Toward the Solution Structure of Large (>30 kDa) Proteins and Macromolecular Complexes

Cheryl H. Arrowsmith, Weontae Lee, Matthew Revington

Division of Molecular and Structural Biology, Ontario Cancer Institute and Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada M5X 1K9

and

Toshio Yamazaki and Lewis E. Kay

Protein Engineering Network of Centers of Excellence and Departments of Medical Genetics, Biochemistry and Chemistry, University of Toronto, Toronto, Ontario, Canada M5S 1A8

I. Introduction

Advances in NMR technology over the last 5 years, particularly the development of 3 and 4 dimensional (3D and 4D) heteronuclear NMR (1-4), have provided the tools to determine the solution conformations of medium sized proteins in the range 15-25 kDa. Recent applications of this technology include the structures of a 23 kDa calmodulin-peptide complex (5), IL1-β (17 kDa) (6), domain IIA of glucose permease (18 kDa) (7) and protein S (19 kDa) (8). Backbone assignments have been reported for two 28 kDa proteins (9,10) and most recently for a 38 kDa protein (Copie and Torchia, personal communication). However, a number of laboratories have pointed out the limitations of these techniques for many protein systems with molecular weights greater than about 25 kDa (10-13). The problem arises from the short transverse relaxation times (T_2) of nuclei in larger molecules due to slower rotational correlation times and the resulting efficient dipolar relaxation between covalently bound nuclei. Rapid T_2 relaxation, especially of the carbon nuclei, during evolution periods and scalar transfer steps of nD pulse sequences reduces the sensitivity of many triple resonance techniques. In addition, many of the isotope filtered pulse sequences (14-17), which are crucial for identifying intermolecular interactions, also contain extensive delays which reduce their sensitivity for larger systems. In practice, many of the currently used NMR experiments will fail for molecular weights in excess of ~30 kDa. There are at least two strategies that can be used to overcome this problem and to extend the usefulness of NMR to larger macromolecules and complexes. First, one can modify current sequences to minimize the length of the pulse sequence and secondly one can improve the sensitivity of pulse schemes by increasing the T_2 relaxation times of nuclei involved through random incorporation of deuterium into the protein (12,18,19).

We discuss here our application of these strategies to the 37 kDa complex between *E.coli trp* repressor and *trp* operator. The resonances of this complex have been assigned and a structure determined using a combination of selective deuteration and a series of heteronuclear NMR experiments which did not rely on the small Cα- N couplings (11). We discuss the limitations of this method and promising recent results suggesting that the strategies mentioned above will extend the usefulness of NMR to molecular weights in the range of 30-40 kDa.

TECHNIQUES IN PROTEIN CHEMISTRY VI

II. Materials and Methods

A. Protein Preparation

Uniformly ^{13}C and ^{15}N labelled *trp* repressor was isolated from *E. coli* strain CY15070 (20) containing the overproducing plasmid pJPR2 grown in 1-2 L of minimal media with 1g/L $^{15}NH_4Cl$ and 2.5g/L D- glucose-$^{13}C_6$ as the sole nitrogen and carbon sources, respectively. It typically takes 12-16 hours at 37°C for the cells to reach an OD of 0.6-0.9 at 600 nm at which point protein production was induced with 1 mM IPTG. Cells were harvested after another 8-9 hours, when the final OD was typically double that at the time of induction. Purification is as described previously (20-21). Typical final protein yields were 10-20 mg /L growth media.

For triply labelled $^{2}H,^{13}C,^{15}N$-*trp* repressor the bacteria were first adapted to growth in D_20 as follows. A single colony from an LB agar plate was used to inoculate 12 ml of M9 media (22) with 200 μg/ml ampicillin and 33% D_20 as the solvent. After overnight growth at 37°C, the cells were plated out onto an M9 agar plate made with 33% D_20. The 33% plate was incubated at 37°C for two days resulting in very small colonies. A colony from the 33% plate was used to inoculate 12 ml of M9 media with 56% D_20 as solvent. After overnight growth the cells were plated onto a 56% D_20 M9 plate and incubated for 24 hours. For large scale purification of triply labelled trp repressor a single colony from the 56% D_2O plate was used to inoculate 10 ml of M9 media containing $^{15}NH_4Cl$, D-glucose-$^{13}C_6$ and 70% D_20 (Cambridge Isotope Labs.) as solvent. After overnight growth this culture was dense and was used to inoculate 1.5 L of the same M9 media. After shaking at 37°C for 18.5 hours the O.D. (600 nm) had reached 0.6, at which point the culture was induced with IPTG and grown for another 15 hours. The final O.D. was 1.2, yielding 3.9g of wet cells. Purification as usual yielded 60 mg of pure triply labelled protein.

For NMR samples the protein was concentrated by ultrafiltration to a final concentration of 1-2.4 mM *trp* repressor monomer in 500 mM NaCl, 50 mM sodium phosphate, pH6. The high salt concentration was necessary to prevent aggregation of the protein. The corepressors, either L-tryptophan or 5-methyl-L-tryptophan, were added at a concentration of 1.5-2 times the protein subunit concentration to form *trp* holorepressor. Protein-DNA complexes were prepared by adding the appropriate amount of synthetic operator DNA (23) to the above sample followed by dialysis and concentration into a pH6 solution of 50 mM (or less) sodium phosphate. Higher salt concentrations destabilized the protein-DNA complex and resulted in much poorer quality NMR spectra.

B. NMR Spectroscopy

For non-deuterated complexes, NMR experiments were performed on either Varian Unity600 or Varian Unity+500 spectrometers. The 600 MHz instrument was equipped with a triple resonance probe and a PTS synthesizer as a pseudo fourth channel. The 500 MHz spectrometer was a four channel instrument with a triple resonance probe with an actively shielded pulsed field gradient coil. All experiments were performed at 37°C. The heteronuclear experiments shown in Figure 1 were performed as described in Zhang et al (11) and Revington et al (24). For all 3D experiments 32 transients were required for sufficient signal to noise. This necessitated the use of fewer increments and the use of linear prediction (25)

to obtain sufficient resolution in one or both of the indirectly detected dimensions. An improved version of the 3D ^{13}C-F_3-filtered-HMQC-NOESY (5) was recorded as reported by Lee et al (27).

An HNCA experiment designed for $^{2}H,^{15}N,^{13}C$ labelled proteins was recorded as described by Yamazaki et al (19). This experiment was acquired on a three channel Varian Unity500 spectrometer modified to perform the ^{15}N pulses, ^{15}N decoupling and ^{2}H decoupling on a single channel. Alternatively, this could be accomplished on a four channel instrument without modification.

III. Results and Discussion

The solution structure of the *trp* repressor-operator complex was recently determined using a combination of selective deuteration and heteronuclear NMR assignment strategies. Figure 1 shows the heteronuclear experiments used in our laboratory to assign ^{15}N, ^{13}C and ^{1}H resonances of the repressor-operator complex. We started off with gradient enhanced ^{15}N HSQC (26) and NOESY-HMQC (28) spectra of an ^{15}N labelled protein bound to natural abundance DNA and corepressor. Since only those proton and ^{15}N resonances in the DNA-binding or ligand binding regions showed significant changes upon binding DNA, approximately 70% of the amide N-H pairs could be assigned based on similarity with spectra of the holorepressor. Assignment of the backbone carbon resonances using conventional triple resonance techniques (2,4) was not possible due to the poor quality of spectra involving the CA-NH correlation. The NOE-based ^{15}N assignments were confirmed and further backbone carbon and nitrogen assignments identified from the HNCO (29) and (HB)CBCACO(CA)HA (30) experiments. The HNCO is the most sensitive triple resonance experiment (2) and provides sequential connectivities through the peptide bond. The (HB)CBCACO(CA)HA relys on relatively large C-H and C-C couplings and also gives a reasonable signal on this size of complex. By matching the carbonyl resonances from these two experiments it is possible, in principle, to connect the Cα and Cβ of each side chain with the NH resonances of the next residue in the sequence. In practice, however, there will be ambiguities which would ideally be resolved by additional triple resonance experiments. Although we were not successful in acquiring additional triple resonance experiments, the fact that approximately 50% of the Cα resonances remained the same as in the free protein allowed us to assign ~80% of the backbone and Cβ resonances from the two spectra described above. The Cα and Cβ

Labelled Protein		
^{15}N	$^{15}N/^{13}C$	^{13}C
HSQC	3D HNCO	3D HCCH-TOCSY
3D NOESY-HMQC	3D (HB)CBCACO(CA)HA	3D NOESY-HSQC
^{15}N-F_1-filtered NOESY	$^{15}N/^{13}C$-F_1, F_2-filtered NOESY	3D ^{13}C-F_3-filtered HMQC-NOESY

Figure 1 NMR Experiments used to assign the *trp* repressor-operator complex (11,24).

assignments served as a starting point for analysis of the 3D ^{13}C-NOESY-HSQC (31) and HCCH-TOCSY (32). In our hands the HCCH-TOCSY gave only partial connectivities, presumably due to rapid T_2 relaxation. Therefore a complete side chain assignment via scalar connectivities was not possible, and we relied heavily on NOE spectra for assignment of the side chain resonances.

Using this combination of experiments it was possible to assign approximately 80% of the backbone and side chain resonances of DNA-bound *trp* repressor. The proton assignments were confirmed by comparison with the results of selective deuteration experiments (11). The success of this strategy relied, in part, on the fact that a large portion of the protein did not undergo significant changes in chemical shift upon binding to DNA. This meant that assignments for the smaller species based on the more reliable triple resonance experiments (rather than NOESYs) could be carried over to the complex. Figure 2 shows a comparison of "strips" from the HNCA of the holorepressor with those of the (HB)CBCACO(CA)HA of the complex for residues 94-106 whose alpha carbons do not change significantly upon interaction with DNA.

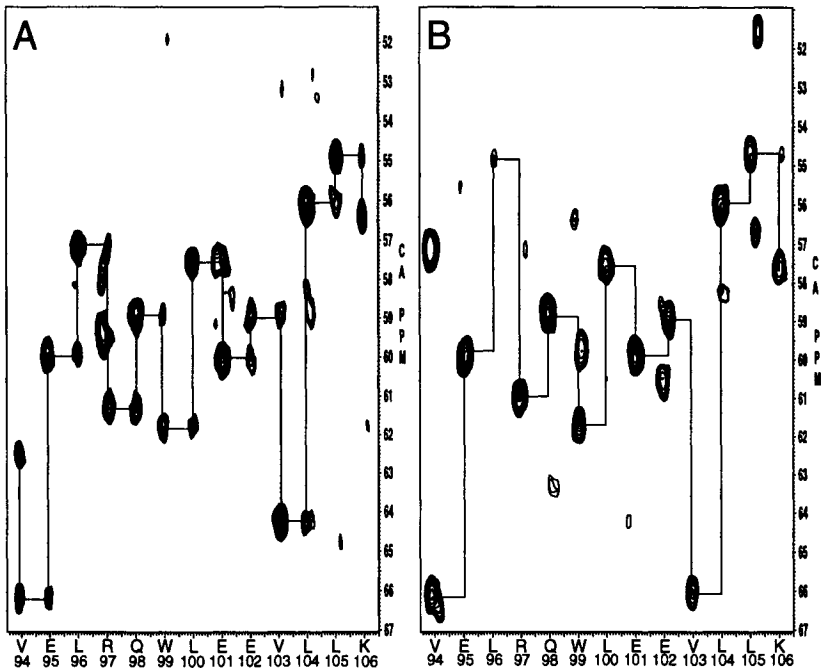

Figure 2. Strips from the HNCA of *trp* repressor (A) and the (HB)CBCACO(CA)HA of *trp* repressor bound to DNA (B) showing how residues 94-106 have very similar chemical shifts. Vertical strips from each residue in (A) contain correlations to the intraresidue Cα as well as the Cα of the previous residue. Vertical strips in (B) show only intraresidue Cα correlations.

In addition to assigning the resonances of the protein we needed to assign those of the DNA and ligand in the complex and identify intermolecular NOEs from the protein to corepressor and DNA. The assignment of the resonances of unlabelled *trp* operator DNA and the corepressor, L-tryptophan, were accomplished using 2D F1/F2 isotope-filtered NOESY experiments (15). These spectra show only NOEs within the bound DNA and ligand, and can be analyzed in the conventional manner for sequence specific assignment of DNA (33). NOEs between the ^{13}C-labelled protein and unlabelled DNA and corepressor were identified from the F_3-filtered-HMQC-NOESY. This spectrum proved to be extremely valuable for the structural analysis of the protein-DNA contact surface. However, all the isotope filtered experiments require "purge sequences" to eliminate magnetization from the labelled species which requires the spins involved to spend additional time in the transverse plane. This reduces the sensitivity of these experiments relative to a normal 2D or 3D NOESY. In particular, the 3D F_3-filtered-HMQC-NOESY sequence as originally reported (5) gave useful signal on our complex only when it was run in the 2D mode (*i.e.* the carbon evolution time was eliminated). Although the carbon evolution time reduces the sensitivity relative to the 2D experiment, it is essential for the unambiguous assignment of the protein contribution to the cross peaks. Therefore, we improved upon the existing 3D pulse sequence by reducing the total time that ^1H magnetization spends in the transverse plane by combining the proton evolution and $^1H/^{13}C$ scalar transfer times (4, 27, 34, 35). In addition, pulsed field gradients (PFG) were used to eliminate artifacts rather than extensive phase cycling. A double purging scheme with delays optimized for two different values of carbon-proton coupling was also used to insure complete filtering of all ^{13}C-bound protons. Figure 3 shows a slice of the 3D F_3-filtered-HMQC-NOESY spectrum with that of the analogous HMQC-NOESY. It is clear from this comparison that the identification of solely intermolecular NOEs in the F_3-filtered spectrum is very useful for interpreting the NOEs in the non-filtered spectrum.

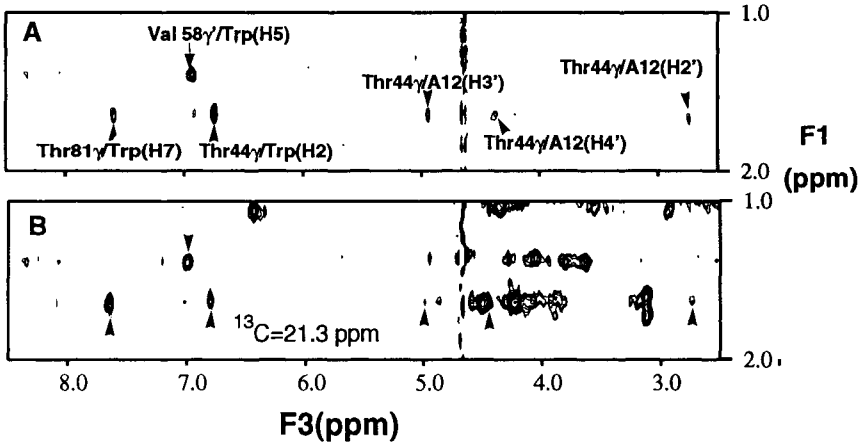

Figure 3. A comparison of carbon slices from the F_3-filtered-HMQC-NOESY spectrum(A) with those of the ^{13}C HMQC-NOESY (B). Intermolecular NOEs in (B) are indicated.

Although an assignment strategy such as that outlined in Figure 1 may work for certain favorable cases, it is clear that a more general assignment strategy is needed for many protein systems with molecular weights in excess of 25 kDa. Since the major problem with larger systems is rapid transverse relaxation during the pulse sequence, longer T_2 relaxation times of the nuclei involved will improve the sensitivity of most NMR experiments. The major mechanism of carbon transverse relaxation involves dipolar relaxation with its covalently bound protons. Therefore, we have sought to increase the carbon T_2 by substitution of deuterium for hydrogen. Toward this end we have prepared ~70%-^{2}H/^{15}N/^{13}C labelled *trp* repressor as described above. From preliminary NMR measurements it appears that the level of deuterium incorporation at the Cα position is approximately that of the deuterium level in the growth media. The incorporation of deuterium within the amino acid side chains is currently under investigation. Deuterium labelling at this level is only slightly more expensive than a protonated ^{15}N/^{13}C growth because high isotopic purity D2O is not necessary. Moreover, although the bacteria grew more slowly in deuterated media, the final yield of protein was as good or better than other protonated preparations in our lab. Preliminary relaxation measurements indicate that this level of deuteration increases the Cα T_2 relaxation times by approximately 8-fold (Yamazaki et al, submitted). This has allowed us to perform triple resonance experiments with excellent sensitivity in an effort to more systematically and thoroughly assign all backbone residues.

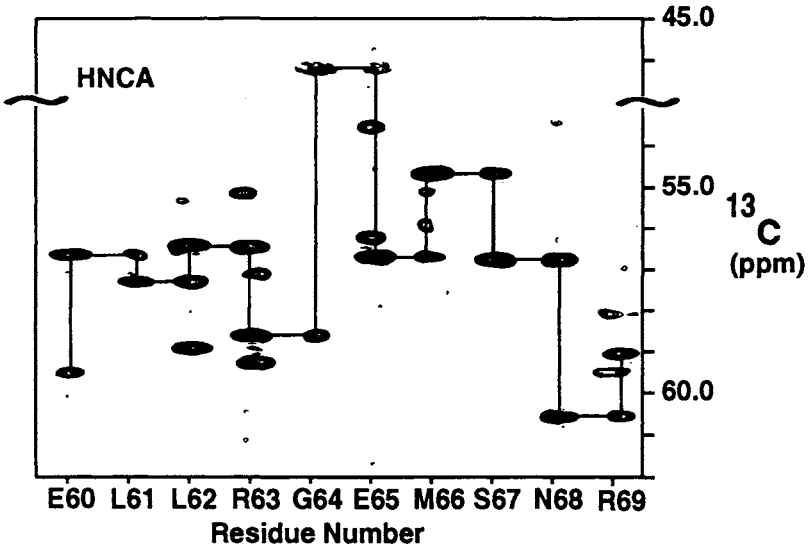

Figure 4. Strips from the HNCA of 70% deuterated ^{15}N,^{13}C labelled *trp* repressor bound to DNA in the presence of the corepressor 5-methyl-L-tryptophan. Each vertical strip contains correlations to the intraresidue Cα as well as the Cα of the previous residue.

Figure 4 shows strips from the HNCA spectrum of $^{2}H/^{15}N/^{13}C$ labelled repressor bound to DNA. Due to the excellent sensitivity and resolution of this spectrum, 99% of the expected inter and intra residue crosspeaks were observed. Thus, a high proportion of the HN, N and Cα resonances of the DNA-bound protein could be assigned from this spectrum alone. It is likely that a complete backbone assignment will be possible with complementary triple resonance experiments on this deuterated complex. This is a promising result indicating that the use of fractional deuteration may enable one to assign proteins and complexes as large as perhaps 30-40 kDa.

Acknowledgments

The protein chemistry, much of the spectroscopy and NMR resonance assignments for *trp* repressor were carried out in the Arrowsmith laboratory with support from the NCI of Canada and the Human Frontier Science Program. We acknowledge our very fruitful collaboration with the laboratory of O. Jardetzky on the structure of the *trp* repressor complex. The development and implementation of the HNCA and isotope filtered pulse sequences was carried out in the laboratory of L.E. Kay with support from NSERC and NCI of Canada. T. Yamazaki acknowledges fellowship support from the Human Frontier Science Program.

References

1. Montelione, G.T., Wagner, G. (1990) *J. Magn. Reson.***87** 183-188.
2. Ikura, M., Kay, L.E., Bax, A. (1990) *Biochem.* **29** 4659-4667.
3. Kay, L. E., Clore, G. M., Bax, A. and Gronenborn, A. M. (1990) *Science* **249**, 411-414.
4. Bax, A., and Grzesiek, S. (1993) *Acc. Chem. Res.* **26**, 131-138.
5. Ikura, M; Clore, G.M.; Gronenborn, A.M.; Zhu, G.; Klee, C.B.; Bax, A. (1992)*Science* **256**, 632-638.
6. Clore, G.M., Wingfield, P.T. and Gronenborn, A.M. (1991) *Biochem.* **30**, 2315-2323.
7. Fairbrother, W.J., Gippert, G.P. Reizer, J., Seier, M.H. and Wright, P.E. (1992) *FEBS Lett.* **296**, 148-152.
8. Bagby, S., Harvey, T.S., Kay, L.E., Eagle, S.G., Inouye, S., and Ikura, M. (1994) *Structure* **2**, 107-122.
9. Remerowski, M.L., Komke, T., Groenewegen, A., H.A.M. Pepermans, Hilbers, C.W. and van de Ven, F.J.M (1994) *J. Biomol. NMR.* **4**, 257-278.
10. Fogh, R.H., Schipper, D. Boelens, R., and Kaptein, R. (1994) *J. Biomol. NMR* **4**, 123-128.
11. Zhang, H., Zhao, D., Revington, M., Lee, W., Jia, X., Arrowsmith, C. and Jardetzky, O. (1994) *J.Mol.Biol.* **238**, 592-614.
12. Grzesiek, S., Anglister, J., Ren, H. and Bax, A. (1993) *J. Am. Chem. Soc.* **115**, 4369.
13. Kushlan, D.M. and LeMaster, D.M. (1993) *J. Biomol. NMR* **3**, 701.
14. Wider, G., Weber, C., Traber, R., Widmer, H., and Wuthrich, K. (1990) *J. Am. Chem. Soc.* **112**, 9015-9016; Otting, G, Wuthrich, K., (1990) *Q. Rev. Biophys.* **23**, 39-96.
15. Ikura, M., Bax, A., (1992)*J. Am. Chem. Soc.*, **114**, 2433-2440.
16. Gemmecker, G., Olejniczak, E.T., Fesik, S.,(1992) *J. Magn. Reson.*, **96**, 199-204.
17. Folkers, P.J.M., Folmer, R.H.A., Konings, R.N.H., Hilbers, C.W. (1993) *J. Am. Chem. Soc.* **115**, 3798 3799.
18. LeMaster, D.M. and Richards, F.M. (1988) Biochem. **27**, 142-150.
19. Yamazaki, T. Lee, W., Revington, M., Mattiello, D. L., Dahlquist, F. W., Arrowsmith, C. H. & Kay, L. E. (1994) *J. Amer. Chem. Soc.***116** 6464-6465.

20. Paluh, J.L., and Yanofsky, C. (1986) *Nuc. Acids. Res.* **14**, 7851-7861.
21. Arrowsmith, C.H., Pachter, R., Altman, R.B., Iyer, S.B. and Jardetzky, O. (1990) *Biochem.* **29**, 6332-6341.
22. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning*, 2nd Ed. Cold Spring Harbor Press, Cold Spring Harbor, N.Y.
23. Lefevre, J.F., Lane A.N. and Jardetzky, O. (1986) *Biochem.* **26**, 5076-5090.
24. Revington, M.J., Lee, W. and Arrowsmith, C.H. submitted.
25. Marion, D. and Bax, A. (1989) *J. Magn. Res.* **84**, 72-84.
26. Kay, L. E., Keifer, P., & Saarinen, T. (1992) *J. Am. Chem. Soc.* **114**, 10663-10665.
27. Lee, W., Revington, M.J., Arrowsmith, C.H. and Kay, L.E.(1994) *FEBS. Lett. in press.*
28. Zuiderweg, E. R. P. & Fesik, S. W. (1989) *Biochemistry* **28**, 2387-2391.
29. Kay, L. E., Ikura, M. & Bax, A. (1991) *J. Magn. Reson.* **91**, 84-91.
30. Kay, L. E. (1993) *J. Am. Chem. Soc.* **115**, 2055-2057.
31. Muhandiram, D. R., Farrow, N., Xu, G. Y., Smallcombe, S.J. and Kay, L. E. (1993) *J. Magn. Reson. Series B.* **102**, 317-321.
32. Kay, L. E., Xu, G. Y., Singer, A. U., Muhandiram, D. R. and Forman-Kay, J. D. (1993) *J. Magn. Reson. B* **101**, 333-337.
33. Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids.* John Wiley & Sons, N.Y.
34. Muhandiram, D. R., Xu, G.Y. and Kay, L. E. (1993) *J. Biomol. NMR*, 3, 463-470.
35. Logan, T. M., Olejniczak, E. T., Xu, R. X. and Fesik, S. W. (1993)*J. Biomol. NMR*, **3**, 225-231.