Biol. Chem., Vol. 393, pp. 915–931, September 2012 • Copyright © by Walter de Gruyter • Berlin • Boston. DOI 10.1515/hsz-2012-0168

A catalogue of putative HIV-1 protease host cell substrates

Francis Impens[1,2], Evy Timmerman[1,2], An Staes[1,2], Kathleen Moens[1,2], Kevin K. Ariën[3,a], Bruno Verhasselt[3], Joël Vandekerckhove[1,2] and Kris Gevaert[1,2,*]

[1] Department of Medical Protein Research and Biochemistry, VIB, B-9000 Ghent, Belgium
[2] Department of Biochemistry, Ghent University, A. Baertsoenkaai 3, B-9000 Ghent, Belgium
[3] Department of Clinical Chemistry, Microbiology and Immunology, Ghent University Hospital, De Pintelaan 185, B-9000 Ghent, Belgium

*Corresponding author
e-mail: kris.gevaert@vib-ugent.be

Abstract

Processing of human immunodeficiency virus (HIV) proteins by the HIV-1 protease is essential for HIV infectivity. In addition, several studies have revealed cleavage of human proteins by this viral protease during infection; however, no large-scale HIV-1 protease degradomics study has yet been performed. To identify putative host substrates in an unbiased manner and on a proteome-wide scale, we used positional proteomics to identify peptides reporting protein processing by the HIV-1 protease, and a catalogue of over 120 cellular HIV-1 protease substrates processed *in vitro* was generated. This catalogue includes previously reported substrates as well as recently described interaction partners of HIV-1 proteins. Cleavage site alignments revealed a specificity profile in good correlation with previous studies, even though the ELLE consensus motif was not cleaved efficiently when incorporated into peptide substrates due to subsite cooperativity. Our results are further discussed in the context of HIV-1 infection and the complex substrate recognition by the viral protease.

Keywords: AIDS; COFRADIC; positional proteomics; protease degradomics; protein neo-N-termini; subsite cooperativity.

Introduction

Many pathogens encode their own protease(s) or use host proteases for their replication cycle. For the human immunodeficiency virus-1 (HIV-1), causing the acquired immunodeficiency syndrome (AIDS) pandemic, the processing of the viral Gag and Gag-Pol polyprotein precursors to generate the mature Gag and Pol proteins is well described (Freed, 2001). These processing events are catalyzed by the HIV-1 protease, a homodimeric aspartic protease, and take place during or shortly after virion release from the plasma membrane in a process called virus maturation. Failure of the HIV-1 virion to mature properly is associated with a complete loss of infectivity (Kohl et al., 1988); therefore, HIV-1 protease inhibitors are amongst the most effective protease inhibitor drugs used in health care (Drag and Salvesen, 2010).

Next to the processing of viral proteins, several host cell proteins have also been identified as substrates of the HIV-1 protease. In fact, according to the HIV-1 Human Protein Interaction Database at NCBI [http://www.ncbi.nlm.nih.gov/RefSeq/HIVInteractions/ (Fu et al., 2009)], more than 60 human proteins are known to be cleaved by the viral protease (interaction types 'cleaves,' 'degrades' or 'activates'). One example is the cleavage of human procaspase-8, which results in caspase activation and the induction of apoptosis in infected cells (Nie et al., 2002, 2007, 2008). Another example is the processing of eukaryotic translation initiation factor 4 gamma 1 (eIF4GI) and polyadenylate-binding protein 1 (PABP), which may lead to inhibition of cellular protein synthesis by blocking cap- and poly(A)-dependent initiation of translation in infected cells (Castello et al., 2009). Although host proteins might not be the primary substrates for the HIV-1 protease, these data indicate that the processing of host cell factors should be considered as more than just bystander events given the potentially profound effect on cellular homeostasis.

At present, no studies have yet been undertaken to systematically identify cellular substrates of the HIV-1 protease. A recent interactomics screen identified 54 human proteins as interaction partners of a catalytically inactive variant of the HIV-1 protease; however, only the eukaryotic translation initiation factor 3 subunit D (eIF3D) was confirmed as a substrate of the wild-type protease (Jager et al., 2012). As the current data on protein processing of host proteins during HIV-1 infection result from rather scattered studies, we opted for a large-scale positional proteomics approach to expand the present knowledge on putative host cell HIV-1 protease substrates. Several proteomics-based methods were developed to screen for protease cleavage events in complex samples [also referred to as degradomics (Lopez-Otin and Overall, 2002)], and one of these is the N-terminal COFRADIC technology developed in our lab (Gevaert et al., 2003; Staes et al., 2008, 2011). In contrast to shotgun proteomics, these technologies specifically isolate and identify protein terminal peptides and, thus, also neo-N-terminal peptides generated upon protease cleavage that serve as proxies to identify protease processing events [reviewed in (Impens et al., 2010a)].

Here, we used N-terminal COFRADIC to screen for protein processing in lysates of human Jurkat T-cells that were incubated *in vitro* with recombinant HIV-1 protease. We identified

[a] Present address: Department of Biomedical Sciences, Virology Unit, Institute of Tropical Medicine, Antwerp, Belgium.

more than 140 HIV-1 protease cleavage sites in about 120 host proteins, including known substrates such as actin and eIF3D (Tomasselli et al., 1991; Jager et al., 2012). Many substrates are involved in protein translation or are known interaction partners of HIV-1 proteins. Cleavage site alignments further showed a subsite-specificity profile corresponding to previous observations, but did not allow delineation of an efficiently cleaved consensus motif.

Results

Experimental setup

To identify putative novel host cell substrates of the HIV-1 protease, we incubated freeze-thaw lysates of Jurkat T-cells with recombinant HIV-1 (rHIV-1) protease. As this protease is more active under conditions of high ionic strength and low pH, a buffer containing 300 mM NaCl at pH 5.5 was used (Wondrak et al., 1991). Different protease concentrations and incubation times were first tested by monitoring the processing of vimentin, a known HIV-1 protease substrate (Shoeman et al., 1990) (Figure 1). Incubation of the Jurkat cell lysate for 60 min with 200 nM rHIV-1 protease led to almost complete vimentin processing, which was prevented by the HIV-1 protease inhibitor Ac-pepstatin, but not by the pan-caspase inhibitor zVADfmk. Based on these results, we decided to screen for processing events in lysates treated for 75 min with 200 nM rHIV-1 protease in the presence of 50 μM zVADfmk. The latter was added to inhibit possible activated caspases and thus to prevent detection of caspase-induced processing events (Nie et al., 2002, 2007, 2008).

The COFRADIC proteomics technology was used to isolate N-terminal peptides, including those generated by protease-mediated processing (Gevaert et al., 2003). In addition, stable isotope labeling by amino acids in cell culture (SILAC) labeling (Ong et al., 2002) with arginine residues was used to compare a protease-treated sample ($^{12}C_6$-Arg) with a control sample ($^{13}C_6$-Arg). As such, identification of newly formed N-termini by the HIV-1 protease was straightforward as such peptides were only picked up in their $^{12}C_6$-Arg form (Figure 2), different from other identified proteins that are not substrates. To increase the overall coverage of identified cleavage sites, the experiment was repeated three times. In total, we identified 148 rHIV-1 protease cleavage sites in 123 proteins (Table 1).

HIV-1 protease *in vitro* processed host substrates

The number of substrates identified in our screen almost doubles the number of host cell proteins that are known to be cleaved by the HIV-1 protease (Fu et al., 2009). Nevertheless, to our knowledge, only two proteins were previously reported as substrates of the viral protease. In the cytoskeletal protein actin we picked up three cleavage sites, one of which (between Leu 105 and Thr 106) was reported by Tomasselli et al. (1991) more than 20 years ago (Figure 3). In contrast, the processing of eIF3D was described recently by Jager et al. (2012) and found to occur between Met 114 and Leu 115, a position close to the one reported here (between Gln 120 and Ile 121). In the latter study, eIF3D was identified as a high confidence interaction partner of a catalytically inactive variant of the HIV-1 protease, together with 53 other human proteins. Interestingly, two of these proteins were also confirmed in our screen as substrates of the HIV-1 protease. Moreover, we found that seven host proteins reported as high confidence interaction partners of other HIV-1 proteins are also substrates of the HIV-1 protease [(Jager et al., 2012) and Table 1].

Given the *in vitro* nature of the experiment using total cell lysates, it is not surprising that substrates were derived from different cellular locations as revealed by gene ontology analysis using PIGOK (Jacob and Cramer, 2006). Substrates were distributed over the nucleus, cytoplasm and mitochondria, similar to all human proteins identified in our screen (including non-substrates) and all human proteins present in the Swiss-Prot database (Figure 4, top panel). Almost no substrates were annotated to be part of the endoplasmic reticulum, Golgi apparatus, lysosomes or plasma membrane. RNA-binding proteins were clearly enriched among the substrates, including several ribosomal proteins, ribonucleoproteins, RNA helicases and splicing factors. No such enrichment was observed for DNA-binding proteins, which is reflected in the higher fraction of substrates involved in translation compared to transcription (Figure 4, bottom panel). Further, visualization of functional interactions between substrate proteins using STRING (Jensen et al., 2009) pointed to ribosomal proteins and spliceosome components as two major clusters of RNA-binding proteins in our dataset (data not shown).

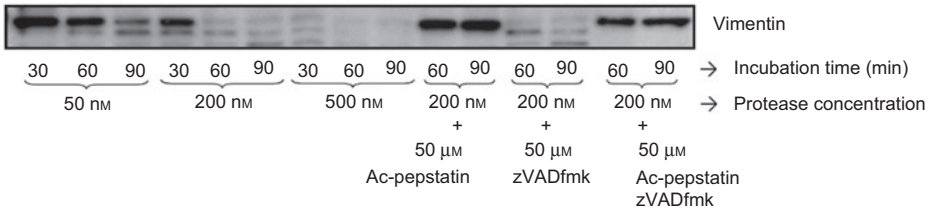

Figure 1 Immunoblot-based optimization of protease concentration and incubation time by monitoring degradation of the known HIV-1 protease substrate vimentin (Ac-pepstatin: HIV-1 protease inhibitor; zVADfmk: pan-caspase inhibitor).
Progressive vimentin processing was observed with increasing protease concentration and incubation time. Addition of Ac-pepstatin blocked this process, whereas the addition of zVADfmk had no effect.

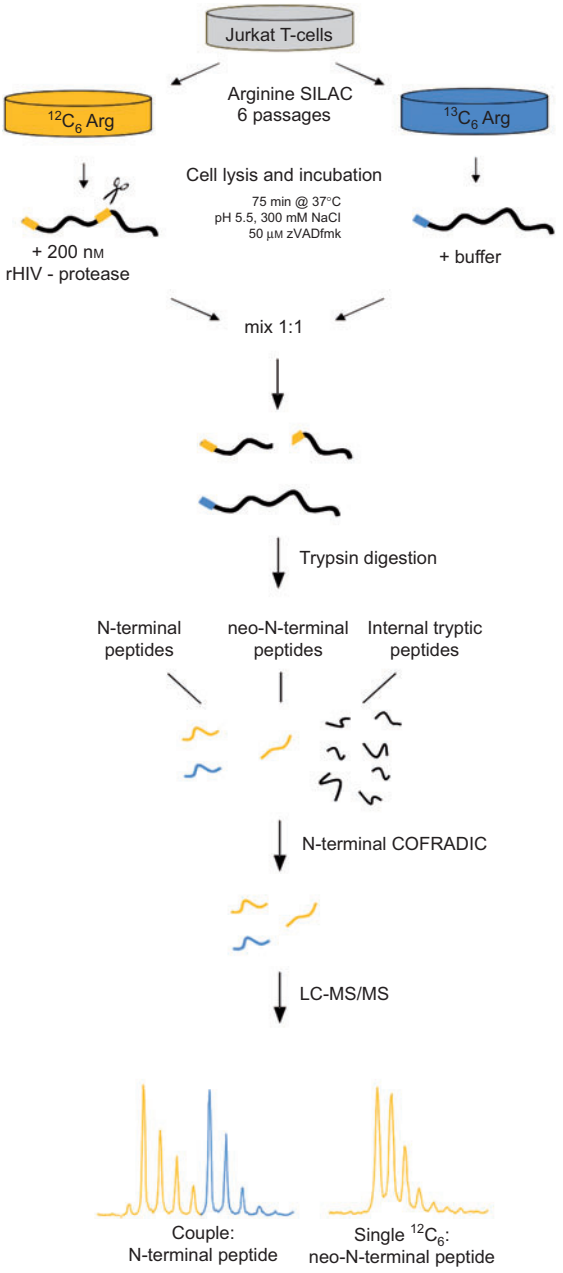

Figure 2 Outline of the COFRADIC experimental design. Light ($^{12}C_6$) and heavy ($^{13}C_6$) arginine SILAC-labeled Jurkat T-cells were incubated with 200 nM rHIV-1 protease and buffer, respectively, for 75 min at 37°C, following which both proteomes were mixed in a 1:1 ratio. Subsequent trypsin digestion resulted in a mixture of internal, N-terminal and neo-N-terminal peptides. N-terminal COFRADIC (Gevaert et al., 2003; Van Damme et al., 2008; Staes et al., 2008, 2011) was then used to isolate N-terminal and neo-N-terminal peptides prior to LC-MS(/MS) analysis, leading to MS-spectra showing one (light, neo-N-terminal peptides) or two (light and heavy, protein N-terminal peptides) isotopically labeled peptides.

Using Western blotting, we confirmed HIV-1 protease mediated processing of the probable ATP-dependent RNA helicase DDX5 (DDX5), a host RNA helicase involved in gene transcription and RNA splicing (Fuller-Pace and Moore, 2011). Several DDX5 fragments were observed when HIV-1 protease was added to Jurkat cell lysates, and DDX5 processing was prevented by ritonavir, a potent HIV-1 protease inhibitor (Figure 5). The size of the major DDX5 fragment observed matched with the expected molecular weight of the newly formed C-terminal fragment based on the DDX5 cleavage site found here (58 kDa, note that the antibody epitope is situated near the C-terminus of the protein), whereas lower MW bands probably indicate secondary cleavage products by host proteases.

HIV-1 protease cleavage site specificity

Information about cleavage site specificity of the HIV-1 protease was extracted using iceLogo (Colaert et al., 2009). This software tool was used to visualize cleavage site alignments and compare the amino acid frequency at positions before and after the cleavage site to their frequency found in all human proteins stored in the Swiss-Prot database (background control). Significantly over- or underrepresented residues are then depicted at the top or the bottom of the iceLogo, respectively. IceLogo representation of all cleavage sites identified in our screen showed an overall high overlap with the iceLogo obtained from 327 cleavage sites in peptide substrates that were reported using another proteomics approach, PICS (Schilling and Overall, 2008), which indicates no major differences in cleavage site specificity upon processing of protein and peptide substrates (Figure 6). Both iceLogos show an enrichment of residues two positions before and after the cleavage site (P2 to P2′), whereas positions beyond this region are clearly less important for cleavage site recognition, in line with previous observations (Beck et al., 2000, 2002). The enriched amino acids correspond well with the literature and show a striking symmetry between primed and unprimed sites that fits with the homodimeric nature of the protease (Tozser and Oroszlan, 2003). Whereas at P2 and P2′, negatively charged and hydrophobic, β-branched residues are preferred, large non-β branched hydrophobic and aromatic residues are enriched at P1 and P1′ (Beck et al., 2000, 2002) (Figure 6). Basic residues, as well as glycine and proline, are generally disfavored at all four positions. At P2, arginine residues were completely absent in both COFRADIC and PICS datasets (depicted pink at the lower part of the iceLogo). The same is true for proline residues at P1; however, the latter residue is allowed at the P1′ position, a rather unique feature of the HIV-1 protease required to cleave several sites in the viral polyproteins (Tozser and Oroszlan, 2003).

To test whether the palindromic consensus motif ELLE is efficiently cleaved by the HIV-1 protease, we incorporated the sequence into synthetic peptides and compared it with the IFLE-motif that was found to be the optimal sequence using phage display libraries (Beck et al., 2000). Fluorescently quenched peptides were initially used, and their processing was monitored by fluorimetry. However, significant inner filter effects interfered with these analyses, and therefore RP-HPLC separations of precursor peptides and their fragments were performed (Figure 7). The GEL**ELLE**LKDG peptide represents the consensus sequence of the COFRADIC sites, the

Table 1 List of 148 rHIV-1 protease cleavage sites in 123 human proteins.

Swiss-Prot accession	Peptide sequence	Start	End	m/z	z	delta m/z	Score/ threshold	# Spectra exp. 1	# Spectra exp. 2	# Spectra exp. 3	Protein name	Cleavage site	Isoforms	Interacting HIV-1 protein (Jager et al., 2012)
P62258	LVYQAKLAEQAER	7	19	804.952	2	0.0126	55/44	1			14-3-3 protein epsilon	DRED↓LVYQ		IN
P43686	VEKAQDEIPALSVSR	8	22	866.440	2	-0.0260	64/48	1			26S protease regulatory subunit 6B	IGIL↓VEKA		
P08708	QVTQPTVGMNFKTPR	117	131	905.582	2	0.1140	66/44			5	40S ribosomal protein S17	LSNL↓QVTQ		
P39019	LEGLKMVEKDQDGGR	106	120	916.434	2	-0.0711	92/44		1		40S ribosomal protein S19	VLQA↓LEGL		
P62266	IEENDEVLVAGFGR	93	106	796.862	2	-0.0340	66/44	1			40S ribosomal protein S23	CLNF↓IEEN		
P62847	YEKKKTSR	97	104	613.390	2	-0.0032	48/43		1		40S ribosomal protein S24	RHGL↓YEKK		
P61247	EVSLADLQNDEVAFR	67	81	875.916	2	-0.0146	59/44	1	1		40S ribosomal protein S3a	GRVF↓EVSL		
P61247	LADLQNDEVAFR	70	81	718.237	2	-0.1221	84/43	1			40S ribosomal protein S3a	FEVS↓LADL		
P62081	LELEMNSDLKAQLR	28	41	883.338	2	-0.1219	47/44	1			40S ribosomal protein S7	SQAL↓LELE		
P62081	EMNSDLKAQLR	31	41	705.897	2	0.0425	62/43	1	1		40S ribosomal protein S7	LLEL↓EMNS		
P10809	VEEGIVLGGGCALLR	432	446	794.352	2	-0.0744	61/48	1			60 kDa heat shock protein, mitochondrial	TRAA↓VEEG		
P27635	DEYEQLSSEALEAAR	54	68	881.441	2	-0.0041	47/44		1		60S ribosomal protein L10	HMVS↓DEYE	Q96L21	
P27635	EQLSSEALEAAR	57	68	677.921	2	0.0423	54/43		1		60S ribosomal protein L10	SDEY↓EQLS	Q96L21	
Q02543	KVEEIAASKCR	128	138	713.437	2	0.0548	56/43	1			60S ribosomal protein L18a	IQIM↓KVEE		
P62829	ATVKKGKPELR	63	73	706.929	2	-0.0484	60/43		1		60S ribosomal protein L23	DMVM↓ATVK		
P39023	LKGCVVGTKKR	332	342	713.447	2	0.0189	56/43	1			60S ribosomal protein L3	DFVM↓LKGC		
P39023	EEKKAFMGPLKKDR	382	395	640.084	3	0.0674	72/44	1	1		60S ribosomal protein L3	FQTM↓EEKK		
P62888	LAIIDPGDSDIIR	93	105	721.957	2	0.0643	54/43	1			60S ribosomal protein L30	RVCT↓LAII		
P42766	TVINQTQKENLR	58	69	767.483	2	0.0615	77/43	2	1	2	60S ribosomal protein L35	ARVL↓TVIN		
P46777	GQPGAFTCYLDAGLAR	136	151	871.533	2	0.1164	56/43			1	60S ribosomal protein L5	ESID↓GQPG		
Q02878	ATVTKPVGGDKNGGTR	89	104	847.507	2	0.0574	63/44	1	1		60S ribosomal protein L6	EKVL↓ATVT		

Table 1 (Continued)

Swiss-Prot accession	Peptide sequence	Start	End	m/z	z	delta m/z	Score/ threshold	# Spectra exp. 1	# Spectra exp. 2	# Spectra exp. 3	Protein name	Cleavage site	Isoforms	Interacting HIV-1 protein (Jager et al., 2012)
P18124	VEEKKKEVPAVPETLKKKR	4	22	851.159	3	-0.0200	47/44	2			60S ribosomal protein L7	xMEG↓VEEK		
P32969	VELSLLGKKKKR	43	54	812.506	2	-0.0216	47/44	1			60S ribosomal protein L9	NHIN↓VELS		
P32969	SLLGKKKKR	46	54	641.929	2	-0.0011	52/43	1			60S ribosomal protein L9	NVEL↓SLLG		
Q9BTT0	VEEGEEEEEEEEGGLR	236	251	947.507	2	0.1151	49/44			1	Acidic leucine-rich nuclear phosphoprotein 32 family member E	DDDY↓VEEG		
P60709	LVVDNGSGMCKAGFAGDDAPR	8	28	748.723	3	0.0529	66/44	2	1	3	Actin, cytoplasmic 1	DIAA↓LVVD		
P60709	TEAPLNPKANR	106	116	650.971	2	0.1183	64/43	4	1	1	Actin, cytoplasmic 1	PVLL↓TEAP	P62736, P63261, P63267, P68032, P68133, Q6S8J3, Q9BYX7	
P60709	TVLSGGTTMYPGIADR	297	312	850.374	2	-0.0423	88/44	1			Actin, cytoplasmic 1	LYAN↓TVLS	P63261	
O43823	YEEPDTKLAR	311	320	656.308	2	-0.0232	44/43	1			A-kinase anchor protein 8	QFQL↓YEEP		
P06733	AVCKAGAVEKGVPLYR	116	131	927.134	2	0.1227	51/44			1	α-enolase	GVSL↓AVCK		
Q01433	PEESPIEQLEER	148	159	753.780	2	-0.1223	68/43		1		AMP deaminase 2	PYEF↓PEES		
Q9BVC5	TLEQKNIAVETDVR	30	43	854.047	2	0.0968	60/48	1			Ashwin	FLLL↓TLEQ		
P25705	ILGADTSVDLEETGR	59	73	810.942	2	0.0380	65/32	1			ATP synthase subunit α, mitochondrial	LEER↓ILGA		
Q9NVP1	SETQNGDVSEETMGSR	36	51	897.399	2	-0.0119	62/44		1		ATP-dependent RNA helicase DDX18	NLTL↓SETQ		
Q8WYA6	TVVEEADDDKKR	47	58	770.372	2	-0.0167	72/43	1			β-catenin-like protein 1	EEEM↓TVVE		
Q9P203	EILSSLLPFVR	511	521	659.927	2	0.0403	32/31	1			BTB/POZ domain-containing protein 7	EELR↓EILS		
O43570	EMINNFR	269	275	484.834	2	0.1028	35/30	1			Carbonic anhydrase 12	PSPR↓EMIN		
P46092	GGSCPSGPQPR	327	337	576.289	2	-0.0048	49/31		1		C-C chemokine receptor type 10	RLLR↓GGSC		
Q99459	YQALDADFR	232	240	572.361	2	0.0892	68/43	1			Cell division cycle 5-like protein	SEEN↓YQAL		
Q99459	PENAEKELEER	495	505	717.464	2	0.1164	46/43	1			Cell division cycle 5-like protein	EIVL↓PENA		
Q6P1N0	LEVMVRIR	784	791	530.940	2	0.1249	78/31			2	Coiled-coil and C2 domain-containing protein 1A	TGGR↓LEVM		

Table 1 (Continued)

Swiss-Prot accession	Peptide sequence	Start	End	m/z	z	delta m/z	Score/ threshold	# Spectra exp. 1	# Spectra exp. 2	# Spectra exp. 3	Protein name	Cleavage site	Isoforms	Interacting HIV-1 protein (Jager et al., 2012)
Q96MW1	SDKLLFR	160	166	484.798	2	0.0176	52/42	1			Coiled-coil domain-containing protein 43	MNIG↓SDKL		
P33316	EVQALDDTER	232	241	610.829	2	0.0410	90/43			1	Deoxyuridine 5′-triphosphate nucleotidohydrolase, mitochondrial	PEIE↓EVQA		
P09622	LEYGASCEDIAR	471	482	717.786	2	-0.0788	74/43		1		Dihydrolipoyl dehydrogenase, mitochondrial	AALA↓LEYG		
P33992	YSDSFGGDAQADEGQAR	11	27	909.921	2	0.0445	120/44	1	2		DNA replication licensing factor MCM5	PGIF↓YSDS		
P33992	SFGGDAQADEGQAR	14	27	727.377	2	0.0616	79/43	1			DNA replication licensing factor MCM5	FYSD↓SFGG		
Q99615	AATEPELLDDQEAKR	12	26	888.504	2	0.0612	56/44	1			DnaJ homolog subfamily C member 7	DVVM↓AATE		Vif
Q99615	TILSDPKKKTR	436	446	733.910	2	-0.0347	50/43	1			DnaJ homolog subfamily C member 7	GEAF↓TILS		Vif
O75937	QIDPEVTDEEIKKR	63	76	918.071	2	0.0875	62/44			1	DnaJ homolog subfamily C member 8	FEVL↓QIDP		
Q7Z6Z7	AENVVIVASQKR	3687	3698	702.411	2	0.0085	64/48	1			E3 ubiquitin-protein ligase HUWE1	RFDM↓AENV		Vif
P26641	AQYSGAQVR	21	29	515.776	2	-0.0261	46/43		1		Elongation factor 1-gamma	ALIA↓AQYS		
P13639	YISKMVPTSDKGR	396	408	819.936	2	-0.0342	68/44	2	1		Elongation factor 2	PLMM↓YISK		
P49411	AEGGGAKFKKYEEIDNAPEER	82	102	840.103	3	0.0226	55/44	1			Elongation factor Tu, mitochondrial	TKIL↓AEGG		
P49411	VVAANDGPMPQTR	150	162	708.941	2	0.0957	60/43	1	1		Elongation factor Tu, mitochondrial	GCIL↓VVAA		
Q14240	FVTEEDKR	376	383	560.390	2	0.0657	45/42		1		Eukaryotic initiation factor 4A-II	VAIN↓FVTE		
O43324	LEKSLGLSKGNKYSAQGER	10	28	749.024	3	-0.0473	86/44	1			Eukaryotic translation elongation factor 1 epsilon-1	ELSL↓LEKS		MA
O15371	ILPKSAKQKER	121	131	742.516	2	0.0201	50/43		1		Eukaryotic translation initiation factor 3 subunit D	FNLQ↓ILPK		POL;PR
P63241	FETGDAGASATFPMQCSALR	6	25	726.566	3	-0.0872	73/44	1			Eukaryotic translation initiation factor 5A-1	DDLD↓FETG	Q6IS14	
P63241	YLSLLQDSGEVR	97	108	712.950	2	0.0807	52/43			1	Eukaryotic translation initiation factor 5A-1	IQDG↓YLSL		

Table 1 (Continued)

Swiss-Prot accession	Peptide sequence	Start	End	m/z	z	delta m/z	Score/ threshold	# Spectra exp. 1	# Spectra exp. 2	# Spectra exp. 3	Protein name	Cleavage site	Isoforms	Interacting HIV-1 protein (Jager et al., 2012)
P63241	LSLLQDSGEVR	98	108	631.399	2	0.0614	73/43			1	Eukaryotic translation initiation factor 5A-1	QDGY↓LSLL		
P63241	LQDSGEVR	101	108	474.780	2	0.0424	72/43			1	Eukaryotic translation initiation factor 5A-1	YLSL↓LQDS		
Q02790	VDISPKQDEGVLKVIKR	22	38	1052.661	2	0.0446	64/49	1			FK506-binding protein 4	PMEG↓VDIS		PR
P04075	DESTGSIAKR	33	42	580.385	2	0.0471	58/43		1		Fructose-bisphosphate aldolase A	ILAA↓DEST		
Q13630	VSSKDADLTDTAQTR	41	55	849.416	2	-0.0033	58/44	1			GDP-L-fucose synthetase	DWVF↓VSSK		
P00390	LVIGGGSGGLASAR	68	81	633.383	2	-0.0138	89/43		2		Glutathione reductase, mitochondrial	SYDY↓LVIG		
Q9BVP2	EETGEALSEETTAGEQSTR	510	528	1038.532	2	0.0332	51/44		1		Guanine nucleotide-binding protein-like 3	MFAA↓EETG		
Q9NX24	LVNQNPIAQPLASR	28	41	783.446	2	0.0078	65/44	1			H/ACA ribonucleoprotein complex subunit 2	YQEL↓LVNQ		
P07900	VSLKDYCTR	474	482	616.353	2	0.0437	51/43	1			Heat shock protein HSP 90-alpha	GDEM↓VSLK		
P07900	YITGETKDQVANSAFVER	492	509	706.685	3	0.0001	76/44	1			Heat shock protein HSP 90-α	KHIY↓YITG		
P08238	YITGESKEQVANSAFVER	484	501	708.722	3	0.0083	82/44		1		Heat shock protein HSP 90-β	KSIY↓YITG	Q58FF7	
Q99729	IELPMDPKLNKR	182	193	802.960	2	0.0124	59/43			1	Heterogeneous nuclear ribonucleoprotein A/B	EIEA↓IELP		
P14866	VEMADGYAVDR	393	403	643.843	2	0.0586	61/43	1			Heterogeneous nuclear ribonucleoprotein L	GAAM↓VEMA		
Q00839	FEYIEENKYSR	256	266	787.415	2	-0.0018	70/48	1	1		Heterogeneous nuclear ribonucleoprotein U	GRGY↓FEYI		
Q00839	IEENKYSR	259	266	564.845	2	0.0586	59/43	1			Heterogeneous nuclear ribonucleoprotein U	YFEY↓IEEN		
P22626	GKIDTIEIITDR	136	147	735.431	2	-0.0197	71/43		1		Heterogeneous nuclear ribonucleoproteins A2/B1	FEEY↓GKID		
Q13601	FLKANQKKR	287	295	656.911	2	0.0084	52/43	1			KRR1 small subunit processome component homolog	SGEY↓FLKA		
Q9HBX8	LLPSGMCQQLPR	203	214	722.904	2	0.0337	38/36	1			Leucine-rich repeat-containing G-protein coupled receptor 6	AGIR↓LLPS		

Table 1 (Continued)

Swiss-Prot accession	Peptide sequence	Start	End	m/z	z	delta m/z	Score/ threshold	# Spectra exp. 1	# Spectra exp. 2	# Spectra exp. 3	Protein name	Cleavage site	Isoforms	Interacting HIV-1 protein (Jager et al., 2012)
P05455	FGDFNLPR	25	32	505.867	2	0.1140	55/42			1	Lupus La protein	IEYY↓FGDF		
P40926	VVVIPAGVPR	95	104	526.363	2	0.0316	66/42			1	Malate dehydrogenase, mitochondrial	KGCD↓VVVI		
P62072	SMQDEELMKR	73	82	694.850	2	0.0385	48/43			1	Mitochondrial import inner membrane translocase subunit Tim10	LTEL↓SMQD		
Q9BYG3	ILQKTESISKTNR	189	201	826.964	2	-0.0087	46/44	1			MKI67 FHA domain-interacting nucleolar phosphoprotein	FPSL↓ILQK		
O15226	VLSKPSKGQKR	22	32	707.430	2	-0.0450	44/43		1		NF-κ-B-repressing factor	SYDL↓VLSK		
Q15233	LTIDLKNFR	55	63	605.385	2	0.0332	62/43	2			Non-POU domain-containing octamer-binding protein	QNEG↓LTID		
Q15233	IEMEKQQQDQVDR	275	287	879.969	2	0.0124	65/44		2		Non-POU domain-containing octamer-binding protein	WKAL↓IEME		
Q15233	EKQQQDQVDR	278	287	685.345	2	-0.0305	61/43	2	2		Non-POU domain-containing octamer-binding protein	LIEM↓EKQQ		
P49321	SSTSGFTPGGGGSSVSMIASR	677	697	996.021	2	0.0640	54/44			1	Nuclear autoantigenic sperm protein	TLVE↓SSTS		
P49321	EEEAENQAESR	760	770	668.973	2	0.1921	52/47			1	Nuclear autoantigenic sperm protein	SENM↓EEEA		
Q14980	NTPKKLGNSLLR	2054	2065	738.311	2	-0.1298	59/47	1			Nuclear mitotic apparatus protein 1	FSIL↓NTPK		
P23511	MVPGAGSVPAIQR	239	251	672.381	2	0.0257	68/43			1	Nuclear transcription factor Y subunit α	GMVM↓MVPG		
Q8N6M0	TEELDEEEQLLR	6	17	777.828	2	-0.0850	61/43		1		OTU domain-containing protein 6B	EAVL↓TEEL		
P62937	FDIAVDGEPLGR	7	18	667.380	2	0.0424	91/43	2		2	Peptidyl-prolyl *cis-trans* isomerase A	PTVF↓FDIA		
P62937	DIAVDGEPLGR	8	18	593.835	2	0.0316	75/42			2	Peptidyl-prolyl *cis-trans* isomerase A	TVFF↓DIAV		
P62937	IAVDGEPLGR	9	18	536.326	2	0.0361	53/42			1	Peptidyl-prolyl *cis-trans* isomerase A	VFFD↓IAVD		
P62937	FADKVPKTAENFR	24	36	829.522	2	0.0810	64/43	1		1	Peptidyl-prolyl *cis-trans* isomerase A	SFEL↓FADK		
O43447	VVPKTAENFR	70	79	625.875	2	0.0281	52/43			1	Peptidyl-prolyl *cis-trans* isomerase H	LFAD↓VVPK	P30405	

Table 1 (Continued)

Swiss-Prot accession	Peptide sequence	Start	End	m/z	z	delta m/z	Score/ threshold	# Spectra exp. 1	# Spectra exp. 2	# Spectra exp. 3	Protein name	Cleavage site	Isoforms	Interacting HIV-1 protein (Jager et al., 2012)
Q06830	FVCPTEIIAFSDR	50	62	800.354	2	-0.0381	61/43			1	Peroxiredoxin-1	LDFT↓FVCP		
P32119	FVCPTEIIAFSNR	48	60	799.904	2	0.0040	60/48			1	Peroxiredoxin-2	LDFT↓FVCP		
O15305	FDVDGTLTAPR	11	21	618.803	2	-0.0082	60/43	1			Phosphomannomutase 2	ALCL↓FDVD		
P09874	LATEDKEALKKQLPGVKSEGKR	186	207	901.150	3	-0.0609	53/48		1		Poly [ADP-ribose] polymerase 1	GFSL↓LATE		
P26599	AASAAAVDAGMAMAGQSPVLR	165	185	674.760	3	0.1017	109/44	4	1		Polypyrimidine tract-binding protein 1	NLAL↓AASA		
Q9Y257	LRAQLLQR	43	50	499.390	2	0.0804	46/42	1			Potassium channel subfamily K member 6	ELET↓LRAQ		
A5A3E0	LVIDNGSGMCKAGFAGDDAPR	8	28	1131.935	2	-0.1293	96/44	2	3		POTE ankyrin domain family member F	DTAV↓LVID	P63261, Q6S8J3	
Q92841	QEYGAGTYGASSTTSTGR	574	591	922.770	2	-0.1811	59/44	2	1		Probable ATP-dependent RNA helicase DDX17	SYTA↓QEYG		
P17844	YEANFPANVMDVIAR	97	111	885.927	2	0.0027	96/44	2	3		Probable ATP-dependent RNA helicase DDX5	VLNF↓YEAN		
P51531	LEMNPGYEVAPR	617	628	718.909	2	0.0668	50/43	1			Probable global transcription activator SNF2L2	LDAW↓LEMN	P51532	
O00155	ARALDGACGR	316	325	546.304	2	0.0350	35/31			1	Probable G-protein coupled receptor 25	RSFR↓ARAL		
Q99848	LEGDQKPLAQR	213	223	672.930	2	0.0642	50/43	1			Probable rRNA-processing protein EBP2	KLDF↓LEGD		
Q9UQ80	EVQDAELKALLQSSASR	347	363	967.827	2	-0.1843	49/48	1			Proliferation-associated protein 2G4	KSEM↓EVQD		
P60900	AINQGGLTSVAVR	31	43	668.944	2	0.0286	55/47		1		Proteasome subunit α type-6	YAFK↓AINQ		
P28065	VEFDGGVVMGSDSR	26	39	758.396	2	0.0585	66/43	1			Proteasome subunit β type-9	TIMA↓VEFD		
P41223	IEPTLDELDQKMR	18	30	847.488	2	0.0625	46/44	1			Protein BUD31 homolog	GWEL↓IEPT		
O60610	EEAKELVGR	1238	1246	560.904	2	0.1019	50/43	1			Protein diaphanous homolog 1	PTIL↓EEAK		PR
P30101	LELTDDNFESR	28	38	692.492	2	0.1723	71/43	1			Protein disulfide-isomerase A3	ASDV↓LELT		GP41; GP160
Q96PU8	AEIKLKR	164	170	499.883	2	0.0209	55/31	1	1		Protein quaking	AQNR↓AEIK		
Q9P258	LELDGAPGGGKR	57	68	630.411	2	0.0716	72/43	1			Protein RCC2	DEDG↓LELD		
Q9Y5G7	LLETKGEPR	797	805	566.851	2	0.0307	67/43			1	Protocadherin γ-A6	ITQD↓LLET		

Table 1 (Continued)

Swiss-Prot accession	Peptide sequence	Start	End	m/z	z	delta m/z	Score/ threshold	# Spectra exp. 1	# Spectra exp. 2	# Spectra exp. 3	Protein name	Cleavage site	Isoforms	Interacting HIV-1 protein (Jager et al., 2012)
P55786	FVKDVFSPIGER	678	689	742.476	2	0.0765	61/43			1	Puromycin-sensitive aminopeptidase	EIQE↓FVKD		
O60231	VEEESGAPGEEQR	315	327	734.378	2	0.0119	64/43		1		Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX16	AVDL↓VEEE		
P14618	LAQKMMIGR	307	315	585.317	2	0.0063	65/42	1			Pyruvate kinase isozymes M1/M2	EKVF↓LAQK	P30613	
P14618	LSGETAKGDYPLEAVR	360	375	901.526	2	0.0192	76/44		1		Pyruvate kinase isozymes M1/M2	DCIM↓LSGE		
P14618	SGETAKGDYPLEAVR	361	375	841.944	2	0.0225	44/44	1			Pyruvate kinase isozymes M1/M2	CIML↓SGET		
P18754	TLGQGDVGQLGLGENVMER	39	57	678.796	3	0.1318	57/44	1			Regulator of chromosome condensation	GLVL↓TLGQ		
Q92900	YEGSLQNGVTAADR	728	741	763.907	2	0.0527	81/43	3			Regulator of nonsense transcripts 1	SNIF↓YEGS		
P13489	LQALEKDKPSLR	446	457	767.483	2	0.0450	33/32	1			Ribonuclease inhibitor	MEDR↓LQAL		NC
Q15050	QVAKVSTASVGR	223	234	646.917	2	0.0487	98/43	1			Ribosome biogenesis regulatory protein homolog	GRAM↓QVAK		
Q9NW13	FQTKAEVEQVELPDGKKR	649	666	761.322	3	-0.0894	49/44	1			RNA-binding protein 28	SWTG↓FQTK		
Q9Y5S9	VEQDGDEPGPQR	57	68	686.435	2	0.1279	51/43			1	RNA-binding protein 8A	DYDS↓VEQD		
Q8IYB3	LELKKEEIKQR	124	134	800.503	2	-0.0166	66/44		2		Serine/arginine repetitive matrix protein 1	PSAF↓LELK		
Q13523	LEDLEKQR	106	113	563.836	2	-0.0094	54/43		1		Serine/threonine-protein kinase PRP4 homolog	DLAL↓LEDL		
P10768	VVIAPDTSPR	77	86	550.373	2	0.0674	56/46	1			S-formylglutathione hydrolase	EHGL↓VVIA		
O76094	LENSAGATYIR	520	530	623.393	2	0.0331	64/43	1	2		Signal recognition particle 72 kDa protein	DVEA↓LENS		
Q13573	LEKSVSQKVAAAMPVR	151	166	933.034	2	0.0122	50/49	1			SNW domain-containing protein 1	TRVA↓LEKS		
Q99865	QQTRAAAGR	12	20	502.821	2	0.0591	48/43	1			Spindlin-2A	EAEG↓QQTR	Q9BPZ2	
Q13435	AEFQTKTEEEEIDR	669	682	907.946	2	0.0216	56/44	2	1		Splicing factor 3B subunit 2	GTNA↓AEFQ		

Table 1 (Continued)

Swiss-Prot accession	Peptide sequence	Start	End	m/z	z	delta m/z	Score/ threshold	# Spectra exp. 1	# Spectra exp. 2	# Spectra exp. 3	Protein name	Cleavage site	Isoforms	Interacting HIV-1 protein (Jager et al., 2012)
Q7KZF4	PVLEEKER	671	678	545.398	2	0.0988	60/46	1			Staphylococcal nuclease domain-containing protein 1	EEVM↓PVLE		
Q9UH65	TEAELEER	567	574	514.315	2	0.0316	70/43		1		Switch-associated protein 70	PAAF↓TEAE		
Q9Y2W1	LVYAPPGKEKQR	474	485	763.959	2	-0.0197	61/43	1	1		Thyroid hormone receptor-associated protein 3	KWEG↓LVYA		GAG
P49711	VEVSKLKR	361	368	550.413	2	0.0271	63/43	1	1		Transcriptional repressor CTCF	DYAS↓VEVS		
P25490	TLVTVAAAGKSGGGGSSSSGGGR	149	171	670.976	3	-0.0322	55/44	1			Transcriptional repressor protein YY1	YIEQ↓TLVT		
P29401	VAILDINR	179	186	479.930	2	0.1457	45/42			1	Transketolase	LDNL↓VAIL		
P68363	VDLEPTVIDEVR	68	79	718.423	2	0.0028	70/43	2	1		Tubulin α-1B chain	RAVF↓VDLE	Q71U36, Q9BQE3	
P68363	MVDNEAIYDICR	203	214	783.405	2	0.0200	55/43		1		Tubulin α-1B chain	DCAF↓MVDN	P68366, Q13748, Q6PEY2, Q71U36, Q9BQE3, Q9NY65	
O43290	VTFKKTKR	397	404	594.392	2	0.0189	45/43	1			U4/U6.U5 tri-snRNP-associated protein 1	PEEM↓VTFK		
P42768	FIEDQGGLEAVR	292	303	689.891	2	0.0426	57/47			1	Wiskott-Aldrich syndrome protein	LIYD↓FIED		
Q7Z2W4	LASASASAER	261	270	504.325	2	0.0689	69/43	1			Zinc finger CCCH-type antiviral protein 1	SQEF↓LASA		
O95125	ELCHQWLR	69	76	594.450	2	0.1684	40/35	1			Zinc finger protein 202	IRLR↓ELCH	O43309, P17029, Q3MJ62, Q96JS3, Q96LW9	
O60281	QKASNLKR	1771	1778	541.436	2	0.1298	38/31	1			Zinc finger protein 292	QSER↓QKAS		
Q86UK7	EVAAAVR	344	350	760.583	1	0.1512	40/36	1			Zinc finger protein 598	EEDR↓EVAA		

Peptides are ranked by the Swiss-Prot annotated name of their corresponding proteins and by increasing cleavage site position. Columns from left to right contain the Swiss-Prot accession number of the substrate protein, the sequence of the identified neo-N-terminal peptide, the start and end position of the peptide in the protein sequence, m/z, charge (z), delta m/z, Mascot score and threshold of the best scoring MS/MS spectrum (95% confidence level), the number of MS/MS spectra by which the neo-N-terminal peptide was identified in the individual experiments, the protein name, the actual cleavage site position including the surrounding amino acid sequence, possible protein isoforms in which the same peptide sequence can be found and the HIV protein the substrate was found to interact with by Jager et al. (2012).

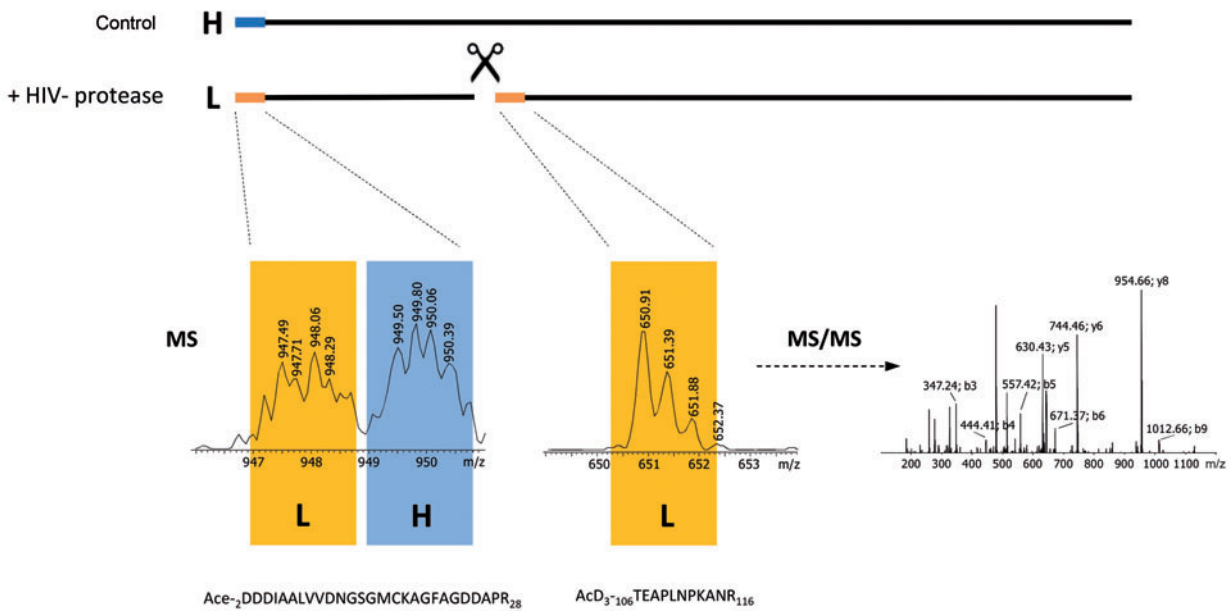

Figure 3 Identification of the known HIV-1 protease cleavage site in actin between Leu 105 and Thr 106.
Two different peptides from actin were identified: actin's N-terminal peptide, 2-DDDIAALVVDNGSGMCKAGFAGDDAPR-28 (N-terminus acetylated, lysine trideutero-acetylated, methionine oxidized and cysteine carbamidomethylated) was present in both light and heavy labeled proteome digests (left ion trap MS spectrum of its triply charged precursor), and a second peptide that was only present in the heavy labeled cell lysate treated with HIV-1 protease (middle ion trap MS spectrum of its doubly charged precursor). Upon MS/MS analysis (right MS/MS spectrum, b and y fragment ions indicated), this peptide was identified as 106-TEAPLNPKANR-116 (N-terminus and lysine were both trideuteroacetylated) and pointed to a previously reported cleavage site of the HIV-1 protease (Tomasselli et al., 1991).

GSG**IFLE**TSLY peptide was previously reported by Beck et al. (2000) as an optimal substrate for the HIV-1 protease, and GSG**ELLE**TSL contains the ELLE-motif surrounded by the residues of the sequence reported by Beck et al. Remarkably, much less efficient cleavage of the ELLE-motif was observed, even though Ile at P2 and Phe at P1 showed lower degrees of enrichment in both iceLogos compared to Glu and Leu at these positions (Figure 6).

Discussion

We here provide a catalogue of 123 human proteins cleaved *in vitro* by the HIV-1 protease. This host substrate list was generated by incubating lysates of human Jurkat T-cells with recombinant HIV-1 protease, followed by identification of neo-N-terminal, proteolysis-reporter peptides by mass spectrometry. In this way, information on both substrate identity and exact cleavage site location was obtained.

Two proteins on our list were reported before as substrates of the HIV-1 protease. Although this overlap might seem small, it validates our approach and its results, and it should further be noted that certain cleavage sites might be missed by mass spectrometry-based degradomics screens for different reasons. Obviously, processing should result in protein fragments and neo-N-terminal peptides that are stable and identifiable by mass spectrometry. In the case of vimentin, four HIV-1 protease cleavage sites have been reported (Shoeman et al., 1990), but three of these sites will lead to tryptic neo-N-terminal peptides that are too short to allow confident identification. Furthermore, only a fraction of the peptides injected on LC-MS/MS instruments will be fragmented for identification, a phenomenon generally known as undersampling (Liu et al., 2004). Interestingly, ten of the identified substrates recently came out of an interactomics study as host interaction partners of HIV-1 proteins (Jager et al., 2012). This overlap suggests an interesting and unexpected interplay between the viral protease, the web of its host cell substrates and other HIV-1 proteins. PIGOK (GO) analysis revealed that our substrate catalogue is enriched for RNA-binding proteins often involved in RNA splicing and protein translation. The latter observation suggests that modulation of protein translation during infection is probably achieved by proteolytic processing of additional host factors other than eIF4GI, eIF3D and PABP (Castello et al., 2009; Jager et al., 2012). We further validated our substrate catalogue by confirming DDX5 as an *in vitro* substrate of the HIV-1 protease by western blot. Interestingly, DDX5 was recently found to be associated with the HIV-1 Rev protein by co-immunoprecipitation, and silencing of this host RNA helicase in reporter HeLa cells strongly increased viral production and virion infectivity (Naji et al., 2012). It might be worth further investigating if proteolytic inactivation of host DDX5 is a strategy used by the virus to enhance its own infectivity and replication.

In contrast to consensus sequences derived from catalogues of caspases, granzymes and cathepsins (Demon et al., 2009; Van Damme et al., 2009; Impens et al., 2010b), we did not observe efficient cleavage of the iceLogo-suggested

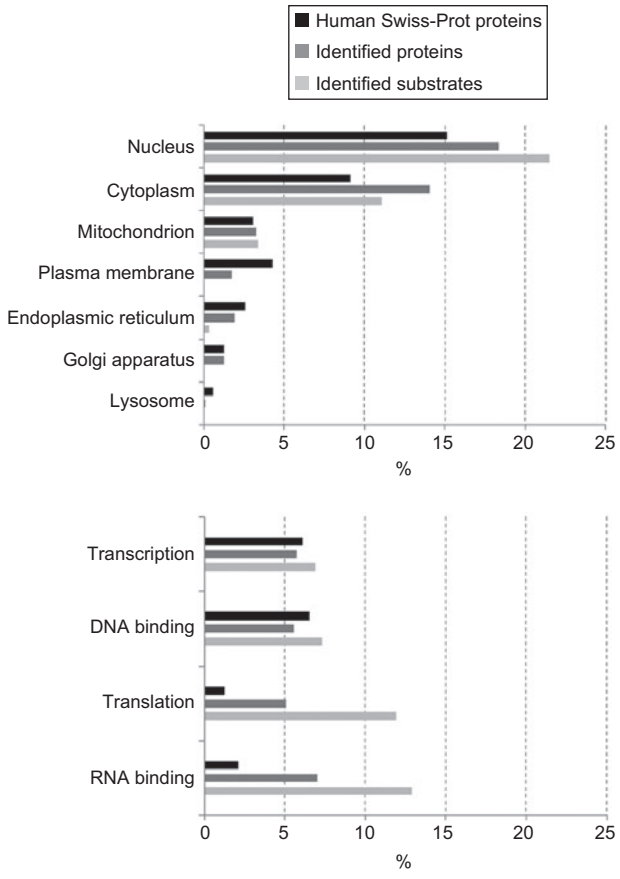

Figure 4 Substrate Gene Ontology analysis using PIGOK (Jacob and Cramer, 2006).
Gene ontology annotation was compared for all human proteins present in the Swiss-Prot database (black bars), all proteins identified in our screen (substrates and non-substrates, dark grey bars) and all identified substrates (light grey bars). Substrates were distributed over different cellular compartments as expected (top), but showed a remarkable enrichment of RNA-binding proteins involved in translation but not of DNA-binding proteins and proteins involved in transcription.

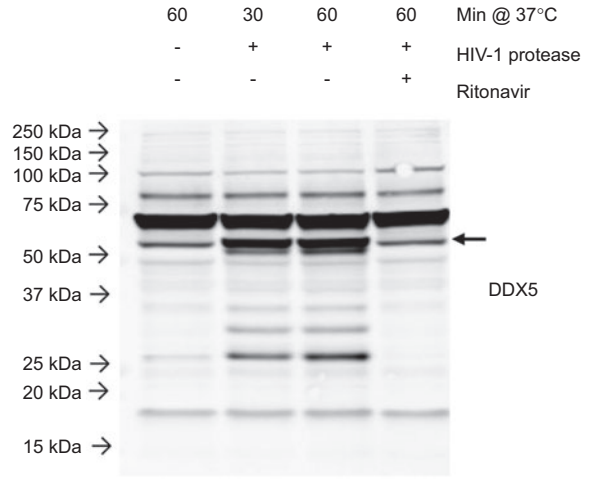

Figure 5 DDX5 is directly processed by the HIV-1 protease.
Jurkat cell lysates were incubated for the indicated time with 200 nM recombinant HIV-1 protease [or with HIV-1 protease (control)] in the absence or presence of 50 μM ritonavir. The molecular weight of the DDX5 fragment indicated with an arrow corresponds to the size of the C-terminal fragment (58 kDa) generated by HIV-1 protease cleavage after Phe 96 (Table 1).

ELLE-motif when tested in peptide substrates (Figure 7). We attribute this observation to subsite cooperativity [a substrate residue at one subsite influencing the binding of residues at other positions (Ng et al., 2009)], which has been described for the HIV-1 protease (Tozser and Oroszlan, 2003). A similar observation was recently reported for astacin proteases where the presence of Asp at P1′ and Pro at P2′ was found to be mutually exclusive when incorporated into peptide substrates. Also in this study, Asp and Pro were preferred residues at the P1′ and P2′ positions as revealed by sequence logo representation (Becker-Pauly et al., 2011). In the case of the HIV-1 protease, several types of cooperativity have been suggested, with *cis* and *trans* interactions between subsites located on the same or opposite subunits, respectively, and *coupled* interactions influence binding in multiple adjacent subsites (Ridky et al., 1996; Tozser et al., 1997; Tozser and Oroszlan, 2003). We believe that *trans* cooperativity between the P1 and P2 subsites is responsible for the difference in cleavage observed between the IFLE and ELLE motifs, in line with previous observations (Ridky et al., 1996). In the complete dataset, Glu (20.3%) and Leu (25.7%) are the most observed amino acids at P2 and P1, respectively, whereas Ile (12.2%) and Phe (11.5%) show an overall lower abundance at these positions (Supplementary Table 1). However, when only considering sites with a P2 Glu, the presence of both P1 Leu (20.0%) and P1 Phe (6.7%) residues is decreased with about 5%, which is not the case for sites with a P2 Ile [P1 Leu (27.8%) and P1 Phe (11.1%)]. This cooperativity is reciprocal, as sites with a P1 Leu and P1 Phe both show a decrease in P2 Glu (respectively, 15.8% and 11.8%), whereas, in the presence of P2 Ile, these frequencies are largely unchanged (13.2% and 11.8%, respectively). In general, our data indicate that mainly P2′ Glu residues are involved in positive interactions with the adjacent subsites (P1 and P1′), whereas the inverse is true for P2 Glu residues (Supplementary Table 1).

It should further be noted that both tested motifs (i.e., IFLE and ELLE) were never observed in any of the identified cleavage sites. Unlike the ELLE sequence, processing sites in the Gag and Gag-Pol polyproteins are structurally asymmetric and crystallographic studies with peptides containing these sites showed that the viral protease achieves its substrate specificity by recognition of shapes and volumes rather than specific amino acids (Prabu-Jeyabalan et al., 2002). A previous study with palindromic peptide substrates also showed no obvious symmetrical preference of the HIV-1 protease (Tozser et al., 1997). Moreover, recent studies have shown that substrate dynamics (Ozen et al., 2011) and interactions outside the active site cleft influence substrate cleavage (Lee et al., 2012), phenomena that are ignored by sequence site alignments using tools such as iceLogo. Finally, it should be mentioned that we could only partially confirm cooperativity between the P3 and P2 positions that was observed in the

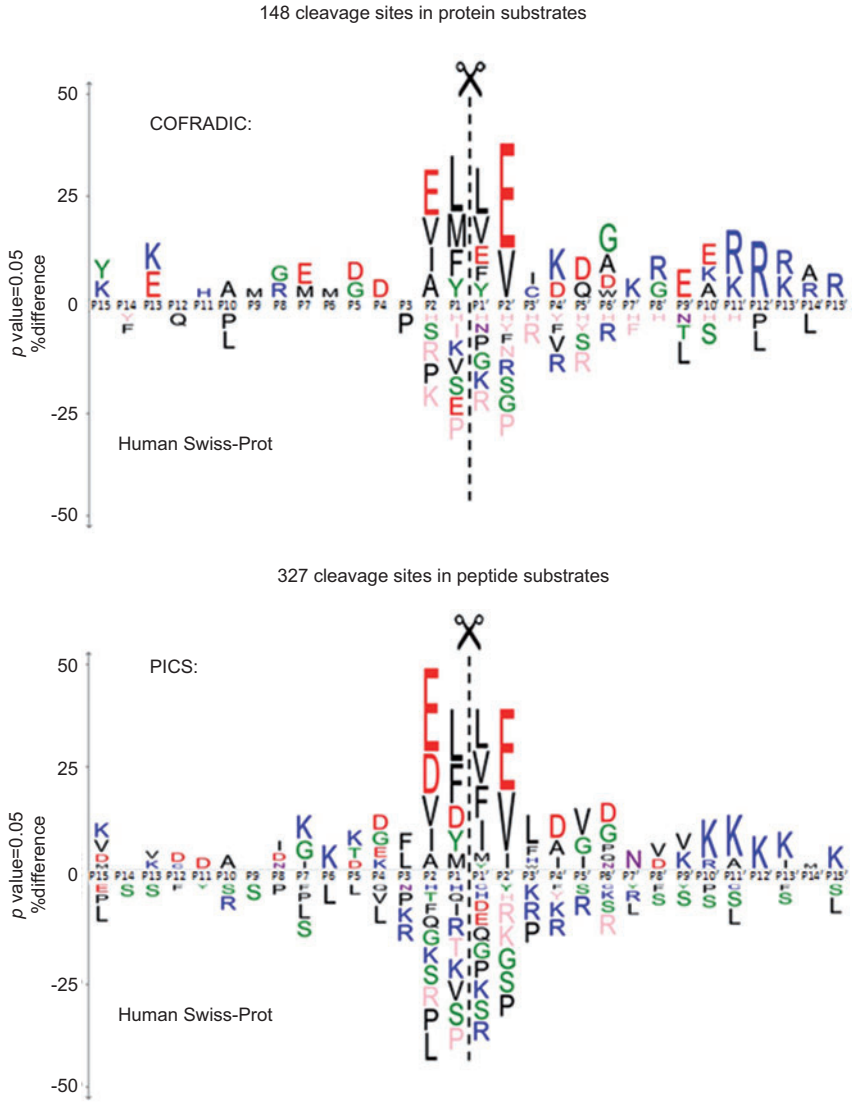

Figure 6 iceLogo (Colaert et al., 2009) alignment and visualisation of all processed sequences identified by N-terminal COFRADIC (n=148) and PICS (n=327) (Schilling and Overall, 2008).
The amino acid frequency at every subsite is compared to their frequency in the human proteins stored in the Swiss-Prot database (negative control). Only residues that are statistically over-represented (upper part of the iceLogo) or under-represented (lower part of the iceLogo) at the 95% confidence level are depicted. Residues that were never observed at specific positions are shown in pink. With both COFRADIC and PICS, the palindromic ELLE-motif was revealed as consensus sequence, reflecting the homodimeric nature of the HIV-1 protease.

aforementioned PICS dataset (Schilling and Overall, 2008). In the latter study, the authors report positive and negative cooperativity between P3 Leu and P1 Ala, and P3 Leu and P1 Phe residues, respectively, the dataset reported here, however, shows a slight preference for both amino acid pairs.

Taken together, we provide a list of human proteins cleaved *in vitro* by the HIV-1 protease. The identification of known host substrates and the overlap with previous studies suggests processing of several proteins during HIV infection with a hint to factors involved in protein translation and RNA splicing. Our failure to delineate an efficient substrate sequence based on the primary amino acid sequence illustrates the complex mechanisms involved in substrate recognition by the HIV-1 protease. Nevertheless, as the protein structures of several substrates in our catalogue have been resolved, our data could be of great value to the study of substrate recognition in detail by structural modeling and docking studies.

Materials and methods

Cell culture and SILAC labeling

Human Jurkat T-cells were purchased from the ATCC (TIB-152, American Type Culture Collection, Manassas, VA, USA) and grown in RPMI-1640 medium supplemented with 10% dialyzed foetal bovine serum (Invitrogen, Carlsbad, CA, USA), 100 units/ml penicillin (Invitrogen) and 100 µg/ml streptomycin (Invitrogen). Cells were cultured in SILAC media containing $^{12}C_6$ (light) or $^{13}C_6$ (heavy)

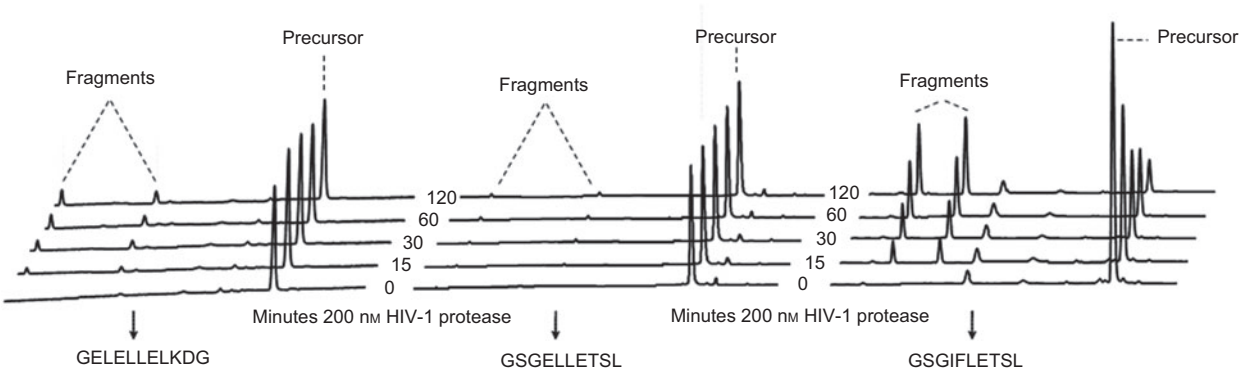

Figure 7 Cleavage of peptides synthesized for consensus motifs.
Cleavage of the ELLE-motif was tested on peptide substrates that were analyzed by RP-HPLC and compared with the IFLE-motif discovered from phage display libraries (Beck et al., 2000). The recorded chromatograms (UV absorbance at 214 nm) are shown, clearly indicating less efficient cleavage of the ELLE-motif. The sequences of the fluorescently quenched peptides were Abz-GELELLELKDGY(NO_2)R, Abz-GSGIFLETSLY(NO_2)R, Abz-GSGELLETSLY(NO_2)R (Abz: amino benzoic acid; Y(NO_2): nitrotyrosine).

L-arginine (Cambridge Isotope Labs, Andover, MA, USA) (Ong et al., 2002) at a reduced concentration (20% of the normal concentration present in RPMI medium). Cells were held at 37°C in 5% CO_2 for at least six population doublings to ensure complete incorporation of the arginine label.

N-terminal sample preparation

Samples were prepared as previously described (Van Damme et al., 2008; Staes et al., 2011) and outlined in Figure 2. To maximize the number of identified cleavage sites, the experiment was repeated three times, including label swapping. Briefly, equal amounts of light- and heavy-labeled Jurkat T-cells were harvested, washed with PBS and lysed by three freeze-thaw cycles in 50 mM 2-(*N*-morpholino)ethanesulfonic acid (MES), 50 mM sodium phosphate, pH 7.4, 150 mM NaCl, 1 mM dithiothreitol (DTT) supplemented with an EDTA-free protease inhibitor tablet (Roche Diagnostics, Mannheim, Germany, 1 tablet per 100 ml). Heavy- and light-labeled lysates were centrifuged for 15 min at 16 000 *g* to remove insoluble components and the protein concentration in the supernatant was measured using the Biorad Protein Assay (Biorad Laboratories, Munich, Germany). Both lysates (each containing 1.5 mg total protein) were acidified with 2 N HCl to pH 5.5, and NaCl was added to a final concentration of 300 mM from a 5 M stock solution. Recombinant HIV-1 protease (ProteinOne, Bethesda, MD) was added to a final concentration of 200 nM to either the light- or heavy-labeled lysate (treated sample), whereas no protease was added to the other lysate (control sample). Both samples were incubated for 75 min at 37°C, after which guanidinium hydrochloride was added to a final concentration of 4 M to denature proteins and inhibit protease activity. The pH was again increased to 7.5 using 2 M NaOH followed by protein reduction and S-alkylation with 1 mM tris(2-carboxyethyl)phosphine (TCEP) and 2 mM iodoacetamide, respectively, and incubation for 60 min at 30°C. Excess reagents were removed by desalting on NAP™-10 desalting columns (Amersham Biosciences, Uppsala, Sweden) in 2 M guanidinium hydrochloride in 50 mM sodium phosphate buffer at pH 8.0 (sample volume was now 1.5 ml). The sample volume of both samples was reduced to 1 ml by vacuum drying and free amines were modified by addition of an N-hydroxysuccinimide ester of trideutero-acetate [synthesized as described before (Staes et al., 2011)] to a final concentration of 7.5 mM, followed by incubation for 2 h at 30°C. In both samples, O-acetylation was reversed by addition of hydroxylamine to a final concentration of 30 mM and incubation for 10 min at room temperature, followed by quenching the excess of the N-hydroxysuccinimide ester by adding glycine to a final concentration of 15 mM and incubation for another 10 min at room temperature. Equal amounts from both samples were mixed and this mixture was immediately desalted on PD-10 desalting columns (Amersham Biosciences) in 20 mM ammonium bicarbonate pH 8.0 (the sample was eluted in 3.5 ml). The sample volume was reduced to 2 ml by vacuum drying, and the protein mixture was boiled for 5 min and placed on ice for 5 min, after which 5 µg of sequencing-grade trypsin (Promega, Madison, WI) was added for overnight digestion at 37°C. Subsequent selection of N-terminal and neo-N-terminal peptides by strong cation exchange chromatography and COFRADIC were performed as described previously (Staes et al., 2008, 2011).

LC-MS/MS analyses, peptide identification and quantification

For processing of all MS data, the ms_lims software platform was used (Helsens et al., 2010). N-terminal peptides were analyzed on a Bruker Esquire HCT ion trap (Bruker Daltonics, Bremen, Germany) that was operated as described previously using identical parameters for MS/MS peak list generation by Bruker's DataAnalysis software (version 3.1) (Ghesquiere et al., 2006). In total, 88,997 MS/MS spectra were recorded and searched with MASCOT using the MASCOT Daemon interface (version 2.1.0) in the Swiss-Prot database (version 51 of UniProtKB/Swiss-Prot protein database containing 14 987 human protein sequences), as well as in N-terminally truncated peptide databases as previously described (Van Damme et al., 2005). For the protein database search, the taxonomy was set to human and the enzyme setting was ArgC/P, allowing for one missed cleavage. Mass tolerance of the precursor ions as well as fragment ions was set to 0.5 Da. The peptide charge parameter was set to 1+, 2+ or 3+ and instrument settings were ESI-TRAP. Variable modifications were oxidation of methionine (sulfoxide), deamidation of asparagine and glutamine, pyroglutamate formation of N-terminal glutamine, pyrocarbamidomethyl formation of N-terminal cysteine and acetylation (both normal and trideutero-acetyl) of the N-terminus. Fixed modifications were carbamidomethylation of cysteines and trideutero-acetylation of lysine side-chains. To enable identification of $^{13}C_6$-arginine containing peptides, searches were repeated with $^{13}C_6$-labeling of arginine as an additional fixed modification. 3581 MS/MS spectra

were identified and peptides were checked individually for the singleton status of potential neo-N-terminal peptides. Note that in all analyses, only peptides that were ranked first and that scored above the threshold score set at 95% confidence were withheld. The FDR was calculated for every experiment as described previously (Kall et al., 2008) and was always found to be lower than 3.7%. Sequences of neo-N-terminal peptides assigned to HIV-1 protease activity (see main text) were re-mapped against a recent version of the Swiss-Prot database (version 57 of UniProtKB/Swiss-Prot protein database containing 20,330 human protein sequences) to generate Table 1. For every neo-N-terminal peptide, the highest scoring MS/MS spectrum (n=148) was stored in the PRIDE database (experiment accession number 16377) using the PRIDE converter (Barsnes et al., 2009).

Immunoblotting

Jurkat T-cell lysates were prepared and treated with recombinant HIV-1 protease (ProteinOne) as described previously with or without addition of Ac-pepstatin (Merck, Darmstadt, Germany), ritonavir (obtained through the NIH AIDS Research and Reference Reagent Program) and/or zVADfmk (Promega) at the indicated concentrations. Criterion XT gels and buffers (Bio-Rad) were used for SDS-PAGE. Equal amounts of total protein from the incubated cell lysates were loaded on individual gel lanes (20 to 30 µg as measured with the Bio-Rad Protein Assay). Proteins were separated on 12% or 4–12% Bis-Tris gels in MOPS or MES buffer at 150 V and transferred overnight in 50 mM Tris and 50 mM boric acid at 10 V onto a PVDF membrane. Protein detection was performed using the LI-COR Odyssey Infrared Imaging System according to the manufacturer's instruction (LI-COR Biosciences, Lincoln, NE). Primary antibodies against vimentin (Abcam, ab7783) and DDX5 (Abcam, ab21696) were diluted 500× and 1000× for protein detection, respectively.

Peptide cleavage assays

The following peptides were synthesized using Fmoc (N-(9-fluorenyl) methoxycarbonyl) chemistry on an Applied Biosystems 443A Peptide Synthesizer: Abz-GELELLELKDGY(NO2)R (COFRADIC consensus), Abz-GSGIFLETSLY(NO_2)R (Beck et al., 2000), Abz-GSGELLETSLY(NO_2)R (COFRADIC/Beck hybrid) (Abz: amino benzoic acid; Y(NO_2): nitrotyrosine). Peptides were dissolved in assay buffer (50 mM MES pH 5.5, 300 mM NaCl, 1 mM DTT) at a concentration of 100 µM and incubated for the time indicated with 200 nM rHIV-1 protease (ProteinOne, Bethesda, MD) at 37°C. Reactions were quenched in RP-HPLC solvent A (0.1% trifluoroacetic acid, 2% acetonitrile in water) and monitored by recording the UV absorbance signal (214 nm) of the precursor peptide and its fragments during RP-HPLC separation.

Acknowledgements

FI and KKA are postdoctoral fellows and BV is a senior clinical investigator of the Research Foundation-Flanders (FWO-Vlaanderen). KG acknowledges support by research grants from the Fund for Scientific Research-Flanders (Belgium) (project number G.0048.08), the Concerted Research Actions (project BOF07/GOA/012) from Ghent University, the Interuniversity Attraction Poles (IUAP06) of the Belgian Science Policy and the UGent Multidisciplinary Research Partnership N2N (from Nucleotides to Networks). BV acknowledges support by research grants from the Fund for Scientific Research-Flanders (Belgium) (project number G.0091.10), the Concerted Research Actions (project BOF11/GOA/013) from Ghent University, the Interuniversity Attraction Poles (IUAP06) project HIV-STOP of Belgian Science Policy, the SBO CellCoVir grant from the agency for Innovation by Science and Technology (IWT) Flanders, Belgium and European Union FP7 Health-2007-2.3.2-1 Collaborative Project iNEF (201412).

References

Barsnes, H., Vizcaino, J.A., Eidhammer, I., and Martens, L. (2009). PRIDE Converter: making proteomics data-sharing easy. Nat. Biotechnol. *27*, 598–599.

Beck, Z.Q., Hervio, L., Dawson, P.E., Elder, J.H., and Madison, E.L. (2000). Identification of efficiently cleaved substrates for HIV-1 protease using a phage display library and use in inhibitor development. Virology *274*, 391–401.

Beck, Z.Q., Morris, G.M., and Elder, J.H. (2002). Defining HIV-1 protease substrate selectivity. Curr. Drug Targets Infect. Disord. *2*, 37–50.

Becker-Pauly, C., Barre, O., Schilling, O., Auf dem Keller, U., Ohler, A., Broder, C., Schutte, A., Kappelhoff, R., Stocker, W., and Overall, C.M. (2011). Proteomic analyses reveal an acidic prime side specificity for the astacin metalloprotease family reflected by physiological substrates. Mol. Cell Proteomics *10*, M111.009233.

Castello, A., Franco, D., Moral-Lopez, P., Berlanga, J.J., Alvarez, E., Wimmer, E., and Carrasco, L. (2009). HIV- 1 protease inhibits Cap- and poly(A)-dependent translation upon eIF4GI and PABP cleavage. PLoS One *4*, e7997.

Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., and Gevaert, K. (2009). Improved visualization of protein consensus sequences by iceLogo. Nat. Methods *6*, 786–787.

Demon, D., Van Damme, P., Vanden Berghe, T., Deceuninck, A., Van Durme, J., Verspurten, J., Helsens, K., Impens, F., Wejda, M., Schymkowitz, J., et al. (2009). Proteome-wide substrate analysis indicates substrate exclusion as a mechanism to generate caspase-7 versus caspase-3 specificity. Mol. Cell Proteomics *8*, 2700–2714.

Drag, M. and Salvesen, G.S. (2010). Emerging principles in protease-based drug discovery. Nat. Rev. Drug Discov. *9*, 690–701.

Freed, E.O. (2001). HIV-1 replication. Somat. Cell Mol. Genet. *26*, 13–33.

Fu, W., Sanders-Beer, B.E., Katz, K.S., Maglott, D.R., Pruitt, K.D., and Ptak, R.G. (2009). Human immunodeficiency virus type 1, human protein interaction database at NCBI. Nucleic Acids Res. *37*, D417–422.

Fuller-Pace, F.V. and Moore, H.C. (2011). RNA helicases p68 and p72: multifunctional proteins with important implications for cancer development. Future Oncol. *7*, 239–251.

Gevaert, K., Goethals, M., Martens, L., Van Damme, J., Staes, A., Thomas, G.R., and Vandekerckhove, J. (2003). Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. Nat. Biotechnol. *21*, 566–569.

Ghesquiere, B., Goethals, M., Van Damme, J., Staes, A., Timmerman, E., Vandekerckhove, J., and Gevaert, K. (2006). Improved tandem mass spectrometric characterization of 3-nitrotyrosine sites in peptides. Rapid Commun. Mass Spectrom. *20*, 2885–2893.

Helsens, K., Colaert, N., Barsnes, H., Muth, T., Flikka, K., Staes, A., Timmerman, E., Wortelkamp, S., Sickmann, A., Vandekerckhove, J., et al. (2010). ms_lims, a simple yet powerful open source laboratory information management system for MS-driven proteomics. Proteomics *10*, 1261–1264.

Impens, F., Colaert, N., Helsens, K., Ghesquiere, B., Timmerman, E., De Bock, P.J., Chain, B.M., Vandekerckhove, J., and Gevaert, K. (2010a). A quantitative proteomics design for systematic identification of protease cleavage events. Mol. Cell Proteomics *9*, 2327–2333.

Impens, F., Colaert, N., Helsens, K., Plasman, K., Van Damme, P., Vandekerckhove, J., and Gevaert, K. (2010b). MS-driven protease substrate degradomics. Proteomics *10*, 1284–1296.

Jacob, R.J. and Cramer, R. (2006). PIGOK: linking protein identity to gene ontology and function. J. Proteome Res. *5*, 3429–3432.

Jager, S., Cimermancic, P., Gulbahce, N., Johnson, J.R., McGovern, K.E., Clarke, S.C., Shales, M., Mercenne, G., Pache, L., Li, K., et al. (2012). Global landscape of HIV-human protein complexes. Nature *481*, 365–370.

Jensen, L.J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M., et al. (2009). STRING 8–a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. *37*, D412–416.

Kall, L., Storey, J.D., MacCoss, M.J., and Noble, W.S. (2008). Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. J. Proteome Res. *7*, 29–34.

Kohl, N.E., Emini, E.A., Schleif, W.A., Davis, L.J., Heimbach, J.C., Dixon, R.A., Scolnick, E.M., and Sigal, I.S. (1988). Active human immunodeficiency virus protease is required for viral infectivity. Proc. Natl. Acad. Sci. USA *85*, 4686–4690.

Lee, S.K., Potempa, M., Kolli, M., Ozen, A., Schiffer, C., and Swanstrom, R. (2012). Context surrounding processing sites is crucial in determining cleavage rate of a subset of the processing sites in the HIV-1 Gag and Gag-Pro-Pol precursors by the viral protease. J. Biol. Chem. *287*, 13279–13290.

Liu, H., Sadygov, R.G., and Yates, J.R. 3rd. (2004). A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal. Chem. *76*, 4193–4201.

Lopez-Otin, C. and Overall, C.M. (2002). Protease degradomics: a new challenge for proteomics. Nat. Rev. Mol. Cell Biol. *3*, 509–519.

Naji, S., Ambrus, G., Cimermancic, P., Reyes, J.R., Johnson, J.R., Filbrandt, R., Huber, M.D., Vesely, P., Krogan, N.J., Yates, J.R., 3rd, et al. (2012). Host cell interactome of HIV-1 Rev includes RNA helicases involved in multiple facets of virus production. Mol. Cell Proteomics *11*, M111.015313.

Ng, N.M., Pike, R.N., and Boyd, S.E. (2009). Subsite cooperativity in protease specificity. Biol. Chem. *390*, 401–407.

Nie, Z., Phenix, B.N., Lum, J.J., Alam, A., Lynch, D.H., Beckett, B., Krammer, P.H., Sekaly, R.P., and Badley, A.D. (2002). HIV-1 protease processes procaspase 8 to cause mitochondrial release of cytochrome c, caspase cleavage and nuclear fragmentation. Cell Death Differ. *9*, 1172–1184.

Nie, Z., Bren, G.D., Vlahakis, S.R., Schimnich, A.A., Brenchley, J.M., Trushin, S.A., Warren, S., Schnepple, D.J., Kovacs, C.M., Loutfy, M.R., et al. (2007). Human immunodeficiency virus type 1 protease cleaves procaspase 8 in vivo. J. Virol. *81*, 6947–6956.

Nie, Z., Bren, G.D., Rizza, S.A., and Badley, A.D. (2008). HIV protease cleavage of procaspase 8 is necessary for death of HIV-infected cells. Open Virol. J. *2*, 1–7.

Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol. Cell Proteomics *1*, 376–386.

Ozen, A., Haliloglu, T., and Schiffer, C.A. (2011). Dynamics of preferential substrate recognition in HIV-1 protease: redefining the substrate envelope. J. Mol. Biol. *410*, 726–744.

Prabu-Jeyabalan, M., Nalivaika, E., and Schiffer, C.A. (2002). Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes. Structure *10*, 369–381.

Ridky, T.W., Cameron, C.E., Cameron, J., Leis, J., Copeland, T., Wlodawer, A., Weber, I.T. and Harrison, R.W. (1996). Human immunodeficiency virus, type 1 protease substrate specificity is limited by interactions between substrate amino acids bound in adjacent enzyme subsites. J. Biol. Chem. *271*, 4709–4717.

Schilling, O. and Overall, C.M. (2008). Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. Nat. Biotechnol. *26*, 685–694.

Shoeman, R.L., Honer, B., Stoller, T.J., Kesselmeier, C., Miedel, M.C., Traub, P., and Graves, M.C. (1990). Human immunodeficiency virus type 1 protease cleaves the intermediate filament proteins vimentin, desmin, and glial fibrillary acidic protein. Proc. Natl. Acad. Sci. USA *87*, 6336–6340.

Staes, A., Van Damme, P., Helsens, K., Demol, H., Vandekerckhove, J., and Gevaert, K. (2008). Improved recovery of proteome-informative, protein N-terminal peptides by combined fractional diagonal chromatography (COFRADIC). Proteomics *8*, 1362–1370.

Staes, A., Impens, F., Van Damme, P., Ruttens, B., Goethals, M., Demol, H., Timmerman, E., Vandekerckhove, J., and Gevaert, K. (2011). Selecting protein N-terminal peptides by combined fractional diagonal chromatography. Nat. Protoc. *6*, 1130–1141.

Tomasselli, A.G., Hui, J.O., Adams, L., Chosay, J., Lowery, D., Greenberg, B., Yem, A., Deibel, M.R., Zurcher-Neely, H., and Heinrikson, R.L. (1991). Actin, troponin C, Alzheimer amyloid precursor protein and pro-interleukin 1 beta as substrates of the protease from human immunodeficiency virus. J. Biol. Chem. *266*, 14548–14553.

Tozser, J. and Oroszlan, S. (2003). Proteolytic events of HIV-1 replication as targets for therapeutic intervention. Curr. Pharm. Des. *9*, 1803–1815.

Tozser, J., Bagossi, P., Weber, I.T., Louis, J.M., Copeland, T.D., and Oroszlan, S. (1997). Studies on the symmetry and sequence context dependence of the HIV-1 proteinase specificity. J. Biol. Chem. *272*, 16807–16814.

Van Damme, P., Martens, L., Van Damme, J., Hugelier, K., Staes, A., Vandekerckhove, J., and Gevaert, K. (2005). Caspase-specific and nonspecific in vivo protein processing during Fas-induced apoptosis. Nat. Methods *2*, 771–777.

Van Damme, P., Impens, F., Vandekerckhove, J., and Gevaert, K. (2008). Protein processing characterized by a gel-free proteomics approach. Methods Mol. Biol. *484*, 245–262.

Van Damme, P., Maurer-Stroh, S., Plasman, K., Van Durme, J., Colaert, N., Timmerman, E., De Bock, P.J., Goethals, M., Rousseau, F., Schymkowitz, J., et al. (2009). Analysis of protein processing by N-terminal proteomics reveals novel species-specific substrate determinants of granzyme B orthologs. Mol. Cell Proteomics *8*, 258–272.

Wondrak, E.M., Louis, J.M., and Oroszlan, S. (1991). The effect of salt on the Michaelis Menten constant of the HIV-1 protease correlates with the Hofmeister series. FEBS Lett. *280*, 344–346.

Received March 29, 2012; accepted June 19, 2012