Journal of Biomolecular NMR, 3 (1993) 375–385
ESCOM

J-Bio NMR 127

Prospects for NMR of large proteins

Gerhard Wagner

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, MA 02115, U.S.A.

Received 2 March 1993
Accepted 22 March 1993

Keywords: Protein NMR; Protein structure; Heteronuclear multidimensional NMR; Relaxation

SUMMARY

During the last decade, solution structures of many small proteins have been solved by NMR. The size of proteins that are being analyzed by NMR seems to increase steadily. Protein structures up to 18 kD have been solved sofar, and spectra of proteins up to 30 kD have been assigned. Thus, NMR emerges as an attractive technique, in particular for structural studies of proteins that cannot by crystallized. However, the application of the technology is limited by relaxation properties of the proteins. If relaxation would only be determined by Stokes–Einstein-type rotational diffusion, the effects of the molecular size on relaxation properties of proteins and thus on the performance of multi-dimensional multiple-resonance experiments could readily be estimated. From this perspective, solving two- or three-fold larger structures seems possible. However, most larger proteins exhibit serious line broadening due to aggregation or other still unknown effects. Sample conditioning to minimize these effects is presently the challenge in the work with large proteins.

Can one break the molecular weight barrier, what is the largest protein structure that can be solved by NMR, should we focus on solving larger and larger protein structures by NMR, should one compete with X-ray crystallography? These are frequently asked questions from within or from outside the protein NMR community. Breaking a record is a highly motivating goal, and it is an easily understandable scientific achievement. More importantly it will break new grounds to learn about protein structure and function. If we approach protein structure analysis from the perspective of solving a biological problem we want to select a protein to be studied according to its biological interest rather than its molecular weight. However, it often happens that interesting proteins are just a little too large or, more precisely, they have just a little too broad lines so that they cannot be handled by NMR. The present challenge of protein NMR is to perform spectroscopy of proteins with large line widths, and to identify and eliminate causes of line broadening other than large molecular weight.

An attempt to solve a large protein structure is time consuming, expensive in terms of sample preparation and instrument time. To solve a large protein structure by NMR is particularly

attractive if the structure cannot be solved by crystallographic techniques. However, we cannot predict whether or not the protein we have selected will be crystallized. At any point in an NMR structure analysis, crystallization may be successful. From this point onwards, solving the structure crystallographically is usually fast. The efficiency of crystallography is higher for larger proteins, while structures of small proteins can be solved with competitive speed by the two methods. Therefore, one should seriously consider whether an endeavor to solve a structure of a large protein will be rewarding. It is worth identifying proteins that would be difficult to handle by X-ray crystallography and to focus on such problems. Proteins seem to be difficult to crystallize if they have large mobile regions, such as unstructured polypeptide tails or carbohydrate moieties. NMR can often cope with and describe large-scale mobility in proteins. Although it is clearly more rewarding to solve a new protein structure first, there is also merit solving structures in solution after an X-ray structure has been determined. The availability of a second technique besides single crystal diffraction methods provides a control for the correctness of a structure. Indeed, a few discrepancies detected between NMR and X-ray structures have created an awareness of possible mistakes in X-ray and NMR structures. Crystal contacts may perturb protein structures and sometimes provide an incorrect description of a protein surface. Obviously, the chance to perturb the conformation globally by crystal contacts is more significant in small rather than in large proteins. Therefore, the impact of NMR may be higher for small proteins. However, also in larger proteins the information from NMR may be significant since active sites of proteins often reside on protein surfaces that might be perturbed by crystallization. Therefore, structure analysis by NMR has merits even if an X-ray structure of the same protein has already been solved.

Since we are ultimately interested in protein function we want to study interactions of proteins with target molecules or even study protein complexes. Knowledge of the 3D structure of a single protein does not immediately lead to an understanding of its function. Naturally, most proteins interact specifically with other molecules, such as ligands, substrates, inhibitors, receptors or polynucleotides. Identification of the interaction sites and solving a complex structure is a first important step towards an understanding of protein function, based on the structure. Obviously, protein complexes tend to be large, around or above the limits of NMR. If we want to learn about such interactions even in the smallest systems, it is desirable to push the technology to handle larger proteins, in particular protein complexes.

Protein NMR is becoming an attractive tool for rational drug design. It is expected that knowledge of a 3D structure can be used to identify or design molecules that inhibit its function. Enzymes, proteins involved in signal transduction, extracellular or intracellular receptor domains, cell adhesion proteins, proteins involved in blood coagulation, are examples of such drug targets. For receptors with a single transmembrane strand, functional extracellular or cytoplasmatic domains can often be produced. Many of these proteins are at or beyond the present molecular weight limit of NMR. Crystallization of receptor domains seems difficult because these are often glycosylated. To use structure information for drug design, it will be insufficient to identify just the topology and the architecture of the protein core; a good characterization of surface conformations for such large proteins will be necessary.

After isotope labeling of proteins became a widely available expertise and multidimensional and multiple resonance experiments were developed, it seemed feasible that some structures of up to 30 or 35 kDa could be determined in the near future. Indeed, resonance assignments for

proteins as large as 22 kDa for monomeric proteins (Stockman et al., 1992) and even 31 kDa for a dimeric protein (Grzesiek et al., 1992) have been reported. However, the largest structures actually solved are yet only 18 kDa in size (Clore et al., 1991; Fairbrother et al., 1992).

Solving a protein structure by NMR depends to a large extent also on other aspects than molecular size. The main factor is whether a protein exhibits good NMR spectra. In my experience, only a fraction of the proteins that fall within the suitable molecular weight range behave that well that they can be assigned and a structure can be determined. Even if assignments can be made, this is no guarantee that a precise structure can be obtained. Some proteins just show fewer NOEs than others, independent of their size. This fact reflects intrinsic properties of proteins, such as internal mobility, aggregation and partial unfolding. The proteins for which precise and well-defined 3d structures can be determined by NMR or crystallography are a subset and not a representative sample.

It is worth reviewing the problems associated with solving structures of large proteins.

(i) The spectra are crowded due to the large number of resonances. The peak overlap problem has been solved to a large extent by the development of multidimensional and multiple resonance experiments, at least for proteins up to 25 kDa that can be isotope labeled. However, some notorious spectral regions, such as aromatic resonances suffer from overlap problems even in multidimensional spectra since δ-carbon resonances of tyrosines and all phenylalanine carbons show little chemical shift dispersion. Thus, NOEs to aromatic resonances are often difficult to assign.

(ii) The fast transverse relaxation rates of large proteins impose limits to the lengths of the pulse sequences that can be applied. Unfortunately, molecular size is not the only source of line broadening in proteins. Obviously, we have no means to increase transverse relaxation times; however, we can design pulse sequences to use coherences that relax most slowly.

(iii) The increasing size of the protein limits the maximum concentration that can be used, and the sensitivity is limited. At high concentration all proteins tend to aggregate and exhibit additional line broadening beyond that expected just from the molecular size.

(iv) Many proteins tend to aggregate already far below this critical concentration and exhibit a severely broadened spectrum. Electrostatic or hydrophobic interactions are possible reasons for the aggregation. Sometimes proteins with mobile tails have higher tendencies to aggregate. There is insufficient understanding of this phenomenon and the only strategies to overcome this problem are screening of pH, temperature, buffers, ionic strength, detergents or even mutations and trimming of the polypeptide chains. The art of this 'protein conditioning' is in its infancy. Progress in this field is desperately needed.

(v) In our experience, large proteins often have significant moieties that do not show NOEs; sometimes a number of residues do not show detectable resonances at all. Most likely this is due to multiple conformations in intermediate exchange. Such a lack of a well-defined conformation may be an important property of a protein. On the other hand, this hampers assignments as well as the determination of the structure of the well-defined part. The scientist who is faced with such a situation may be blamed for poor performance. Indeed it is difficult to judge whether a structure appears ill defined because of real mobility or incompetence of the scientist. Relaxation time measurements can provide valuable means to assure oneself that mobility is causing problems for structure analysis.

(vi) Limited protein stability is another common problem. It may lead to loss of samples

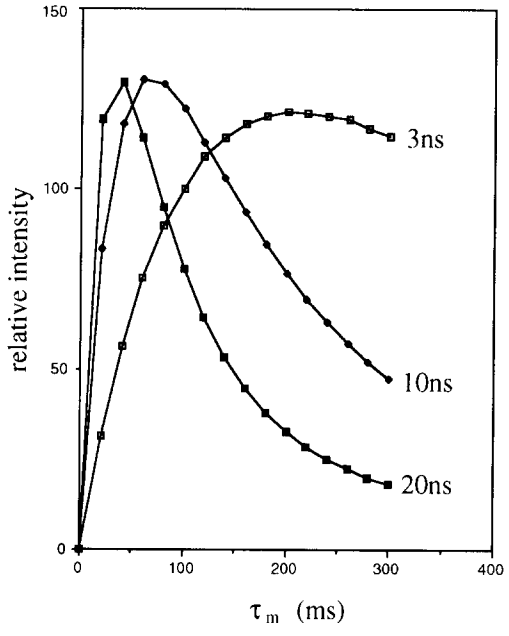

Fig. 1. (a) Simulation of the dependence on the mixing time of a typical $d_{\alpha N}(i,i+1)$ cross peak in a β-sheet secondary structure for correlation times of 3, 10 and 30 ns, for a 500-MHz spectrometer frequency. The points were obtained from a complete relaxation matrix calculation of the protein eglin c.

prepared with a lot of effort and at high cost. In many cases this may be a consequence of insufficient purification and presence of proteinases. Thus more extensive purification may solve the problem.

The low sensitivity of NMR experiments is a major problem for solving large protein structures. To discuss this it is worthwhile to reflect on the architecture of NMR experiments used for structural studies of proteins. It is safe to say that NMR structures of larger proteins can only be solved if the proteins can be labeled with stable isotopes, such as ^{15}N and ^{13}C. Heteronuclear multidimensional NMR experiments are essential to solve the overlap and crowding problem in protein NMR. Whether such spectra can be recorded depends on the relaxation properties of the protein and the signal-to-noise ratio achievable in a given time. The well-known architecture of an n-dimensional NMR experiment is:

$$\text{preparation} - (\text{evolution} - \text{mixing})_{n-1} - \text{detection}$$

The crucial point is whether there is enough signal left at the beginning of the detection period. Mixing periods require a certain minimal length to be effective. On the other hand, the signal decays during the mixing times due to relaxation. The mixing periods in the most interesting multidimensional NMR experiments have similar characteristics. The transfer function initially increases linearly with the mixing time, τ_m, reaches a maximum and decays. In NOESY-type experiments the cross-peak intensity S_{AB} follows approximately the function:

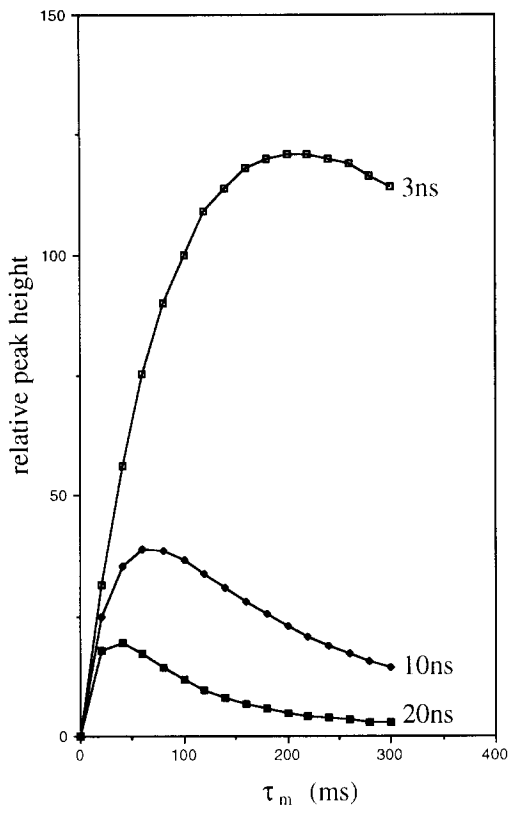

Fig. 1. (b) Plot of the peak heights as expected in a 1D transient NOE experiment. To obtain the relative peak heights the intensities from (a) are scaled with τ_c^{-1}.

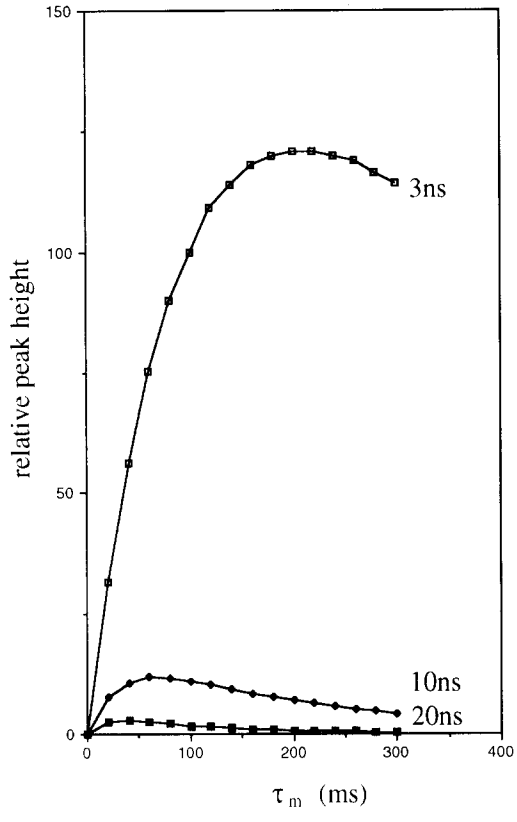

Fig. 1. (c) Plot of the expected peak heights in a 2D NOE experiment. To obtain the relative peak heights the intensities from (a) are scaled with τ_c^{-2}.

$$S_{AB}(\tau_m) = C\, e^{-R\tau_m}(1 - e^{2\sigma_{AB}\tau_m}) \quad (1)$$

R is a relaxation rate, σ_{AB} is the cross-relaxation rate and C is a constant. This is the well-known transient NOE curve. That the maximum of this curve reaches out of the noise floor is the first prerequisite that NOE cross peaks can be observed and used as structural constraints. For large proteins the initial slope (σ_{AB}) increases proportional to the correlation time τ_c. The maximum of the NOE build-up curve moves to shorter mixing times (see for example, Neuhaus and Williamson, 1989) with a τ_c^{-1} dependence, as can be shown readily from Solomon's equations (Solomon, 1955) or by simulation (see Fig. 1). As a rule of thumb, the maximum of the transient NOE curve is located approximately at the inverse auto-relaxation rate, $\tau_m^{max} \approx R_H(H_z)^{-1}$, moving to longer mixing times when spin diffusion is contributing to the cross peak. Here the auto-relaxation rate, $R_H(H_z)$, is the rate obtained after selective inversion of the proton as relevant for decay of diagonal peaks in NOESY spectra (see below). In theory, the value of the maximum NOE observable in a transient NOE experiment (NOESY) should be relatively independent of the correlation time for $\omega\tau_c \gg 1$. Figure 1a shows a complete relaxation matrix simulation of a

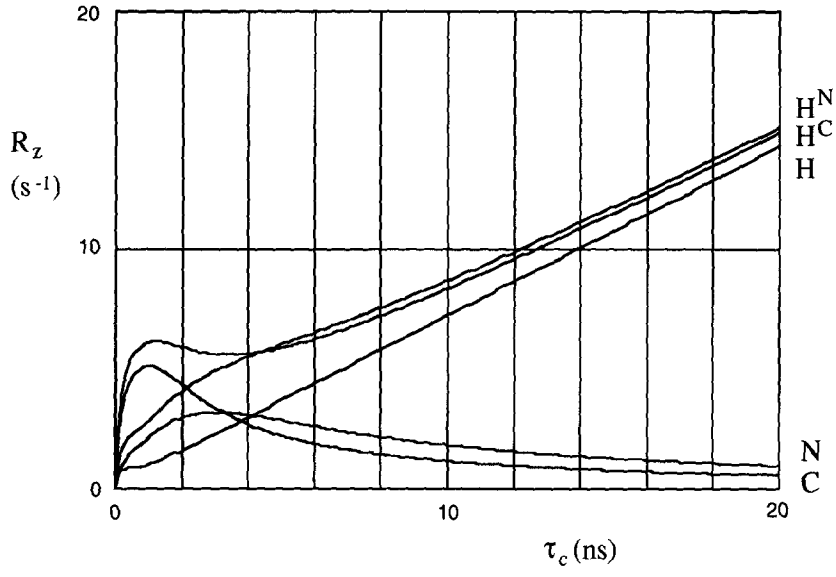

Fig. 2. (a) Simulation of longitudinal relaxation rates, R_z, vs. correlation time assuming Lorentzian spectral density functions, for a 500-MHz spectrometer. Simulations for a methine carbon (C), an amide nitrogen (N), an amide proton (H^N), a methine proton (H^C) and a proton neither attached to a ^{13}C or ^{15}N are given. The simulations include dipole-dipole and chemical shift anaisotropy relaxation. The proton included in the calculations is assumed to be in dipolar contact with three other protons with distances of 2.2, 2.6 and 3.5 Å, respectively.

sequential $d_{\alpha N}$ (i, 1 + 1) NOE cross peak for the protein eglin c with different hypothetical correlation times of 3, 10 and 20 ns, respectively. Empirically, the position of the NOE maximum for this $d_{\alpha N}$ cross peak in eglin c (Tyr^{56}-Phe^{55}) follows the linear relation:

$$\tau_m^{max} = 0.64 * \tau_c^{-1} \qquad (\tau_m^{max} \text{ in s, } \tau_c \text{ in ns}) \tag{2}$$

It appears that the height of the maximum intensity in the transient NOE curve, i.e. the maximum achievable NOE intensity, is essentially independent of the correlation time. However, the intensity of a cross peak is not really relevant from the aspect of detectability of a cross peak. It is the height of a cross peak that is crucial. This quantity is proportional to the transverse relaxation time, T_2. For large proteins, T_2^H is proportional to the inverse correlation time, τ_c^{-1}. Figure 1b shows a plot of the same relative NOE time dependence, but normalized with the inverse correlation time. This corresponds to the peak heights in a 1D transient NOE experiment. In a 2D NOESY, the relative cross peak heights are proportional to $(T_2^H)^2$. Figure 1c shows the same data but normalized with τ_c^{-2}. In a heteronuclear 3D NOE experiment, the peak heights scale down in proportion with the transverse relaxation times of the heteronucleus involved. Generally, in a nD experiment the peak heights scale with τ_c^{-n}. Of course, the method of data processing will have some influence on the actual peak heights. The present wide-spread attitude to truncate evolution periods early in multidimensional NMR experiments lets linewidths in these dimensions appear similar and independent of molecular weight. A more thoughtful strategy for data acquisition and advanced data processing routines represents a non-exhausted resource for boosting the sensitiv-

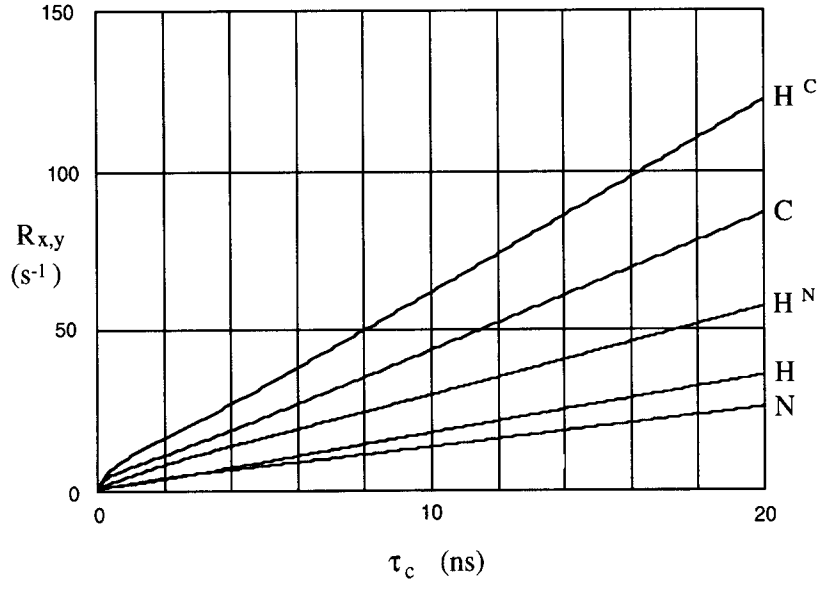

Fig. 2. (b) Simulation of transverse relaxation rates, $R_{x,y}$, of in-phase coherences vs. correlation time for the same systems as in (a).

ity of multidimensional NMR experiments and may have a major impact for NMR spectroscopy of larger proteins.

In coherence transfer experiments the crucial part of the transfer function is usually of the following type:

$$S_{AB}(\tau_m) = e^{-R\tau_m} * \sin(\pi J_{AB} \tau_m) * \Pi_C \cos(\pi J_{AC} \tau_m) \quad (3)$$

The position of the first maximum of this transfer function is determined by the relaxation properties. The tuned delay τ_m can be adjusted to match the maximum. There are few possibilities to influence the relaxation properties of a protein, such as to design the experiments in a way that coherences or spin orders are used that have the most favorable relaxation behavior. Otherwise, the only help to overcome the molecular weight limit is to increase the sensitivity. Improvements of sensitivity will come from higher field strengths, and more importantly from improved probe technology, larger sample volumes, better rf homogeneity and reduced amplifier noise. Probably gradient-enhanced spectroscopy will have some impact on increasing the sensitivity for studies of large proteins by elimination of subtraction noise.

In order to optimally design pulse sequences it is worthwhile to recall the expected dependence of relaxation rates of protons H and heteronuclei S on the molecular size. The functional dependence of relaxation rates on spectral density functions are well established (see for example Peng and Wagner, 1992, 1993). Figure 2 shows some estimates of relaxation rate dependence on the correlation time, which were calculated using those relations with simple Lorentzian spectral density functions. In this estimate ^1H-^{15}N and ^1H-^{13}C dipolar interactions are considered, as well as CSA relaxation for ^{15}N and ^{13}C (Peng and Wagner, 1993). Relaxation of the H^N or H^C protons is assumed to be due to heteronuclear dipolar relaxation as well as dipole interaction with three

protons at distances of 2.2, 2.6 and 3.5 Å, respectively. This may be an underestimation of the number of neighboring protons, and the relaxation of the protons in a real protein could be faster. Nevertheless it may serve for discussing the effect of molecular size on relaxation. Figure 2 shows simulations of relaxation rates vs. τ_c. Typical correlation times are 3.4 ns for an 8-kDa protein (eglin c; Peng and Wagner, 1992) or 13 ns for a 21.5-kDa protein (human dihydrofolate reductase; Kördel et al., 1993). Thus all protein structures solved so far with high precision, with molecular weights of 18 kDa or less, fall in the left-hand half of the diagrams in Fig. 2.

In large proteins, longitudinal proton relaxation rates, $R(H_z)$, are dominated by the values of spectral density functions at the differences of proton frequencies, $J(\omega(H_A) - \omega(H_B))$. This quantity increases with molecular size and is responsible for the fact that the longitudinal proton relaxation rates increase with the size of the protein (Fig. 2a). This is favorable for rapid repetition of pulse sequences that start with proton polarization. However, one has to be aware that the increase of the proton longitudinal relaxation rates with the correlation time is valid only for selective inversion of individual protons. This is the relaxation rate relevant for the decay of diagonal peaks in NOESY experiments. It is due to a distribution of energy to other spins (spin diffusion) rather than to the lattice. Non-selective inversion would lead to a similar dependence of the longitudinal proton relaxation rates as for ^{13}C and ^{15}N (see for example Kalk and Berendsen, 1976). This means, the non-selective proton relaxation rates will decrease with increasing molecular weight. Nevertheless, the longitudinal proton relaxation will always be faster than that of the carbon or nitrogen they are attached to, due to the additional dipolar interaction with other protons.

Longitudinal relaxation rates of ^{15}N and ^{13}C in large proteins are dominated by $J(\omega_N)$ and $J(\omega_C)$, respectively, and decrease with increasing size of the protein. Therefore, any start of a pulse sequence with polarization of a heteronucleus is unfavorable because they would require long recycling delays. The main foe of sophisticated pulse sequences is transverse relaxation. It is dominated by J(0) which increases with molecular weight. The dependence on τ_c for relaxation rates of in-phase coherences is shown in Fig. 2b. The interaction with the heteronucleus has a considerable influence on the proton relaxation. It is worth mentioning that the protons have always larger line widths than the ^{13}C or ^{15}N nuclei they are attached to. In-phase coherence will generally relax more slowly than antiphase signals (not shown) since the relaxation rates of the nuclei involved in antiphase coherence add up. Heteronuclear H-X two-spin coherence also relaxes faster than inphase X-coherence. While this multiple quantum relaxation rate does not contain a $J_{HX}(0)$ term, it is dominated by the fast transverse relaxation of the proton involved, which is caused by the $J_{HH}(0)$ terms (Peng and Wagner, 1993). Approximate estimates for the relaxation rates of anti-phase and multiple-quantum coherences can be obtained by adding the corresponding relaxation rates of Fig. 2. More precise expressions are given for example in Peng and Wagner (1993). Overall, the line width predicted from the tumbling of a rigid sphere characterized by a single correlation time seems to be not so large that they would be the major obstacle for working with larger proteins. The major factor hampering spectroscopy of large proteins is that lines are broadened by other effects, and not finding the right conditions for a protein has a much more severe effect on the quality of NMR spectra than the size of the protein.

Conditioning of the samples is *the* key to solving larger protein structures. While the size of a protein is a limitation for NMR spectroscopy, there are many cases where proteins show very poor spectra although they are within the molecular weight range below 20 kDa. Understanding

the reasons for this behavior and finding means to handle this will be as important as pushing the molecular weight limit for the well-behaved proteins. Empirically the behavior of proteins can be modulated by changing the pH, the ionic strength or the buffer. Proteins that like to bind a cofactor, an inhibitor, a substrate or a metal, seem to exhibit significantly better spectra after binding the counterpart. Thus, proteins that exhibit poor spectra may be missing an essential counterpart that has not yet been identified. The resulting less compact structure seems to lead to increased interaction between the protein molecules (non-specific aggregation) and increase the overall correlation time. Another means to reduce aggregation can be sought by addition of solubilizing molecules such as detergents. Little systematic work has appeared in the literature so far. In a few cases the tendency to aggregate could be reduced by cutting off flexible tails or by mutating single amino acids. This approach was successful, for example, in improving the quality of spectra for insulin (Weiss et al., 1991), guided by the knowledge of the crystal structures of oligomeric forms of insulin. It remains to be seen whether such an approach will be successful for larger proteins without prior knowledge of the structure and without knowledge about residues involved in forming intermolecular contacts.

Efficient data analysis is crucial for large proteins. Multiple-resonance and multidimensional experiments largely reduce the overlap problem. Nevertheless, assigning the spectrum of a protein above 15 kDa is still a major effort, and only a few proteins of this size have been assigned. Handling significantly larger proteins will only be reasonable if the assignment process and the extraction of the conformational constraints can be aided by computer tools. The first promising software packages are already available (see, for example, Eccles et al., 1991), and additional or improved software can be expected. In theory, it seems straightforward to perform a peak picking in a number of triple resonance experiments and to obtain assignments from a smart sorting routine. Similarly, peak picking in 3D and 4D NOESY spectra could provide extensive lists of constraints. In practice, spectra artifacts render this process difficult. Thus development of methods for elimination or recognition of artifacts will be equally important to further improvement of the assignment software.

There is hope that a major gain in sensitivity can be achieved from non-linear sampling and appropriate data processing routines (see for example Barna et al., 1987). Since the beginning of 2D NMR, sampling of the time domain signals in the indirect dimensions was almost exclusively done by linearly incrementing the evolution delays. A typical 3D NOE experiment may consist of $256 \times 64 \times 1024$ linearly sampled data points. This set of ca. 16 million data points is several orders of magnitude larger than the largest set of constraints we could hope to use for the best NMR structures. Considering this, there is hope that the number of time points to be collected in the indirect dimensions can be significantly reduced to free time for additional signal averaging at each time point. With the methods available now it appears that in some 2D NMR experiments the number of time points acquired in the indirect dimension can easily be reduced by a factor of four when linear sampling and the discrete Fourier transformation are abandoned and replaced with non-linear sampling and alternative data processing routines (Schmieder et al., 1993). Larger reduction in measuring times or larger improvements of sensitivity, respectively, can be expected for heteronuclear 3D and 4D experiments.

Isotope labeling with ^{13}C and ^{15}N is essential for studies of proteins above 10 kDa. Essentially only labeling in *E. coli* was used so far. This is relatively inexpensive if the expression level is high. Proteins with many disulfide bonds sometimes do not fold correctly in *E. coli* expression systems.

Significant progress has been reported recently using other expression systems, such as Chinese hamster ovary cells (Hansen et al., 1992; Archer et al., 1993), insect cells or yeast. While the first attempts to label in such systems seem horribly expensive, these costs may come down to the present costs of labeling in *E. coli*. Considering past developments, it seems possible that more efficient expression systems will become available in the future. This will be important expanding the technique to proteins that have been out of reach for structural NMR studies. Deuteration of protein samples has been proposed as a method to facilitate assignments and to reduce line widths in proteins (Markley et al., 1968). It was demonstrated in a few cases that deuteration can be used to assign proteins (LeMaster and Richards, 1988; Torchia et al., 1988). Surprisingly, this method has found little general application, probably because it is more expensive than other labeling methods; isotope shifts may cause complications and the coupling of the protons with the quadrupolar nucleus deuterium may cause line broadening due to scalar relaxation. Furthermore, assignments and collection of NOE distance constraints with 3D ^{15}N or ^{13}C dispersed NOE and TOCSY experiments, as well as with triple resonance experiments have worked well for proteins up to 20 or 25 kDa. It may be that the time for deuteration as a method to facilitate spectroscopy is yet to come when really large and important protein structures need to be solved that cannot be elucidated with any other method.

Should NMR compete with X-ray crystallography? The development of NMR started almost half a century later than X-ray crystallography. To date, the technology for NMR structure analysis of proteins has reached a certain maturity; however, it is not yet as efficient as crystallography. The two hurdles in X-ray crystallography are growing good-quality crystals and solving the phase problem. As soon as these hurdles are crossed an X-ray structure can be solved relatively quickly, while solving an NMR structure for a sizeable protein, even after resonance assignments are obtained, is still a major and time-consuming effort. To date, as we are still far from semi-automatic spectra analysis, tackling very large protein structures seems to be rewarding only when there is a strong evidence that crystallography would not succeed. A more favorable situation is when NMR and crystallography can provide complementary information. NMR is more powerful in describing proteins that have not yet found their target, such as a substrate, an inhibitor or a DNA binding site. In a way this represents active forms of proteins that show more mobility to be able to adapt to a target molecule. An example is the structure analysis of the DNA binding domain of the transcriptional activator GAL4 (Baleja et al., 1992). Only 30 out of 65 residues adopt a well-defined structure. Nevertheless, the structure of this part of the protein could be solved. However, crystallization of this protein was not possible so far. When the protein binds DNA it folds up; it can be crystallized and a structure of the complex could be solved (Marmorstein et al., 1992). This example may illustrate the power of NMR with respect to X-ray crystallography, as well as the complementarity of the two techniques. Crystallography is more suited to solve structures that have found their target. In a way these could be considered 'resting' or 'satisfied' structures. Leaving the field of structure analysis for larger proteins all to X-ray crystallography could mean that we would miss many aspects of active protein states.

Several years ago it seemed that solving structures of proteins in the range of 30 kDa would certainly come soon. NMR structures of this size have not yet found their way into the literature. The problems seem to be larger than anticipated. However, basic developments along several lines are on the way that will lead to solving large protein structures. Increased spectrometer sensitivity, computer-assisted data analysis, novel sampling and data-processing routines, and

most importantly, a better understanding of what causes line broadening beyond the molecular size may be key developments in solving structures of larger proteins. What will be the molecular weight limit? Who knows. After the initial excitement about the fact that NMR *can* indeed solve protein structures at atomic resolution an awareness of the unique information obtainable only from NMR will probably shift the focus from straight structure determination to aspects of mobility, ligand binding, or mapping of interaction sites.

ACKNOWLEDGEMENTS

The author is grateful to Dr. Ju-Xing Yang for providing the data of Fig. 1 and Dr. Peter Schmieder for critically reading the manuscript, as well as to Drs. Jonathan Lee and Jeffrey W. Peng for interesting discussions. Research on which this article is based was supported by NIH and NSF.

REFERENCES

Archer, S.J., Bax, A., Roberts, A.B., Sporn, M.B., Ogawa, Y., Piez, K.A., Weatherbee, J., Tsang, M., Lucas, R., Zheng, B., Wenker, J. and Torchia, D.A. (1993) *Biochemistry,* **32,** 1152–1163.

Baleja, J.D., Marmorstein, R., Harrison, S.C. and Wagner, G. (1992) *Nature,* **356,** 450–453.

Barna, J.C.J., Laue, E.D., Mayger, M.R., Skilling, J. and Worrall, S.J.P. (1987) *J. Magn. Res.,* **73,** 69–77.

Clore, G.M., Wingfield, P.T. and Gronenborn, A. (1991) *Biochemistry,* **30,** 2315–2323.

Eccles, C., Güntert, P., Billeter, M. and Wüthrich, K. (1991) *J. Biomol. NMR,* **1,** 111–130.

Fairbrother, W.J., Gippert, G.P., Reizer, J., Saier, M.H. and Wright, P.E. (1992) *FEBS Lett.,* **296,** 148–152.

Grzesiek, S., Döbel, H., Genz, R., Garotta, G., Labhardt, A.M. and Bax, A. (1992) *Biochemistry,* **31,** 8180–8190.

Hansen, A.P., Petros, A.M., Mazar, A.P., Pederson, T.M., Rueter, A. and Fesik, S.W. (1992) *Biochemistry,* **31,** 12713–12718.

Kalk, A. and Berendsen, H.J.C. (1976) *J. Magn. Reson.,* **24,** 343–366.

Kördel, J., Peng, J.W. and Wagner, G. (1993) in preparation.

LeMaster, D.M. and Richards, F.M. (1988) *Biochemistry,* **27,** 142–150.

Markley, J.L., Potter, I. and Jardetzky, O. (1968) *Science,* **161,** 1249–1251.

Marmorstein, R., Carey, M., Ptashne, M. and Harrison, S.C. (1992) *Nature,* **356,** 408–414.

Neuhaus, D. and Williamson, M.P. (1989) *The Nuclear Overhauser Effect in Structural and Conformational Analysis,* VCH Publishers, New York.

Peng, J.W. and Wagner, G. (1992) *Biochemistry,* **31,** 8571–8586.

Peng, J.W. and Wagner, G. (1993) *Methods in Enzymology,* in press.

Schmieder, P., Stern, A.S., Wagner, G. and Hoch, J.C. (1993) *J. Biomol. NMR,* in press.

Solomon, I. (1955) *Phys. Rev.,* **99,** 559–565.

Stockman, B.J., Nirmala, N.R., Wagner, G., Delcamp, T.J., DeYarman, M.T. and H. Freisheim, J.H. (1992) *Biochemistry,* **31,** 218–229.

Torchia, D.A., Sparks, S.W. and Bax, A. (1988) *J. Am. Chem. Soc.,* **110,** 2320–2321.

Weiss, M.A., Hua, Q.-X., Lynch, C.S., Frank, B.H. and Shoelson, S.E. (1991) *Biochemistry,* **30,** 7373–7389.