Protein Science (2000), *9*:138–144. Cambridge University Press. Printed in the USA.

Anti-HIV and anti-tumor protein MAP30, a 30 kDa single-strand type-I RIP, shares similar secondary structure and β-sheet topology with the A chain of ricin, a type-II RIP

YUN-XING WANG,[1] JAISON JACOB,[1] PAUL T. WINGFIELD,[2] IRA PALMER,[2] STEPHEN J. STAHL,[2] JOSHUA D. KAUFMAN,[2] PHILIP LIN HUANG,[3] PAUL LEE HUANG,[3] SYLVIA LEE-HUANG,[3] AND DENNIS A. TORCHIA[1]

[1]Molecular Structural Biology Core, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, Maryland 20892
[2]Protein Expression Laboratory, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, Maryland 20892
[3]Department of Biochemistry, School of Medicine, New York University, New York, New York 10016

(RECEIVED June 8, 1999; ACCEPTED August 20, 1999)

Abstract

MAP30 is a 30 kDa single-stranded, type-I ribosome inactivating protein (RIP) possessing anti-tumor and anti-HIV activities. It binds both ribosomal RNA and the HIV-1 long-terminal repeat DNA. To understand the structural basis for MAP30 activities, we undertook the study of MAP30 by solution NMR spectroscopy. We report nearly complete ^{1}H, ^{13}C, and ^{15}N chemical shift assignments of its 263 amino acids. Based upon an analysis of secondary ^{13}C chemical shifts, $^{3}J_{HNHA}$ coupling constants, hydrogen exchange data, and nuclear Overhauser effect patterns, we find that the secondary structure and β-sheet topology of MAP30 are very similar to those of the ricin A chain, a subunit of the well-known type-II RIP, even though two proteins display distinct activities. We therefore suggest that MAP30 and ricin A chain share a similar three-dimensional fold, and that the reported functional differences between two proteins arise primarily from differences in local three-dimensional structure and other structural properties such as surface electrostatic potentials.

Keywords: anti-cancer; anti-HIV; NMR; RIP proteins, secondary structure

MAP30 is a single-stranded protein containing 263 amino acids (~30 kDa), reported to possess anti-HIV and anti-tumor activity (Lee-Huang et al., 1995a). It belongs to the family of type-I ribosomal inactivating proteins (RIPs) (Barbieri et al., 1993). Like every protein in the RIP family, MAP30 is an N-glycosidase that depurinates the adenine base at position 2543 of 28S rat liver rRNA, thereby inhibiting ribosomal protein synthesis (Barbieri et al., 1993). The reported anti-tumor and anti-HIV activities of MAP30 and other RIPs are believed to be distinct from their ribosome inactivating activity (Lee-Huang et al., 1990, 1991; Zarling et al., 1990; Tumer et al., 1997). In addition to its activity as an N-glycosidase, MAP30 has been reported to (1) inhibit HIV integrase (Lee-Huang et al., 1995b); (2) topologically inactivate viral DNA irreversibly (Lee-Huang et al., 1995a); (3) recognize both DNA and RNA as substrates (Lee-Huang et al., 1995c); and (4) selectively attack tumor transformed and HIV infected cells (Lee-Huang et al., 1995a). These observations stimulated us to determine the structure of the protein in solution, with the goal of obtaining a structural basis for understanding its functions. In addition, MAP30 is among the largest single-chain proteins studied by solution NMR spectroscopy; it therefore serves to test the utility of the repertoire of triple resonance NMR experiments that have been successfully used to assign the spectra of numerous smaller proteins.

Reprint requests to: Dennis A. Torchia, OD NIDCR, Molecular Structural Biology Unit, Bldg. 30, Rm. 106, Bethesda, Maryland 20892; e-mail: dtorchia@dir.nidcr.nih.gov.

Abbreviations: ^{15}N-HSQC, ^{15}N-edited hetero-single quantum coherence spectroscopy; COSY, correlation spectroscopy; HMQC, hetero multiple quantum coherence spectroscopy; NOE, nuclear Overhauser effect; RIP, ribosome inactivating protein.

Results and discussion

Chemical shift assignments

As expected for a 30 kDa single-stranded protein, assigning the chemical shifts was challenging because of short T_2 values and signal overlap. In addition, there were two other obstacles in as-

signing the chemical shifts. First, there is a total of 38 aromatic residues (Tyr17, Phe17, His3, and Trp1) in MAP30. The side chains of aromatic residues, particularly those of phenylalanine, exhibit extensive signal overlap in both ^{1}H and ^{13}C dimensions and have short T_2 values. Second, about 20 residues at the C-terminus of MAP30 are flexible, and their signal intensities were either (1) much stronger than those of residues in the structured region of the protein, due to fast internal motion, or (2) too weak to detect, due to exchange broadening. The strong signals often obscured nearby signals with normal intensity in ^{15}N-edited spectra such as three-dimensional (3D) CCONH (Grzesiek & Bax, 1993). Using the strategy described below, we assigned more than 95% of the

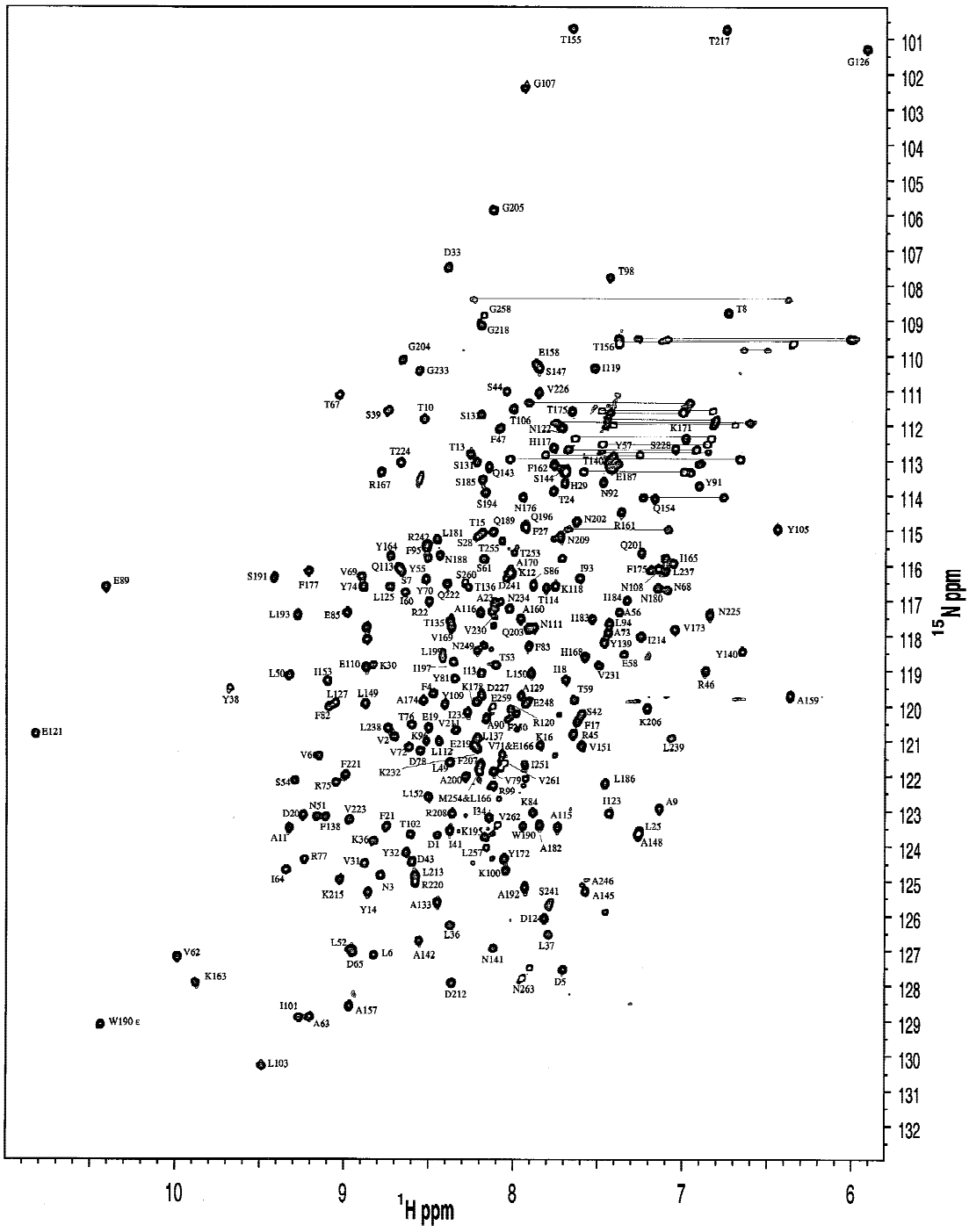

Fig. 1. A two-dimensional ^{15}N HSQC 750 MHz spectrum of MAP30 recorded with a data matrix of 200 (t_1) × 1,024 (t_2) complex points. The cross peaks are labeled with the one-letter amino acid code and residue ID number. No attempt was made to assign the side-chain $^{1}H/^{15}N$ signals of N, Q, K, and R residues. The spectrum was acquired at 40 °C on a Bruker DMX 750 spectrometer. $^{1}H/^{15}N$ signals of G97, N240, A243, S244, T245, and D229 were not observed, presumably because of chemical exchange broadening.

$^{1}H/^{13}C/^{15}N$ chemical shifts of MAP30, including over 90% of the side-chain chemical shifts of the aromatic residues. The table listing the signal assignments is available in supplementary materials.

Because of the relatively good dispersion of amide $^{1}H/^{15}N$ signals (Fig. 1), we used the standard array of amide $^{1}H/^{15}N$-edited through-bond connectivity experiments to make sequential backbone assignments (Bax & Grzesiek, 1993). HNCACB, CBCA(CO)NH, HNCO(CT), and HNCA experiments provide sequential and intraresidue connectivities between the amide $^{1}H/^{15}N$ and $^{13}C'$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$. Strips taken from the 3D HNCACB spectrum are shown in Figure 2. Backbone chemical shifts were extended to $^{1}H/^{13}C$ nuclei in side chains using 3D amide $^{1}H/^{15}N$-edited C(CO)NH, HBHA(CO)NH, and HC(CO)NH experiments in conjunction with a 3D HCCH-TOCSY experiment. Stereospecific assignments for Val and Leu methyl groups were made with the protocol by Neri et al. using a 10% randomly ^{13}C labeled MAP30 sample (Neri et al., 1989). A 750 MHz ^{1}H-^{15}N-HSQC spectrum with backbone amide assignments is shown in Figure 1. No attempt was made to assign labile side-chain protons (or their directly bonded ^{15}N partners) of Lys, Arg, Asn, and Gln residues. However, we did assign 16 side-chain carboxyl ^{13}C signals of aspartate and asparagine residues using a modified HNCO experiment optimized for hydrogen bond detection (Wang et al., 1999a).

As mentioned earlier, MAP30 contains a large number of aromatic residues. Such residues improve spectral resolution because aromatic ring currents disperse chemical shifts of nearby nuclei. However, both ^{1}H and ^{13}C signals of the aromatic side chains are themselves often poorly resolved and have short T_2 values that reduce their sensitivity. To partially remedy the sensitivity prob-

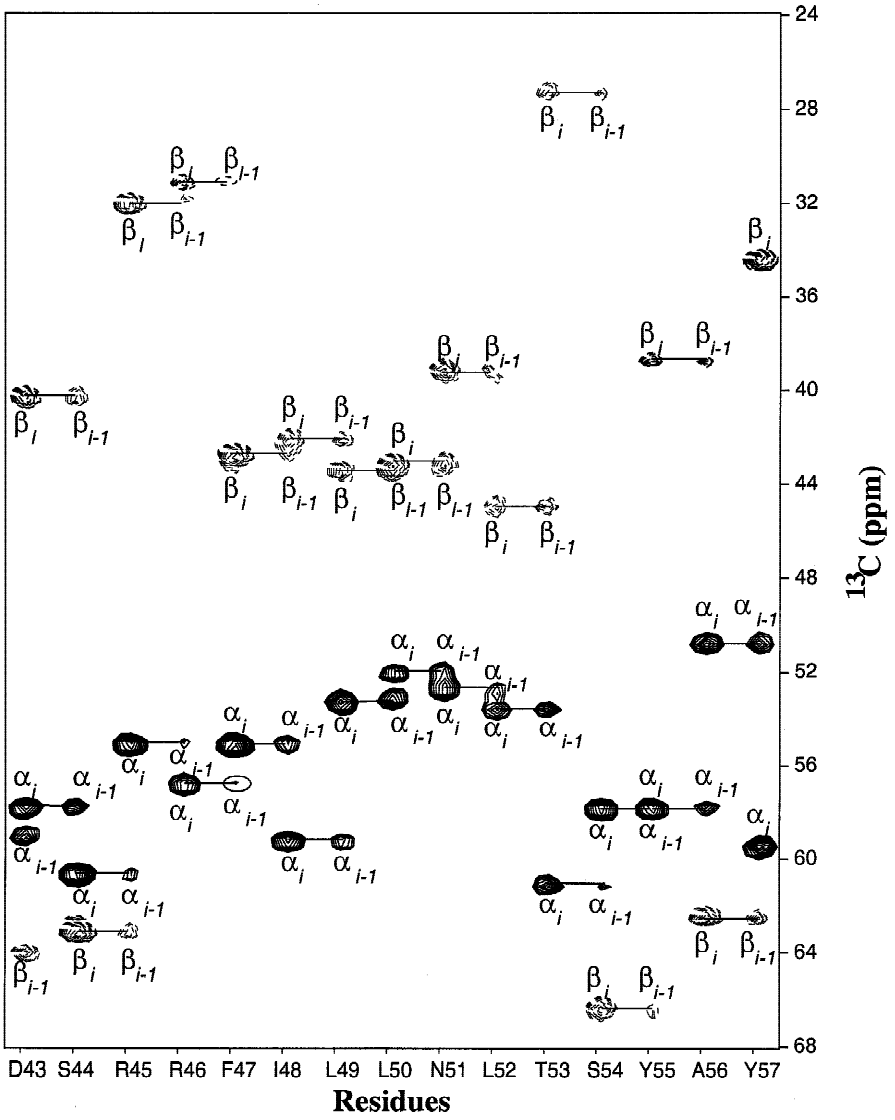

Fig. 2. Representative strips taken from the 750 MHz 3D HNCACB spectrum. The data matrix consisted of 61 (^{15}N) × 54 (^{13}C) × 512 (^{1}H) complex points with 6.4 and 26.8 ms acquisition times in ^{13}C and ^{15}N dimensions, respectively. Solid and broken lines are used to draw cross peaks associated with α- and β-carbons, respectively. The strips typically show intraresidue and weaker sequential through-bond connectivities. Occasional missing sequential cross peaks, such as the $\alpha(i-1)$ and $\beta(i-1)$ cross peaks of F47 were observed in a CBCA(CO)NH experiment, which typically yields strong sequential HN(i) and $C^{\alpha}(i-1)/C^{\beta}(i-1)$ cross peaks.

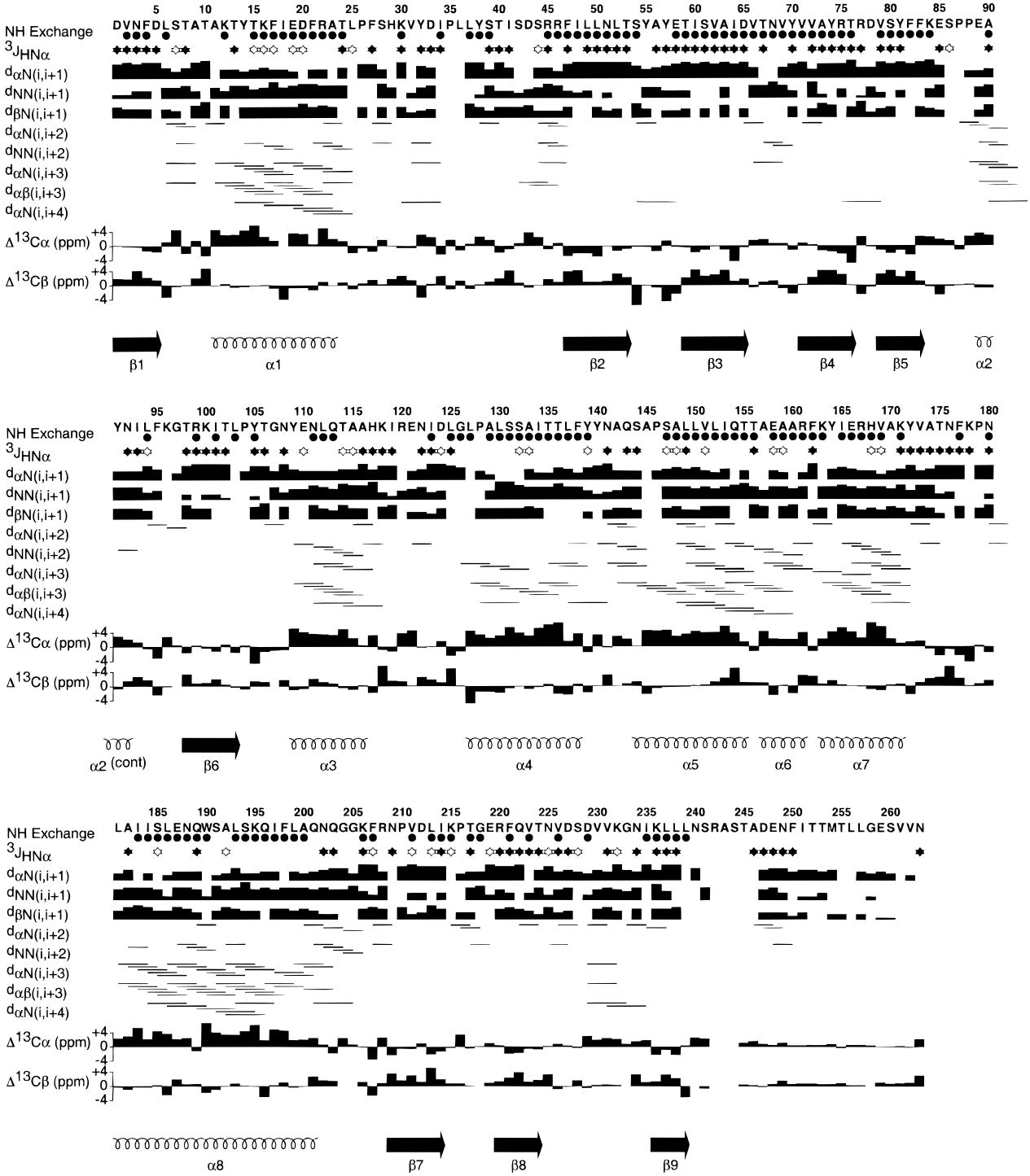

Fig. 3. A diagram of the MAP30 sequence, NH exchange data, three-bond coupling constants ($^3J_{HN\alpha}$), sequential and short-range NOE connectivities, secondary chemical shifts of $^{13}C^{\alpha}$ and $^{13}C^{\beta}$, and the derived protein secondary structure. Slowly exchanging amide protons with exchange lifetime greater than 20 min were found in residues labeled with filled circles. Residues labeled with open and closed stars have $^3J_{HN\alpha} \leq 6$ Hz and $^3J_{HN\alpha} > 6$ Hz, respectively. Otherwise $^3J_{HN\alpha}$ was not detected. Secondary carbon shifts are relative to random-coiled $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts of Wishart et al. (1995).

lem, we developed the 3D CT-HMQC-COSY experiment in which both t_1 and t_2 evolves during the same constant time period, $CT(t_1, t_2)$ (Fig. 1, see Supplementary material in Electronic Appendix) (Zerbe et al., 1996). The spectrum acquired with this scheme correlates the chemical shifts of directly bonded ring carbons and their attached protons. This scheme reduces the magnetization loss due to transverse spin relaxation when compared with the more conventional approach using sequential t_1, t_2 evolution periods. During

the compact $CT(t_1, t_2)$ period, ^{13}C-^{1}H heteronuclear-multiple-quantum coherence evolves and, to first-order, transverse relaxation due to heteronuclear vicinal dipolar coupling vanishes, resulting in a significant increase in sensitivity. We used this scheme and 3D-relayed HMQC-TOCSY (Zerbe et al., 1996) and 3D HCHC-NOESY experiments to assign $^{1}H/^{13}C$ chemical shifts of side chains of all Tyr, His, and Trp residues, and 10 of 14 Phe residues. Assignments were not obtained for either the $^{1}H/^{13}C_{\zeta}$ of Phe27, Phe95, Phe177, and Phe221, or for the $^{1}H/^{13}C$ chemical shifts of the Phe162 side chain, probably due to exchange broadened signals and/or signal overlap.

Secondary structure of MAP30

Using our signal assignments we derived the secondary structure of MAP30 based upon hydrogen exchange data, $^{3}J_{HN\alpha}$ coupling constants, secondary $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts (Spera & Bax, 1991) and short-range NOE patterns (Fig. 3). MAP30 contains eight α-helixes and nine β-strands. Roughly speaking, the N- and C-terminal halves of the protein are rich in β-strands and α-helices, respectively. Strand β1 (residues 1–5) initiates the structure and is followed by the long helix, α1 (residues 11–23). Helix α2 leads directly into four β-strands connected by short turns, β2 (residues 47–53), β3 (residues 59–65), β4 (residues 72–76), and β5 (residues 79–83). The four-stranded motif is followed by helix, α2 (residues 89–92) and strand β6 (residues 98–103). The protein structure is then dominated by six successive alpha helices, α3 (residues 109–116), α4 (residues 127–138), α5 (residues 144–155), α6 (residues 157–61), α7 (residues 163–171), and α8 (residues 181–201). Three short β-strands, i.e., β7 (residues 209–214), β8 (residues 220–224), and β9 (residues 236–238), complete the regular secondary structure of MAP30. Strand β9 is followed by ca. 20 C-terminal residues that are highly flexible and lack regular secondary structure.

β-Sheet topology of MAP 30

As noted above, strands β2–β5 are connected by short turns, suggesting that they form a sheet. In fact, the pattern of interstrand NOEs (Fig. 4) and hydrogen exchange data (Fig. 3) show that strand β1, together with strands β2 through β6, form a six-stranded β-sheet. In addition to the six-stranded β-sheet, β7 and β8 are connected by a three-residue turn and form an antiparallel two-strand sheet. Evidently the two sheets do not interact, because no intersheet NOEs were observed.

Secondary structural comparison of MAP30 with ricin A chain

Because the ricin A chain is the catalytic subdomain of the most extensively studied RIP, ricin (a type-II RIP), we compare its secondary structure to that of MAP30, a type-I RIP (Fig. 5). Although type-I RIPs are single-strand proteins, whereas type-II RIPs are heterodimers, MAP30 and the ricin A chain share about 25% sequence identity in a 240-residue alignment. Furthermore, the residues responsible for the N-glycosidase activity are conserved. These considerations lead us to expect that the proteins have similar structures. To compare their secondary structures, we used PROCHECK (Laskowski et al., 1993) to delineate the secondary structure of the ricin A chain. It is clear from Figure 5 that the proteins share similar secondary structures and the gaps in the

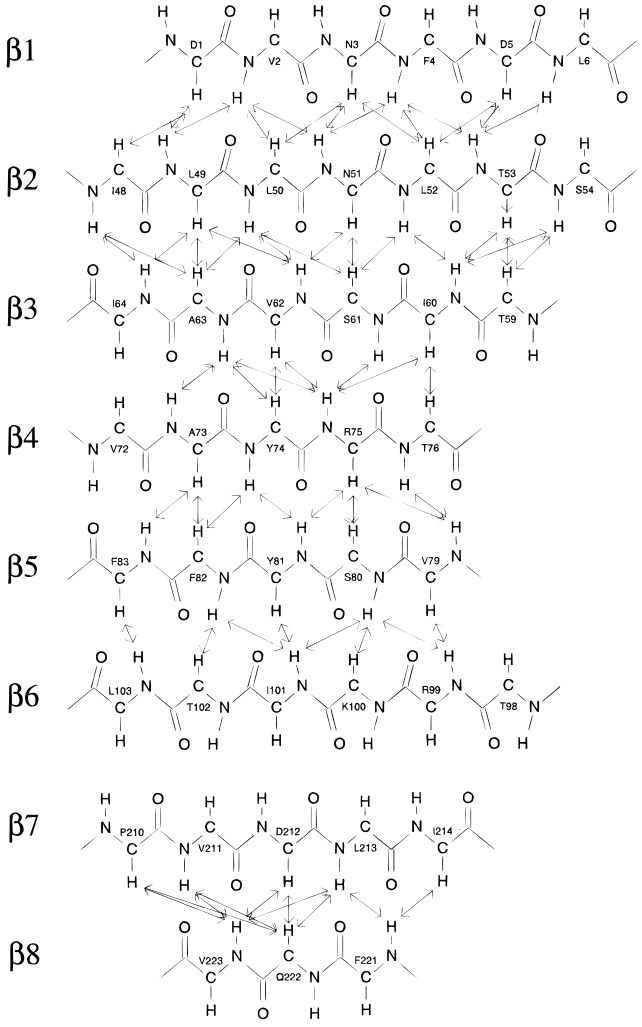

Fig. 4. The topologies of the six- and two-stranded β-sheets of MAP30. Each arrow represents a NOE observed between a pair of protons, without regard to cross peak intensity. Strands β1–β2 and β5–β6 are parallel, whereas strands β2–β3, β3–β4, β4–β5, and β7–8 are antiparallel.

sequential alignment mainly occur at loops of the two proteins. In addition, their β-sheet topologies are nearly the same, as they each contain a six- and a two-stranded β-sheet (Mlsna et al., 1993), each sheet being composed of residues in corresponding positions in the sequence alignment. A closer examination of Figure 5 does reveal some differences between the protein secondary structures. Short strands β2, β3 and short helices α2, α5 in the ricin A chain do not have counterparts in MAP30. In addition, several larger elements of secondary structure have slightly different lengths. However, these differences should be regarded as minor, when one considers that (1) PROCHECK and Figure 3 use different criteria to assign secondary structure, so some differences in lengths of structural elements are expected and, (2) short elements of secondary structure (involving four residues or less) are difficult to identify using the criteria of Figure 3, and often become manifest only after a 3D structure is determined. Finally we note that, although helix α7 is characterized as a long continuous helix in the 3D structure of the ricin A chain (according to PROCHECK), it is distorted precisely at the break between corresponding helices α7 and α8 in MAP30.

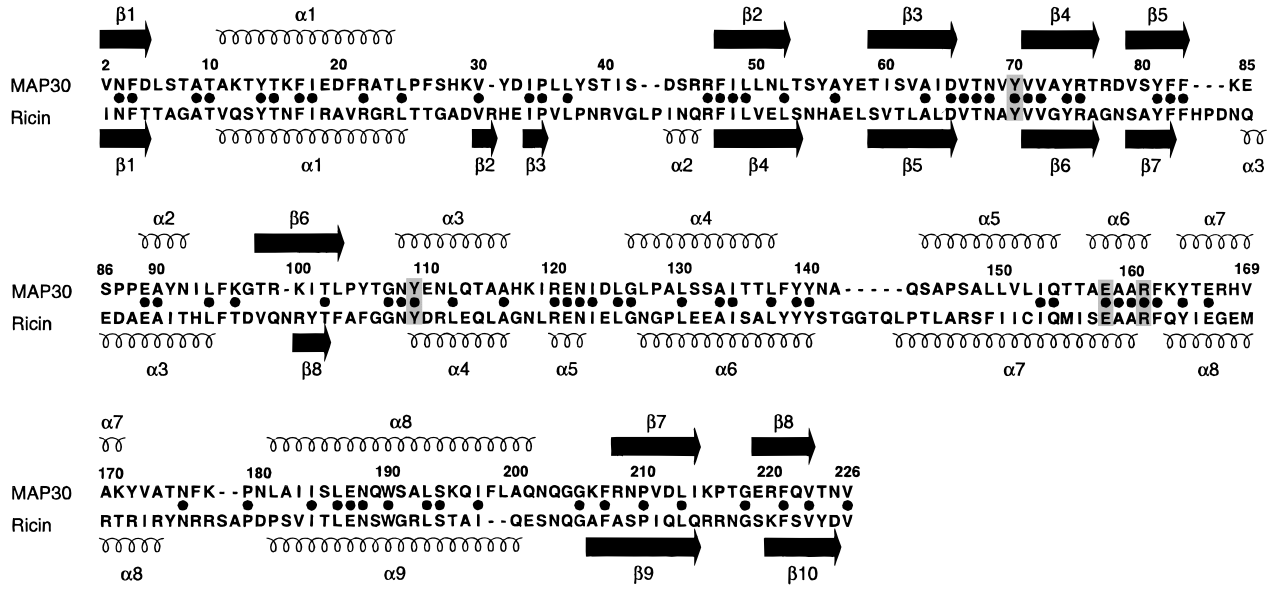

Fig. 5. Sequence alignment and secondary structure comparison of MAP30 with the ricin A chain. Note that only the aligned portions of the full protein sequences are shown in the figure. The numbers above the MAP30 sequence are as in Figure 3. Filled circles indicate identical residues in the alignment, while shaded residues Y70, Y109, E158, and R161 (MAP30 numbering) are the putative active site residues responsible for the N-glycosidase activity (Mlsna et al., 1993). We used PROCHECK to extract the secondary structural information of the ricin A chain from the X-ray coordinates, accession number 1RTC (Mlsna et al., 1993).

Hence, these results indicate the overall folds of MAP30 and the ricin A chain are essentially the same, despite their reported functional differences. Possibly, differences in surface conformations and/or charge distributions contribute to functional differences reported for the two proteins. This will be examined when the determination of the MAP30 3D solution structure, currently underway, is complete.

Conclusion

We have assigned more than 95% of $^{1}H/^{13}C/^{15}N$ chemical shifts of MAP30. The major difficulty in assigning the 263-residue single-chain protein, signal overlap, was overcome without specific labeling or deuteration, by recording triple resonance experiments at 750 MHz and at 40°. We have derived the secondary structure and β-sheet topology of MAP30 based upon signal assignments together with hydrogen exchange data, $^{3}J_{HNHA}$, couplings, the secondary ^{13}C chemical shifts, and NOE patterns. Overall, the secondary structure and β-sheet topology of MAP30 are very similar to those of the ricin A chain. This observation suggests that the reported functional differences between the proteins may arise primarily from differences in local structure and surface electrostatic potentials. We are currently determining the 3D solution structure of MAP30 to ascertain what insight its structure can provide about its reported multiple functions.

Materials and methods

Recombinant MAP30 was overexpressed in *Escherichia coli* to obtain the large quantities of uniformly ^{15}N and $^{13}C/^{15}N$ protein required for NMR structural studies. Recombinant MAP30 has the same activities as the natural protein extracted from bitter melon (Lee-Huang et al., 1995a). The detailed sample preparation and purification are described elsewhere (Wang et al., 1999b).

All NMR spectra were acquired at a protein concentration of ~0.7 mM in Shigemi microcells at 40 °C, in 10 mM sodium phosphate buffer, pH ~5.5. Under these conditions, the average amide proton T_2 is ~17 ms at 500 MHz, indicating that MAP30 behaves as a monomer. Above concentrations of 0.7 mM, the average T_2 of its amide protons decreases as a consequence of aggregation; therefore, NMR spectra were acquired at concentrations below ~0.7 mM.

Supplementary material in Electronic Appendix

$^{1}H/^{13}C/^{15}N$ chemical shift assignment table and 3D CT-HMQC-COSY pulse diagram are available in the Electronic Supplementary Material.

Acknowledgments

We thank Dan Garrett and Frank Delaglio for software, Rolf Tschudin for hardware support, and Stephan Grzesiek for assistance with pulse sequences. This work was supported by the AIDS Targeted Antiviral Program of the Office of the Director of National Institutes of the Health to D.A.T. and by RO1 AI31343 to S.L.H.

References

Barbieri L, Battelli MG, Stirpe F. 1993. Ribosome-inactivating proteins from plants. *Biochim Biophys Acta 1154*:237–282.

Bax A, Grzesiek S. 1993. Methodological advances in protein NMR. *Acc Chem Res 26*:131–138.

Grzesiek S, Bax A. 1993. Amino acid type determination in the sequential assignment procedure of uniformly $^{13}C/^{15}N$-enriched proteins. *J Biomol NMR 3*:185–204.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr 26*:283–291.

Lee-Huang S, Huang PL, Chen H-C, Huang PL, Bourinbaiar A, Huang HI, Kung H-F. 1995a. Anti-HIV and anti-tumor activities of recombinant MAP30 from bitter melon. *Gene 161*:151–156.

Lee-Huang S, Huang PL, Huang PL, Bourinbaiar AS, Chen H-C, Kung H-F. 1995b. Inhibition of the integrase of human immunodeficiency virus (HIV) type 1 by anti-HIV plant proteins MAP30 and GAP31. *Proc Natl Acad Sci USA 92*:8818–8822.

Lee-Huang S, Huang PL, Nara PL, Chen HC, Kung HF, Huang P, Huang HI. 1990. MAP 30: A new inhibitor of HIV-1 infection and replication. *FEBS Lett 272*:12–18.

Lee-Huang S, Kung H-F, Huang PL, Bourinbaiar AS, Morell JL, Brown JH, Huang PL, Tsai W-P, Chen AY, Huang HI, Chen H-C. 1995c. Human immunodeficiency virus type 1 (HIV-1) inhibition, DNA-binding, RNA-binding, and ribosome inactivation activities in the N-terminal segments of the plant anti-HIV protein GAP31. *Proc Natl Acad Sci USA 91*:12208–12212.

Lee-Huang S, Kung H-F, Huang PL, Huang PL, Li B-Q, Huang P, Huang HI, Chen H-C. 1991. A new class of anti-HIV agents: GAP31, DAPs30 and 32. *FEBS Lett 291*:139–144.

Mlsna D, Monzingo AF, Katzin BJ, Ernst S, Robertus JD. 1993. Structure of recombinant ricin A chain at 2.3 A. *Protein Sci 2*:429–435.

Neri D, Szyperski T, Otting G, Senn H, Wuthrich K. 1989. Stereospecific nuclear magnetic resonance assignments of methyl groups of valine and leucine in DNA-binding domain of the 434 repressor by biosynthetically directed fractional ^{13}C labelling. *Biochemistry 28*:71510–71516.

Spera S, Bax A. 1991. An empirical correlation between proton backbone conformation and CA and CB chemical shifts. *J Am Chem Soc 113*:5490–5492.

Tumer NE, Hwang DJ, Bonness M. 1997. C-Terminal deletion mutant of pokeweed antiviral protein inhibits viral infection but does not depurinate host ribosomes. *Proc Natl Acad Sci USA 94*:3866–3871.

Wang Y-X, Jacob J, Cordier F, Wingfield P, Stahl SJ, Lee-Huang S, Torchia DA, Grzesiek S, Bax A. 1999a. Measurement of 3hJNC′ connectivities across hydrogen bonds in a 30 kDa protein. *J Biomol NMR 14*:181–184.

Wang YX, Neamati N, Jacob J, Palmer I, Stahl SJ, Kaufman JD, Huang PL, Winslow HE, Pommier Y, Wingfield PT, Lee-Huang S, Bax A, Torchia DA. 1999b. Solution structure of anti-HIV-1 and anti-tumor protein MAP30: Structural insights into its multiple functions. *Cell 99*:433–442.

Wishart DS, Bigam CG, Holm A, Hodges RS, Sykes BD. 1995. 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J Biomol NMR 5*:67–81.

Zarling JM, Moran PA, Haffar O, Sias J, Richman DD, Spina CA, Myers DE, Kuebelbeck V, Ledbetter JA, Uckun FM. 1990. Inhibition of HIV replication by pokeweed antiviral protein targeted to CD4+ cells by monoclonal antibodies. *Nature 347*:92–95.

Zerbe O, Szyperski T, Ottiger M, Wuthrich K. 1996. Three-dimensional 1H-TOCSY-relayed ct-[^{13}C,^{1}H]-HMQC for aromatic spin system identification in uniformly ^{13}C-labeled proteins. *J Biomol NMR 7*:99–106.