Local and spatial factors determining HIV-1 protease substrate recognition

Stephane HAZEBROUCK, Valerie MACHTELINCKX-DELMAS, Jean-Jacques KUPIEC and Pierre SONIGO[1]

Genetique des Virus, ICGM-CNRS UPR 415, 22 rue Mechain, 75014 Paris, France

Insertional mutagenesis of the *Escherichia coli* thymidylate synthase (TS) was used to address substrate recognition of HIV-1 protease in a well characterized structural context. By modifying the TS conformation while maintaining its enzymic activity, we investigated the influence of protein folding on protease–substrate recognition. A slight destabilization of the TS structure permitted the cleavage of a target site, which was resistant in the native TS. This result supports a dynamic interpretation of HIV-1 protease specificity. Exposure time of the potential cleavage site, which depends on the stability of the global conformation, must be compatible with the cleavage kinetics, which are determined by the local sequence. Cleavage specificity has been described as the consequence of cumulative interactions, globally favourable, between at least six amino acids around the cleavage site. To investigate influence of local sequence, we introduced insertions of variable lengths in two exposed loops of the TS. In both environments, insertion of only two amino acids could determine specific cleavage. We then inserted libraries of dipeptides naturally cleaved by the HIV-1 protease in order to assess the limitations of established classifications of substrates in different conformational contexts.

Key words: cleavage specificity, *Escherichia coli*, thymidylate synthase.

INTRODUCTION

The HIV aspartyl protease processes the gag and gag-pol polypeptide precursors into structural proteins and enzymes (see Figure 1A) [1,2]. This processing step is required for acquisition of infectivity by the newly assembled virion, and thus represents a critical target for chemotherapy against AIDS. Selective inhibitors of the HIV-1 protease have been synthesized [3] and, in association with inhibitors of the reverse transcriptase, have been shown to reduce viral replication efficiently in human patients. Unfortunately, utilization of protease inhibitors can lead to the selection of virus mutants, whose protease displayed a lower affinity for drugs while maintaining cleavage rates necessary for viral replication. Mutations causing resistance were also described in cleavage sites of gag and gag-pol precursors, resulting in more efficiently cleaved substrates [4,5].

The protease appears highly specific in that it catalyses the cleavage of a limited number of sites in the gag and gag-pol precursors (Figure 1A). On the other hand, the HIV-1 protease can hydrolyse a large variety of peptides and fails to show a clear preference for particular amino acid residues at any position flanking the scissile bond. Substrates are, nevertheless, divided into two broad classes, those having a proline residue at P1′ (according to the nomenclature of Schecter and Berger [6]) and those having two hydrophobic amino acids at P1 and P1′ [7]. Depending on the type of cleavage site, the protease has been described to have different preferences at the P2 and P2′ positions. For cleavage sites with a Pro at P1′, the presence of Asn at the P2 position and β-branched Val or Ile at the P2′ position were found to favour the hydrolysis. β-branched Val at the P2 position and Glu or Gln at P2′ were preferred for the processing of bonds between two hydrophobic residues. Enzymic characterization of the HIV-1 protease has mainly been carried out with synthetic peptides and their derivatives systematically altered in their chemical structure [8]. Thus the dependence of the protease specificity on the conformational context remains largely unexplored, though this aspect is critical to the understanding of the kinetically driven mechanism of the particle maturation.

We wished to develop an experimental approach which allows the sequence and spatial structure of the substrate to be varied in a controlled manner. In a previous work, we have performed extensive insertional mutagenesis of the *Escherichia coli* thymidylate synthase (TS) gene [9], and found several sites which could expose exogenous sequences at the surface of the protein. By inserting natural substrate sequences, we observed the *in vivo* cleavage of the engineered TS by HIV-1 protease.

In the present work, we analysed the influence of the structural context on substrate recognition. First, we confirmed the importance of the structural context for the proteolytic specificity by identifying a site whose processing is dependent on the TS conformation. Then, we showed that insertion of only two amino acids, Phe-Pro, within different surface loops induced specific cleavage by the HIV-1 protease. Finally, we tested the cleavage of sites varying in size, sequence or conformational context. These results revealed the limitations of the previous classification of substrates and define better the relationships between local and spatial factors affecting cleavage efficiency.

EXPERIMENTAL

Plasmids

The 1.3 kb *E. coli thyA* fragment PTZ*thyA* was first subcloned from pBTAH2 [10] into the phagemid pTZ18R (Amersham Pharmacia Biotech) using *Hin*dIII sites. The *Hin*dIII sites were replaced by *Xho*I sites by site-directed mutagenesis. The 2.9 kb HIV-1 *pol* gene fragment pSU*prt*+, corresponding to the protease and reverse transcriptase domains, was subcloned from pBRT1*prt*+[11] into pSU19 [12], using the *Eco*RI and *Sal*I sites. The 1.5 kb HIV-1 *pol* fragment pSU*prt*−, corresponding solely to the reverse transcriptase domain, was also subcloned from pBRT3*prt*− [11] into pSU19 using the *Eco*RI and *Sal*I sites. pBRT1*prt*+ and pBRT3*prt*− clones were provided by the

Abbreviations used: CA, capsid protein; IN, integrase; IS, insertion site; MA, matrix protein; $p6^{pol}$, preprotease; RT, reverse transcriptase; TS; thymidylate synthase..

[1] To whom correspondence should be addressed (e-mail sonigo@cochin.inserm.fr).

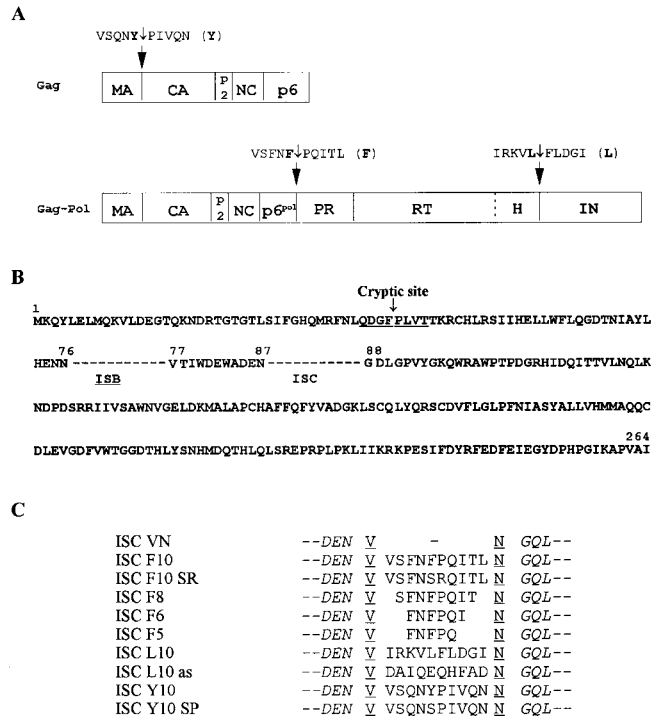

Figure 1 Conversion of the *E. coli* TS into HIV-1 protease substrate

(**A**) Structure of HIV-1 gag and gag-pol polyproteins. The matrix, capsid, nucleocapsid, 2-kDa, 6-kDa, preprotease, protease, reverse transcriptase, ribonuclease H and integrase proteins are abbreviated to MA, CA, NC, p2, p6, $p6^{pol}$, PR, RT, H and IN respectively. The abbreviation Y refers to the MA–CA junction, F to the $p6^{pol}$–protease junction and L to the RT–IN junction of the gag and gag-pol precursors. These sites of proteolytic maturation were inserted in site C of the *E. coli* TS (see **B** and **C**). (**B**) Amino acid sequence of the *E. coli* TS. Sites B and C, mutagenized by insertion, are designated as ISB and ISC respectively. The cryptic site, located 45 amino acids upstream of the flanking residue of ISC, is underlined. (**C**) Amino acid sequence of exogenous insertions in site C. Nomenclature of the mutants indicates the insertion site (ISC), the origin of the sequence from the gag-pol precursor [F, L, or Y; see (**A**)], the length of the inserted sequence as a number of amino acids (10, 8, 6 or 5), and additional mutagenesis (for example SR indicates the modified residues at the cleaved site, by means of antisense insertion). Residues that belong to the primary structure of the wild-type *E. coli* TS are *italicized*. Val and Asn residues, which are underlined, are encoded by the *Hpa*I restriction site introduced during the mutagenesis (as described in the Experimental section).

National Institutes of Health (AIDS Research and Reference Reagent Program).

Insertion sequences and nomenclature of *E. coli* thyA mutants

Mutagenesis was performed according to Kunkel et al. [13]. Annealing and extension reactions were performed on the single stranded form of the pTZ*thyA* phagemid DNA. Oligonucleotides were chosen to introduce a unique blunt-end *Hpa*I restriction site to facilitate subsequent insertions at site B and C (Figure 1B). For the 5–10 amino acid insertions, complementary oligonucleotides coding for the HIV-1 protease targets were synthesized, hybridized and cloned in both orientations (designated as sense and antisense) within the *Hpa*I restriction site. The ligation mixtures were digested with the *Hpa*I enzyme which no longer had a recognition sequence within the desired clones. The ligation mixtures were then used to transform competent *E. coli* DH5α cells. For smaller insertions, oligonucleotides coding for the mutations were annealed on the single-stranded form of the modified pTZ*thyA*. The Val and Asn encoded by the *Hpa*I restriction site were either conserved or deleted, according to the sequence context required. Mutants were identified by nucleotide sequencing using a commercial kit (Amersham International). The first three letters designate the site of insertion (i.e. ISC means insertion site C), the following characters designate either the sequence inserted and its length (i.e. ISC F10 means the ten amino acid in length F sequence inserted at site C) or simply the amino acids inserted (Figure 1C).

E. coli strains

Expression of the mutants was performed in the β-1308: Δ-*thyA*::Em derivative of the wild-type MG 1655 *E. coli* K12 ($F^-\lambda^-$). Cells were grown on Luria–Bertani broth complemented with thymidine (100 μg/ml), ampicillin (50 μg/ml) and chloramphenicol (25 μg/ml) at 30 °C.

Antibodies and protein extracts for immunoblotting experiments

Pellets corresponding to 1.5 ml cultures of β-1308 bacteria were resuspended in 150 μl of buffer [0.5 mM 4-(2-aminoethyl)-benzenesulphonyl fluoride, 4 mM 2-mercaptoethanol and 50 mM Tris/HCl, pH 7.5], frozen in liquid nitrogen and thawed three times. Cellular debris was removed by centrifugation (30 min at 12 000 ***g***), and soluble proteins were analysed by SDS/PAGE and Western blotting. Polyclonal antibodies against *E. coli* TS were prepared by injecting a rabbit with a synthetic peptide corresponding to the 24 amino acids of the C-terminus of TS. Total protein concentration was determined in the extracts by the Biuret method (BCA kit, Pierce).

Immunopurification of cleavage products and microsequencing

Soluble proteins corresponding to 500 ml of culture of transformed β-1308 bacteria were resolved by SDS/PAGE. Proteins were revealed by Coomassie Blue staining and bands containing the cleavage products, around 27 kDa and 21 kDa, were excised from the gel. The bands were incubated in 0.1 % SDS and 100 mM Tris/HCl, pH 8.0, overnight at 37 °C. Gel was removed by filtering through a 0.45 μm membrane (Millipore) and the supernatant was concentrated on centriprep-10 (Amicon). SDS and Coomassie blue were eliminated by precipitation with an excess of 10 % cetyltrimethylammonium bromide. Samples were centrifuged (30 min at 150 000 ***g***). The supernatants were diluted in PBS and loaded on the top of a 1-ml affinity column. The affinity column was prepared by covalently coupling the purified polyclonal antibody directed against the C-terminal region of the *E. coli* TS to Protein A–Sepharose, using the Immunopure IgG orientation kit (Pierce). Unbound material was removed by washing with 30 ml of PBS and specific binding was disrupted by elution with 0.1 M glycine, pH 2.5. The eluate was neutralized by addition of 1 M Tris, pH 9, and concentrated on centricon-10 (Amicon). The proteins were separated by SDS/PAGE and electroblotted on to PVDF membrane. After Coomassie Blue staining, discrete protein bands were closely excised from the membrane. The N-terminal amino acid sequences of the samples was determined on a 473A sequencer (Applied Biosystem).

RESULTS

Conversion of the *E. coli* TS into HIV-1 protease substrate

E. coli TS was shown previously to be permissive to exogenous insertions in two exposed loops designated as B and C [9]. By introducing the preprotease ($p6^{pol}$)–protease ('PR') junction (**F**), SFN**F**↓PQIT, the matrix protein (MA)–capsid protein (CA) junction (**Y**), SQN**Y**↓PIVQ, or the reverse transcriptase (RT)–

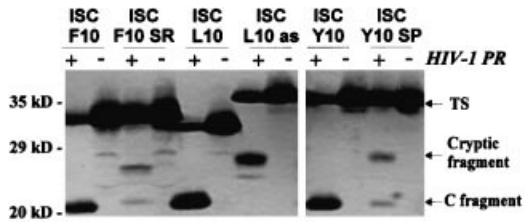

Figure 2 ***In vivo*** **cleavage of target sequences inserted within TS**

Analysis by immunoblotting using TS-specific antibodies. β-1308 cells were transformed by two constructs, a pTZ*ThyA* mutant and a pSU*prt* plasmid expressing either an active (+) or a defective (−) HIV-1 protease protein. The HIV-1 protease was also able to recognize and cleave the altered target sequences (ISC F10 SR and ISC Y10 SP), which have been described previously as non-cleavable. Abbreviation: kD, kDa (in this and subsequent Figures).

integrase (IN) junction (**L**), RKV**L**↓FLDG, we converted the engineered TS into a substrate for the HIV-1 protease. The cleavage point is indicated in substrate sequence by an arrow (↓). The *in vivo* processing was revealed by the appearance of a cleavage product, which migrated at the predicted size of 21 kDa (Figure 2). The C-terminal fragment of the TS ISC F10 mutant was purified using an immuno-affinity column, and its amino acid sequencing confirmed that the cleavage occurred at the Phe–Pro bond of the target sequence SFNF↓PQIT.

Cleavage of two altered substrates

Two cleavage sites, altered at P1 and P1′, were introduced at site C. The selected mutations, Ser-Arg in place of Phe-Pro in the $p6^{pol}$–protease sequence context SFNSRPQIT [14] and Ser-Pro in place of Tyr-Pro in the MA–CA sequence context SQNSPIVQ [15], have been described previously as blocking proteolysis, even in the presence of an excess of purified HIV-1 protease. In particular, substitution of Ser for Tyr in the MA–CA junction was shown to inhibit cleavage of the synthetic peptide SQNSPIVQ, as well as proteolytic maturation of the gag precursor [15]. By contrast, in the structural context of the TS, the two altered target sequences were cleaved, albeit at a very low

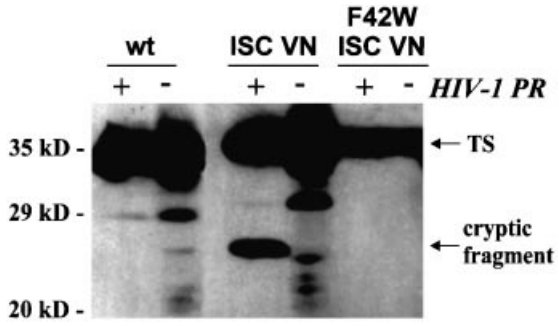

Figure 3 Insertion of only two amino acids in site C resulted in the unmasking of a cleavage site located 45 amino acids upstream of site C

Analysis by immunoblotting using TS-specific antibodies. β-1308 cells were transformed by two constructs, a pTZ*ThyA* mutant and a pSU*prt* plasmid expressing either an active (+) or a defective (−) HIV-1 protease protein. The wild-type TS (wt) is resistant to the action of the HIV-1 protease. Insertion in site C of two amino acids, Val and Asn, induced the cleavage of the ISC VN mutant at a cryptic site, identified by protein sequencing as QDGF↓PLVT. Substitution of the native Phe^{42} flanking the scissile bond by Trp blocked the processing of the resulting mutant F42W ISC VN.

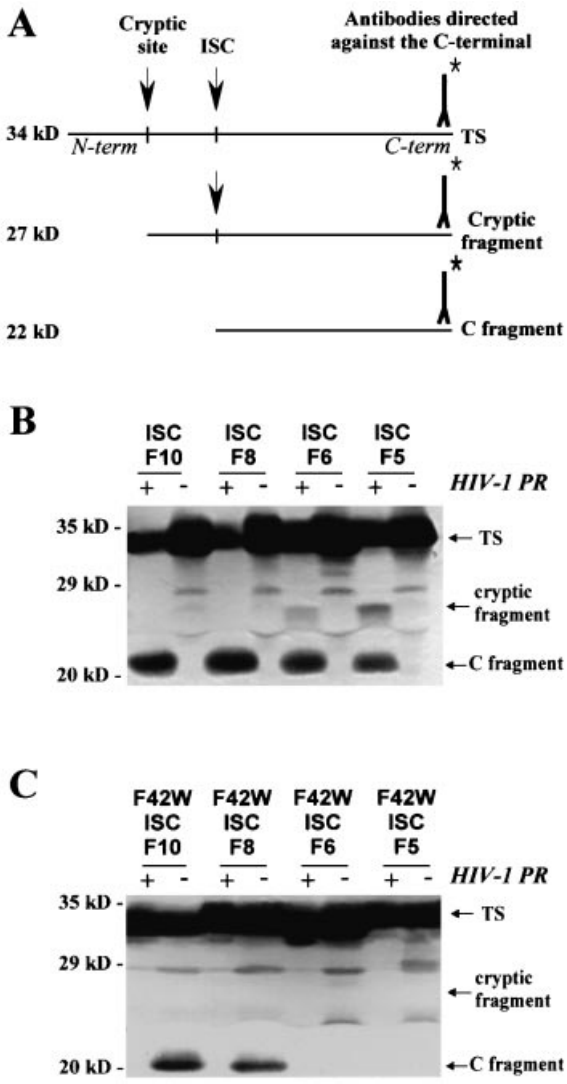

Figure 4 Influence of the length of the inserted target sequence on its recognition by the HIV-1 protease

(**A**) Representation of the proteolytic degradation of the engineered TS and its visualization by immunoblotting using antibodies directed against the C-terminus of TS. The cryptic fragment can be detected by immunoblotting only if cleavage at site C is partial. (**B** and **C**) Analysis by immunoblotting using TS-specific antibodies. β-1308 cells were transformed by two constructs, a pTZ*ThyA* mutant and a pSU*prt* plasmid expressing either an active (+) or a defective (−) HIV-1 protease protein. The HIV-1 protease could cleave insertions of decreasing length from ten to five residues in the structural context of the wild-type TS (**B**), while the protease was not able to cleave the six and five amino acid long targets in the structural context of the F42W TS mutant (**C**). Abbreviations: *C-term*, C-terminus; *N-term*, N-terminus.

efficiency, by the HIV-1 protease, as revealed by the presence of cleavage products (Figure 2). In this regard, the surface loop C certainly provided a suitable conformation for hydrolysis of target sequences and therefore, a sensitive method for revealing cleavage events. Insertion of unrelated peptide did not result in cleavage, as shown with the inserted sequence DAIQEQHFAD.

Identification of a conformation-dependent cleavage site

Wild-type TS is resistant to the action of the viral protease (Figure 3) and only one cleavage product at 21 kDa was detected

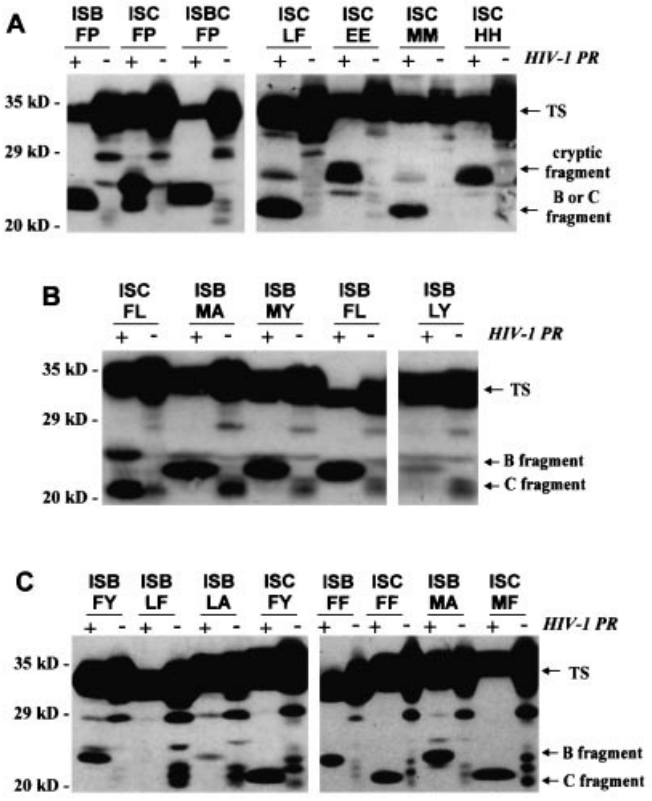

Figure 5 ***In vivo*** **cleavage of dipeptides inserted in sites B or C**

Analysis by immunoblotting using TS-specific antibodies. β-1308 cells were transformed with two constructs, a pTZ*ThyA* mutant and a pSU*prt* plasmid expressing either an active (+) or a defective (−) HIV-1 protease protein.

with TS mutants exposing HIV-1 protease substrates (Figure 2). However, after insertion of an unrelated peptide at site C, another cleavage product migrated at 27 kDa (Figure 2). This fragment was purified by affinity chromatography and its N-terminus was sequenced, yielding the amino acid sequence XXVTTKR, localizing the cleavage site to Phe[42] (Figure 1B). The substrate sequence QDGF↓PLVT was consistent with the known property of the HIV-1 protease to cleave the Phe–Pro bond. We observed that insertion of as few as two amino acids, Val-Asn, at site C was sufficient to render this cleavage site, designated as cryptic, susceptible to the HIV-1 protease (Figure 3). Substitution of Phe[42] by Trp, another aromatic residue that has never been observed in the scissile bond of proteolytic sites, blocked the processing of the cryptic site (Figure 3).

Determination of the minimal length of insertion that could induce cleavage

Studies with synthetic peptides showed that the 6–8 residues occupying P4-P4′ were necessary to give specific and efficient cleavage [8,16,17]. On the other hand, some results support the strong interaction of chemically modified tetra- and even tri-peptides with the active site of the HIV-1 protease [3]. We tested several insertions of decreasing length from ten to five residues to determine the critical core which was sufficient for cleavage in the heterologous context of the engineered *E. coli* TS (Figure 4B). At five amino acids, an efficient cleavage occurred, although with a reduced efficiency for the smaller insertions, as revealed by visualization of the cryptic fragment. Indeed, only TS C-terminal fragments can be detected with our anti-peptide antibody. Therefore, as site C is located 45 amino acids downstream of the cryptic site, the cryptic fragment could be revealed only if the cleavage at site C occurred with a reduced efficiency, as happened with insertion of six and five residues (Figure 4).

We engineered the same constructions in the F42W TS mutant in which the cryptic cleavage site was blocked. HIV-1 protease was still able to cleave the ten and eight amino acid long targets (Figure 4C), while the smaller insertions of six and five amino acids were no longer processed. Considering that the target sequences were strictly identical in the wild-type TS and F42W mutant, conformation of the inserted sites should be responsible for the observed dissimilarities in processing efficiency. These differences were observed under physiological conditions (i.e. at neutral pH) and without previous denaturation steps such as treatment with SDS [18] or acidification [19].

Dipeptide insertions induced specific cleavage in two different structural contexts

We pursued the investigation of the minimal insertion that allowed proteolytic processing of the TS. We introduced only two amino acids, Phe-Pro, at site B, ENNF↓PVTI, and at site C, DENF↓PGDL. The presence of the flanking Asn at P2 in both sites B and C and the β-branched Val at P2′ in site B, while fortuitous, is known to define an appropriate sequence context

Table 1 **Cleavage by the HIV-1 protease of dipeptides inserted in sites B and C**

HIV-1 protease substrates*		Dipeptides inserted in site B		Dipeptides inserted in site C	
Protein	Cleavage site	Mutant	Cleavage site	Mutant	Cleavage site
gag	ATI **M** ↓**M** QRG		nd	ISC MM	ENV **M**↓**M** NGD
gag†	ARL **M**↓**M** EAL	ISB MA	ENN **M**↓**A** VTI		nd
actin	TQI **M** ↓**F** ETF		nd	ISC MF	ENV **M**↓**F** NGD
	M↓**Y**	ISB MY	ENN **M**↓**Y** VTI		nd
gag	PGN **F**↓**L**QSR	ISB FL	ENN **F** ↓**L** VTI	ISC FL	ENV **F** ↓**L** NGD
prolL1b	DDL **F** ↓**F** EAD	ISB FF	ENN **F** ↓**F** VTI	ISC FF	ENV **F** ↓**F** NGD
pol	AET **F** ↓**Y** VDG	ISB FY	ENN **F** ↓**Y** VTI	ISC FY	ENV **F** ↓**Y** NGD
vimentin	SRS **L** ↓**Y** ASS	ISB LY	ENN **L** ↓**Y** VTI		nd
pol	RKI **L** ↓**F** LDG	ISB LF	ENN **L** ↓**F** VTI	ISC LF	ENV **L** ↓**F** NGD
gag	ARV **L** ↓**A** EAM	ISB LA	ENN **L** ↓**A** VTI		nd

↓, Cleaved bond; *, from Poorman et al. [21]; †, hydrolysed by the HIV-2 protease; nd, not determined; = , uncleaved bond.

for the cleavage of peptide bonds containing a Pro at the P1′ position.

The engineered surface loops resulted in specific cleavage at the two sites with relatively high efficiency, especially at site B (Figure 5A). Microsequencing of the C-terminal fragment confirmed that cleavage of the TS ISB FP actually occurred between the inserted Phe and Pro residues. Cleavage at site C was less efficient as revealed by the presence of the cryptic fragment in addition to the C fragment (see Figure 4). Proteolysis was more efficient following insertion of the dipeptide Phe↓Pro at both sites B and C (ISBC FP), than after insertion of Phe↓Pro in only one of the two sites (Figure 5A).

Finally, we inserted dipeptides cleaved in gag and gag-pol precursors or in non-viral substrates (Table 1). In order to study the context dependence of HIV-1 protease specificity, we introduced Val and Asn at P2 and P2′ respectively in site C, while conserving the naturally occurring Asn and Val in site B. All combinations containing a Leu at the P1 position, LF, LY and LA, were hydrolysed poorly or not at all in the context of site B (Figures 5B and 5C). In contrast, LF was cleaved efficiently in site C (Figure 5A). Otherwise, we observed the cleavage of all dipeptides tested in the heterologous context of the TS loops B and C (Table 1 and Figure 5).

DISCUSSION

The exploration of substrate specificity of HIV-1 protease has principally been carried out using synthetic peptides. According to available results, it remains difficult to establish consensus sequences or rules defining protease specificity. Moreover, the use of free peptides does not address the role of global structures and interactions. The experimental system presented here allowed us to dissociate the influence of the target sequence (controlled by mutagenesis) and the conformational context (controlled by the TS structure at sites B and C).

In a previous study [9], we reported that the HIV-1 protease is able to recognize and process target sequences inserted into different surface loops of *E. coli* TS, whose structure is well characterized. Here, we confirmed that the conformational context exhibited by the TS is suitable to study the cleavage specificity of the HIV-1 protease. This system is sensitive enough to detect low cleavage efficiency as shown with modified targets described previously to be resistant to the retroviral protease. The TS system allowed us to assess the influence of the conformation upon the protease substrate recognition. We identified a cryptic site whose cleavage by the HIV-1 protease was dependent on the protein folding. The processing of the cryptic site, located 45 amino acids upstream of loop C (Figure 1B), was detected only in TS mutants destabilized by peptide insertion. In fact, proteases are known to degrade some proteins extensively only after these proteins loose their structural organization. Thus, albeit conservative, insertion mutagenesis of only two amino acids may be destabilizing enough to allow sufficient interaction between the enzyme and its potential substrate. Insertion in site C may have produced a new stable conformation of TS combining both TS activity and exposure of the previously hidden cryptic site. Alternatively, insertion at site C may have produced a dynamic instability resulting in structural diversity in the population of TS molecules. At any given time, cleaved molecules might be different from active molecules. The cleaved TS might be recruited from the most destabilized and inactive fraction of the enzymic population. This second interpretation does not require the existence of a TS conformation which is both active and different from the wild-type.

The presence of two amino acids, Phe-Pro, at the cryptic site or at the surface loops B and C was sufficient to induce a specific cleavage by the HIV-1 protease. It is noteworthy that the three loops cleaved specifically by the HIV-1 protease shared only two characteristics: the Phe-Pro dipeptide and their exposure at the protein surface [20]. It is unlikely that these three sites have additional structural features in common. Thus we may ask whether a neutral or even an unfavourable sequence context (i.e. in the absence of P4, P3, P3′ and P4′ residues known to stabilize the protease–substrate complex) could be compensated by an appropriate conformation such as those provided by the TS surface loops. Subsequently, we verified that other dipeptides, cleaved in the natural sequence context of the gag and gag-pol precursors, were also processed in the heterologous context of the TS. In our system, the class I sequence context, with Pro at P1′, was displayed by loop B of the TS while the sequence of loop C favoured the processing of sites belonging to class II, between two hydrophobic amino acids. It appeared that substrates from class II did not exhibit a particular preference for a given sequence context. Most of the tested combinations of two hydrophobic amino acids were cleaved in loops B and C.

The classification into two types of substrates, mainly established by studies on short peptides, does not allow predictions of cleavage in a protein context. The limitations of such classification [14] are certainly due to the variable exposure of cleavable bonds in a larger protein. In this context, the apparent cleavage specificity observed in viral proteins might result from the combination of the dynamics of exposure and the kinetics of cleavage. A compact globular protein remains unprocessed, even at favourable sequences, as we observed for the TS cryptic site. A short exposure would allow cleavages occurring with rapid kinetics; such may be the case for the conformation-dependent cleavage that we observed at a Phe–Pro bond, in an enzymically active, and thus relatively compact, TS mutant. A long exposure would allow more cleavage to occur, even at less favourable dipeptides. Very long exposure, corresponding to denaturation of the target, would allow cleavage at many sites, leading to complete degradation of the target protein. These considerations can be applied for the analysis of mutants emerging during exposure to protease inhibitors. Until now, mutations of resistance were identified exclusively in the protease gene or in sequences encoding cleavage sites. As shown in this report, some distant mutations can alter the presentation of the maturation sites and consequently could result in an improved apparent affinity of the HIV-1 protease for its substrates. Such resistance against drugs for protease binding could be associated with a relatively weak loss of viral fitness [22] and are, consequently, of critical interest.

We thank Dr Jennifer Richardson and J.-L. Guillaume for critical reading of the manuscript, Luc Camoin for help in protein sequencing and Sylvain Paisant for technical assistance. We also thank Professor Arthur Donny Strosberg for constant help and encouragement. This work was supported by Agence Nationale de Recherches sur le SIDA and Sidaction/Ensemble contre le SIDA.

REFERENCES

1 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. U.S.A. **85**, 4686–4690

2 Ashorn, P., McQuade, T. J., Thaisrivongs, S., Tomasselli, A. G., Tarpley, W. G. and Moss, B. (1990) An inhibitor of the protease blocks maturation of human and simian immunodeficiency viruses and spread of infection. Proc. Natl. Acad. Sci. U.S.A. **87**, 7472–7476

3 Tomasselli, A. G., Thaisrivongs, S. and Heinrikson, R. L. (1996) Discovery and design of HIV protease inhibitors as drugs for treatment of AIDS. Adv. Antiviral. Drug Des. **2**, 173–228

4 Doyon, L., Croteau, G., Thibeault, D., Poulin, F., Pilote, L. and Lamarre, D. (1996) Second locus involved in human immunodeficiency virus type 1 resistance to protease inhibitors. J. Virol. **70**, 3763–3769
5 Zhang, Y. M., Imamichi, H., Imamichi, T., Lane, H. C., Falloon, J., Vasudevachari, M. B. and Salzman, N. P. (1997) Drug resistance during indinavir therapy is caused by mutations in the protease gene and in its Gag substrate cleavage sites. J. Virol. **71**, 6662–6670
6 Schechter, I. and Berger, A. (1967) On the size of the active site in proteases. I. Papain. Biochem. Biophys. Res. Commun. **27**, 157–162
7 Pettit, S. C., Simsic, J., Loeb, D. D., Everitt, L., Hutchison, C. A. and Swanstrom, R. (1991) Analysis of retroviral protease cleavage sites reveals two types of cleavage sites and the structural requirements of the P1 amino acid. J. Biol. Chem. **266**, 14539–14547
8 Dunn, B. M., Gustchina, A., Wlodawer, A. and Kay, J. (1994) Subsite preferences of retroviral proteinases. Methods Enzymol. **241**, 254–278
9 Kupiec, J.-J., Hazebrouck, S., Leste-Lasserre, T. and Sonigo, P. (1996) Conversion of thymidylate synthase into an HIV protease substrate. J. Biol. Chem. **271**, 18465–18470
10 Belfort, M., Maley, G., Pedersen, L. J. and Maley, F. (1983) Primary structure of the *Escherichia coli thyA* gene and its thymidylate synthase product. Proc. Natl. Acad. Sci. U.S.A. **80**, 4914–4918
11 Farmerie, W. G., Loeb, D. D., Casavant, N. C., Hutchison, C. A., Edgell, M. H. and Swanstrom, R. (1987) Expression and processing of the AIDS virus reverse transcriptase in *Escherichia coli*. Science (Washington, D.C.) **236**, 305–308
12 Martinez, E., Bartolome, B. and de la Cruz, F. (1988) pACYC184-derived cloning vectors containing the multiple cloning site and lacZ alpha reporter gene of pUC8/9 and pUC18/19 plasmids. Gene **68**, 159–162
13 Kunkel, T. A., Roberts, J. D. and Zakour, R. A. (1987) Rapid and efficient site-specific mutagenesis without phenotypic selection. Methods Enzymol. **154**, 367–382
14 Gottlinger, H. G., Dorfman, T., Sodroski, J. G. and Haseltine, W. A. (1991) Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release. Proc. Natl. Acad. Sci. U.S.A. **88**, 3195–3199
15 Partin, K., Krausslich, H. G., Ehrlich, L., Wimmer, E. and Carter, C. (1990) Mutational analysis of a native substrate of the human immunodeficiency virus type 1 proteinase. J. Virol. **64**, 3938–3947
16 Tozser, J., Weber, I. T., Gustchina, A., Blaha, I., Copeland, T. D., Louis, J. M. and Oroszlan, S. (1992) Kinetic and modeling studies of S3-S3′ subsites of HIV proteinases. Biochemistry **31**, 4793–4800
17 Darke, P. L., Nutt, R. F., Brady, S. F., Garsky, V. M., Ciccarone, T. M., Leu, C. T., Lumma, P. K., Freidinger, R. M., Veber, D. F. and Sigal, I. S. (1988) HIV-1 protease specificity of peptide cleavage is sufficient for processing of gag and pol polyproteins. Biochem. Biophys. Res. Commun. **156**, 297–303
18 Hansen, J., Billich, S., Schulze, T., Sukrow, S. and Moelling, K. (1988) Partial purification and substrate analysis of bacterially expressed HIV protease by means of monoclonal antibody. EMBO J. **7**, 1785–1791
19 Tomaszek, T. J., Moore, M. L., Strickler, J. E., Sanchez, R. L., Dixon, J. S., Metcalf, B. W., Hassell, A., Dreyer, G. B., Brooks, I., Debouck, C. and Meek, T. D. (1992) Proteolysis of an active site peptide of lactate dehydrogenase by human immunodeficiency virus type 1 protease. Biochemistry **31**, 10153–10168
20 Hardy, L. W., Finer-Moore, J. S., Montfort, W. R., Jones, M. O., Santi, D. V. and Stroud, R. M. (1987) Atomic structure of thymidylate synthase: target for rational drug design. Science (Washington, D.C.) **235**, 448–455
21 Poorman, R. A., Tomasselli, A. G., Heinrikson, R. L. and Kezdy, F. J. (1991) A cumulative specificity model for proteases from human immunodeficiency virus types 1 and 2, inferred from statistical analysis of an extended substrate database. J. Biol. Chem. **266**, 14554–14561
22 Mammano, F., Petit, C. and Clavel, F. (1998) Resistance-associated loss of viral fitness in human immunodeficiency virus type 1: phenotypic analysis of protease and gag coevolution in protease inhibitor-treated patients. J. Virol. **72**, 7632–7637

Received 22 February 2001/18 May 2001; accepted 12 June 2001