinavir and darunavir. *J Mol Biol* **381** 102-115

Liu Z, Yedidi RS, Wang Y, Dewdney TG, Reiter SJ, Brunzelle JS, Kovari IA & Kovari LC 2013a Crystallographic study of multi-drug resistant HIV-1 protease lopinavir complex: mechanism of drug recognition and resistance. *Biochem Biophys Res Commun* **437** 199-204

Liu Z, Yedidi RS, Wang Y, Dewdney TG, Reiter SJ, Brunzelle JS, Kovari IA & Kovari LC 2013b Insights into the mechanism of drug resistance: X-ray structure analysis of multi-drug resistant HIV-1 protease ritonavir complex. *Biochem Biophys Res Commun* **431** 232-238

Louis JM, Oroszlan S & Tozser J 1999 Stabilization from autoproteolysis and kinetic characterization of the human T-cell leukemia virus type 1 proteinase. *J Biol Chem* **274** 6660-6666

Lv Z, Chu Y & Wang Y 2015 HIV protease inhibitors: a review of molecular selectivity and toxicity. *HIV AIDS (Auckl)* **7** 95-104

Martinez-Picado J, Savara AV, Sutton L & D'Aquila RT 1999 Replicative fitness of protease inhibitor-resistant mutants of human immunodeficiency virus type 1. *J Virol* **73** 3744-3752

Marx PA, Li Y, Lerche NW, Sutjipto S, Gettie A, Yee JA, Brotman BH, Prince AM, Hanson A, Webster RG & et al. 1991 Isolation of a simian immunodeficiency virus related to human immunodeficiency virus type 2 from a west African pet sooty mangabey. *J Virol* **65** 4480-4485

Masse S, Lu X, Dekhtyar T, Lu L, Koev G, Gao F, Mo H, Kempf D, Bernstein B, Hanna GJ & Molla A 2007 In vitro selection and characterization of human immunodeficiency virus type 2 with decreased susceptibility to lopinavir. *Antimicrob Agents Chemother* **51** 3075-3080

McKinnon JE, Delgado R, Pulido F, Shao W, Arribas JR & Mellors JW 2011 Single genome sequencing of HIV-1 gag and protease resistance mutations at virologic failure during the OK04 trial of simplified versus

standard maintenance therapy. *Antivir Ther* **16** 725-732
McWilliam H, Li W, Uludag M, Squizzato S, Park Y M, Buso N, Cowley AP & Lopez R 2015 Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res* **41** W597-600
Miller M, Schneider J, Sathyanarayana BK, Toth MV, Marshall GR, Clawson L, Selk L, Kent SB & Wlodawer A 1989 Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 A resolution. *Science* **246** 1149-1152
Mocroft A, Ruiz L, Reiss P, Ledergerber B, Katlama C, Lazzarin A, Goebel FD, Phillips AN, Clotet B & Lundgren JD 2003 Virological rebound after suppression on highly active antiretroviral therapy. *Aids* **17** 1741-1751
Nijhuis M, Schuurman R, de Jong D, Erickson J, Gustchina E, Albert J, Schipper P, Gulnik S & Boucher CA 1999 Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy. *Aids* **13** 2349-2359
Ovchinnikov S, Kim DE, Wang RY, Liu Y, DiMaio F & Baker D 2016 Improved de novo structure prediction in CASP11 by incorporating coevolution information into Rosetta. *Proteins* **84** 67-75
Pazhanisamy S, Stuver CM, Cullinan AB, Margolin N, Rao BG & Livingston D J 1996 Kinetic characterization of human immunodeficiency virus type-1 protease -resistant variants. *J Biol Chem* **271** 17979-17985
Pearl LH & Taylor WR 1987 Sequence specificity of retroviral proteases. *Nature* **328** 482
Pecon-Slattery J, Troyer J L, Johnson WE & O'Brien SJ 2008 Evolution of feline immunodeficiency virus in Felidae: implications for human health and wildlife ecology. *Vet Immunol Immunopathol* **123** 32-44
Peeters M, Honore C, Huet T, Bedjabaga L, Ossari S, Bussi P, Cooper RW & Delaporte E 1989 Isolation and partial characterization of an HIV-related virus occurring naturally in chimpanzees in Gabon. *Aids* **3** 625-630
Pei J, Kim BH & Grishin NV 2008 PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res* **36** 2295-2300
Pettit SC, Henderson GJ, Schiffer CA & Swanstrom R 2002 Replacement of the P1 amino acid of human immunodeficiency virus type 1 Gag processing sites can inhibit or enhance the rate of cleavage by the viral protease. *J Virol* **76** 10226-10233
Pettit SC, Lindquist JN, Kaplan AH & Swanstrom R 2005 Processing sites in the human immunodeficiency virus type 1 (HIV-1) Gag-Pro-Pol precursor are cleaved by the viral protease at different rates. *Retrovirology* **2** 66
Pettit SC, Sanchez R, Smith T, Wehbie R, Derse D & 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
Phillips RE, Rowland-Jones S, Nixon DF, Gotch FM, Edwards JP, Ogunlesi AO, Elvin JG, Rothbard JA, Bangham CR, Rizza CR & et al. 1991 Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. *Nature* **354** 453-459
Prabu-Jeyabalan M, Nalivaika E & Schiffer CA 2002 Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes. *Structure* **10** 369-381
Prabu-Jeyabalan M, Nalivaika EA, King NM & Schiffer CA 2004 Structural basis for coevolution of a human immunodeficiency virus type 1 nucleocapsid-p1 cleavage site with a V82A drug-resistant mutation in viral protease. *J Virol* **78** 12446-12454
Prince JL, Claiborne DT, Carlson JM, Schaefer M, Yu T, Lahki S, Prentice HA, Yue L, Vishwanathan SA, Kilembe W, Goepfert P, Price MA, Gilmour J, Mulenga J, Farmer P, Derdeyn CA, Tang J, Heckerman D, Kaslow RA, Allen SA & Hunter E 2012 Role of transmitted Gag CTL polymorphisms in defining replicative capacity and early HIV-1 pathogenesis. *PLoS Pathog* **8** e1003041
Proietti FA, Carneiro-Proietti AB, Catalan-Soares BC & Murphy EL 2005 Global epidemiology of HTLV-I infection and associated diseases. *Oncogene* **24** 6058-6068
Ratner L, Haseltine W, Patarca R, Livak KJ, Starcich B, Josephs SF, Doran ER, Rafalski JA, Whitehorn EA, Baumeister K & et al. 1985 Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature* **313** 277-284
Rhee SY, Gonzales MJ, Kantor R, Betts BJ, Ravela J & Shafer RW 2003 Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res* **31** 298-303
Richman DD, Morton SC, Wrin T, Hellmann N, Berry S, Shapiro MF & Bozzette SA 2004 The prevalence of antiretroviral drug resistance in the United States. *Aids* **18** 1393-1401
Rodes B, Sheldon J, Toro C, Jimenez V, Alvarez MA & Soriano V 2006 Susceptibility to protease inhibitors in HIV-2 primary isolates from patients failing antiretroviral therapy. *J Antimicrob Chemother* **57** 709-713
Rose RB, Craik CS, Douglas NL & Stroud RM 1996a Three-dimensional structures of HIV-1 and SIV protease product complexes. *Biochemistry* **35** 12933-12944
Rose RE, Gong YF, Greytok JA, Bechtold CM, Terry BJ, Robinson BS, Alam M, Colonno RJ & Lin PF 1996b Human immunodeficiency virus type 1 viral background plays a major role in development of resistance to protease inhibitors. *Proc Natl Acad Sci U S A* **93** 1648-1653
Rosenbloom DI, Hill AL, Rabi SA, Siliciano RF & Nowak MA 2012 Antiretroviral dynamics determines HIV evolution and predicts therapy outcome. *Nat Med* **18** 1378-1385
Santiago ML, Rodenburg CM, Kamenya S, Bibollet-Ruche F, Gao F, Bailes E, Meleth S, Soong SJ, Kilby JM, Moldoveanu Z, Fahey B, Muller MN, Ayouba A,

Nerrienet E, McClure HM, Heeney JL, Pusey AE, Collins DA, Boesch C, Wrangham RW, Goodall J, Sharp PM, Shaw GM & Hahn BH 2002 SIVcpz in wild chimpanzees. *Science* **295** 465

Schock HB, Garsky VM & Kuo LC 1996 Mutational anatomy of an HIV-1 protease variant conferring cross -resistance to protease inhibitors in clinical trials. Compensatory modulations of binding and activity. *J Biol Chem* **271** 31957-31963

Seibert SA, Howell CY, Hughes MK & Hughes AL 1995 Natural selection on the gag, pol, and env genes of human immunodeficiency virus 1 (HIV-1). *Mol Biol Evol* **12** 803-813

Servais J, Lambert C, Fontaine E, Plesseria JM, Robert I, Arendt V, Staub T, Schneider F, Hemmer R, Burtonboy G & Schmit JC 2001 Variant human immunodeficiency virus type 1 proteases and response to combination therapy including a protease inhibitor. *Antimicrob Agents Chemother* **45** 893-900

Strickler JE, Gorniak J, Dayton B, Meek T, Moore M, Magaard V, Malinowski J & Debouck C 1989 Characterization and autoprocessing of precursor and mature forms of human immunodeficiency virus type 1 (HIV 1) protease purified from Escherichia coli. *Proteins* **6** 139-154

Tang C, Ndassa Y & Summers MF 2002 Structure of the N-terminal 283-residue fragment of the immature HIV-1 Gag polyprotein. *Nat. Struct. Biol.* **9** 537-543

Tozser J, Blaha I, Copeland TD, Wondrak EM & Oroszlan S 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

Tritch RJ, Cheng YE, Yin FH & Erickson-Viitanen S 1991 Mutagenesis of protease cleavage sites in the human immunodeficiency virus type 1 gag polyprotein. *J Virol* **65** 922-930

Voevodin AF, Johnson BK, Samilchuk EI, Stone GA, Druilhet R, Greer WJ & Gibbs CJ, Jr. 1997 Phylogenetic analysis of simian T-lymphotropic virus Type I (STLV-I) in common chimpanzees (Pan troglodytes): evidence for interspecies transmission of the virus between chimpanzees and humans in Central Africa. *Virology* **238** 212-220

Wang JY, Ling H, Yang W & Craigie R 2001 Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. *EMBO J* **20** 7333-7343

Wang Y, Dewdney TG, Liu Z, Reiter SJ, Brunzelle J S, Kovari IA & Kovari L C 2012 Higher Desolvation Energy Reduces Molecular Recognition in Multi-Drug Resistant HIV-1 Protease. *Biology (Basel)* **1** 81-93

Wensing AM, Calvez V, Gunthard HF, Johnson VA, Paredes R, Pillay D, Shafer RW & Richman DD 2015 2015 Update of the Drug Resistance Mutations in HIV -1. *Top Antivir Med* **23** 132-141

Witvrouw M, Pannecouque C, Switzer WM, Folks T M, De Clercq E & Heneine W 2004 Susceptibility of HIV-2, SIV and SHIV to various anti-HIV-1 compounds: implications for treatment and postexposure prophylaxis. *Antivir Ther* **9** 57-65

Wlodawer A, Miller M, Jaskolski M, Sathyanarayana BK, Baldwin E, Weber IT, Selk LM, Clawson L, Schneider J & Kent SB 1989 Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science* **245** 616-621

Wolfe ND, Heneine W, Carr JK, Garcia AD, Shanmugam V, Tamoufe U, Torimiro JN, Prosser A T, Lebreton M, Mpoudi-Ngole E, McCutchan FE, Birx DL, Folks TM, Burke DS & Switzer WM 2005 Emergence of unique primate T-lymphotropic viruses among central African bushmeat hunters. *Proc Natl Acad Sci U.S.A.* **102** 7994-7999

Yedidi RS, Proteasa G, Martin PD, Liu Z, Vickrey JF, Kovari IA & Kovari LC 2014 A multi-drug resistant HIV-1 protease is resistant to the dimerization inhibitory activity of TLF-PafF. *J Mol Graph Model* **53** 105-111

Yewdell J, Anton L C, Bacik I, Schubert U, Snyder HL & Bennink JR 1999 Generating MHC class I ligands from viral gene products. *Immunol Rev* **172** 97-108

Zervoudi E, Saridakis E, Birtley JR, Seregin SS, Reeves E, Kokkala P, Aldhamen YA, Amalfitano A, Mavridis IM, James E, Georgiadis D & Stratikos E 2013 Rationally designed inhibitor targeting antigen-trimming aminopeptidases enhances antigen presentation and cytotoxic T-cell responses. *Proc Natl Acad Sci U.S.A.* **110** 19890-19895