J. Mol. Biol. (1991) **221**, 1081–1090

Secondary Structure of the Ribonuclease H Domain of the Human Immunodeficiency Virus Reverse Transcriptase in Solution using Three-Dimensional Double and Triple Resonance Heteronuclear Magnetic Resonance Spectroscopy

Robert Powers[1], G. Marius Clore[1]†, Ad Bax[1], Daniel S. Garrett[1], Stephen J. Stahl[2] Paul T. Wingfield[2] and Angela M. Gronenborn[1]†

[1]*Laboratory of Chemical Physics, Building 2*
National Institute of Diabetes and Digestive and Kidney Diseases
National Institutes of Health, Bethesda, MD 20892, U.S.A.

[2]*Protein Expression Laboratory, Building 6B*
Office of the Director, National Institutes of Health
Bethesda, MD 20892, U.S.A.

(Received 13 June 1991; accepted 11 July 1991)

The solution structure of the ribonuclease H domain of HIV-1 reverse transcriptase has been investigated by three-dimensional double and triple resonance heteronuclear magnetic resonance spectroscopy. The domain studied has 138 residues and comprises residues 427 to 560 of the 66 kDa reverse transcriptase with an additional four residues at the N terminus. Initial studies on the wild-type protein were hindered by severe differential line broadening, presumably due to conformational averaging. Mutation of the single tryptophan residue located in a loop at position 113 (position 535 in the reverse transcriptase sequence) to an alanine resulted in much improved spectral properties with no apparent change in structure. ^{1}H, ^{15}N and ^{13}C backbone resonances were assigned sequentially using a range of three-dimensional double and triple resonance heteronuclear experiments on samples of uniformly (>95%) ^{15}N and $^{15}N/^{13}C$-labeled protein, and the secondary structure was elucidated from a qualitative analysis of data derived from three-dimensional ^{15}N- and ^{13}C-edited nuclear Overbauser enhancement spectra. The secondary structure comprises three α-helices and five strands arranged in a mixed parallel/antiparallel β-sheet with a +1, +1, −3x, −1x topology. The C-terminal region from residue 114 onwards appears to be conformationally disordered in solution as evidenced by an almost complete absence of sequential and medium range nuclear Overhauser effects.

Key words: HIV-1; RNase H domain; reverse transcriptase; solution secondary structure; 3D heteronuclear n.m.r.; double and triple resonance n.m.r.

The ribonuclease H (RNase H‡) domain of human immunodeficiency virus (HIV-1) reverse transcriptase plays a crucial role in viral replication as evidenced by the failure of mutant provirus, defective for RNase H function, to produce infective virus particles (Schatz *et al.*, 1989). The RNase H domain catalyses the cleavage of the RNA portion of a DNA/RNA hybrid, a process that is

† Authors to whom all correspondence should be addressed.

‡ RNase H, the ribonuclease H domain comprising residues 427 to 560 of the 66 kDa reverse transcriptase of HIV-1 together with the four amino acid sequence Met-Asn-Glu-Leu at the N terminus; HIV-1, human immunodeficiency virus-1; AIDS, acquired immunodeficiency syndrome; n.m.r. nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser enhancement spectroscopy; HOHAHA, homonuclear Hartmann-Hahn spectroscopy; 3D, three-dimensional; HNCO, amide proton to nitrogen to carbonyl correlation; HNCA, amide proton to nitrogen to α-carbon correlation; HN(CO)CA, amide proton to nitrogen (*via* carbonyl) to α-carbon correlation; HCACO, alpha-proton to α-carbon to carbonyl correlation; HCA(CO)N, α-proton to α-carbon (*via* carbonyl) to nitrogen correlation; c.d., circular dichroism; p.p.m., parts per million; 2D, two-dimensional.

0022-2836/91/201081-10 $03.00/0

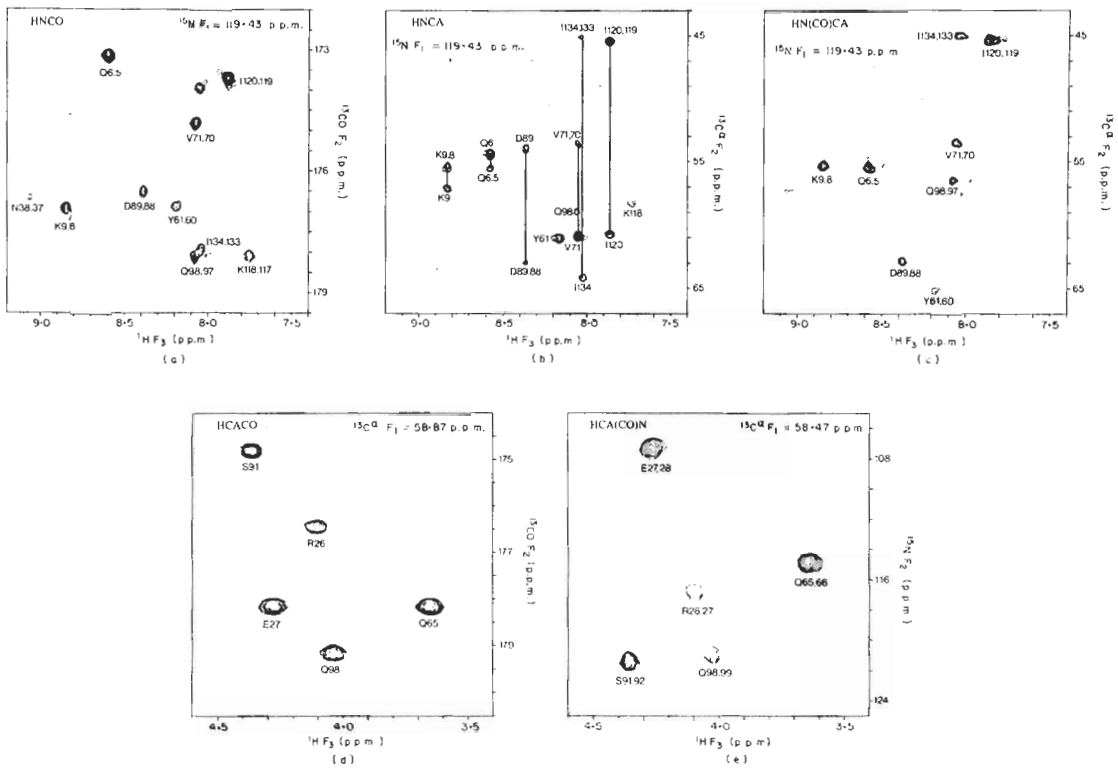
HNCO
HNCA
HN(CO)CA
HCACO
HCA(CO)N
(a)
(b)
(c)
(d)
(e)

Figure 1. Selected F_2-F_3 planes for the 3D HNCO (a), HNCA (b), HN(CO)CA (c), HCACO (d) and HCA(CO)N (e) triple resonance spectra of a sample of uniformly (>95%) $^{15}N/^{13}C$-labeled Trp113→Ala RNase H from HIV-1 at 26°C. All spectra were recorded at 600 MHz on a Bruker AM600 spectrometer equipped with a triple resonance probe and modified with additional hardware, as described by Kay *et al.*, (1990*a*). The HNCO and HNCA experiments were recorded as described by Kay *et al.* (1990*a*), the HN(CO)CA experiment as described by Ikura & Bax (1991), and the HCACO and HCA(CO)N experiments as described by Powers *et al.* (1991). The latter 2 experiments employ a constant time evolution of the $^{13}C^{\alpha}$ magnetization in the F_1 dimension, which results in in-phase $^{13}C^{\alpha}$ signals. The Trp113→Ala mutant of the RNase H domain was purified as described by Becerra *et al.* (1990) and the sample contained 1·1 mM-Trp113→Ala RNase H in 100 mM-sodium phosphate (pH 5·4), dissolved in 90% H_2O/10% ^2H_2O for the HNCO, HNCA and HN(CO)CA experiments and in 99·996% ^2H_2O for the HCACO and HCA(CO)N experiments. The spectral widths in the ^{15}N, $^{13}C^{\alpha}$ and ^{13}CO dimensions were 29·16 p.p.m., 33·13 and 12·05 p.p.m., respectively, with the carrier positions placed at 118·5 p.p.m., 56 p.p.m. and 177 p.p.m., respectively. The spectral width in the ^1H dimension was 13·44 p.p.m. with the carrier at 4·76 p.p.m. for the HNCO, HNCA and HN(CO)CA experiments. For the HCACO and HCA(CO)N experiments, the ^1H spectral width was 8·33 p.p.m. with the carrier at 4·76 p.p.m. For the HNCO, HNCA and HN(CO)CA experiments the number of points acquired in the various dimensions was 32 complex in F_1 (^{15}N), 64 complex in F_2 (^{13}CO or $^{13}C^{\alpha}$) and 1024 real in F_3 (^1H). For the HCACO and HCA(CO)N experiments, the number of points acquired was 32 complex in F_1 ($^{13}C^{\alpha}$), 64 complex in F_2 (^{13}CO) for the HCACO experiment and 32 complex in F_2 (^{15}N) for the HCA(CO)N experiment, and 512 real in F_3. Zero-filling was used in all dimensions, and in the case of the HCACO and HCA(CO)N experiment, linear prediction by means of the mirror image technique (Zhu & Bax, 1990) was used to further extend the data. The final 3D spectra consisted of $64 \times 128 \times 1024$ data points for the HNCO, HNCA and HN(CO)CA experiments, and $128 \times 128 \times 512$ points for the HCACO and HCA(CO)N experiments. All the spectra in this Figure, as well as in Fig. 2, were processed on a Sun Sparc Workstation using in-house routines for Fourier transformation (Kay *et al.*, 1989) and linear prediction (Zhu & Bax, 1990), together with the commercially available software package NMR2 (New Methods Research, Inc, Syracuse, NY). Analysis of the 3D spectra and peak picking was carried out using the in-house programs CAPP and PIPP (Garrett *et al.*, 1991).

required at several stages during reverse transcription, and displays both endonuclease and 3′→5′ exonuclease activity (Krug & Berger, 1989; Mizrahi, 1989; Schatz *et al.*, 1990). Thus, the RNase H domain presents a potential site for the design of drugs for the treatment of AIDS (Mitsuya *et al.*, 1990). In a recent paper we described the overexpression, purification and physical characterization of the RNase H domain comprising residues 427 to 560 of the 66 kDa reverse transcriptase with an additional four residue sequence at the N terminus (Becerra *et al.*, 1990). In this paper we present initial multi-dimensional heteronuclear studies on the RNase H domain undertaken with the eventual aim of determining its high-resolution three-dimensional structure in solution. Specifically, the backbone ^{1}H, ^{15}N and ^{13}C resonances are assigned in a sequential manner using a combination of 3D double and triple resonance heteronuclear n.m.r. experiments, and the secondary structure is elucidated from a qualitative analysis of NOE connectivities derived from 3D heteronuclear-edited NOESY spectra (for reviews, see Clore & Gronenborn 1991*a*,*b*,*c*). As the work presented in this paper was being prepared for publication, a 2·4 Å (1 Å = 0·1 nm) resolution X-ray structure of RNase H from a different HIV-1 virus strain was published (Davies *et al.*, 1991). The results from both studies are essentially in agreement, although significant differences are noted at the C terminus.

Initially n.m.r. studies were carried out on the wild-type RNAse H domain from strain HXB2 of HIV-1. However, it rapidly became apparent that the assignment procedure was being impaired by suspected conformational exchange processes. Thus, we were unable to see many connectivities in the various 3D triple resonance experiments due to severe line broadening. This was also manifested in the 3D ^{15}N-edited NOESY (Fesik & Zuiderweg, 1988; Marion *et al.*, 1989*a*,*b*) and HOHAHA (Marion *et al.*, 1989*b*; Driscoll *et al.*, 1990*a*; Clore *et al.*, 1991) spectra, as well as in the 2D ^{1}H–^{15}N and ^{1}H–^{13}C correlation spectra (Bodenhausen & Ruben, 1980; Bax *et al.*, 1990*a*). In the latter spectra, a large variation in cross-peak intensities occurred. At a very early stage of this work, we found that, under all conditions tried, we did not observe the ^{1}H–^{13}C correlations for the aromatic ring of the single Trp residue at position 113, while those for the other eight aromatic rings (6 Tyr, 1 Phe and 1 His) were clearly detectable. We therefore postulated that Trp113 was located in a segment of the polypeptide chain that exhibited conformational flexibility, resulting in severe line broadening of adjacent protons through large differences in ring current shifts between the conformers. On the basis of this hypothesis, we proceeded to construct a Trp113→Ala mutant by primer-directed mutagenesis (Oostra *et al.*, 1983). No difference in stability of the native and mutant proteins was found, both of which had a T_m of ~60°C (determined by differential scanning calorimetry), and the location of many of the cross-peaks in the ^{1}H–^{15}N correlation spectra of the two proteins were identical. Further, there was little difference in the c.d. spectrum of the two proteins. This implied that the mutation caused only minimal structural perturbation, which would be entirely consistent with the predicted location of Trp113 on the protein surface based on sequence alignment with the known crystal structure of *Escherichia coli* RNase H (Yang *et al.*, 1990; Katayanagi *et al.*, 1990). However, the spectra of the mutant enzyme were both qualitatively and quantitatively vastly superior to those of the wild-type protein, so that further detailed study was restricted to the mutant.

The sequential assignment strategy was based on a series of 3D double and triple resonance n.m.r. experiments. In particular, we made use of five triple resonance experiments to establish connectivities along the chain *via* one- and two-bond heteronuclear couplings. The 3D HNCO, HNCA (Ikura *et al.*, 1990*a*) and HN(CO)CA (Ikura & Bax, 1991) experiments recorded in water were used to establish $NH(i)$-$^{15}N(i)$-$^{13}CO(i-1)$, $NH(i)$-$^{15}N(i)$-$^{13}C^{\alpha}(i,i-1)$ and $NH(i)$-$^{15}N(i)$-$^{13}C^{\alpha}(i-1)$ correlations, respectively. $^{13}C(F_2)$–$NH(F_3)$ planes of these three experiments at a single $^{15}N(F_1)$ frequency of 119·43 p.p.m. are illustrated in Figure 1(a) to (c). As the ^{15}N–$^{13}C^{\alpha}$ intraresidue one-bond coupling is larger than the interresidue two-bond coupling, it is usually the case that the intraresidue correlations in the HNCA experiment are more intense than the interresidue ones (Clore *et al.*, 1990; Kay *et al.*, 1990*b*). However, in the case of RNase H where complications arise from exchange line broadening this rule is not generally applicable. Thus, for example, the interresidue correlations for Lys9 and Ile120 seen in Figure 1(a) are actually more intense than the intraresidue ones. Possible ambiguities are resolved by analysis of the HN(CO)CA spectrum, which only displays the interresidue $^{13}C^{\alpha}(i-1)$-$^{15}N(i)$ correlations. The 3D HCACO and HCA(CO)N experiments (Ikura *et al.*, 1990*a*; Powers *et al.*, 1991) recorded in $^{2}H_2O$ were used to establish $C^{\alpha}H(i)$-$^{13}C^{\alpha}(i)$-$^{13}CO(i)$ and $C^{\alpha}H(i)$-$^{13}C^{\alpha}(i)$-$^{15}N(i+1)$ correlations, respectively. These experiments are illustrated in Figure 1(d) to (e), which show a set of $^{13}CO(F_2)$–$C^{\alpha}H(F_3)$ and $^{15}N(F_2)$–$C^{\alpha}H(F_3)$ planes at the same $^{13}C(F_1)$ frequency of 58·87 p.p.m. In addition, a 3D ^{15}N-edited HOHAHA spectrum recorded in H_2O with a DIPSI-2 mixing scheme (Clore *et al.*, 1991) was used to identify $^{15}N(i)$–$NH(i)$-$C^{\alpha}H(i)$ correlations (Fig. 2(a)). Interpretation of these six scalar correlation experiments is sufficient to sequentially assign the backbone ^{1}H, ^{15}N and ^{13}C resonances. Confirmation of the sequential assignment was based on 3D ^{15}N and ^{13}C-edited (Ikura *et al.*, 1990*b*; Zuiderweg *et al.*, 1990) NOESY spectroscopy to identify through-space (<5 Å) connectivities of the type $NH(i)$–$NH(i+1,2)$ ($C^{\alpha}H(i)$–$NH(i+1,2,3,4)$, $C^{\beta}H(i)$–$NH(i+1)$, $C^{\alpha}H(i)$–$C^{\beta}H(i+3)$, which have been extensively used in conventional protein resonance assignment by 2D methods (Wüthrich, 1986; Clore & Gronenborn, 1987). Examples of selected amide strips through

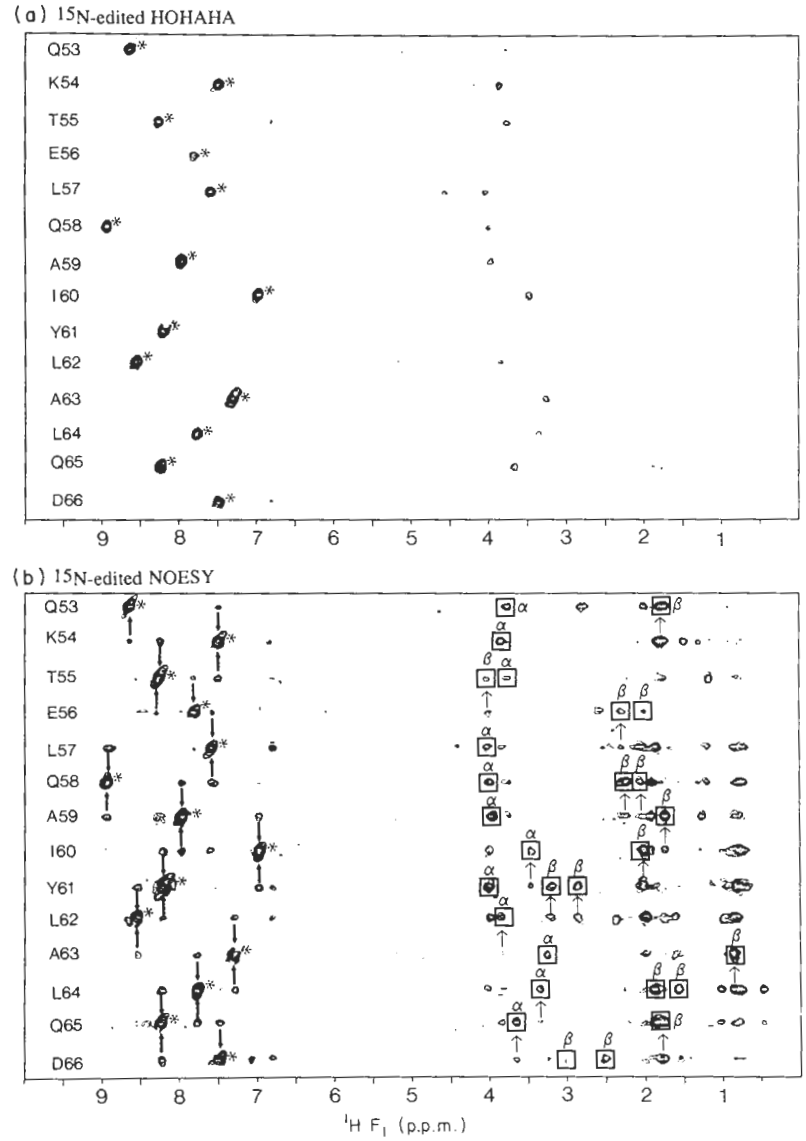

Figure 2. Amide strips extending from Gln53 to Asp66 taken from the 600 MHz 3D ^{15}N-edited HOHAHA (a) and NOESY (b) spectra of uniformly (>95%) ^{15}N-labeled Trp113→Ala RNase H from HIV-1 recorded at 26°C. The mixing times for the 2 experiments were 30 ms and 100 ms, respectively, and the experiments were recorded as described by Clore *et al.* (1991) and Driscoll *et al.* (1990*a*), respectively. The Figure is composed of narrow strips taken from different ^{1}H(F_1)-NH(F_3) planes of the 3D spectrum, as described by Driscoll *et al.* (1990*a*). Asterisks indicate the position of the diagonal peak for each residue and boxes enclose the intraresidue NH–C$^\alpha$H and NH–C$^\beta$H cross-peaks observed in the NOESY spectrum. Sequential NH-NH($i+1$) NOEs are indicated by thick arrows, sequential C$^\alpha$H(i)-NH($i+1$) and C$^\beta$H(i)-NH($i+1$) NOEs with thin arrows. Note that while the NH–C$^\alpha$H cross peaks in the HOHAHA spectrum are observed for all the residues in this Figure (with the exception of Glu56, Tyr61 and Asp66), NH–C$^\beta$H cross-peaks are only seen for Gln65 and Asp66 and these are of very weak intensity. The spectral widths in the F_1 (^{1}H), F_2 (^{15}N) and F_3 (^{1}H) dimensions were 11·41 p.p.m., 29·16 p.p.m. and 13·44 p.p.m. with the carrier placed at 118·5 p.p.m. in the ^{15}N dimension and at 4·76 p.p.m. in the ^{1}H dimensions. The total number of points acquired were 128 complex in F_1 (^{1}H), 32 complex in F_2 (^{15}N) and 1024 real in F_3 (^{1}H), and the final absorptive part of the 3D spectra, after appropriate zero-filling, consisted of 256 × 64 × 1024 points.

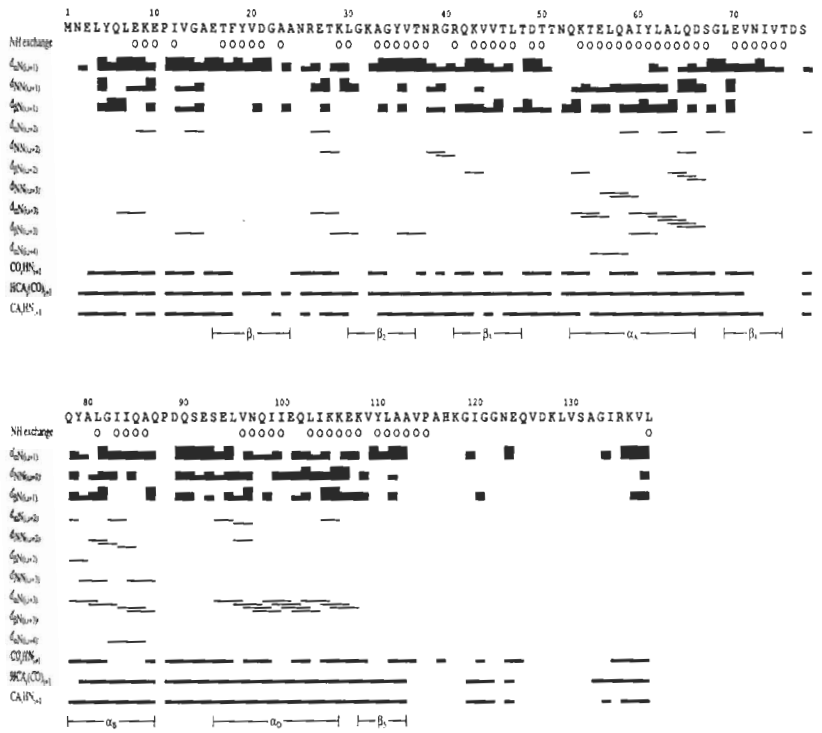

Figure 3. Summary of the sequential scalar connectivities observed in the 3D triple resonance experiments and the sequential and medium (up to i, $i+4$) range through-space connectivities observed in the 3D heteronuclear-edited NOESY spectra for Trp113→Ala RNase H from HIV-1. The NOE intensities are classified into strong, medium and weak, according to the thickness of the lines. Slowly exchanging amide protons are indicated by open circles, and the elements of regular secondary structure are displayed at the bottom of the Figure. The sequence of the RNase H domain starts at Tyr5, which corresponds to residue 427 of the 66 kDa HIV-1 reverse transcriptase.

the ^{15}N-edited NOESY spectrum (Driscoll *et al.*, 1990*a*,*b*) are shown in Figure 2(b), which displays, the sequential NOE connectivities observed along the segment of chain extending from Gln53 to Asp66, corresponding to helix α_A. Finally, spin systems were identified, where possible, using 3D HCCH-COSY and HCCH-TOCSY experiments that establish scalar connectivities along the chain *via* the one-bond ^{1}H–^{13}C and ^{13}C–^{13}C couplings (Bax *et al.*, 1990*b*,*c*; Clore *et al.*, 1990; Kay *et al.*, 1990*a*).

A summary of the sequential scalar and the sequential and medium range (up to i, $i+4$) NOE connectivities observed for the Trp113→Ala mutant of RNase H domain is shown in Figure 3, and the backbone ^{1}H, ^{15}N and ^{13}C assignments are given in Table 1. We were able to confidently obtain backbone resonance assignments for 123 of the 138 residues for the mutant protein compared to 110 tentative assignments for the wild-type. The number of residues for which a complete set of correlations was observed in the triple resonance experiments and the ^{15}N-edited HOHAHA spectrum was 60 for the mutant compared to only 31 for the wild-type. Further NOEs involving 113 NH protons could be assigned in the 3D ^{15}N-edited NOESY spectrum of the mutant compared to only 86 for the wild-type. It is also interesting to note that with the exception of Met1 and Asp76, almost all residues in the mutant protein for which no backbone assignments could be ascertained are located in the C terminus, whereas in the wild-type unassigned residues are spread throughout the sequence.

The secondary structure was deduced from a qualitative analysis of the NOE data involving the backbone NH, C$^\alpha$H and C$^\beta$H protons (Wüthrich, 1986; Clore & Gronenborn, 1987) derived from the 3D ^{15}N- and ^{13}C-edited NOESY spectra, in conjunction with data on slowly exchanging amide protons (summarized in Fig. 3). The latter was

Table 1

Backbone ^{1}H, ^{13}C and ^{15}N resonance assignments for the RNase H domain of HIV-1 reverse transcriptase at pH 5·4 and 26°C

	Chemical shift (p.p.m.)†				
	^{15}N	$^{13}C^{\alpha}$	^{13}CO	NH	$C^{\alpha}H$
Met1	—	—	—	—	—
Asn2	—	53·6	176·5	—	5·14
Glu3	114·8	56·8	176·5	8·42	4·22
Leu4	121·5	55·8	176·1	8·49	4·28
Tyr5	113·9	55·7	173·1	7·18	4·71
Gln6	119·3	54·5	174·7	8·58	4·55
Leu7	123·9	53·9	178·0	8·55	4·93
Glu8	121·5	55·7	176·9	9·06	4·29
Lys9	119·4	57·3	176·0	8·84	4·44
Glu10	116·8	53·0	171·8	7·55	4·59
Pro11	—	62·4	175·7	—	—
Ile12	125·0	61·2	178·3	9·43	3·88
Val13	132·3	65·1	177·4	8·69	3·65
Gly14	114·4	45·1	173·4	8·79	—
Ala15	121·6	50·6	176·5	6·96	4·57
Glu16	125·7	57·6	174·7	8·97	4·56
Thr17	121·7	61·4	172·6	8·56	5·33
Phe18	127·4	55·6	—	9·41	4·34
Tyr19	124·7	57·1	176·0	9·19	5·18
Val20	116·0	59·1	175·0	8·40	5·60
Asp21	119·0	53·2	172·6	8·55	5·05
Gly22	108·6	45·3	—	8·28	4·94, 3·50
Ala23	122·9	52·2	175·1	8·52	4·64
Ala24	120·5	50·4	175·2	8·52	5·09
Asn25	120·7	52·5	176·6	—	4·83
Arg26	127·6	58·8	176·5	9·10	4·15
Glu27	117·4	58·6	178·3	8·41	4·34
Thr28	107·6	61·8	176·0	8·06	4·29
Lys29	115·4	58·0	173·4	8·08	3·89
Leu30	118·3	54·2	177·1	7·45	4·86
Gly31	111·9	44·8	—	8·02	5·03, 3·61
Lys32	—	55·0	174·0	—	5·46
Ala33	121·6	50·0	175·5	8·49	5·27
Gly34	106·0	4[illegible]	182·4	8·70	4·40, 4·32
Tyr35	110·9	5[illegible]	174·2	8·48	6·04
Val36	115·2	60·9	175·5	8·69	5·24
Thr37	114·5	57·5	176·6	8·94	6·28
Asn38	119·0	54·2	175·4	9·05	4·62
Arg39	118·0	55·2	176·2	8·12	4·68
Gly40	107·4	45·6	174·3	7·87	4·21, 3·82
Arg41	117·7	56·3	175·1	7·20	4·71
Gln42	117·7	54·9	173·8	8·50	5·22
Lys43	123·2	56·5	172·7	7·77	4·23
Val44	123·1	60·8	174·5	8·15	4·90
Val45	123·6	59·7	174·4	9·26	4·69
Thr46	117·9	62·0	173·8	8·41	4·96
Leu47	127·7	53·8	174·9	8·92	4·78
Thr48	111·3	60·2	173·5	8·13	4·67
Asp49	121·9	55·0	174·9	8·57	4·32
Thr50	114·4	59·8	173·6	8·91	4·88
Thr51	108·9	58·6	174·9	8·18	4·63
Asn52	118·5	57·8	177·0	8·14	4·15
Gln53	118·1	59·9	177·7	8·59	3·74
Lys54	116·7	60·6	179·4	7·44	3·88
Thr55	111·5	66·7	176·6	8·21	3·79
Glu56	121·3	59·6	179·8	7·79	3·94
Leu57	117·6	57·5	178·6	7·53	4·06
Gln58	120·4	58·6	177·7	8·90	4·02
Ala59	120·6	55·8	178·4	7·93	3·99
Ile60	115·2	65·5	176·9	6·91	3·49
Tyr61	119·4	61·2	176·8	8·17	4·03
Leu62	118·5	57·7	177·9	8·51	3·84
Ala63	117·6	53·8	179·3	7·24	3·27
Leu64	116·8	57·6	177·6	7·72	3·37
Gln65	117·5	59·0	178·3	8·21	3·69
Asp66	115·4	54·5	175·1	7·45	4·74
Ser67	112·9	56·1	175·7	7·05	4·54
Gly68	104·6	43·8	173·8	8·32	—
Leu69	117·7	57·6	176·4	8·35	4·07

Table 1 *(continued)*

	Chemical shift (p.p.m.)†				
	^{15}N	$^{13}C^{\alpha}$	^{13}CO	NH	$C^{\alpha}H$
Glu70	115·2	53·7	174·8	7·57	5·83
Val71	119·6	61·1	169·9	8·06	4·64
Asn72	122·8	51·8	174·0	7·92	5·93
Ile73	125·2	61·1	—	9·75	—
Val74	126·2	60·6	175·4	8·93	5·00
Thr75	120·4	58·1	171·6	8·76	5·18
Asp76	—	—	—	—	—
Ser77	—	56·9	174·8	—	4·63
Gln78	128·6	58·2	—	8·70	3·96
Tyr79	120·3	60·2	176·9	8·40	4·26
Ala80	121·6	54·7	178·8	8·40	3·59
Leu81	114·9	57·9	178·6	7·54	3·80
Gly82	104·5	46·7	176·6	7·90	3·80
Ile83	120·7	62·9	177·3	7·39	3·84
Ile84	117·5	64·2	178·0	7·39	3·65
Gln85	116·7	57·4	176·2	8·03	4·42
Ala86	121·0	52·5	176·4	7·38	4·35
Gln87	115·9	55·0	173·1	7·95	4·40
Pro88	—	63·4	176·5	—	4·43
Asp89	119·0	54·1	176·3	8·37	4·61
Gln90	118·5	56·3	175·7	8·21	4·32
Ser91	115·5	59·0	174·9	8·50	4·40
Glu92	122·2	56·6	176·3	8·34	4·39
Ser93	114·4	58·5	175·1	7·98	4·31
Glu94	126·7	59·6	178·0	8·76	4·10
Leu95	120·0	57·4	178·5	7·84	4·10
Val96	117·9	67·6	177·4	7·68	3·37
Asn97	117·2	56·8	178·1	8·00	4·40
Gln98	119·9	59·1	179·2	8·07	4·07
Ile99	121·7	66·3	177·3	8·06	3·52
Ile100	121·3	65·9	177·1	8·53	3·36
Glu101	116·7	59·6	179·2	7·91	3·93
Gln102	114·5	57·7	178·6	7·46	3·94
Leu103	120·9	58·6	179·3	8·42	3·58
Ile104	115·3	64·7	176·6	7·91	3·79
Lys105	116·6	57·1	177·6	6·97	4·22
Lys106	116·5	54·3	176·8	7·19	4·27
Glu107	120·8	58·5	177·7	9·03	4·37
Lys108	114·8	56·1	174·1	7·69	4·93
Val109	121·2	60·3	172·7	9·07	4·98
Tyr110	129·6	56·3	172·2	8·44	4·80
Leu111	129·1	52·8	174·0	7·93	5·31
Ala112	125·1	50·9	174·2	8·38	4·32
Ala113	120·8	49·6	177·8	8·21	5·50
Val114	—	—	—	—	—
Pro115	—	—	—	—	—
Ala116	—	—	—	—	—
His117	—	—	—	—	—
Lys118	119·4	58·5	178·2	7·74	4·06
Gly119	—	45·4	—	—	4·03, 3·83
Ile120	119·5	60·9	176·8	7·87	4·15
Gly121	113·8	45·6	175·0	8·66	3·93
Gly122	108·8	—	—	8·13	—
Asn123	—	53·8	174·2	—	4·61
Glu124	120·5	57·0	175·6	8·57	4·24
Gln125	—	—	—	—	—
Val126	—	—	—	—	—
Asp127	—	—	—	—	—
Lys128	—	—	—	—	—
Leu129	—	—	—	—	—
Val130	—	—	—	—	—
Ser131	—	—	—	—	—
Ala132	—	—	—	—	—
Gly133	110·0	45·2	178·0	8·59	3·95, 3·80
Ile134	119·7	64·3	177·0	8·03	4·22
Arg135	114·0	55·8	175·4	7·10	4·40
Lys136	122·7	56·3	175·7	8·25	4·35
Val137	122·3	62·4	174·9	8·15	4·11
Leu138	130·6	56·6	—	7·83	4·20

† ^{1}H and ^{13}C chemical shifts are expressed relative to (trimethylsilyl)-propionic-d_4 acid, and ^{15}N shifts relative to liquid NH_3.

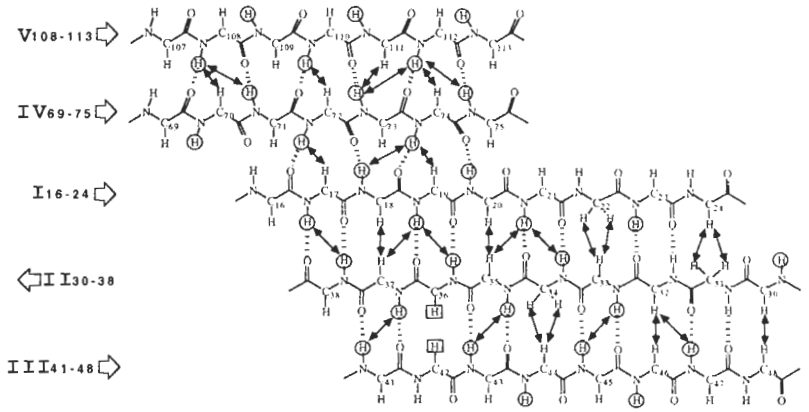

Figure 4. β-Sheet structure of Trp113→Ala RNase H from HIV-1 as determined from a qualitative analysis of NOE and amide exchange data. The β-strands are indicated on the left by roman numerals and the residue number range. Long-range NOEs are indicated by arrows, and hydrogen bonds derived from the NH/solvent exchange and NOE data are shown as broken lines. The slowly exchanging amide protons are encircled. The ^{1}H chemical shifts of the C^2H protons of Val36 and Glu42 are degenerate and indicated by a square box around them.

obtained by recording a series of ^{15}N–^1H Overbodenhausen correlation spectra (Bax *et al.*, 1990*a*) over a period of 80 hours starting within five minutes of dissolving an unexchanged sample of lyophilized protein in ^2H_2O (Driscoll *et al.*, 1990*b*). We find three clearly defined helices, which extend from residues 52 to 66 (α_A), 78 to 87 (α_B) and 93 to 106 (α_D), and five β-strands from residues 16 to 24 (strand I), 30 to 38 (strand II), 41 to 48 (strand III), 69 to 75 (strand IV) and 108 to 113 (strand V). These five strands form a mixed parallel–antiparallel β-sheet, which is illustrated in Figure 4. Strands I, II and III are antiparallel and connected *via* a short loop and β-turn, respectively. Strand I is parallel to strand IV, which in turn is parallel to strand V. The connecting element between strands III and IV is formed by α-helix α_A while that between strands IV and V is formed by the other two helices, α_B, and α_D. Thus, using the notation of Richardson (1981), the topology of the sheet can be described as $+1, +1, -3x, -1x$. It is interesting to note that the slowest exchanging amide protons are located in strands I, IV and V and helices α_A and α_D. Further, while there is an extensive network of NOE connectivities between strands I, II and III and between strands IV and V, the number of NOE connectivities between strands I and IV is very limited and their intensities are weak. The latter may be due to fraying at the ends of these two strands, accompanied by conformational mobility. Taken together, these observations suggest that β-strands IV and V and the connecting α-helices α_A and α_D constitute the most tightly folded portion of the RNase H structure.

The last well-defined residue in solution is Ala113, and the C-terminal region from Val114 to Gly133 appears to be conformationally disordered as evidenced by an almost complete absence of sequential and medium range NOEs (Fig. 3). In the X-ray structure residues 116 to 120 are also disordered and not visible in the electron density map. However, there is a helix (α_E) from Gly121 to Ser131 in the X-ray structure, which is absent in solution. This is probably due to greater conformational flexibility in the solution state and implies that this helix is not required to stabilize the overall polypeptide fold.

It should be noted that carboxy-terminal proteolytic cleavage of the RNase H domain results in a polypeptide approximately 3000 Da smaller than the native form (P. T. Wingfield, A. M. Gronenborn, G. M. Clore, unpublished results). This shortened cleavage product no longer exhibits a folded structure, as judged by both c.d. and n.m.r. spectra, which are characteristic of a random coil conformation. The most likely location of the cleavage site is in the sequence Lys105-Lys106-Glu107-Lys108, located at the beginning of β-strand V. Thus, removal of part or all of β-strand V destroys the folded form of RNase H, which is consistent with the above interpretation.

With the exception of the C-terminal helix α_E, which is not observed in solution, the secondary structure found in the liquid and crystal states are in agreement. It should be pointed out that the RNase H domains studied by n.m.r. and crystallography were derived from two different HIV-1 strains, HXB2 and BH10, respectively, and differ by several amino acid substitutions (i.e. Ile46→Pro, Asp49→Asn, Gln90→Lys and Val137→Ile). In addition, the sequence of the protein used for the n.m.r. studies contains the engineered Trp113→Ala mutation and the tetrapeptide Met-Asn-Glu-Leu

appended at the N terminus, while these four residues in the crystallized protein are Tyr-Ala-Ser-Arg. With the exception of the Ile46→Pro change, which is located at the end of strand III, the differences are conservative, located in loops or at the disordered N and C termini, and clearly do not affect the structure in any significant manner.

Comparison of the secondary structure of HIV-1 RNase H with that of *E. coli* (Yang *et al.*, 1990; Katayanagi *et al.*, 1990; Yamazaki *et al.*, 1991) is also of interest. There is an almost perfect match of the five β-strands and helices α_A, α_B and α_D between HIV-1 and *E. coli* RNase H. The α_B helix in the HIV-1 RNase H is about two to three residues longer than that of the corresponding helix in the *E. coli* enzyme, and the α_C helix and the subsequent 12-residue loop that serve to connect helices α_B and α_D in the *E. coli* protein are replaced by a short loop comprising residues 88 to 92 in the HIV-1 enzyme.

In summary we have made use of 3D double and triple resonance heteronuclear n.m.r. spectroscopy to obtain ^1H, ^{15}N and ^{13}C backbone sequential assignments for the RNase H domain of HIV-1 reverse transcriptase. The observed secondary structure in solution is consistent with that found in the crystal structure but there appears to be substantially more disorder at the C terminus in the solution structure. Further, extensive line broadening in the n.m.r. spectra of the wild-type protein is strongly suggestive of extensive conformational heterogeneity, which is in part alleviated by the mutation of Trp113 to Ala.

This work was supported by the AIDS Targeted Anti-viral Program of the Office of the Director of the National Institutes of Health. We thank Joshua Kaufman and Ira Palmer for technical assistance with the protein purification, and Joseph Shiloah for fermentations.

References

Bax, A., Ikura, M., Kay, L. E., Torchia, D. A. & Tschudin, R. (1990*a*). Comparison of different modes of two-dimensional reverse correlation NMR for the study of proteins. *J. Magn. Reson.* **86**, 304–318.

Bax, A., Clore, G. M., Driscoll, P. C., Gronenborn, A. M., Ikura, M. & Kay, L. E. (1990*b*). Practical aspects of proton-carbon-carbon-proton three-dimensional correlation spectroscopy of ^{13}C-labeled proteins. *J. Magn. Reson.* **87**, 620–628.

Bax, A., Clore, G. M. & Gronenborn, A. M. (1990*c*). ^1H-^1H correlation *via* isotropic mixing of ^{13}C magnetization: a new three-dimensional approach for assigning ^1H and ^{13}C spectra of ^{13}C-enriched proteins. *J. Magn. Reson.* **88**, 425–431.

Becerra, S. P., Clore, G. M., Gronenborn, A. M., Karlström, A., Stahl, S. J., Wilson, S. H. & Wingfield, P. T. (1990). Purification and characterization of the RNase H domain of HIV-1 reverse transcriptase expressed in recombinant *Escherichia coli*. *FEBS Letters*, **270**, 76–80.

Bodenhausen, G. & Ruben, D. J. (1980). Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Letters*, **69**, 185–189.

Clore, G. M. & Gronenborn, A. M. (1987). Determination of three-dimensional structures of proteins in solution by nuclear magnetic resonance spectroscopy. *Protein Engineer.* **1**, 275–288.

Clore, G. M. & Gronenborn, A. M. (1991*a*). Structures of larger proteins in solution: three- and four-dimensional heteronuclear NMR spectroscopy. *Science*, **252**, 1390–1399.

Clore, G. M. & Gronenborn, A. M. (1991*b*). Two, three and four-dimensional NMR methods for obtaining larger and more precise three-dimensional structures of proteins in solution. *Annu. Rev. Biophys. Biophys. Chem.* **20**, 29–63.

Clore, G. M. & Gronenborn, A. M. (1991*c*). Applications of three- and four-dimensional heteronuclear NMR spectroscopy to protein structure determination. *Progr. Nucl. Magn. Reson.* **23**, 43–92.

Clore, G. M., Bax, A., Driscoll, P. C., Wingfield, P. T. & Gronenborn, A. M. (1990). Assignment of the side chain ^1H and ^{13}C resonances of interleukin-1β using double and triple resonance heteronuclear three-dimensional NMR spectroscopy. *Biochemistry*, **29**, 8172–8184.

Clore, G. M., Bax, A. & Gronenborn, A. M. (1991). Stereospecific assignment of β-methylene protons using three-dimensional ^{15}N-separated Hartmann-Hahn and ^{13}C-separated rotating frame Overhauser spectroscopy. *J. Biomol. NMR*, **1**, 13–22.

Davies, J. F., Hostomska, Z., Hostomsky, Z., Jordan, S. R. & Matthews, D. A. (1991). Crystal structure of the ribonuclease H domain of HIV-1 reverse transcriptase. *Science*, **252**, 88–95.

Driscoll, P. C., Clore, G. M., Marion, D., Wingfield, P. T. & Gronenborn, A. M. (1990*a*). Complete resonance assignment for the polypeptide backbone of interleukin-1β using three-dimensional heteronuclear NMR spectroscopy. *Biochemistry*, **29**, 3542–3556.

Driscoll, P. C., Gronenborn, A. M., Wingfield, P. T. & Clore, G. M. (1990*b*). Determination of the secondary structure and molecular topology of interleukin-1β using two- and three-dimensional heteronuclear ^{15}N-^1H NMR spectroscopy. *Biochemistry*, **29**, 4468–4682.

Fesik, S. W. & Zuiderweg, E. R. P. (1988). Heteronuclear three-dimensional NMR spectroscopy: a strategy for the simplification of homonuclear two-dimensional NMR spectra. *J. Magn. Reson.* **78**, 588–593.

Garrett, D. S., Powers, R., Gronenborn, A. M. & Clore, G. M. (1991). A common sense approach to peak picking using two, three- and four-dimensional contour diagrams for automated computer analysis. *J. Magn. Reson.* In the press.

Ikura, M. & Bax, A. (1991). An efficient three-dimensional NMR technique for correlating the proton and nitrogen-15 backbone amide resonance with the alpha carbon of the preceeding residue in uniformly ^{15}N-^{13}C enriched proteins. *J. Biomol. NMR*, **1**, 99–104.

Ikura, M., Kay, L. E. & Bax, A. (1990*a*). A novel approach for sequential assignment of ^1H, ^{13}C and ^{15}N spectra of larger proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry*, **29**, 4659–4667.

Ikura, M., Kay, L. E., Tschudin, R. & Bax, A. (1990*b*). Three-dimensional NOESY-HMQC spectroscopy of a ^{13}C-labeled protein. *J. Magn. Reson.* **86**, 204–209.

Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Ikehara, M., Matsuzaki, T. & Morikawa, K. (1990). Three-dimensional structure of ribonuclease H from *E. coli*. *Nature (London)*, **347**, 306–309.

Kay, L. E., Marion, D. & Bax, A. (1989). Practical aspects of 3D heteronuclear NMR of proteins. *J. Magn. Reson.* **84**, 72–84.

Kay, L. E., Ikura, M. & Bax, A. (1990*a*). Proton-proton correlation *via* carbon-carbon couplings: a three-dimensional NMR approach for the assignment of aliphatic resonances in proteins labeled with carbon-13. *J. Amer. Chem. Soc.* **112**, 888–889.

Kay, L. E., Ikura, M., Tschudin, R. & Bax, A. (1990*b*). Three-dimensional triple resonance NMR spectroscopy of isotopically enriched proteins. *J. Magn. Reson.* **89**, 496–514.

Krug, M. S. & Berger, S. L. (1989). Ribonuclease H activities associated with viral reverse transcriptases are endonucleases. *Proc. Nat. Acad. Sci., U.S.A.* **86**, 3539–3543.

Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A. & Bax, A. (1989*a*). Three-dimensional heteronuclear NMR of ^{15}N labeled proteins. *J. Amer. Chem. Soc.* **111**, 1515–1517.

Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M. & Clore, G. M. (1989*b*). Overcoming the overlap problem in the assignment of ^{1}H-NMR spectra of larger proteins using three-dimensional heteronuclear ^{1}H-^{15}H Hartmann-Hahn and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin-1β. *Biochemistry*, **28**, 6150–6156.

Mitsuya, H., Yarchoan, R. & Broder, S. (1990). Molecular targets for AIDS therapy. *Science*, **249**, 1533–1544.

Mizrahi, V. (1989) Analysis of ribonuclease H activity of HIV-1 reverse transcriptase using RNA/DNA hybrid substrates derived from the *gag* region of HIV-1. *Biochemistry*, **28**, 9088–9094.

Oostraa, B. M., Harvey, R., Ely, B. K., Markham, A. F. & Smith, A. E. (1983). Transforming activity of polyoma virus middle-T antigen probed by site-directed mutagenesis. *Nature (London)*, **304**, 456–459.

Powers, R., Gronenborn, A. M., Clore, G. M. & Bax, A. (1991). Three-dimensional triple resonance NMR of $^{13}C/^{15}N$ enriched proteins using constant-time evolution. *J. Magn. Reson.* **94**, 209–213.

Richardson, J. S. (1981). The anatomy and taxonomy of protein structure. *Advan. Protein Chem.* **34**, 167–339.

Schatz, O., Cromme, F. V., Grünniger-Leitch, F. & LeGrice, S. F. J. (1989). Point mutations in conserved amino acid residues within the C-terminal domain of HIV-1 reverse transcriptase specifically repress RNase H function. *FEBS. Letters*, **257**, 311–314.

Schatz, O., Mous, J. & LeGrice, S. F. J. (1990). HIV-1 RT-associated ribonuclease H displays both endonuclease and 3′→5′ exonuclease activity. *EMBO J.* **9**, 1171–1176.

Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*. John Wiley, New York.

Yamazaki, T., Yoshida, M., Kanaya, S., Nakamura, H. & Nagayama, K. (1991). Assignment of the backbone ^{1}H, ^{13}C and ^{15}N resonances and secondary structure of ribonuclease H from *Escherichia coli* by heteronuclear three-dimensional NMR spectroscopy. *Biochemistry*, **30**, 6036–6047.

Yang, W., Hendrickson, W. A., Crouch, R. J. & Satow, Y. (1990). Structure of ribonuclease H phased at 2 Å resolution by MAD analysis of the selenomethyonyl protein. *Science*, **249**, 1398–1405.

Zhu, G. & Bax, A. (1990). Improved linear prediction for truncated signals of known phase. *J. Magn. Reson.* **90**, 405–410.

Zuiderweg, E. R. P., McIntosh, L. P., Dahlquist, F. W. & Fesik, S. W. (1990). Three-dimensional ^{13}C-resolved proton NOE spectroscopy of uniformly ^{13}C-labeled proteins for NMR assignment and structure determination of larger molecules. *J. Magn. Reson.* **86**, 210–216.

Edited by P. E. Wright