STRUCTURAL BIOLOGY

ISSN 2059-7983

A substrate selected by phage display exhibits enhanced side-chain hydrogen bonding to HIV-1 protease

Ian W. Windsor and Ronald T. Raines*

Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. *Correspondence e-mail: rtraines@mit.edu

Received 17 January 2018
Accepted 1 May 2018

Edited by A. Berghuis, McGill University, Canada

Keywords: molecular recognition; HIV-1 protease; hydrogen bonds; phage display.

PDB reference: HIV-1 protease (D25N, inactive) in complex with the phage-display-optimized substrate SGIFLETS, 6bra

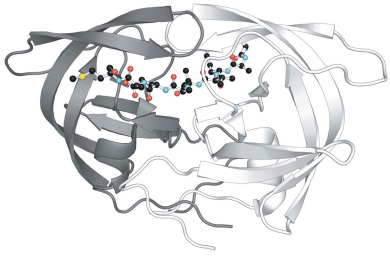
Supporting information: this article has supporting information at journals.iucr.org/d

Crystal structures of inactive variants of HIV-1 protease bound to peptides have revealed how the enzyme recognizes its endogenous substrates. The best of the known substrates is, however, a nonnatural substrate that was identified by directed evolution. The crystal structure of the complex between this substrate and the D25N variant of the protease is reported at a resolution of 1.1 Å. The structure has several unprecedented features, especially the formation of additional hydrogen bonds between the enzyme and the substrate. This work expands the understanding of molecular recognition by HIV-1 protease and informs the design of new substrates and inhibitors.

1. Introduction

Elaboration of how HIV-1 protease recognizes its endogenous substrates has been a triumph of structural biology (Prabu-Jeyabalan *et al.*, 2002, 2003; Liu *et al.*, 2011; Tie *et al.*, 2005). The homodimeric protease is known to bind peptidic substrates between its core and flaps through the formation of a mixed β-sheet-like motif. The conserved interactions with the main chain diminish the reliance on substrate side chains for recognition. The side chains of bound substrates are buried in subsites (Fig. 1*a*) through hydrophobic and nonconserved hydrogen-bonding interactions. Accordingly, HIV-1 protease substrates lack a rigid consensus sequence (Table 1). This variability could provide spatial and temporal regulation of proteolytic processing (Lee *et al.*, 2012).

Endogenous substrates exhibit a modest affinity for HIV-1 protease, having K_m values in the millimolar to high-micromolar range. Despite extensive efforts, few good substrates for HIV-1 protease have emerged from rational design (Altman *et al.*, 2008). In contrast, directed evolution has generated excellent substrates (Beck *et al.*, 2000; Szeltner & Polgár, 1996). In previous work, we employed a substrate for HIV-1 protease with a low micromolar K_m value, SGIFLETS, as the basis for a hypersensitive assay of catalytic activity (Windsor & Raines, 2015). Here, we report the high-resolution X-ray crystal structure of the complex of this substrate with an inactivated protease variant.



2. Materials and methods

2.1. Protein

The expression plasmid for D25N HIV-1 protease was prepared as described previously (Windsor & Raines, 2015) with modifications. An initiating methionine codon was placed directly before the native N-terminal proline residue, and an AAC codon was used for residue 25. D25N HIV protease was

© 2018 International Union of Crystallography

produced heterologously in *Escherichia coli* cells grown in Luria–Bertani medium. Expression was induced when the OD reached 1.5 at 600 nm, and the cells were grown for an additional 4 h. The cell pellets were suspended in 20 m*M* Tris–HCl buffer pH 7.4 containing 1 m*M* EDTA, lysed using a cell disrupter from Constant Systems and collected by centrifugation at 10 500*g* for 30 min. The cell pellet was dissolved in 20 m*M* Tris–HCl buffer pH 8.0 containing 9 *M* urea and the solution was clarified by centrifugation at 30 000*g* for 1 h. The supernatant was flowed through a 0.2 µm filter and a HiTrap Q column from GE Healthcare, which removed anionic contaminants (Velazquez-Campoy *et al.*, 2001). To fold the protease, the resulting solution was diluted 20-fold by dropwise addition to 50 m*M* sodium acetate buffer pH 5.0 containing 100 m*M* NaCl, 5%(*v*/*v*) ethylene glycol and 10%(*v*/*v*) glycerol. The solution of folded protease was concentrated using a stirred-cell concentrator from Amicon and applied onto a G75 gel-filtration chromatography column (GE Healthcare) that had been equilibrated with the folding buffer. The protease, which eluted near 0.5 column volumes, was concentrated to 10 mg ml^{-1}. The purity of the ensuing protein was verified by SDS–PAGE.

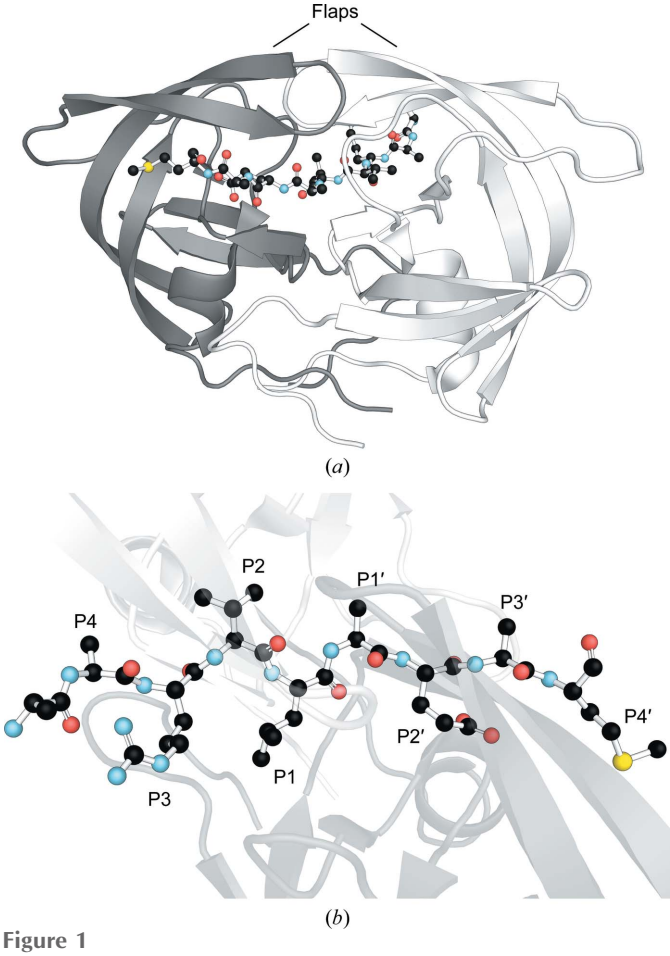


Figure 1
Structure of the D25N HIV-1 protease–CA/p2 complex. Protease residues from chain *A* are shown in white and those from chain *B* in gray; the substrate CA/p2 is shown in ball-and-stick representation (PDB entry 1f7a; Prabu-Jeyabalan *et al.*, 2000). (*a*) The substrate CA/p2 binds in the active site of the protease (white and gray) in an extended conformation between the two flaps and the core domain. (*b*) The substrate side chains are numbered relative to the scissile bond.

Table 1
Endogenous and optimized HIV-1 protease substrate sequences.

Residues in bold are shared with the substrate identified by phage display.

Substrate†	P4	P3	P2	P1	P1′	P2′	P3′	P4′
MA/CA	**S**	Q	N	Y	P	I	V	Q
CA/p2	A	R	V	L	A	**E**	A	M
p2/NC	T	A	**I**	M	M	Q	K	G
NC/p1	R	Q	A	N	F	L	G	K
p1/p6gag	P	**G**	N	**F**	**L**	Q	S	R
NC/TFP	R	Q	A	N	F	L	R	E
TFP/p6pol	N	L	A	**F**	Q	Q	G	E
p6pol/PR	**S**	F	S	**F**	P	Q	I	T
PR/RTp51	T	L	N	**F**	P	I	S	P
RT/RTp66	A	E	T	**F**	Y	V	D	G
RTp66/INT	R	K	V	L	F	L	D	G
Nef	D	C	A	W	**L**	**E**	A	Q
Phage display	**S**	**G**	**I**	**F**	**L**	**E**	**T**	**S**

† de Oliveira *et al.* (2003).

2.2. Peptide

The SGIFLETS peptide with free N- and C-termini was synthesized and purified to >99% purity by Biomatik, Wilmington, Delaware, USA. Stock solutions in DMSO containing 0.1%(*v*/*v*) TFA were prepared at a concentration of 1 m*M* for crystallization.

2.3. Crystallization

Protease and peptide stock solutions were mixed in a 4:1 volume ratio. Crystals were grown by vapor diffusion in 2 µl drops over a mother liquor consisting of 100 m*M* sodium acetate buffer pH 5.0 containing 1.0 *M* NaCl. The crystals, which grew within 24 h, were cryoprotected in mother liquor containing 10%(*v*/*v*) glycerol before flash-cooling with liquid nitrogen.

2.4. Data collection and processing

Single-crystal diffraction data were collected at Station G in Sector 21 (LS-CAT) of the Advanced Photo Source at Argonne National Laboratory. The data were indexed, integrated and scaled using *HKL*-2000 (HKL Research). Details of diffraction and data reduction can be found in Table 2.

2.5. Structure solution and refinement

Molecular replacement was conducted with *Phaser* as implemented in *PHENIX* (Adams *et al.*, 2010) using PDB entry 1kjf (Prabu-Jeyabalan *et al.*, 2002) as a starting model. Model building was conducted with *Coot* (Emsley *et al.*, 2010). Refinement with *PHENIX* following initial substrate placement revealed an additional antiparallel orientation of the substrate. Subsequent refinement estimated occupancies of approximately 0.6 (conformation *A*) and 0.4 (conformation *B*) for the major and minor orientations and revealed other alternative conformations for residues Ser1 and Gly2 in conformation *A* (Figs. 2*a* and 2*b*). Because of the complexity

Table 2
Crystallographic data-collection and refinement statistics.

Values in parentheses are for the last shell.

PDB code	6bra
Data collection	
X-ray source	LS-CAT 21-ID-G
Detector	MAR 300 CCD
Wavelength (Å)	0.97857
Resolution (Å)	26.0–1.11 (1.15–1.11)
Space group	$P2_12_12$
a, b, c (Å)	58.033, 85.767, 46.130
α, β, γ (°)	90, 90, 90
No. of reflections	612996
No. of unique reflections	90088 (8193)
Multiplicity	6.8 (3.9)
Mean $I/\sigma(I)$	33.6 (1.5)
Completeness (%)	98.63 (91.12)
R_{meas}	0.065 (0.691)
Wilson B factor (Å^2)	10.99
Refinement	
Reflections in working set	90066 (8181)
Reflections in test set	1998 (181)
R_{work}	0.1708 (0.2536)
R_{free}	0.1840 (0.2732)
R.m.s.d., bond lengths (Å)	0.004
R.m.s.d., bond angles (°)	0.78
No. of protein residues	206
No. of atoms	
Total	2101
Protein	1822
Ligand	4
Water	275
Average B factor (Å^2)	
Overall	15.25
Protein	13.46
Ligand	19.58
Water	27.04
Ramachandran statistics (*MolProbity*)	
Favored (%)	99
Allowed (%)	1
Outliers (%)	0

of constraining alternative conformations of some residues simultaneously with other residues that occupy subsites fully (*i.e.* 1.0), occupancies were set manually. The residues in conformation *A* with alternative conformations were assigned an occupancy of 0.4 (conformation *C*), with the original conformer retaining 0.2 of the total occupancy (0.6) of conformation *A*. Details of refinement and the statistics of the final model are listed in Table 2.

3. Results and discussion

3.1. SGIFLETS binds in alternative orientations

Unlike in analogous complexes, the substrate in the D25N HIV-1 protease–SGIFLETS complex lies in two antiparallel orientations (Figs. 2*a* and 2*b*). These orientations are not of equal occupancy (0.6 and 0.4 for *A* and *B*, respectively). Chemical symmetry (Table 1) of the residues in the P3 through P3′ positions and a serine residue at both the P4 and the P4′ positions are characteristics of SGIFLETS that could have led to this redundancy. Moreover, the protease flaps in the complex with SGIFLETS are in a previously unreported conformation in which a bridging water molecule (wat254) accepts hydrogen bonds from the main-chain amides of both Ile50 and Gly51, which are residues in the flaps. In other HIV-1 protease structures an intersubunit hydrogen bond forms between the main-chain atoms of Ile50 and Gly51. Despite unique interactions with its side chains, SGIFLETS is recognized by the protease through conserved interactions (Fig. 2*c*).

3.2. Ser1*A* and Gly2*A* at the P4 and P3 positions occupy alternative conformations

Beck and coworkers used directed evolution (*i.e.* phage display) with the intent of diversifying the P3 to P3′ residues (Beck *et al.*, 2000). Instead, they found that the residues of SGIFLETS varied in the P2 to P4′ positions. Beck and

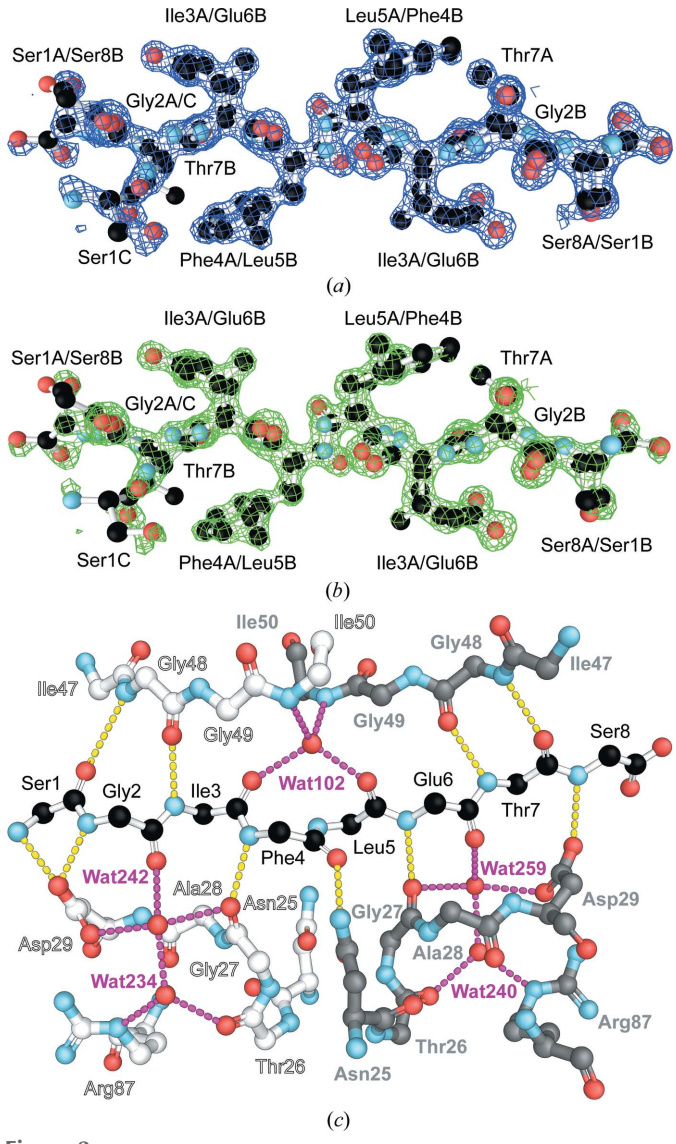


Figure 2
Electron density and interactions of SGIFLETS bound in the active site of D25N HIV-1 protease. Protease residues from chain *A* are labeled in white, those from chain *B* in gray and those from SGIFLETS in black. A $2F_o - F_c$ map (contoured at 1σ) (*a*) and an $F_o - F_c$ map after simulated-annealing refinement with the substrate excised (contoured at 3σ) (*b*) are depicted as a mesh around the substrate in the final structure. (*c*) Conformation *A* of SGIFLETS showing hydrogen bonds to HIV-1 protease residues (yellow, direct hydrogen bonds; magenta, water-mediated hydrogen bonds).

coworkers postulated that a marked preference for serine in the P4 position led to a high frequency of serine and glycine at the P4 and P3 positions in the selected substrates. Yet, both of these residues occupy alternative conformations in the major substrate orientation (conformation *A*) of the protease–SGIFLETS structure.

Few endogenous substrates exhibit alternative binding modes for P3 and P4 residues (Prabu-Jeyabalan *et al.*, 2002; Liu *et al.*, 2011). Unlike the conserved recognition strategy, in which the side chain of Asp29 accepts a hydrogen bond from the P3 main-chain NH and the NH of Gly48 donates a hydrogen bond to the P4 main-chain carbonyl O atom (Fig. 3*a*), the P3/P4 amide of the p1/p6 substrate interacts with the carbonyl O atom of Gly48 and the side-chain $N^{\eta 2}H$ of Arg8 (Fig. 3*b*). The protease–SGIFLETS complex employs both recognition strategies (in conformations *A*/*B* and *C*). Although found in opposite orientations relative to the protease, conformations *A* and *B* of Ser1 and Gly2 share the conserved β-sheet mode of main-chain recognition, with the side-chain hydroxyl group of Ser1*A*/*B* forming a unique hydrogen bond to Lys45 in the protease flap (Fig. 3*c*). The alternative conformer (conformation *C*) of the major orientation is reminiscent of the alternative recognition mode observed in the p1/p6 complex, with Ser1*C* instead forming a unique hydrogen bond to the carbonyl O atom of Gly49 (Fig. 3*d*). Alternative recognition of the P4 main-chain carbonyl group by p1/p6 and selected substrates occurs through both direct and water-mediated interactions with Arg8. The unique hydrogen bonding exhibited by Ser1 provides a structural explanation for the preference for serine and glycine at the P4 and P3 positions.

The tips of the protease flaps were also resolved in a previously unreported interaction in which a bridging water molecule accepts hydrogen bonds from the main-chain NH of both Ile50 and Gly51 (Fig. 3*d*). The occupancy of this novel water bridge correlates with the previously unreported hydrogen bond between the side chain of Ser1*C* and the carbonyl O atom of Gly49*B*. Rotation of Gly49*B* to accept the hydrogen bond appears to move the tip of the flap into a conformation that is incompatible with the inter-flap hydrogen bond, thus enabling water-bridge formation.

3.3. Glu6 and Ser8 form a network of hydrogen bonds

Weber and coworkers identified hydrogen bonds between the side-chain carboxyl group of a glutamic acid residue at position P2′ of the CA/p2 substrate and the side chain of Asp30 of the protease (Weber *et al.*, 1997). This interaction is also apparent in the protease–SGIFLETS complex (Fig. 4*a*). Given the pH of 5 at which these crystals were grown, a plausible explanation for the interatomic distance of 2.7 Å between $O^{\varepsilon 1}$ of Glu6 (P2′) and $O^{\delta 2}$ of Asp30 (chain *A*) is the formation of an intra-residue hydrogen bond (Fig. 4*b*). Such a hydrogen bond is consistent with the substantial increases in the Michaelis constant (K_m) of the peptide substrate and the inhibition constant (K_i) of an analogous inhibitor upon increasing the pH from 5.6 to 6.7 (Beck *et al.*, 2000). The different interatomic distances (2.7 and 3.3 Å) between $O^{\varepsilon 1}$ and $O^{\varepsilon 2}$ of Glu6 and $O^{\delta 2}$ of Asp30 suggest a single hydrogen bond to the proximal O atom and not a bifurcated hydrogen bond (Feldblum & Arkin, 2014).

The serine residues at P4/P4′ also form hydrogen bonds to the carboxyl group of Asp30 (Fig. 4*a*). Few polar interactions have been revealed between Asp30 and residues in the P4/P4′ positions, including arginine and serine. Neither of these interactions occur alongside a P2/P2′-interacting side chain

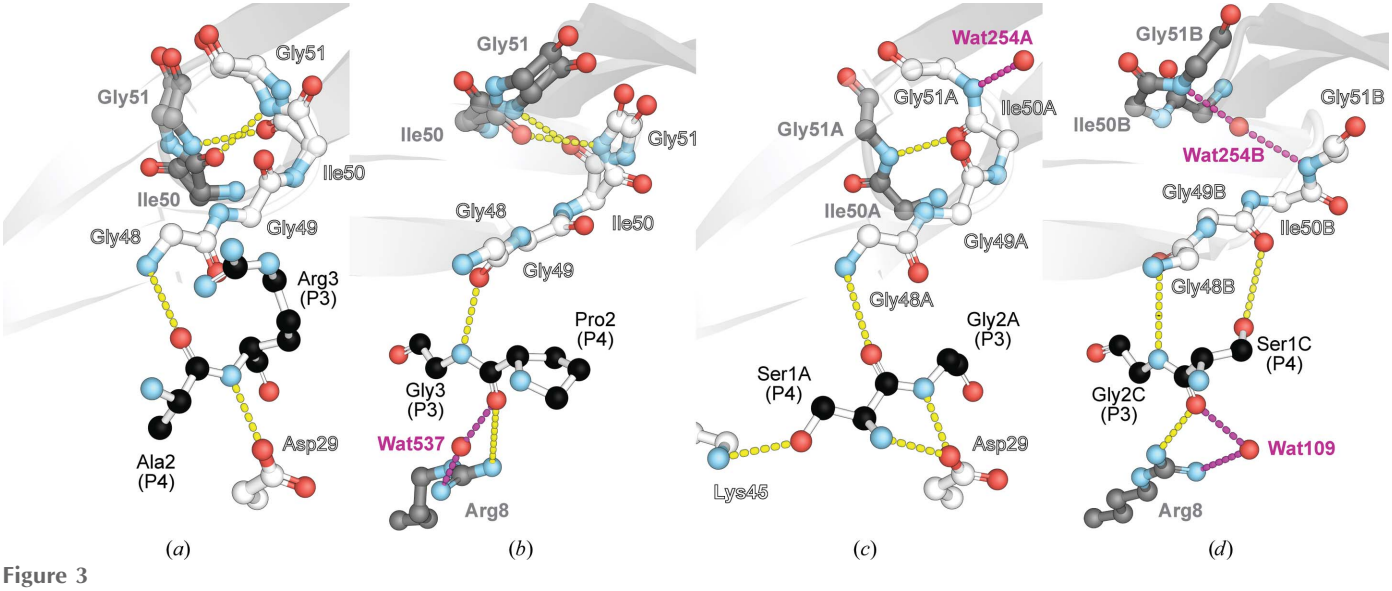


Figure 3
Alternative conformations of P3 and P4 residues. Protease residues from chain *A* are shown in white, those from chain *B* in gray and substrates in black. (*a*) The CA/p2 complex (PDB entry 1f7a). Ala2 (P4) and Arg3 (P3) form β-sheet-like interchain hydrogen bonds to Asp29 and Gly48. (*b*) The p1/p6 complex (PDB entry 1kjf). Gly48 and Arg8 alternatively recognize the P3/P4 amide. (*c*) The alternative orientations *A* and *B* of SGIFLETS exhibit the conserved β-sheet conformation and a unique hydrogen bond between the side chains of Ser1*A*/*B* (P4) and Lys45. (*d*) Alternative conformation *C* is similar to that in (*b*) and has a unique hydrogen bond between the side-chain hydroxyl group of Ser1*C* (P4) and the main-chain O atom of Gly49.

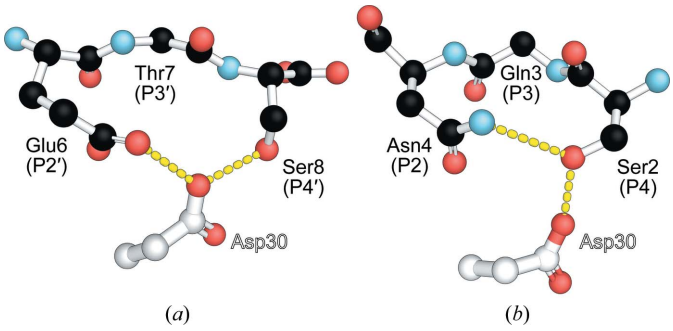


Figure 4
Role of Asp30. (*a*) Network of hydrogen bonds formed by Glu6 (P2′) and Ser8 (P4′) of SGIFLETS. Protease residues from chain *A* are shown in white and SGIFLETS in orientation *B* is shown in black. (*b*) Analogous hydrogen bonds formed by Ser2 (P4) and Asn4 (P2) of the MA/CA substrate (PDB entry 1kj4), although these residues interact with each other and only Ser2 interacts with Asp30.

(Fig. 4*b*). In the β-strand conformation of bound substrates, the side chains of adjacent residues ($i\cdots i+1$) are farther from each other than are the side chains of two residues with an intervening residue ($i\cdots i+2$) (Ridky *et al.*, 1996). Bulky groups can lead to a steric clash between side chains and only spatially compatible amino acids are found at the i and $i+2$ positions. The structure of the protease–SGIFLETS complex reveals the interdependence of P2′ and P4′ residues where, in addition to sterics, the identity of the residue is constrained by donor–acceptor interactions.

3.4. Thr7 plays a limited role

Endogenous protease substrates employ 2–5 polar residues in the core recognition sequence. Yet, only some of these side chains participate in hydrogen bonds, with an average utilization of about 60% (Özen *et al.*, 2011). In the protease–SGIFLETS complex three of the four polar side chains of SGIFLETS form hydrogen bonds. The exception is Thr7. The protease buries little of the P3/P3′ side chain, leaving residues in this position largely exposed to solvent. Although threonine is a residue in the P3′ position of HIV-1 protease substrates identified by phage display (Beck *et al.*, 2000), this position seems to have a limited role in substrate specificity (Tözsér *et al.*, 1992) and could be a site for further optimization.

4. Conclusion

The endogenous substrates of HIV-1 protease represent only a small subset of sequences that can be cleaved by the enzyme. Among the best of the known substrates, SGIFLETS, was derived by phage display. Its structure bound to the protease reveals the formation of many hydrogen bonds to its glutamic acid and serine side chains. Thus, hydrogen-bond formation could serve as a basis for the design of optimal substrates and perhaps of inhibitors of HIV-1 protease.

Funding information

IWW was supported by Biotechnology Training Grant T32 GM008349 (NIH) and a Genentech Predoctoral Fellowship. This work was supported by Grant R01 GM044783 (NIH).

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Altman, M. D., Nalivaika, E. A., Prabu-Jeyabalan, M., Schiffer, C. A. & Tidor, B. (2008). *Proteins*, **70**, 678–694.
Beck, Z. Q., Hervio, L., Dawson, P. E., Elder, J. H. & Madison, E. L. (2000). *Virology*, **274**, 391–401.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Feldblum, E. S. & Arkin, I. T. (2014). *Proc. Natl Acad. Sci. USA*, **111**, 4085–4090.
Lee, S. K., Potempa, M. & Swanstrom, R. (2012). *J. Biol. Chem.* **287**, 40867–40874.
Liu, Z., Wang, Y., Brunzelle, J., Kovari, I. A. & Kovari, L. C. (2011). *Protein J.* **30**, 173–183.
Oliveira, T. de, Engelbrecht, S., Janse van Rensburg, E., Gordon, M., Bishop, K., zur Megede, J., Barnett, S. W. & Cassol, S. (2003). *J. Virol.* **77**, 9422–9430.
Özen, A., Haliloğlu, T. & Schiffer, C. A. (2011). *J. Mol. Biol.* **410**, 726–744.
Prabu-Jeyabalan, M., Nalivaika, E. A., King, N. M. & Schiffer, C. A. (2003). *J. Virol.* **77**, 1306–1315.
Prabu-Jeyabalan, M., Nalivaika, E. & Schiffer, C. A. (2000). *J. Mol. Biol.* **301**, 1207–1210.
Prabu-Jeyabalan, M., Nalivaika, E. & Schiffer, C. A. (2002). *Structure*, **10**, 369–381.
Ridky, T. W., Cameron, C. E., Cameron, J., Leis, J., Copeland, T., Wlodawer, A., Weber, I. T. & Harrison, R. W. (1996). *J. Biol. Chem.* **271**, 4709–4717.
Szeltner, Z. & Polgár, L. (1996). *J. Biol. Chem.* **271**, 32180–32184.
Tie, Y., Boross, P. I., Wang, Y.-F., Gaddis, L., Liu, F., Chen, X., Tozser, J., Harrison, R. W. & Weber, I. T. (2005). *FEBS J.* **272**, 5265–5277.
Tözsér, J., Weber, I. T., Gustchina, A., Bláha, I., Copeland, T. D., Louis, J. M. & Oroszlan, S. (1992). *Biochemistry*, **31**, 4793–4800.
Velazquez-Campoy, A., Todd, M. J., Vega, S. & Freire, E. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 6062–6067.
Weber, I. T., Wu, J., Adomat, J., Harrison, R. W., Kimmel, A. R., Wondrak, E. M. & Louis, J. M. (1997). *Eur. J. Biochem.* **249**, 523–530.
Windsor, I. W. & Raines, R. T. (2015). *Sci. Rep.* **5**, 11286.