Biochemistry

 Volume 27, Number 15 July 26, 1988

Perspectives in Biochemistry

Protein Structures from NMR

R. Kaptein,*,‡ R. Boelens,‡ R. M. Scheek,§ and W. F. van Gunsteren§

Organic Chemistry Laboratory, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands, and Physical Chemistry Laboratory, University of Groningen, Nijenborg 16, 9747 AG Groningen, The Netherlands

Received February 2, 1988; Revised Manuscript Received March 16, 1988

In the last few years it has become possible to determine the three-dimensional structure of small proteins in solution by NMR methods. This is mainly due to advances in NMR technology such as the development of high-field spectrometers and new two-dimensional (2D) NMR techniques (Ernst et al., 1987). Progress has also been made to overcome the main bottleneck in the structure analysis of a protein, i.e., the assignment of ^{1}H resonances (Wüthrich, 1986). Finally, methods have been developed for structure determination on the basis of distance constraints from NMR such as distance geometry (Havel et al., 1979, 1983; Braun & Go, 1985) and restrained molecular dynamics (van Gunsteren et al., 1983; Kaptein et al., 1985; Clore et al., 1985). As a result, the structures of about 10 proteins, for which no crystallographic data were previously available, have now been determined by NMR.

In this paper we will review the methods for protein structure determination on the basis of high-resolution NMR data. These will be illustrated with results from our laboratory on the DNA-binding domain (or headpiece) of the *Escherichia coli lac* repressor, a protein domain consisting of the 51 N-terminal amino acid residues.

The primary source of information, on which NMR structures of proteins are based, is the nuclear Overhauser effect (NOE) (Noggle & Schirmer, 1971). NOE's between protons in a protein can be most conveniently measured by 2D NOE spectroscopy (or NOESY) (Jeener et al., 1979). The origin of the NOE is dipolar cross-relaxation between protons. Because of the weak proton magnetic moment and the r^{-6} distance dependence of the effect, NOE's can only be measured between protons that have relatively short distances (<5 Å). By use of suitable calibration procedures the NOE's can be translated into proton–proton distance constraints.

Another useful NMR parameter is the J-coupling constant between vicinal protons. Via relationships such as those first proposed by Karplus (1959) [cf. Pardi et al. (1984)] information is obtained about dihedral angles. However, only for large (8–10-Hz) J couplings are these relations unambiguous. For intermediate values uncertainties may arise, first, because several dihedral angles may belong to a certain J coupling and, second, because they may be the result of motional averaging.

Thus, the structural constraints derived from NMR, i.e., proton–proton distances and dihedral angles, are short range in nature. For a small protein of, say, 60 amino acid residues, typically 200–300 of these constraints can be obtained. Is this enough for a structure determination? One might argue that the conformation of a 60 amino acid protein is fully specified by ca. 240 dihedral angles (for each amino acid ϕ, and ψ for the backbone and on average two χ's for the side chain). Therefore, a few hundred NMR constraints well distributed over the molecule in addition to those from the covalent structure would seem sufficient. However, since the constraints from NMR come with errors, the propagation of these errors may lead to inaccuracies in overall properties. For globular proteins, in which the peptide chain falls back on itself, this does not seem to be a problem. Model calculations on bovine pancreatic trypsin inhibitor (BPTI) with data from the crystal structure simulated to represent a NMR data set with a few hundred approximate proton–proton distances have shown that the structure can be defined to within 1–2-Å rms difference from the crystal structure (Havel & Wüthrich, 1985). Global properties such as the radius of gyration were also in good agreement. For nonglobular molecules, for instance, double-stranded oligonucleotides, this may be quite different. Here, one would expect that local conformations can be determined well from NMR but long range properties less accurately.

Procedure for Structure Determination from NMR Data

It should be noted that a single generally accepted method for structure determination of biomolecules by NMR does not exist. Different authors have used different procedures, and new algorithms are still being developed. The protocol that

‡ University of Utrecht.
§ University of Groningen.

0006-2960/88/0427-5389$01.50/0

we have applied to *lac* repressor headpiece is shown in Scheme I. The first two steps, consisting of ^{1}H resonance assignment and determination of distance and dihedral angle constraints, are common to all procedures. Steps 3 and 4 are suitable to address questions such as how unique are the structures obtained, how well do they satisfy the experimentally derived constraints, and how reasonable are they from the point of view of energetics?

Scheme I: Protocol for Protein Structure Determination from NMR

(1) Assign ^{1}H resonances.

(2) Determine proton–proton distance constraints and dihedral angle constraints from NOE's and J couplings, respectively.

(3) Calculate family of structures by using geometric constraints only (experimental constraints plus covalent structure), for instance, with distance geometry (DG) and distance bounds driven dynamics (DDD).

(4) Refine these structures by using geometric constraints and potential energy functions, for instance, with restrained energy minimization (REM) and restrained molecular dynamics (RMD).

In the following we shall discuss the various steps of Scheme I in some detail.

^{1}H Resonance Assignments. A necessary requirement for the structural analysis of a protein is the assignment of the great majority of its proton resonances. For small proteins (M_r <10 000) that do not aggregate at millimolar concentrations this can be accomplished by using a combination of various 2D NMR experiments. The procedure for the so-called sequential assignment of protein ^{1}H NMR spectra is extensively described by Wüthrich (1986). Briefly, two main classes of 2D experiments are used. In the first class off-diagonal cross-peaks arise only between protons connected through J-coupling networks, with COSY (correlated spectroscopy) as the prime example. Another very useful experiment in this category is TOCSY (total correlation spectroscopy; Braunschweiler & Ernst, 1983) or HOHAHA (homonuclear Hartmann–Hahn; Davis & Bax, 1985). In these spectra patterns of cross-peaks can be traced between pairs of J-coupled protons as in COSY or between several protons within an amino acid chain as in HOHAHA.

In the second class of 2D NMR experiments (2D NOE) cross-peaks connect protons that are spatially in close proximity (distance <5 Å). The analysis usually starts with a search for cross-peak patterns belonging to the spin systems of types of amino acids. These are then connected through cross-peaks in a 2D NOE spectrum between neighboring amino acids in the polypeptide chain. Useful short distances involving backbone protons that are manifested in 2D NOE spectra are those between C_α, C_β, and amide protons of one residue and the amide proton of the next residue ($d_{\alpha N}$, $d_{\beta N}$, and d_{NN}, respectively). Often the sequential assignment procedure is redundant so that many internal checks are possible. This makes the assignment unambiguous. However, it does not lead to stereospecific assignments for diastereotopically related protons such as those of methylene groups or methyl groups of valine and leucine. Sometimes stereospecific assignments of these protons are possible by using a combination of vicinal J-coupling and NOE information (Zuiderweg et al., 1985a; Hyberts et al., 1987).

Distance and Dihedral Angle Constraints. Proton–proton distance constraints are most conveniently derived from cross-peak intensities in 2D NOE spectra. The initial buildup rate of these cross-peaks in spectra taken at different short mixing times is proportional to the cross-relaxation rate σ_{ij} between protons i and j. Therefore, these cross-relaxation rates can be measured either from a single 2D NOE spectrum taken with a sufficiently short mixing time or, more accurately, from a buildup series recorded with various mixing times. For a rigid isotropically tumbling molecule σ_{ij} is simply related to the distance d_{ij} and the correlation time τ_c:

$$\sigma_{ij} \propto \tau_c d_{ij}^{-6} \tag{1}$$

Therefore, with a known calibration distance d_{cal} the proton–proton distances follow from the relation

$$d_{ij} = d_{cal}(\sigma_{cal}/\sigma_{ij})^{1/6} \tag{2}$$

In practice eq 1 and 2 are only approximately valid. There are two main problems associated with accurate determination of proton–proton distances. The first is that of indirect magnetization transfer or "spin diffusion". In reality the NOE cross-peaks are the result of multispin relaxation and only in the limit of extremely short mixing times (where the signal-to-noise ratio is poor) is the two-spin approximation of eq 1 valid. In principle the effect of spin diffusion can be calculated, and procedures based on a full relaxation matrix treatment are being developed to solve this problem (Keepers & James, 1984; Boelens et al., 1988). The second is that proteins are not rigid bodies, and intramolecular mobility leads to nonlinear averaging of distances and to different effective correlation times for different interproton vectors in the molecule. Internal motions occur over a wide range of time scales, and only the fast fluctuations (up to a few hundred picoseconds) can presently be simulated by molecular dynamics calculations. The slower motions are difficult to handle and therefore constitute the most serious source of error in distance determination from NOE's.

For this reason, the approach that is usually taken is that of translating the NOE information into distance ranges (e.g., 2–3, 2–4, and 2–5 Å for strong, medium, and weak NOE's, respectively) rather than attempting to obtain precise distances. Alternatively, as was done in the case of *lac* headpiece (Kaptein et al., 1985), a single cutoff distance was used at 3.5 Å (corresponding to a range of 2–3.5 Å).

In case stereospecific assignments of methylene and methyl groups are not known, the distances involving these protons have to refer to pseudoatoms, and a corresponding correction term has to be added that allows for the maximum possible error. For instance, for a methylene group the pseudoatom position is defined at the geometric mean of the CH_2 proton positions, and a correction of 0.9 Å is added to the distance constraints involving these protons.

The absence of NOE's between assigned protons also contains useful information as it means that the distance between the protons is larger than ca. 4–5 Å. This non-NOE information has been used in the work on the *lac* headpiece structure (de Vlieg et al., 1986). Great care should be taken, however, in interpreting the absence of NOE's, since it may also result from local mobility.

In favorable cases dihedral angle constraints can be obtained from three-bond J couplings. These can be obtained from the fine structure of cross-peaks in COSY-like spectra recorded with high digital resolution (Hyberts et al., 1987). An important example is the three-bond coupling $^3J_{HN\alpha}$ between amide and C_α protons, which gives a measure of the backbone torsion angle ϕ. For helical regions $^3J_{HN\alpha}$ is small (ca. 4 Hz), while for extended chain conformations such as in β-sheets

it takes large values (9–10 Hz). Usually the large J couplings (8–10 Hz) are the most useful source of information, because J couplings smaller than the linewidth (typically 5 Hz or larger) cannot be reliably measured due to cancellation effects in antiphase multiplets (Neuhaus et al., 1985). Furthermore, as was mentioned above, the interpretation of the larger coupling constants in terms dihedral angles is less ambiguous.

Structure Calculations Based on Geometric Constraints (Distance Geometry and Distance Bounds Driven Dynamics). The metric matrix distance-geometry (DG) algorithm (Blumenthal, 1970; Havel et al., 1979, 1983; Havel & Wüthrich, 1985) was known well before protein structure determination by NMR became possible. Thus far, it is the only method that does not rely on some starting conformation and is therefore free from operator bias. The DG procedure amounts to the following. First, upper and lower bound matrices **U** and **L** are set up for all atom–atom distances of the molecule. Some of the elements u_{ij} and l_{ij} follow from standard bond lengths and bond angles of the covalent structure and from experimentally found distance ranges from NOE's and J-coupling constants. A bound smoothening procedure using triangle inequalities extends the constraints to all elements of **U** and **L**. Then, a distance matrix **D** is set up with distances chosen randomly between upper and lower bounds, $l_{ij} \leq d_{ij} \leq u_{ij}$. The so-called "embedding" algorithm then finds a 3D structure corresponding to the distances d_{ij}. This structure must be optimized with an error function consisting of chirality constraints (chiral centers sometimes come out the wrong way) and a distance constraint error function. This forces the amino acid side chains to adopt the correct chirality and the distances to satisfy the upper and lower bound criteria, although usually the DG structures still contain a number of violations of the distance bounds. By repetition of the DG calculations several times the random step of choosing the distance matrix **D** between upper and lower bounds allows different structures to be obtained, and therefore it can be judged how uniquely the structure is determined by the constraints.

Another method, also termed distance geometry but using a quite different mathematical procedure, was suggested by Braun and Go (1985). Here, the protein conformation is calculated by minimizing a distance constraint error function. Special features of the method are that dihedral angles are used as independent variables rather than Cartesian coordinates and that it uses a variable target function, first satisfying local constraints (between amino acids nearby in the polypeptide chain) and later including long-range constraints. Usually one starts with various initial conformations obtained by taking random values for the dihedral angles. A comparison between a metric matrix distance-geometry algorithm DISGEO (Havel & Wüthrich, 1984) and the variable target function algorithm DISMAN (Braun & Go, 1985) has shown that the efficiency and convergence properties of both methods are rather similar (Wagner et al., 1987).

Although the distance-geometry method does not need starting structures and is therefore not subject to operator bias, this does not mean that it samples the allowed conformational space (consistent with the bounds) in a truly random fashion. In fact, it was noted by Havel and Wüthrich (1985) in model calculations on bovine pancreatic trypsin inhibitor that the DG structures tend to be somewhat more expanded than the crystal structure, from which the constraints had been derived. We have also noticed this effect in our work on *lac* headpiece (Scheek et al., 1988). In regions that are relatively unconstrained the DG algorithm tends to produce extended conformations for the backbone and side chains (vide infra). It was further noted that in restrained molecular dynamics refinement a larger variation of conformations was found in regions with few NOE's in spite of the fact that the structures now have to satisfy criteria of low potential energy as well. This led one of us (R.M.S.) to the idea of improving on the sampling properties of the DG procedure by adding a simplified MD calculation with only geometric constraints as the driving potential. In this so-called distance bounds driven dynamics (DDD) algorithm (Scheek & Kaptein, 1988) Newton's equations of motion

$$m_i \ddot{\mathbf{r}}_i = \mathbf{F}_i \tag{3}$$

are solved with the forces given by

$$\mathbf{F}_i = -\partial V / \partial \mathbf{r}_i \tag{4}$$

where the potential function now does not contain any energy terms but is taken proportional to the DG error function

$$V = K_{dc}[\sum_{d_{ij}>u_{ij}} (d_{ij}^2 - u_{ij}^2)^2 + \sum_{l_{ij}>d_{ij}} (l_{ij}^2 - d_{ij}^2)^2] \tag{5}$$

The time step for integration of Newton's equations (eq 3) should be chosen in accordance with the magnitude of the "force constant" K_{dc} (taken somewhat arbitrarily as 10000 $kJ \cdot mol^{-1} \cdot nm^{-4}$). Of course, the entirely nonphysical nature of the potential V means that the calculations cannot be interpreted as a simulation of a physical process. A DDD run using a DG structure as the starting conformation has the effect of "shaking up" the DG structure and thereby improving the sampling of conformation space, which is especially important in regions with few constraints.

Structure Refinement Including Energy Terms (Restrained Energy Minimization and Molecular Dynamics). The quality of the protein structures based on geometric contraints can be improved by taking energy considerations into account. For instance, in DG structures amino acid side chains often adopt eclipsed conformations, while in alkyl chains the energy of the staggered conformation is 10–15 $kJ \cdot mol^{-1}$ lower. Also, hydrogen bonds and salt bridges may not be formed unless they are specifically introduced as constraints. In restrained molecular dynamics (RMD) refinement structures are optimized simultaneously with respect to a potential energy function and to a set of experimental constraints. Of course the success of this method now depends to a large extent on the quality of the force field used. It is therefore important to realize the limitations and approximate nature of this force field, especially when the calculations do not include solvent molecules.

In RMD calculations eq 3 and 4 are integrated with the potential energy functions given by

$$V = V_{bond} + V_{angle} + V_{dihedr} + V_{vdW} + V_{coulomb} + V_{dc} \tag{6}$$

The first two terms tend to keep bond lengths and bond angles at their equilibrium values. V_{dihedr} is a sinusoidal potential describing rotations about bonds; for V_{vdW} (the van der Waals interaction) usually a Lennard–Jones potential is taken, and $V_{coulomb}$ describes the electrostatic interactions. The extra distance constraint term V_{dc} distinguishes RMD from more conventional MD simulations. It has the effect of pulling protons within the distance d_{ij}^0 (or the distance range) in accordance with the NOE observations. Usually a harmonic pseudopotential is chosen

$$V_{dc} = \tfrac{1}{2} K_{dc} (d_{ij} - d_{ij}^0)^2 \tag{7}$$

Other forms of V_{dc} are a half-harmonic potential (Kaptein et al., 1985), forms that contain a linear part at long distances, and repulsive terms for non-NOE's (de Vlieg et al., 1986).

Similarly, sinusoidal terms describing dihedral angle constraints from J-coupling constants can be included (de Vlieg et al., 1986).

A restrained dynamics calculation is usually preceded by restrained energy minimization (REM) using steepest descent or conjugate gradient methods to bring the energy down to an acceptable level. REM using the same potential energy function (eq 6) usually changes the conformation only slightly and cannot take it out of local minima. By contrast, RMD is able to overcome barriers of the order of kT because of the kinetic energy in the system and therefore has a much larger radius of convergence. RMD works as an efficient minimizer, since excess potential energy, converted to kinetic energy, is drained off by coupling the system to a thermal bath of constant temperature. It has been suggested (Brünger et al., 1986) that the RMD procedure could be used to obtain folded protein structures starting, for instance, from a fully extended polypeptide chain. Although apparently this was successful for model calculations on crambin (Brünger et al., 1986), it is our experience that this procedure does not work in general and is certainly not cost effective in terms of best use of computer time. In our view RMD should be considered as a structure refinement tool using DG or DDD structures as starting conformations.

Application to lac Headpiece

lac headpiece is a separate domain of the *lac* repressor that is responsible for recognition of a specific sequence of base pairs of *E. coli* DNA called the *lac* operator. No crystal structure is known for repressor or headpiece. On the basis of sequence homology it had been suggested that headpiece contains the helix–turn–helix structural motif present in other DNA-binding proteins such as cro and cI repressors of phage λ (Ohlendorf et al., 1982; Sauer et al., 1982).

The first step in the structure determination of *lac* headpiece was the sequential assignment of its ^{1}H resonances according the procedure described by Wüthrich (1986). For all amide and C_α protons except those of Ile-48 (due to an overlap problem) assignments could be made (Zuiderweg et al., 1983a, 1985b). Most of the side-chain protons have also been assigned, although for some of the long side chains of Lys, Arg, and Gln the assignments do not extend beyond the C_β protons (Zuiderweg et al., 1983a, 1985b). By making use of a combination of J coupling ($^3J_{\alpha\beta}$) and NOE's the prochiral methyl groups of valines-9, -20, -23, and -38 could be stereospecifically assigned (Zuiderweg et al., 1985a).

Inspection of assigned cross-peaks in the 2D NOE spectra of *lac* headpiece showed that it is a typical α-helical protein with no β-sheets. In three regions of the protein relatively strong NOE's were found that correspond to short distances d_{NN}, $d_{\alpha N}(i, i+3)$, and $d_{\alpha N}(i, i+4)$ prevailing in α-helices (Zuiderweg et al., 1983b). The three α-helices are found in the regions 6–13, 17–25, and 34–45 of the polypeptide. Consistent with this is the observation of slow H–D exchange for the amide protons in the helical hydrogen bonds (Boelens et al., 1985). It should be noted that with the combination of NOE and exchange data the secondary structure of proteins can be established quite reliably.

The next step was the tertiary structure determination. This was done on the basis of 169 NOE's observed in 2D NOE spectra taken at a relatively short mixing time of 50 ms. As a distance calibration we used the NOE intensities of a series of cross-peaks corresponding to the distances $d_{\alpha N}$ and $d_{\alpha N}(i, i+3)$ of α-helical regions, which are both approximately 3.5 Å. Thus, the NOE's were essentially converted to distance ranges of 1.8–3.5 Å (the lower limit being the sum of the van der Waals radii). To these ranges pseudoatom corrections were applied for groups that show dynamic averaging effects (methyl groups and tyrosine rings) and in cases where the stereospecific assignments of diastereotypically related protons were not known. To this data set 17 hydrogen-bond constraints were added for those slowly exchanging amide protons for which the H-bond acceptor was known, i.e., for the α-helical regions. The initial structure determination was carried out with the restrained molecular dynamics method, with a model-built conformation as the starting structure (Kaptein et al., 1985; Zuiderweg et al., 1985c). A relatively low value for the force constant K_{dc} of 250 kJ·mol^{-1}·nm^{-2} was used, allowing for excess distances ($d_{ij} - d_{ij}^0$) of ca. 1 Å before an energy penalty is felt at the level of kT. However, larger values for K_{dc} up to 4000 kJ·mol^{-1}·nm^{-2} were sometimes found to speed up convergence. At a later stage the absence of NOE's between assigned protons was also taken into account (de Vlieg et al., 1986). Thus, repulsive pseudopotential terms for a carefully selected set of 9529 non-NOE's were included. Although the number of non-NOE's may seem large, the information content is in fact rather low, since most non-NOE's are trivial in a structure that is already approximately correct. After a RMD run of 60 ps (without solvent) the resulting structure satisfied the experimental constraints very well and at the same time had a low energy (the energy dropped from +4074 kJ·mol^{-1} for the starting structure to −3091 kJ·mol^{-1} after RMD followed by REM). The remaining violations of the constraints were not larger than 0.5 Å, while the sum of all violations was 5.8 Å (de Vlieg et al., 1986). Figure 1 shows a stereo picture of a snapshot taken from the RMD run. The helix–turn–helix region consisting of helices I and II of the headpiece can be clearly seen with the third helix packing against these forming a hydrophobic core. The RMD run also indicated that the three-helical core of the protein is rather rigid, whereas the N-terminal and C-terminal region and also the loop between helix II and III showed higher mobility.

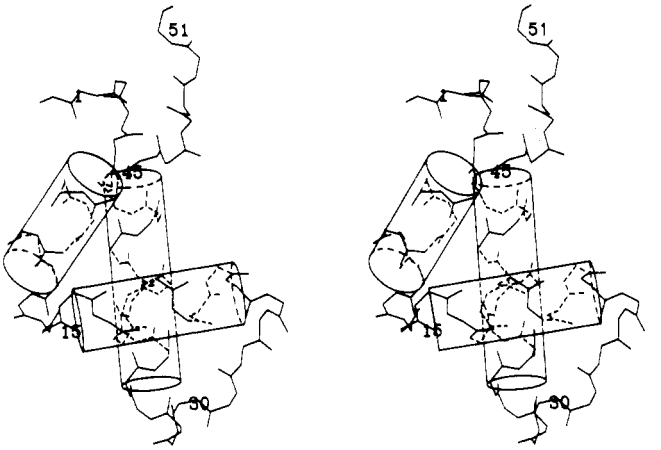

FIGURE 1: Stereo diagram showing the backbone atoms of *lac* repressor headpiece taken as a snapshot during a restrained molecular dynamics run. The cylinders represent the three α-helices at positions 6–13, 17–25, and 34–45.

Next, the question of uniqueness was addressed or, in other words, what is the range of conformations consistent with the constraints? A series of DG calculations were performed with the same set of distance constraints (Scheek et al., 1988). An overlay of 10 DG structures is shown in Figure 2A. The variation among these structures can be expressed as an average rms difference of the C_α-atom coordinates, which was 1.4 Å for all C_α's and 1.1 Å for those of residues 4–47. It can be noticed in Figure 2A that the N-terminal and C-terminal

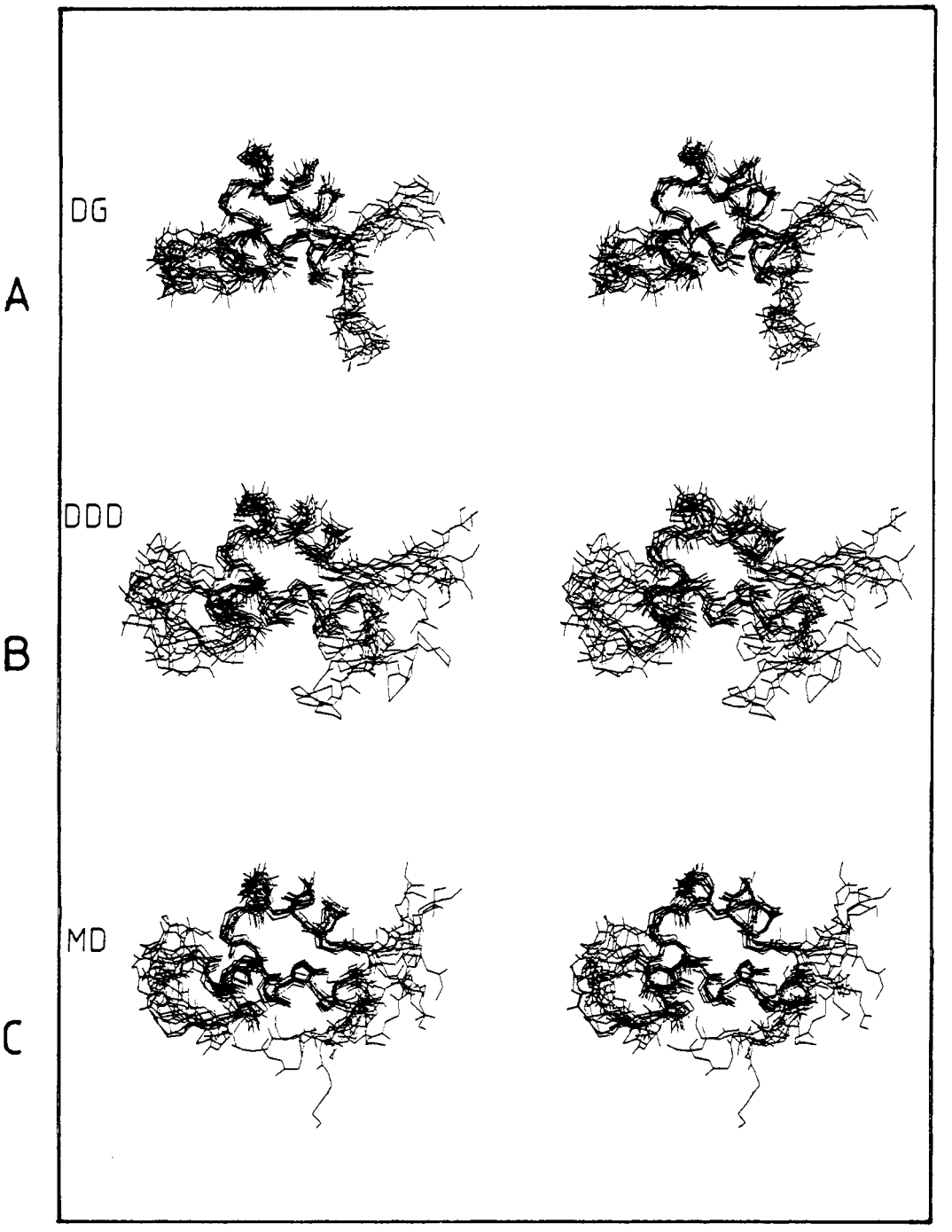

FIGURE 2: Family of 10 conformations of *lac* repressor headpiece obtained after distance geometry (A), distance bounds driven dynamics (B), and restrained molecular dynamics refinement (C). Stereo diagrams of the backbone atoms are shown.

peptide regions have a preference for extended backbone conformations in spite of the fact that there are no long-range NOE's in these regions that would fix the conformation with respect to the helical core. This is clearly an artifact of the metric matrix DG procedure. Starting with the DG structures, DDD runs were carried out with 1000 integration steps each (formally corresponding to 2 ps). Figure 2B clearly shows that the N-terminal and C-terminal peptide fragments adopt a much wider range of conformations, while the structures satisfied the constraints equally well. The average rms differences of the DDD structures were 3.0 Å for all C atoms and 2.0 Å for the C_α's of residues 4–47. Thus, the DDD procedure in combination with DG greatly improves the sampling properties compared to DG alone.

Finally, RMD refinement of these structures resulted in the family of structures shown in Figure 2C. Some convergence can be noticed in helical regions; the rms difference for the C_α atoms of residues 4–47 now becomes 1.7 Å (for the C_α's of the helices this value is 0.8 Å, showing that the helical core of the protein is particularly well determined). Calculated for all C_α atoms, the rms difference is still 3.0 Å, which is mainly caused by the large spread in conformations of the N-terminal and C-terminal peptides.

The *lac* headpiece structure formed the basis for further studies aimed at determining the way in which *lac* repressor recognizes its operator. From 2D NOE spectra of the complex of *lac* headpiece with a 14 base pair operator fragment, a low-resolution structural model for this complex could be derived (Boelens et al., 1987a,b). Although a full discussion of this work is outside the scope of this paper, we just mention the surprising result that the orientation of *lac* headpiece on the operator is opposite to that of all other repressors studied

so far by X-ray crystallography. Some of the specific interactions between base pairs and amino acid residues predicted by the NMR model (Boelens et al., 1987a) are now guiding genetic studies (Müller-Hill, private communication) that will be discussed elsewhere.

Perspectives for Protein Structure Determination in Solution

How much trust can one have in protein structures determined by NMR? There are a number of proteins now for which both X-ray and NMR structures are available. These include bovine pancreatic trypsin inhibitor (Wagner et al., 1987), the α-amylase inhibitor tendamistat (Kline et al., 1986), and potato carboxypeptidase inhibitor (Clore et al., 1987a). In these globular proteins the structures from NMR and X-ray agree to within 1–2 Å for the backbone atoms. The BPTI study further showed that the errors were not larger for long distances than for short ones in the protein in spite of the fact that NMR structures are based on short distance information. In one case there is a discrepancy between NMR and X-ray results. For two related metallothioneins different cysteine ligands seemed to be involving in binding the metal ions (Braun et al., 1986; Furey et al., 1986). The origin of this disagreement is presently not known.

Proteins for which structure determinations by NMR have been reported and for which the crystal structures are not known include (apart from *lac* headpiece) insectotoxin I_5A (Arseniev et al., 1984), proteinase inhibitor II from bull seminal plasma (Williamson et al., 1985), phoratoxin (Clore et al., 1987b), hirudin (Clore et al., 1987c), human epidermal growth factor (Cooke et al., 1987), and the globular domain of histone H5 (Clore et al., 1987d). All these proteins have a molecular weight below 10000. One may expect, however, that in the future this limit can be pushed to the 15000–20000 molecular weight range. This is because progress is being made in various stages involved in structure determination. Thus, work is going on in several laboratories on automated procedures for the peak assignment problem, which up to now has to be solved by hand. With respect to NMR methodology the development of magnets operating at higher fields is rather slow, but improvements come from new pulse sequences allowing more efficient 2D and even 3D experiments (Vuister & Boelens, 1987; Griesinger et al., 1987). Practical applications of 3D NMR to biomolecules are around the corner.

The accuracy of proton–proton distance measurements can be improved by using in the analysis of 2D NOE spectra relaxation matrix procedures that take spin diffusion into account (Keepers & James, 1984; Boelens et al., 1988). Weak NOE's can then also be interpreted more reliably so that longer distances can be measured and full advantage taken of the sensitivity of the 2D NOE experiment.

Finally, we should mention the potentially very powerful combination of NMR with the techniques of modern molecular biology. Proteins can be engineered to contain varous sorts of NMR labels such as enriched magnetic isotopes, which facilitate the study of their functionality. Once the spectra of a native protein has been analyzed, mutant proteins can be screened rather quickly by NMR. These possibilities have not yet been systematically explored, but there is no doubt this will change in the near future.

References

Arseniev, A. S., Kondakov, V. I., Maiorov, V. N., & Bystrov, V. F. (1984) *FEBS Lett. 165*, 57–62.

Blumenthal, L. M. (1970) *Theory and Applications of Distance Geometry*, Chelsea, New York.

Boelens, R., Gros, P., Scheek, R. M., Verpoorte, J. A., & Kaptein, R. (1985) *J. Biomol. Struct. Dyn. 3*, 269–280.

Boelens, R., Scheek, R. M., van Boom, J. H., & Kaptein, R. (1987a) *J. Mol. Biol. 193*, 213–216.

Boelens, R., Scheek, R. M., Lamerichs, R. M. J. N., de Vlieg, J., van Boom, J. H., & Kaptein, R. (1987b) in *DNA–Ligand Interactions* (Guschlbauer, W., & Saenger, W., Eds.) pp 191–215, Plenum, New York.

Boelens, R., Koning, T. M. G., & Kaptein, R. (1988) *J. Mol. Struct.* (in press).

Braun, W., & Go, N. (1985) *J. Mol. Biol. 186*, 611–626.

Braun, W., Wagner, G., Worgotter, E., Vasak, M., Kagi, J., & Wüthrich, K. (1986) *J. Mol. Biol. 187*, 125–129.

Braunschweiler, L., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521–528.

Brünger, A. T., Clore, G. M., Gronenborn, A. M., & Karplus, M. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3801–3805.

Clore, G. M., Gronenborn, A. M., Brünger, A. T., & Karplus, M. (1985) *J. Mol. Biol. 186*, 435–455.

Clore, G. M., Gronenborn, A. M., Nilges, M., & Ryan, C. A. (1987a) *Biochemistry 26*, 8012–8023.

Clore, G. M., Sukumaran, D. K., Nilges, M., & Gronenborn, A. M. (1987b) *Biochemistry 26*, 1732–1745.

Clore, G. M., Sukumaran, D. K., Nilges, M., Zarbock, J., & Gronenborn, A. M. (1987c) *EMBO J. 6*, 529–537.

Clore, G. M., Gronenborn, A. M., Nilges, M., Sukumaran, D. K., & Zarbock, J. (1987d) *EMBO J. 6*, 1833–1842.

Cooke, R. M., Wilkinson, A. J., Baron, M., Pastore, A., Tappin, M. J., Campbell, I. D., Gregory, H., & Sheard, B. (1987) *Nature (London) 327*, 339–341.

Davis, D. G., & Bax, A. (1985) *J. Am. Chem. Soc. 107*, 2821–2822.

De Vlieg, J., Boelens, R., Scheek, R. M., Kaptein, R., & van Gunsteren, W. F. (1986) *Isr. J. Chem. 27*, 181–188.

Ernst, R. R., Bodenhausen, G., & Wokaun, A. (1987) *Principles of Nuclear Magnetic Resonance in One and Two* Dimensions, Clarendon Press, Oxford.

Furey, W. F., Robbins, A. H., Clancy, L. L., Winge, D. R., Wang, B. C., & Stout, C. D. (1986) *Science (Washington, D.C.) 231*, 704–710.

Griesinger, C., Sorensen, O. W., & Ernst, R. R. (1987) *J. Am. Chem. Soc. 109*, 7227–7228.

Havel, T. F., & Wüthrich, K. (1984) *Bull. Math. Biol. 45*, 673–698.

Havel, T. F., & Wüthrich, K. (1985) *J. Mol. Biol. 182*, 281–294.

Havel, T. F., Crippen, G. M., & Kuntz, I. D. (1979) *Biopolymers 18*, 73–81.

Havel, T. F., Kuntz, I. D., & Crippen, G. M. (1983) *Bull. Math. Biol. 45*, 665–720.

Hyberts, S. G., Märki, W., & Wagner, G. (1987) *Eur. J. Biochem. 164*, 625–635.

Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546–4553.

Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R., & van Gunsteren, W. F. (1985) *J. Mol. Biol. 182*, 179–182.

Karplus, M. (1959) *J. Chem. Phys. 30*, 11–15.

Keepers, J. W., & James, T. L. (1984) *J. Magn. Reson. 57*, 404–426.

Kline, A. D., Braun, W., & Wüthrich, K. (1986) *J. Mol. Biol. 183*, 503–507.

Neuhaus, D., Wagner, G., Vasak, M., Kägi, J. H. R., & Wüthrich, K. (1985) *Eur. J. Biochem. 151*, 257–273.

Noggle, J. H., & Schirmer, R. E. (1971) *The Nuclear Overhauser Effect—Chemical Applications*, Academic, New York.
Ohlendorf, D. H., Anderson, W. F., Fischer, R. G., Takeda, Y., & Matthews, B. (1982) *Nature (London) 298*, 718–723.
Pardi, A., Billeter, M., & Wüthrich, K. (1984) *J. Mol. Biol. 180*, 741–751.
Sauer, R. T., Yocum, R. R., Doolittle, R. F., Lewis, M., & Pabo, C. O. (1982) *Nature (London) 298*, 447–451.
Scheek, R. M., & Kaptein, R. (1988) in *NMR in Enzymology* (Oppenheimer, N. J., & James, T. L., Eds.) Academic, New York (in press).
Scheek, R. M., de Vlieg, J., van Gunsteren, W. F., Boelens, R., Kaptein, R., Thomason, J., & Kuntz, I. D. (1988) *Biopolymers* (submitted for publication).
Van Gunsteren, W. F., Kaptein, R., & Zuiderweg, E. R. P. (1983) in *Nucleic Acid Conformation and Dynamics* (Olson, W. K., Ed.) Report of NATO/CECAM Workshop, pp 79–92, Orsay, France.
Vuister, G. W., & Boelens, R. (1987) *J. Magn. Reson. 73*, 328–333.
Wagner, G., Braun, W., Havel, T. F., Schaumann, T., Go, N., & Wüthrich, K. (1987) *J. Mol. Biol. 196*, 611–639.
Williamson, M. P., Havel, T. F., & Wüthrich, K. (1985) *J. Mol. Biol. 182*, 295–315.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Zuiderweg, E. R. P., Kaptein, R., & Wüthrich, K. (1983a) *Eur. J. Biochem. 137*, 279–292.
Zuiderweg, E. R. P., Kaptein, R., & Wüthrich, K. (1983b) *Proc. Natl. Acad. Sci. U.S.A. 80*, 5837–5841.
Zuiderweg, E. R. P., Boelens, R., & Kaptein, R. (1985a) *Biopolymers 24*, 601–611.
Zuiderweg, E. R. P., Scheek, R. M., & Kaptein, R. (1985b) *Biopolymers 24*, 2257–2277.
Zuiderweg, E. R. P., Scheek, R. M., Boelens, R., van Gunsteren, W. F., & Kaptein, R. (1985c) *Biochimie 67*, 707–715.

Articles

Identification and Properties of an Oxoferryl Structure in Myeloperoxidase Compound II†

W. Anthony Oertling,‡ Hans Hoogland,§ Gerald T. Babcock,‡ and Ron Wever*,§

Department of Chemistry and MSU Shared Laser Laboratory, Michigan State University, East Lansing, Michigan 48824, and Laboratory of Biochemistry, University of Amsterdam, P.O. Box 20151, 1000 HD, Amsterdam, The Netherlands

Received February 9, 1988

ABSTRACT: Myeloperoxidase compound II has been characterized by using optical absorption and resonance Raman spectroscopies. Compared to compounds II in other peroxidases, the electronic and vibrational properties of this intermediate are strongly perturbed due to the unusual active-site iron chromophore that occurs in myeloperoxidase. Despite this difference in prosthetic group, however, other properties of myeloperoxidase compound II are similar to those observed for this intermediate in the more common peroxidases (horseradish peroxidase in particular). Two forms of the myeloperoxidase intermediate species, each with distinct absorption spectra, are recognized as a function of pH. We present evidence consistent with interconversion of these two forms via a heme-linked ionization of a distal amino acid residue with a $pK_a \simeq 9$. From resonance Raman studies of isotopically labeled species at pH 10.7, we identify an iron–oxygen stretching frequency at 782 cm^{-1}, indicating the presence of an oxoferryl ($O{=}Fe^{IV}$) group in myeloperoxidase compound II. We further conclude that the oxo ligand is not hydrogen bonded above the pK_a but possibly exhibits oxygen exchange with the medium at pH values below the pK_a due to hydrogen bonding of the oxo ligand to the distal protein group.

Peroxidases are utilized by the immune response systems of mammals (Klebanoff & Clark, 1978). Most peroxidase enzymes contain protoheme IX; however, the prosthetic group of myeloperoxidase (from granulocytes) is uncertain. Optical spectra of the pyridine hemochrome of myeloperoxidase suggest a heme *a* structure (Schultz & Shmukler, 1964; Newton et al., 1965), but resonance Raman measurements do not detect the expected formyl substituent (Babcock et al., 1985). Both Raman and magnetic circular dichroism studies were taken to suggest the presence of a saturated pyrrole ring, thus indicating an iron chlorin chromophore (Eglington et al., 1982; Sibbett & Hurst, 1984; Babcock et al., 1985; Ikeda-Saito et al., 1985; Stump et al., 1987). Still neither of these suggestions is sufficient to explain the distinctly red-shifted absorption spectra of several forms of the enzyme (Wever & Plat, 1981).

During turnover, myeloperoxidase (MPO) reacts with H_2O_2 to form a compound I intermediate which is capable of cat-

† This study was supported by grants from the Netherlands Organization for the Advancement of Pure Research (ZWO) under the auspices of the Netherlands Foundation for Chemical Research (SON) to R.W. and from the National Institutes of Health (GM 25480) to G.T.B. R.W. acknowledges receipt of a travel grant from the Netherlands Organization for the Advancement of Pure Research. This collaboration was made possible by Nato Research Grant 86/734.
‡ Michigan State University.
§ University of Amsterdam.

0006-2960/88/0427-5395$01.50/0