RESEARCH ARTICLE

HIV-1 protease tethered heterodimer–pepstatin-A complex: NMR characterization

S. C. Panchal[†], Bindu Pillai*, M. V. Hosur* and R. V. Hosur[†,]**

[†]Department of Chemical Sciences, Tata Institute of Fundamental Research, Mumbai 400 005, India
*Solid State Physics Division, Bhabha Atomic Research Centre, Mumbai 400 085, India

Wild type HIV-1 protease is a homodimer that plays a crucial role in the function of the virus and hence has been a target of anti HIV (human immunodeficiency virus) drug design. However, heterodimers generated by different mutations in the individual monomers have been seen to have different characteristics and thus have been used as macromolecular inhibitors of enzyme activity. Further, during the course of clinical treatment, it is conceivable that more than one mutant species exist inside a cell, resulting in the production of both homo- and heterodimers of the enzyme. In this context we have investigated here by NMR, the characteristic features of a particular heterodimer complexed to the peptide inhibitor pepstatin-A. The heterodimer has a GGSSG linker joining the two monomers head to tail and one of the monomers has a C95M mutation that lies in the dimerization domain. NMR backbone assignments have revealed that there is an asymmetry between the two monomer units. The secondary structural characteristics of the protein in the complex are almost identical to those in the crystal structure of the wild type homodimer protein complexed to the inhibitor, acetyl pepstatin. Thus neither the covalent linker nor the mutation seems to affect the gross solution structural features of the protease. Some local differences are, however, seen near the site of mutation. The inhibitor sitting in the active site cavity exhibits a flip-flop motion. Amide proton deuterium exchange studies reveal different local stabilities of the individual residues and also differences between the two monomer units. To our knowledge, this is the first NMR report characterizing a heterodimer of HIV-1 protease and forms a basis for future detailed investigation with different heterodimers, inhibitors, etc.

HUMAN immunodeficiency virus (HIV) is the etiological agent of Acquired Immunodeficiency Syndrome (AIDS) and HIV-1 protease is an aspartyl protease that catalyses the cleavage of several polypeptides yielding mature proteins that are required for the function of the virus. Therefore this protein has been the therapeutic target for several anti-HIV agents[1–3]. The functional protease enzyme is a 22 kDa homodimer that self-assembles from two identical polypeptide chains of 99 residues each, and resembles other monomeric aspartyl proteases[4–7]. Extensive X-ray crystallographic structural studies on the protein and its complexes with a variety of inhibitors have been carried out with a view to understand the molecular mechanisms of inhibition and to be able to design better drugs for the dreaded disease (reviewed in refs 8–11). The structure of the protein, which is seen to be highly conserved, contains a catalytic site at the interface of the two monomers, a flap region formed by two loops, one each from the two monomers, and the two monomers are held together by a short *b* sheet formed by the amino terminal of one monomer and the carboxyl terminal of the other. The inhibitor sits in the catalytic site or the active site of the enzyme. It has been observed that in the free protein the flap is in an open conformation, but gets closed when the protein binds to an inhibitor. This suggests that flexibility in the protein conformation plays a major role in the functioning of the enzyme.

The structural studies described above have been the basis for design of new inhibitors for the enzyme. Specific interactions between the protein and different types of symmetric and asymmetric inhibitors have been investigated, which in turn have suggested possibilities for new syntheses. There has been some success in these efforts and several drugs have been approved for clinical use. However a major hurdle in all these efforts has been the emergence of drug-resistant variants of the virus[12]. It has been observed that on continued treatment with a particular drug or a combination of drugs, several mutations appear in the HIV protease and this depends on the length of the treatment[13]. While many of these mutations are at the catalytic site, there are also some at places very remote. Wild type and mutant homodimers have different activities and binding affinities to inhibitors. Besides, there could be different species *in vivo* with different numbers and types of mutations arising out of possible integration of genetically variant[14] viral DNA into the host chromosomes, and the cellular soup may contain both homo- and heterodimers of the protease enzyme. Thus it follows that the design of drugs would have to take care of the additional asymmetry

**For correspondence. (e-mail: hosur@tifr.res.in)

factor, and a lot remains to be investigated, like the effects of asymmetry in the sequence, asymmetry of internal motions, asymmetry of interactions and so on.

The study of heterodimers is significant from another point of view. There are reports in the literature, which indicate that some of the heterodimers are less active compared to the wild type homodimers. The inactivity of the heterodimers has been related to instabilities[15] in the dimerization domains. These have led to artificial generation of heterodimers inside the cell as macromolecular inhibitors of viral maturation[15–17]. Also, single site mutations in the heterodimers may aid in the understanding of the interactions of inhibitors with the protease molecule. Thus a detailed knowledge of the structure and interactions of heterodimers is of considerable interest for understanding inhibitory actions of this dimeric protease.

In this context, the X-ray structure of a covalently tethered dimer–inhibitor complex[18] indicates that the structure is identical in every respect to that of the wild type dimeric molecule, except at the linker region. The tethered dimer has also been found to be functionally similar to the wild type enzyme[19,20]. Thus the tethered dimer may be a useful surrogate enzyme for studying the effects of single site mutations[21–23] on substrate and inhibitor binding as well as on enzyme asymmetry, and for simulating independent mutational drift of the two domains, which has been proposed to have led to the evolution of modern-day, single-chain aspartic proteinases. Thus we have initiated structural and dynamical studies by NMR on such a tethered dimer which has the linker GGSSG joining the two units and a single point mutation C95M in the first monomer only. This mutation may have significance for the dimerization process that is very crucial for the stability and regulation of the wild type protein[6,24–26]. Also, at this position methionine is present in HIV-2 which is homologous to HIV-1 (ref. 27). In this paper we report the NMR-derived features of the tethered heterodimer complexed to the general aspartyl protease inhibitor, pepstatin-A[1]; which is a hexapeptide (isovaleryl–Val–Val–Sta–Ala–Sta), where statine (Sta) is the rare amino acid (4S,3S)-4-amino-3-hydroxy-6-methylheptanoic acid. We present our results on secondary structure characterization, local stabilities and inhibitor dynamism in the complex. It so happens that the NMR spectral dispersions and patterns are similar to those of complexes of wild type protein with some other inhibitors, DMP323 and KNI-272 (refs 28 and 29). This indicates that the gross topology of the molecule is conserved. The detailed secondary structural characteristics of the protein are seen to be similar to those seen in the crystal structure of the wild type homodimeric protein complexed to a related inhibitor, acetyl pepstatin[30]. The specific mutation is however, seen to introduce some local differences. These results provide a basis for future detailed investigations on the protein with different mutations, different inhibitors, etc.

Materials and methods

Expression of HIV-1 protease tethered dimer

For HIV-1 protease expression, *Escherichia coli* strain BL21 (DE3) transformed with the expression vector pET11a containing the gene for HIV-1 protease as 29 kDa protein was obtained from John W. Erickson, NCI-Frederick Cancer Research and Development, USA. The uniformly ^{15}N-labelled HIV-1 Protease was obtained by growing bacteria at 37°C in M9 minimal medium, containing $^{15}NH_4Cl$ (1 g/l) as the sole nitrogen source. At an $OD_{600\ nm}$ of ~0.6–0.8, protein expression was induced by the addition of isopropyl *b*-D-thiogalactoside (IPTG) to a final concentration of 1 mM. After 5 h of further growth at 37°C, cells were harvested, washed and stored at –32°C. Uniformly $^{13}C/^{15}N$-labelled HIV-1 protease was obtained by growing cells in M9 medium containing $^{15}NH_4Cl$ (1 g/l) and ^{13}glucose (2 g/l) as the sole nitrogen and carbon sources, respectively.

Protein purification

For HIV-1 protease purification, the frozen cell pellets were thawed and resuspended in 15 ml of 10X TE buffer (100 mM Tris, 10 mM EDTA, pH 7.5) per litre of culture. The cell suspension was kept on ice for 30 min. The cells were lysed by sonication. The suspension was then centrifuged at 12000 *g* for 15 min at 4°C. The recombinant protein was seen to be expressed in inclusion bodies, hence was present in the cell pellet. The cell pellet was resuspended in 15 ml of 10X TE buffer, stirred over an ice-bath for 45 min and then again sonicated. This procedure was repeated three times. The pellet finally obtained was suspended in 25 ml of 63% glacialacetic acid to dissolve inclusion bodies and stirred over on ice-bath for 1 h. This suspension was subjected to sonication and then centrifuged at 35000 *g* for 1 h at 4°C. The supernatant was diluted 33-fold with water, dialysed against water overnight, and dialysed against a pH 6.5 buffer consisting of 20 mM Mes, 1 mM dithiothreitol (DTT), 100 mM NaCl, 5 mM EDTA, 10% (vol/vol) glycerol at 4°C for a few hours. During the latter dialysis, the 29 kDa polypeptide undergoes autolyses to form 22 kDa HIV-1 protease of correct amino acid sequence and the precipitate appears in the dialysis bag. The precipitate was removed by centrifuging the dialysed solution at 17400 *g* for 20 min at 4°C and pure HIV-1 protease was obtained as supernatant. The yield of the protein was 20 mg/l. For NMR sample preparation, the protein solution was concentrated and

exchanged with pH 5.2 NMR buffer consisting of 50 mM Na-acetate, 5 mM EDTA, 150 mM DTT by ultrafiltration.

Sample preparation for NMR experiments

Overall, the stability of the free protein was concentration-dependent and quite less, just like the wild type protease, to record 3D NMR experiments. However, it is reported that wild type protease–pepstatin-A complex is well formed and stable at pH less than 6.0 (ref. 31). Considering this and the fact that the protein is well folded at low concentrations, a complex of the heterodimeric protein with the inhibitor was formed by adding small aliquots (100 µl) of 5 times excess of inhibitor dissolved in methanol (4 mg of pepstatin-A dissolved in 2 ml of methanol) to the protein solution containing protein at µM concentration, pH 5.5 and at room temperature. The addition process was completed over a period of half to one hour and care was taken that the pH of the solution remained between 5.5 and 6.0. The solution was incubated at room temperature for about 10–12 h. The solution containing the complex was concentrated by ultrafiltration at 4°C. After each 50 ml reduction in volume, 200 µl of stock solution of pepstatin-A was added to the ultrafiltration cell. Since the presence of oxygen in the NMR sample tube may cause oxidation of cysteine residues, the sample tube was well purged with argon gas and another 10–20 µl of pepstatin-A stock solution was added before putting the protein sample inside the tube. The NMR tube was finally well sealed with parafilm.

NMR spectroscopy

The formation of protein–inhibitor complex did increase the protein stability[32], which was quite evident by no change in the ^{15}N-heteronuclear single quantum correlation (HSQC)[33] spectra even after a week. This made it possible for us to record a battery of 3D triple resonance experiments, namely HNCA[34], HN(CO)CA[35], CBCANH[36], CBCA(CO)NH[37] and HNCO[34] which are essential to correlate backbone $^{1}H^{N}$, ^{15}N, $^{1}H^{a}$, $^{13}C^{a}$, $^{13}C^{b}$ and ^{13}CO spins and also 3D ^{15}N-TOCSY-HSQC[38], ^{15}N-NOESY-HSQC[39]. All NMR experiments were carried out on a Varian Unity^{+} 600 MHz NMR spectrometer equipped with pulsed field-gradients, using 0.6 ml samples containing 1.0–1.5 mM protein–inhibitor complex in NMR buffer at 32°C, in a mixed solvent of 90% H_2O and 10% D_2O. NMR spectra recorded for resonance assignments are shown in Table 1. Data transformation and processing were done using the FELIX 97.0 software (Molecular Simulations Inc, San Diego, CA). Briefly, all the data were apodized with sine squared window function shifted by 60° and zero-filled before Fourier transformation. In all spectra, ^{1}H, ^{13}C, ^{15}N chemical shifts were, respectively, referenced to HDO (4.71 ppm at 32°C), indirectly to 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS), and to 3-(trimethylsilyl) propionate sodium (TSP)[40].

Amide proton–deuterium exchange studies were done by lyophilizing the protein from the acetate buffer (pH 5.2) containing 150 mM DTT and 5 mM EDTA in H_2O, dissolving the protein in D_2O and monitoring the signals in a series of two-dimensional ^{1}H–^{15}N HSQC spectra. The H/D exchange was monitored till there were no visible peaks in the spectrum just above the noise level. In each 2D experiment, 32 transients of 2048 data points were collected for each of the 140 t_1 increments.

Solvent accessibility calculations

For the purpose of interpreting amide proton–deuterium exchange data, the solvent accessibilities of the individual residues in the HIV-1–acetyl pepstatin complex

Table 1. NMR experiments and their parameters

	Time domain complex points[a]			Spectral width (Hz)					
Experiment	t_1	t_2	t_3	SW_1	SW_2	SW_3	Matrix size[b]	*nt*[c]	Time[d]
^{15}N-HSQC	200 (N)	2048 (H)	–	2300	8500	–	2048 × 4096	8	0.5
^{15}N-TOCSY-HSQC[e]	210 (H)	64 (N)	1024 (H)	8500	2300	8500	1024 × 128 × 1024	12	54.0
^{15}N-NOESY-HSQC[f]	210 (H)	64 (N)	1024 (H)	8500	2300	8500	1024 × 128 × 1024	16	76.0
HNCA	128 (C)	58 (N)	1024 (H)	5250	2300	8500	256 × 128 × 1024	12	28.5
HN(CO)CA	60 (C)	58 (N)	1024 (H)	5250	2300	8500	256 × 128 × 1024	12	29.0
HNCO	132 (C)	60 (N)	1024 (H)	2500	2300	8500	256 × 128 × 1024	8	11.0
CBCANH	132 (C)	60 (N)	1024 (H)	11500	2300	8500	256 × 128 × 1024	26	65.0
CBCA(CO)NH	120 (C)	60 (N)	1024 (H)	11500	2300	8500	256 × 128 × 1024	26	62.5

a, N, C, H in the parentheses stand for nitrogen, carbon and proton dimensions, respectively; b, Final matrix size in points after zero-filling; c, *nt* is the number of transients for each fid; d, Experimental time in hours; e, TOCSY mixing time used was 68 ms; f, NOESY mixing time used was 150 ms.

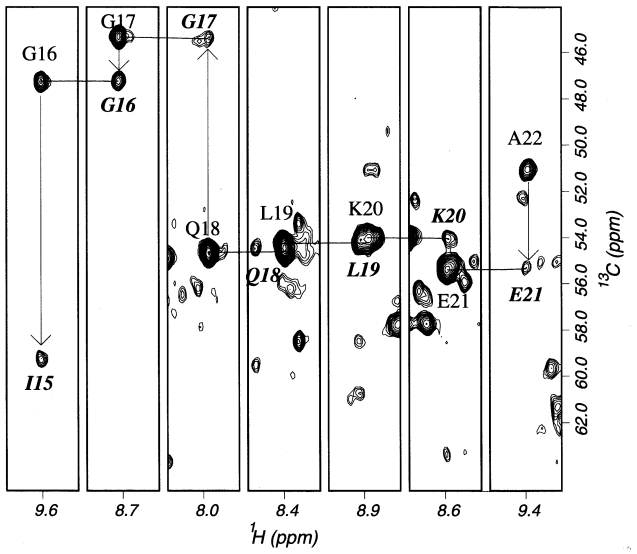

Figure 1. Selected ^{1}H–^{13}C[a] strips from the 3D HNCA spectrum at 32°C of ^{13}C, ^{15}N-labelled protein complexed to pepstatin-A. Sequential connections from G16 to A22 have been shown.

crystal structure[30] were calculated using the computer program NACCESS 2.1.1 (Hubbard, S. J. and Thornton, J. M., Department of Biochemistry and Molecular Biology, University College London, 1993) which employs the Lee–Richards method[41].

Results and discussion

Backbone assignments and dimer asymmetry

The assignments of the backbone resonances obtained following standard procedures based on triple-resonance and double-resonance heteronuclear experiments are listed in Table 1. Figure 1 shows an illustrative walk using the HNCA spectrum. Figure 2 shows the fingerprint, namely the ^{1}H–^{15}N HSQC spectrum of the heterodimer in the complex with residue-specific assignments. For many residues pairs of peaks are seen which come from the two monomer units. Table 2 lists

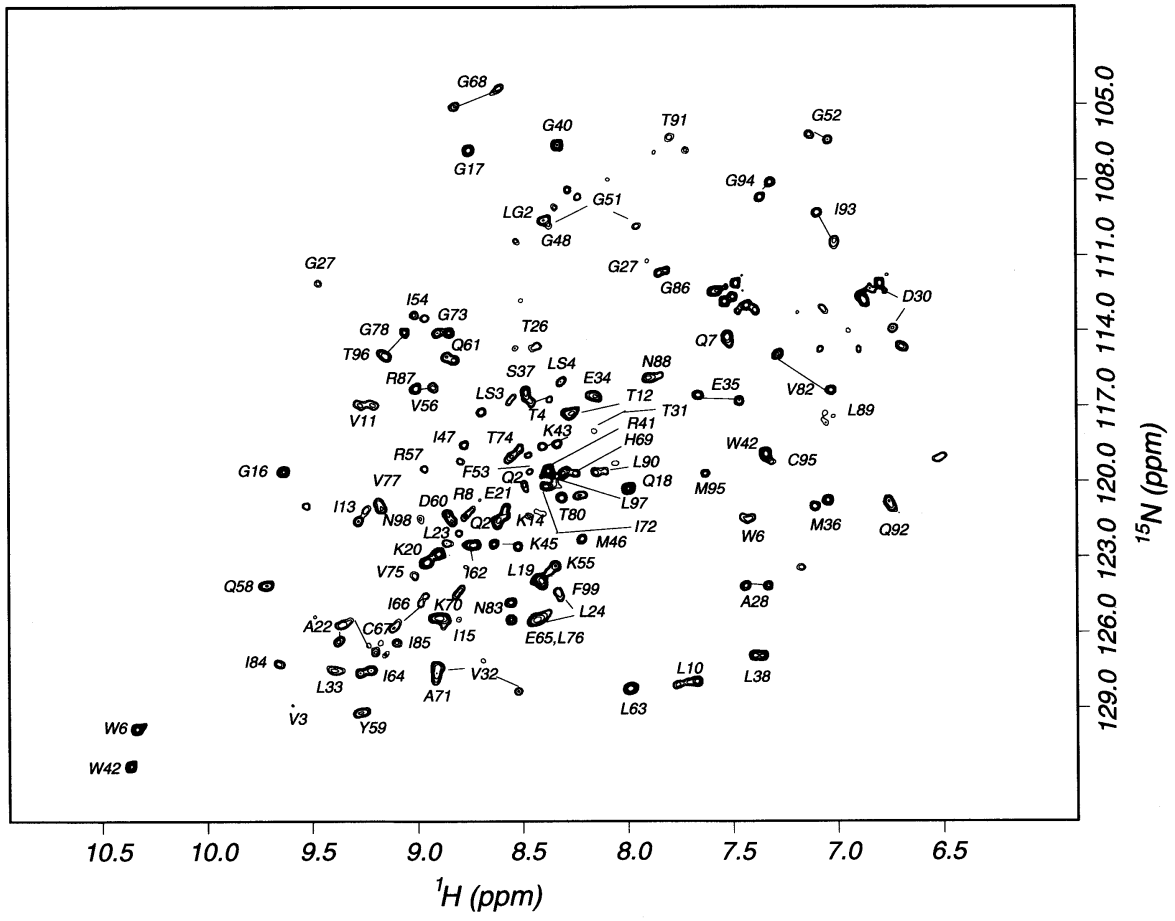

Figure 2. 600 MHz ^{1}H–^{15}N HSQC spectrum at 32°C of uniformly ^{15}N-labelled HIV-1 protease tethered dimer complexed to pepstatin-A with residue-specific assignments.

Table 2. Chemical shifts (in ppm) of backbone nuclei of HIV-1 protease tethered dimer complexed to pepstatin-A. Values in parentheses for some residues identify the second set of resonances arising from the asymmetry in the dimer. Linker residues are listed with a prefix L

Residue	^{15}N	H^N	$^{13}C^a$	$^{13}C^b$	^{13}CO
P1			63.31		173.55
			(62.38)	(33.24)	(170.95)
Q2	121.65	8.59	55.24	32.06	174.98
	(120.11)	(8.45)	(55.2)		
V3	128.33	9.54	62.05		176.26
					(176.29)
T4	116.74	8.32	61.28		
	(116.79)	(8.4)	(61.19)		
L5			54.05		176.26
W6	121.48	7.41	59.48	29.19	175.99
Q7	114.38	7.47	53.17	31.69	175.88
R8	121.1	8.71	56.26		
P9			61.92		173.52
			(61.94)		(173.52)
L10	127.98	7.66	53.92	43.97	177.57
	(127.98)	(7.72)	(53.92)	(43.97)	(177.54)
V11	116.86	9.24	59.04	36.26	174.54
	(116.86)	(9.14)	(59.04)	(36.26)	
T12	117.28	8.21	62.78	69.75	174.27
I13	121.58	9.25	58.68	41.45	173.55
K14	121.05	8.57	55.13	36.03	175.91
I15	125.44	8.83	59.23	40.76	175.29
G16	119.62	9.59	47.18		175.5
G17	106.81	8.71	45.35		173.69
Q18	120.3	7.98	54.61	31.67	174.28
L19	123.92	8.37	54.39	43.1	177.49
K20	122.89	8.86	54.09	37.35	174.25
E21	121.44	8.55	55.39	32.32	176.32
A22	126.41	9.34	50.97	24.87	173.6
	(125.25)	(9.32)	(50.97)	(24.87)	
L23	122.48	8.85	54.0	46.22	176.32
					(176.93)
L24	125.03	8.26	54.88		175.91
	(125.13)	(8.42)	(54.9)		
D25			53.15		175.91
T26	114.63	8.42	65.23		175.03
					(175.07)
G27	112.08	9.43	45.6		172.32
	(111.64)	(7.79)	(45.03)		(171.52)
A28	124.16	7.29	49.41	21.95	
	(124.16)	(7.41)	(49.49)		
D29			57.98	42.67	175.12
			(57.48)		
D30	113.9	6.69	52.12	45.73	172.73
	(112.42)	(6.77)	(51.69)	(45.63)	
T31	117.96	8.11	63.58	74.0	173.99
			(63.32)		(173.89)
V32	128.35	8.5	60.05		173.23
	(127.63)	(8.87)	(59.36)		(172.21)
L33	127.51	9.38	52.23	45.13	175.12
E34	116.61	8.14	55.49	30.89	175.54
			(55.51)		(174.72)
E35	116.82	7.43	58.67	30.55	175.79
	(116.58)	(7.64)	(58.48)		(175.62)
M36	121.0	7.02	54.88	34.49	173.52
	(120.75)	(7.07)	(54.88)	(34.49)	
S37	116.4	8.44	57.7	63.37	173.93
L38	126.39	7.34	51.99	43.53	
P39			62.92	32.78	177.04
G40	106.61	8.29	44.69		174.31
R41	119.59	8.35	56.34	31.53	175.0
W42	118.91	7.3	53.91	31.45	175.56
					(175.50)
K43	118.49	8.31	53.17	35.02	
	(118.59)	(8.37)	(53.25)	(35.02)	
P44			63.25	32.96	176.52
			(63.32)	(32.96)	(176.46)
K45	122.51	8.59	55.5	37.21	173.81
	(122.59)	(8.51)	(55.96)	(37.05)	(173.81)
M46	122.29	8.2	54.33	35.02	175.88
I47	118.53	8.76	59.51		173.72
G48	109.84	8.36	44.97		
I50			63.79		175.07
			(65.86)		(174.59)
G51	109.17	8.37	44.5		173.14
	(109.76)	(7.93)	(45.08)		(173.05)
G52	106.44	7.01	44.53		171.3
	(106.21)	(7.09)	(44.56)		
F53	119.63	8.44	56.86	42.46	176.81
			(56.84)		
I54	113.51	8.97	59.88	42.8	173.84
	(113.31)	(8.92)	(59.88)		
K55	123.63	8.33	56.18	33.47	176.64
			(56.39)		(176.61)
V56	116.32	8.96	58.39	35.57	174.89
		(8.89)	(58.45)		(174.89)
R57	119.49	8.93	55.7	32.63	175.12
Q58	124.13	9.71	56.0	30.59	174.89
Y59	129.96	9.25	57.63	41.28	174.37
D60	121.58	8.8	54.04	43.89	175.15
Q61	115.08	8.8	56.77	27.02	174.13
I62	122.55	8.7	57.66	37.07	174.83
L63	128.23	7.98	59.78	42.92	175.71
					(175.76)
I64	127.56	9.23	59.78	42.38	173.69
	(127.5)	(9.19)	(59.89)		
E65	125.51	8.38	54.89	32.31	176.29
			(55.04)	(31.69)	(176.20)
I66	125.6	9.08	60.87		175.07
	(124.54)	(8.94)	(60.16)		(174.83)
C67	126.49	9.18	59.67	27.26	175.56
	(125.25)	(9.29)	(59.66)	(26.67)	(175.09)
G68	105.06	8.76	45.77		174.07
	(104.44)	(8.54)	(45.57)		(173.86)
H69	119.64	8.21	54.66	30.17	174.31
	(119.64)	(8.09)	(54.71)	(30.42)	(173.99)
K70	125.43	8.85	57.56	33.72	174.92
	(124.38)	(8.78)	(57.05)	(33.14)	
A71	127.93	8.88	51.0	22.71	174.3
I72	120.2	8.37	59.18	40.72	176.93
			(59.22)		(176.89)
G73	114.07	8.87	46.17		171.15
	(114.07)	(8.83)	(46.17)		
T74	118.99	8.52	63.43		175.0
	(118.89)	(8.5)	(63.33)		
V75	123.23	8.98	61.09		172.73
L76	125.14	8.37	52.49	41.79	176.2
V77	121.1	9.15	60.14		177.19
			(60.13)		(176.99)

continued

Table 2. *Continued*

Residue	^{15}N	H^N	$^{13}C^a$	$^{13}C^b$	^{13}CO
G78	115.0	9.13	46.11		
	(113.82)	(9.02)	(46.11)		
P79			63.75	28.78	175.94
			(63.76)	(32.79)	(175.59)
T80	120.7	8.27	57.2		
	(120.7)	(8.18)	(54.9)		
P81			64.02		175.65
			(64.39)		(175.97)
V82	114.94	7.26	59.55	35.76	173.31
	(116.37)	(7.02)	(60.51)	(37.11)	(174.28)
N83	124.79	8.52	54.37	38.72	
	(125.48)	(8.52)	(54.37)		
I84	127.27	9.64	60.95	39.12	174.54
I85	126.39	9.07	58.55		174.8
G86	111.64	7.79	44.67		177.16
R87	116.3	8.95	61.43		176.58
N88	115.83	7.89	57.14	36.93	175.33
L89	117.59	7.01	55.64	42.7	178.71
L90	119.61	8.0	58.22		179.21
T91	106.24	7.77	63.96	68.17	178.13
Q92	121.04	6.69	58.1	29.49	177.43
				(29.82)	(177.37)
I93	109.25	7.06	60.74	37.57	174.89
	(110.37)	(7.0)	(60.63)		(175.36)
G94	108.08	7.29	46.18		174.25
	(108.68)	(7.35)	(46.67)		(174.77)
C95(M)	118.91	7.29	55.69		174.98
	(119.71)	(7.59)	(59.03)		
T96	115.0	9.14	59.99		173.72
L97	120.0	8.3	53.88		176.72
N98	121.3	8.94	53.34	42.91	172.7
F99	124.38	8.3	58.52	38.92	
LG1			46.07		174.89
LG2	109.6	8.35	45.02		175.33
LS3	116.65	8.51	60.66	63.78	175.44
LS4	116.0	8.26	57.12	64.28	

all the residue-specific backbone assignments. Figure 3 summarizes through-bond backbone connectivities observed in the triple-resonance experiments. The residues for which two cross peaks are observed in the ^1H–^{15}N HSQC spectra are marked by empty circles in Figure 3.

The heterodimer has two sources of asymmetry, namely the linker and the C95M mutation in one of the monomers. The linker results in non-equivalence of few residues near the N and C terminals of the individual monomer units. Since the mutation is also near the C terminal, it extends this asymmetry to a few more residues. The protein–pepstatin-A complex has another source of asymmetry, namely the asymmetric inhibitor. Thus it must be expected that the interactions of the inhibitor with the two monomers would be different and one would get different resonances for some residues from the two monomers. Indeed we observe from Table 2 and Figure 2 that for about 20% of the residues we have two sets of chemical shifts. Figure 4 shows the amide and ^{15}N chemical shift differences of the individual residues in the two monomer units of the protein in the complex. The magnitudes of differences are seen to be quite varied and most noteworthy are the changes in the catalytic site, e.g. G27.

Chemical shift indices and secondary structure

The above assignments form the basis for structural and dynamic investigations on the protein and its interactions with different inhibitors. Figure 3 also shows the chemical shift indices[42] for $^{13}C^a$, $^{13}C^b$ and ^{13}CO. A negative value in the case of $^{13}C^a$ and ^{13}CO and a positive value in the case of $^{13}C^b$ are indicative of a *b*-sheet backbone conformation. On this basis it is seen from Figure 3, that the secondary structure characteristics of the heterodimer in solution are very similar to those seen in the crystal structure of the complex of wild type HIV-1 protease with acetyl-pepstatin. Thus neither the linker nor the specific mutation seems to affect the gross secondary structural characteristics of the protein.

One of the two peaks of I93 (Figure 2) is stronger than the other, implying that the two amides have different exchange rates with the solvent. This, in turn, reflects on different solvent exposures or different flexibilities of the two components of the dimerization domain. Thus the specific mutation, though does not prevent the dimerization process, does introduce some local structural differences.

Deuterium exchange and local stabilities

H/D exchange studies provide valuable information on the local stabilities of individual residues in a given protein. Amide proton–deuterium exchange reaction in the folded state has been extensively studied for several proteins and the process is described by a two-state model as follows[43–46]:

$$(N–H)_{cl} \leftrightarrow (N–H)_{op} \rightarrow (N–D).$$

The protein is believed to be in two sets of conformations, closed and open, in equilibrium with each other. However, solvent exchange occurs only from the open state of the protein. The exchange characteristics are a reflection on the solvent accessibilities of the individual amide protons in the folded protein on the one hand, and on the local stabilities of the conformation on the other[46]. The hydrogen bonded amide protons exchange much slower compared to the non-hydrogen bonded ones. Thus hydrogen–deuterium (H/D) exchange studies give considerable insight into structure, stability, folding, dynamics and intermolecular interactions in protein systems in solution. In favourable situations where protection factors can be quantitatively measured for a

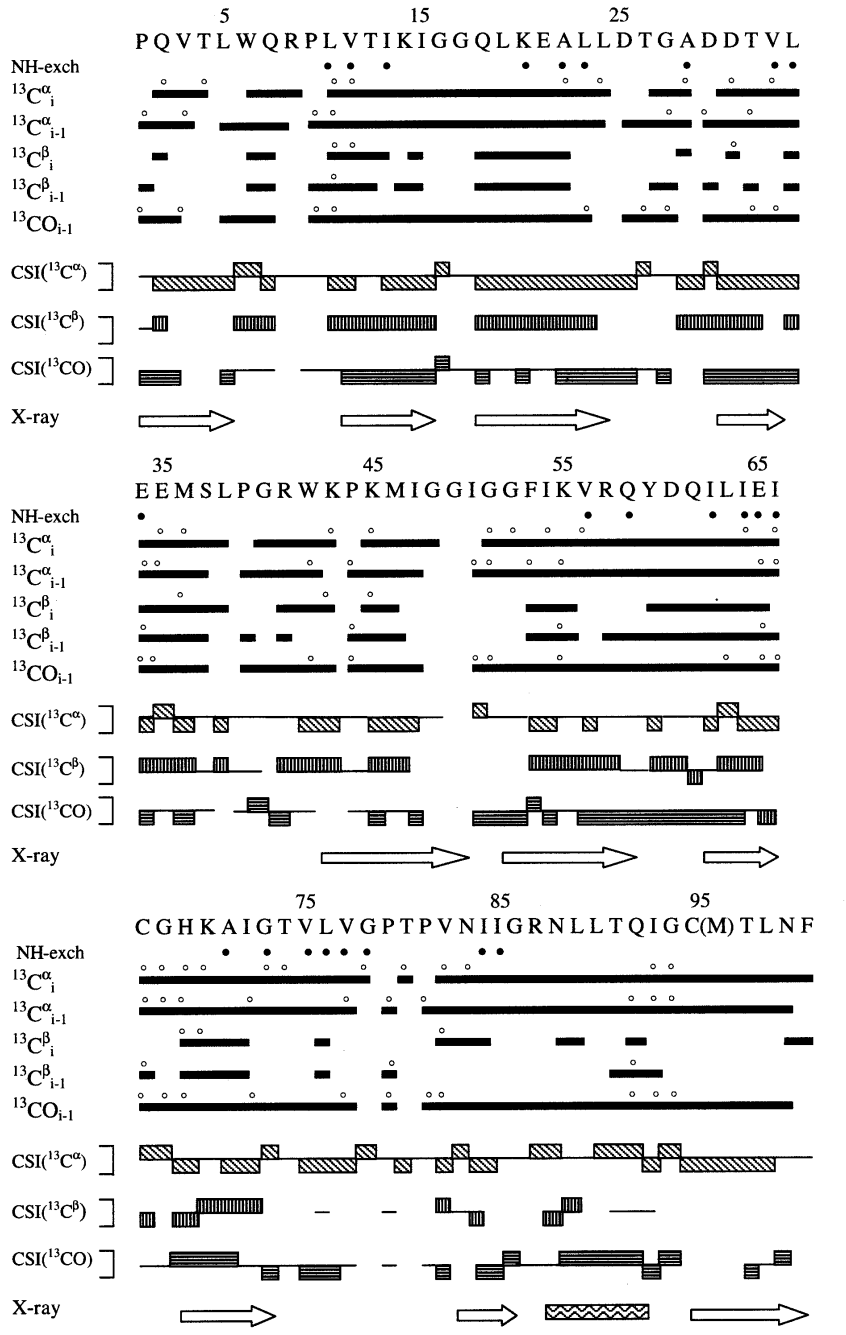

Figure 3. Summary of the through-bond connectivities and chemical shift indices (CSI) identified for assignment of backbone resonances of HIV-1 protease tethered dimer complexed to pepstatin-A. Four through-bond correlations between the $^{15}N_i$ nucleus and ^{13}C nuclei in amino acids *i* and *i*–1, provided the assignments, as follows: $^{13}C^{a}_{i}$ in HNCA; $^{13}C^{a}_{i-1}$ in HN(CO)CA and HNCA; $^{13}C^{b}_{i}$ in CBCANH; $^{13}C^{b}_{i-1}$ in CBCA(CO)NH and CBCANH; $^{13}CO_{i-1}$ in HNCO. Solid bars in each row indicate the presence of a cross peak in the spectra for these correlations. The consensus CSI are based on the changes from random-coil values for $^{13}C^{a}$, $^{13}C^{b}$ and ^{13}CO. These values are displayed in the bottom row, labelled CSI. A negative value in $^{13}C^{a}$ and ^{13}CO and a positive value in $^{13}C^{b}$ are indicative of a *b*-sheet backbone conformation. Below CSI are given the secondary structure elements (*b*-sheet by thick arrow and *a*-helix by cylinder) present in the X-ray crystal structure of HIV-1 protease–acetyl pepstatin complex for comparison. The residues where two cross peaks are observed in HSQC spectra are marked by empty circles. Filled circles indicate the residues whose amides survived after first 40 minutes after adding D_2O.

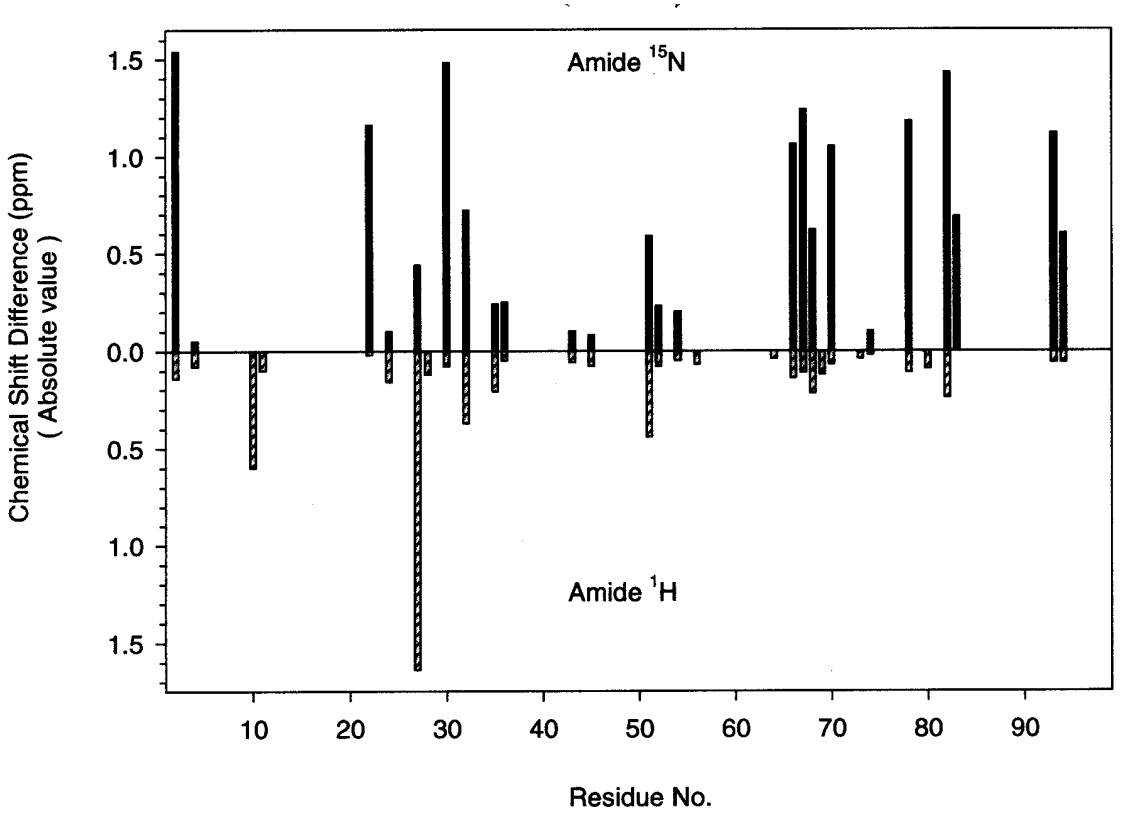

Figure 4. Residue-wise comparison of the amide proton and amide ^{15}N chemical shift differences between the two monomers in protein–pepstatin-A complex.

large number of residues, these studies have the potential of defining the energy landscape of a protein.

In the present case, considering the asymmetry in the structure in some regions of the protein–pepstatin-A complex (refer to Table 2 of supplementary material), one could in principle monitor H/D exchange kinetics of 124 backbone amide protons (see Figure 2). However, of these, only 24 were seen to survive after the first 40 minutes after addition of D_2O, 18 survived after 140 minutes and so on, as shown in Figure 5 *a*. This directly displays the variation of exchange time along the polypeptide chain. Four residues, namely **V32**, E34, A71 and V75 are protected till 26 h. The other 20 amide protons which do not exchange out in the first 40 minutes are **L10**, **V11**, I13, K20, **A22**, L23, **A28**, L33, **V56**, Q58, I62, **I64**, E65, **I66**, **G73**, L76, V77, **G78**, I84 and I85. The residues indicated in bold are those for which separate peaks are seen for the two monomer units, but only one of them is retained in the H/D exchange experiments. A quantitative characterization of the exchange kinetics was not feasible because of insufficient time resolution in the data. However, useful qualitative information on the local stabilities of the heterodimeric protein has been derived as described next in the article.

Taking the X-ray structure of HIV-1 protease with acetyl pepstatin to be a valid solution structure for the present system, the differently exchanging residues are shown in a colour-coded manner in Figure 6. It is seen that even some of the residues involved in *b*-sheet conformation also exchange significantly within one hour. This includes Q2 to L5, K43 to G49, G52 to K55, H69, K70, C95 to F99. These residues belong to the dimerization domain (Q2–L5, C95–C99), hinge region (K43–G49) and flap regions (G52–K55), respectively. This is significant considering that the dimerization domain is believed to be playing an important role in the control mechanism of the enzyme and the hinge and flap regions are important for inhibitor binding.

The amide proton–deuterium exchange results have to be interpreted by first considering their solvent accessibilities in the structure and then in terms of the local stabilities. To this end, we calculated the solvent accessibilities of the main chain atoms of the individual residues from the crystal structure as mentioned in Materials and methods section. Data are also shown in Figure 5 *b*. We see a wide variation in the solvent accessibilities among the different residues and also differences are seen between the two monomer units. The peaks which are seen in the H/D exchange spectra are those for which the solvent accessibility is nearly zero, except in L10, for which the accessibility value is 69.5. This peak should have been absent in the H/D exchange

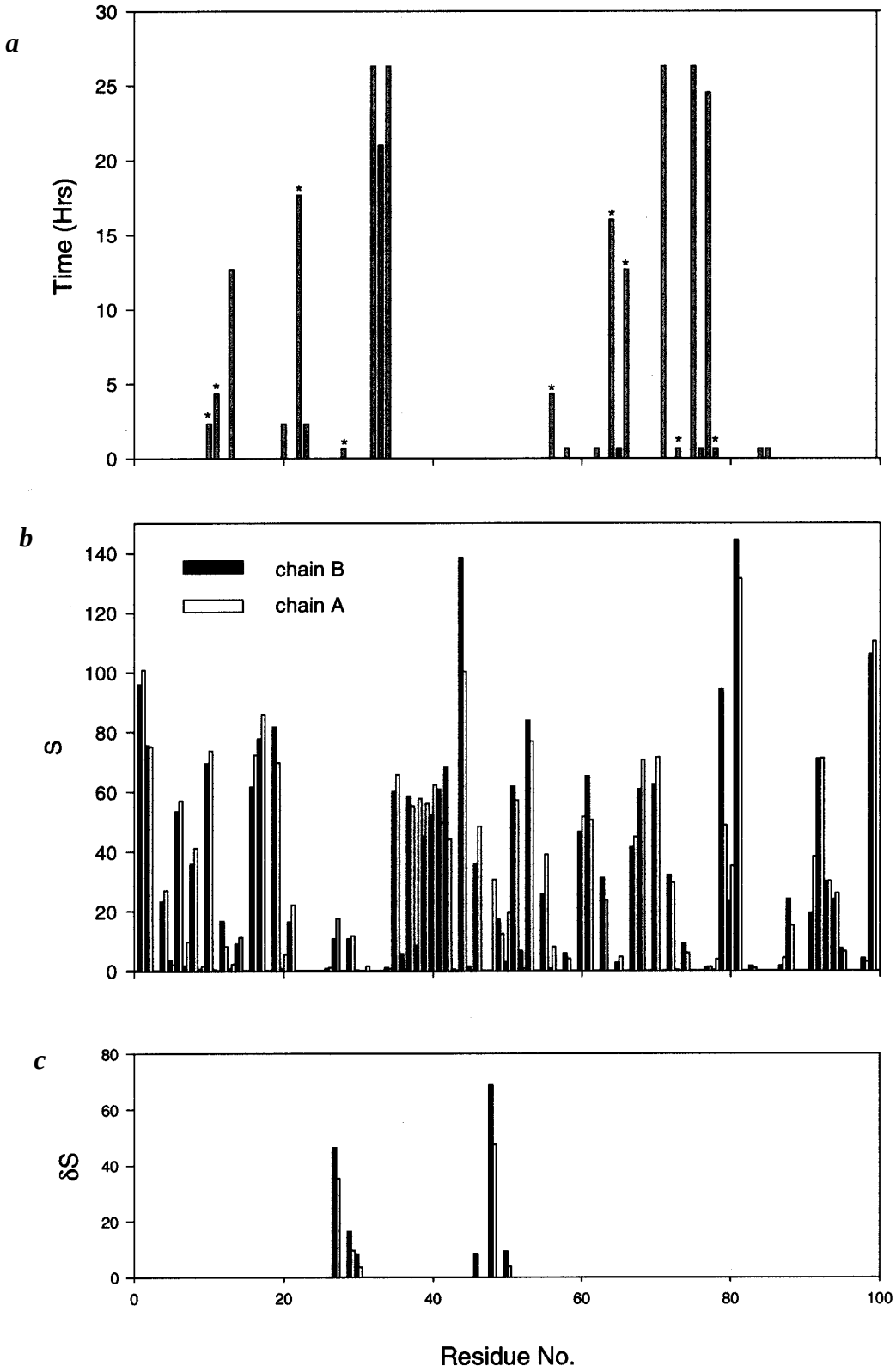

Figure 5. ***a***, Variation of the exchange time along the polypeptide chain in amide proton–deuterium exchange experiments. Residues for which only one peak is seen in the exchange experiment (D_2O ^{15}N-HSQC) as against two in H_2O ^{15}N-HSQC have an asterisk (*) on top of the bar; ***b***, Relative solvent accessibilites of the residues calculated using the X-ray structure of HIV-1 protease–acetyl pepstatin complex in both chains. ***c***, Difference plot of solvent accessibilities to indicate the residues protected by the inhibitor (see text).

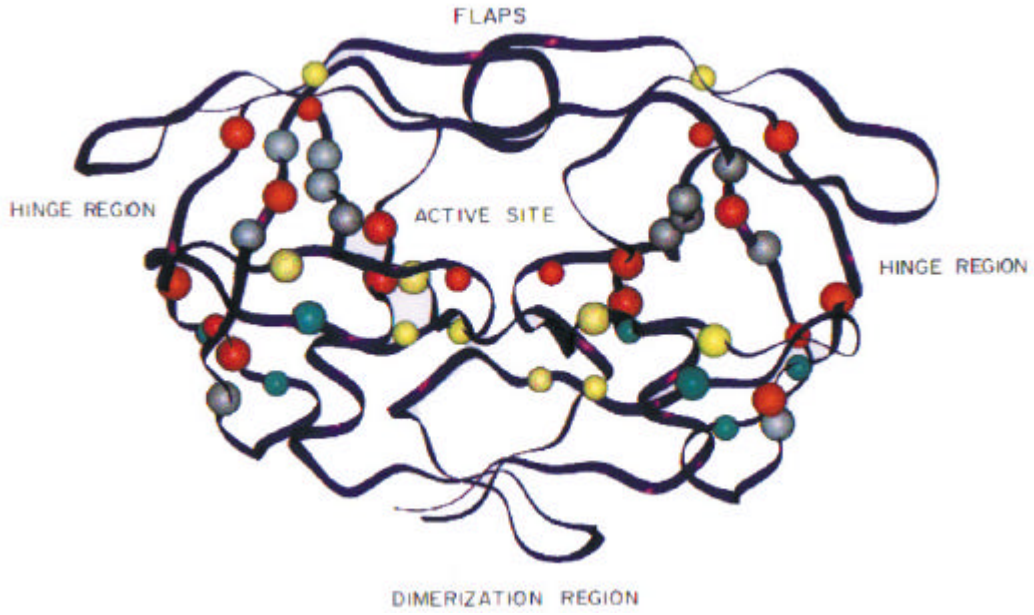

Figure 6. Residues retained during the amide proton–deuterium exchange experiments are shown in colour-coded manner. Residues which survive for 0.7 h from the start of experiment are shown in red, for 4.3 h in yellow, for 18 h in green and for 21 h in grey. Smaller size of the ball within a given set indicates that only one peak is seen in the experiment compared to the two peaks present in H_2O ^{15}N-HSQC.

spectra. Though the solvent accessibilities are nearly zero for all the residues retained in H/D exchange spectra, there is substantial gradation in their retention times. Similarly, for residues 3, 7, 15, 18, 26, 30, 31, 43, 45, 47, 48, 54, 57, 59, 69, 76, 82, 83, 86, 87, 89, 90, 96 and 97 the solvent accessibility values are nearly zero, but they are not seen in the H/D exchange spectra. These residues are well protected, but show high exchange rates. These differences could either mean that the solution structure is different from the crystal structure or that there is a high degree of open–close local flexibility in the structure. The secondary structural information derived from the chemical shift indices data described earlier, shows a close parallel with the X-ray structure, and hence it is reasonable to think that the overall fold and tertiary structures in solution and crystal are also not very different. Moreover, in a large number of structures with different inhibitors, mutations, etc. which have been reported in the literature, the backbones of the polypeptide chains are nearly superimposable[8]. Thus the above discrepancies between solvent accessibilities and exchange results must be interpreted in terms of the different degrees of local stabilities in the molecule. The highly protected residues V32, E34, A71 and V75 must have very rigid structures.

We notice that for 10 residues, **L10**, **V11**, **A22**, **A28**, **V56**, **I64**, **I66**, **G73**, **G78**, where two sets of peaks are seen in the H_2O spectra, only one of them is seen in the H/D exchange spectra. Since the solvent accessibilities for all of these in both the monomers are zero, except for L10, which anyway does not have good correlation with solvent accessibility, this difference must indicate differences in the local open–close conformational flexibilities in the two monomer units. Interestingly, these are distributed throughout the sequence of the polypeptide chain. Thus while the chemical shift data identify the asymmetry in the structure, the H/D exchange data identify the asymmetry in the stabilities in the two monomer units.

In order to probe the active site directly and to evaluate the contribution of the inhibitor to amide protection, we removed the inhibitor coordinates from the crystal structure coordinates of the complex and again calculated the solvent accessibilities. A difference plot of the solvent accessibilities with and without the inhibitor is shown in Figure 5 *c*. We notice that six residues are well protected (two of them, G27 and G48 more significantly) by the inhibitor. The H/D exchange data, however, show a remarkable difference. One would have expected the above residues to be slowly exchanging. In contrast, they exchange out within the first 40 minutes indicating, as mentioned earlier, a high degree of local flexibility in the active site (G27) and flap region (G48) of the complex. The above amide peaks are also weak in the H_2O spectrum, which supports the above conclusion.

Flip-flop motion of inhibitor in the complex

During the course of the assignments, we observed cross peaks between the amide protons of V82 in the

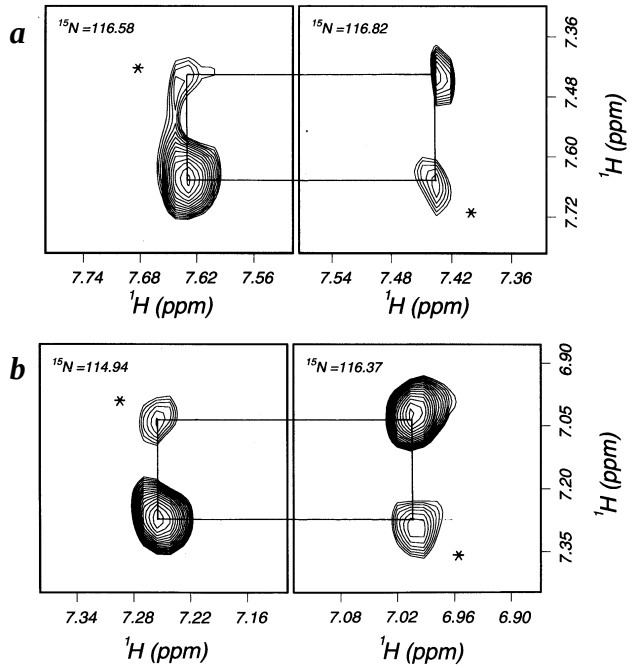

Figure 7. Strips from a 3D ^{15}N-separated 600-MHz NOESY spectrum of the protease complexed to pepstatin-A at the ^{15}N chemical shifts of (***a***) E35 and (***b***) V82 residues. Left and right panels contain the peaks from the individual monomers and the connections drawn identify chemical exchange cross peaks (*) arising due to dynamism of pepstatin-A in the active site.

two monomers and also between the amide protons of E35 in the ^{15}N edited NOESY spectrum (Figure 7). In the crystal structure of the complex between acetyl pepstatin and homodimeric protease, the distance between the V82 amide protons is 17.5 Å and that between the amide protons of E35 is 29 Å. Therefore, an NOE cross peak between the two sets of amide protons is not possible. Similar NOE cross peaks were also observed in the case of the homodimer–KNI-529 (a variant of KNI-272) complex and the peaks were attributed to be originating from chemical exchange due to a flip-flop motion of the inhibitor. On the basis of the temperature dependence of the exchange rates in KNI-529 complex[47] and the measured enthalpy changes by calorimetric methods in the case of acetyl pepstatin complex[48], it was concluded that the inhibitor KNI-529 undergoes the flip-flop motion in the cavity of the enzyme without dissociating from the complex[47]. We believe that similar conclusions would be applicable to pepstatin-A as well, in the present situation.

The flip-flop motion and the associated conformational transitions would influence the widths of the resonances and could result in non-observation of some peaks in the spectra. Specifically, we could not observe the amide-^{15}N cross peaks of residues D25, D29, G49 and I50 in the HSQC spectra of the complex. The first two of these residues are located in the active site, while the other two are in the flap.

Concluding remarks

Structure, stability, inhibitor binding and dynamical characteristics of HIV-1 protease continue to draw wide attention in the literature[11,26,29,31,47–49,50–53]. While NMR is the most suited technique for such investigations, these studies have not been as extensive as the X-ray crystallographic studies. Further, from the point of view investigating the roles of individual monomer units in enzyme action and consequently from the viewpoint of drug design, the tethered dimers show great promise. They permit creation of monomer-specific mutations in the dimeric structure. In this context, we have described in this paper qualitative characterization of secondary structure and local stabilities in a tethered heterodimer of HIV-1 protease complexed to the general aspartyl protease inhibitor, pepstatin-A, by NMR spectroscopy. We observed that the mutation did not prevent proper dimer formation, but did influence the dynamical properties of the neighbouring residues. The NMR chemical shift indices indicated that the secondary structure elements of the heterodimer protein in the complex were similar to those in the crystal structure of the homodimer–acetyl pepstatin complex. Amide proton–deuterium exchange studies, in conjunction with solvent accessibility calculations indicated different local stabilities in different regions of the protein in the complex and also indicated differences in the two monomer units. The NOE data on the complex showed exchange cross peaks between the amide protons of V82 and E35 residues in the individual monomers and these were attributed to a flip-flop dynamic motion of the inhibitor in the complex. These observed dynamisms in the protein and the inhibitor, reflect on the ability of the protein to accommodate different types of inhibitors/substrates. Thus, one may speculate that for a drug binding at the active site to be efficient, it may be necessary not only to have a high binding affinity but also to induce high rigidity in the structure or it should be able to adjust itself to the demands of a mutated protein.

1. Seelmeier, S., Schmidt, H., Turk, V. and Helm, K. V. D., *Proc. Natl. Acad. Sci. USA*, 1988, **85**, 6612–6616.
2. Kohl, N. E., Emini, E. A., Schlief, W. A., Davis, L. J., Heimbach, J., Dixon, R. A. F., Scolnick, E. M. and Sigal, I. S., *Proc. Natl. Acad. Sci. USA*, 1988, **85**, 4686–4690.
3. McQuade, T. J., Tomasselli, G., Liu, L., Karacostas, V., Moss, B., Sawyer, T. K., Heinrkson, R. L. and Tarpley, W. G., *Science*, 1990, **247**, 454–456.
4. Lapatto, R., Blundell, T., Hemmings, A., Overington, J., Wilderspin, A. F., Wood, S., Merson, J. R., Whittle, P. J., Danley, D. E., Geoghegan, K. F., Hawrylik, S. J., Lee, S. E., Scheld, K. G. and Hobart, P. M., *Nature*, 1989, **342**, 299–302.
5. Navia, M. A., Fitzgerald, P. M. D., McKeever, B. M., Leu, C., Heimbach, J. C., Herber, W. K., Sigal, I. S., Darke, P. L. and Springer, J. P., *Nature*, 1989, **337**, 615–620.
6. Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B. K., Baldwin, E., Webber, I. T., Selk, L. M., Clawson, L., Schneider, J. and Kent, S. B. H., *Science*, 1989, **245**, 616–621.

7. Davies, D. R., *Annu. Rev. Biophys. Chem.*, 1990, **19**, 189–215.
8. Wlodawer, A. and Erickson, J. W., *Annu. Rev. Biochem.*, 1993, **62**, 543–585.
9. Clercq, E. D., *J. Med. Chem.*, 1995, **38**, 2491–2517.
10. Erickson, J. W. and Burt, S. K., *Annu. Rev. Pharmacol. Toxicol.*, 1996, **36**, 545–571.
11. Wlodawer, A. and Vondrasek, J., *Annu. Rev. Biophys. Biomol. Struct.*, 1998, **27**, 249–284.
12. Ridky, T. and Leis, J., *J. Biol. Chem.*, 1995, **270**, 29621–29623.
13. Condra, J. H., Schleif, W. A., Blahy, O. M., Gabryelski, L. J., Graham, D. J., Quintero, J. C., Rhodes, A., Robbins, H. L., Roth, E., Shivapraksh, M., Titus, D., Yang, T., Teppler, H., Squires, K. E., Deustsch, P. J. and Emini, E. A., *Nature*, 1995, **374**, 569–571.
14. Coffin, J. M., *Science*, 1995, **267**, 483–489.
15. Babé, L. M., Pichuantes, S. and Craik, C. S., *Biochemistry*, 1991, **30**, 106–111.
16. McPhee, F., Good, A. C., Kuntz, I. D. and Craik, C. S., *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 11477–11481.
17. Babé, L. M., Rosé, J. and Craik, C. S., *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 10069–10073.
18. Bhat, T. N., Baldwin, E. T., Liu, B., Chen, Y. E. and Erickson, J. W., *Natl. Struct. Biol.*, 1994, **1**, 552–556.
19. Cheng, Y. E., Yin, F. H., Foundling, S., Blomstrom, D. and Kettner, C. A., *Proc. Natl. Acad. Sci. USA*, 1990, **87**, 9660–9664.
20. Dilanni, C. L., Davis, L. J., Holloway, M. K., Herber, W. K., Darke, P. L., Kohl, N. E. and Dixon, R. A. F., *J. Biol. Chem.*, 1990, **265**, 17348–17354.
21. Kräusslich, H., *Proc. Natl. Acad. Sci. USA*, 1991, **88**, 3213–3217.
22. Patterson, C. E., Seetharam, R., Kettner, C. A. and Cheng, Y. E., *J. Virol.*, 1992, **66**, 1228–1231.
23. Griffiths, J. T., Tomchak, L. A., Mills, J. S., Graves, M. C., Cook, N. D., Dunn, B. M. and Kay, J., *J. Biol. Chem.*, 1994, **269**, 4787–4793.
24. Darke, P. L., Jordan, S. P., Hall, D. L., Zugay, J. A., Shafer, J. A. and Kuo, L. C., *Biochemistry*, 1994, **33**, 98–105.
25. Grant, S. K., Deckman, I. C., Culp, J. S., Minnich, M. D., Brooks, I. S., Hensley, P., Debouck, C. and Meek, T. D., *Biochemistry*, 1992, **31**, 9491–9501.
26. Louis, J. M., Clore, G. M. and Gronenborn, A. M., *Natl. Struct. Biol.*, 1999, **6**, 868–875.
27. Tomasselli, A. G., Hui, J. O., Sawyer, T. K., Staples, D. J., Bannow, C., Reardon, I. M., Howe, W. J., DeCamp, D. L., Craik, C. S. and Heinrkson, R. L., *J. Biol. Chem.*, 1990, **265**, 14675–14683.
28. Yamazaki, T., Nicholson, L. K., Torchia, D. A., Stahl, S. J., Kaufman, J. D., Wingfield, P. T., Domaille, P. J. and Campbell-Burk, S., *Eur. J. Biochem.*, 1994, **219**, 707–712.
29. Freedberg, D. I., Wang, Y., Stahl, S. J., Kaufman, J. D., Wingfield, P. T., Kiso, Y. and Torchia, D. A., *J. Am. Chem. Soc.*, 1998, **120**, 7916–7923.
30. Fitzgerald, P. M. D., McKeever, B. M., VanMiddlesworth, J. F., Springer, J. P., Heimbach, J. C., Leu, C., Herber, W. K., Dixon, R. A. F. and Darke, P. L., *J. Biol. Chem.*, 1990, **265**, 14209–14219.
31. Loo, J. A., Holler, T. P., Foltin, S. K., McConnell, P., Banotai, C. A., Horne, N. M., Mueller, W. T., Stevenson, T. I. and Mack, D. P., *Proteins: Struct. Funct. Genet.*, 1998, **2**, 28–37.
32. Todd, M. J., Semo, N. and Freire, E., *J. Mol. Biol.*, 1998, **283**, 475–488.
33. Bodenhausen, G. and Ruben, D. J., *Chem. Phys. Lett.*, 1980, **69**, 185–189.
34. Kay, L. E., Ikura, M., Tschudin, R. and Bax, A., *J. Magn. Reson.*, 1990, **89**, 496–514.
35. Grzesiek, S. and Bax, A., *J. Magn. Reson.*, 1992, **96**, 432–440.
36. Grzesiek, S. and Bax, A., *J. Magn. Reson.*, 1992, **99**, 201–207.
37. Grzesiek, S. and Bax, A., *J. Am. Chem. Soc.*, 1992, **114**, 6291–6293.
38. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M. and Clore, G. M., *Biochemistry*, 1989, **28**, 6150–6156.
39. Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A. and Bax, A., *J. Am. Chem. Soc.*, 1989, **111**, 1515–1517.
40. Bax, A. and Subramanian, S., *J. Magn. Reson.*, 1986, **67**, 565–569.
41. Lee, B. and Richards, F. M., *J. Mol. Biol.*, 1971, **55**, 379–400.
42. Wishart, D. S. and Sykes, B. D., *J. Biomol. NMR*, 1994, **4**, 171–180.
43. Hvidt, A. and Nielsen, S. O., *Adv. Protein. Chem.*, 1966, **21**, 287–386.
44. Bahar, I., Wallqvist, A., Covell, D. G. and Jernigan, R. L., *Biochemistry*, 1998, **37**, 1067–1075.
45. Sivaraman, T., Kumar, T. K. S. and Yu, C., *Biochemistry*, 1999, **28**, 9899–9905.
46. Beatrice, M. P., Huyghues-Despointes, B. M. P., Scholtz, J. M. and Pace, C. N., *Natl. Struct. Biol.*, 1999, **6**, 910–912.
47. Katoh, E., Yamazaki, T., Kiso, Y., Wingfield, P. T., Stahl, S. J., Kaufman, J. D. and Torchia, D. A., *J. Am. Chem. Soc.*, 1999, **121**, 2607–2608.
48. Luque, I., Todd, M. J., Gómez, J., Semo, N. and Freire, E., *Biochemistry*, 1998, **37**, 5791–5797.
49. Ishima, R., Wingfield, P. T., Stahl, S. J., Kaufman, J. D. and Torchia, D. A., *J. Am. Chem. Soc.*, 1998, **120**, 10534–10542.
50. Todd, M. J. and Freire, E., *Proteins: Struct. Funct. Genet.*, 1999, **36**, 147–156.
51. Ishima, R., Freedberg, D. I., Wang, Y., Louis, J. M. and Torchia, D. A., *Structure*, 1999, **7**, 1047–1055.
52. Tawa, G. J., Topol, I. A., Burt, S. K. and Erickson, J. W., *J. Am. Chem. Soc.*, 1998, **120**, 8856–8863.
53. Persichini, T., Colasanti, M., Lauro, G. M. and Ascenzi, P., *Biochem. Biophys. Res. Commun.*, 1998, **250**, 575–576.

ACKNOWLEDGEMENTS. The facilities provided by the National Facility for High Field NMR, and at the Tata Institute of Fundamental Research, Mumbai, supported by the Government of India are gratefully acknowledged. The clone of the tethered dimer was a kind gift from Dr John W. Erickson and Beishan Li of NCI-Frederick Cancer Research and Development Center, Frederick, Maryland 21702, USA.

Received 23 June 2000; revised accepted 4 November 2000