DETERMINATION OF THREE-DIMENSIONAL STRUCTURES OF PROTEINS AND NUCLEIC ACIDS IN SOLUTION BY NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY

Authors: **G. Marius Clore**
Angela M. Gronenborn
Laboratory of Chemical Physics
National Institute of Diabetes and Digestive and Kidney Diseases
National Institutes of Health
Bethesda, Maryland

Referee: Stanley J. Opella
Department of Chemistry
University of Pennsylvania
Philadelphia, Pennsylvania

I. INTRODUCTION*

Over the last decade, modern molecular biology, and in particular recombinant DNA technology, has had a major impact on a large number of biological disciplines. If one considers the area of protein biochemistry, for example, the number of protein sequences that has become available has reached over 10,000, the majority being derived from the nucleotide sequences of their respective genes. While this wealth of data are clearly impressive, amino acid sequences *per se* are of limited value in understanding protein function. In this respect, it is essential to determine the three-dimensional (3D) structures of proteins, and use this structural information in conjunction with functional data to study catalysis, ligand binding, assembly, and other processes in which proteins play a key role.

Until recently, the only experimental technique available for determining 3D protein structures has been single crystal X-ray diffraction. One of the rate-limiting factors in solving an X-ray structure lies not only in obtaining suitable crystals that diffract to sufficient resolution, but also appropriate heavy atom derivatives to determine the phases of the reflections accurately. Consequently, it is not suprising that the number of protein X-ray structures solved to date (a few hundred) is several orders of magnitude smaller than the number of available proteins sequences.

Over the last few years a second method of determining protein structures has been developed that makes use of nuclear magnetic resonance (NMR) spectroscopy. Unlike crystallography, NMR measurements are carried out in solution under potentially physiological conditions and therefore are not hampered by the ability or the inability of a protein to crystallize.

Although the field of macromolecular structure determination by NMR is still young, a reasonable number of NMR protein structures have already been published (Table 1). With more and more proteins being produced by recombinant DNA methods, there is little doubt that this number will rapidly increase. At the present time, however, the application of NMR to protein structure determination is limited to proteins up to $M_r \sim 20{,}000$, a restriction that may not be as severe as it might appear at first glance. It is known, for example, from protein structures solved by X-ray crystallography that the size of protein domains generally lies between 10 and 20 kDa. Such domains are therefore clearly amenable to the NMR approach. Thus, it is likely that the expression and production of protein domains by recombinant DNA technology will make larger proteins accessible to NMR structure determination by studying their functional domains one by one.

* A list of abbreviations used throughout this article appears following the text.

TABLE 1
Summary of Proteins and Polypeptides Whose Three-Dimensional Structures Have Been Determined by NMR

Protein[a] (residues)	Method[b]	Ref.
Micelle bound glucagon (29)	DG	2
Insectotoxin I_5A (35)	DG	3
BUSI IIa (57)	DG	4
Lac repressor headpiece (51)	Model building and RD	5
Helix F of CRP (17)	RD	6
α1-Purothionin (45)	DG and RD	7
α-Amylase inhibitor Tendamistat (74)	RLST	8
Rabbit metallothionein-2 (61)	RLST	9, 10
Rat metallothionein-2 (62)	RLST	11
Phoratoxin (46)	DG and RD	12
Hirudin (65)	DG and RD	13
	Hybrid DG-SA	14
GH5 (79)	DG and RD	15
	Hybrid DG-SA	16
hEGF (50)	DG	17
	RLST	18
mEGF (53)	RLST	19
BPTI (56)	DG and RLST	20
CPI (39)	DG and RD	21
	SA	22, 23
BSPI-2 (64)	DG and RD	24
	SA	22
hGHRF (29)	RD	25
Secretin (27)	RD	26
ACP-SH (77) fragments	EM	27
ACP-SH (77)	Hybrid DG-SA	28
	DG and EM	29
Cecropin (37)	SA	30
δ-Hemolysin	RLST and RD	31
BDS-I (43)	Hybrid DG-SA	32
C5A (74)	Hybrid DG-SA	33
	RLST	34
Mellitin (26)	RLST and RD	35
Cardiotoxin CTXIIb (60)	RLST	36
Defensin NP-5 (33)	DG	37
Plastocyanin (99)	DG and RD	38
Rat atrial natriuretic factor 7-23	DG and RD	39
Model calculations[c]		
BPTI (56)	DG	40
BPTI (56)	RLST	41—43
Crambin (46)	RD	44, 45
	DG	46
	Hybrid DG-SA	16
	SA from extended strand	22
	SA from random array of atoms	23

[a] The abbreviations for the various proteins are as follows: BUSI IIA, proteinase inhibitor IIA from bull seminal plasma; CRP, cAMP receptor protein of *Escherichia coli*; hGHRF, human growth hormone releasing factor; GH5, globular domain of histone H5; BPTI, basic pancreatic trypsin inhibitor; CPI, potato carboxypeptidase inhibitor; hEGF, human epidermal growth factor; mEGF, murine epidermal growth factor; BSPI-2, barley serine proteinase inhibitor 2; ACP-SH, acyl carrier protein (free sulfhydryl species).

[b] The abbreviations for the various methods are DG, metric matrix distance geometry; RD, restrained molecular dynamics; RLST, restrained least squares minimzation in torsion space with a variable target function (this, in general, refers to the DISMAN program of Braun and Go[41]); SA, dynamical simulated annealing; EM, energy minimization.

[c] In the model calculations, a set of interproton distance restraints that could realistically be determined experimentally were derived from the X-ray structures.

TABLE 2
Flow Chart of the Steps Involved in Determining the Three-Dimensional Structure of a Protein in Solution by NMR

Sequential resonance assignment
- Identification of spin systems by means of experiments demonstrating through-bond connectivities (COSY, MQF-COSY, HOHAHA; PS-COSY; relayed HMQC-COSY; relayed HMQC-HOHAHA)
- Identification of neighboring amino acids by means of NOE measurements demonstrating through-space (≤5 Å) short-range ($|i-j| \leq 5$) interproton connectivities (2D-NOESY, 2D relayed HMQC-NOESY; 3D HOHAHA-NOESY; 3D NOESY-NOESY; 3D HMQC-NOESY)

Assignment of tertiary long-range ($|i-j| > 5$) NOEs (2D-NOESY; 2D isotope edited NOESY; 2D relayed HMQC-NOESY; 3D HOHAHA-NOESY; 3D NOESY-NOESY; 3D HMQC-NOESY)

Quantification or classification of NOEs to yield approximate distance restraints

Measurement of three-bond coupling constants where feasible using experiments such as DQF-COSY, E-COSY, PE-COSY, and z-COSY, in order to derive approximate torsion angle restraints

Identification of regular secondary structure elements by means of a qualitative interpretation of the NOE and backbone amide exchange data

Determination of the three-dimensional structure on the basis of the approximate interproton distance and torsion angle restraints using a combination of the following approaches:
- Manual and semi-automatic model building methods
- Methods relying on data bases derived from X-ray structures of proteins
- Metric matrix distance geometry
- Restrained least squares minimization in torsion angle space with a variable target function or a series of ellipsoids
- Restrained molecular dynamics
- Dynamical simulated annealing
- Hybrid metric matrix distance geometry-dynamical simulated annealing

The principal source of information used to solve 3D structures of macromolecules by NMR spectroscopy resides in short (<5 Å) interproton distance data. To obtain these distances, the ^{1}H-NMR spectrum must first be assigned using a variety of experiments, in particular correlation and nuclear Overhauser effect (NOE) experiments to demonstrate through-bond and through-space connectivities, respectively. Because the NOE is proportional to r^{-6}, distance information can then be derived. Although conceptually simple, it is only relatively recently that the goal of solving 3D structures of macromolecules has been realized. This was due to the introduction of three key developments. The first was the extension of the traditional one-dimensional (1D) NMR experiments into a second dimension, thereby enabling one to detect and interpret effects that would not have been possible previously owing to extensive spectral overlap and much reduced information (see Ernst et al.[1] for a detailed review of 2D-NMR). The second was the introduction of high-field (500 MHz or higher) spectrometers, resulting in an increase both in spectral resolution and signal-to-noise. The third was the development of suitable mathematical algorithms and computational approaches for determining 3D structures on the basis of interproton distance data.

In this article, the various stages involved in the determination of 3D structures of macromolecules by NMR are reviewed as summarized by the flowchart in Table 2.

II. THE BASIS OF TWO-DIMENSIONAL (2D) NMR

Each proton posseses a property known as magnetization. When a sample is placed in a magnetic field B_o, the magnetization lies parallel to B_o (i.e., along the z-axis of the rotating frame). To record a conventional 1D NMR spectrum, a radiofrequency pulse B_1 is applied that

rotates the magnetization away from the z-axis toward the x-y plane. A free induction decay is then recorded immediately after the pulse. Fourier transformation of the free induction decay yields the conventional 1D spectrum. To obtain additional information on interactions between spins, double or multiple irradiation experiments must be carried out. This necessitates the use of a second selective radiofrequency pulse B_2 at the position of a particular resonance. As long as there is little or no resonance overlap, the application of such experiments is relatively simple. In the case of macromolecules, and proteins in particular, there is extensive spectral overlap rendering this approach impractical in the majority of cases.

The limitations of 1D NMR can be overcome by extending the measurements into a second dimension. All 2D-NMR experiments comprise the same basic scheme. This consists of a preparation period, an evolution period (t_1) during which the spins are labeled according to their chemical shift, a mixing period during which the spins are correlated with each other, and, finally, a detection period (t_2). A number of experiments are recorded with successively incremented values of the evolution period t_1 to generate a data matrix $s(t_1,t_2)$. Two-dimensional Fourier transformation of $s(t_1,t_2)$ then yields the desired 2D frequency spectrum $S(\omega_1,\omega_2)$. This is summarized schematically in Figure 1. In most homonuclear 2D experiments, the diagonal corresponds to the 1D spectrum, and the symmetrically placed cross-peaks on either side of the diagonal indicate the existence of an interaction between two spins. The nature of the interaction depends on the type of experiment. Thus, in a homonuclear correlated spectroscopy (COSY) experiment, the cross-peaks arise from through-bond scalar correlations, while in a NOESY experiment they arise from through-space correlations. For optimal sensitivity, the 2D experiment should be acquired with amplitude modulation as a function of t_1[47,48] as opposed to phase modulation. The latter cannot be phased and hence is displayed in an absolute value mode. The former, on the other hand, results in a pure phase absorption spectrum that is of much higher quality, both in terms of sensitivity and resolution, than the corresponding absolute value spectrum.

III. THE NOE EFFECT

The main source of geometrical information employed in determining 3D structures of macromolecules by NMR spectroscopy is derived from nuclear Overhauser enhancement (NOE) measurements that enable one to demonstrate the proximity of protons in space and to deduce their approximate separation.[49-52] The principle of the NOE is relatively simple and is summarized in Figure 2. Considering the simplest system with only two protons, we note that each proton possesses a property known as magnetization. Magnetization is exchanged between the protons by a process known as cross-relaxation. Because the cross-relaxation rates in both directions are equal, the magnetization of the two protons at equilibrium is equal. The approximate chemical analogy of such a system would be one with two interconverting species with an equilibrium constant of 1. The cross-relaxation rate is proportional to two variables: r^{-6}, where r is the distance between the two protons, and τ_{eff}, the effective correlation time of the interproton vector. It follows that if the magnetization of one of the spins is perturbed, the magnetization of the second spin will change. In the case of macromolecules, the cross-relaxation rates are positive and the leakage rate from the system is very small, so that, in the limit, the magnetization of the two protons would be equalized. The change in magnetization of proton i upon perturbation of the magnetization of proton j is known as the nuclear Overhauser effect (NOE). The initial build-up rate of the NOE is equal to the cross-relaxation rate and hence proportional to r^{-6}.

In 1D NMR, the NOE can be observed in a number of ways, all of which involve the application of a selective radiofrequency pulse at the position of one of the resonances. The simplest experiment involves the irradiation of resonance i for a time t, followed by acquisition of the spectrum. If proton j is close in space to proton i, its magnetization will be reduced. This

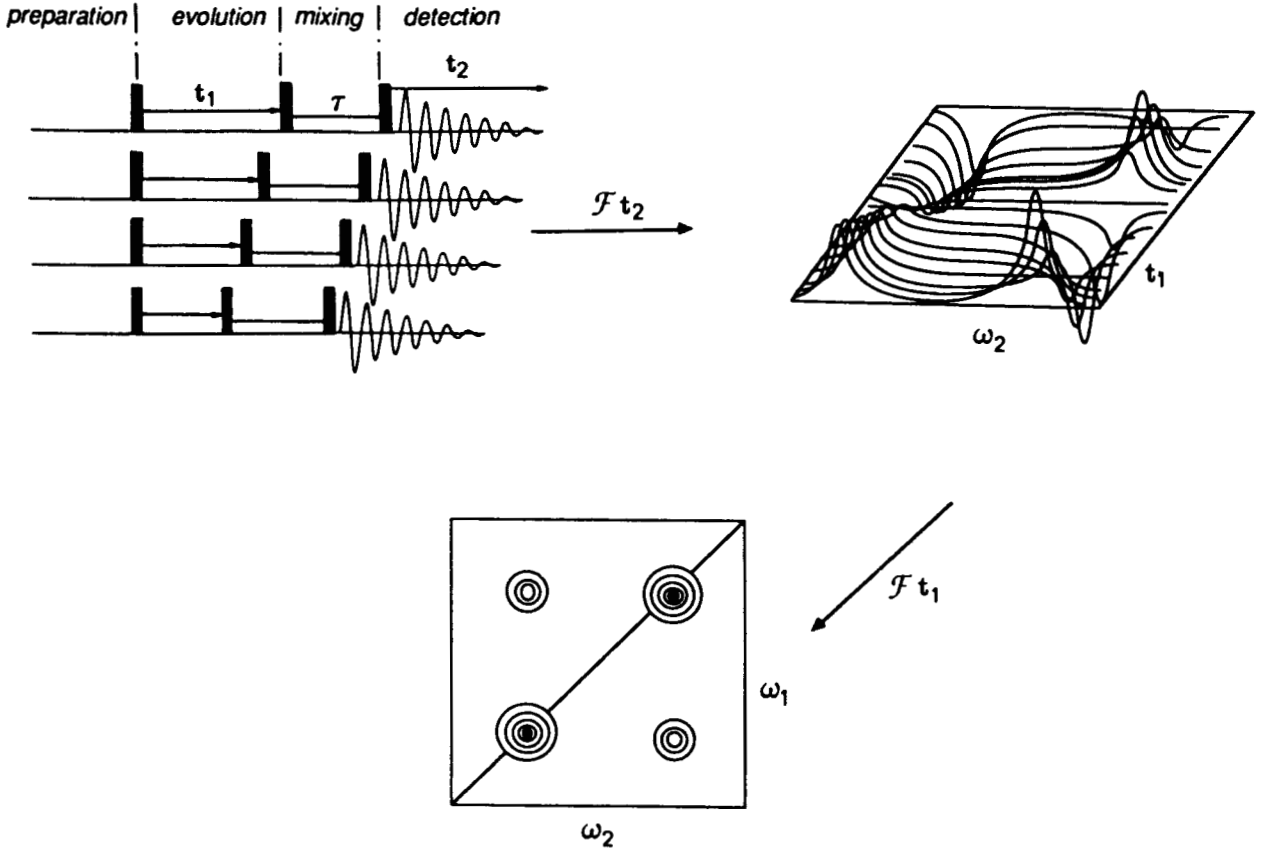
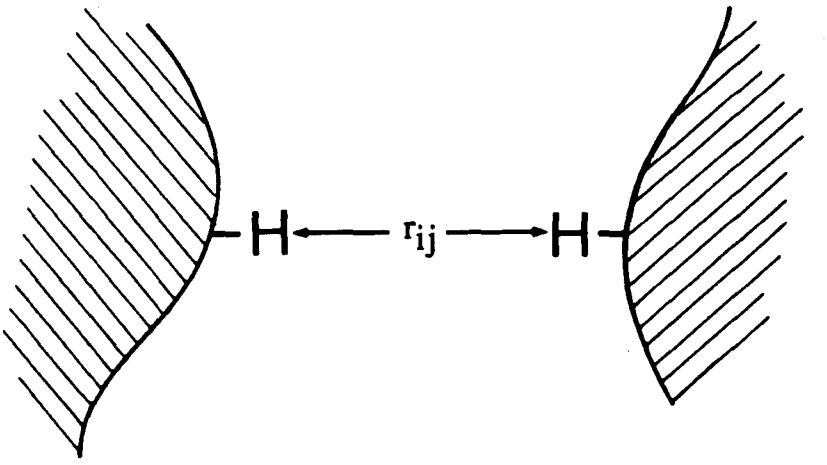


FIGURE 1. Generalized representation of the 2D-NMR experiment.



$$N_{ij}(t) \sim \sigma_{ij}\, t$$

$$\sigma_{ij} = \frac{\gamma^4 \hbar^2}{10\, r_{ij}^6} \left(\tau_{app} - \frac{6\tau_{app}}{1 + 4\omega^2 \tau_{app}^2} \right)$$

FIGURE 2. The basis of the NOE. (Symbols: r_{ij}, distance between the protons i and j; σ, cross-relaxation rate; N, NOE, τ_{app}, correlation time.)

is best observed in a difference spectrum, subtracting a spectrum without irradiation from one with selective irradiation. An alternative approach involves the selective inversion of resonance i followed by acquisition after a time t. This particular experiment is the 1D analog of the 2D experiment. In the 2D experiment,[53-55] cross-peaks between proton resonances i and j are observed when the two protons are close in space, and thus exchange magnetization via cross-relaxation.

IV. SEQUENTIAL RESONANCE ASSIGNMENT

Sequential resonance assignment of the ^{1}H-NMR spectra of protein and nucleic acids relies on two sorts of experiments: (1) those demonstrating through-bond scalar connectivities, and (2) those demonstrating through-space (<5 Å) connectivities.[56] The former, which are generally referred to as correlation experiments, serve to group together protons belonging to the same residue. The latter involve the detection of NOEs and serve to connect one residue with its immediate neighbors in the linear sequence of amino acids or nucleotides. Pulse schemes for a number of commonly used proton experiments are shown in Figures 3 and 4. Those that the authors have found most useful in terms of resonance assignment are discussed below.

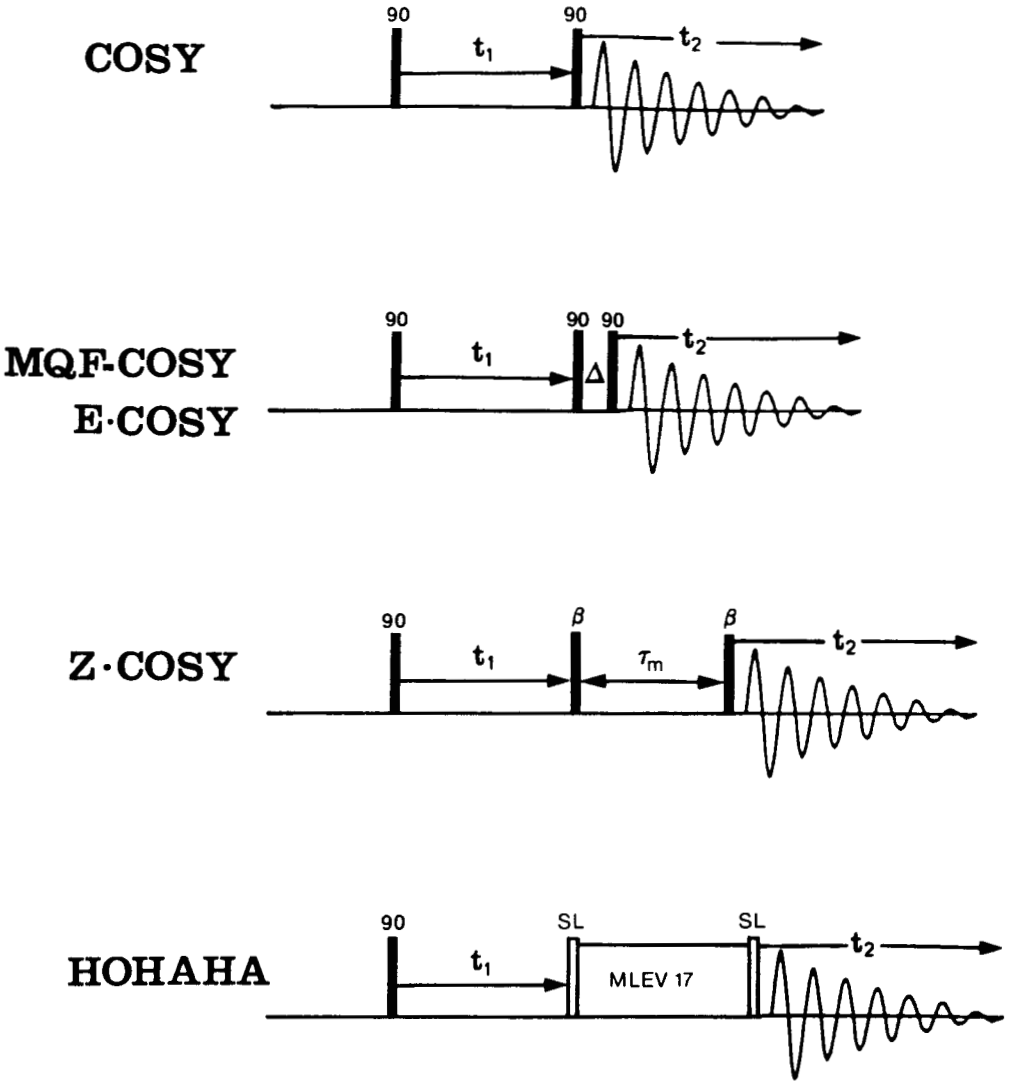


FIGURE 3. Pulse schemes for some commonly used homonuclear correlation experiments which demonstrate through-bond connectivities. (Symbols: τ_m, mixing time; SL, spin lock.)

A. SEQUENTIAL RESONANCE ASSIGNMENT OF PROTEIN NMR SPECTRA

The first step of the assignment procedure lies in identifying spin systems, that is to say protons belonging to one residue unit in the polypeptide chain. Such experiments have to be carried out both in H_2O and D_2O, the former to establish connectivities involving the exchangeable NH protons, and the latter to identify connectivities between nonexchangeable protons. Some spin systems are characteristic of single amino acids. This is the case for Gly, Ala, Thr, Leu, Ile, and Lys. Others are characteristic of several different amino acids; for example, Asp, Asn, Cys, Ser, and the aliphatic protons of all aromatic amino acids belong to the AMX spin system (i.e., they all have one $C^{\alpha}H$ and two $C^{\beta}H$ protons).

The simplest experiment used to delineate spin systems via scalar correlations is the COSY experiment, which was first described in 1976 by Aue et al.[57] and demonstrates direct through-bond connectivities. Thus, for a residue that has NH, $C^{\alpha}H$, $C^{\beta}H$, and $C^{\gamma}H$ protons, connectivities will only be manifested between the NH and $C^{\alpha}H$, $C^{\alpha}H$ and $C^{\beta}H$, and $C^{\beta}H$ and $C^{\gamma}H$ protons. This

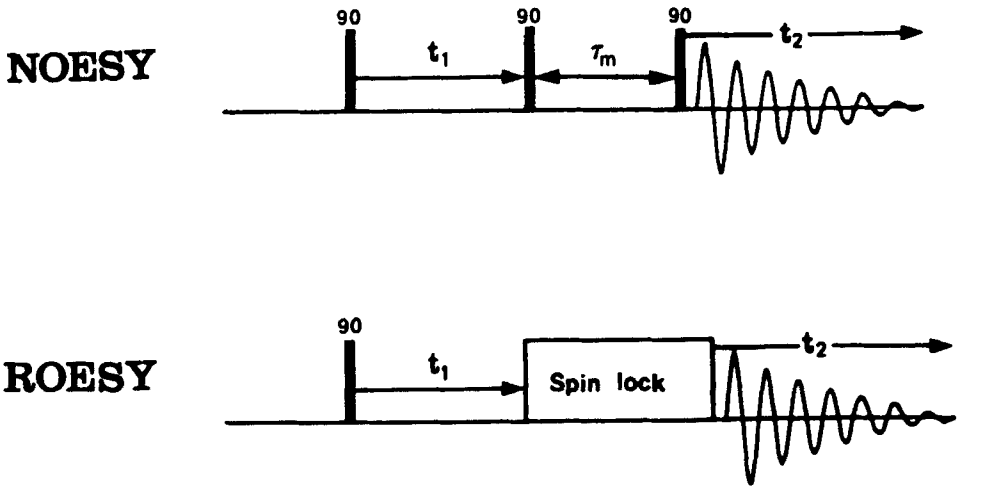


FIGURE 4. Pulse schemes for the NOESY and ROESY experiments which demonstrate through-space (<5 Å) connectivities.

basic COSY experiment has now been superseded by slightly more sophisticated experiments such as double-quantum filtered COSY (DQF-COSY)[58] and P.COSY[59] which have the advantage of exhibiting pure phase absorption diagonals when the spectra are recorded in the pure phase absorption mode. This enables one to detect cross-peaks close to the diagonal. As the sensitivity of the P.COSY experiment is approximately double that of the DQF-COSY experiment, it is found to be generally more useful to record the former rather than the latter.

Experiments that demonstrate only direct through-bond connectivities are of limited value if taken alone, owing to problems of spectral overlap. In a protein spectrum, the degree of spectral overlap tends to increase as one progresses from the NH and $C^{\alpha}H$ protons to the side chain protons. For this reason, experiments that also demonstrate indirect or relayed through-bond connectivities, such as between the NH and $C^{\beta}H$ protons, are invaluable. In this respect, the most useful experiment is the homonuclear Hartmann-Hahn (HOHAHA) experiment[61-64] (also referred to as TOCSY for Total Correlated Spectroscopy[65]). By adjusting the experimental mixing time, one can obtain successively direct, single, double, and multiple relayed connectivities. Further, the multiple components of the cross-peaks are all in-phase in HOHAHA spectra, in contrast to COSY-type spectra, where they are in anti-phase. As a result, the HOHAHA experiment is in general more sensitive and affords better resolution than the COSY-type experiment. A schematic representation of cross-peak patterns observed in HOHAHA spectra for the various spin systems is illustrated in Figure 5. Examples of protein HOHAHA spectra in H_2O and D_2O are shown in Figures 6 and 7, and an example of a DQF-COSY spectrum is shown in Figure 8.

Once a few spin systems have been identified, one can then proceed to identify sequential through-space connectivities involving the NH, $C^{\alpha}H$, and $C^{\beta}H$ protons by means of 2D NOE spectroscopy (NOESY). For the purpose of sequential assignment, the most important connectivities are the $C^{\alpha}H(i)$-NH(i+1,2,3,4), $C^{\beta}H(i)$-NH(i+1), NH(i)-NH(i+1), and $C^{\alpha}H(i)$-$C^{\beta}H(i+3)$ NOEs.[56,66-69] This is illustrated schematically in Figure 9, and examples of NOESY spectra are shown in Figures 10 and 11.

In the case of large molecules where the NH proton resonances are broad, the sensitivity of the conventional COSY spectrum can be improved by a factor of ~2 by recording a ^{15}N-filtered COSY spectrum[60] (also known as PS-COSY for pseudo single quantum COSY; see Figure 12). Labeling the protein uniformly with ^{15}N permits efficient generation of heteronuclear zero- and

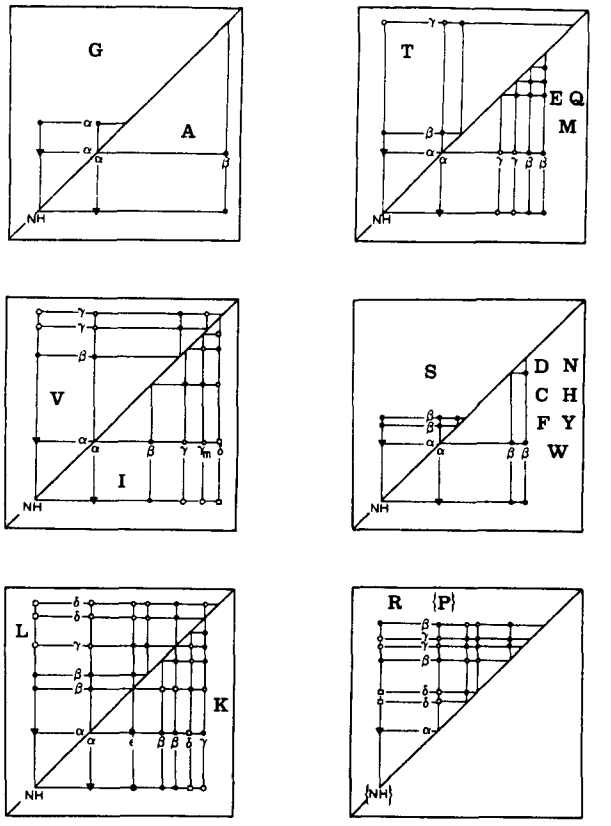


FIGURE 5. Schematic representation of cross-peak patterns observed in HOHAHA spectra for various spin systems. In a COSY spectrum, only direct connectivities are observed.

double-quantum coherences whose relaxation rates, to a first-order approximation, are not affected by heteronuclear dipolar coupling. This permits one to eliminate one of the major line broadening mechanisms for amide protons in proteins, namely, heteronuclear dipolar coupling to the nitrogen nucleus.[1] As a result, the multiple-quantum resonances are significantly narrower than the corresponding NH resonances. The ^{15}N chemical shift contribution is easily removed from the multiple-quantum frequency, yielding spectra that are similar in appearance to a regular COSY spectrum apart from the line narrowing of the NH resonances (Figure 13).

As proteins get larger, problems associated with chemical shift dispersion become increasingly severe. One approach for alleviating such problems in the sequential assignment of proteins involves correlating proton-proton through-space and through-bond connectivities with the chemical shift of a directly bonded NMR active nucleus. In the case of completely ^{15}N labeled protein, two sorts of experiments are particularly useful. The first are relayed experiments combining the heteronuclear multiple-quantum coherence (HMQC) scheme[70-75] with experiments such as NOESY, COSY, or HOHAHA (Figure 14).[76] These experiments yield the same information as the homonuclear NOESY, COSY, and HOHAHA experiments, but the NH proton chemical shift axis is replaced by that of the ^{15}N chemical shift (Figure 15). Because it is very rare to find that both the ^{1}H and ^{15}N chemical shifts of two NH groups are degenerate, NOEs and through-bond correlations involving NH protons with the same chemical shift can readily be resolved in this manner. The second type of experiment involves the detection of long-range correlations between ^{15}N and $C^{\alpha}H$ atoms using ^{1}H-detected heteronuclear multiple-bond

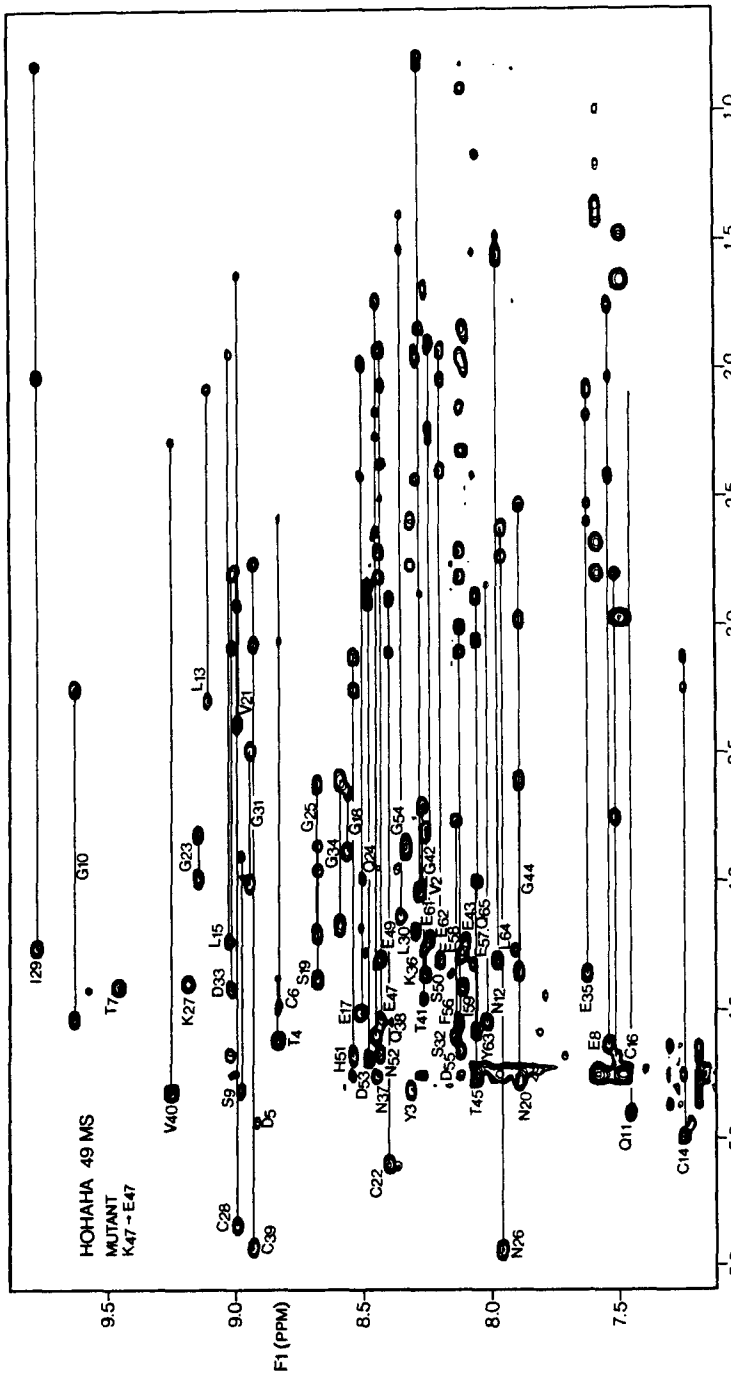


FIGURE 6. 600 MHz HOHAHA spectrum of Hirudin in H_2O showing the NH(F2)-aliphatic(F1) region of the spectrum. Peaks are labeled at the positions of the direct NH-C^{α}H connectivities. (From Folkers, P. J. M., Clore, G. M., Driscoll, P. C., Dodt, J., Köhler, S., and Gronenborn, A. M., *Biochemistry*, 28, 2178, 1989. With permission.)

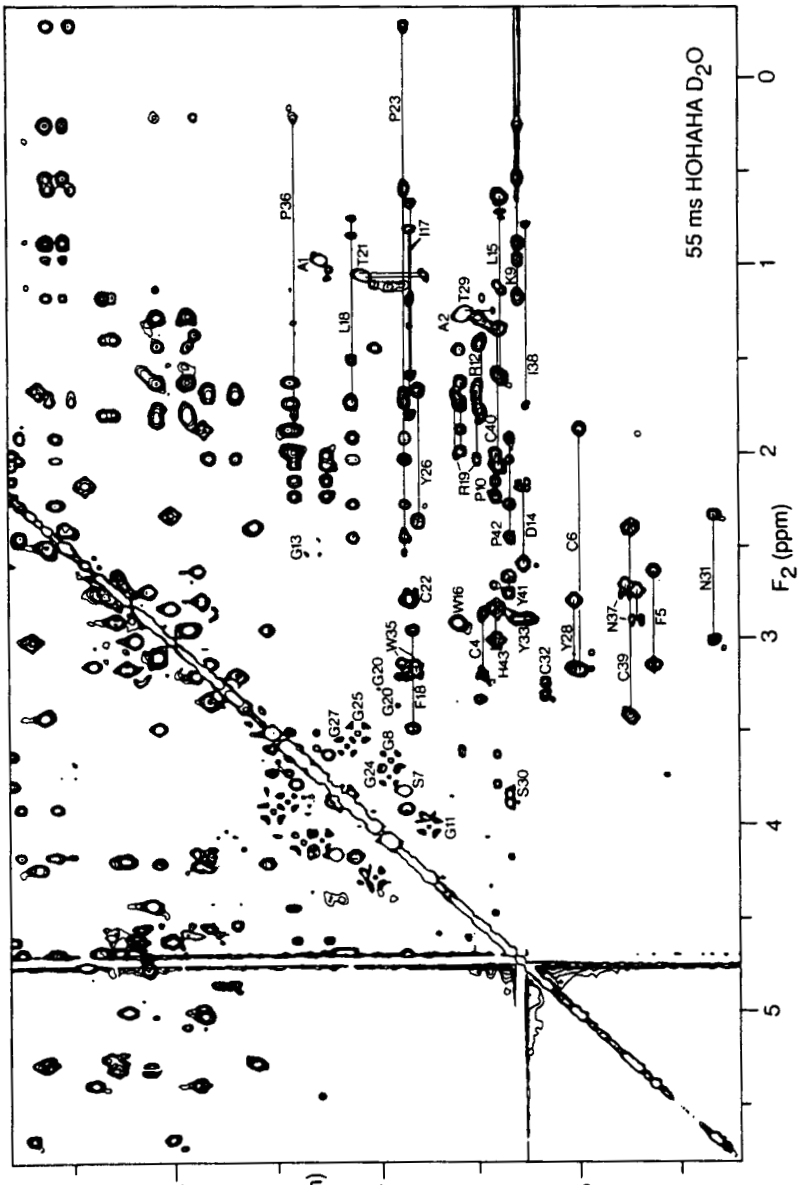


FIGURE 7. 600 MHz HOHAHA spectrum of BDS-1 in D_2O showing the $C^{\alpha}H(F1)$-aliphatic(F2) region of the spectrum. A number of spin systems are indicated. (From Driscoll, P. C., Clore, G. M., Beress, L., and Gronenborn, A. M., *Biochemistry*, 28, 2178, 1989. With permission.)

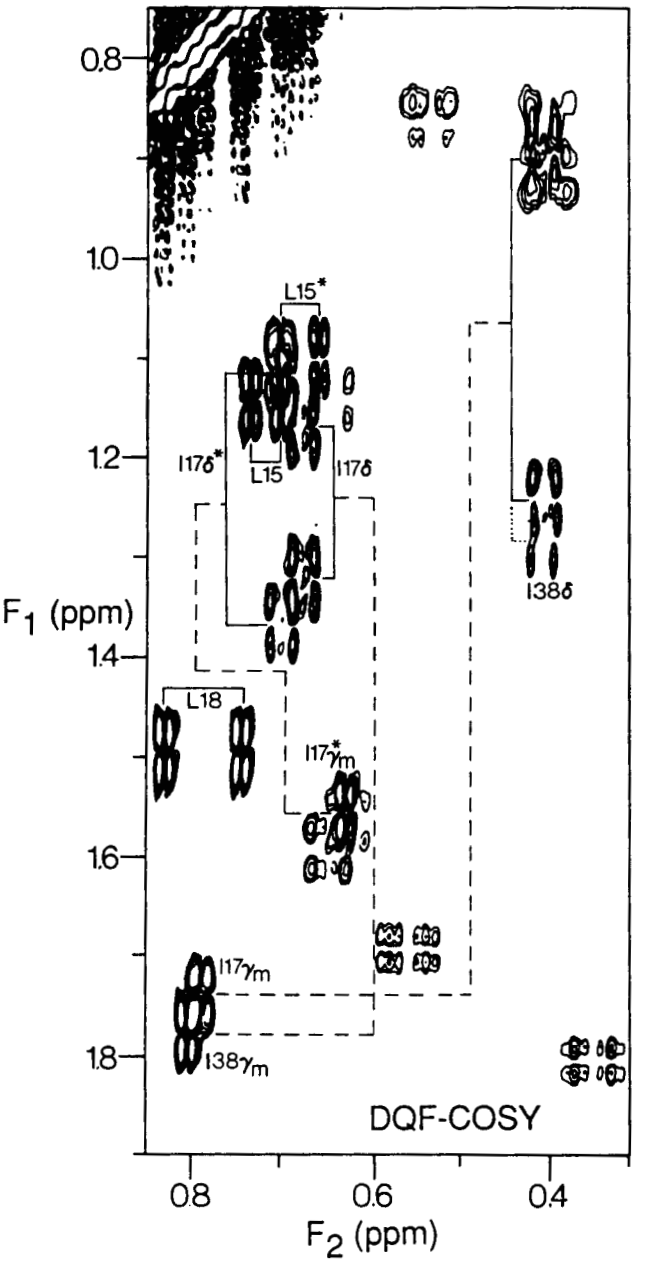


FIGURE 8. 600 MHz DQF-COSY spectrum of BDS-I in D_2O showing the methyl region of the spectrum. (From Driscoll, P. C., Clore, G. M., Beress, L., and Gronenborn, A. M., *Biochemistry*, 28, 2178, 1989. With permission.)

correlation (HMBC) spectroscopy (Figure 14).[77,78] In particular, the observation of two-bond $^{15}N(i)$-$C^{\alpha}H(i)$ and three-bond $^{15}N(i)$-$C^{\alpha}H(i-1)$ correlations enables one to connect one residue with the next.[79] Additionally, because the size of the $^{15}N(i)$-$C^{\alpha}H(i-1)$ coupling is very sensitive to the ψ backbone torsion angle, qualitative structural information can be readily derived. An example of this type of experiment is shown in Figure 16.

Yet another avenue for large proteins involves the application of 3D NMR. The extension of

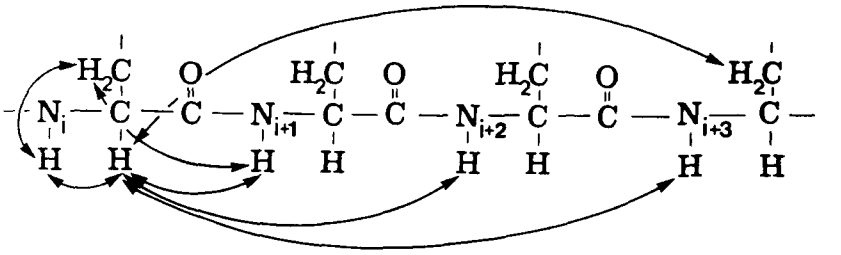


FIGURE 9. Schematic illustration of the connectivities used for sequential resonance assignment of protein spectra.

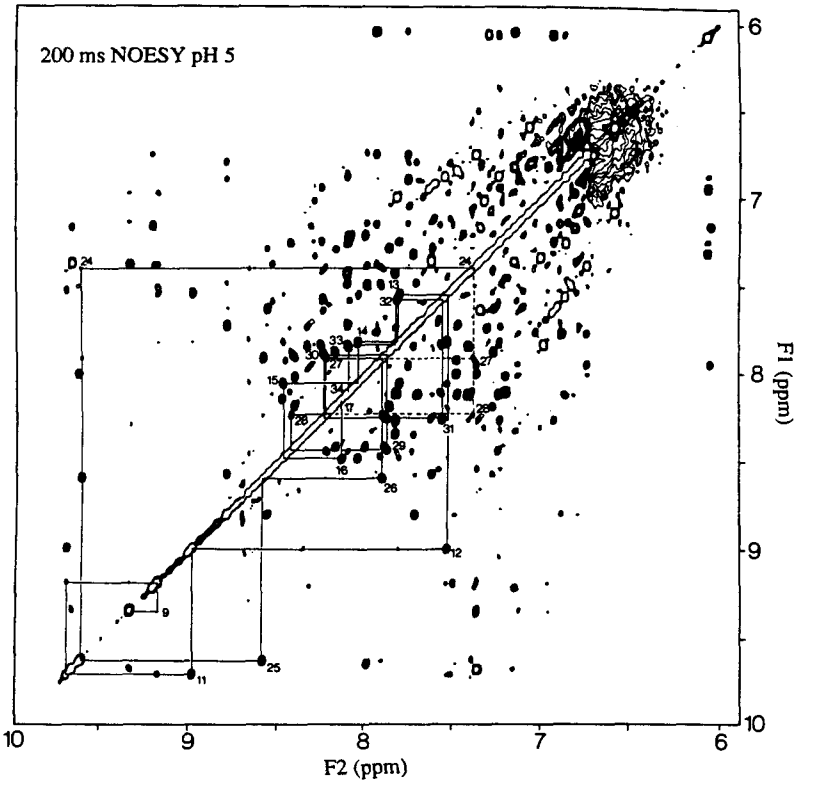


FIGURE 10. 600 MHz NOESY spectrum of the protein Ner in H_2O showing the NH(F1)-NH(F2) region of the spectrum. Two stretches of NH(i)-NH(i + 1) connectivities indicative of α-helical regions are indicated. (From Gronenborn, A. M., Wingfield, P. T., and Clore, G. M., *Biochemistry*, in press.. With permission.)

NMR experiments from two to three dimensions is relatively straightforward. Any 3D-NMR experiment can be conceived by simply combining two 2D-NMR experiments, leaving out the detection period of the first experiment and the preparation period of the second experiment (Figure 17).[80-83] The location of a peak in 3D space is described by the chemical shifts of the three spins, i, j, and k. The information contained within a peak depends on the nature of the experiment. For example, in a 3D experiment combining a correlation experiment with a NOE experiment, the peak at position F1(i), F2(j), and F3(k) will arise from through-bond scalar coupling between spins i and j, and through-space correlation between spins j and k. Thus, for

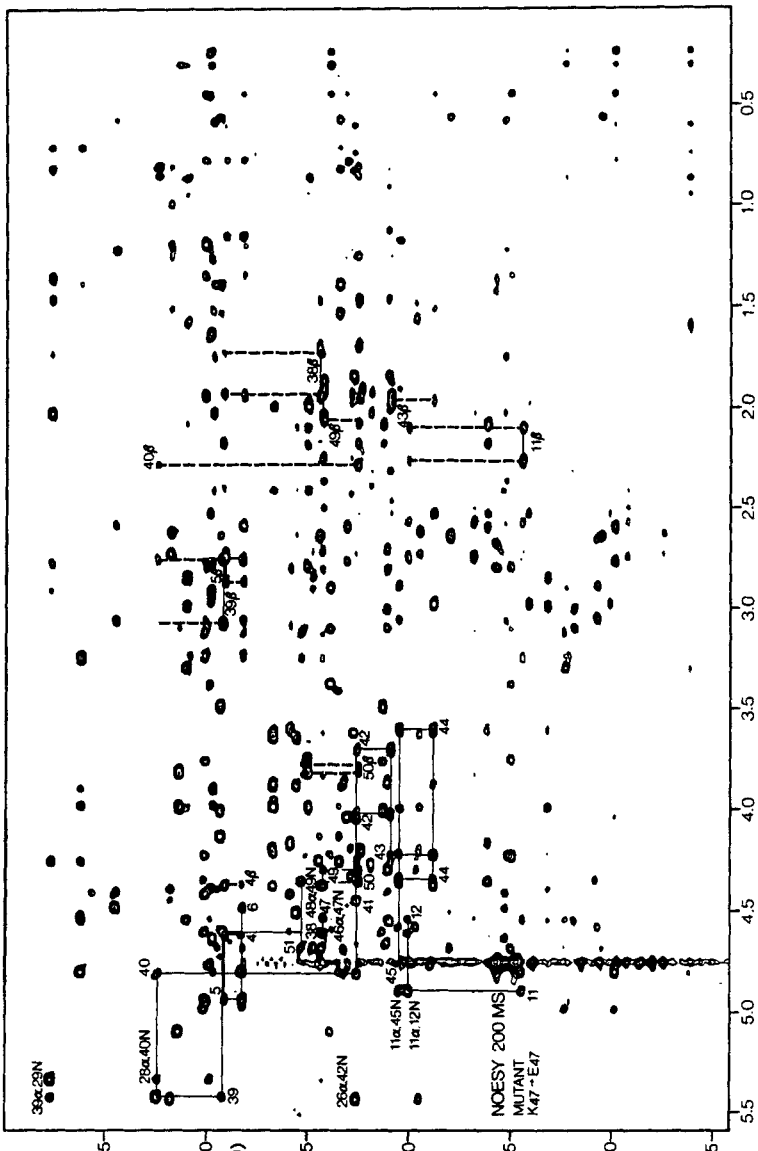


FIGURE 11. 600 MHz NOESY spectrum of Hirudin in H_2O showing the NH(F2)-aliphatic(F1) region of the spectrum. Some $C^{\alpha}H(i)$-NH(i + 1) as well as $C^{\beta}H(i)$-NH(i + 1) connectivities are indicated. (From Folkers, P. J. M., Clore, G. M., Driscoll, P. C., Dodt, J., Köhler, S., and Gronenborn, A. M., *Biochemistry*, 28, 2601, 1989. With permission.)

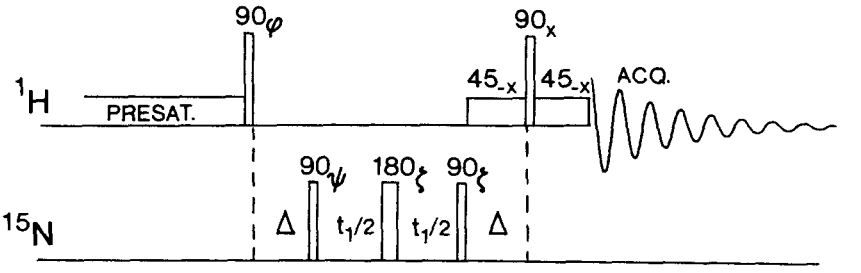


FIGURE 12. Pulse scheme for the ^{15}N-filtered pseudo single quantum (PS) COSY experiment that eliminates one of the major line-broadening mechanisms for amide protons in proteins, namely, heteronuclear dipolar coupling to the nitrogen nucleus. Effective application of this pulse scheme requires the use of proteins that are completely ^{15}N labeled.

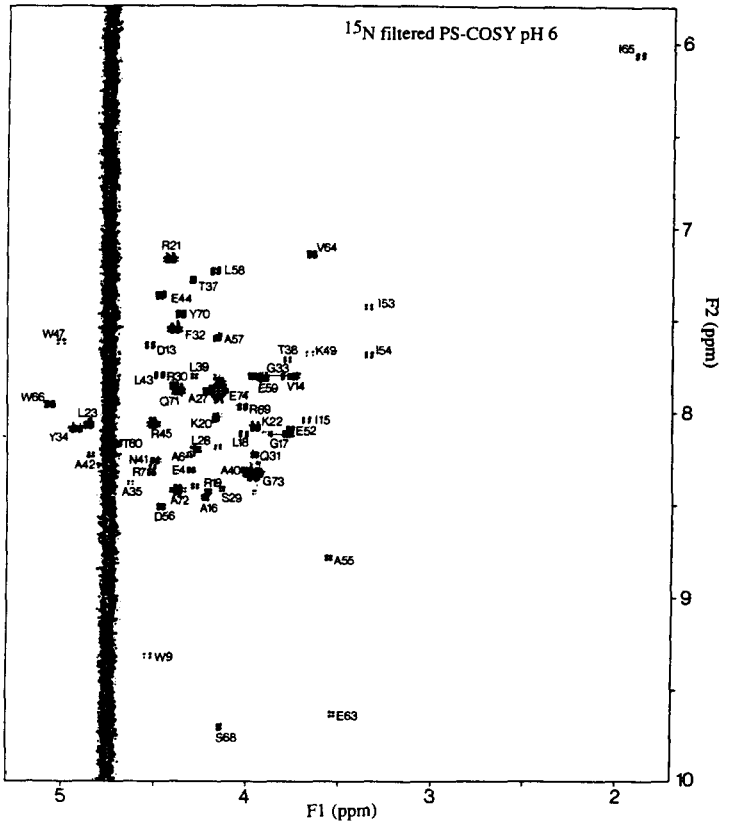


FIGURE 13. Example of a 600 MHz ^{15}N-filtered PS-COSY spectrum of the protein Ner from phage Mu in H_2O, showing the NH(F1)-C$^{\alpha}$H(F2) region of the spectrum. Because the pseudo single quantum has to be present during the t_1 period, the C$^{\alpha}$H resonances have to be detected during t_2, necessitating the use of presaturation to suppress the H_2O resonance. The ^{15}N enrichment is approximately 95%. (From Gronenborn, A. M., Wingfield, P. T., and Clore, G. M., *Biochemistry*, in press. With permission.)

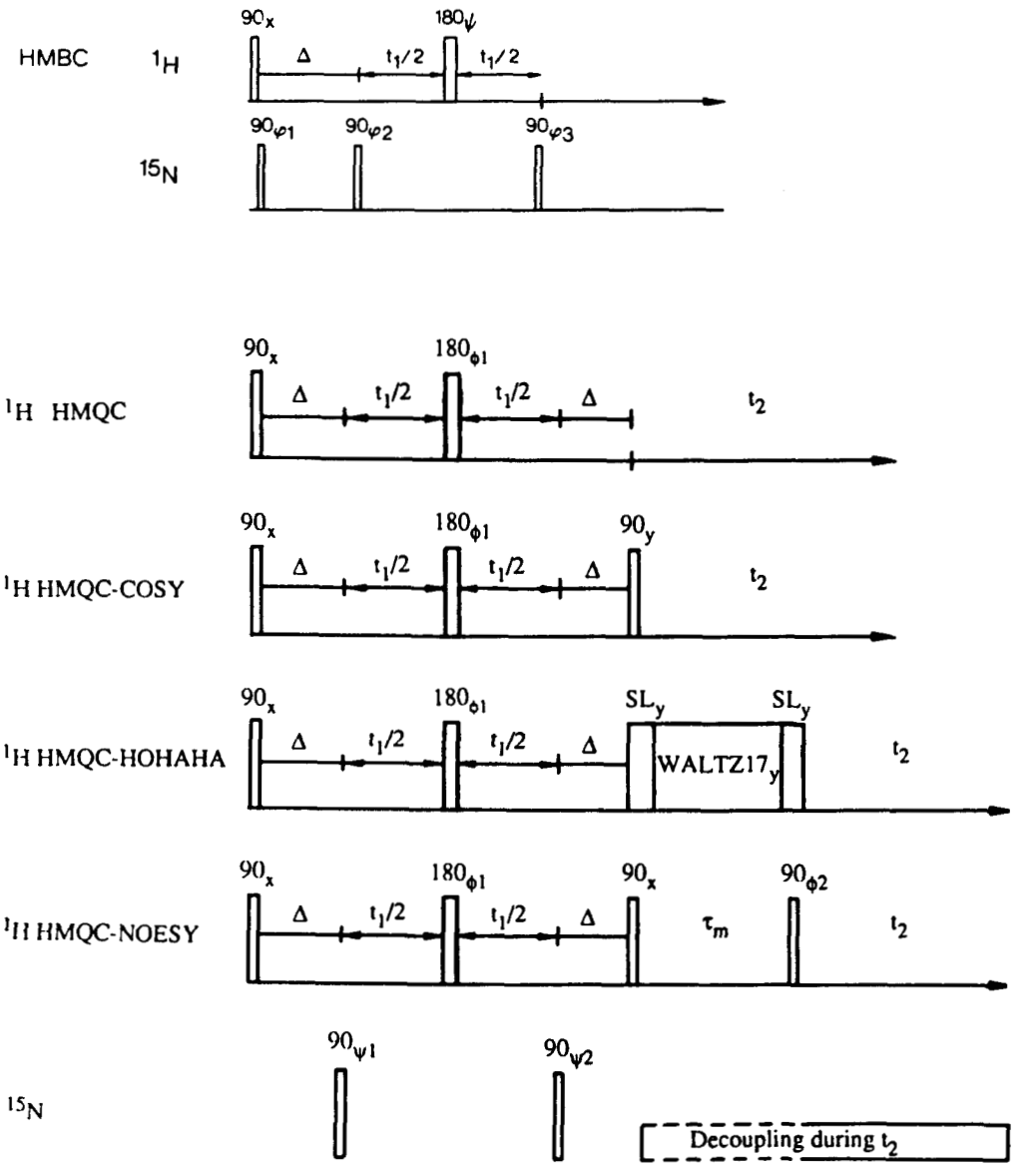


FIGURE 14. Pulse scheme for ^{1}H detected HMQC, relayed HMQC, and HMBC experiments.

example, in a homonuclear 3D-HOHAHA-NOESY experiment employing a selective pulse centered around the NH and $C^{\alpha}H/C^{\beta}H$ regions in the F1 and F2 dimensions, respectively, cross-peaks will arise from the pathway

$$NH \xrightarrow{J + \text{relayed } J} C^{\alpha}H/C^{\beta}H \xrightarrow{NOE} NH/\text{aliphatics}$$

The pulse scheme for such an experiment is depicted in Figure 18, and an example of a 3D HOHAHA-NOESY spectrum is shown in Figure 19. Probably the most useful 3D-NMR experiments, in terms of both sensitivity and information content, are the 3D extensions of the relayed heteronuclear HMQC experiments described above. In such a spectrum, each slice at a given ^{15}N chemical shift gives rises to a NOESY, COSY, or HOHAHA spectrum involving only those NH protons that are bonded to the ^{15}N atoms at that chemical shift. Thus, these 2D slices

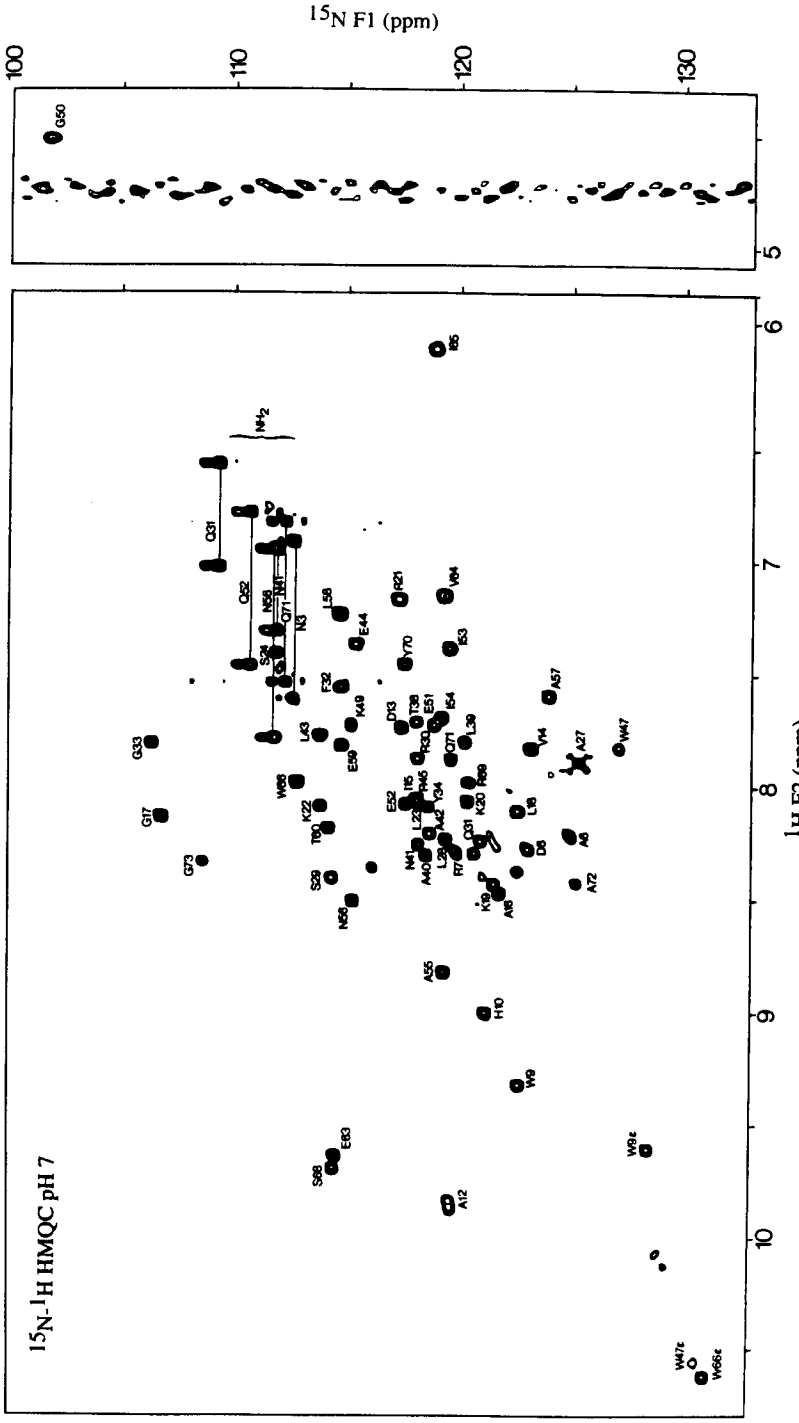


FIGURE 15. 600 MHz ^{15}N-^{1}H HMQC correlation (A) and ^{15}N-^{1}H HMQC-NOESY (B and C) spectra of uniformly labeled ^{15}N Ner protein in H_2O. The ^{15}N(F1)-NH(F2) and ^{15}N(F1)-aliphatic(F2) regions of the relayed ^{15}N-^{1}H HMQC-NOESY spectrum are shown in (B) and (C), respectively. Selected NH(i)-NH(i + 1), C^{α}H(i)-NH(i + 1), and C^{β}H(i)-NH(i + 1) NOE connectivities are shown in (B) and (C). (From Gronenborn, A. M., Wingfield, P. T., and Clore, G. M., *Biochemistry*, in press. With permission.)

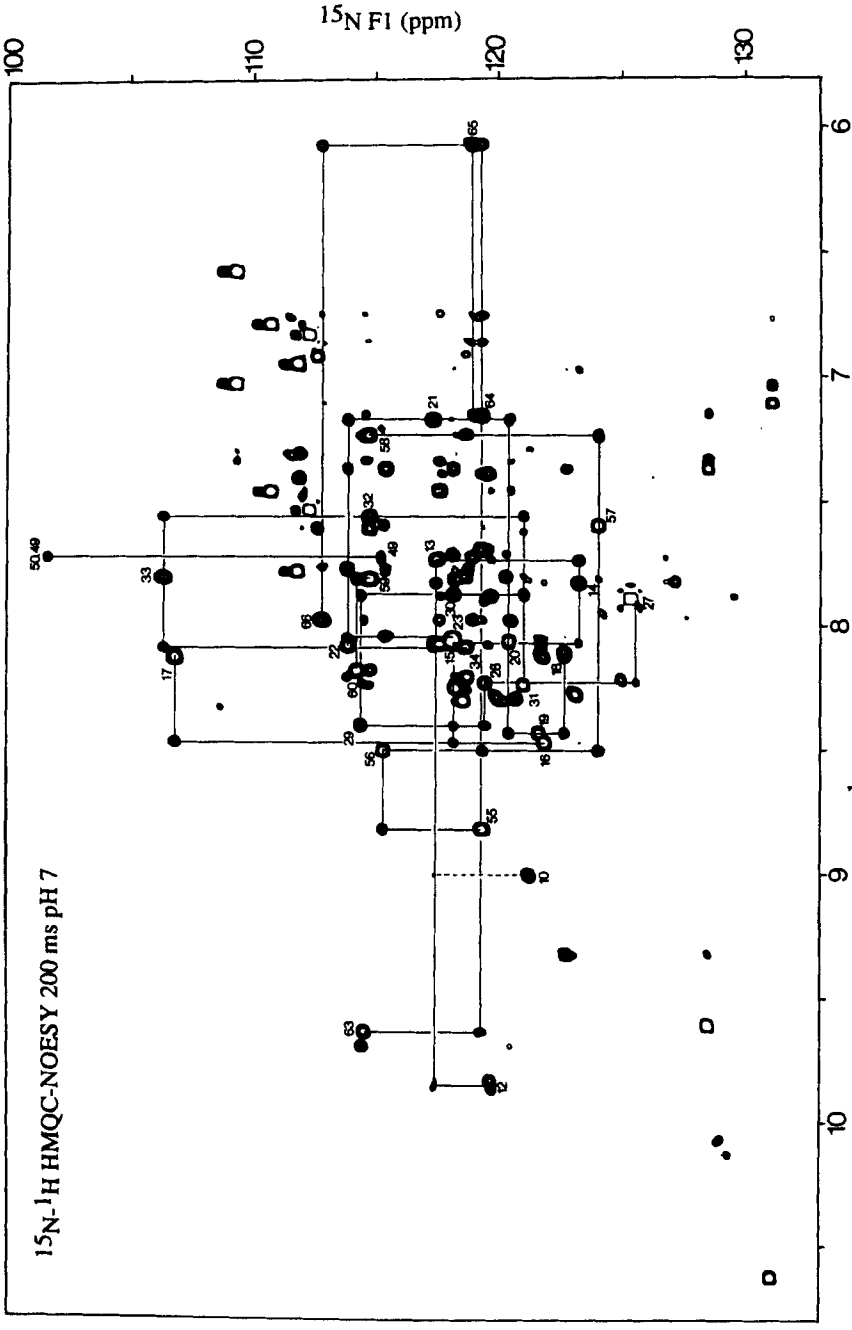


FIGURE 15B

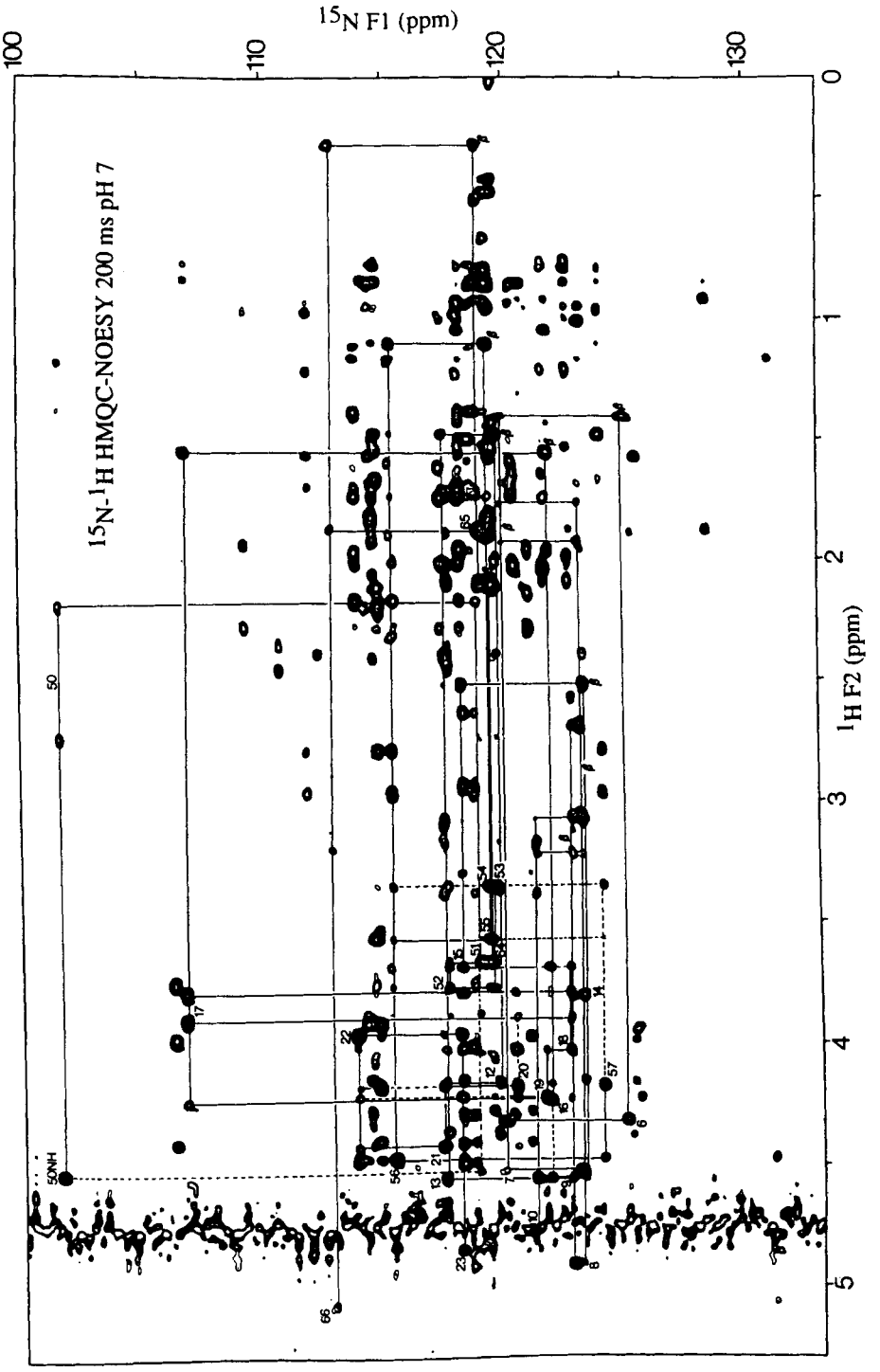


FIGURE 15C

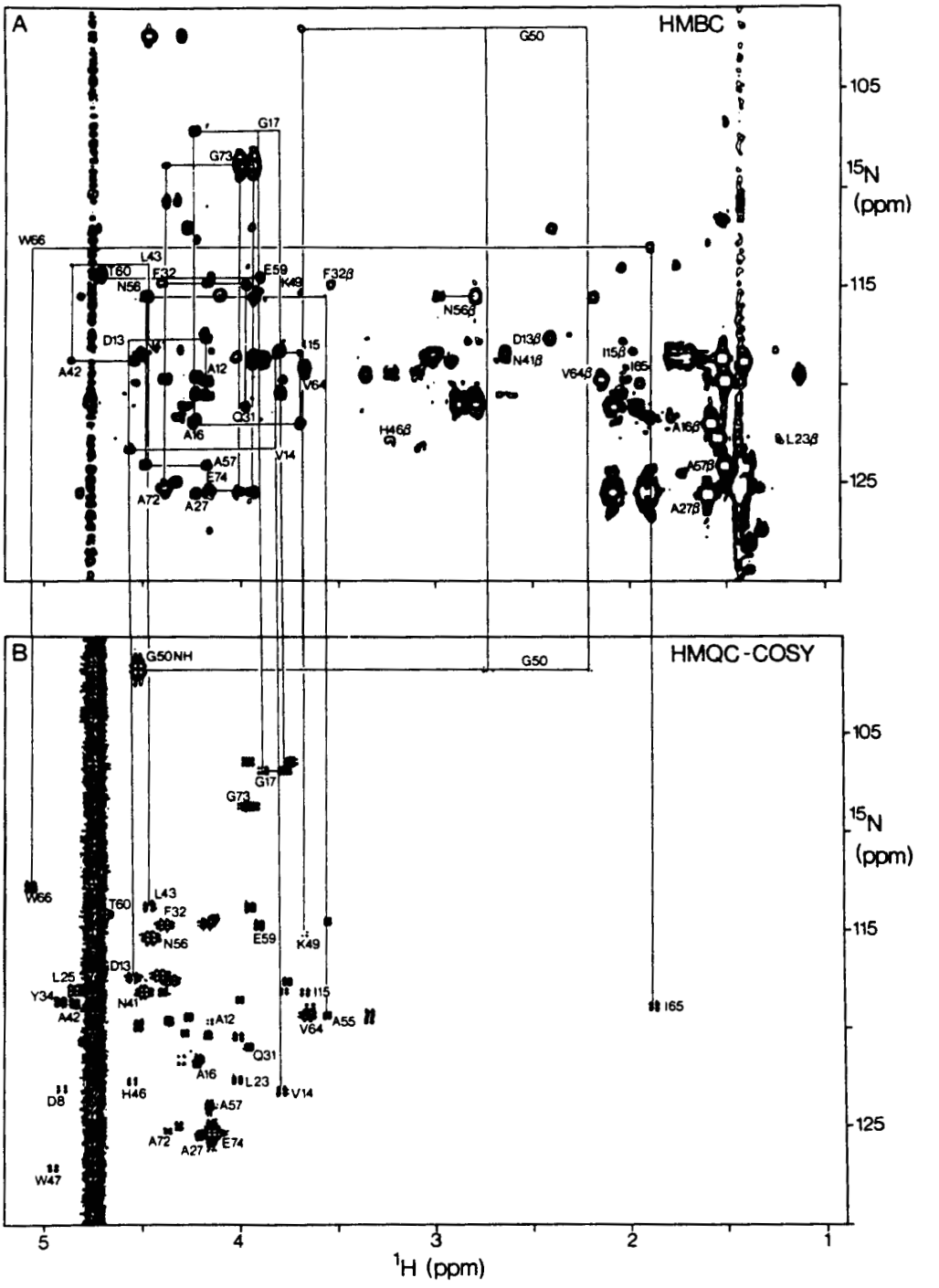


FIGURE 16. 600 MHz ^{15}N-^{1}H HMBC spectrum in D_2O (A) and ^{15}N-^{1}H HMQC spectrum in H_2O (B) of 2 m*M* uniformly ^{15}N labeled Ner protein. Labels are at the positions of the intraresidue correlations, and a number of sequential connectivities are indicated. The peaks in the HMBC spectrum (A) arise from long-range two- and three-bond correlations of the type $C^{\alpha}H(i)$-$^{15}N(i)$, $C^{\beta}H(i)$-$^{15}N(i)$ and $C^{\alpha}H(i)$-$^{15}N(i+1)$, while those in the relayed HMQC-COSY spectrum (B) arise only from $C^{\alpha}H(i)$-$^{15}N(i)$ connectivities via an indirect pathway involving two successive one-bond scalar connectivities, namely, $^{15}N(i)$-NH(i) and NH(i)-$C^{\alpha}H(i)$. (From Clore, G. M., Bax, A., Wingfield, P. T., and Gronenborn, A. M., *FEBS Lett.*, 238, 17, 1988. With permission.)

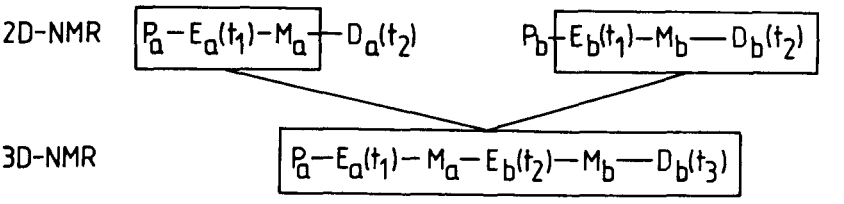


FIGURE 17. Generalized representation of the 3D-NMR experiment illustrating how a 3D-NMR experiment is derived from two 2D-NMR experiments.

are analogous to ^{15}N-filtered spectra of proteins containing only selected ^{15}N labeling (see Section IV.C), with the obvious difference that the selection in the former is by chemical shift, while in the latter it is by residue. Some reports on such experiments have already appeared.[84,85]

B. SEQUENTIAL ASSIGNMENT OF NUCLEIC ACID SPECTRA

There are only four bases to consider in oligonucleotides, compared with 20 amino acids in proteins. This renders the assignment of oligonucleotide spectra in general much easier than that of proteins. In terms of the actual assignment, nucleic acids differ from proteins in several respects. There are two independent scalar coupling networks within the basic nucleotide unit, one comprising the sugar, the other the base. The coupling networks are best identified by means of HOHAHA spectroscopy demonstrating direct and relayed through-bond connectivities along the H1′ → H2′/H2″ → H3′ → H4′ → H5′/H5″ pathway in each sugar unit, and between the H6 and H5 protons of cytosine and the H6 and CH_3 protons of thymine within the pyrimidine bases. The base and sugar moieties of the same nucleotide are connected by means of NOESY spectroscopy via the intraresidue H1′/H2′(i)-H8/H6(i) cross-relaxation pathway, while through-space connectivities along the H1′/H2′/H2″(i–1) → H8/H6(i) → H1′/H2′(i) and the H8/H6(i) → H5/CH_3(i+1) pathways yield sequential assignments along one strand.[86-98] In virtually all cases, complete assignments can be obtained by means of NOESY spectroscopy alone, without recourse to any scalar correlation experiments. This is in marked contrast to the situation in proteins where the spectral complexity necessitates the mandatory use of both scalar correlation and NOESY experiments. Assignment of the imino and amino protons, as well as the adenine H2 protons, is also carried out sequentially[89,98-108] on the basis of NOE spectra in H_2O. The expected NOE connectivities for right-handed A- and B- DNA, as well as for left-handed Z-DNA, are summarized in Figure 20, and examples of NOESY spectra in D_2O, as well as in H_2O, are shown in Figure 21.

C. SELECTIVE ISOTOPE LABELING FOR EXTENSION TO LARGER SYSTEMS

As all COSY, multiple-quantum, and HOHAHA experiments are limited to cases where the linewidths are not significantly larger than the size of the J couplings, it becomes increasingly difficult to identify proton spin systems as the size of the molecule increases. Thus, for proteins larger than about 20 kDa, much of the information on proton-proton scalar correlations is no longer available.

To some extent, the proton linewidth problem can be partially overcome by randomly incorporating about 80% deuterium into the protein. The dipolar broadening effect of deuterium on neighboring protons is quite small, as the gyromagnetic ratio of deuterium is seven times smaller than that of the proton. Of course, this approach also results in a reduced concentration of protons, but this is partially compensated for by their reduced linewidths. The application of this method has been demonstrated in the case of *E. coli* thioredoxin,[109,110] where random deuteration resulted in improvements in the quality of the COSY spectra of both the NH-$C^{\alpha}H$ and aliphatic-aliphatic regions. Additionally, a gain in the sensitivity of the NOESY spectrum

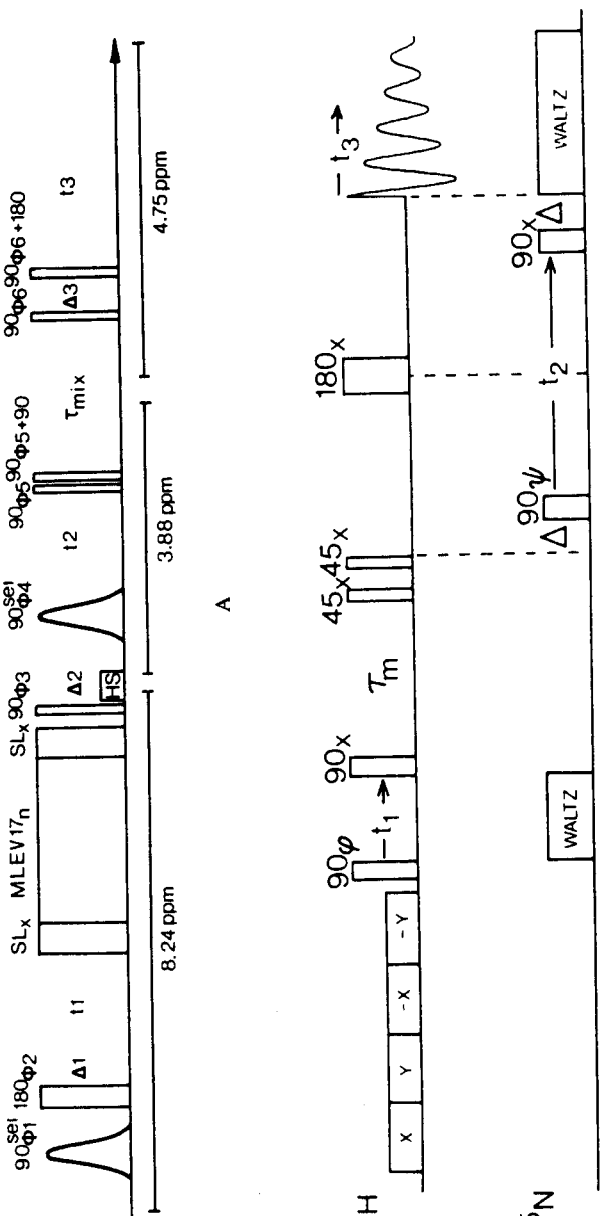
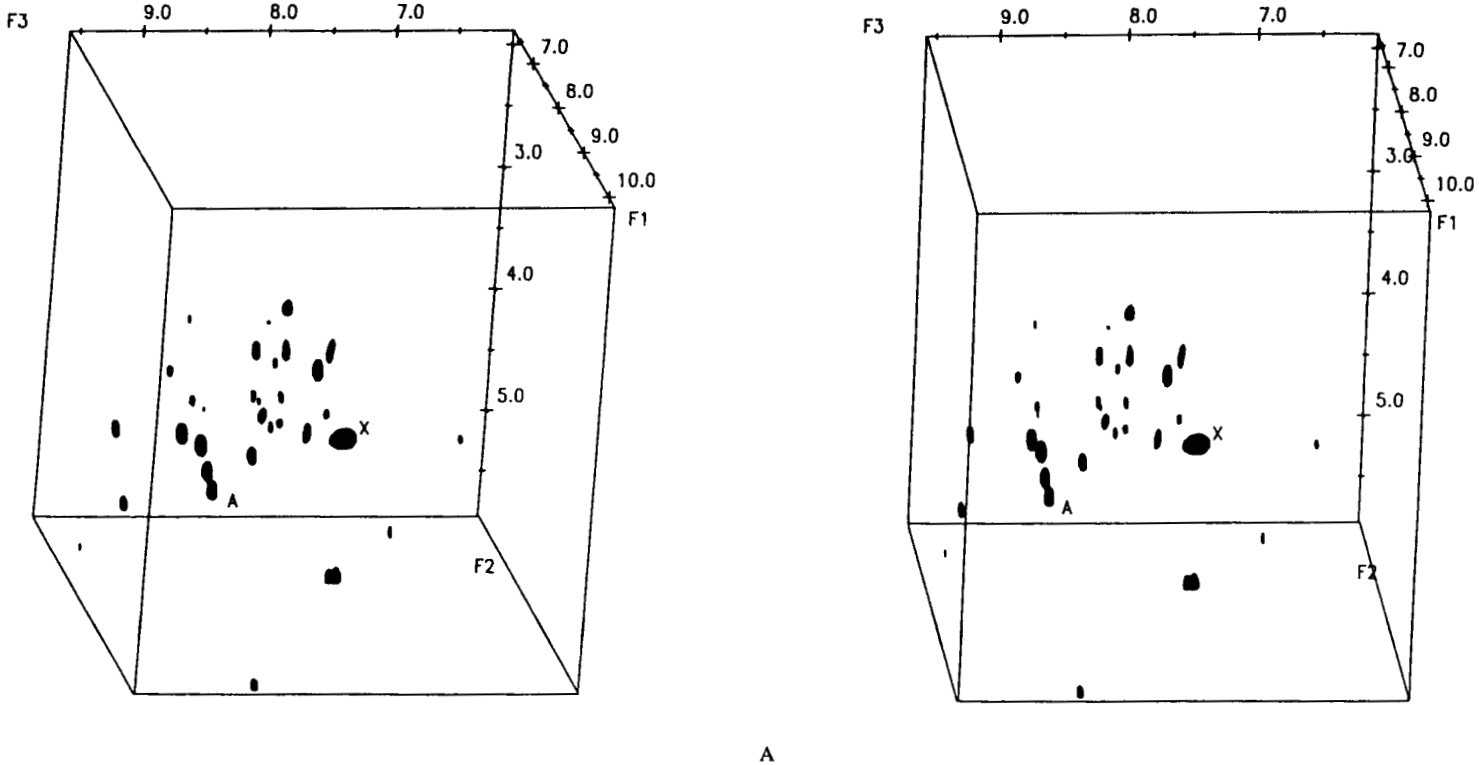


FIGURE 18. Pulse scheme for (A) the 3D homonuclear HOHAHA-NOESY and (B) the 3D ^{1}H-^{15}N HMQC-NOESY experiments. (From Oschkinat, H., Cieslar, C., Gronenborn, A. M., and Clore, G. M., *J. Magn. Reson.*, 82, 212, 1989; Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A., and Bax, A., *J. Am. Chem. Soc.*, 111, 1515, 1989.)



A

FIGURE 19. 500 MHz 3D HOHAHA-NOESY spectrum of α1-purothionin in H_2O. The NH(F1)-$C^{\alpha}H$(F2)-NH(F3) and NH(F1)-$C^{\alpha}H/C^{\beta}H$(F2)-aliphatic regions are shown in (A) and (B), respectively. The spectra are displayed as stereoviews. (From Oschkinat, H., Cieslar, C., Gronenborn, A. M., and Clore, G. M., *J. Magn. Reson.*, 81, 212, 1989. With permission.)

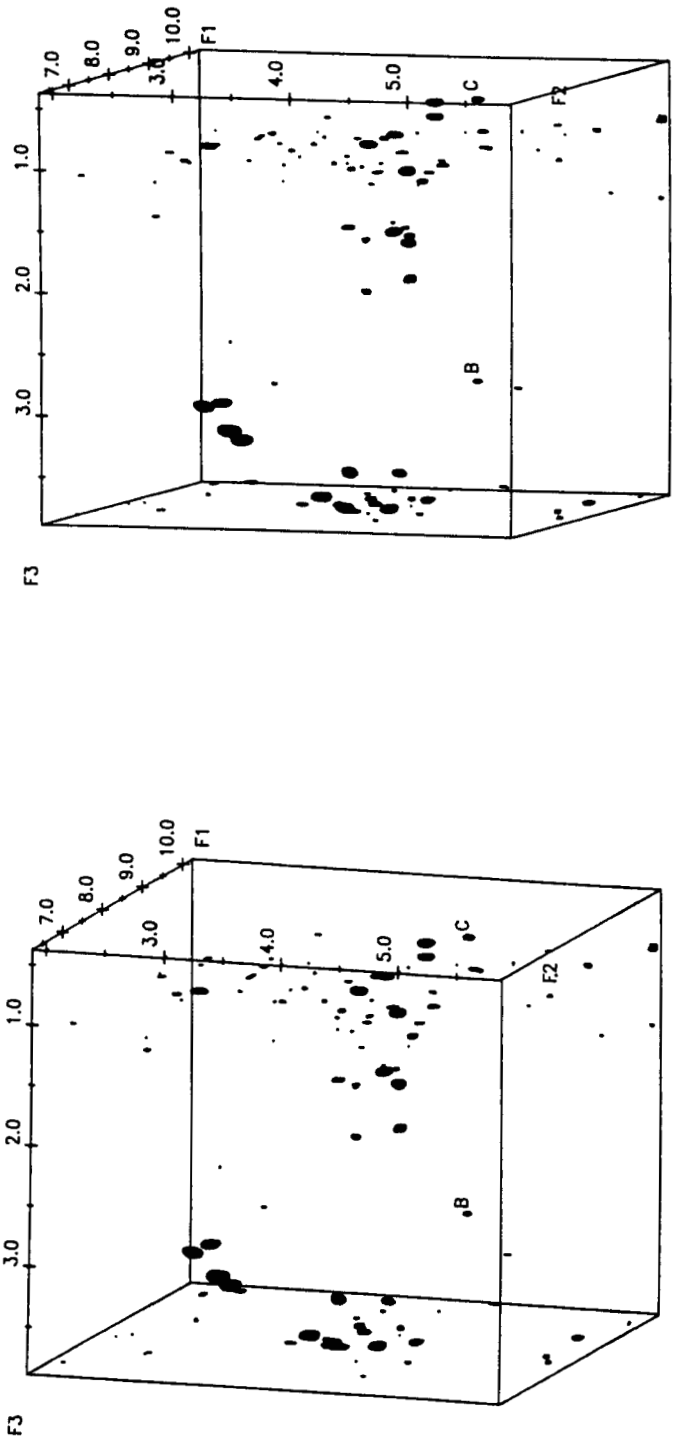


FIGURE 19B

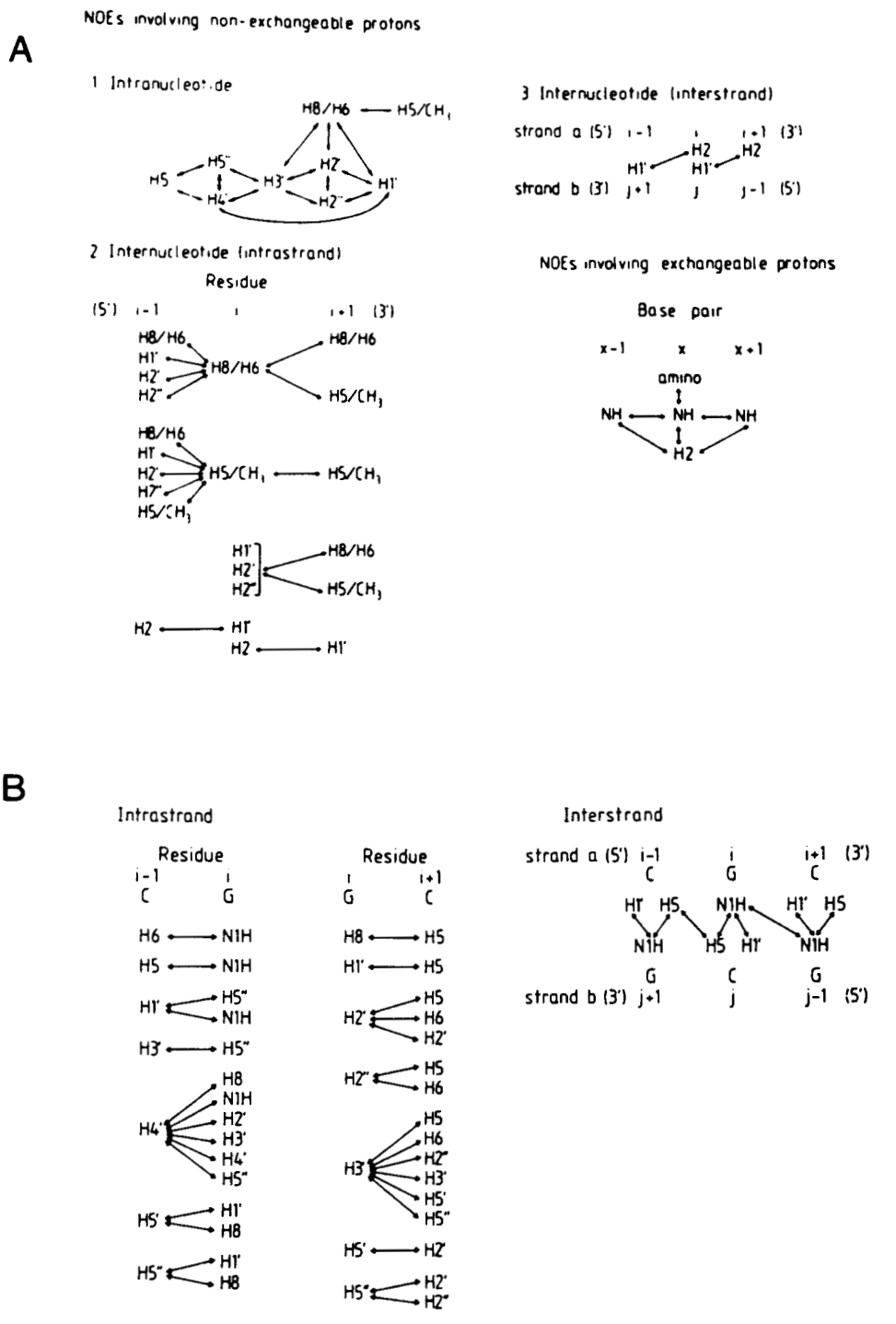


FIGURE 20. NOE sequential connectivities used for the assignment of (A) right-handed A- or B-DNA and (B) of left-handed Z-DNA.

in the NH-NH region is observed as longer mixing times can be used without engendering loss of magnetization via spin diffusion.[109-111]

An alternative procedure, however, can be envisaged for the unambiguous identification of

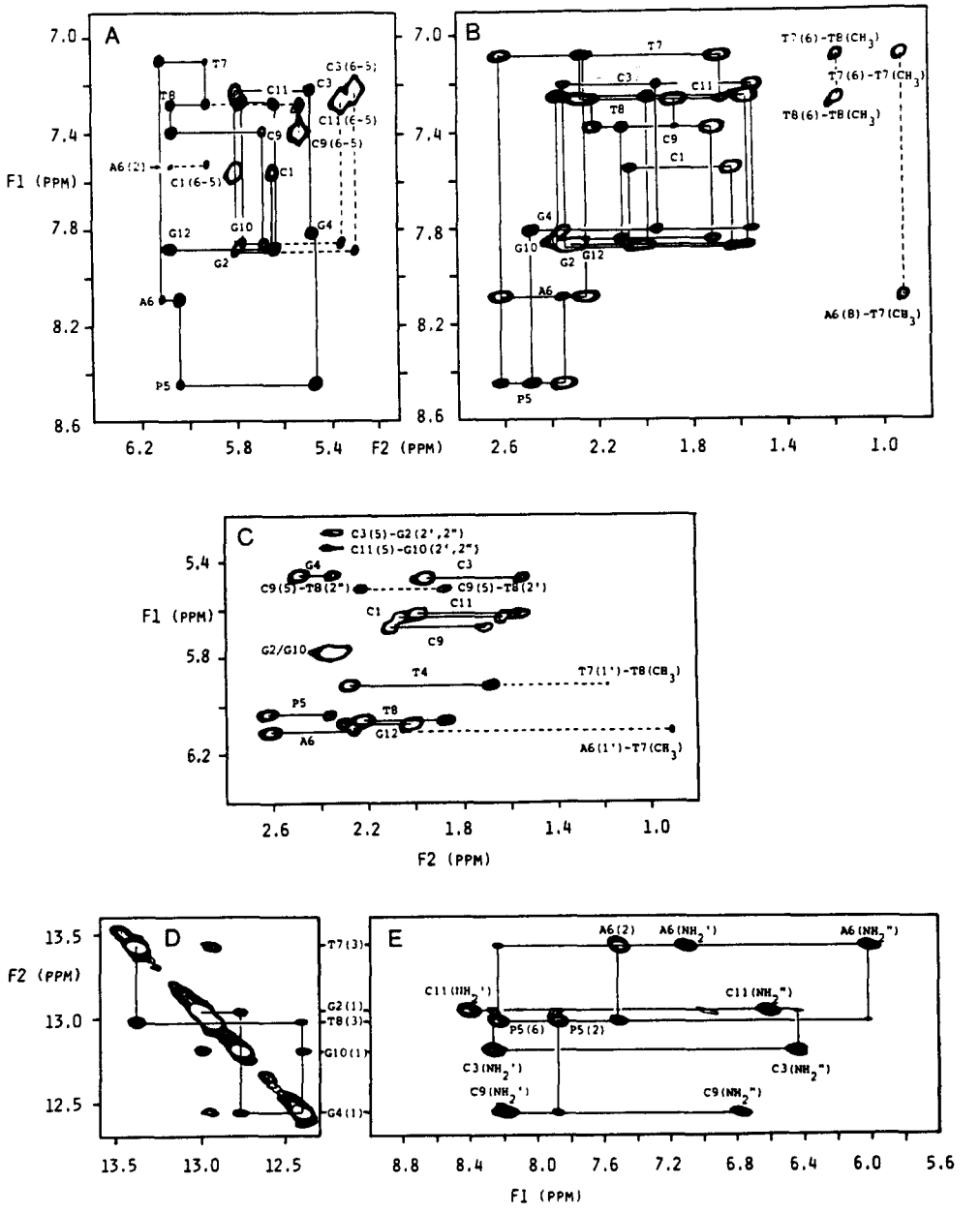


FIGURE 21. 500 MHz NOESY spectra of the DNA dodecamer 5′d(CGCGPATTCGCG)$_2$ in D_2O (A to C) and H_2O. The spectrum in D_2O was recorded with a mixing time of 40 ms and that in H_2O with a mixing time of 150 ms. The regions displayed are as follows: (A) H8/H6(F1)-H1′/H5(F2); (B), H8/H6(F1)-H2′/H2″/CH_3(F2); (C) H1′/H5(F1)-H2′/H2″/CH_3(F2); (D) NH(F2)-NH(F1); and (E) NH(F2)-aromatic/NH_2(F2). (From Clore, G. M., Oschkinat, H., McLaughlin, L. W., Benseler, F., Scalfi Happ, C., Happ, E., and Gronenborn, A. M., *Biochemistry*, 27, 4185, 1988. With permission.)

spin systems in large proteins. This makes use of ^{1}H-detected HMQC experiments to establish one-bond scalar correlations between protons and directly bonded ^{13}C or ^{15}N atoms. Because these one-bond couplings are large (90 Hz for ^{1}H-^{15}N and 125 to 160 Hz for ^{1}H-^{13}C), the method can be extended to proteins probably up to ~40 kDa in size. The requirement for this method is a set of amino acids labeled either with ^{15}N or with ^{13}C at selected positions (e.g., C^α or C^β). To

cover all 20 amino acids, a suitable combination involving only five protein samples for each type of labeling would be required. In this way, it would be possible to establish unambiguously the type of amino acid to which each NH, $C^{\alpha}H$, and $C^{\beta}H$ proton belonged. This information could then be used in conjunction with the 3D heteronuclear ^{15}N-^{1}H HMQC-NOESY experiment described in Section IV.A to establish sequential connectivities. The specifically labeled samples could also be used to exclusively observe protons directly attached to ^{15}N or ^{13}C. In this way isotopically edited NOESY spectra could be obtained[75,112-117] that would complement the information obtained by the 3D-experiment on a fully labeled ^{15}N sample.

D. USE OF SITE-SPECIFIC MUTAGENESIS

Comparison of the 2D spectra of homologous proteins may offer an additional approach for obtaining sequence-specific assignments, particularly for larger proteins for which ambiguities arising from chemical shift overlap cannot be resolved in the spectra of a single variant. Until recently, this approach has been restricted to the use of "natural variants" isolated from different species, and an example for the successful use of protein variants is the comparative NMR study of ovomucoid third domains.[118] Due to progress in modern molecular biology, the limitation of having to rely on natural variants has been overcome. It is now possible, in principle, to produce any protein and variant thereof by cloning its gene and carrying out site-directed mutagenesis. This strategy has been used for the sequence-specific assignment of tyrosine, histidine, and tryptophan residues of human interleukin-1β.[119,120] Further, in the case of hirudin, a small protein of only 65 residues, the availability of a mutant (Lys 47→Glu) in addition to the wild-type protein proved to be essential for the correct sequence specific assignment.[14]

It should also be mentioned at this point that NMR spectroscopy may have at least two important applications with respect to genetically engineered proteins. First, the structural identity of a natural and recombinant protein can be rapidly and easily ascertained from a comparison of the appropriate (NOESY, HOHAHA, etc.) 2D spectra, as these constitute a fingerprint providing a direct reflection of the 3D structure of the protein in solution. Second, NMR can establish whether the structure of a mutant is essentially unchanged with respect to that of the wild type. This application of NMR may develop into a useful tool aiding genetic engineers to correlate functional to structural changes.

V. STRUCTURAL RESTRAINTS

A. INTERPROTON DISTANCES

The initial slope of the time-dependent NOE, $N_{ij}(t)$, between two protons i and j is equal to the cross-relaxation rate σ_{ij} between the two protons.[121-124]

$$\left.\frac{dN_{ij}}{dt}\right|_{t=0} = \sigma_{ij} \tag{1}$$

σ_{ij} is simply the rate constant for exchange of magnetization between the two protons and is proportional to $\langle r_{ij}^{-6} \rangle$ and $\tau_{eff}(ij)$, where r_{ij} is the distance between the two protons, and $\tau_{eff}(ij)$ the effective correlation time of the i-j vector.[51]

$$\sigma_{ij} = \frac{\gamma^4 \hbar^2}{10 r_{ij}^6} \left(\tau_{eff}(ij) - \frac{6\tau_{eff}(ij)}{1 + 4\omega^2 \tau_{eff}(ij)^2} \right) \tag{2}$$

(γ is the gyromagnetic ratio of the proton, $\hbar$ Planck's constant divided by 2 π, and ω the spectrometer frequency.) It therefore follows that at short mixing times τ_m, ratios of NOEs can yield either ratios of distances or actual distances, if one distance is already known, through the relationship

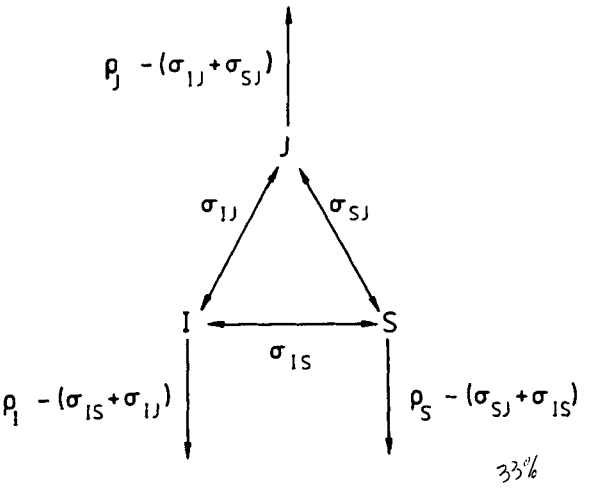


FIGURE 22. Three-spin system used for simulating effects of spin diffusion in NOE experiments.

$$r_{ij}/r_{kl} = (\sigma_{kl}/\sigma_{ij})^{1/6} \tag{3}$$

providing the effective correlation times for the two interproton vectors are approximately the same.

In practice, initial slope measurements are not entirely trivial. First, the magnitude of the NOEs at very short mixing times are small, inevitably posing a signal-to-noise problem. Second, the measured NOE at short mixing times may not reflect the true magnitude of the NOE. In the 2D experiment, for example, it may not be possible to completely remove the contribution from zero-quantum coherence transfer to a cross-peak between two scalar coupled spins. Although the zero-quantum coherence contribution is dispersive and hence has zero integrated intensity, the superposition of large zero-quantum coherence peaks on top of small NOE ones invariably leads to large integration errors. Similarly, in the 1D experiment involving the saturation of resonance j for a short time t followed by the observation of other proton resonances, it may not be feasible, owing to resonance overlap, to use a sufficiently high power to achieve instantaneous saturation. Third, it may not be possible, for reasons of limitations in measuring time, to obtain a sufficient number of points covering the time course of the NOE with adequate signal-to-noise ratios. Finally, the measured initial slope for the NOE between protons i and j will not reflect the true initial slope if both $\sigma_{ij} \leq \sigma_{is}$ and $\sigma_{ij} \leq \sigma_{js}$, where s represents a third proton. Therefore, from the experimental point of view, it is important to ask two questions: (1) how many time points for the NOE time course does one really need to obtain, and (2) what is the time dependence of the magnitude of the error Δr_{kl} introduced into the determination of the unknown interproton distance r_{kl}, when r_{kl} is calculated from the approximate relationship

$$r_{kl} \sim r_{ij}[N_{ij}(t)/N_{kl}(t)]^{1/6} \sim r_{ij}[a_{ij}(\tau_m)/a_{kl}(\tau_m)]^{1/6} \tag{4}$$

as opposed to exact expression given in Equation 3, where r_{ij} is a known fixed internal reference distance, $N_{ij}(t)$ and $N_{kl}(t)$ are the NOEs observed between protons i and j and between protons k and l, respectively, in the 1D selective saturation experiment at time t, and $a_{ij}(\tau_m)$ and $a_{kl}(\tau_m)$ are the corresponding NOE cross-peak intensities in the 2D NOESY experiment at a mixing time τ_m.

These questions can be answered by considering the three-spin system illustrated in Figure 22 that provides a perfectly adequate description for understanding the behavior of a multiple-

spin system. The z magnetization of the three spins is governed by three coupled ordinary differential equations that can be conveniently expressed in matrix form:

$$\begin{pmatrix} \dot{I} \\ \dot{J} \\ \dot{S} \end{pmatrix} = \begin{pmatrix} -\rho_I & \sigma_{IJ} & \sigma_{IS} \\ \sigma_{IJ} & -\rho_J & \sigma_{SJ} \\ \sigma_{IS} & \sigma_{SJ} & -\rho_S \end{pmatrix} \begin{pmatrix} I - I_o \\ J - J_o \\ S - S_o \end{pmatrix} \quad (5)$$

I is the magnetization of proton I at time t, I_o the equilibrium magnetization of proton I prior to the perturbation that permits one to detect exchange through cross-relaxation, ρ_I the total spin-lattice relaxation rate of proton I given by $\Sigma_{J \neq I} \sigma_{IJ} + R_{1I}$ (where R_{1I} is the external relaxation rate, i.e., the leakage rate from the system), and similarly for protons J and S.

For the 1D selective saturation experiment in which resonance J is selectively saturated, $I_o = J_o = S_o = 1$ and the initial conditions (i.e., at t = 0) are I = S = 1 and J = 0. The magnitude of the observed NOE $N_{IJ}(t)$ observed on resonance I following saturation of resonance J is simply given by $[I(t) - I_o]/I_o$, and similarly for the NOE, N_{JS} observed on resonance S.

In the case of the 2D experiment (cf. Figures 4, 12, and 13), the diagonal peaks represent the conventional 1D spectrum and pairs of cross-peaks at symmetrical locations with respect to the diagonal peaks arise from cross-relaxation (cf. Figure 1). Bearing symmetry considerations in mind, the mathematics of the dynamics of the intensity a_{JJ} of the diagonal peak at location (F_{1J}, F_{2J}) and the intensities of the cross-peaks a_{JI} and a_{JS} at locations (F_{1J}, F_{2I}) and (F_{1J}, F_{2S}), respectively, are identical in form to the 1D selective inversion experiment $180°_{s} - t - 90°_{ns}$ with the selective 180° pulse applied to resonance J. In the latter case, $I_o = J_o = S_o = 1$ with initial conditions I = S = 1 and J = −1. In the 2D experiment, all that needs to be done to obtain the dynamics of $a_{JJ}(\tau_m)$, $a_{JI}(\tau_m)$, and $a_{JS}(\tau_m)$ is to modify Equation 5 by setting $I_o = J_o = S_o = 0$, replacing J by a_{JJ}, I by a_{JI} and S by a_{JS}, with initial conditions $a_{JI} = a_{JS} = 0$, and $a_{JJ} = 1$.

The results of a series of calculations for both the 1D and 2D experiments are summarized in Figure 23, illustrating the time dependence of the error Δr_{SJ} in the estimate of the distance r_{SJ} and the magnitudes of the NOEs for different values of the distances . $\omega\tau_{eff}$ is set to 15.7 for the three interproton vectors, which corresponds to an effective correlation time of 5 ns at a spectrometer frequency of 500 MHz; the value of the reference distance r_{IJ} is 2.4 Å, which corresponds to a cross-relaxation rate of 1.47 s^{-1}, and all three external relaxation rates have been set to 0.5 s^{-1}. The error in calculating r_{SJ} from the approximation given by Equation 4 as a function of time is given by

$$\Delta r_{SJ}(t) = r_{IJ}\{[N_{IJ}(t)/N_{JS}(t)]^{1/6} - [\sigma_{IJ}/\sigma_{JS}]^{1/6}\} \quad (6)$$

for the 1D experiment and by

$$\Delta r_{SJ}(\tau_m) = r_{IJ}\{[a_{JI}(\tau_m)/a_{JS}(\tau_m)]^{1/6} - [\sigma_{IJ}/\sigma_{JS}]^{1/6}\} \quad (7)$$

for the 2D experiment. Four main features emerge from Figure 23:

1. The initial slope of the time development of the 1D and 2D NOEs is the same, as expected.
2. The approximation given by Equation 4 introduces only small errors (≤ 0.2 Å) in the estimate of r_{SJ} up to relatively long times (~0.4 s in the present example) for both the 1D and 2D experiments, *providing* either $\sigma_{SJ} \geq \sigma_{IJ}$ or $\sigma_{JI} \geq \sigma_{IS}$ (Figures 23A and B). This is despite the fact that the initial rate approximation ($\sigma_{JS} \sim N_{JS} \cdot t \sim a_{JS} \cdot \tau_m$) breaks down at much shorter mixing times (50 ms in the present example). In other words, the approximation

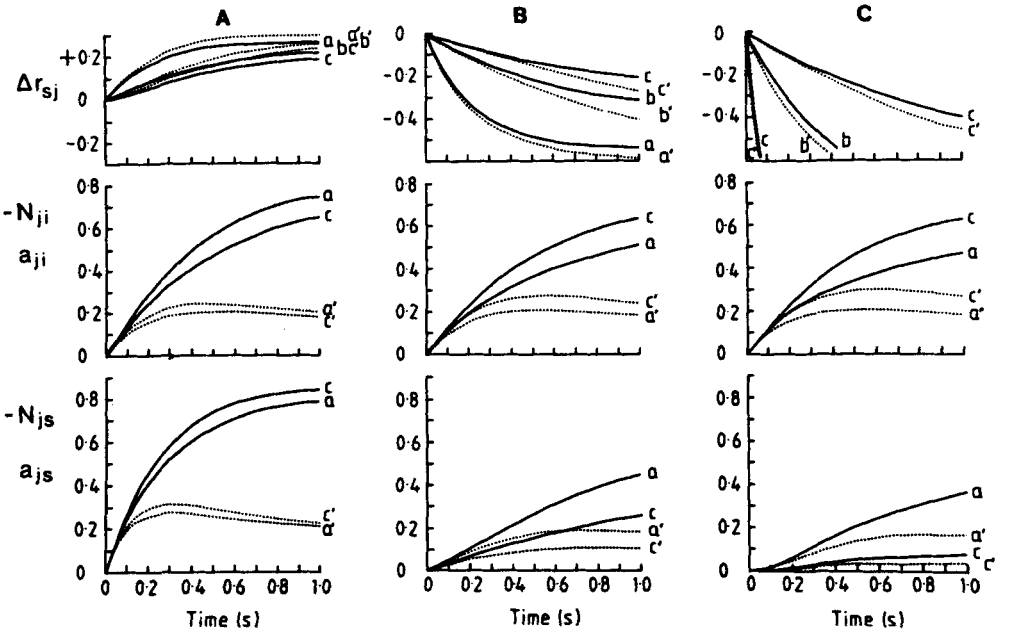


FIGURE 23. Time dependence of (1) the error of Δr_{SJ} (Å) in the estimate of r_{SJ} using the approximation given by Equation 4 with r_{IJ} as the internal reference distance; (2) the magnitudes of the NOEs, N_{JI}, and N_{JS}, observed on resonances I and S, respectively, following irradiation of resonance J in the 1D NOE experiment; and (3) the cross-peak intensities a_{JS} and a_{JI} for the 2D NOESY experiment, for different values of r_{SJ} and r_{IS}. The continuous lines are the time courses for the 1D experiment, whereas the dashed lines are those for the 2D experiment. Δr_{SJ} is calculated using Equation 6 for the 1D experiment and Equation 7 for the 2D experiment. $\omega\tau_{eff}$ has a value of 15.7, which corresponds to a value of 5 ns for τ_{eff} at 500 MHz, and the external relaxation rates have a value of 0.5 s^{-1} for all protons. The internal reference distance r_{IJ} is set to 2.4 Å, which corresponds to a value of 1.47 s^{-1} for the cross-relaxation rate σ_{IJ}. r_{SJ} has values of 2.1 Å (A), 3.0 Å (B), and 3.8 Å (C). r_{IS} has values of 2.1 Å (curves a and a'), 3.0 Å (curves b and b'), and 4.6 Å (curves c and c'). Thus, the relationships between the cross-relaxation rates are as follows: (A) $\sigma_{SJ} > \sigma_{IJ}$ and $\sigma_{SJ} \geq \sigma_{IS}$; (B) $\sigma_{SJ} < \sigma_{IJ}$ and $\sigma_{SJ} < \sigma_{IS}$ (curves a and a'), $\sigma_{SJ} \geq \sigma_{IS}$ (curves b, b', c, and c'); (C) $\sigma_{SJ} < \sigma_{IJ}$, $\sigma_{SJ} < \sigma_{IS}$ (curves a, a', b, and b'), $\sigma_{SJ} > \sigma_{IS}$ (curves c and c'). (a, b, and c refer to the 1D experiment, while a', b', and c' refer to the 2D experiment.) (From Clore, G. M. and Gronenborn, A. M., *J. Magn. Reson.*, 61, 158, 1985. With permission.)

given by Equation 4 remains valid for values of τ_m up to eight times longer than the initial rate approximation.

3. The errors in the 2D experiment arising from the use of Equation 4 are *always larger* at any given time t than those in the 1D experiment. This is obvious from the nature of the time courses of the NOEs in the two cases. Whereas the intensity of the 1D NOE increases in magnitude to a maximum value of –1, the intensity of the 2D NOE cross-peak first exhibits an increase followed by a decrease back toward zero.
4. If the cross-relaxation rate of the S-J vector is *greater* than that of the reference I-J vector (Figure 23A), Equation 4 leads to an *overestimation* of r_{SJ}. If, on the other hand, the cross-relaxation rate of the S-J vector is *less* than that of the reference I-J vector (Figure 23B and C), the value of r_{SJ} will be *underestimated* through the use of Equation 4.

In order to apply Equation 4 one needs suitable fixed internal reference distances. These are available in the case of oligonucleotides and include those between the H5 and H6 protons of cytosine, and between the H2′ and H2″ protons and H1′ and H2″ protons of the sugar. Further, the variations in effective correlation time are relatively small in terms of their effect on distance estimates (due to the r^{-6} dependence of the NOE) and are easily anticipated. As a result, reasonably accurate interproton distances can be obtained for oligonucleotides.[95-97,125-137] The

general accuracy of such measurements is of the order of –0.2/+0.3 Å for distances less than 2.5 Å, and –0.3/+0.4 Å for distances between 2.5 and 5 Å. The reason that the error estimates are asymmetric is that there is a systematic error involved in the determination of interproton distances using Equation 1, due to the presence of a small amount of unavoidable spin diffusion (cf. Figure 23). This effect, as well as corrections dealing with dynamical averaging should in principle be included in the distance estimates by a full multiple-spin treatment.[138] For practical purposes, however, the three-spin approximation discussed above is sufficient. From Figure 23 it is clear that the apparent value of the unknown distance will become closer to that of the reference distance as the extent of spin diffusion increases. In other words, if the unknown distance is *smaller* than the reference distance, its value will be systematically *overestimated*; in contrast, if the unknown distance is *larger* than the reference distance, its value will be systematically *underestimated.*

In some cases, problems associated with spin-diffusion can be avoided by using 2D rotating frame nuclear Overhauser Spectroscopy (ROESY) spectra[139-141] for deriving interproton distances. This is because spin-diffusion in the rotating frame is easily detected as second-order effects are opposite in sign to first-order effects. Quantification, however, of ROEs between scalar coupled protons is difficult due to the inevitable contribution of Hartmann-Hahn effects (which are opposite in sign to those of the ROE).

An alternative method for avoiding problems associated with spin-diffusion is to record a conventional NOESY spectrum with irradiation of selective resonances during the mixing period. For example, if the H2″ protons in an oligonucleotide are irradiated, spin-diffusion via the H2″ proton is completely removed. In this manner, it is possible to measure cross-relaxation rates without the need to correct for spin-diffusion.[142]

Some attempts have also been made to carry out a full relaxation matrix analysis of a trial structure, comparing the calculated NOE intensities with the experimental ones and adjusting the distance estimates appropriately.[143-146] The structure is then recalculated on the basis of the new set of distances and the procedure is repeated until the calculated NOE intensities match up with the experimental ones. This is obviously an elegant approach, but suffers at present from a number of pitfalls, so that the results have to be treated with some degree of caution. As the method ignores any potential variation in effective correlation times for different interproton vectors, it gives the impression that the final distances are estimated to a greater degree of accuracy than they may in fact be. Second, the whole procedure is *model dependent* and does *not* exclude the possibility of different distance sets yielding the same computed NOESY spectra. Because of these drawbacks, we favor at the present time the use of the approximation given by Equation 4, providing the errors involved are appropriately taken into account in the ensuing structure calculations.

A model independent method of obtaining distances considering all the cross-relaxation pathways would probably be the method of choice. One possible approach involves carrying out the analysis in n-dimensional distance space employing the same mathematical techniques that are used in metric matrix distance geometry algorithms (cf. Section VIII.A). The idea is a relatively simple one. A metric matrix is constructed by choosing distances at random between all pairs of atoms within the triangle limits imposed by the covalent geometry, the hard-sphere van der Waals radii, and the experimental NMR restraints. The distances are then adjusted within these limits in an iterative manner until the computed NOE spectrum matches up with the experimental one. By repeating the calculation several times with different random number seeds, error limits on the distance estimates are easily obtained.

In contrast to oligonucleotides where the differences in correlation times for different interproton vectors are relatively small, significant variations in effective correlation times are observed in the case of proteins as one progresses outward from the main chain atoms to the side chain atoms.[147] This is associated with higher mobility of side chains, especially long ones, and is most marked at the protein surface. Further, there are relatively few suitable internal reference

distances available due to extensive spectral overlap. Consequently, any attempt to use Equation 4 to measure distances is fraught with uncertainties. Nevertheless, because of the $\langle r^{-6} \rangle$ dependence of the NOE, approximate interproton distance restraints can clearly be derived even in the presence of large variations in effective correlation times. Empirically, the type of classification generally used is one in which strong, medium, and weak NOEs correspond to distance ranges of approximately 1.8 to 2.7, 1.8 to 3.3, and 1.8 to 5.0 Å, where the lower limit of 1.8 Å corresponds to the sum of the van der Waals radii of two protons.[4,7] By using such a scheme, variations in effective correlation times do not introduce errors into the distance restraints. Rather, they only result in an increase in the estimated range for a particular interproton distance. Thus, the effect of a decrease in the effective correlation time of an interproton vector i-j is simply manifested by a reduction in the magnitude of the corresponding NOE, N_{ij}, and therefore by a reduction in the precision with which that distance is defined. Thus, for example, an interproton distance of, say, 2.3 Å for a very mobile side chain is likely to appear as a weak NOE, and hence be classified in the 1.8 to 5 Å range instead of in the 1.8 to 2.7 Å range.

B. TORSION ANGLE RESTRAINTS

Vicinal spin-spin coupling constants can provide useful information supplementing the interproton distance restraints derived from NOE data. In particular, ranges of torsion angles can be estimated from the size of the coupling constants. The latter may be obtained by analyzing the multiplet pattern in COSY and COSY-like (e.g., DQF-COSY, exclusive COSY [E-COSY], and z-COSY) spectra.

1. ϕ Backbone Torsion Angle Restraints

The easiest coupling constants to determine in proteins are the $^3J_{HN\alpha}$ coupling constants that can be obtained by simply measuring the peak-to-peak separation of the anti-phase components of the $C^\alpha H$-NH COSY cross-peaks. The size of the $^3J_{HN\alpha}$ coupling constant is related to the ϕ backbone torsion angle through a Karplus-type relationship[148] of the form:[149]

$$^3J_{HN\alpha} = 6.4\cos^2(\phi - 60°) - 1.4\cos(\phi - 60°) + 1.9 \quad (8)$$

Consequently, values of $^3J_{HN\alpha} < 6$ Hz and > 8 Hz correspond to ranges of $-10°$ to $-90°$ and $-80°$ to $-180°$, respectively, for the ϕ backbone torsion angles.[149] Considerable care, however, has to be taken in deriving ϕ backbone torsion angle ranges from apparent values of $^3J_{HN\alpha}$ coupling constants measured in this way, as the minimum separation between the anti-phase components of a COSY cross-peak is equal to ~0.58 times the linewidth at half-height.[150] That is to say that small coupling constants can only be determined for relatively sharp resonances.

2. Side Chain Torsion Angle Restraints and Stereospecific Assignments

χ_1 side chain torsion angle restraints and stereospecific assignments can be obtained by analyzing the pattern of $^3J_{\alpha\beta}$ coupling constants and the relative intensities of the intraresidue NOEs from the NH and $C^\alpha H$ protons on the one hand, to the two $C^\beta H$ protons on the other,[20,151] and in the case of Valine to the $C^\gamma H_3$ protons.[152] The $^3J_{\alpha\beta}$ coupling constants are related to the χ_1 torsion angle via the Karplus relationships:[153]

$$^3J_{\alpha\beta}(2) = 9.5\cos^2(\chi_1 - 120°) - 1.6\cos(\chi_1 - 120°) + 1.8 \quad (9)$$

$$^2J_{\alpha\beta}(3) = 9.5\cos^2\chi_1 - 1.6\cos\chi_1 + 1.8 \quad (10)$$

and are best measured from correlation spectra that yield reduced multiplets such as β-COSY,[154] E-COSY,[155,156] primitive exclusive COSY (PE-COSY),[157] or z-COSY.[158,159] In addition, under suitable conditions, they can be qualitatively assessed from the $C^\alpha H$-$C^\beta H$ cross-peak shapes in

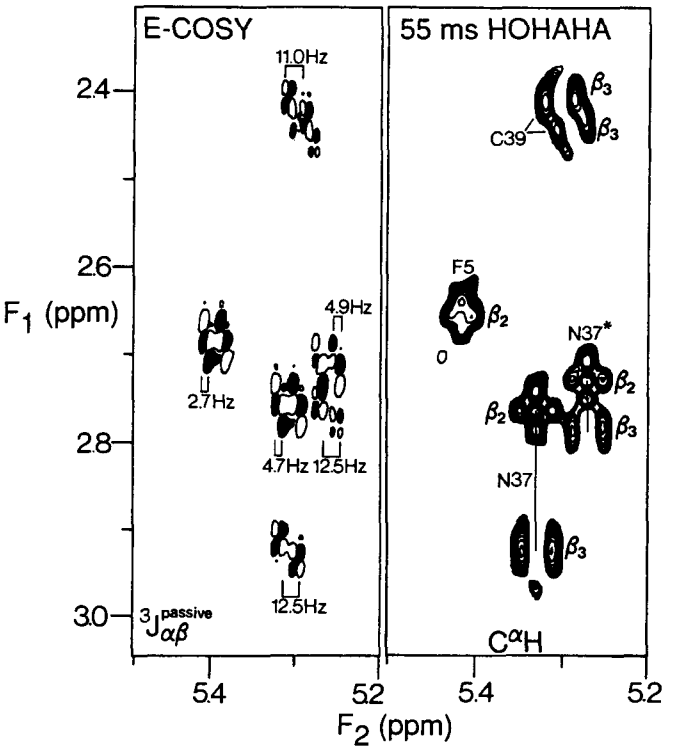
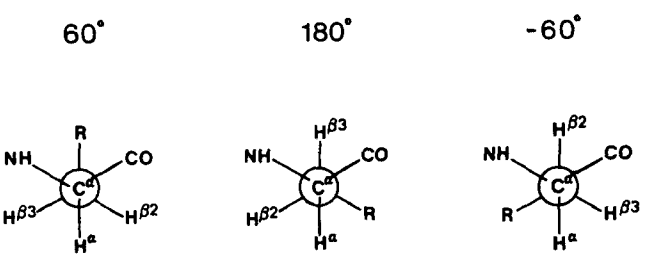


FIGURE 24. 600 MHz 55 ms HOHAHA (27°C) and E-COSY (45°C) spectra displaying cross-peak shapes for selected AMX spin systems of BDS-I. Note that the magnitude of the passive $^3J_{\alpha\beta}$ coupling constants can be directly visualized from the shape of HOHAHA peaks. Positive and negative components of the E-COSY cross-peaks are distinguished by filled and open contour levels. The stereospecific assignments of the $C^\beta H$ protons is indicated on the HOHAHA spectrum. (From Driscoll, P. C., Clore, G. M., Beress, L., and Gronenborn, A. M., *Biochemistry*, 28, 2178, 1989. With permission.)

HOHAHA spectra.[160] This is illustrated in Figure 24. These cross-peak shapes arise from the superposition of resolved anti-phase components created during the mixing process, on a pure phase absorption peak. For a $C^\alpha H$-$C^\beta H$ cross-peak with small active $^3J_{\alpha\beta a}$ and large passive $^3J_{\alpha\beta b}$ coupling constants, the HOHAHA cross-peak exhibits a double-lobed appearance. For the reverse situation, namely, large active $^3J_{\alpha\beta a}$ and small passive $^3J_{\alpha\beta b}$ coupling constants, the HOHAHA cross-peak takes on a four-lobed cross-like shape. For the other common situation of both small active and small passive coupling constants, the HOHAHA cross-peak retains a simple single absorption peak with a relatively small width in the $C^\alpha H$ dimension. The qualitative information derived from the HOHAHA spectrum is often sufficient for the purposes of stereospecific assignments and nicely complements the more precise data available from an E-COSY or PE-COSY spectrum (Figure 24). Further, the HOHAHA spectrum permits one to infer the relevant information even in cases where cross-peaks are partially overlapping, as is the case in the spectra illustrated in Figure 24.

If both $^3J_{\alpha\beta}$ couplings are small (~3 Hz), then χ_1 must lie in the range $60 \pm 60°$. If, on the other hand, one of the $^3J_{\alpha\beta}$ couplings is large and the other small, χ_1 can lie either in the range $180 \pm 60°$ or $-60 \pm 60°$. These two possibilities are easily distinguished on the basis of short mixing time NOESY experiments that yield simultaneously stereospecific assignments of the β-methylene protons. A scheme outlining the information necessary to determine the χ_1 angles and



$^3J\alpha\beta2$ (Hz)	<4	<4	>10
$^3J\alpha\beta3$ (Hz)	<4	>10	<4
NOEs	αβ2~αβ3	αβ2>αβ3	αβ2<αβ3
NOEs	NHβ2<NHβ3	NHβ2~NHβ3	NHβ2>NHβ3

FIGURE 25. Simple scheme for obtaining stereospecific assignments of β-methylene protons on the basis of $^3J_{\alpha\beta}$ coupling constants and intraresidue NOEs involving the NH, $C^\alpha H$, and $C^\beta H$ protons.

stereospecific assignments is illustrated in Figure 25. Clearly, this approach may fail if a side chain has a mixture of conformations or the χ_1 angle deviates by more than ~40° from the staggered rotamer conformations (i.e., 60°, 180°, and –60°). In the former case, the coupling and NOE data are mutually inconsistent, while in the latter, it may not be possible to make an unambiguous distinction between two rotamer conformations. Fortunately, analysis of high-resolution X-ray structures in the protein data bank has shown that 95% of all χ_1 angles lie within ±15° of the staggered rotamer conformations, and that there is a very clear correlation between the values of χ_1 and the degree of refinement, the better refined the structures, the closer the χ_1 angles to the ideal staggered rotamer conformations.[161,162] These results suggest that stereospecific assignments can be potentially obtained for up to 95% of β-methylene protons using this simple approach. In practice, of course, the percentage of stereospecific assignments will be lower due to either spectral overlap or large linewidths, preventing the determination of $^3J_{\alpha\beta}$ coupling constants.

A more rigorous approach for stereospecific assignment involves matching the observed $^3J_{HN\alpha}$ and $^3J_{\alpha\beta}$ coupling constants, together with approximate distances from the intraresidue $C^\alpha H$-$C^\beta H$ and NH-$C^\beta H$ NOEs and the interresidue $C^\alpha H$(i–1)-NH(i), $C^\alpha H$(i)-NH(i+1), $C^\beta H$(i–1)-NH(i), and $C^\beta H$(i)-NH(i+1) NOEs to the expected values for all combinations of ϕ, ψ, and χ_1 torsion angles held in a data base.[163,164] The data base search is carried out for both possible stereospecific assignments, and in those cases where only one of the two assignments satisfies the information in the data base, the correct stereospecific assignment, together with ranges for the ϕ, ψ, and χ_1 torsion angles, are obtained. This approach has been tested using model data derived from 20 crystal structures that were solved at a resolution of ≤2 Å and refined to a R-factor of <20%. Using this method, it is possible to obtain approximately 80% of the possible β-methylene stereospecific assignments using a distance classification of ≤2.7, ≤3.3, and ≤5 Å for strong, medium, and weak NOEs, and the condition that the difference in distance between a strong and weak NOE to members of a given C^β methylene pair is ≥0.5 Å.[163] A key advantage of this method is that it allows one to obtain much stricter limits for the ϕ, ψ, and χ_1 torsion angle restraints than would otherwise be possible.

As an alternative, one could envisage an approach that would make use of a data base of high-resolution X-ray structures. The method would be similar to that described by Kraulis and

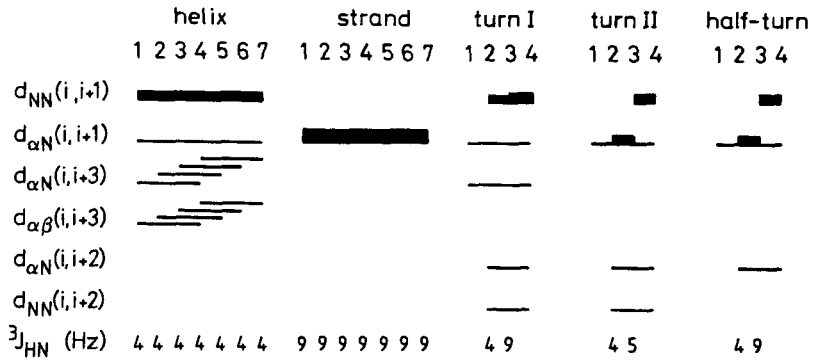


FIGURE 26. Characteristic patterns of short range NOEs involving the NH, $C^{\alpha}H$, and $C^{\beta}H$ protons seen in various regular secondary structure elements. The NOEs are classified as strong, medium, and weak, reflected by the thickness of the lines.

Jones[187] for the identification of secondary structure elements (see Section VIII). In particular, the coupling constant data together with all the NOE data involving the NH, $C^{\alpha}H$, and $C^{\beta}H$ protons (as well as the $C^{\delta}H$ protons of proline) for five residue segments would be matched with the existing structures in the data base. If a unique local structure type is found, then the stereospecific assignments as well as estimates of the ϕ, ψ, and χ_1 angles are obtained. In general, such an approach will only be useful when there are a large number of medium-range NOEs (i.e., $1 < |i-j| \leq 5$) within the five residue segment, and, consequently, will probably be limited to regions of regular secondary structure.

3. Dihedral Angle Restraints in Nucleic Acids

A Karplus-type correlation between the $^3J_{HP}$ coupling constant and the HCOP dihedral angle has been derived, which is of the form:[165]

$$^3J_{HP} = 15.3\cos^2\varphi - 6.1\cos\varphi + 1.6 \quad (11)$$

Using a selective version of the ^{1}H-detected HMQC scheme, $^3J_{C3'H,P}$ coupling constants can be accurately measured for oligonucleotides.[166] For example, in the case of the DNA dodecamer 5′d(CGCGAATTCGCG)$_2$, values for the C4′-C3′-O3′-P backbone torsion angle (ε) were derived on the basis of $^3J_{C3'H,P}$ coupling constants using the simple relationship $\varepsilon = 120° + \varphi$.[166] The resulting values of ε were in good agreement with the values obtained from the single crystal X-ray structure of the same oligonucleotide.[167] With the exception of one residue, G_{10}, where the value of ε in the X-ray structure was in the g^+ conformation rather than in the usual *trans* conformation, all the residues had a *trans* conformation for ε and the average angular rms difference between the solution and X-ray values was only $8 \pm 6°$, which is well within the errors of the two techniques.

VI. SECONDARY STRUCTURE IN PROTEINS

Regular secondary structure elements can be easily identified from a qualitative interpretation of the sequential NOEs, as each type of secondary secondary structure element is characterized by a particular pattern of short-range ($|i-j| \leq 5$) NOEs.[56,67,69,168,169] This is illustrated in Figure 26. Thus, for example, helices are characterized by a stretch of strong or medium NH(i)-NH(i+1) NOEs, and medium or weak $C^{\alpha}H$(i)-NH(i+3), $C^{\alpha}H$(i)-$C^{\beta}H$(i+3) NOEs, and $C^{\alpha}H$(i)-NH(i+1) NOEs, sometimes supplemented by NH(i)-NH(i+2) and $C^{\alpha}H$(i)-NH(i+4) NOEs (Figure 26). Strands, on the other hand, are characterized by very strong $C^{\alpha}H$(i)-NH(i+1)

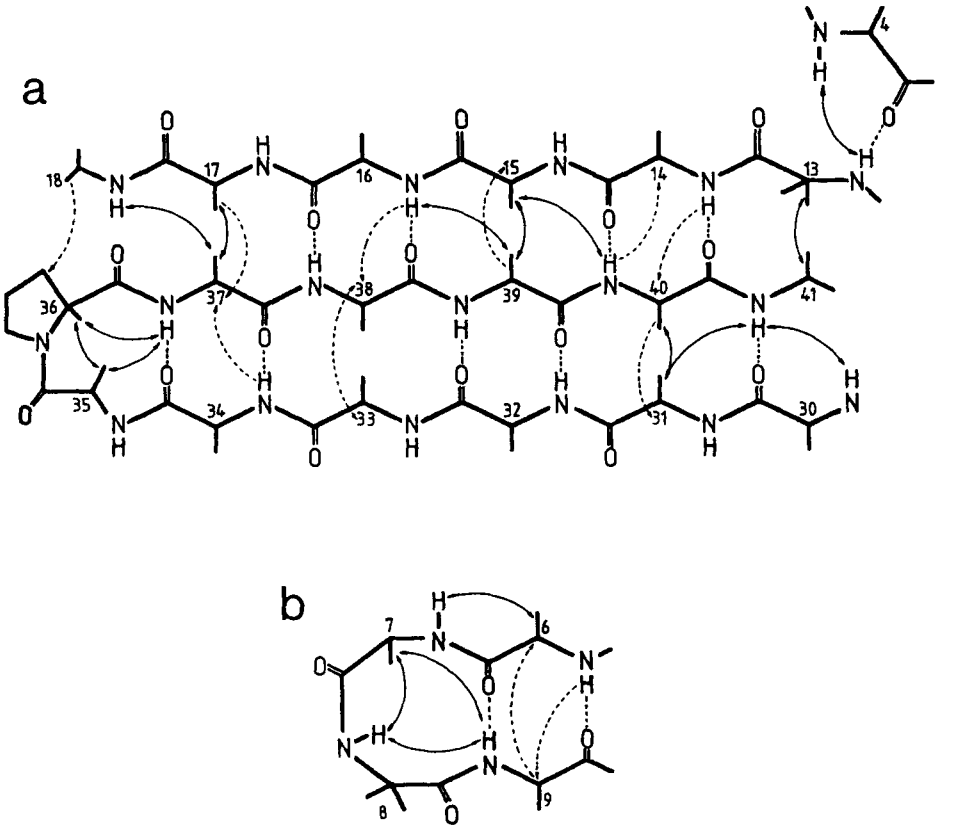


FIGURE 27. Pattern of interstrand NOEs observed in an anti-parallel β-sheet illustrated with respect to the sea anemone protein BDS-I. (From Driscoll, P. C., Clore, G. M., Beress, L., and Gronenborn, A. M., *Biochemistry*, 28, 2178, 1989. With permission.)

NOEs and by the absence of other short range NOEs involving the NH and $C^{\alpha}H$ protons. β-Sheets can be identified and aligned from interstrand NOEs involving the NH, $C^{\alpha}H$, and $C^{\beta}H$ protons (Figure 27). It should also be pointed out that the identification of secondary structure elements is aided by NH exchange data insofar that slowly exchanging NH protons are usually involved in hydrogen bonding, and by $^{3}J_{HN\alpha}$ coupling constant data. The application of this method to the small protein phoratoxin[12] is illustrated in Figure 28. Inspection of the short-range NOE data immediately enables one to identify two helices and five turns. This method of secondary structure determination has been widely applied[160,170-186]for a range of small proteins.

In assessing the accuracy of secondary structure elements deduced using this approach, several factors should be borne in mind. Essentially, it is a data-based approach insofar that the expected patterns of short-range NOE connectivities for different secondary structure elements have been derived by examining the values of all the short-range distances involving the NH, $C^{\alpha}H$, and $C^{\beta}H$ protons in regular secondary structure elements present in protein X-ray structures. Thus, it tends to perform relatively poorly in regions of irregular structure such as loops. In addition, the exact start and end of helices tend to be rather ill-defined, particularly as the pattern of NOEs for turns is not all too dissimilar from that present in helices. Thus, a turn at the end of a helix could be misinterpreted as still being part of the helix. In the case of β-sheets, the definition of the start and end is more accurate as the alignment is accomplished from the interstrand NOEs involving the NH and $C^{\alpha}H$ protons.

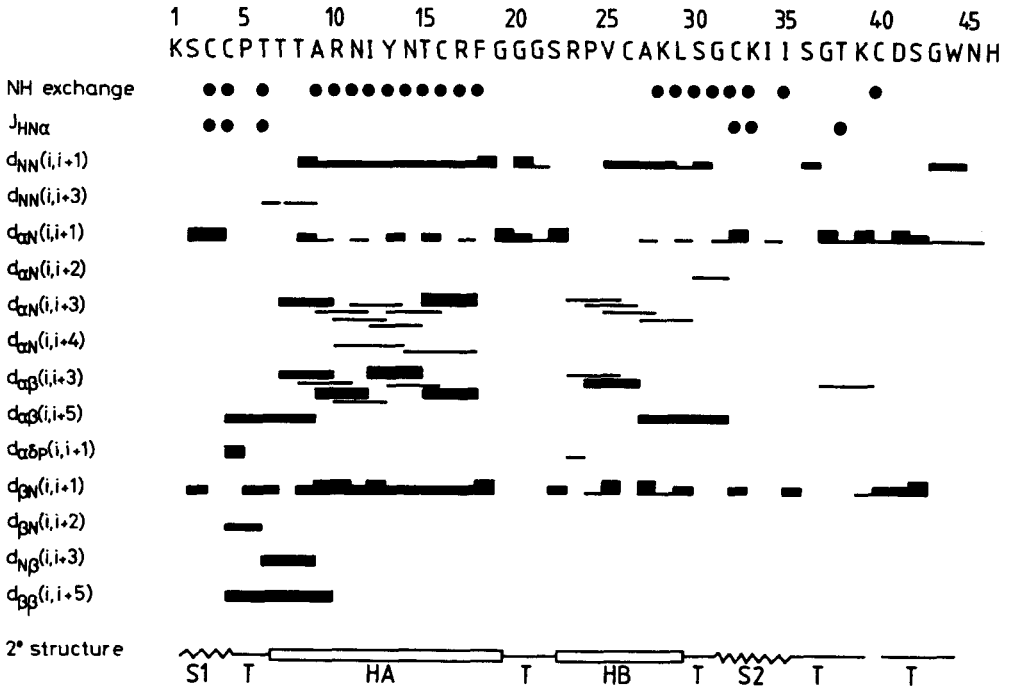


FIGURE 28. Pattern of short-range NOEs involving the NH, $C^{\alpha}H$, and $C^{\beta}H$ protons, as well as the $C^{\delta}H$ protons, for phoratoxin. The NOEs are classified as strong, medium, and weak according to the thickness of the lines. In addition, slowly exchanging backbone amide protons and residues with apparent values of $^{3}J_{HN\alpha} > 9$ Hz are indicated. The secondary structure deduced from this data is shown at the bottom of the figure. (From Clore, G. M., Sukumaran, D. K., Nilges, M., and Gronenborn, A. M., *Biochemistry*, 26, 1732, 1987. With permission.)

VII. ASSIGNMENT OF LONG-RANGE (|i-j|> 5) NOES IN PROTEINS

While oligonucleotides are essentially linear such that interresidue NOEs are only observed between protons of adjacent base pairs, proteins are generally folded into a tertiary structure such that protons far apart in the sequence may be close together in space. These protons give rise to tertiary NOEs that are essential for determining the polypeptide fold. Once complete assignments have been made, many such long-range NOEs can be identified in a straightforward manner. It is usually the case, however, that the assignment of a number of long-range NOE cross-peaks remain ambiguous due to resonance overlap. In some cases, this ambiguity can be removed by recording a 2D or 3D ^{15}N-^{1}H HMQC-NOESY experiment[76,84,85,184] or a 3D HOHAHA-NOESY experiment.[82,83] Where ambiguities still remain, it is often possible to resolve them by deriving a low-resolution structure on the basis of the available data (i.e., the secondary structure and the assignment of a subset of all the long-range NOEs) either by model building or by distance geometry calculations. This low-resolution structure can then be used to test possible assignments of certain long-range NOEs.

VIII. TERTIARY STRUCTURE DETERMINATION

A number of different approaches can be used to determine the 3D structure of a macromolecule from experimental NMR data.

The simplest approach, at least conceptually, is model building. This can be carried out either with real models or by means of interactive molecular graphics. It suffers, however, from the disadvantage that no unbiased measure of the size of the conformational space consistent with

the NMR data can be obtained. Consequently, there is no guarantee that the structure that has been built is the only one consistent with the experimental data. Further, it is incapable of providing anything more than a very low-resolution structure. Nevertheless, model building can play an important role in the early stages of a structure determination, particularly with respect to resolving ambiguities in the assignments of some of the long-range NOEs.

Another approach that may be useful to generate low-resolution structures, as well as to identify secondary structure elements in a reliable manner, is based on a method that combines the NOE data with known substructures taken from the crystallogaphic protein data bank.[187] The underlying philosophy behind this approach is the notion that most local structural features are already well represented in existing protein structures solved by X-ray crystallography to date. The method relies only on short-range NOEs involving the NH, $C^{\alpha}H$, and $C^{\beta}H$ protons. The data base is composed of a series of distance matrices (comprising nine combinations of distances involving the NH, $C^{\alpha}H$, and $C^{\beta}H$ protons) generated from X-ray structures in the Brookhaven protein data bank. A zone of five to eight residues in length in the sequence of the protein under investigation is chosen and a set of mini-distance matrices is generated from the NOE data. These mini-matrices are then simultaneously slid along the diagonals of the precalculated distance matrices in the data base until a best fit(s) is obtained and the coordinates of the associated known fragment(s) are extracted and saved. This procedure is repeated for overlapping fragments of the sequence and the entire protein structure is then built by superimposing sequentially overlapping parts. The result is a polyalanine representation of the protein. Because the method does not make use of any long-range NOEs, it can be applied at an early stage of the investigation. Long-range NOEs that have been unambiguously assigned can then be used to confirm the general correctness of the global fold. In addition, the structure(s) can be used as an aid to resolving ambiguities in the assignment of other long-range NOEs. It should be pointed out, however, that because this is a data base approach it only performs well in regions of regular secondary structure.

A slightly more sophisticated approach, based on similar lines, has been developed by Litharge et al.[188] The first step involves using the short-range NOEs to derive a secondary structure. This is followed by a systematic search involving rigid body movement to obtain the relative orientations of the secondary structure units on the basis of long-range NOE data.

In contrast to the above approaches, the main methods used to generate structures from NMR data do not involve a data base, and, consequently, permit a completely unbiased determination of 3D structure. The common feature of these methods is a conformational search to locate the global minimum of a target function that is made up of stereochemical and experimental NMR restraints. The descent to the global minimum region is not a simple straightfoward path, as the target function is characterized by many false local minima that have to be avoided or surmounted by all the methods.

There are essentially two general classes of methods. The first can be termed real space methods. These include restrained least squares minimization in torsion angle space with either a variable target function[41-43] or a sequence of ellipsoids of constantly decreasing volume, each of which contains the minimum of the target function,[189] and restrained molecular dynamics[5,6,44,45,190] and dynamical simulated annealing[16,22,23] in cartesian coordinate space. All real space methods require initial structures. These can be (1) random structures with correct covalent geometry; (2) structures that are very far from the final structure (e.g., a completely extended strand); (3) structures made up of a completely random array of atoms; and (4) structures generated by distance space methods. They should not, however, comprise structures derived by model building as this inevitably biases the final outcome. Because these methods operate in real space, great care generally has to be taken to ensure that incorrect folding of the polypeptide chain does not occur. A new real space approach involving the use of dynamical simulated annealing, however, has succeeded in circumventing this problem.[23] In contrast to the real space methods, the folding problem does not exist in the second class of methods which

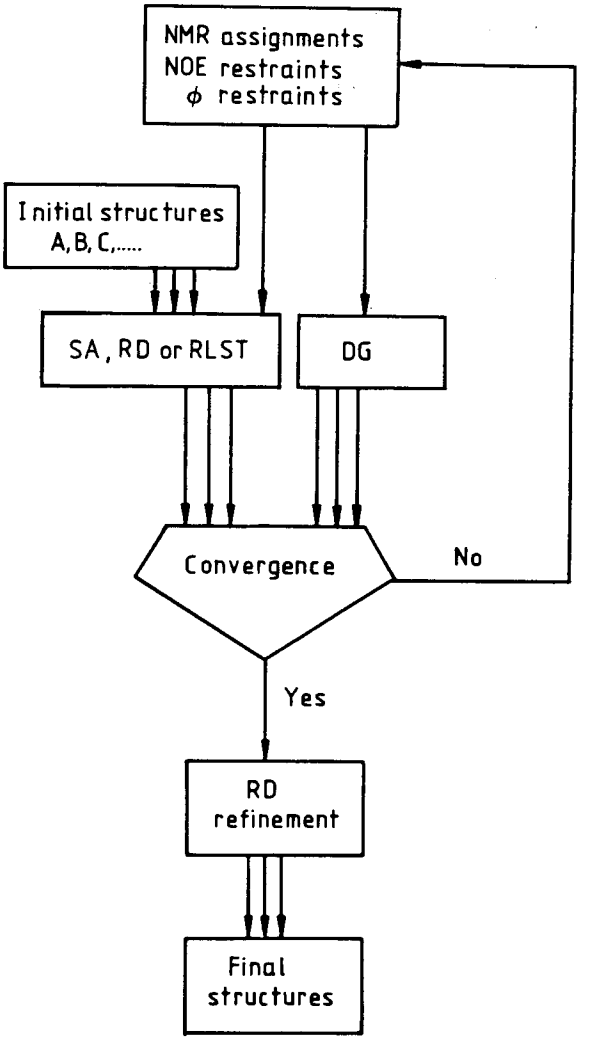


FIGURE 29. Calculational strategy used to solve 3D structures of macromolecules on the basis of NMR data.

operates in distance space and is generally referred to as metric matrix distance geometry.[191-199] Here the coordinates of the calculated structure are generated by a projection from N(N-1)/2 dimensional distance space (where N is the number of atoms) into 3D cartesian coordinate space by a procedure known as embedding.

A flow chart of the calculational strategy that is generally used to solve protein structures is shown in Figure 29. All the methods are comparable in convergence power. In general, however, the structures generated by dynamical simulated annealing or refined by restrained molecular dynamics tend to be better in energetic terms than the structures generated by the other methods, particularly with respect to nonbonded contacts and agreement with the experimental NMR data. A brief description of these various methods is given below.

A. METRIC MATRIX DISTANCE GEOMETRY

A detailed review of metric matrix distance geometry has appeared,[195] so only the basic principles are summarized here. The projection from high-dimensional distance space to 3D coordinate space is generally carried out in four phases.[199] In phase 1, a complete set of bounds

on the distances between all atoms of the molecules is determined by triangulation from the experimental restraints and from the distance and planarity restraints derived from the primary structure. The latter consist of assumed exact distances between all covalently bonded and geminal pairs of atoms, as well as lower limits on the distances between all pairs of atoms more than three bonds apart that are assumed to be no smaller than the sum of the hard sphere van der Waals radii. The complete set of distance bounds is then subjected to a procedure known as bounds smoothing, which involves the selection of the smallest possible intervals between lower and upper bounds consistent with the triangle inequalities. Phase 1 takes a few minutes of CPU time and has to be performed only once for a given set of experimental data. In phase 2, a subset of atoms is embedded (e.g., N, C, C^{α}, $C^{\alpha}H$, C^{β}, and nonterminal C^{γ} and C^{δ} atoms, as well as a pseudoatom for the aromatic rings of Tyr and Phe, in the case of proteins). This phase takes only a few minutes of CPU time per substructure, as it does not require the time-consuming checking of triangle inequalities. The resulting substructures have approximately the correct polypeptide fold or its global mirror image. (These are suitable starting structures for dynamical simulated annealing calculations, see Section VIII.D.) The substructures are then subjected to conjugate gradient minimization similar to that in phase 4 below. The distances between the atoms in the substructure are relaxed and included as additional distance restraints for the embedding of all atoms in phase 3. As bounds smoothing is performed for every single distance chosen within its allowed limit, the third phase is very time consuming. Phase 3 is then usually followed by conventional conjugate gradient minimization of a target function that contains terms for bond lengths, angles, planes, chirality, and NMR restraints (phase 4).

In mathematical terms, the metric matrix distance geometry approach may be formulated as follows. The distance of each point from the center of mass is calculated using Lagrange's theorem:

$$d_{io}^2 = \frac{1}{N}\sum_{j=1}^{N} d_{ij}^2 - \frac{1}{N^2}\sum_{j=2}^{N}\sum_{k=1}^{j-1} d_{jk}^2 \tag{12}$$

The metric matrix $[G_{ij}]$ is then formed where $G_{ij} = (d_{io}^2 + d_{jo}^2 - d_{ij}^2)/2$. The elements of the metric matrix are all the scalar products

$$\mathbf{d}_{io} \cdot \mathbf{d}_{jo} = |d_{io}||d_{jo}|\cos\phi_{ioj} \tag{13}$$

where

$$d_{ij}^2 = d_{io}^2 + d_{jo}^2 - 2|d_{io}||d_{jo}|\cos\phi_{ioj} \tag{14}$$

and

$$|d_{io}||d_{jo}|\cos\phi_{ioj} = d_{io}^2 + d_{jo}^2 - d_{ij}^2 \tag{15}$$

The eigenvalues of $[G_{ij}]$ are found, the three largest eigenvalues, λ_1, λ_2, λ_3, taken, and the three corresponding eigenvectors, given as the three columns of the $N \times 3$ matrix $[W_{ij}]$, determined. The coordinates x_{ij} (points i+1 to N, axes j = 1 to 3) are then calculated from $x_{ij} = \lambda_j^{1/2} W_{ij}$. If by chance λ_j is less than zero due to errors, the absolute value of λ_j, $|\lambda_j|$ is used instead of λ_j. In general, the method is reasonably robust and can sustain relatively large inaccuracies in the values of d_{ij}.

One problem associated with the metric matric distance geometry method is that distances alone cannot define the chirality of the structure so that mirror images (local or global) of the correct structure can occur. In general, these can easily be rejected at the substructure stage as the chirality of single amino acids (L) and helices (right handed) is known. A further problem

associated with the method is that it does not sample efficiently the conformational space consistent with the stereochemical and experimental restraints. As a result, the atomic rms distribution of a series of calculated structures tends to be underestimated, particularly in regions that are poorly defined by the experimental data. Subjecting these structures to restrained molecular dynamics, for example, results in an increase in the atomic rms distribution while improving the agreement with the experimental data and reducing the van der Waals energy.[12,13,15,21,24,46]

B. MINIMIZATION IN TORSION ANGLE SPACE

In methods involving minimization in torsion angle space,[41-43,189] bond lengths and angles are kept fixed during the minimization and only torsion angles are varied. To ensure that correct folding occurs, the restraints have to be introduced gradually during the calculation. This may be achieved using a variable target function[41] or a progressive build-up procedure[42,43] so that the distances between residues further and further apart in the sequence are incorporated in successive cycles of minimization. An alternative procedure is to use the ellipsoid minimization algorithm that makes use of a sequence of ellipsoids of constantly decreasing volume, each of which contains the minimum of the target function.[189] In general, the structures found by these methods tend to have quite unrealistically high energies that can be subsequently reduced by energy minimization with only minimal atomic rms shifts.[200] For a detailed description of these two methods, the reader is referred to the review of Braun.[200]

C. RESTRAINED MOLECULAR DYNAMICS

In contrast to the other methods, restrained molecular dynamics[5,6,44,45,190] makes use of a full empirical energy function (E_{tot}) comprising both bonded and nonbonded interactions, supplemented by an effective potential term(s) representing the experimental NMR data. The nonbonded interactions are made up of van der Waals, electrostatic, and hydrogen bonding interactions. The basis of the method involves the simultaneous integration of Newton's equations of motion:[201-206]

$$\frac{\partial^2 \mathbf{x}_i}{\partial t^2} = -\frac{1}{m_i}\partial E_{tot}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots) \tag{16}$$

for all n atoms in the system whose temperature is given by

$$(\text{Temp})_t = \frac{2}{k_B(3n-6)}\sum_{i=1}^{n} m_i v_i^2/2 \tag{17}$$

Interproton distance constraints may be incorporated as biharmonic effective potentials[6,190] when distances can be reasonably accurately determined, as in the case of DNA:

$$E_{NOE}(r_{ij}) = \begin{cases} c_1(r_{ij} - r_{ij}^o)^2 & \text{if} \quad r_{ij} > r_{ij}^o \\ \\ c_2(r_{ij} - r_{ij}^o)^2 & \text{if} \quad r_{ij} < r_{ij}^o \end{cases} \tag{18}$$

where r_{ij} and r_{ij}^o are the calculated and experimental interproton distances, respectively, and c_1 and c_2 are force constants given by

$$c_1 = k_B TS/[2(\Delta_{ij}^+)^2] \quad \text{and} \quad c_2 = k_B TS/[2(\Delta_{ij}^-)^2] \tag{19}$$

where k_B is the Boltzman constant; T, the absolute temperature; S, a scale factor, and Δ_{ij}^+ and Δ_{ij}^- are the positive and negative error estimates, respectively, of r_{ij}. If, on the other hand, only ranges of distances can be specified, as in the case of proteins, the interproton distance restraints are incorporated in the form of square-well potentials of the form[7]

$$E_{NOE} = \begin{cases} k_{NOE}(r_{ij} - r_{ij}^u)^2 & \text{if} \quad r_{ij} > r_{ij}^u \\ 0 & \text{if} \quad r_{ij}^l \leq r_{ij} \leq r_{ij}^u \\ k_{NOE}(r_{ij} - r_{ij}^l)^2 & \text{if} \quad r_{ij} < r_{ij}^l \end{cases} \tag{20}$$

where r_{ij}^u and r_{ij}^l are the values of the upper and lower limits of the target distances, respectively, and k_{NOE} is the NOE force constant. Similarly, torsion angle restraints are also incorporated in the form of square-well potentials of the form[7]

$$E_{tor} = \begin{cases} k_{tor}(\varphi_i - \varphi_i^u)^2 & \text{if} \quad \varphi_i > \varphi_i^u \\ 0 & \text{if} \quad \varphi_i^l \leq \varphi_i \leq \varphi_i^u \\ k_{tor}(\varphi_i - \varphi_i^l)^2 & \text{if} \quad \varphi_i > \varphi_i^l \end{cases} \tag{21}$$

where φ_i^u and φ_i^l are the upper and lower limits of the target range of a particular torsion angle, and φ_i is its calculated value.

To ensure that the experimental restraints are the dominating factor in determining the conformation of the molecule under consideration, it is very important that the force constants for the restraints are set sufficiently high that the experimental data is satisfied within the precision of the measurements. At the same time, the contribution from the empirical energy function should be such that for any *individual* restrained dynamics structures, the deviations from idealized covalent geometry are small and the nonbonded contacts are good, as evidenced by negative values for the Lennard-Jones van der Waals energy.[12,13,15,21,24,44,45] This can easily be assessed empirically on the basis of a few trial calculations.

Although it is possible to fold an extended strand into a correct 3D structure by restrained molecular dynamics,[44,45] this procedure is not the most efficient and is somewhat inflexible, owing to the overheads generated by the calculation of a full nonbonded interaction potential comprising van der Waals, electrostatic, and hydrogen-bonding terms. Consequently, restrained molecular dynamics is best applied as a tool of refinement. This may, for example, involve the refinement of protein structures generated by metric matrix distance geometry that invariably exhibit rather poor nonbonded contacts and, in general, do not satisfy the experimental data that well.[7,12,13,15,46] Another example involves the refinement of oligonucleotide structures. The handedness of the DNA helix is easily ascertained on the basis of a qualitative interpretation of the NOE data, but the global conformation is not. Thus, in the case of right-handed DNA, restrained molecular dynamics is an ideal method for generating refined structures starting from both A and B forms of DNA. In all cases where this has been applied, the method is easily capable of converting an A-DNA starting structure to B-type DNA, even though large atomic rms shifts are involved.[131-135,190] For example, in the case of a DNA dodecamer, the atomic rms difference between A and B forms of DNA is around 7 Å.

D. DYNAMICAL SIMULATED ANNEALING

Dynamical simulated annealing is similar in spirit to restrained molecular dynamics in that it involves solving Newton's equations of motion for all atoms of the system for a suitable time at a temperature that is directly proportional to the kinetic energy of the system.[16,22,23] It differs,

however, from restrained molecular dynamics in one very important respect, namely, all the nonbonded interaction terms of the usual empirical energy function are replaced by a simple van der Waals repulsion term. This not only speeds up the calculations, but also allows considerably more flexibility in the protocols that can be used.

The basis of simulated annealing involves raising the temperature of the system followed by slow cooling in order to overcome local minima and locate the global minimum region of the target function. In the original description of simulated annealing,[207,208] the Metropolis algorithm[209] was used to simulate a system at a temperature T. In the case of macromolecules, it is simpler to do this by solving Newton's equations of motion. Additionally, it is often more efficient to vary the force constants for the various terms in the target function rather than the temperature for two reasons. First, no variable time step integrator is required if the calculations are done at constant temperature. Second, varying the force constants at differential rates adds further flexibility to the method. For a potential that is harmonic in the coordinates, scaling of a force constant by a factor c corresponds to simultaneously scaling the temperature by $1/c$ and the time by $1/\sqrt{c}$.

The total target function F_{tot} for which the global minimum region is searched comprises the following terms:

$$F_{tot} = F_{covalent} + F_{repel} + F_{expt} \tag{22}$$

F_{expt} is represented by the square-well potentials for the NOE (Equation 20) and torsion angle (Equation 21) restraints derived from the NMR measurements.

$F_{covalent}$ maintains correct bond lengths, angles, chirality, and planes, and is given by

$$F_{covalent} = \sum_{bonds} k_b(r - r_o)^2 + \sum_{angles} k_\phi(\phi - \phi_o)^2 + \sum_{impropers} k_\varphi(\varphi - \varphi_o)^2 \tag{23}$$

The force constants of the energy terms for bonds, angles, and improper torsion angles (which serve to maintain planarity and chirality) are set to uniform values. All dihedral angle force constants are set to 0 as the dihedral potential at rotatable bonds is effectively a nonbonded interaction.

The nonbonded interactions are represented by a simple van der Waals repulsion term with a variable force constant k_{rep}:

$$E_{repel} = \begin{cases} 0 & \text{if} \quad r \geq s.r_{min} \\ k_{rep}(s^2 r_{min}^2 - r^2)^2 & \text{if} \quad r < s.r_{min} \end{cases} \tag{24}$$

The values of r_{min} are given by the sum of the standard values of the van der Waals radii between two atoms as represented by the Lennard-Jones potential used in the CHARMM empirical energy function.[210] A van der Waals radius scale factor, s, should typically be set to 0.8 to account for the fact that interatomic separations slightly smaller than the sum of the hard-sphere van der Waals radii can easily occur due to the attractive component of the van der Waals interaction.

One of the advantages of dynamical simulated annealing is its generality and ease of implementation. The design of an annealing schedule for practical purposes, however, is still something of a black art.[208] Hence, the three strategies that have been developed for the application of simulated annealing to protein structure determination from NMR data are described in detail.

1. Folding an Extended Strand by Dynamical Simulated Annealing

In the first approach, simulated annealing is used to fold a structure from an extended strand.[22] This involves using a soft asymptotic NOE potential and varying the NOE target function so that interproton distances between protons far apart in the sequence are gradually incorporated into the calculation. This is similar in spirit to the variable target function method of Braun and Go.[41] To achieve this, the NOE target function F_{NOE} is a complex term made up of three terms, F_{long}, F_{short}, and F_{final}, whose functional form depends on the difference between the calculated (r_{ij}) and target values of a particular interproton distance, as well as on the value of the variable force constant for the F_{short} term (see below).

A simplified flow chart of the calculational strategy is shown in Figure 30. Its design is based on three guidelines: (1) that secondary structure elements should be formed prior to tertiary structure folding; (2) that converged NOE restraints and/or local structure elements once formed should be retained; and (3) that the incorporation of NOE restraints into the NOE target function F_{NOE} should be completely automatic. To achieve these aims, the NOE violations are analyzed and grouped into different classes, depending on their size (see description below), at the end of each cycle of simulated annealing, and a special potential form is used that places greater weight on smaller violations.

The calculations have to start from unfolded structures (e.g., an extended strand) rather than from entirely random structures that may already be folded. The reason for this is that once the polypeptide chain has folded incorrectly, it can no longer converge to the correct global minimum region with this particular protocol. The random number seed, however, used for the assignment of the initial velocities is sufficient to ensure good sampling of the available conformational space consistent with the interproton distance data.

The NOE restraints are initially classified into two classes, *long* and *short*, depending on the difference ($viol_{ij}^{u}$) between the calculated (r_{ij}) and upper limit of the target (r_{ij}^{u}) distances. Class *long* contains NOE restraints that are violated by more than $rdist_{LS}$, and class *short* contains all the others. In class *long*, the target function is switched off (i.e., $F_{long} = 0$). In class *short*, the target function has the following functional form:

$$F_{short} = \begin{cases} k_s\left(c.viol_{ij}^{u} + \dfrac{b}{viol_{ij}^{u}} + a\right) & \text{if} \quad viol_{ij}^{u} > rswitch \\ k_s(viol_{ij}^{u})^4 & \text{if} \quad r_{ij} > r_{ij}^{u} \text{ and } viol_{ij}^{u} \leq rswitch \\ 0 & \text{if} \quad r_{ij}^{l} \leq r_{ij} \leq r_{ij}^{u} \\ k_s(viol_{ij}^{l})^4 & \text{if} \quad r_{ij} < r_{ij}^{l} \end{cases} \tag{25}$$

The values of a and b are chosen so that F_{short} is continuous and differentiable at *rswitch*. The slope of the assymptote, c, is set to 0 in the present calculations. The initial values chosen for *rswitch*, $rdist_{LS}$, and the force constant k_s are 3 Å, 10 Å, and 0.05 kcal.mol^{-1}.Å^{-2}, respectively.

A diagram of the functional form of F_{short} is shown in Figure 31. This potential form is designed to ensure that secondary structure elements, defined by interproton distances between residues close together in the sequence, are formed prior to the incorporation of NOEs between residues far apart in the sequence. The gradient of F_{short} is largest at $rswitch + r_{ij}^{u}$ so that NOEs that are violated by about the value of *rswitch* experience the largest force. At the beginning of the simulation, *rswitch* is set to a low value (~3 Å); by progressively increasing its value, the maximum of the driving force is shifted to larger violations. Thus, once the formation of local secondary structures such as α-helices has occurred, turns can be formed and tertiary structure folding can gradually take place. For the same reason, NOEs that are violated by more than $rdist_{SL}$ are placed in class *long*, where they experience no force from the NOE restraints at all.

At the beginning of the calculations, the hard sphere radii of the atoms are chosen as in the

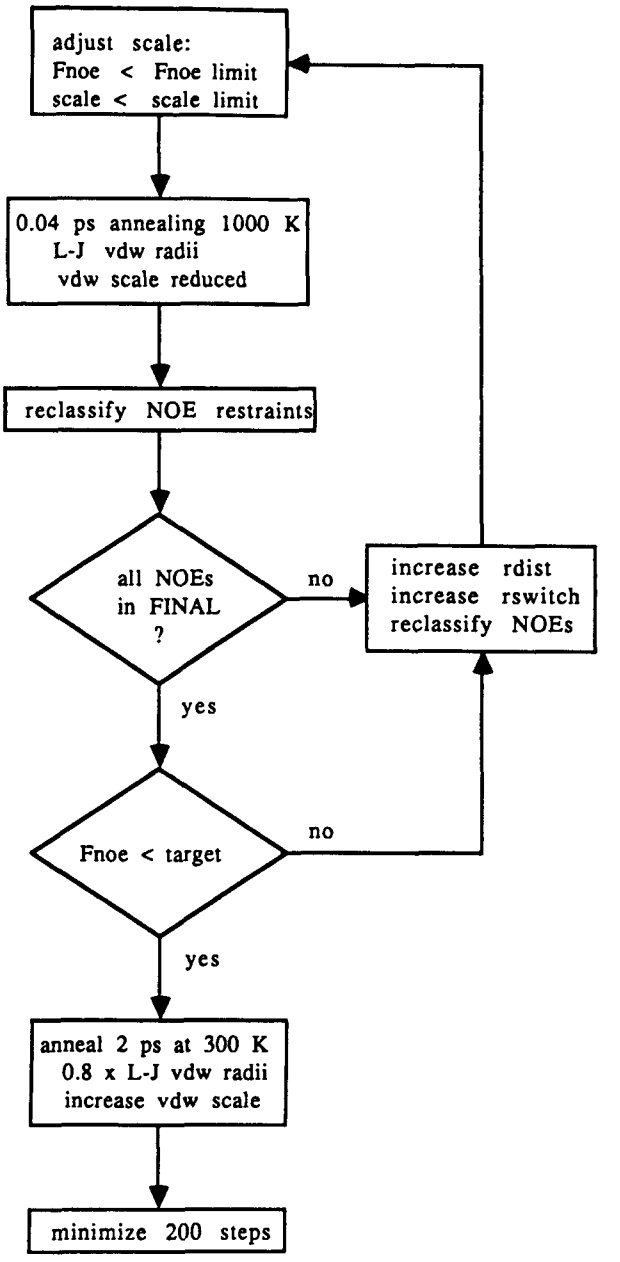


FIGURE 30. Flowchart of the dynamical simulated annealing protocol used to fold structures from an extended strand. (From Nilges, M., Gronenborn, A. M., Brünger, A. T., and Clore, G. M., *Protein Eng.*, 2, 27, 1988. With permission.)

Lennard-Jones potential (i.e., s in Equation 24 is set to 1.0). The calculations are initiated with 20 steps of Powell minimization to remove some bad nonbonded contacts. This is followed by *Phase 1* of the simulated annealing protocol. The initial velocities at t = 0 ps are chosen from a Maxwellian distribution at 1000 K. This temperature is chosen to ensure that local minima along the convergence pathway toward the global minimum region of the target function F_{tot} can be overcome. Each cycle of annealing comprises 40 steps with a time step of 1 fs in which the

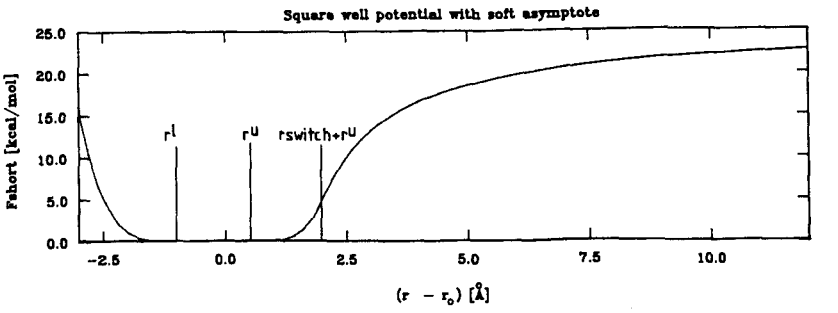


FIGURE 31. Functional form of the NOE potential F_{short} used in the dynamical simulated annealing protocol to fold structures from an extended strand. (From Nilges, M., Gronenborn, A. M., Brünger, A. T., and Clore, G. M., *Protein Eng.*, 2, 27, 1988. With permission.)

nonbonded contact list is updated every 8 steps, and the velocities are rescaled to 1000 K every 20 steps. After every cycle of annealing the force constant k_s for F_{short} is increased up to a maximum value of 15.8 kcal.mol^{-1}.Å^{-2} by multiplying its value by $10^{0.1}$. The value of F_{NOE} is then evaluated with the new value of k_s, and if $F_{NOE} > 2500$/kcal.mol^{-1}, k_s is divided by $10^{0.1}$ until $F_{NOE} < 2500$ kcal.mol^{-1}. Once k_s has reached its maximum value of 15.8 kcal.mol^{-1}.Å^{-2}, the NOE restraints are further reclassified between class *short* and class *final* if $viol_{ij}^{u}$ is less than $rdist_{FS}$. The value of $rdist_{FS}$ chosen is 1.0 Å, and F_{final} is a square well-function with a force constant k_f of 60 kcal.mol^{-1}.Å^{-2}, given by Equation 20. Thus, class *final* contains all the NOE restraints that have converged. In addition, the restraints in the *long* and *short* classes are counted and the smallest violation in class *long* is calculated. To proceed further, three cases are distinguished:

1. If both the *long* and *short* classes are empty and F_{NOE} has a value below its target value (in this case 120 kcal.mol^{-1}), global as well as local convergence has occurred, the *Phase 1* calculation is stopped, and the simulation proceeds directly to *Phase 2* of the annealing protocol.
2. If only the *short* class is empty and the value of F_{NOE} lies below its target value, "local convergence" has been achieved, and the value of $rdist_{LS}$ for the reclassification between classes *long* and *short* is set to just above (0.02 Å) the smallest violation in class *long*, so that one or only a few additional violations go into class *short*.
3. If neither case 1 nor case 2 applies, then $rdist_{LS}$ is increased by 0.02 Å.

Additionally, in cases (2) and (3), the value of *rswitch* is increased by 0.01 Å and the NOE restraints are reclassified between classes *long* and *short*.

The rationale behind the grouping of all "converged NOE restraints" in class *final* is to ensure that once secondary structure elements have formed, they are preserved and not disrupted at a later stage during the course of the simulation. This is required since there is no force other than the NOE restraints to stabilize such secondary structure elements, and the scale of the *short* potential has to be reduced drastically at times as longer-range NOEs are incorporated into F_{short}. For this reason, the force constant for the *final* potential is never reduced. To ensure that the reclassification only takes place once NOEs have really converged and to allow some rearrangement of the local structure, NOEs are only reclassified between the *final* and *short* classes when the force constant for F_{short} is at its maximum value.

The maximum number of cycles for *Phase 1* depends on the size of the protein, the larger the protein, the more cycles are required. Thus, for calculations on crambin (46 residues) 250 cycles were required, while for barley serine proteinase inhibitor-2 (65 residues) 350 cycles were needed. If at this stage there are still violations in class *long*, failure of convergence is presumed

and the calculation comes to a complete halt. If, on the other hand, there are no violations in class *long*, the NOE restraints are once again reclassified between classes *short* and *final* at 10 Å in order to place all the NOEs into the *final* class. This is followed by *Phase* 2 of the annealing protocol that comprises 20 cycles of 100 steps of annealing at 300 K. The velocities are rescaled to 300 K after every cycle, and the force constant k_{rep} for the repulsion target function F_{repel} (cf. Equation 24) is increased in steps of 0.2 from an initial value of 0.4 kcal.mol^{-1}.Å^{-4} to a maximum value of 4 kcal.mol^{-1}.Å^{-4}. The values of the hard-sphere atomic radii are set to 0.8 times their Lennard-Jones values (i.e., s = 0.8 in Equation 24). Finally, *Phase 2* is followed by 200 steps of restrained Powell minimization to complete the simulation.

To date, this method has been applied to carboxypeptidase inhibitor, crambin, and barley serine proteinase inhibitor-2,[22] as well as on the linear helical 36 residue peptide cecropin.[30] The convergence properties are reasonably good, and the computational requirements are similar to metric matrix distance geometry. The atomic rms distributions of the final converged structures is comparable to that sampled by either restrained molecular dynamics or the hybrid distance geometry-dynamical simulated annealing method and slightly larger than that sampled by metric matrix distance geometry. Although all the dynamical simulated annealing calculations performed to date using this protocol have started out from the same initial structure, namely, an extended strand, it must be emphasized that varying the random number seed used in the assignment of the initial velocities ensures that significantly different convergence pathways are followed, so that the different trajectories do *not* posess any common intermediate structures. That is to say that during the initial stages of the simulation, the different trajectories diverge. For example, in the case of calculations carried out on crambin, the maximum average and maximum absolute backbone atomic rms differences are 5.4 and 8.1 Å, respectively. As the simulation proceeds, and more and more NOEs are satisfied, convergence between the different trajectories occurs. The divergence of trajectories, however, cannot occur to the extent that the distribution of the structures between different trajectories would be totally random (with an expected mean backbone atomic rms difference of ~10 Å for a protein the size of crambin). The reason is twofold. First, local convergence, driven by short-range NOEs, starts already at the beginning of the calculations. Second, the structures have a tendency to stay extended in the absence of tertiary folding forces (i.e., long-range NOEs) due to their intrinsic inertia arising from the fact that the masses of the atoms enter explicitly into the calculations (cf. Equation 16). All calculations carried out to date using this method indicate, however, that this procedure does not introduce any bias with respect to the end result, and that the use of different random number seeds for the assignment of initial velocities is equivalent to using a set of randomly chosen initial structures in static real space methods.[41,189]

2. The Hybrid Metric Matrix Distance Geometry-Dynamical Simulated Annealing Method

The second strategy is a hybrid of the metric matrix distance geometry and simulated annealing methods.[16] This is summarized in the flowchart of Figure 32. Substructures that contain only about a third of the atoms and have approximately the correct fold are first generated by projection from *n*-dimensional distance space employing a metric matrix distance geometry algorithm (without checking the triangle inequalities) and are subsequently used as starting structures for the simulated annealing calculation. By this means, the folding problem inherent in real space methods is avoided, and the concept of simulated annealing is mainly used to improve the conformation of the substructure, which may be accompanied by quite large atomic rms shifts (up to 5 Å). Typically, amino acids in an extended conformation ($\phi = -120°$, $\psi = 120°$, and $\chi_i = 180°$) are best fitted to the substructures residue by residue. The force constants for the bond and angular terms in $F_{covalent}$ are set to uniformly high values of 500 kcal.mol^{-1}.Å^{-2} and 500 kcal.mol^{-1}.rad^{-2}, respectively. A few cycles of unrestrained (i.e., no experimental potential) Powell minimization with a very low force constant (k_{vdW} = 0.01 kcal.mol^{-1}.Å^{-4}) on the van der

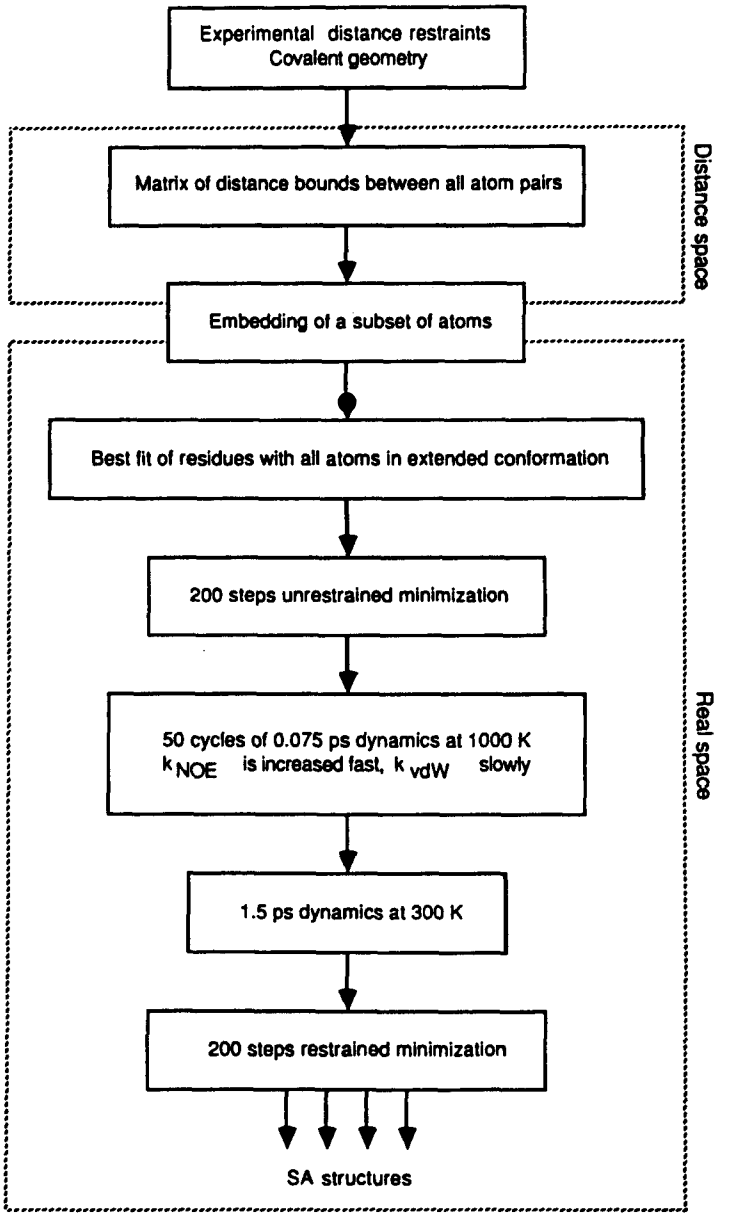


FIGURE 32. Flowchart of the hybrid metric matrix distance geometry-dynamical simulated annealing method. (From Nilges, M., Gronenborn, A. M., and Clore, G. M., *FEBS Lett.*, 229, 317, 1988. With permission.)

Waals repulsion term are then carried out to improve the covalent geometry prior to dynamical simulated annealing. The annealing schedule is carried out in two steps. Step 1 comprises 50 cycles of 75 fs dynamics each. The initial velocities are chosen from a Maxwell distribution at 1000 K. During the first few cycles, the value of the force constant k_{NOE} is doubled at the beginning of each new cycle from an initial value of 0.1 to a maximum value of 50 kcal.mol^{-1}.Å^{-2}. If there are torsion angle restraints, the same procedure is applied to k_{tor}. To make rearrangements possible, the initial value of k_{vdW} is very low (0.01 kcal.mol^{-1}.Å^{-4}). This is increased slowly by multiplying its value by a factor of $40^{1/50}$ prior to each cycle up to a final value of 4 kcal.mol^{-}

$Å^{-4}$. These relative values of the force constants are somewhat arbitrary. They do, however, maintain nearly perfect covalent geometry, ensuring that no unduly close contacts occur and that the NOE restraints are effectively introduced and satisfied. Step 2 of the simulated annealing schedule consists of 1.5 ps dynamics at 300 K with values of the force constants equal to their final values at the end of step 1. This is followed by 200 cycles of restrained Powell minimization.

In terms of computational speed, it is probable that the hybrid distance geometry-dynamical simulated annealing approach is the fastest method available at the present time. Comparative studies indicate that for a protein of approximately 80 residues, the hybrid method is double the speed of either full metric matrix distance geometry calculations or of minimization in torsion angle space with a variable target function, and approximately 5 to 10 times the speed of restrained molecular dynamics. Typical total computational times on a VAX8550 for the hybrid method are 45 min per structure for crambin (46 residues) and 3 h per structure for the globular domain of histone H5 (79 residues).

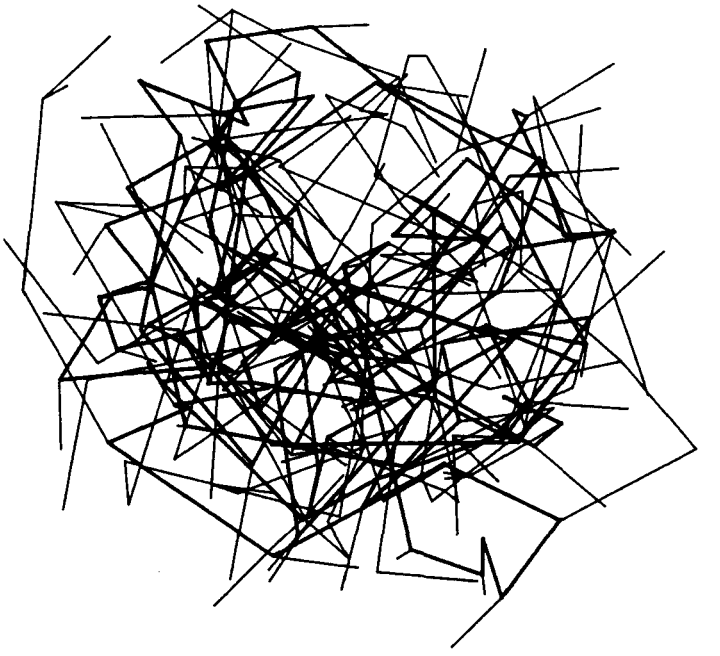
3. Dynamical Simulated Annealing Starting from a Random Array of Atoms

The third and final strategy involves an essentially new real space approach that circumvents the folding problem entirely and sets out to obtain correctly folded structures starting out from a completely random array of atoms.[23]

The calculational approach is relatively straightforward. First, a starting conformation with a completely random array of atoms is generated by assigning random values to the x, y, and z coordinates of the atoms according to a Gaussian distribution with a standard deviation of 1.0 Å centered about the coordinate origin. (Note that the exact form of the distribution is of no relevance.) One of the starting conformations for crambin is shown in Figure 33A. The starting structure can be envisaged as a very high-temperature conformation of the system. Very close nonbonded contacts are first removed by a few cycles of Powell minimization with all force constants set to very low values (0.001 $kcal.mol^{-1}.Å^{-2}$ for the bond and NOE terms, 0.001 $kcal.mol^{-1}.Å^{-4}$ for the F_{repel} term, and 0.001 $kcal.mol^{-1}.rad^{-2}$ for the angular terms). The force constants for $F_{covalent}$ (Equation 23) and F_{NOE} (Equation 20) are then set to values such that the initial potential energy F_{tot} (Equation 22) is approximately equal to the kinetic energy at 1000 K. All force constants are set to identical values, except the repulsion force constant k_{rep}, which is set to 0.001 $kcal.mol^{-1}.Å^{-4}$ (cf. Equation 24). Initial velocities are assigned according to a Maxwell distribution at 1000 K, and all masses are set to uniform values (10 a.u.).

The first phase of the calculation consists of ~45 cycles of dynamical simulated annealing, each comprising 1 ps dynamics with a time step of 1 fs at 1000 K. The exact number of cycles depends on the initial values of the force constants (see above). The velocities are rescaled whenever the temperature is lower than 500 or higher than 1250 K. At the beginning of each cycle, the force constants (covalent and NOE) are increased by multiplying them by a factor of 1.25, up to maximum values of 100 $kcal.mol^{-1}.Å^{-2}$ for k_{NOE}; 500 $kcal.mol^{-1}.Å^{-2}$ for k_b; and 500 $kcal.mol^{-1}.rad^{-2}$ for the angle, dihedral, and improper terms. The repulsion force constant k_{rep} is left unchanged until the bond force constant k_b reaches a value of 100 $kcal.mol^{-1}.Å^{-2}$, at which time k_{rep} is increased by multiplying its value by a factor of 1.5 at the beginning of each subsequent cycle, up to a maximum value of 0.25 $kcal.mol^{-1}.Å^{-4}$. Although the value of k_{rep} is very low during the early stages of the protocol, it is sufficient to define and maintain the global structure while allowing atoms to get very close to each other and even move through each other to improve the structure locally.

The path of the calculations is illustrated in Figures 33 and 34 for a typical crambin trajectory. The increase in the bond and repulsion force constants, together with the backbone (C, C^{α}, N, and O) atomic RMS difference vs. the crystal structure of crambin, is shown in Figure 34 as a function of time. Figure 33B shows snapshots of the "trajectory" every 4 ps, and Figure 33C shows the best fit superposition of the final structure with the X-ray structure. Note that the



A

FIGURE 33. Path of the dynamical simulated annealing calculations for crambin starting from a completely random array of atoms. (A) Initial structure comprising a random array of atoms with the backbone and side chain atoms shown as thick and thin lines, respectively; (B) snapshots at 4 ps intervals during the first phase of the dynamical simulated annealing calculation (only the N, C^{α}, and C backbone atoms are shown); and (C) best fit superposition of the backbone (N, C^{α}, C, and O) atoms of a final dynamical simulated annealing structure (thick lines) resulting from the trajectory shown in (B) on the X-ray structure of crambin (thin lines). The calculations are based on 240 approximate interproton distance restraints. (From Nilges, M., Clore, G. M., and Gronenborn, A. M., *FEBS Lett.*, 239, 129, 1988. With permission.)

global features of the structure begin to emerge at remarkably low values of the force constants. The atoms first arrange themselves roughly in their correct global position, prior to the evolution of local structural elements.

Stage 2 of the protocol consists of 20 cycles of 0.1 ps dynamics with the parameters of the van der Waals repulsion term F_{repel} (Equation 24) set to their final values (s = 0.825, k_{rep} = 4 kcal.mol^{-1}.Å^{-4}), and the temperature lowered to 300 K. The NOE force constant is set to 50 kcal.mol^{-1}.Å^{-2}. Stage 2 is followed by 200 cycles of restrained Powell minimization.

The convergence properties of this method are excellent, reflecting the fact that the folding problem generally associated with real space methods is efficiently avoided. Indeed, trial calculations for crambin taking the global mirror image of the crystal structure (i.e., a structure containg only D-amino acids) as the initial conformation converge to within 1.5 Å of the crystal structure. The only term that discerns such a starting conformation from the correct solution is the improper term maintaining chirality at the tetrahedral carbon atoms. Thus, a mirror image of the true solution represents a deep false minimum, and, as the whole structure has to be inverted, very high-energy barriers have to be overcome to reach the correct global fold. The pathway of such a calculation is shown in Figure 35. The key to the success of the method lies in reducing the force constants of all terms in the target function to values such that the barriers between different folds are of the order of the kinetic energy of the system. Thus, because the atoms are initially only weakly coupled, they can move virtually independently of each other during the early stages of the calculation to satisfy the restraints, thereby avoiding any problems associated with folding.

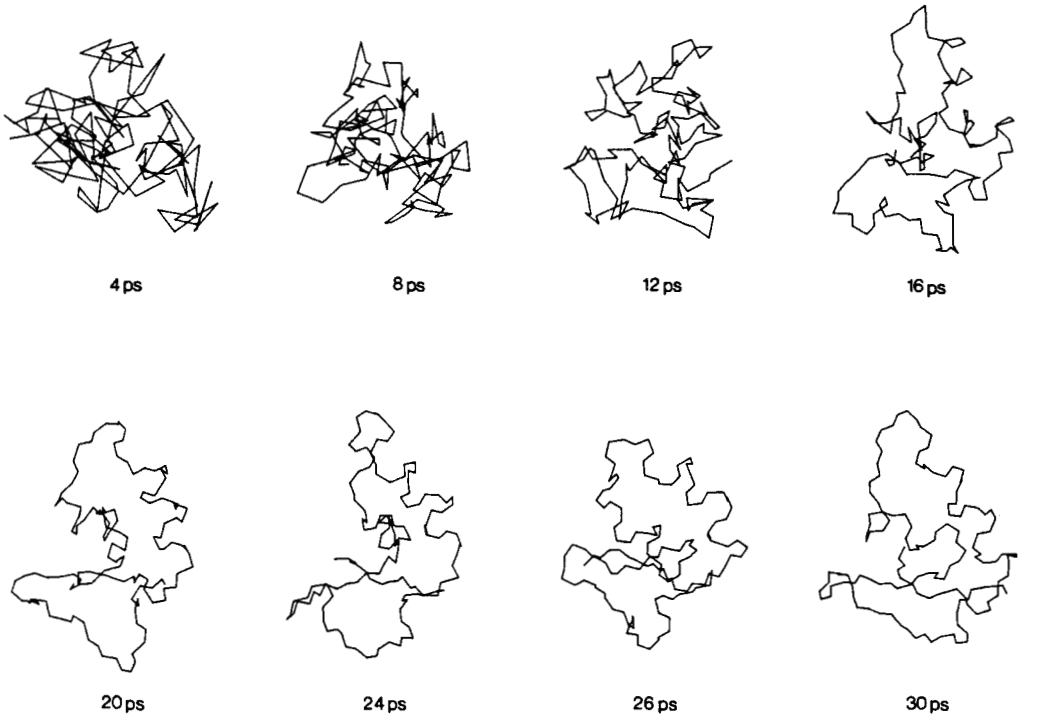


FIGURE 33B

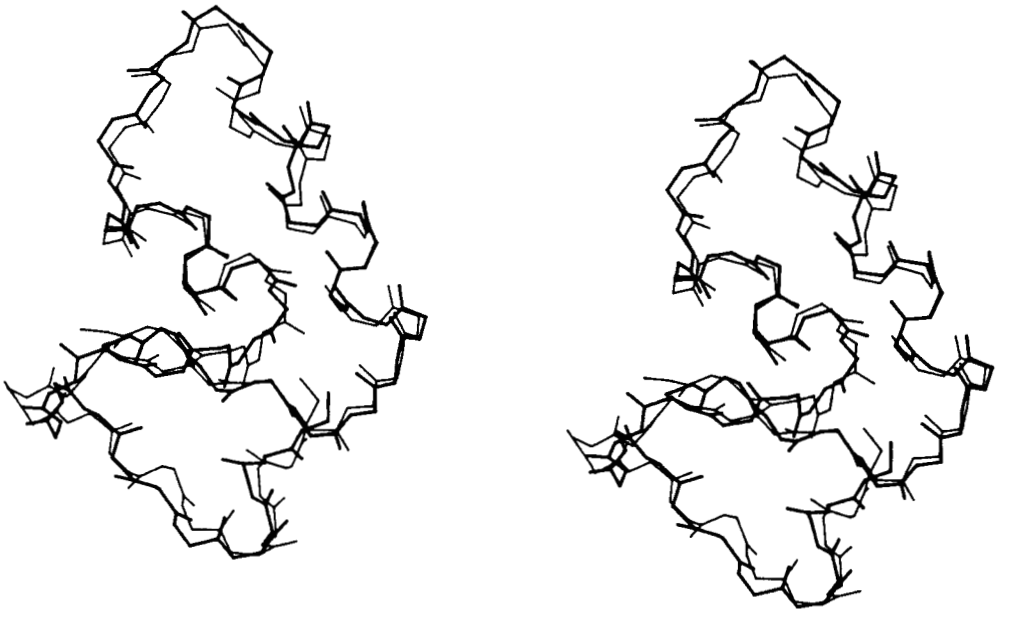


FIGURE 33C

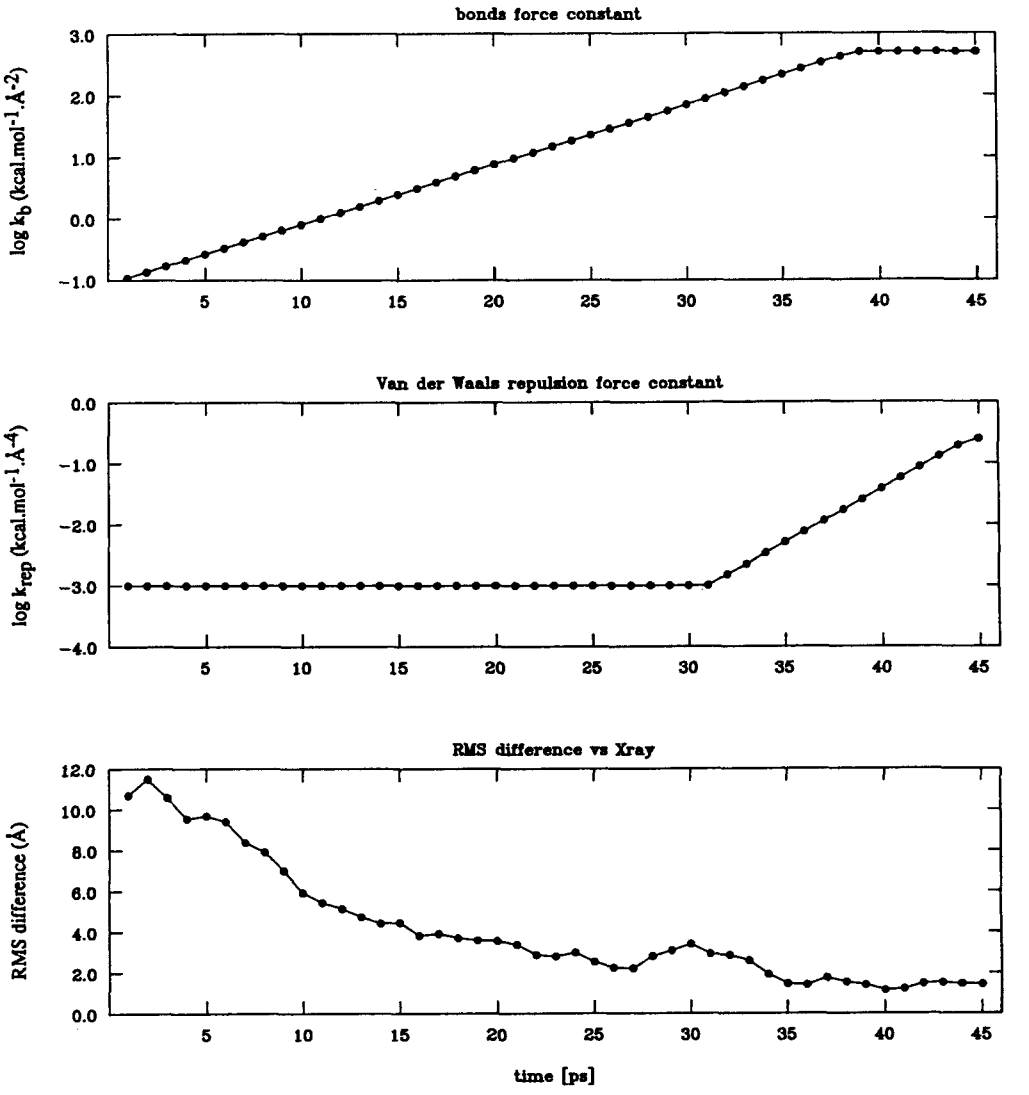


FIGURE 34. Time dependence of the bond force constant k_b, the van der Waals repulsion force constant k_{rep}, and the backbone (N, C^{α}, C, and O) atomic rms difference vs. the crystal structure during the course of the dynamical simulated annealing trajectory shown in Figure 33. (From Nilges, M., Clore, G. M., and Gronenborn, A. M., *FEBS Lett.*, 239, 129, 1988. With permission.)

4. Advantages of Dynamical Simulated Annealing

The implementation of dynamical simulated annealing described here has a number of major advantages over other methods. First, as the nonbonded cutoff radius is only 4 Å, computational times rise only linearly with the number of atoms. As a result, the time requirements for the dynamical simulated annealing method become increasingly favorable with increasing size of the protein, relative to other methods. Second, the quality of the structures is high. That is to say, the method generates structures that satisfy the experimental restraints, have very small deviations from idealized covalent geometry, and have good nonbonded contacts, as judged by negative Lennard-Jones van der Waals energies. Third, the method offers an easy and straightforward approach for dealing with the problem of obtaining stereospecific assignments at prochiral centers, similar to that suggested by Pardi et al[37]. When two separate signals are

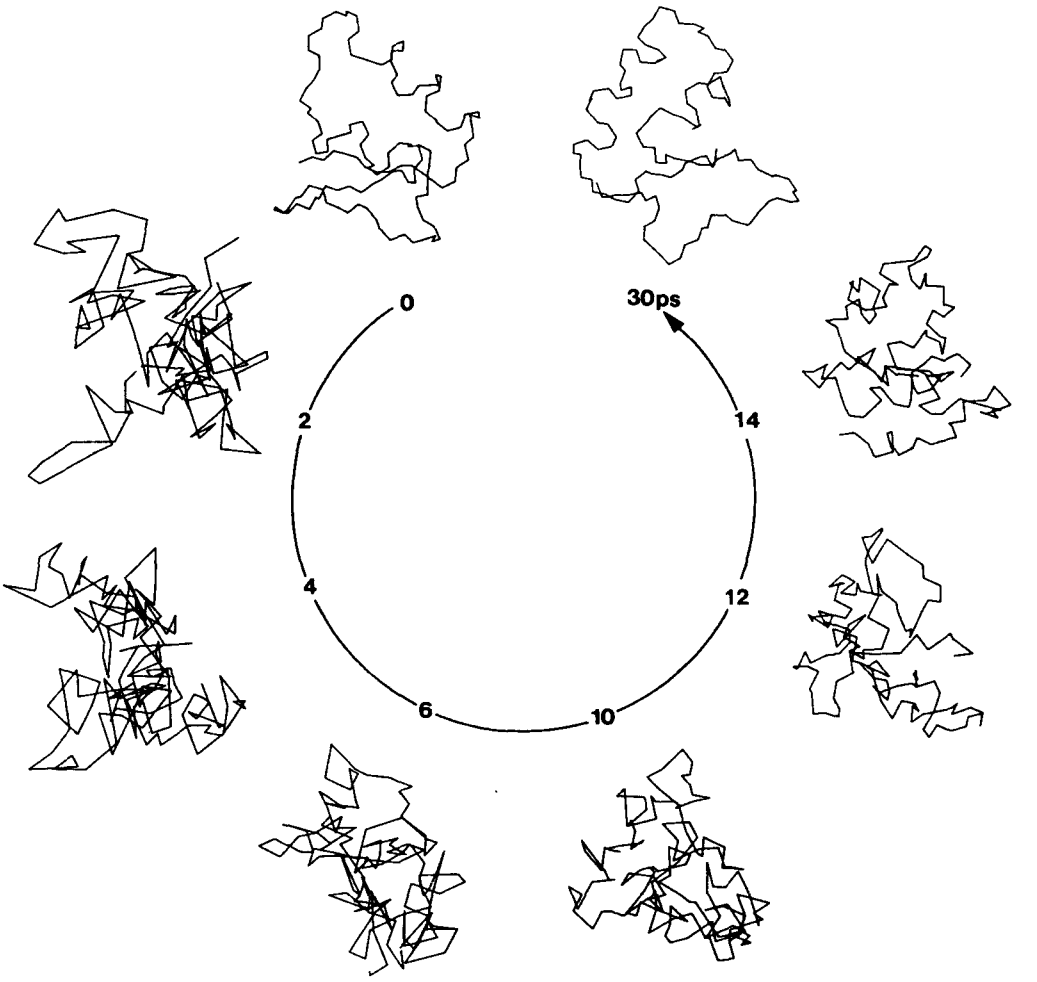


FIGURE 35. Path of the dynamical simulated annealing calculations for crambin starting from the global mirror image of the X-ray structure. The calculation is based on 240 approximate interproton distance restraints.

observed for two β-methylene protons that cannot be stereospecifically assigned in the straightforward manner described in Section V.B.2, they may be arbitrarily assigned to β_2 and β_3. For the first two strategies that start out from structures with perfect chirality at prochiral centers, all that is required to use the procedure for finding the correct stereospecific assignment is to have a low value for the $C^{\beta 2}H\text{-}C^{\beta}\text{-}C^{\beta 3}H$ angle term that is slowly increased during the course of the calculation. For the third strategy, the tetrahedral improper for the prochiral center is simply left out. The end result is that the assignment that best satisfies the NOE data is automatically selected during the course of the calculation.

From the practical viewpoint, different simulated annealing protocols are often best suited to different sorts of problems. In the case of globular proteins, the hybrid distance geometry-dynamical simulated annealing protocol (Section VIII.D.2) is the most efficient computationally and has a very high convergence rate (≥90%). For extended structures (e.g., linear helical peptides), the protocol starting from different initial structures such as an extended strand (Section VIII.D.1) appears to be more satisfactory as the embedding procedure used to obtain the substructures in the initial stage of the hybrid method performs rather poorly in the absence of long-range NOEs. Finally, the protocol starting from a random array of atoms is suitable for both sets of problems and has a high convergence rate (≥80%), but is computationally more time-

consuming than the other two protocols. With the advent of personal mini-supercomputers, this consideration will become less and less important. An important advantage of this protocol is that random factors are introduced at two levels, namely, in the starting coordinates and in the assignment of the initial velocities. This ensures almost perfect random sampling of the conformational space consistent with the experimental restraints.

IX. QUALITY OF STRUCTURES GENERATED FROM NMR DATA

There are two crucial questions concerning structures determined by NMR, namely, how unique are they and how accurately have they been determined. To answer these questions, it is essential to calculate a number of structures and to examine their atomic rms distribution. In the case of metric matrix distance geometry calculations, this is done by using different random number seeds for the computation of the distances matrices. For real space methods involving only minimization (e.g., least-squares minimization in torsion angle space with a variable target function), random starting structures with perfect covalent geometry are used; for real space methods involving restrained molecular dynamics, structures far from the final structure (e.g., an extended β-strand for proteins, different forms of DNA) together with different random number seeds for the assignment of initial velocities are employed; finally, for the dynamical simulated annealing calculations, one can use different sets of random atom arrays, random structures with perfect covalent geometry, structures far from the final structure, or substructures generated during the early phase of metric matrix distance geometry calculations, combined with different random number seeds for the assignment of initial velocities. If such a series of calculations results in a number of different structural types, all of which satisfy the experimental and stereochemical restraints within their error limits, then the information content of the experimental data can be deemed insufficient to determine the 3D structure of the macromolecule. Conversely, if convergence to a "unique" structural set satisfying the experimental data, exhibiting very small deviations from idealized covalent geometry and exibiting good nonbonded contacts is achieved, then one can be confident that a realistic and accurate picture of the actual solution structure has been obtained and that the region of conformational space occupied by the global energy minimum has been located. The accuracy of different regions of the structure can simply be examined by plotting the atomic rms distribution of the individual structures about the mean structure as a function of residue number. Finally, if convergence does not occur and, in addition, the experimental restraints are not satisfied, it is likely that the restraints contain errors, for example, due to incorrect assignment of one or more long-range NOEs.

Once a reasonable number of structures have been calculated for a given set of data, it is clearly essential to have some quantitative means of judging the quality of the structure. Two separate questions have to be addressed in this respect: (1) how well do the structures satisfy the experimental data as well as the constraints imposed by covalent geometry and van der Waals radii, and (2) how well are the structures determined by the experimental data.

The extent of agreement with the experimental data is easily assessed by computing the rms difference between the experimental and target values of the restraints. So, for example, in the case where the interproton distance restraints are represented by distance ranges, the rms difference between the calculated (r_{ij}) and target restraints, R_{rmsd}, would be calculated with respect to the upper (r_{ij}^u) and lower (r_{ij}^l) limits:[7]

$$R_{rmsd} = \begin{cases} [\Sigma(r_{ij} - r_{ij}^u)^2/n]^{1/2} & \text{if } r_{ij} > r_{ij}^u \\ 0 & \text{if } r_{ij}^l \leq r_{ij} \leq r_{ij}^u \\ [\Sigma(r_{ij} - r_{ij}^l)^2/n]^{1/2} & \text{if } r_{ij} < r_{ij}^l \end{cases} \quad (26)$$

This quantity is essentially the NMR equivalent of the crystallographic R-factor. At the same time, it is also quite useful to count the number of interproton distances that deviate by more than, say, 0.5 Å from the target values, as this often enables one to pinpoint problems with local geometry. The quality of the covalent geometry can be determined by computing the deviations from idealized geometry (i.e., idealized bond lengths, angles, chirality, and planes). Finally, the nonbonded contacts should be judged not only by the value of the van der Waals repulsion term (cf. Equation 24), but also by the value of the 6-12 Lennard-Jones van der Waals energy computed using the parameters from one of the available energy minimization and molecular dynamics programs. The latter is particularly useful, as the van der Waals energy in an energetically good structure should always be negative. In this respect, it is worth pointing out that structures computed using either DISGEO[199] or DISMAN[41] invariably tend to have quite large positive van der Waals energies that can, however, easily be reduced with only small accompanying atomic rms shifts by restrained energy minimization.

The extent to which the conformation of the macromolecule has been determined on the basis of the experimental NMR data is best assessed in a qualitative manner by inspecting stereoplots showing best fit superpositions of the calculated structures, and in a quantitative manner by computing the average atomic rms difference between the individual converged structures and the mean structure obtained by averaging their coordinates. To assess which regions are well-determined by the data and which are poorly determined by the data, it is best to plot the latter as a function of residue number. Considering the ill-defined regions, it is important to distinguish two cases. The ill-defined parts of the structure can be fully disordered. Alternatively, they can be locally well-defined but globally ill-defined such that the position of a whole group of residues varies from structure to structure with respect to the remainder of the protein. This, for example, is often the case with exposed loops. A simple iterative method based on both atomic rms differences and interatomic distances between equivalent atoms in all residue pairs has been developed that enables one to distinguish well-defined from variable regions and to identify several locally well-defined regions whose relative positions are not well-defined globally.[211] It is also very important to remember that the degree of precision with which a particular region is defined is solely a function of the experimental restraints and cannot be correlated with factors such as increased mobility. In many cases, it will be found that poorly defined regions comprise exposed turns and loops, which, because of their location, are quite likely to be more mobile. Even if these regions were completely rigid, however, they would still be poorly defined if the density of short interproton distances for these structural elements is low. These conclusions are borne out by model calculations using crystallographic structures to obtain a set of distance restraints.[16,22,23,44,45]

A. MODEL CALCULATIONS

1. Proteins

The accuracy of the method is best judged from model calculations in which a set of interproton distances that could be realistically obtained from NMR measurements has been derived from X-ray structures and then used to compute a set of structures employing one of the methods discussed in Section VIII. The metric matrix distance geometry program DISGEO[199] has been tested on basic pancreatic trypsin inhibitor (BPTI)[40] and crambin;[46] the DISMAN program for minimization in torsion angle space with a variable target function has been tested on BPTI;[41] and the restrained molecular dynamics protocol[45,46] and all three dynamical simulated annealing protocols[16,22,23] have been tested on crambin. A number of distance sets were used for the model calculations, but all included at least one set that could realistically be obtained experimentally and did not necessitate stereospecific assignments of β-methylene protons. In such cases, the overall size, shape, and polypeptide fold are well reproduced. The conformational space consistent with the NOE data sampled by DISMAN, restrained molecular dynamics, and dynamical simulated annealing is comparable and somewhat larger than that sampled

by DISGEO. This is particularly marked in regions that are poorly defined by the data. Thus, for example, subjecting structures generated by DISGEO to restrained molecular dynamics invariably results in an increase in the atomic rms distributions while improving the quality of the structures both with respect to agreement with the NOEs and the nonbonding contacts.[7,12,13,15,21,24] In the case of DISMAN and DISGEO, the structures tend to be slightly expanded relative to the X-ray structures, quite large deformations in the local backbone structures are apparent, and the non-bonded contacts are poor. In contrast, the structures produced by restrained molecular dynamics tend to be slightly compressed relative to the X-ray structures, the local structure tends to be better reproduced, and the nonbonded contacts are good. The structures produced by dynamical simulated annealing are similar in quality to those generated by restrained molecular dynamics, but like DISGEO and DISMAN structures tend to be slightly expanded.

In quantitative terms, the average atomic rms difference between the calculated structures and the corresponding X-ray structures for a data set without stereospecific assignments is 1.5 to 2.0 Å for the backbone atoms and 2.0 to 3.0 Å for all atoms. This is illustrated in Figure 36 by the superpositions of a number of crambin structures obtained by restrained molecular dynamics. Significant improvements, however, can be obtained by averaging the coordinates of the calculated structures best fitted to each other. This invariably results in a structure that is closer to the X-ray structure than any of the individual structures. Thus, for example, in the case of the dynamical simulated annealing calculations on crambin starting out from a random array of atoms the atomic rms difference was reduced by averaging from 1.2 ± 0.2 Å to 0.9 Å for the backbone atoms and from 1.7 ± 0.2 Å to 1.3 Å for all atoms.

The average structure itself has no physical significance insofar as it does not represent a structure that actually exists and further is extremely poor in energetic terms (i.e., both with respect to covalent and nonbonded interactions). The concept of the average structure, however, is very useful. It represents the mean structure about which the individual structures are randomly distributed, and can be characterized in terms of the atomic rms error in its coordinates, which is simply given by ~rmsd/ $\sqrt{n}$ (where rmsd is the average atomic rms difference between the n structures and the average structure). In addition, it provides a reference point for measuring the atomic rms distribution of the individual structures. This measure is more useful than the average atomic rms difference between all pairs of individual structures, as its value is independent of the total number of structures. The latter, however, is related to the former by a factor of $\sim[2n/(n-1)]^{1/2}$.

A structure with physical significance that is closer to the mean structure than any of the individual structures and is energetically reasonable can be obtained from the mean structure by restrained minimization (either using the simplified target function of dynamical simulated annealing or the full empirical energy function of restrained molecular dynamics). Care, however, has to be taken when doing this as the nonbonded energy of the mean structure is very high. Consequently, it is essential to slowly increase the force constant on the van der Waals term from an initial low value to its final value during the course of the minimization.[16,45] The atomic rms shifts resulting from this procedure are generally small (0.2 to 0.3 Å) and the atomic rms difference with respect to the X-ray structures is only slightly larger than that of the mean structure. The quality of the restrained minimized mean structure in terms of nonbonded contacts depends to a large extent on the atomic rms distribution of the individual structures. The smaller this distribution (i.e., the better the structures are determined by the NMR data), the higher the quality, in energetic terms, of the restrained minimized mean structure.

A further test of the quality of structures that can be obtained from NMR data, as well as the assessment of the usefulness of the concept of the average structure, comes from a study in which the NMR structures were used to solve the X-ray structure of crambin directly by molecular replacement.[212] It was shown that the correct solution of the translation and rotation functions of the Patterson search could only be obtained using either the restrained minimized mean

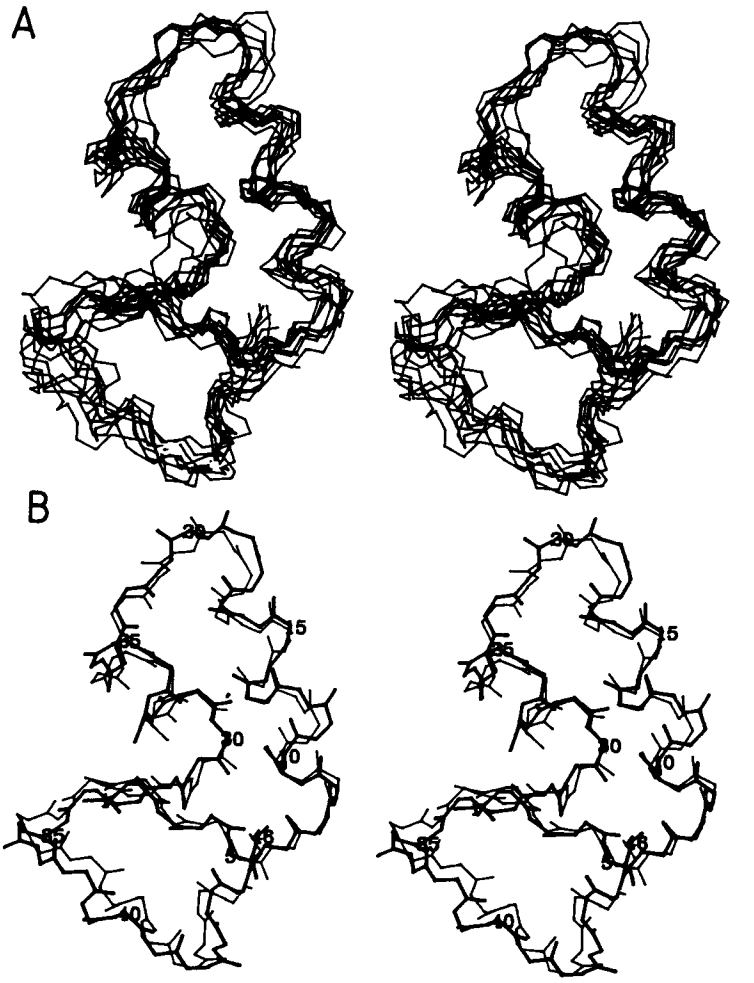


FIGURE 36. (A) Superposition of 9 crambin structures obtained by dynamical simulated annealing on the basis of 240 approximate interproton distance restraints derived from the crystal structure. (B) Best fit superposition of the restrained minimized mean structure (thick line) and the X-ray structure (thin line). The restrained minimized mean structure is derived by averaging the coordinates of the individual simulated annealing structures best fitted to each other and subjecting the resulting coordinates to restrained minimization. (From Nilges, M., Gronenborn, A. M., Brünger, A. T., and Clore, G. M., *Protein Eng.*, 2, 27, 1988. With permission.) (The X-ray structure was solved by Hendrickson, W. A. and Teeter, M. M., *Nature (London)*, 290, 107, 1981. With permission.)

structure as a starting model or by averaging the individual translation and rotation functions obtained using the individual structures as starting models. It was also shown that such starting models could be refined by conventional refinement techniques that reduced the R-factor for the restrained minimized mean structure from 0.43 at 4 Å resolution to 0.27 at 2 Å resolution without the inclusion of water molecules. This compares to an R-factor of 0.25 at 2 Å resolution for the published X-ray structure[213] when all the water molecules are omitted in the computation of the calculated structure factors. A similar study has been recently carried with α-amylase inhibitor in which the NMR structure[8] was used to solve the X-ray structure by molecular replacement.[218a] The results are in agreement with the X-ray structure solved by conventional isomorphous replacement.[214]

B. OLIGONUCLEOTIDES

Several model calculations have been carried out for oligonucleotides using restrained molecular dynamics.[190] To establish the feasibility of the method, a series of calculations were carried out on the hexamer 5'd$(CGTACG)_2$ without restraints (i.e., free dynamics) and with a set of 192 distance restraints taken from classical A- and B-DNA comprising the same 192 proton pairs that had been measured experimentally.[97] The interproton target distances (r_{ij}^{o}) were set equal to the values calculated from ideal B- and A-DNA structures and the force constants (cf. c_1 and c_2 in Equations 18 and 19) were evaluated using estimated errors of –0.1 Å/+0.2 Å for $r < 2.4$ Å, –0.15 Å/+0.3 Å for 2.4 Å $\leq r \leq$ 3.4 Å, and –0.2 Å/+0.4 Å for 3.4 Å $< r \leq$ 4.4 Å. In the absence of restraints, A-DNA remained A-type and B-DNA B-type. In the presence of B-DNA restraints, A-DNA was converted to B-DNA, and similarly in the presence of A-DNA restraints, B-DNA was converted to A-DNA. In the case of the restrained dynamics calculation starting from A-DNA, the resulting B-type structure was equally close to both classic B-DNA and the B-DNA free dynamics structure (rms difference ~1 Å). Similar results were obtained for the reverse calculation, although in that case the resulting restrained dynamics structure was closer to the A-DNA free dynamics structure than to classic A-DNA. This is simply due to the fact that there are some poor nonbonded contacts in classic A-DNA reflected by a positive Lennard Jones van der Waals energy, whereas in classic B-DNA the van der Waals energy is negative.

When either experimental or idealized interproton distances are used as restraints, the restrained molecular dynamics simulations converge to final structures independent of the choice of the initial structure (e.g., with both model and experimental NOE data, ideal A- and B-DNA starting structures yield a B-type DNA with atomic rms differences between the two restrained dynamics structures comparable to the rms fluctuations of the atoms about their average positions). This is in marked contrast to the results of restrained energy minimization that alters only the local structure and leaves the global structure unchanged.

Examination of the time dependence of the NOE, van der Waals, and electrostatic energy components during the A to B transition reveal that the principal driving force arises from a decrease in the restraints energy term coupled with a smaller improvement in the van der Waals energy term. There is *no* significant change in the electrostatic energy component during the course of the transition. This suggests that considerable caution is required concerning structure determination approaches for oligonucleotides in which the nonbonded interaction is represented solely by a hard-sphere van der Waals repulsion term. Indeed, matrix distance geometry calculations[125-130,136] invariably yield structures that are systematically underwound and have α, β, ε, and ζ backbone torsion angles well outside the ranges characteristic of all right-handed DNAs studied by either fiber diffraction or single crystal X-ray crystallography.[167,215] Problems with systematic underwinding were also encountered using the dynamical simulated annealing approach, and, in addition, the resulting structures did not satisfy the experimental restraints as well as the restrained dynamics structures (unpublished data). The results with algorithms using a hard-sphere van der Waals repulsion term can easily be rationalized, since base-base stacking is probably the key feature contributing to the stability and conformation of oligonucleotides, and unlike in proteins where the tertiary structure ensures that NOEs are observed between residues far apart in the sequence, all measured interresidue NOEs for oligonucleotides are restricted to adjacent base pairs so that only short range NOEs are available for a structure determination.

X. PROTEIN STRUCTURES

A. COMPARISON OF SOLUTION AND X-RAY STRUCTURES OF GLOBULAR PROTEINS

To date, there are four globular proteins whose structures have been solved both by X-ray crystallography and by NMR spectroscopy. These are BPTI,[20,216] potato carboxypeptidase inhibitor (CPI),[21,217] barley serine proteinase inhibitor-2 (BSPI-2),[24,218,219] and α-amylase inhib-

itor.[8,214] The results obtained provide a yardstick to assess the extent to which differences between the "solution" and X-ray structures arise from (1) genuine differences as reflected by differences between the experimental interproton distance restraints and a corresponding set of X-ray-derived restraints on the one hand, and (2) from inadequacies in the NMR data as reflected by the limitations of number, range (<5 Å), and accuracy of the experimental data.

A detailed comparison in the case of CPI[21] and BSPI-2[218] indicates that there are probably some genuine differences between the solution and X-ray structures as manifested by noticeably larger interproton distance deviations between the calculated and experimental distances for the X-ray structures than for the computed solution structures. These differences, however, are clearly of a minor nature, as most of the deviations can be corrected by restrained minimization of the X-ray structures with only minor atomic rms shifts (of the order of ~0.5 Å for all atoms). Nevertheless, a few interproton distance deviations larger than 0.5 Å still remain after this procedure and reflect more substantial differences. Thus, in the case of CPI, there are two regions where the backbone atomic rms differences are significant, as can readily be appreciated from the best fit superposition shown in Figure 37. The first region comprises the segment including residues 18 to 20, which are part of a short helix extending from residues 16 to 21 in the solution structures, whereas in the X-ray structure this region is slightly more distorted. The second region comprises the turn formed by residues 28 to 31, which has a slightly different orientation with respect to the rest of the protein in the solution and X-ray structures. In contrast, no significant atomic rms differences between the solution and X-ray structures of BSPI-2 could be detected within the resolution of that particular NMR determination (Figure 38).

It should also be pointed out that quite large differences at the N and C termini are frequently observed. In general, however, these cannot be regarded as significant, as in most cases there is a paucity of experimental NMR data at both ends of the polypeptide chain, especially if the termini are not folded back onto the main structure.

In quantitative terms, the X-ray and solution structures are very similar. Thus, the average backbone atomic rms differences between the individual solution structures and the corresponding X-ray structures are 2.1 ± 0.2 Å for CPI,[21] 1.9 ± 0.2 Å for BSPI-2,[218] and 2.2 ± 0.5 Å for BPTI.[20] In the case of CPI and BSPI-2, the individual structures were also averaged to yield a mean structure that was then subjected to restrained minimization to restore good covalent geometry and non-bonded contacts. In both these cases, this resulted in a structure that was closer to the X-ray structure than any of the individual structures, namely, 1.7 Å for CPI and 1.5 Å for BSPI-2.

No detailed comparison has been published for the fourth protein, α-amylase inhibitor, which was solved independently by NMR[8] and X-ray crystallography.[214] Nevertheless, it would appear that the atomic rms differences between the computed solution structures and the X-ray structure are similar in size to that reported for the three other proteins. Further, recent work on improving the quality of the NMR structures by obtaining many more distances, as well as stereospecific assignments, suggests that the backbone atomic rms difference between the solution and X-ray structures can be reduced to the 1 Å level.[218a]

B. COMPARISON OF SOLUTION STRUCTURES WITH X-RAY STRUCTURES OF RELATED PROTEINS

Several examples exist where NMR structures have been determined for proteins that are related by sequence homology to proteins whose X-ray structures have been solved. These are the plant toxins α1-purothionin[7] and phoratoxin,[12] which are related to crambin,[213] proteinase inhibitor IIA from bull seminal plasma (BUSI),[4] which is related to the third domain of Japanese quail ovomucoid (OMJPQ3)[220] and pancreatic secretory trypsin inhibitor (PSTI),[221] and the anaphylatoxin C5a,[33,34] which is related to C3a.[222] In general, there is good agreement between the solution structures and the corresponding related X-ray structures.

The sequence homology between α1-purothionin, phoratoxin, and crambin is quite high: α1-purothionin and phoratoxin are 33 and 39%, respectively, homologous to crambin and 46%

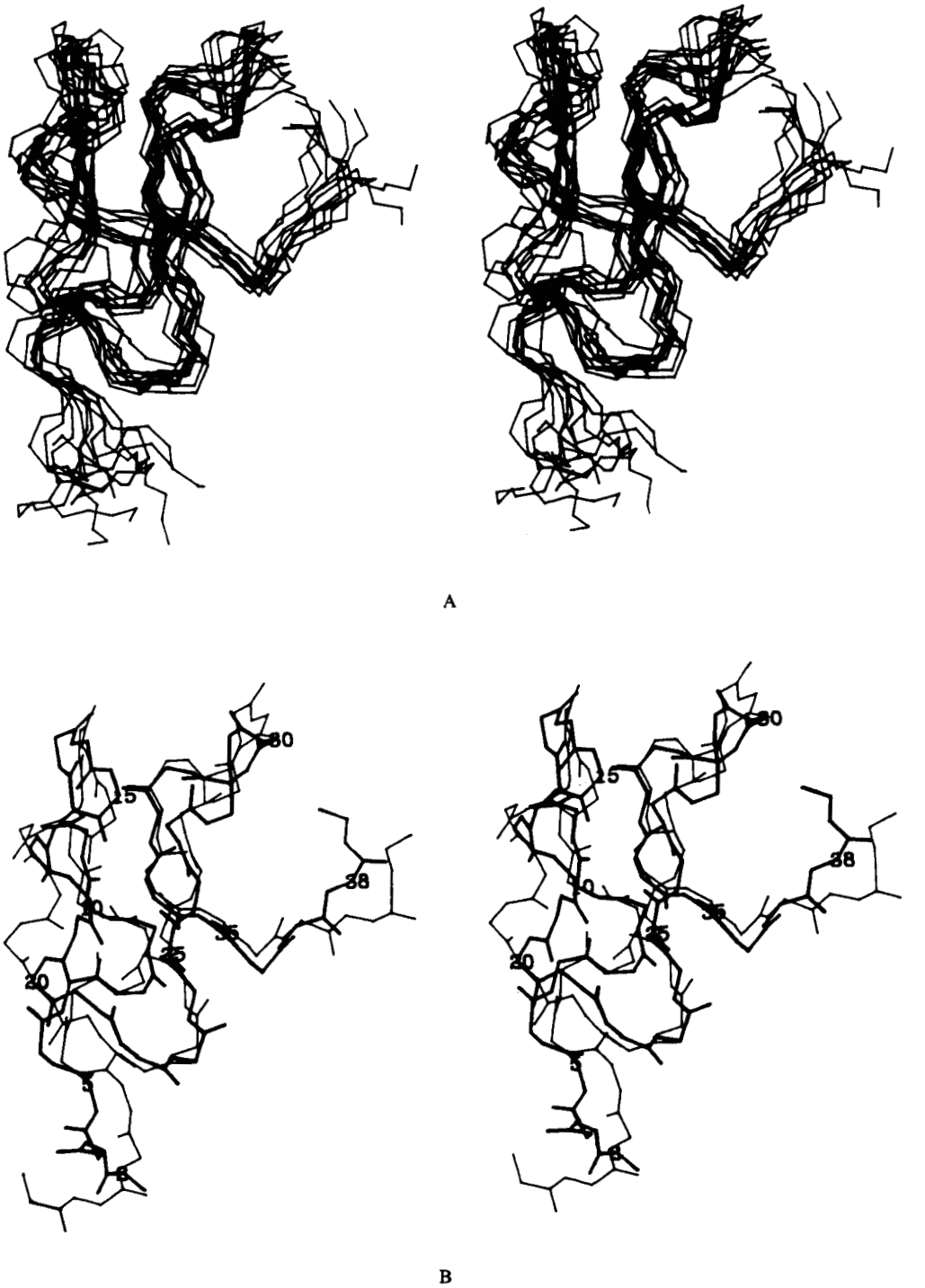


FIGURE 37. (A) Best fit superposition of the backbone atoms of 11 restrained molecular dynamics structures of CPI determined on the basis of 309 interproton distance restraints. (b) Best fit superposition of the backbone atoms of the restrained energy minimized mean structure of CPI (thick lines) and the crystal structure (thin lines). The restrained energy minimized mean structure is derived by averaging the coordinates of the individual restrained dynamics structures best fitted to each other and subjecting the resulting coordinates to restrained energy minimization. (From Clore, G. M., Gronenborn, A. M., Nilges, M., and Ryan, C. A., *Biochemistry*, 26, 8012, 1987. With permission.) (The X-ray structure of CPI was solved by Rees, D. C. and Lipscomb, W. N., *J. Mol. Biol.*, 160, 475, 1982.)

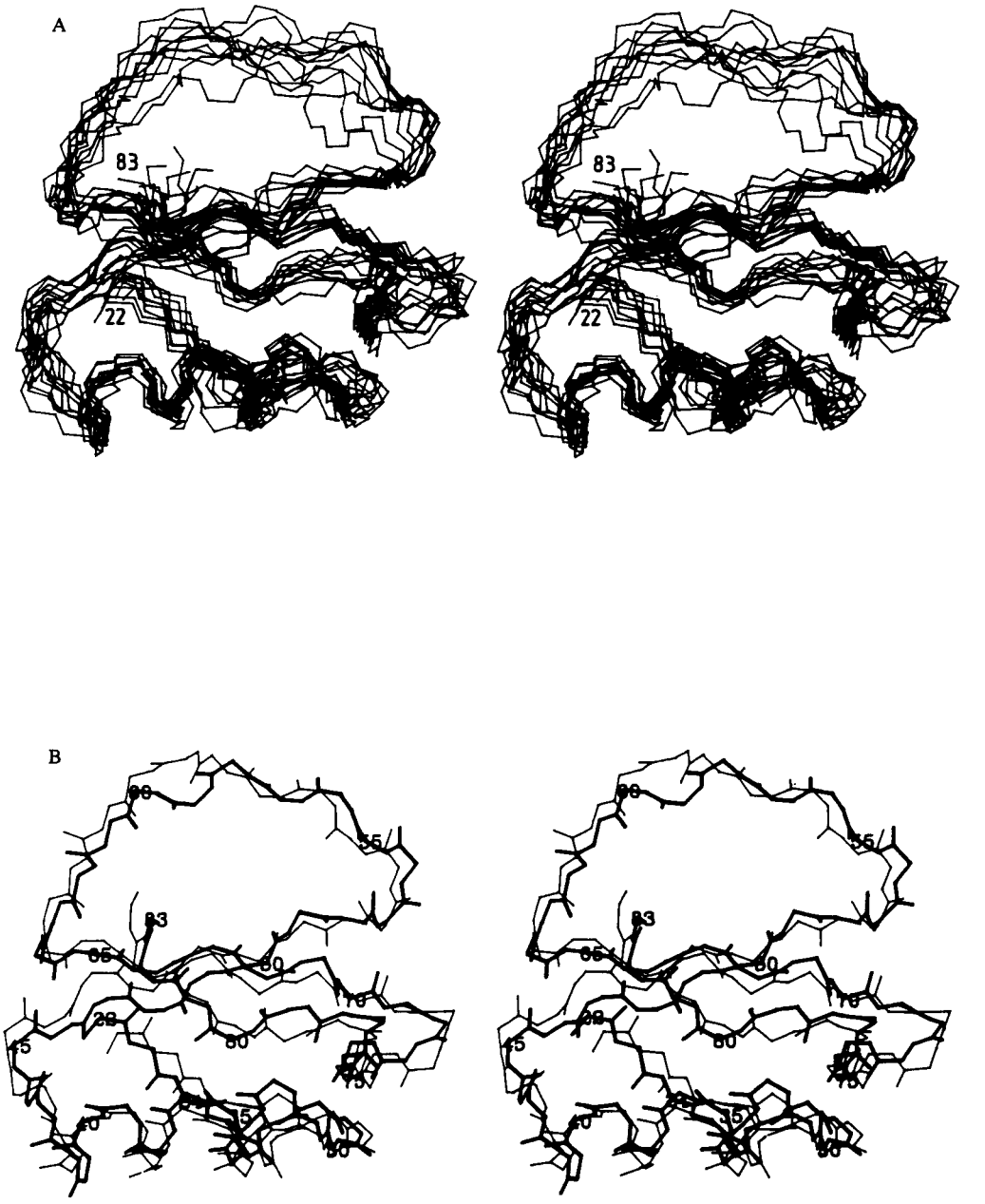


FIGURE 38. (A) Best fit superposition of the backbone atoms of 11 restrained molecular dynamics structures of BSPI-2 determined on the basis of 403 interproton distance restraints. (B) Best fit superposition of the backbone atoms of the restrained minimized mean structure of BSPI-2 (thick lines) and the crystal structure (thin lines). (From Clore, G. M., Gronenborn, A. M., Kjaer, M., and Poulsen, F. M., *Protein Eng.*, 1, 305, 1987; Clore, G. M., Gronenborn, A. M., James, M. N. G., Kjaer, M., McPhalen, C. A., and Poulsen, F., *Protein Eng.*, 1, 313, 1987. X-ray structure of BSPI-2 was solved by McPhalen, C. A. and James, M. N. G., *Biochemistry*, 26, 261, 1987.)

homologous to each other. Considering only the restrained minimized mean structures, the backbone atomic rms difference with respect to crambin is 2.6 Å for α-purothionin[7] and 1.6 Å for phoratoxin.[12] A best fit superposition of the phoratoxin solution and crambin X-ray structures is shown in Figure 39. Given that the quality and extent of the NMR data used to compute the solution structures of purothionin and phoratoxin are comparable, it seems likely that the larger

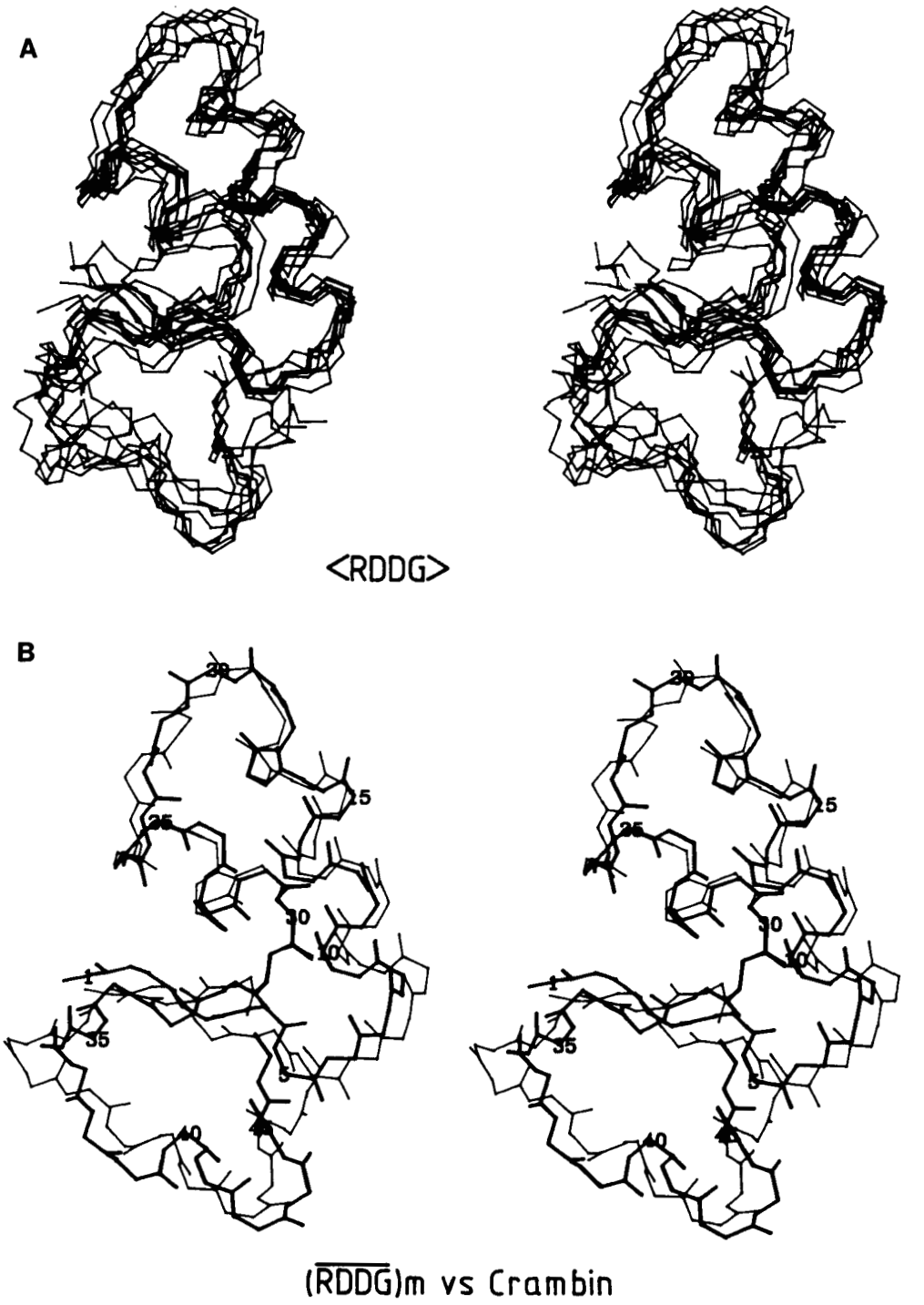


FIGURE 39. (A) Best fit superposition of the backbone atoms of eight restrained molecular dynamics structures of phoratoxin calculated on the basis of 331 interproton distance restraints. (B) Best fit superposition of the the backbone atoms of the restrained minimized mean structure of phoratoxin (thick lines) and the crystal structure of crambin (thin lines). The sequence homology between these two proteins is 39%. (From Clore, G. M., Sukumaran, D. K., Nilges, M., and Gronenborn, A. M., *Biochemistry*, 26, 1732, 1987. With permission.)

atomic rms difference between α1-purothionin and crambin may reflect the deletion of residue 24 and the presence of an extra disulfide bridge between residues 12 and 29 in α1-purothionin.

A quantitative comparison has also been carried out for the C-terminal fragment (residues 23 to 57) of BUSI, which was superimposed on residues 22 to 56 of the two related inhibitors solved by X-ray crystallography. (Note the difference in alignment is due to a deletion.) The C^{α} atomic rms difference between BUSI on the one hand and OMJPQ3 and PSTI on the other is 1.9 and 2.2 Å, respectively, which compares to a C^{α} atomic rms difference of 1.3 Å between the two X-ray structures.[4]

No quantitative comparison has been carried out in the case of the solution structure of C5a and the X-ray structure of C3a. The polypeptide fold of the two proteins is clearly similar, but there are a number of significant differences between the two structures. The solution structure has four helices comprising residues 5 to 11, 15 to 25, 33 to 39, and 46 to 61. The first 12 residues of C3a were not visible in the electron density map due to disorder, while in the case of C5a there clearly is a N-terminal helix. In contrast, the C-terminal helix extends far beyond the globular head in the crystal structure of C3a and is 10 residues longer than that of C5a in solution. This is due to helix stabilization by packing interactions as C3a is a dimer in the crystal with the interface formed by the C-terminal helices arranged in an antiparallel manner.[222] Interestingly, the secondary structure of C3a in solution is virtually identical to that of C5a.[186]

In addition to these four examples, there are two other proteins whose polypeptide fold, as determined by NMR, is similar to a related X-ray structure, but where no quantitative comparison has been presented, namely, the short scorpion insectotoxin $I_5A^{3,}$ which is similar to the helix and antiparallel β-sheet fragment in the crystal structure of the long scorpion toxin v-3,[223] and the lac repressor headpiece,[5] which is similar to the central portion (helices 2 to 4) of the X-ray structure of the sequence related λCI repressor.[224]

C. COMPARISON OF SOLUTION AND X-RAY STRUCTURES OF NONGLOBULAR PROTEINS AND POLYPEPTIDES

Considering nonglobular proteins and polypeptides, there are two cases available where structures have been solved by NMR and X-ray crystallography, namely, glucagon[2] and metallothionein.[9-11] In both cases there are large and significant differences.

The X-ray structure of glucagon is almost entirely helical,[225] whereas that of micelle-bound glucagon is composed of an irregular N-terminal strand (residues 1 to 10) followed by two helices (residues 10 to 13 and 17 to 29) connected by a half-turn (residues 14 to 16). Interestingly, the solution conformations of two sequence-related hormone polypeptides, growth hormone releasing factor,[25,174] and secretin[26,183] in 30 to 40% trifluoroethanol, a solvent that mimics a membrane environment by reducing the water activity, are very similar to that of micelle-bound glucagon.

In the case of the metallothionein, a two-domain structure is found by X-ray crystallography[226] and NMR. Although the polypeptide fold is broadly similar, there are quite distinct differences. Further, there is severe disagreement over the ligand coordination of the seven metal ions. The coordinating cysteine residues for the seven Cd^{2+} ions were unambiguously determined in solution by identifying through-bond scalar couplings between the $CysC^{\beta}H$ protons and the NMR active (spin = 1/2) $^{113}Cd^{2+}$ ions using heteronculear ^{1}H-^{113}Cd correlation spectroscopy.[227] Thus, it is quite possible that selective crystallization of a minor species may have occurred under the experimental conditions employed.[224]

It would therefore appear that for nonglobular proteins and polypeptides, particular care should be taken in deriving structure-function relationships from the results of X-ray analysis alone as the structure in the crystal state may not represent the major species in solution.

D. SOLUTION STRUCTURES FOR WHICH NO X-RAY STRUCTURES EXIST: PUSHING THE LIMITS OF PRECISION

From the model calculations, as well as from results on globular proteins where either an X-

ray structure of the same protein or a related protein is available, it is clear that NMR can be used to determine reliably the 3D structures of globular proteins within certain well-defined limits. These limits depend crucially on the number and distribution of the experimental restraints. The precision with which the majority of NMR structures have been determined to date, as measured by the atomic rms difference between the individual structures and the mean structure obtained by averaging their coordinates, has been of the order of 1 to 2 Å for the backbone atoms and 1.5 to 3 Å for all atoms. One approach to improve the precision of such structure determinations lies in extracting a larger number of structural restraints from the NMR data. In particular, stereospecific assignment of methylene proton resonances, as well as the increased sensitivity and stability offered by 600 MHz spectrometers, should enable a larger number of interproton distances to be obtained. Two examples are discussed below, where attempts at obtaining what could be termed high-resolution NMR structures in this manner have proven successful.

1. BDS-I

BDS-I is a small 43-residue protein from the sea anemone *Anemonia sulcata* with both antihypertensive and antiviral properties.[228] The NMR structure determination was based on a set of 489 approximate interproton distance restraints, 23 ϕ backbone torsion angle restraints and 21 χ_1 side chain torsion angle restraints, derived from NOESY, DQF-COSY, HOHAHA, and E-COSY spectra.[32,160] These restraints, together with 12 backbone hydrogen bonds deduced from a qualitative interpretation of the NOEs and amide hydrogen exchange rates, provided the experimental data for the 3D structure calculations using the hybrid distance geometry-dynamical simulated annealing approach (cf. Section VIII.D.1). Forty-two structures were computed, all of which satisfied the experimental restraints within experimental error, had very small deviations from covalent geometry, and had good nonbonded contacts (Table 3). These structures have been deposited in the Brookhaven Protein Data Bank. The average atomic rms difference between the individual structures and the mean structure is 0.67 ± 0.12 Å for the backbone atoms and 0.90 ± 0.17 Å for all atoms. The definition of the backbone is illustrated by the superpositions of all structures shown in Figure 40 and by the atomic rms distribution plots in Figure 41. The quality of the backbone conformation is also evident from the Ramachandran plot shown in Figure 42, which shows that all nonglycine residues lie within the allowed region of ϕ,ψ conformational space.

The main secondary structure feature of BDS-I is a triple-stranded antiparallel β-sheet, which posesses a right-handed twist (Figure 43). The first (residues 14 to 16) and second (residues 30 to 34) strands of the sheet run parallel to each other and are connected by a long loop extending from residues 17 to 30. The third strand (residues 37 to 41) is antiparallel to and lies between the other two strands, forming backbone hydrogen bonds to them. The second and third strands of this β-sheet are connected by a classic type VIa turn (residues 34 to 37) with Pro-36 in the cis conformation, ϕ,ψ angles of $(-52 \pm 15°, 157 \pm 32°)_{35}$ and $(-79 \pm 6°, 39 \pm 38°)_{36}$, and a hydrogen bond between the amide proton of Asn-37 and the carbonyl oxygen atom of Lys-34. The other element of regular secondary structure consists of a mini-antiparallel β-sheet and a classic type II turn extending from residues 6 to 9 with ϕ,ψ angles of $(-91 \pm 7°, 110 \pm 40°)_7$ and $(101 \pm 47°, -29 \pm 21°)_8$ and a hydrogen bond between the amide proton of Lys-9 NH and the carbonyl oxygen atom of Cys-6. The hydrogen bonds between residues 4 and 13 are arranged in an antiparallel fashion. Residues 10 to 12 form a bulge opposite residue 6.

Of the side chains, only 8 of the 36 side chains have atomic rms distribution values above 1 Å (Figure 41). Without exception, these eight side chains lie on the surface of the protein and display the largest surface accessibilities. Examples of side chain conformations, illustrating the excellent definition, are depicted in Figure 44 A, which shows the best fit superposition of a selected region of BDS-I. With respect to the side chains, it should be borne in mind that the definition of the side chain positions is directly related to the number of NOE distance restraints to them, as well as to the presence or absence of χ_1 torsion angle restraints. Thus, when the

TABLE 3
Structural Statistics for the BDS-I Structures[a]

	<SA>	$\overline{SA}$	$\overline{(SA)}r$
Rms deviations from experimental restraints (Å)[b]			
All (513)	0.085 ± 0.002	0.059	0.079
Interresidue short range (li-jl≤5) (150)	0.083 ± 0.005	0.053	0.086
Interresidue long range (li-jl>5) (105)	0.111 ± 0.007	0.082	0.100
Intraresidue (234)	0.074 ± 0.006	0.053	0.065
hbond (24)[c]	0.051 ± 0.010	0.038	0.043
F_{NOE} (kcal/mol)[d]	187 ± 11	92	160
F_{tor} (kcal/mol)[e]	26 ± 6	13	24
F_{repel} (kcal/mol)[f]	91 ± 7	2417	72
E_{L-J} (kcal/mol)[g]	−103 ± 11	$>10^6$	−117
Deviations from idealized geometry [h]			
bonds (Å) (646)	0.014 ± 0.006	0.329	0.013
angles (°) (1157)	2.910 ± 0.363	28.373	2.517
impropers (°) (242)	0.830 ± 0.060	2.629	0.797

Atomic rms Differences

	Atomic rms difference (Å)	
Structures	Backbone atoms	All atoms
<SA> vs. $\overline{SA}$[i]	0.67 ± 0.12	0.90 ± 0.17
<SA> vs. <SA>[i]	0.96 ± 0.19	1.29 ± 0.23
<SA> vs. $\overline{(SA)}r$	0.73 ± 0.13	1.00 ± 0.18
$\overline{(SA)}r$ vs. $\overline{SA}$	0.21	0.45

[a] The notation of the structures is as follows: <SA> are the 42 final dynamical simulated annealing structures; $\overline{SA}$ is the mean structure obtained by averaging the coordinates of the 42 individual SA structures best fitted to each other; $\overline{(SA)}$ r is the structure obtained by restrained minimization of $\overline{SA}$.

[b] The rms deviations from the experimental restraints are calculated with respect to the upper and lower limits of the distance restraints.[7] None of the structures exhibited violations greater than 0.5 Å. The number of distances in each category is given in parentheses next to the category name.

[c] For each backbone hydrogen bond there are two restraints: r_{NH-O} <2.3 Å and r_{N-O} <3.3 Å. The lower limits are given by the sum of the van der Waals radii of the relevant atoms. On the basis of the NOE and NH exchange data,[32] 12 backbone hydrogen bonds within regular elements of secondary structure were identified.

[d] The values of the square-well NOE potential F_{NOE} (Equation 20) are calculated with a force constant of 50 kcal.mol^{-1}.Å^{-2}.

[e] The values of F_φ are calculated with a force constant of 200 kcal.mol^{-1}.rad^{-2}. F_φ is a square-well dihedral potential (Equation 21) that is used to restrict the ranges of 23ϕ and 21χ_1 torsion angles, and the ω peptide bond torsion angles of the five proline residues (Pro 36 and 42 being restrained to the cis conformation and the others to the trans).

[f] The values of the van der Waals repulsion term F_{repel} (cf. Equation 24) are calculated with a force constant of 4 kcal.mol^{-1}.Å^{-4} with the hard-sphere van der Waals radii set to 0.8 times the standard values used in the CHARMM empirical energy function.[210]

[g] E_{L-J} is the Lennard-Jones van der Waals energy calculated using the CHARMM empirical energy function.[210]

[h] The number of bond, angle, and improper terms is given in parentheses. The improper terms serve to maintain planarity and appropriate chirality; they also maintain the peptide bond of all residues (with the exception of prolines) in the *trans* conformation. In the dynamical simulated annealing calculations, the restraints for the disulfide bridges are included in the bond and angle terms.

[i] Note that using standard statistical theory, it is easily shown that the average rms difference between all pairs of SA structures is related to the average rms difference between the individual SA structures and the mean SA structure by a factor of ~$[2n/(n-1)]^{1/2}$, which in this case is equal to 1.43. Further, the standard atomic rms error, s_{mean}, in the coordinates of the average mean structure $\overline{SA}$ is given by $[\Sigma(\text{rmsd}_i)^2/n(n-1)]^{1/2}$, where rmsd_i is the atomic rms difference between the ith SA structure and the mean $\overline{SA}$ structure, and n is the number of structures. The values for s_{mean} are 0.13 and 0.15 Å for the backbone atoms and all atoms, respectively.

FIGURE 40. Superposition of the backbone (N, C^{α}, and C) atoms of 42 dynamical simulated annealing structures of BDS-I calculated on the basis of 489 interproton distance restraints, 23ϕ and 21χ_1 torsion angle restraints, and 24 distance restraints for 12 hydrogen bonds identified on the basis of the NOE and amide exchange data. (From Driscoll, P. C., Gronenborn, A. M., Beress, L., and Clore, G. M., *Biochemistry*, 28, 2188, 1989. With permission.)

calculations are repeated without stereospecific assignments and χ_1 torsion angle restraints, the definition of the side chain positions deteriorates significantly[229] as illustrated in Figure 44 B.

2. Hirudin

Hirudin is a small 65-residue protein from the leech and is the most potent natural inhibitor of coagulation known.[230] It acts by binding to α-thrombin, thereby abolishing its ability to cleave fibrinogen. Two recombinant hirudin variants have been examined by NMR, namely, wild-type hirudin and the Lys-47→Glu mutant.[14] Analysis of the NMR data indicated that hirudin consists of a N-terminal compact domain (residues 1 to 49) held together by three disulfide linkages and

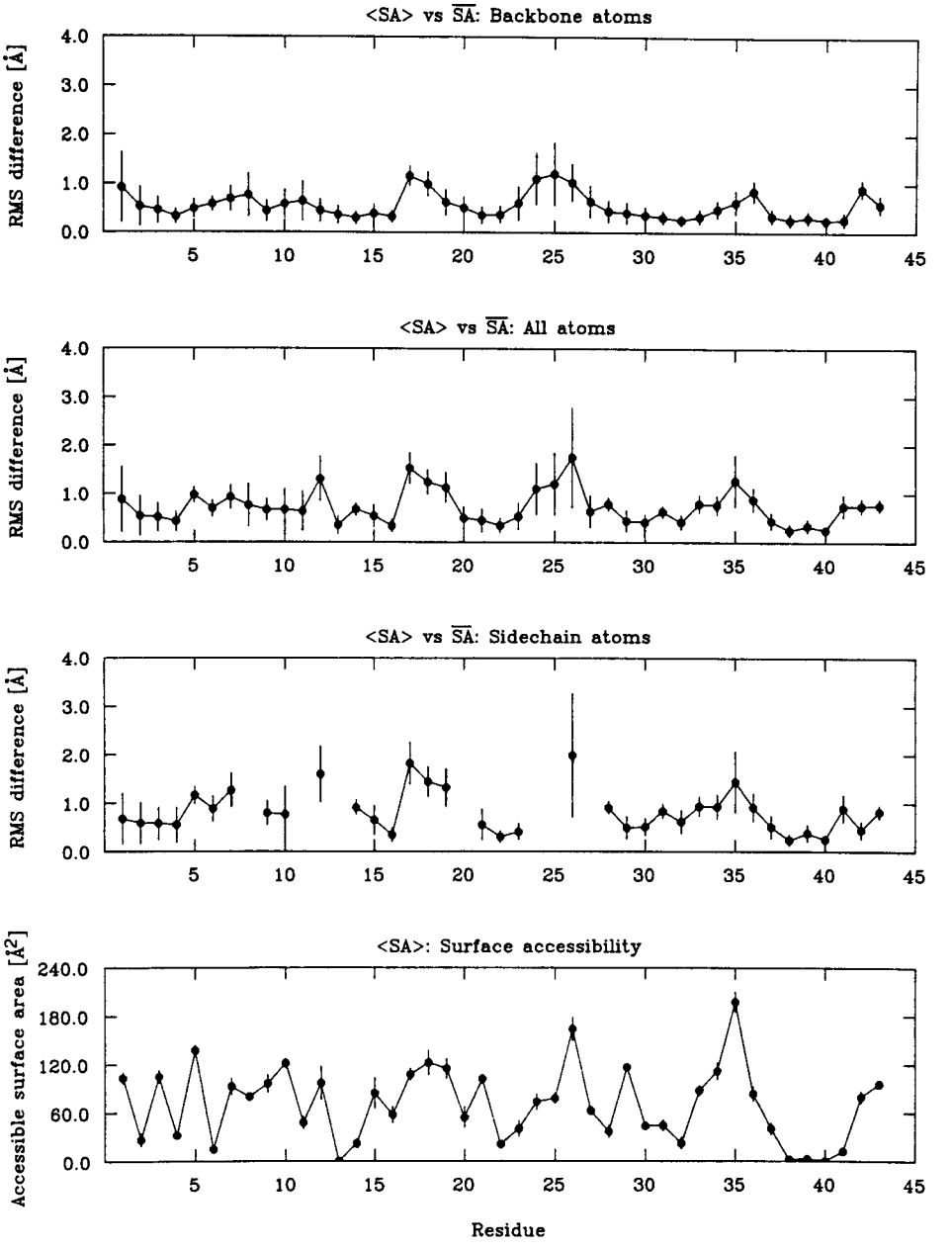


FIGURE 41. Atomic rms distribution of the 42 dynamical simulated annealing structures of BDS-I about the mean. The average deviations from the mean structure are shown by a filled-in circle (•) and the standard deviations in these values by vertical bars. (From Driscoll, P. C., Gronenborn, A. M., Beress, L., and Clore, G. M., *Biochemistry*, 28, 2188, 1989. With permission.)

a disordered C-terminal tail (residues 50 to 65), which does not appear to fold back on the rest of the protein. Structure calculations on the N-terminal domain using the hybrid distance geometry-dynamical simulated annealing method were based on 701 and 677 approximate interproton distance restraints derived from NOE data for the wild-type and mutant hirudin, respectively, 26 ϕ backbone and 18 χ_1 torsion angle restraints derived from NOE and three-bond coupling constant data, and 8 backbone hydrogen bonds identified on the basis of NOE and

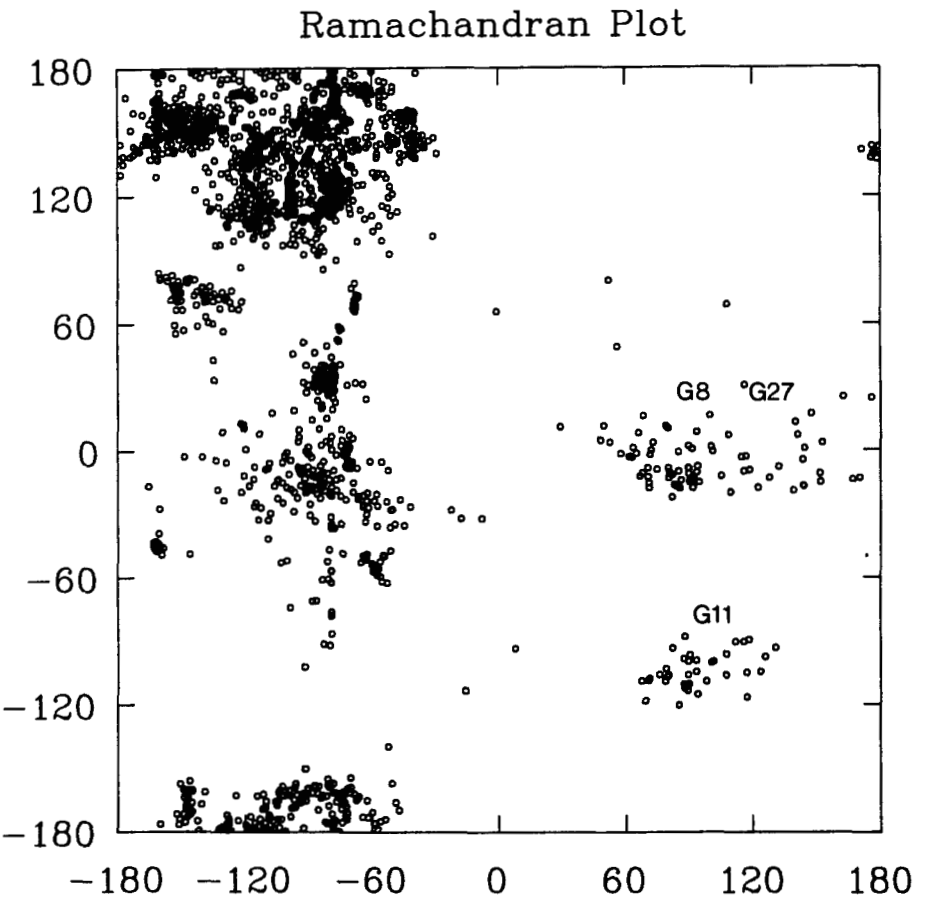


FIGURE 42. Ramachandran plot for the 42 dynamical simulated annealing structures of BDS-I. (From Driscoll, P. C., Gronenborn, A. M., Beress, L., and Clore, G. M., *Biochemistry*, 28, 2188, 1989. With permission.)

amide exchange data. A total of 32 structures were computed for both the wild-type and mutant hirudins and have been deposited in the Brookhaven Protein Data Bank. The structure of residues 2 to 30 and 37 to 48 form the core of the N-terminal domain and have an atomic rms difference between the individual structures and the mean structures of ~0.7 Å for the backbone atoms and ~1 Å for all atoms (Figure 45). The orientation of the exposed finger of antiparallel β-sheet (residues 31 to 36) with respect to the core could not be determined as no long range (|i-j|>5) NOEs were observed between the exposed finger and the core. Locally, however, the polypeptide fold of residues 31 to 36 is reasonably well-defined, despite the obvious existence of local mirror images.

The first five residues form an irregular strand that leads into a loop closed off at its base by the disulfide bridge between Cys-6 and Cys-14. This is followed by a mini-antiparallel β-sheet formed by residues 14 to 16 (strand I) and 21 to 22 (strand I′) connected by a type II turn. This β-sheet is distorted by a β-bulge at Cys-16. Strand I′ leads into a second antiparallel β-sheet formed by residues 27 to 31 (strand II) and 36 to 40 (strand II′) connected by a β-turn (residues 32 to 35). Additionally, residues 10 to 11 exhibit features of a β bulge with the amide of Gly-10 and the carbonyl oxygen atom of Glu-11 hydrogen bonded to the carbonyl and amide groups, respectively, of Cys-28. Finally, strand II′ leads into an irregular strand that folds back onto the protein so that residue 47 (Lys in the wild type, Glu in the mutant) is in close proximity to residues in the loop closed off by the disulfide bridge between Cys-6 and Cys-14.

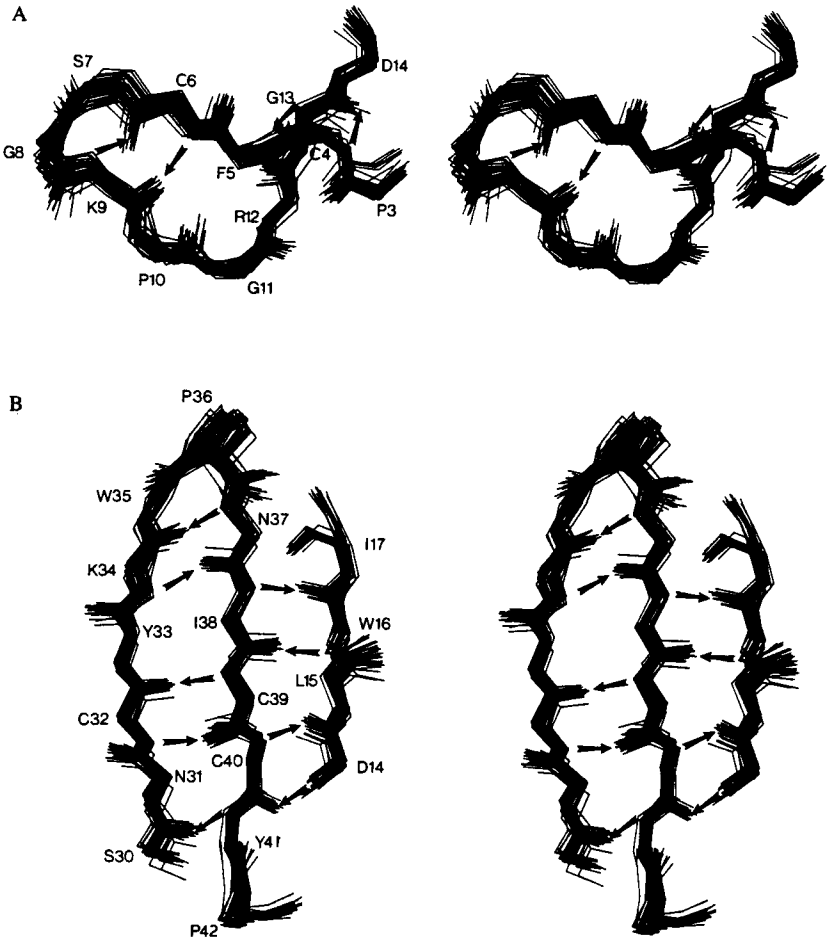


FIGURE 43. Stereoviews showing best fit superpositions of the backbone (N, C^{α}, C, and O) atoms of the 42 dynamical simulated annealing structures of BDS-I for the two regions of regular secondary structure. Backbone hydrogen bonds are indicated by arrows directed from the N to the O atom. (From Driscoll, P. C., Gronenborn, A. M., Beress, L., and Clore, G. M., *Biochemistry*, 28, 2188, 1989. With permission.)

As in the case of BDS-I, many of the side chain conformations are well defined. This is illustrated by the superpositions in Figure 46, which show that the side chain conformations in the vicinity of residue 47 are essentially unaffected by the Lys-47→Glu mutation.

A superposition of the core (residues 1 to 30 and 37 to 49) of the restrained minimized mean structures of the wild-type and mutant hirudin provides a good representation of the differences between the two structures (Figure 47). Regions of noticeable difference can be identified where the atomic rms difference between the two mean structures is larger than the atomic rms distribution of the individual structures about their respective means. This analysis indicates the presence of clear differences for the backbone atoms of residues 3, 5, 8, 11 to 15, 22, 26, and 27. In the mutant, the backbone atoms of residues 3, 4, and 11 to 15 are slightly closer to that of residues 45 to 47 than in the wild-type, a change that can be rationalized in terms of the shorter length of the Glu side chain relative to that of Lys. Concomitantly, the backbone of residue 8 appears to be pushed away in the wild-type structure. Residues 22, 26, and 27 also move in the same direction as residues 11 to 15. This is secondary to the perturbation of residues 11 to 15

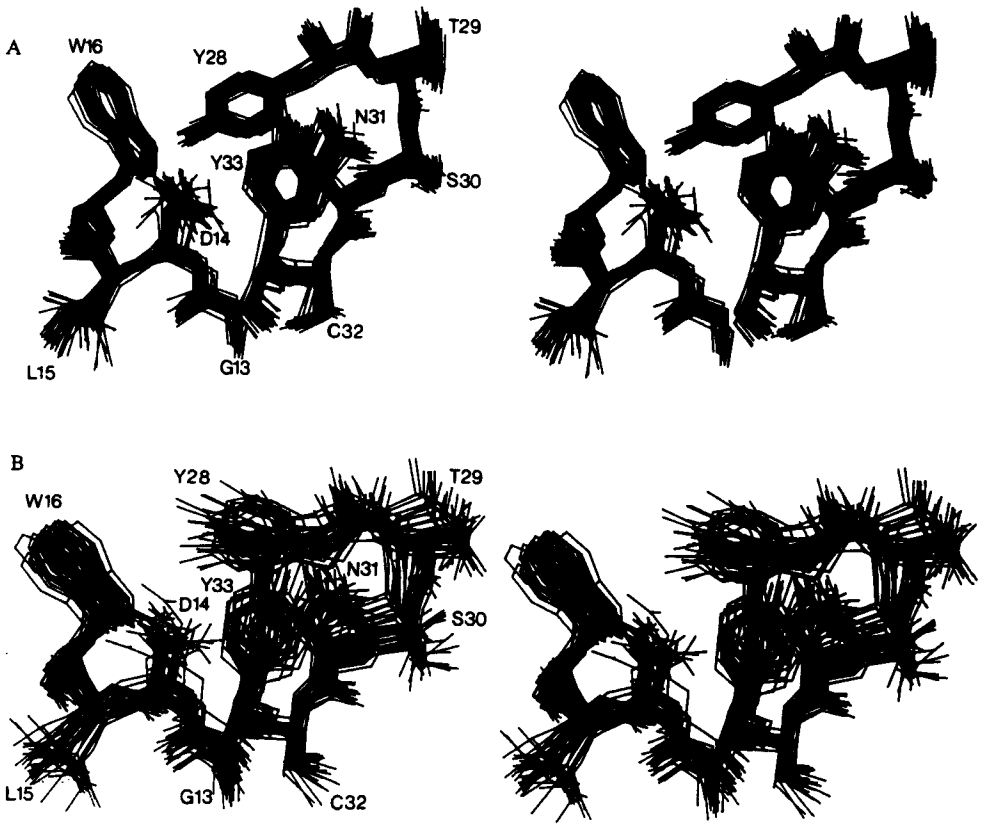


FIGURE 44. Best fit superposition of the 42 dynamical simulated annealing structures of BDS-I for a selected region illustrating the definition of side chain conformations. The structures in A were calculated with stereospecific restraints, while those in B were calculated without. Note the significant increase in definition resulting from the inclusion of stereospecific restraints. (From Driscoll, P. C., Gronenborn, A. M., Beress, L., and Clore, G. M., *Biochemistry*, 28, 2188, 1989; Driscoll, P. C., Clore, G. M., and Gronenborn, A. M., *FEBS Lett.*, 243, 223, 1989. With permission.)

and can be attributed to the presence of numerous contacts between residues 8 to 11 on the one hand and 28 to 30 on the other, including hydrogen bonds between Cys-28 and Gly-10 and between Cys-28 and Glu-11. There seems to be no significant differences, however, with respect to the side chain positions within the errors of the coordinates, even within the immediate vicinity of residue 47 (Figures 46 and 47B).

3. Attainable Precision

With only a few exceptions, both the backbone and side chain positions of BDS-I and hirudin are well-defined. At the present time, they probably represent the limits of precision attainable by NMR. This is achieved in spite of the fact that the NOE interproton distance restraints are only loosely classified into three broad distance ranges (≤2.7, ≤3.3, and ≤5 Å), thus avoiding any potential problems associated with variable internal mobility and r^{-6} averaging. The success of obtaining such well-defined structures is due to the fact that many of the distances are highly correlated and that short (<5 Å) distances between residues far apart in the sequence are conformationally restrictive, even if they are only approximate. This is fortunate, as accurate interproton distance measurements in proteins, particularly those involving side chains, are likely to prove technically very difficult for the reasons discussed in Section V.A.

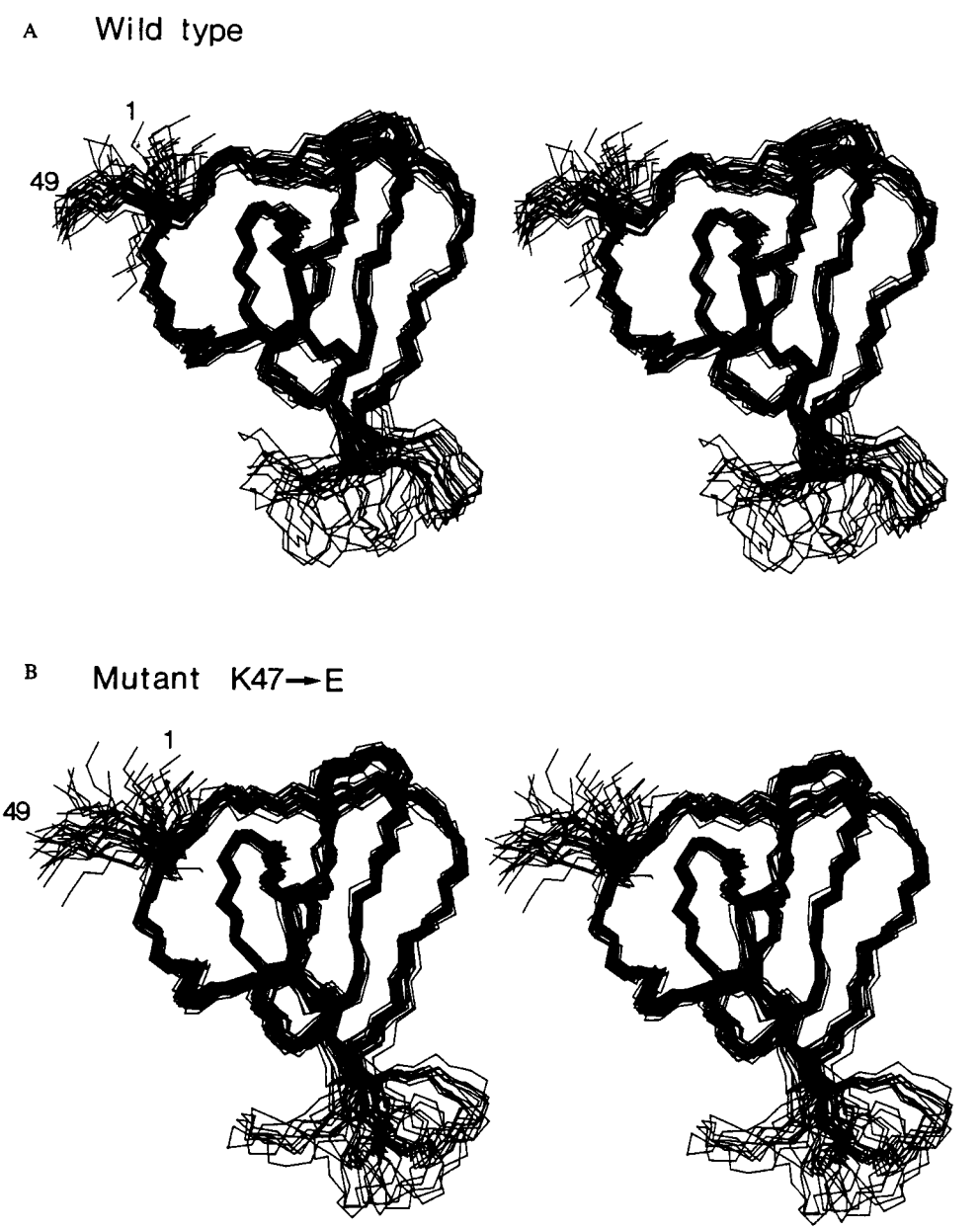


FIGURE 45. Superposition of the backbone (N, C^{α}, and C) atoms of 32 dynamical simulated annealing structures of wild-type hirudin (A) and the Lys-47→Glu mutant (B) for the first 49 amino acids. The wild-type and mutant structures were calculated on the basis of 701 and 677 interproton distance restraints, respectively, 26ϕ and 18χ_1 torsion angle restraints. (From Folkers, P. J. M., Clore, G. M., Driscoll, P. C., Dodt, J., Köhler, S., and Gronenborn, A. M., *Biochemistry*, 28, 2601, 1989. With permission.)

E. OLIGONUCLEOTIDE STRUCTURES

Unlike proteins, the interproton distances that can be measured by NMR in oligonucleotides are restricted to adjacent base pairs (i.e., they are all short-range distances). Consequently, the description of the van der Waals interaction becomes extremely important. The use of a simple van der Waals repulsion term that is so successful for proteins invariably leads to distorted structures.[125-130,136,137] In contrast, the use of a Lennard-Jones van der Waals term with an

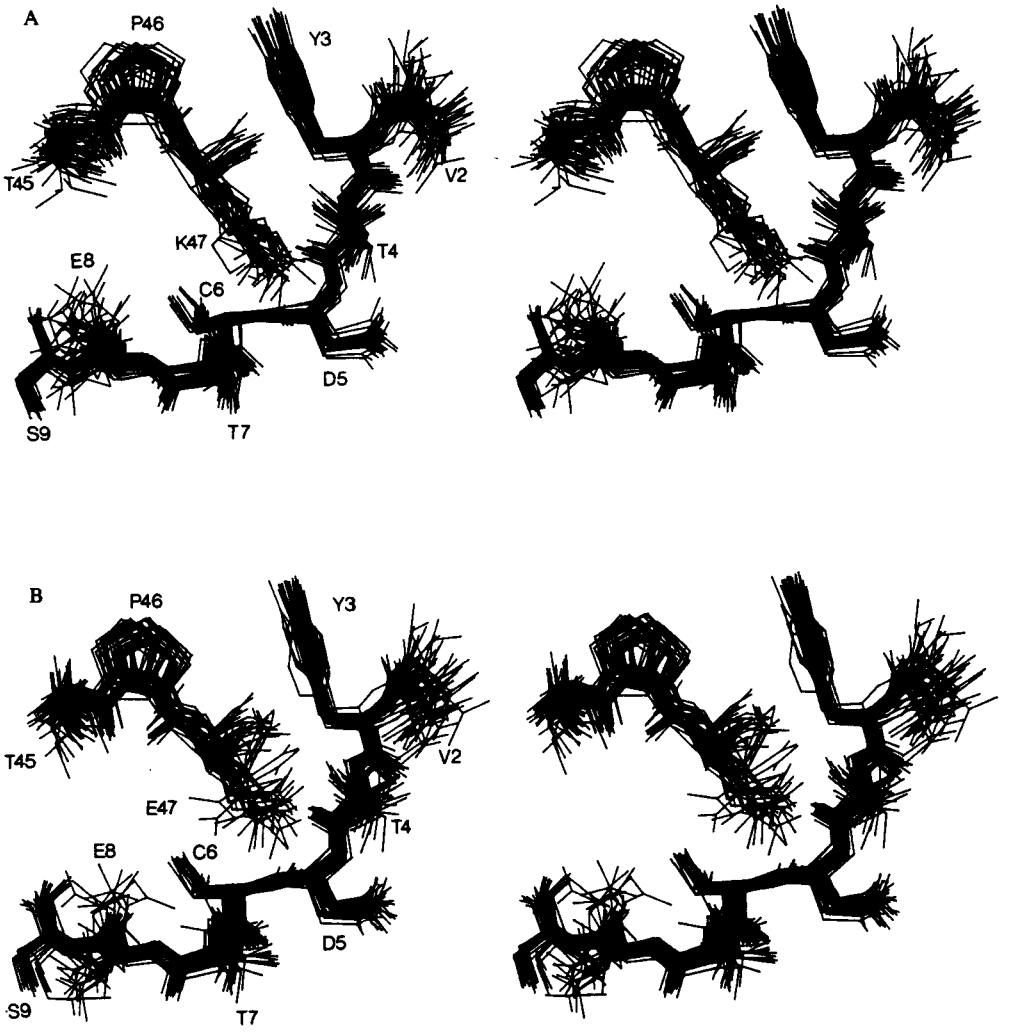


FIGURE 46. Best fit superposition of the 32 dynamical simulated annealing structures of wild type and Lys-47→Glu mutant hirudin for the region surrounding residue 47. (From Folkers, P. J. M., Clore, G. M., Driscoll, P. C., Dodt, J., Köhler, S., and Gronenborn, A. M., *Biochemistry*, 28, 2601, 1989. With permission.)

attractive and repulsive component seems to avoid distortions and ensures optimal base stacking interactions.[190] It is the view of the authors, therefore, that the best computational method at the present time for determining the 3D structures of oligonucleotides on the basis of NMR interproton distance data is restrained molecular dynamics. Like dynamical simulated annealing (cf. Section VIII.D), restrained molecular dynamics is a powerful method for locating the global minimum of the target function (cf. Section VIII.C). For oligonucleotide structures, it is essential to include a full empirical energy function that ensures that the local stereochemistry and nonbonded interactions are approximately correct, while the presence of effective potentials representing the experimental restraints (interproton distances, and torsion angles if available) drives the simulation into the global minimum region. In this manner, convergence to essentially the same structures, both locally and globally, can be achieved starting from both A- and B-DNA. Further, when the calculations are repeated using different random number seeds for the assignments of the initial velocities, it is found that the atomic rms differences between the final structures are independent of the starting structures.[133-135] These final structures all satisfy the

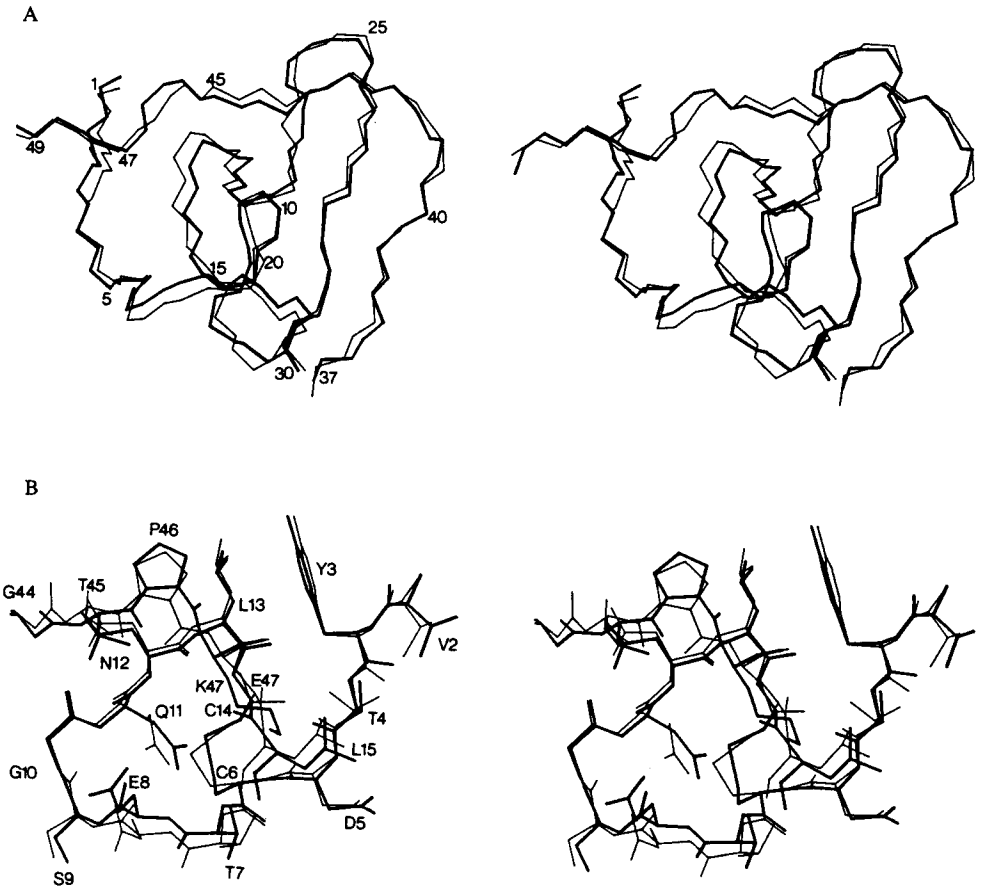


FIGURE 47. Superposition of the core (residues 1 to 30 and 37 to 49) of the restrained minimized mean structure of wild-type (thick lines) and Lys-47→Glu mutant (thin lines) hirudin. (From Folkers, P. J. M., Clore, G. M., Driscoll, P. C., Dodt, J., Köhler, S., and Gronenborn, A. M., *Biochemistry*, 28, 2601, 1989. With permission.)

NMR restraints within experimental error, display very small deviations from idealized covalent geometry, and have good non-bonded interactions as judged by large negative values for the Lennard-Jones van der Waals energy. In the absence of experimental restraints, however, convergence from A- and B-type starting structures does not occur.

The restrained molecular dynamics method has been applied to a number of DNA oligonucleotides,[131,132,135,190] as well as to a RNA oligonucleotide,[133] and a RNA/DNA hybrid.[134] The DNA oligonucleotides have all been found to have a B-type conformation, while the RNA and RNA/DNA hybrid oligonucleotides had A-type conformations. Stereoviews of the solution structures of the DNA dodecamer 5′d(CGCGPATTCGCG)$_2$ and the RNA hexamer 5′r(GCAUGC)$_2$ are shown in Figures 48 and 49, respectively. Interestingly, a number of common structural features have emerged from these studies. The most striking feature seen in the three DNA oligonucleotides analyzed in detail, namely, 5′d(GCATGC)$_2$, 5′d(CTGGATCCAG)$_2$, and 5′d(CGCGPATTCGCG)$_2$, is the behavior of the nonterminal CpG steps. They all have values of slide and roll in the A-type region of the slide-roll diagram, A-type interstrand stacking of the purines, and a decrease in helical twist. This tends to result in kinking of the DNA at CpG steps. The second most obvious common feature is the variation in helical twist, which appears to follow reasonably well, but by no means perfectly, the trend predicted by Calladine[231] and Dickerson.[232] Interestingly, the variation in helical twist observed for the the

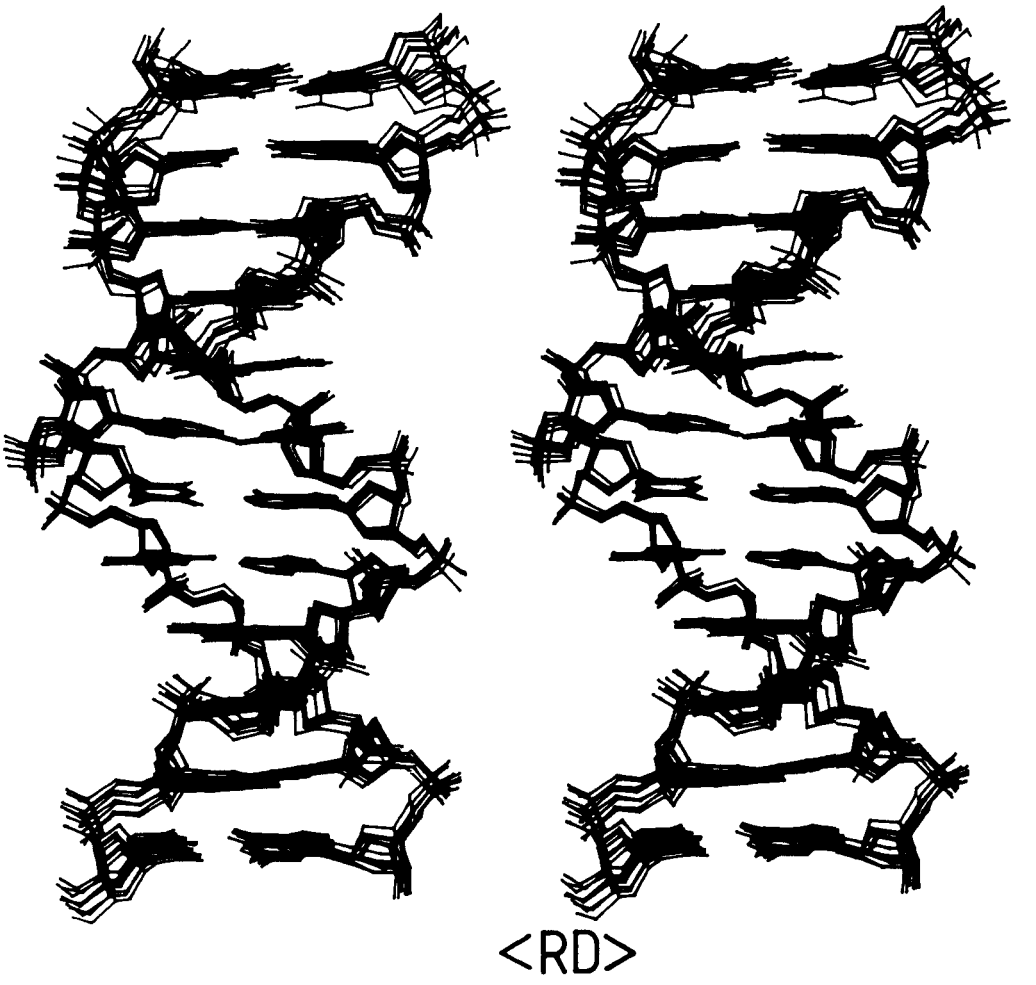


FIGURE 48. Superposition of eight restrained molecular dynamics structures for the oligonucleotide 5′d(CGCGPATTCGCG)$_2$ calculated on the basis of 228 interproton distance restraints. Four of the structures were obtained starting from a B-type structure and four from an A-type structure. (From Clore, G. M., Oschkinat, H., McLaughlin, L. W., Benseler, F., Scalfi Happ, C., Happ, E., and Gronenborn, A. M., *Biochemistry*, 27, 4185, 1988. With permission.)

RNA oligonucleotide 5′r(GCAUGC)$_2$ and the RNA/DNA hybrid (5′[r(GCA)d(TGC)]$_2$ is the opposite of what is observed for the analog DNA hexamer 5′d(GCATGC)$_2$. Namely, the RNA and RNA/DNA hybrid oligonucleotides are characterized by an increase in helical twist at the Pyr$_p$Pur steps, while the DNA oligonucleotide is characterized by a decrease in helical twist at these points.

Other features are more difficult to predict as they depend on the actual sequence of bases rather than the sequence written purely in terms of purines and pyrimidines. Thus, for example, the central six base pairs of the decamer 5′d(CTGGATCCAG)$_2$ and the dodecamer 5′d(CGCGPATTCGCG)$_2$ have the same sequence of purine and pyrimidines: Pur-Pur-Pur-Pyr-Pyr-Pyr. They differ, however, by the substitution of a G for a P at position 2 and of a C for a T at position 5. The sequence dependence of the helical parameters is similar but not identical, and these differences are sufficient to ensure that the central 6 base pairs of the decamer are straight, while those of the dodecamer are smoothly bent. Particularly gratifying is the finding that these features of the NMR structures are in agreement with the results on DNA curvature based on the observation of anomalous electrophoretic migration in polyacrylamide gels.[23,36]

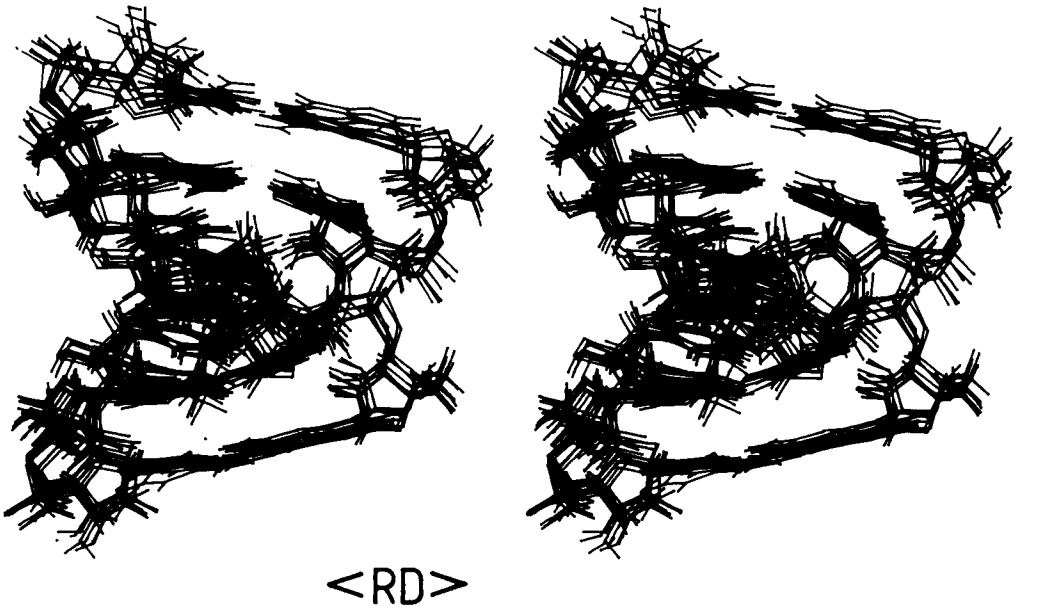


FIGURE 49. Superposition of eight restrained molecular dynamics structures for the oligoribonucleotide 5′r(CGAUGC)$_2$ calculated on the basis of 110 interproton distance restraints. Four of the structures were obtained starting from a B-type structure and four from an A-type structure. (From Scalfi Happ, C., Happ, E., Nilges, M., Gronenborn, A. M., and Clore, G. M., *Biochemistry*, 27, 1735, 1988. With permission.)

XI. PERSPECTIVES AND CONCLUDING REMARKS

It is clear from the above discussion that NMR is a powerful method of structure determination in solution. What are the limitations of this approach? At present it is limited to proteins of mol wt ≤20,000. Indeed, the three largest proteins whose 3D structures have been published to date are plastocyanin (99 residues),[38] the globular domain of histone H5 (79 residues),[15] and α-amylase inhibitor (74 residues).[8] Virtually complete assignments, however, have been made for a variety of larger systems using conventional homonuclear methods, in particular hen egg white lysozyme (129 residues)[181] and the lac repressor headpiece-operator complex (mol wt ~15,000).[234] In addition, progress using isotopic labeling has permitted a substantial number of assignments to be made for *Staphylococcal nuclease*, a protein of mol wt ~18,000.[85,117] Taking into account the development of new techniques, for example, 3D NMR combined with isotopic labeling, and the introduction of yet more powerful magnets, it may be possible to extend the molecular weight range up to proteins of mol wt ~40,000. This, however, will probably present a fundamental limit as the large linewidths of such proteins significantly reduce even the sensitivity of ^{1}H-detected heteronuclear correlation experiments.

It might be appropriate at this point to add a word of caution concerning the practical limits of structure determination by NMR. It is not always the number of residues of a particular protein that determines the feasibility of a NMR structure determination. Equally important is that the protein should be soluble, nonaggregating, and preferably stable up to at least 40°C, particularly for large proteins. A further important consideration is the chemical shift dispersion of the ^{1}H-NMR spectrum. This depends to a large extent on the structure of the protein under investigation. Proteins that are made up only of α-helices, loops, and turns invariably exhibit a fairly poor chemical shift dispersion, while the chemical shift dispersion in β-sheet proteins is usually very good.

A second potential problem may arise from the fact that although one region or domain of

a protein may be well-defined and therefore amenable to a NMR structure determination, other parts of the same protein may not be. This will be reflected in the absence of long-range tertiary NOEs for the ill-defined regions and leads to the fact that the positioning of these regions with respect to the rest of the protein cannot be achieved satisfactorily (e.g., the case of hirudin discussed in Section X.D.2). In this respect, it should be borne in mind that metric matrix distance geometry algorithms tend to obscure such ill-defined regions and generate structures whose atomic rms distribution is smaller than is warranted on the basis of the experimental data. For this reason, it is advisable to subject such distance geometry structures to either restrained molecular dynamics or simulated annealing, thereby ensuring that the conformational space consistent with the experimental data is properly sampled. In addition, it is necessary to calculate at least 20 structures with the same experimental data set in order to obtain a good representation of the NMR structure of a protein, as it is only by analyzing a reasonably large family of structures that the local and global definition of a structure can be assessed. Choosing only five to ten structures, as has often been the case in the past, may seriously bias the interpretation, as this number is too small to assess whether all possible structures that are compatible with the experimental data have been adequately considered.

X-ray crystallography, of course, also has its limitations, the most obvious being the requirement for a protein to crystallize, yielding suitable crystals that diffract to high resolution and of heavy atom derivatives to solve the phase problem. In practice, therefore, it seems likely that there will be relatively few proteins that will be amenable to both NMR and X-ray crystallography. In such cases, the information afforded by NMR and crystallography is clearly complementary, and the structures obtained by NMR could potentially be used to solve the X-ray structures directly by molecular replacement.

Finally, it is worth emphasizing that in addition to being able to determine 3D structures of proteins, NMR has the potential to address questions concerning the dynamics of the system under investigation. Thus, obtaining a structure by NMR may not be the end point of a particular study, but rather a starting point for investigating such phenomena as dynamics of conformational changes upon ligand binding, unfolding kinetics, conformational equilibria between different conformational states, fast internal dynamics on the nanosecond time scale and below, slow internal dynamics on the second and millisecond time scales, and a whole array of other questions that can be asked in light of the previously determined NMR structure.

ACKNOWLEDGMENTS

We wish to thank our numerous collaborators over the years who have participated in various aspects of the work described in this review, in particular Ad Bax, Axel Brünger, Paul Driscoll, Paul Folkers, Martin Karplus, Michael Nilges, Lennart Nilsson, Flemming Poulsen, Mogens Kjaer, Dinesh Sukumaran, and Jutta Zarbock. Work from the author's laboratory was supported by the Max-Planck Society, the Bundesministerium für Forschung und Technologie (Grant 321-400310318909A), the Deutsche Forschungsgeneinschaft (Grants C1 86/1-1 and Gr 658/4-1), the National Institutes of Health, and the Intramural AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health.

ABBREVIATIONS USED

NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, two-dimensional nuclear Overhauser spectroscopy; ROE, rotating frame nuclear Overhauser effect; ROESY, two-dimensional rotating frame nuclear Overhauser spectroscopy; COSY; homonuclear correlated spectroscopy; DQF-COSY; double quantum filtered COSY; PS-COSY, pseudo single quantum COSY; E-COSY, exclusive COSY; PE-COSY, primitive exclusive COSY; HOHAHA, homonuclear Hartmann-Hahn spectroscopy; HMQC, heteronuclear multiple quan-

tum coherence spectroscopy; HMBC, heteronuclear multiple bond correlation spectroscopy; rms, root mean square; RD, restrained dynamics; SA, simulated annealing.

REFERENCES

1. **Ernst, R. R., Bodenhausen, G., and Wokaun, A.,** *Principles of Nuclear Magnetic Resonance in One and Two Dimensions,* Clarendon Press, Oxford, 1986.
2. **Braun, W., Wider, G., Lee, K. H., and Wüthrich, K.,** Conformation of glucagon in a lipid-water interphase by ^{1}H nuclear magnetic resonance, *J. Mol. Biol.,* 169, 921, 1983.
3. **Arseniev, S. A., Kondakov, V. I., Maiorov, V. N., and Bystrov, V. F.,** NMR solution spatial structure of "short" scorpion insectotoxin I_5A, *FEBS Lett.,* 165, 57, 1984.
4. **Williamson, M. P., Havel, T. F., and Wüthrich, K.,** Solution conformation of proteinase inhibitor IIA from bull seminal plasma by ^{1}H nuclear magnetic resonance and distance geometry, *J. Mol. Biol.,* 182, 295, 1985.
5. **Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R., and van Gunsteren, W. F.,** A protein structure from nuclear magnetic resonance data: lac repressor headpiece, *J. Mol. Biol.,* 182, 179, 1985.
6. **Clore, G. M., Gronenborn, A. M., Brünger, A. T., and Karplus, M.,** The solution conformation of a heptadecapeptide comprising the DNA binding helix F of the cyclic AMP receptor protein of *Escherichia coli*: combined use of ^{1}H-nuclear magnetic resonance and restrained molecular dynamics, *J. Mol. Biol.,* 186, 435, 1985.
7. **Clore, G. M., Nilges, M., Sukumaran, D. K., Brünger, A. T., Karplus, M., and Gronenborn, A. M.,** The three-dimensional structure of α1-purothionin in solution: combined use of nuclear magnetic resonance, distance geometry and restrained molecular dynamics, *EMBO J.,* 5, 2729, 1986.
8. **Kline, A. D., Braun, W., and Wüthrich, K.,** Studies by ^{1}H nuclear magnetic resonance and distance geometry of the solution conformation of the α-amylase inhibitor tendamistat, *J. Mol. Biol.,* 189, 377, 1986; **Kline, A. D., Braun, W., and Wüthrich, K.,** Determination of the complete three-dimensional structure of the α-amylase inhibitor tendamistat in aqueous solution by nuclear magnetic resonance and distance geometry, *J. Mol. Biol.,* 204, 675, 1988.
9. **Braun, W., Wagner, G., Wörgötter, E., Vašák, M., Kägi, J. H. R., and Wüthrich, K.,** Polypeptide fold in the two metal clusters of metallothionein-2 by nuclear magnetic resonance and distance geometry, *J. Mol. Biol.,* 187, 125, 1986.
10. **Arseniev, A., Schultze, P., Wörgötter, E., Braun, W., Wagner, G., Vašák, M., Kägi, J. H. R., and Wüthrich, K.,** Three-dimensional structure of rabbit liver (Cd_7)metallothionein-2a in aqueous solution determined by nuclear magnetic resonance, *J. Mol. Biol.,* 201, 637, 1988.
11. **Schultze, P., Wörgötter, E., Braun, W., Wagner, G., Vasák, M., Kägi, M., and Wüthrich, K.,** Conformation of [Cd_7] metallothionein-2 from rat liver in aqueous solution determined by nuclear magnetic resonance, *J. Mol. Biol.,* 203, 251, 1988.
12. **Clore, G. M., Sukumaran, D. K., Nilges, M., and Gronenborn, A. M.,** The three-dimensional structure of phoratoxin in solution: combined use of nuclear magnetic resonance, distance geometry and restrained molecular dynamics, *Biochemistry,* 26, 1732, 1987.
13. **Clore, G. M., Sukumaran, D. K., Nilges, M., Zarbock, J., and Gronenborn, A. M.,** The conformations of hirudin in solution: a study using nuclear magnetic resonance, distance geometry and restrained molecular dynamics, *EMBO J.,* 6, 529, 1987.
14. **Folkers, P. J. M., Clore, G. M., Driscoll, P. C., Dodt, J., Köhler, S., and Gronenborn, A. M.,** The solution structure of recombinant hirudin and the Lys-47→Glu mutant: a nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing study, *Biochemistry,* 28, 2601, 1989.
15. **Clore, G. M., Gronenborn, A. M., Nilges, M., Sukumaran, D. K., and Zarbock, J.,** The polypeptide fold of the globular domain of histone H5 in solution: a study using nuclear magnetic resonance, distance geometry and restrained molecular dynamics, *EMBO J.,* 6, 1833, 1987.
16. **Nilges, M., Gronenborn, A. M., and Clore, G. M.,** Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations, *FEBS Lett.,* 229, 317, 1988.
17. **Cooke, R. M., Wilkinson, A. J., Baron, M., Pastore, A., Tappin, M. J., Campbell, I. D., Gregory, H., and Sheard, B.,** The solution structure of human epidermal growth factor, *Nature (London),* 327, 339, 1987.
18. **Koda, D., Go, N., Hayashi, K., and Inagaki, F.,** Tertiary Structure of mouse epidermal growth factor determined by two-dimensional NMR, *J. Biochem.,* 103, 741, 1988.

19. **Montelione, G. T., Wüthrich, K., Nice, E. C., Burgess, A. W., and Scheraga, H. A.,** Solution structure of murine epidermal growth factor: determination of the polypeptide backbone chain-fold by nuclear magnetic resonance and distance geometry, *Proc. Natl. Acad. Sci. U.S.A.*, 84, 5226, 1987.
20. **Wagner, G., Braun, W., Havel, T. F., Schaumann, T., Go, N., and Wüthrich, K.,** Protein structures in solution by nuclear magnetic resonance and distance geometry. The polypeptide fold of the basic pancreatic trypsin inhibitor determined using two different algorithms, DISGEO and DISMAN, *J. Mol. Biol.*, 196, 611, 1987.
21. **Clore, G. M., Gronenborn, A. M., Nilges, M., and Ryan, C. A.,** The three-dimensional structure of potato carboxypeptidase inhibitor in solution: a study using nuclear magnetic resonance, distance geometry and restrained molecular dynamics, *Biochemistry*, 26, 8012, 1987.
22. **Nilges, M., Gronenborn, A. M., Brünger, A. T., and Clore, G. M.,** Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints: application to crambin, potato carboxypeptidase inihibitor and barley serine proteinase inhibitor 2, *Protein Eng.*, 2, 27, 1988.
23. **Nilges, M., Clore, G. M., and Gronenborn, A. M.,** Determination of three-dimensional structures of proteins from interproton distance data by dynamical simulated annealing from a random array of atoms, *FEBS Lett.*, 239, 129, 1988.
24. **Clore, G. M., Gronenborn, A. M., Kjær, M., and Poulsen, F. M.,** The determination of the three-dimensional structure of barley serine proteinase inhibitor 2 by nuclear magnetic resonance, distance geometry and restrained molecular dynamics, *Protein Eng.*, 1, 305, 1987.
25. **Brünger, A. T., Clore, G. M., Gronenborn, A. M., and Karplus, M.,** Solution conformations of human growth hormone releasing factor: comparison of the restrained molecular dynamics and distance geometry methods for a system without long-range distance data, *Protein Eng.*, 1, 399, 1987.
26. **Clore, G. M., Nilges, M., Brünger, A. T., and Gronenborn, A. M.,** Determination of the backbone conformation of secretin by restrained molecular dynamics on the basis of interproton distance data, *Eur. J. Biochem.*, 171, 479, 1988.
27. **Holak, T. A., Prestegard, J. H., and Forman, J. D.,** NMR-pseudoenergy approach to the solution structure of acyl carrier protein, *Biochemistry*, 26, 4652, 1987.
28. **Holak, T. A., Nilges, M., Prestegard, J. H., Gronenborn, A. M., and Clore, G. M.,** Three-dimensional structure of acyl carrier protein in solution determined by nuclear magnetic resonance and combined use of dynamical simulated annealing and distance geometry, *Eur. J. Biochem.*, 175, 9, 1988.
29. **Holak, T. A., Kim, Y., Kearsley, S. K., and Prestegard, J. H.,** The three-dimensional structure of acyl carrier protein determined by NMR-pseudoenergy and distance geometry calculations, *Biochemistry*, 27, 1635, 1988.
30. **Holak, T. A., Engström, Å., Kraulis, P. J., Lindeberg, G., Bennich, H., Jones, T. A., Gronenborn, A. M., and Clore, G. M.,** The solution conformation of the antibacterial peptide cecropin A: a nuclear magnetic resonance and dynamical simulated annealing study, *Biochemistry*, 27, 7620, 1988.
31. **Tappin, M. J., Pastore, A., Norton, R. S., Freer, J. H., and Campbell, I. D.,** High resolution ^{1}H-NMR study of the solution structure of δ-hemolysin, *Biochemistry*, 27, 1643, 1988.
32. **Driscoll, P. C., Gronenborn, A. M., Beress, L., and Clore, G. M.,** Determination of the three-dimensional solution structure of the anti-hypertensive and anti-viral protein BDS-I from the sea anemone Anemonia sulcata: a study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing, *Biochemistry*, in press.
33. **Zarbock, J., Gennaro, R., Romeo, D., Clore, G. M., and Gronenborn, A. M.,** A proton nuclear magnetic resonance study of the conformation of bovine anaphylatoxin C5a in solution, *FEBS Lett.*, 238, 289, 1988.
34. **Zuiderweg, E. R. P., Henkin, J., Mollison, K. W., and Greer, J.,** Comparison of model and nuclear magnetic resonance structures for the human inflammatory protein C5a, *Proteins Struct. Function Genet.*, 3, 139, 1988; **Zuiderweg, E. R. P., Nettesheim, D. G., Mollison, K. W., and Carter, C. W.,** Tertiary structure of human complement component C5a in solution from nuclear magnetic resonance data, *Biochemistry*, 28, 172, 1989.
35. **Bazzo, R., Tappin, M. J., Pastore, A., Harvey, T. S., Carver, J. A., and Campbell, I. D.,** The structure of melittin, *Eur. J. Biochem.*, 173, 139, 1988.
36. **Steinmetz, W. E., Bougis, P. E., Rochat, H., Redwine, O. D., Braun, W., and Wüthrich, K.,** ^{1}H nuclear magnetic resonance studies of the three-dimensional structure of the cardiotoxin CTXIIb from Naja mossambica in aqueous solution and comparison with the crystal structures of homologous toxins, *Eur. J. Biochem.*, 172, 101, 1988.
37. **Pardi, A., Hare, D. R., Selsted, M. E., Morrison, R. D., Bassolino, D. A., and Bach, A. C., II,** Solution structures of the rabbit neutrophil defensin NP-5, *J. Mol. Biol.*, 201, 625, 1988.
38. **Moore, J. M., Case, D. A., Chazin, W. J., Gippert, G. P., Havel, T. F., Powls, R., and Wright, P. E.,** Three-dimensional solution structure of plastocyanin from the green alga *Scenedesmus obliquus*, *Science*, 240, 314, 1988.
39. **Olejniczak, E. T., Gampe, R. T., Rockway, T. W., and Fesik, S. W.,** NMR study of the solution conformation of rat atrial natruiretic factor 7-23 in sodium dodecyl sulfate micelles, *Biochemistry*, 27, 7124, 1988.
40. **Havel, T. F. and Wüthrich, K.,** An evaluation of the combined use of nuclear magnetic resonance and distance geometry for the determination of protein conformations in solution, *J. Mol. Biol.*, 182, 281, 1985.

41. **Braun, W. and Go, N.,** Calculation of protein conformations by proton-proton distance constraints. A new efficient algorithm, *J. Mol. Biol.*, 186, 611, 1985.
42. **Vásquez, M. and Scheraga, H. A.,** Calculation of protein conformation by the built up procedure. Application to bovine pancreatic trypsin inhibitor using limited simulated nuclear magnetic resonance data, *J. Biomol. Struct. Dyn.*, 5, 705, 1988.
43. **Vásquez, M. and Scheraga, H. A.,** Variable target function and built up procedures for the calculation of protein conformation. Application to bovine pancreatic trypsin inhibitor using limited simulated nuclear magnetic resonance data, *J. Biomol. Struct. Dyn.*, 5, 757, 1988.
44. **Brünger, A. T., Clore, G. M., Gronenborn, A. M., and Karplus, M.,** Three-dimensional structure of proteins determined by molecular dynamics with interproton distance restraints: application to crambin, *Proc. Natl. Acad. Sci. U.S.A.*, 83, 3801, 1986.
45. **Clore, G. M., Brünger, A. T., Karplus, M., and Gronenborn, A. M.,** Application of molecular dynamics with interproton distance restraints to three-dimensional protein structure determination: a model study of crambin, *J. Mol. Biol.*, 191, 523, 1986.
46. **Clore, G. M., Nilges, M., Brünger, A. T., Karplus, M., and Gronenborn, A. M.,** A comparison of the restrained molecular dynamics and distance geometry methods for determining three-dimensional structures of proteins on the basis of interproton distance restraints, *FEBS Lett.*, 213, 269, 1987.
47. **States, D. J., Haberkorn, R. A., and Ruben, D. J.,** A two-dimensional nuclear Overhauser experiment with pure absorption phase in four quadrants, *J. Magn. Reson.*, 48, 286, 1982.
48. **Marion, D. and Wüthrich, K.,** Application of phase sensitive correlated spectroscopy for measurement of proton-proton spin-spin coupling constants in proteins, *Biochem. Biophys. Res. Commun.*, 113, 967, 1983.
49. **Overhauser, A.,** Paramagnetic relaxation in metals, *Phys. Rev.*, 89, 689, 1953.
50. **Overhauser, A.,** Polarization of nuclei in metals, *Phys. Rev.*, 92, 411, 1953.
51. **Solomon, I.,** Relaxation processes in a system of two spins, *Phys. Rev.*, 99, 559, 1955.
52. **Noggle, J. H. and Schirmer, R. E.,** *The Nuclear Overhauser Effect - Chemical Applications*, Academic Press, New York, 1971.
53. **Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R.,** Investigation of exchange processes by two-dimensional NMR spectroscopy, *J. Chem. Phys.*, 71, 4546, 1979.
54. **Macura, S. and Ernst, R. R.,** Elucidation of cross-relaxation in liquids by two-dimensional NMR spectroscopy, *Mol. Phys.*, 41, 95, 1980.
55. **Macura, S., Huang, Y., Suter, D., and Ernst, R. R.,** Two-dimensional chemical exchange and cross-relaxation spectroscopy of coupled nuclear spins, *J. Magn. Reson.*, 43, 259, 1981.
56. **Wüthrich, K.,** *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York, 1986.
57. **Aue, W. P., Batholdi, E., and Ernst, R. R.,** Two-dimensional spectroscopy. Application to nuclear magnetic resonance, *J. Chem. Phys.*, 64, 2229, 1976.
58. **Rance, M., Sørenson, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., and Wüthrich, K.,** Improved spectral resolution in COSY ^{1}H-NMR spectra of proteins via double quantum filtering, *Biochem. Biophys. Res. Commun.*, 117, 479, 1983.
59. **Marion, D. and Bax, A.,** P.COSY: a sensitive alternative for double quantum filtered COSY, *J. Magn. Reson.*, 79, 352, 1988.
60. **Bax, A., Kay, L. E., Sparks, S. W., and Torchia, D.,** Line narrowing of amide proton resonances in 2D NMR spectra of proteins, *J. Am. Chem. Soc.*, 111, 408, 1989.
61. **Davis, D. G. and Bax, A.,** Assignment of complex ^{1}H NMR spectra via two-dimensional homonuclear Hartmann-Hahn spectroscopy, *J. Am. Chem. Soc.*, 107, 2821, 1985.
62. **Bax, A. and Davis, D. G.,** MLEV-17 based two-dimensional homonuclear magnetization transfer spectroscopy, *J. Magn. Reson.*, 65, 355, 1985.
63. **Bax, A.,** Homonuclear Hartmann-Hahn experiments, in *Methods in Enzymology*, Vol. 176, James, T. L. and Oppenheimer, N., Eds., Academic Press, New York, 1989, 151.
64. **Bax, A., Sklenar, V., Gronenborn, A. M., and Clore, G. M.,** Water suppression in two-dimensional spin-locked NMR experiments using a novel phase cycling procedure, *J. Am. Chem. Soc.*, 109, 6511, 1987.
66. **Wüthrich, K., Wider, G., Wagner, G., and Braun, W.,** Sequential resonance assignments as a basis for determination of spatial protein structures by high resolution proton nuclear magnetic resonance spectroscopy, *J. Mol. Biol.*, 155, 311, 1982.
67. **Billeter, M., Braun, W., and Wüthrich, K.,** Sequential resonance assignments in protein ^{1}H nuclear magnetic resonance spectra: computation of sterically allowed proton-proton distances in single crystal protein conformations, *J. Mol. Biol.*, 155, 321, 1982.
68. **Wagner, G. and Wüthrich, K.,** Sequential resonance assignments in protein ^{1}H NMR spectra: basic pancreatic trypsin inhibitor, *J. Mol. Biol.*, 155, 347, 1982.
69. **Wüthrich, K., Billeter, M., and Braun, W.,** Polypeptide secondary structure determination by nuclear magnetic resonance observation of short proton-proton distances, *J. Mol. Biol.*, 180, 715, 1984.
70. **Mueller, L.,** Sensitivity enhanced detection of weak nuclei using heteronuclear multiple quantum coherence, *J. Am. Chem. Soc.*, 101, 4481, 1979.

71. **Redfield, A. G.,** Stimulated echo NMR spectra and their use for heteronuclear two-dimensional shift correlation, *Chem. Phys. Lett.*, 96, 537, 1983.
72. **Bax, A., Griffey, R. H., and Hawkins, B. L.,** Sensitivity enhanced correlation of ^{15}N and ^{1}H chemical shifts in natural abundance samples via multiple quantum coherence, *J. Am. Chem. Soc.*, 105, 7188, 1983.
73. **Bolton, P. H. and Bodenhausen, G.,** Relayed coherence transfer spectroscopy of heteronuclear systems: detection of remote nuclei in NMR, *Chem. Phys. Lett.*, 89, 139, 1982.
74. **Sklenar, V. and Bax, A.,** Two-dimensional heteronuclear chemical shift correlation of proteins at natural abundance ^{15}N and ^{13}C levels, *J. Magn. Reson.*, 71, 379, 1987.
75. **Griffey, R. H. and Redfield, A. G.,** Proton-detected heteronuclear edited and correlated nuclear magnetic resonance and nuclear Overhauser effect in solution, *Q. Rev. Biophys.*, 19, 51, 1987.
76. **Gronenborn, A. M., Bax, A., Wingfield, P. T., and Clore, G. M.,** A powerful method of sequential proton resonance assignment in proteins using relayed ^{15}N-^{1}H multiple quantum coherence spectroscopy, *FEBS Lett.*, 243, 93, 1989.
77. **Bax, A. and Summers, M. F.,** ^{1}H and ^{13}C assignments from sensitivity enhanced detection of multiple bond CH connectivity by two-dimensional multiple quantum NMR, *J. Am. Chem. Soc.*, 108, 2093, 1986.
78. **Bax, A. and Marion, D.,** Improved resolution and sensitivity in ^{1}H detected heteronuclear multiple bond correlation spectroscopy, *J. Magn. Reson.*, 78, 186, 1988.
79. **Clore, G. M., Bax, A., Wingfield, P. T., and Gronenborn, A. M.,** Long range ^{15}N-^{1}H correlation as an aid to sequential proton resonance assignment of proteins: application to the DNA binding protein Ner from phage Mu, *FEBS Lett.*, 238, 17, 1988.
80. **Griesinger, C., Sørensen, O. W., and Ernst, R. R.,** A practical approach to three-dimensional NMR spectroscopy, *J. Magn. Reson.*, 73, 574, 1987.
81. **Griesinger, C., Sørensen, O. W., and Ernst, R. R.,** Novel three-dimensional NMR techniques for studies of peptides and biological macromolecules, *J. Am. Chem. Soc.*, 109, 7227, 1987.
82. **Oschkinat, H., Griesinger, C., Kraulis, P. J., Sørensen, O. W., Ernst, R. R., Gronenborn, A. M., and Clore, G. M.,** Three-dimensional NMR spectroscopy of a protein in solution, *Nature (London)*, 332, 374, 1988.
83. **Oschkinat, H., Cieslar, C., Gronenborn, A. M., and Clore, G. M.,** Three-dimensional homonuclear Hartmann Hahn-nuclear Overhauser enhancement spectroscopy in H_2O and its applications to proteins, *J. Magn. Reson.*, 81, 212, 1989; **Vuister, G. W., Boelus, R., and Kaptein, R.,** Non-selective three-dimensional NMR spectroscopy: the 30 HOHANA-NOESY experiment, *J. Magn. Reson.*, 80, 176, 1988.
84. **Fesik, S. W. and Zuiderweg, E. R. P.,** Heteronuclear three-dimensional NMR spectroscopy: a strategy for the simplification of homonuclear two-dimensional spectra, *J. Magn. Reson.*, 78, 588, 1988.
85. **Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A., and Bax, A.,** Three-dimensional heteronuclear NMR of ^{15}N labelled proteins, *J. Am. Chem. Soc.*, 111, 1515, 1989.
86. **Reid, D. G., Salisbury, S. A., Bellard, S., Shakked, Z., and Williams, D. H.,** Proton nuclear Overhauser effect study of the structure of a deoxyoligonucleotide duplex in aqueous solution, *Biochemistry*, 22, 2019, 1983.
87. **Feigon, J., Wright, J. M., Leupin, W., Denny, W. A., and Kearns, D. R.,** Use of two-dimensional NMR in the study of a double stranded decamer, *J. Am. Chem. Soc.*, 104, 5540, 1982.
88. **Feigon, J., Leupin, W., Denny, W. A., and Kearns, D. R.,** Two-dimensional proton nuclear magnetic resonance investigation of the synthetic deoxyribonucleic acid decamer $d(ATATCGATAT)_2$, *Biochemistry*, 22, 5943, 1983.
89. **Clore, G. M. and Gronenborn, A. M.,** Sequence dependent structural variations in two right handed alternating pyrimidine-purine DNA oligonucleotides in solution determined by NOE measurements, *EMBO J.*, 2, 2109, 1983.
90. **Scheek, R. M., Boelens, R., Russo, N., and Kaptein, R.,** Sequential resonance assignment in DNA ^{1}H-NMR spectra by two-dimensional NOE spectroscopy, *J. Am. Chem. Soc.*, 105, 2914, 1983.
91. **Hare, D. R., Wemmer, D. E., Chou, S. H., Drobny, G., and Reid, B. R.,** Assignment of the non-exchangeable proton resonances of d(CGCGAATTCGCG) using two-dimensional nuclear magnetic resonance methods, *J. Mol. Biol.*, 171, 319, 1983.
92. **Haasnoot, C. A. G., Westerink, H. P., van der Marel, G. A., and van Boom, J. H.,** Conformational analysis of a hybrid DNA-RNA double helical oligonucleotide in aqueous solution: d(CG)r(CG)d(GC) studied by 1-D and 2-D ^{1}H-NMR spectroscopy, *J. Biomol. Struct. Dyn.*, 1, 31, 1983.
93. **Haasnoot, C. A. G., Westerink, H. P., van der Marel, G. A., and van Boom, J. H.,** Discrimination between A-type and B-type conformations of nucleic acid fragments in solution by means of two-dimensional nuclear Overhauser experiments, *J. Biomol. Ster. Dyn.*, 2, 345, 1984.
94. **Weiss, M. A., Patel, D. J., Sauer, R. T., and Karplus, M.,** Two-dimensional ^{1}H NMR study of the λ operator site OL1: a sequential assignment and its applications, *Proc. Natl. Acad. Sci. U.S.A.*, 81, 130, 1984.
95. **Gronenborn, A. M. and Clore, G. M.,** Investigation of the solution structures of short nucleic acid fragments by means of NOE measurements, *Prog. Nucl. Magn. Reson. Spectrosc.*, 17, 1, 1985.
96. **Clore, G. M. and Gronenborn, A. M.,** Probing the three-dimensional structures of DNA and RNA oligonucleotides in solution by nuclear Overhauser enhancement measurements, *FEBS Lett.*, 179, 187, 1985.

97. **Gronenborn, A. M., Clore, G. M., and Kimber, B. J.,** An investigation into the solution structures of d(CGTACG) and d(ACGCGCGT) by means of nuclear Overhauser effect measurements, *Biochem. J.*, 221, 723, 1984.
98. **Clore, G. M., Gronenborn, A. M., Piper, E. A., McLaughlin, L. W., Gaeser, E., and van Boom, J. H.,** The solution structure of a RNA pentadecamer comprising the anticodon loop and stem of $tRNA^{Phe}$: a 500 MHz ^1H-NMR study, *Biochem. J.*, 221, 735, 1984.
99. **Roy, S. and Redfield, A. G.,** NOE study and assignment of D stem and reverse Hoogsteen base pair resonances in $tRNA^{asp}$, *Nucl. Acids Res.*, 9, 7073, 1981.
100. **Roy, S. and Redfield, A. G.,** Assignment of imino proton resonance spectra of yeast $tRNA^{phe}$, *Biochemistry*, 22, 1386, 1983.
101. **Hare, D. R. and Reid, B. R.,** Nuclear Overhauser assignment of the imino protons of the acceptor helix and the ribothymidine helix in nuclear magnetic resonance spectrum of *Escherichia coli* isoleucine transfer ribonucleic acid: evidence for costacked helices in solution, *Biochemistry*, 21, 5129, 1982
102. **Gronenborn, A. M., Clore, G. M., Jones, M. B., and Jiricny, J.,** A nuclear Overhauser enhancement study of the imino proton resonances of a DNA pentadecamer comprising the specific target site of the cyclic AMP receptor protein in the ara BAD operon, *FEBS Lett.*, 165, 216, 1984.
103. **Chou, S. H., Hare, D. R., Wemmer, D. E., and Reid, B. R.,** Sequence-specific recognition of deoxyribonucleic acid. Chemical synthesis and nuclear magnetic resonance assignment of the imino protons of λ O_{R3} operator deoxyribonucleic acid, *Biochemistry*, 22, 3037, 1983.
104. **Ulrich, E. L., John, E. M. M., Gough, G. R., Brunden, M. J., Gilham, P. T., Westler, W. M., and Markley, J. L.,** Imino proton assignments in the proton nuclear magnetic resonance spectrum of the λ phage O_{R3} deoxyribonucleic acid fragment, *Biochemistry*, 22, 4362, 1983.
105. **Weiss, M. A., Patel, D. J., Sauer, R. T., and Karplus, M.,** Proton NMR study of the λ operator site OL1: assignment of the imino and adenine H2 resonances, *Nucl. Acids Res.*, 12, 4035, 1984.
106. **Boelens, R., Scheek, R. M., Dijkstra, K., and Kaptein R.,** Sequential assignment of imino- and amino-proton resonances in proton NMR spectra of oligonucleotides by two-dimensional NMR spectroscopy: application to a lac operator fragment, *J. Magn. Reson.*, 62, 378, 1985.
107. **Heerschap, A., Mellema, J. R., Janssen, H. G., Walters, J. A., Haasnoot, C. A., and Hilbers, C. W.,** Imino proton resonances of yeast $tRNA^{phe}$ studied by 2D NOE spectroscopy, *Eur. J. Biochem.*, 149, 649, 1985.
108. **Hare, D. R., Ribeiro, N. S., Wemmer, D. E., and Reid, B. R.,** Complete assignment of the imino protons of *Escherichia coli* valine transfer RNA: two-dimensional NMR studies in water, *Biochemistry*, 24 4300, 1985.
109. **LeMaster, D. M. and Richards, F. M.,** NMR sequential assignment of *E. coli* thioredoxin utilizing random fractional deuteration, *Biochemistry*, 27, 142, 1988.
110. **LeMaster, D. M.,** Chiral β and random fractional deuteration for the determination of protein sidechain conformation by NMR, *FEBS Lett.*, 223, 191, 1987.
111. **Torchia, D. A., Sparks, S. A., and Bax, A.,** Delineation of alpha-helical domains in deuteriated *Staphylococcal nuclease* by 2D NOE NMR spectroscopy, *J. Am. Chem. Soc.*, 110, 2320, 1988.
112. **Bax, A. and Weiss, M.,** Simplification of two-dimensional NOE spectra of proteins by ^{13}C labelling, *J. Magn. Reson.*, 71, 571, 1987.
113. **Fesik, S. W., Gampe, R. T., and Rockway, T. W.,** Application of isotope-filtered 2D NOE experiments in the conformational analysis of atrial natriuretic factor (7-23), *J. Magn. Reson.*, 74, 366, 1987.
114. **Griffey, R. H., Jarema, M. A., Kunz, S., Rosevear, P. R., and Redfield, A. G.,** Isotopic-label directed observation of the nuclear Overhauser effect in poorly resolved proton NMR spectra, *J. Am. Chem. Soc.*, 107, 711, 1986.
115. **Weiss, M. A., Redfield, A. G., and Griffey, R. H.,** Isotope-detected ^1H NMR studies of proteins: a general strategy for editing interproton nuclear Overhauser effects by heteronuclear decoupling, with application to λ phage repressor, *Proc. Natl. Acad. Sci. U.S.A.*, 83, 1325, 1986.
116. **McIntosh, L. P., Griffey, R. H., Muchmore, D. G., Nelson, C. P., Redfield, A. G., and Dahlquist, F. W.,** Proton NMR measurements of bacteriophage T_4 lysosyme aided by ^{13}C and ^{15}N isotopic labelling: structural and dynamic studies of large proteins by NMR, *Proc. Natl. Acad. Sci. U.S.A.*, 84, 1244, 1987.
117. **Torchia, D. A., Sparks, S. W., and Bax, A.,** NMR signal assignments of amide protons in the alpha-helical domains of staphylococcal nuclease, *Biochemistry*, 27, 5135, 1988.
118. **Markley, J. L., Croll, D. H., Krishnamoorthi, R., Ortiz-Polo, G., Westler, W. M., Bogard, W. C., and Laskowski, M.,** One- and two-dimensional NMR spectral analysis of the consequences of single amino acid replacements in proteins, *J. Cell. Biochem.*, 30, 291, 1986.
119. **Gronenborn, A. M., Clore, G. M., Schmeissner, U., and Wingfield, P. T.,** Sequence specific assignments of aromatic residues in the ^1H-NMR spectrum of human interleukin-1β using site directed mutant proteins, *Eur. J. Biochem.*, 161, 37, 1986.
120. **Gronenborn, A. M., Wingfield, P. T., McDonald, H. R., Schmeissner, U., and Clore, G. M.,** Site directed mutants of human interleukin-1α: a ^1H-NMR and receptor binding study, *FEBS Lett.*, 231, 135, 1988.
121. **Wagner, G. and Wüthrich, K.,** Truncated driven nuclear Overhauser effect (TOE). A new technique for studies of selective ^1H-^1H Overhauser effects in the presence of spin diffusion, *J. Magn. Reson.*, 33, 675, 1979.

122. **Dobson, C. M., Olejniczak, E. T., Poulsen, F. M., and Ratcliffe, R. G.,** Time development of proton nuclear Overhauser effects in proteins, *J. Magn. Reson.*, 48, 97, 1982.
123. **Keepers, J. W. amd James, T. L.,** A theoretical study of distance determination from NMR: two-dimensional nuclear Overhauser effect spectra, *J. Magn. Reson.*, 57, 404, 1984.
124. **Clore, G. M. and Gronenborn, A. M.,** Assessment of errors involved in the determination of interproton distance ratios and distances by means of one- and two-dimensional NOE measurements, *J. Magn. Reson.*, 61, 158, 1985.
125. **Hare, D. R. and Reid, B. R.,** Three-dimensional structure of a DNA hairpin in solution: two-dimensional NMR and distance geometry calculations on d(CGCGTTTTCGCG), *Biochemistry*, 25, 5341, 1986.
126. **Hare, D., Shapiro, L., and Patel, D. J.,** Wobble dG.dT pairing in right-handed DNA: solution conformation of the d(CGTGAATTCGCG) duplex deduced from distance geometry analysis of nuclear Overhauser effect spectra, *Biochemistry*, 25, 7445, 1986.
127. **Hare, D., Shapiro, L., and Patel, D. J.,** Extrahelical adenosine stacks into the d(CGCAGAGCTCGCG) duplex deduced from distance geometry analysis of nuclear Overhauser effect spectra, *Biochemistry*, 25, 7456, 1986.
128. **Pardi, A., Hare, D. R., and Wang, C.,** Deermination of DNA structures by NMR and distance geometry calculations: a computer simulation, *Proc. Natl. Acad. Sci. U.S.A.*, 85, 8785, 1988.
129. **Patel, D. J., Shapiro, L., and Hare, D.,** Nuclear magnetic resonance and distance geometry studies of DNA structures in solution, *Annu. Rev. Biophys. Biophys. Chem.*, 16, 423, 1987.
130. **Reid, B. R.,** Nucleic acid structural studies, in 13th Int. Conf. Magnetic Resonance in Biological Systems, Abstr. L13, Madison, WI, 1988.
131. **Nilges, M., Clore, G. M., Gronenborn, A. M., Brünger, A. T., Karplus, M., and Nilsson, L.,** Refinement of the solution structure of the DNA hexamer 5′d(GCATGC)$_2$: combined use of nuclear magnetic resonance and restrained molecular dynamics, *Biochemistry*, 26, 3718, 1987.
132. **Nilges, M., Clore, G. M., Gronenborn, A. M., Piel, N., and McLaughlin, L. W.,** Refinement of the solution structure of the DNA decamer 5′d(CTGGATCCAG)$_2$: combined use of nuclear magnetic resonance and restrained molecular dynamics, *Biochemistry*, 26, 3734, 1987.
133. **Scalfi Happ, C., Happ, E., Nilges, M., Gronenborn, A. M., and Clore, G. M.,** Refinement of the solution structure of the ribonucleotide 5′r(GCAUGC): combined use of nuclear magnetic resonance and restrained molecular dynamics, *Biochemistry*, 27, 1735, 1988.
134. **Scalfi Happ, C., Happ, E., Clore, G. M., and Gronenborn, A. M.,** Refinement of the solution structure of the RNA-DNA hybrid 5′[r(GCA)d(TGC)]$_2$: combined use of nuclear magnetic resonance and restrained molecular dynamics, *FEBS Lett.*, 236, 62, 1988.
135. **Clore, G. M., Oschkinat, H., McLaughlin, L. W., Benseler, F., Scalfi Happ, C., Happ, E., and Gronenborn, A. M.,** Refinement of the solution structure of the DNA dodecamer 5′d(CGCGPATTCGCG)$_2$ containing a stable purine-thymine base pair: combined use of nuclear magnetic resonance and restrained molecular dynamics, *Biochemistry*, 27, 4185, 1988.
136. **Reid, B. R.,** Sequence-specific assignments and their use in NMR studies of DNA structure, *Q. Rev. Biophys.*, 19, 1, 1987.
137. **Nerdal, W., Hare, D. R., and Reid, B. R.,** Three-dimensional structure of the wild-type lac Pribnow promoter DNA in solution: two-dimensional nuclear magnetic resonance studies and distance geometry calculations, *J. Mol. Biol.*, 201, 717, 1988.
138. **Olejniczak, E. T., Gampe, R. T., and Fesik, S. W.,** Accounting for spin-diffusion in the analysis of 2D-NOE data, *J. Magn. Reson.*, 67, 28, 1986.
139. **Bothner-By, A. A., Stephens, R. L., Lee, J. M., Warren, C. D., and Jeanloz, R. W.,** Structure determination of a tetrasaccharide: transient nuclear Overhauser effect in the rotating frame, *J. Am. Chem. Soc.*, 106, 811, 1984.
140. **Bax, A. and Davis, D. G.,** Practical aspects of two-dimensional transverse NOE spectroscopy, *J. Magn. Reson.*, 63, 207, 1985.
141. **Bax, A., Sklenar, V., and Summers, M. F.,** Direct identification of relayed nuclear Overhauser effects, *J. Magn. Reson.*, 70, 327, 1986.
142. **Massefski, W. and Redfield, A. G.,** Elimination of multiple step spin diffusion effects in two-dimensional NOE spectroscopy of nucleic acids, *J. Magn. Reson.*, 78, 150, 1988.
143. **Broido, M. S., James, T. L., Zon, G., and Keepers, J. W.,** Investigation of the solution structure of a DNA octamer [d(GGAATTCC)]$_2$ using two-dimensional nuclear Overhauser enhancement spectroscopy, *Eur. J. Biochem.*, 150, 117, 1985.
144. **Suzuki, E., Pattabiraman, N., Zon, G., and James, T. L.,** Solution structure of [d(AT)$_5$]$_2$ via complete relaxation matrix analysis of two-dimensional nuclear Overhauser effect spectra and molecular mechanics calculations: evidence for a hydration tunnel, *Biochemistry*, 22, 6854, 1986.
145. **Zhou, N., Bianucci, A. M., Pattabiraman, N., and James, T. L.,** Solution structure of [d(GGTATACC]$_2$: wrinkled D structure of the TATA moiety, *Biochemistry*, 26, 7905, 1987.

146. **Marion, D., Genest, M., and Ptak, M.,** Reconstruction of NOESY maps: a requirement for a reliable conformational analysis of biomolecules using 2D NMR, *Biophys. Chem.*, 28, 235, 1987;**Borgias, B. A. and James, T. L.,** COMATOSE, a method for constrained refinement of macromolecular structure based on two-dimensional nuclear Overhauser effect spectra, *J. Magn. Reson.*, 79, 493, 1988.
147. **Jardetzky, O. and Roberts, G. C. K.,** *NMR in Molecular Biology*, Academic Press, New York, 1981.
148. **Karplus, M.,** Vicinal proton coupling in nuclear magnetic resonance, *J. Am. Chem. Soc.*, 85, 2870, 1963.
149. **Pardi, A., Billeter, M., and Wüthrich, K.,** Calibration of the angular dependence of the amide proton-C^{α} proton coupling constants, $^{3}J_{HN\alpha}$, in a globular protein, *J. Mol. Biol.*, 180, 741, 1984.
150. **Neuhaus, D., Wagner, G., Vasák, M., Kägi, J. H. R., and Wüthrich, K.,** Systematic application of high-resolution, phase-sensitive two-dimensional ^{1}H-NMR techniques for the identification of the amino-acid proton spin systems in proteins, *Eur. J. Biochem.*, 151, 257, 1985.
151. **Hyberts, S. G., Märki, W., and Wagner, G.,** Stereospecific assignment of side chain protons and characterization of torsion angles in Eglin c, *Eur. J. Biochem.*, 164, 625, 1987.
152. **Zuiderweg, E. R. P., Boelens, R., and Kaptein, R.,** Stereospecific assignments of ^{1}H-NMR methyl lines and conformation of valyl residues in the lac repressor headpiece, *Biopolymers*, 24, 601, 1985.
153. **DeMarco, A., Llinas, M., and Wüthrich, K.,** Analysis of the ^{1}H-NMR spectra of ferrichrome peptides I: the non-amide protons, *Biopolymers*, 17, 617, 1978.
154. **Bax, A. and Freeman, R.,** Investigation of complex networks of spin-spin coupling by two-dimensional NMR, *J. Magn. Reson.*, 44, 524, 1981.
155. **Griesinger, C., Sørensen, O. W., and Ernst, R. R.,** Two-dimensional correlation of connected transitions, *J. Am. Chem. Soc.*, 107, 6394, 1985.
156. **Griesinger, C., Sørensen, O. W., and Ernst, R. R.,** Practical aspects of the E.COSY technique: measurement of scalar spin-spin coupling constants in peptides, *J. Magn. Reson.*, 75, 474, 1987.
157. **Mueller, L.,** PE.COSY: a simple alternative to E.COSY, *J. Magn. Reson.*, 203, 251, 1987.
158. **Oschkinat, H., Pastore, A., Pfändler, P., and Bodenhausen G.,** Two-dimensional correlation of directly and remotely connected transitions by z-filtered COSY, *J. Magn. Reson.*, 69, 559, 1986.
159. **Oschkinat, H., Clore, G. M., Nilges, M., and Gronenborn, A. M.,** Application of the z-COSY technique to macromolecules: measuring coupling constants with a modified pulse sequence, *J. Magn. Reson.*, 75, 534, 1987.
160. **Driscoll, P. C., Clore, G. M., Beress, L., and Gronenborn, A. M.,** A proton nuclear magnetic resonance study of the anti-hypertensive and anti-viral protein BDS-I from the sea anemone *Anemonia sulcata*, *Biochemistry*, 28, 2178, 1989.
161. **Ponder, J. W. and Richards, F. M.,** Tertiary templates for proteins: use of packing criteria in the enumeration of allowed sequences for different structural classes, *J. Mol. Biol.*, 193, 775, 1987.
162. **McGregor, M. J., Islam, S. A., and Sternberg, M. J. E.,** Analysis of the relationship between side-chain conformation and secondary structure in globular proteins, *J. Mol. Biol.*, 198, 295, 1987.
163. **Nilges, M., Clore, G. M., and Gronenborn, A. M.,** ^{1}H-NMR stereospecific assignments by conformational database searches, *Biopolymers*, in press.
164. **Billeter, M., Widmer, G., Günter, P., and Wüthrich, K.,** Improved strategies for the structural interpretation of NMR data on proteins in solution: application to the toxin ATXIa from *Anemonia sulcata*, in Abstr. 13th Int. Conf. Magnetic Resonance in Biological Systems, S22-6, Madison, WI, 1988.
165. **Lankhorst, P. P., Erkelens, C., Haasnoot, C. A., and Altona, C.,** Carbon-13 NMR in conformational analysis of nucleic acid fragments. A reparamatrization of the Karplus equation for vicinal NMR coupling constants in CCOP and HCOP fragments, *J. Biomol. Struct. Dyn.*, 1, 1387, 1984.
166. **Sklenar, V. and Bax, A.,** Measurement of ^{1}H-^{31}P NMR coupling constants in double stranded DNA fragments, *J. Am. Chem. Soc.*, 109, 7525, 1987.
167. **Dickerson, R. E. and Drew, H. R.,** Structure of a B-DNA dodecamer II: influence of base sequence on helix structure, *J. Mol. Biol.*, 149, 761, 1981.
168. **Wagner, G., Neuhaus, D., Wörgötter, E., Vasák, M., Kägi, J. H. R., and Wüthrich, K.,** Nuclear magnetic resonance identification of "half-turn" and 3_{10}-helix secondary structure in rabbit liver metallothionein, *J. Mol. Biol.*, 187, 131, 1986.
169. **Sherman, S. A., Andrianov, A. M., and Akhrem, A. A.,** A method of determining protein conformation by the two-dimensional nuclear Overhauser enhancement spectroscopy data, *J. Biomol. Stereo Dyn.*, 4, 869, 1987.
170. **Zuiderweg, E. R. P., Kaptein, R., and Wüthrich, K.,** Secondary structure of the lac repressor DNA-binding domain by two-dimensional ^{1}H nuclear magnetic resonance in solution, *Proc. Natl. Acad. Sci. U.S.A.*, 80, 5837, 1983.
171. **Williamson, M. P., Marion, D., and Wüthrich, K.,** Secondary structure in the solution conformation of the proteinase inhibitor IIA from bull seminal plasma by nuclear magnetic resonance, *J. Mol. Biol.*, 173, 341, 1984.
172. **Zarbock, J., Clore, G. M., and Gronenborn, A. M.,** Nuclear magnetic resonance study of the globular domain of chicken histone H5: resonance assignment and secondary structure, *Proc. Natl. Acad. Sci. U.S.A.*, 23, 7628, 1986.

173. **Weber, P. L., Wemmer, D. E., and Reid, B. R.,** ^{1}H NMR studies of λ cro repressor: sequential resonance assignments of the ^{1}H NMR spectrum, *Biochemistry*, 24, 4553, 1985.
174. **Clore, G. M., Martin, S. R., and Gronenborn, A. M.,** The solution structure of human growth hormone releasing factor: combined use of circular dichroism and nuclear magnetic resonance spectroscopy, *J. Mol. Biol.*, 191, 553, 1986.
175. **Sukumaran, D. K., Clore, G. M., Preuβ, A., Zarbock, J., and Gronenborn, A. M.,** A proton nuclear magnetic resonance study of hirudin: resonance assignment and secondary structure, *Biochemistry*, 26, 333, 1987.
176. **Weber, P. L., Brown, S. C., and Mueller, L.,** Sequential ^{1}H NMR assignments and secondary structure identification of human ubiquitin, *Biochemistry*, 26, 7282, 1987.
177. **DiStefano, D. L. and Wand, A. J.,** Two-dimensional ^{1}H NMR study of human ubiquitin: a main chain directed assignment and structure analysis, *Biochemistry*, 26, 7272, 1987.
178. **Driscoll, P. C., Hill, H. A. O., and Redfield, C.,** ^{1}H-NMR sequential assignment and cation binding studies of spinach plastocyanin, *Eur. J. Biochem.*, 170, 279, 1987.
179. **Chazin, W. J. and Wright, P. E.,** Complete assignment of the ^{1}H nuclear magnetic resonance spectrum of french bean plastocyanin: sequential resonance assignments, secondary structure and global fold, *J. Mol. Biol.*, 202, 623,
180. **Clore, G. M., Sukumaran, D. K., Gronenborn, A. M., Teeter, M. M., Whitlow, M., and Jones, B. L.,** A nuclear magnetic resonance study of the solution structure of α1-purothionin: sequential resonance assignment, secondary structure and low resolution tertiary structure, *J. Mol. Biol.*, 193, 571, 1987.
181. **Redfield, C. and Dobson, C. M.,** Sequential ^{1}H NMR assignments and secondary structure of hen egg white lysozyme in solution, *Biochemistry*, 27, 122, 1988.
182. **Robertson, A. D., Westler, W. M., and Markley, J. L.,** Two-dimensional NMR study of Kazal Proteinase inhibitors I: sequence specific assignments and secondary structure of turkey ovomucoid third domain, *Biochemistry*, 27, 2519, 1988.
183. **Gronenborn, A. M., Bovermann, G., and Clore, G. M.,** A ^{1}H-NMR study of the solution conformation of secretin: resonance assignment and secondary structure, *FEBS Lett.*, 215, 88, 1987.
184. **Gronenborn, A. M., Wingfield, P. T., and Clore, G. M.,** Determination of the secondary structure of the DNA binding protein Ner from phage Mu using ^{1}H homonuclear and ^{15}N-^{1}H hetereonuclear NMR spectroscopy, *Biochemistry*, in press.
185. **Holak, T. A. and Prestegard, J. H.,** Secondary structure of acyl acrrier protein as derived from two-dimensional ^{1}H NMR spectroscopy, *Biochemistry*, 25, 5766, 1986.
186. **Nettesheim, D. G., Edalji, R. P., Mollison, K. W., Greer, J., and Zuiderweg, E. R. P.,** Secondary structure of complement component C3a anaphylatoxin in solution as determined by NMR spectroscopy: differences between crystal and solution conformations, *Proc. Natl. Acad. Sci. U.S.A.*, 85, 5036, 1988.
187. **Kraulis, P. J. and Jones, T. A.,** Determination of three-dimensional protein structures from nuclear magnetic resonance data using fragments of known structures, *Proteins Struct. Function Genet.*, 2, 188, 1987.
188. **Litharge, O., Cornelius, C. W., Buchanan, B. G., and Jardetzky, O.,** Validation of the first step of the heuristic refinement method for derivation of solution structures of proteins from NMR data, *Proteins Struct. Function Genet.*, 2, 340, 1987.
189. **Billeter, M., Havel, T. F., and Wüthrich, K.,** The ellipsoid algorithm as a method for the determination of polypeptide conformations from experimental distance constraints and energy minimization, *J. Comput. Chem.*, 8, 132, 1987.
190. **Nilsson, L., Clore, G. M., Gronenborn, A. M., Brünger, A. T., and Karplus, M.,** Structure refinement of oligonucleotides by molecular dynamics with nuclear Overhauser effect interproton distance restraints: application to 5′ d(CGTACG)$_2$, *J. Mol. Biol.*, 188, 455, 1986; **Gronenborn, A. M. and Clore, G. M.,** Analysis of the relative contributions of the nuclear Overhauser interproton distance restraints and the emperical energy function in the calculation of oligonucleotide structures using restrained molecular dynamics, *Biochemistry*, in press.
191. **Crippen, G. M.,** A novel approach to calculation of conformation: distance geometry, *J. Comput. Phys.*, 24, 96, 1977.
192. **Crippen, G. M. and Havel, T. F.,** Stable calculation of coordinates from distance information, *Acta Crystallogr.*, A34, 282, 1978.
193. **Kuntz, I. D., Crippen, G. M., Kollman, P. A., and Kimmelman, D.,** Calculation of protein tertiary structure, *J. Mol. Biol.*, 106, 983, 1976.
194. **Kuntz, I. D., Crippen, G. M., and Kollman, P. A.,** Application of distance geometry to protein tertiary structure calculations, *Biopolymers*, 18, 939, 1979.
195. **Havel, T. F., Kuntz, I. D., and Crippen, G. M.,** Theory and practice of distance geometry, *Bull. Math. Biol.*, 45, 665, 1983.
196. **Havel, T. F. and Wüthrich, K.,** A distance geometry program for determining the structures of small proteins and other macromolecules from nuclear magnetic resonance measurements of intramolecular ^{1}H-^{1}H proximities in solution, *Bull. Math. Biol.*, 46, 673, 1984.

197. **Wako, H. and Scheraga, H. A.,** Distance-constraint approach to protein folding. II. Prediction of three-dimensional structure of Bovine pancreatic trypsin inhibitor, *J. Prot. Chem.*, 1, 85, 1982.
198. **Sippl, M. J. and Scheraga, H. A.,** Cayley-Menger coordinates, *Proc. Natl. Acad. Sci. U.S.A.*, 83, 2283, 1986.
199. **Havel, T. F.,** DISGEO, Quantum Chemistry Program Exchange Program No. 507, Indiana University, 1986.
200. **Braun, W.,** Distance geometry and related methods for protein structure determination from NMR data, *Q. Rev. Biophys.*, 19 115, 1987.
201. **McCammon, J. A., Gelin, B. R., and Karplus, M.,** Dynamics of folded proteins, *Nature (London)*, 267, 585, 1977.
202. **McCammon, J. A., Wolynes, P. G., and Karplus, M.,** Picosecond dynamics of tyrosine side chains in proteins, *Biochemistry*, 18, 927, 1979.
203. **Karplus, M. and McCammon, J. A.,** Dynamics of proteins: elements and function, *Annu. Rev. Biochem.*, 52, 263, 1983.
204. **Levitt, M.,** Protein folding by restrained energy minimization and molecular dynamics, *J. Mol. Biol.*, 170, 723, 1983.
205. **Weiner, P. K. and Kollman, P. A.,** Amber: assisted model building with energy refinement. A general program for modeling molecules and their interactions, *J. Comput. Chem.*, 2, 287, 1981.
206. **van Gunsteren, W. F. and Berendsen, H. J. C.,** Molecular dynamics: perspective for complex systems, *Biochem. Soc. Trans.*, 10, 301, 1982.
207. **Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P.,** Optimization by simulated annealing, *Science*, 220, 671 1983.
208. **Bounds, D. G.,** New optimization methods from physics and biology, *Nature (London)*, 329, 215, 1987.
209. **Metropolis, N., Rosenbluth, M., Rosenbluth, A., Teller, A., and Teller, E.,** Equations of state calculations by fast computing machines, *J. Chem. Phys.*, 21, 1087, 1953.
210. **Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M.,** CHARMM: a program for macromolecular energy minimization and dynamics calculations, *J. Comput. Chem.*, 4, 187, 1983.
211. **Nilges, M., Clore, G. M., and Gronenborn, A. M.,** A simple method for delineating well-defined and variable regions in protein structures determined from interproton distance data, *FEBS Lett.*, 219, 11, 1987.
212. **Brünger, A. T., Campbell, R. L., Clore, G. M., Gronenborn, A. M., Karplus, M., Petsko, G. A., and Teeter, M. M.,** Solution of a protein crystal structure with a model obtained from NMR interproton distance restraints, *Science*, 235, 1049, 1987.
213. **Hendrickson, W. A. and Teeter, M. M.,** Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur, *Nature (London)*, 290, 107, 1981.
214. **Pflugrath, J., Wiegand, G., Huber, R., and Vértesy, L.,** Crystal structure determination refinement and the molecular model of the α-amylase inhibitor HOE-467A, *J. Mol. Biol.*, 189, 383, 1986.
215. **Arnott, S. and Hukins, D. W.,** Optimized parameters for A-DNA and B-DNA, *Biochem. Biophys. Res. Commun.*, 47, 1504, 1972.
216. **Walter, J. and Huber, R.,** Pancreatic trypsin inhibitor: a new crystal form and its analysis, *J. Mol. Biol.*, 167, 911, 1983.
217. **Rees, D. C. and Lipscomb, W. N.,** Refined crystal structure of the potato inhibitor complex of carboxypeptidase A at 2.5 Å resolution, *J. Mol. Biol.*, 160, 475, 1982.
218. **Clore, G. M., Gronenborn, A. M., James, M. N. G., Kjær, M., McPhalen, C. A., and Poulsen, F.,** Comparison of the solution and X-ray structures of barley serine proteinase inhibitor 2, *Protein Eng.*, 1, 313, 1987.
218a. **Braun, W.,** personal communication.
219. **McPhalen, C. A. and James, M. N. G.,** Crystal and molecular structure of serine proteinase inhibitor CI-2 from barley seeds, *Biochemistry*, 26, 261, 1987.
220. **Papamakos, E., Weber, E., Bode, W., Huber, R., Empie, M. W., Kato, I., and Laskowski, M.,** Crystallographic refinement of Japanese Quail ovomucoid, a Kazal-type inhibitor, and model building studies of complexes with serine proteases, *J. Mol. Biol.*, 158, 515, 1982.
221. **Bolognesi, M., Gatti, G., Menegatti, E., Guarneri, M., Marquart, M., Papamakos, E., and Huber, R.,** Three-dimensional structure of the complex between pancreatic secretory trypsin inhibitor (Kazal type) and trypsinogen at 1.8 Å resolution structure solution, crystallographic refinement and preliminary structural interpretation, *J. Mol. Biol.*, 162, 839, 1982.
222. **Huber, R., Scholze, H., Paques, E. P., and Deisenhofer, J.,** Crystal structure analysis and molecular model of human C3a anaphylatoxin, *Hoppe Seylers Z. Physiol. Chem.*, 361, 1389, 1980.
223. **Fontecilla-Camps, J. C., Almassy, R. J., Suddath, F. L., and Bugg, C. E.,** The three-dimensional structure of scorpion neurotoxins, *Toxicon*, 20, 1, 1982.
224. **Pabo, C. O. and Lewis, M.,** The operator-binding domain of λ repressor: structure and DNA recognition, *Nature (London)*, 298, 443, 1982.
225. **Sasaki, K., Dockerill, S., Adamiak, D. A., Tickle, I. J., and Blundell, T.,** X-ray analysis of glucagon and its relationship to receptor binding, *Nature (London)*, 257, 751, 1975.

226. **Furey, W. F., Robbins, A. H., Clancy, L. L., Winge, D. R., Wang, B. C., and Stout, C. D.,** Crystal structure of Cd, Zn metallothionein, *Science*, 231, 704, 1986.
227. **Frey, M. H., Wagner, G, Vasák, M., Sørensen, O. W., Neuhaus, D., Wörgötter, E., Kägi, J. H. R., Ernst, R. R., and Wüthrich, K.,** Polypeptide-metal cluster connectivities in metallothionein 2 by novel ^{1}H-^{113}Cd heteronuclear two-dimensional NMR experiments, *J. Am. Chem. Soc.*, 107, 6847, 1985.
228. **Beress, L., Doppelfeld, I. S., Etschenberg, E., Graf, E., Henschen-Edman, A., and Zwick, J.,** Polypeptides, Methods of Production and Their Uses as Antihypertensives, Patent No. DE 3324689 A1 of the Federal Republic of Germany, 1985.
229. **Driscoll, P. C., Clore, G. M., and Gronenborn, A. M.,** Influence of stereospecific assignments on the NOE-based structure determination of small proteins: application to the anti-viral and anti-hypertensive protein BDS-I from the sea anemone *Anemonia sulcata*, *FEBS Lett.*, 243, 223, 1989.
230. **Markwardt, F.,** Pharmacology of hirudin: one hundred years after the first report of the anticoagulant agent in medicinal leeches, *Biomed. Biochim. Acta*, 44, 1007, 1985.
231. **Calladine, C. R.,** Mechanics of sequence dependent stacking, *J. Mol. Biol.*, 161, 343, 1982.
232. **Dickerson, R. E.,** Base sequence and helix structure variations in B and A DNA, *J. Mol. Biol.*, 149, 761, 1983.
233. **Dieckmann, S. and McLaughlin, L. W.,** DNA curvature in native and modifed EcoRI recognition sites and possible influence upon the endonuclease cleavage reaction, *J. Mol. Biol.*, 202, 823, 1988.
234. **Boelens, R., Scheek, R. M., Lamerichs, R. M. J. N., de Vlieg, J., van Boom, J. H., and Kaptein, R.,** The complex of the lac repressor headpiece with a 14 base pair lac operator fragment studied by two-dimensional NMR, *J. Mol. Biol.*, 193, 213, 1987.
235. **Plateau, P. and Gueron, M.,** Exchangeable proton NMR without baseline distortion using new strong pulse sequences, *J. Am. Chem. Soc.*, 104, 7310, 1982.