Pergamon

Prog. Biophys. molec. Biol., Vol. 62, pp. 153–184, 1994
1994 Elsevier Science Ltd
Printed in Great Britain
0079-6107/94 $26.00

0079-6107(94)E0003-Q

STRUCTURES OF LARGER PROTEINS, PROTEIN–LIGAND AND PROTEIN–DNA COMPLEXES BY MULTI-DIMENSIONAL HETERONUCLEAR NMR*

G. MARIUS CLORE and ANGELA M. GRONENBORN

Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, U.S.A.

CONTENTS

I. INTRODUCTION

The last few years have seen a quantum jump both in the size and accuracy of protein structures that can be determined by NMR (Clore and Gronenborn, 1991a). Thus it is now possible to determine the structures of proteins in the 15–25 kDa range at a resolution comparable to 2 Å resolution crystal structures (Clore and Gronenborn, 1991b). This is attributable to the development of three- and four-dimensional heteronuclear NMR techniques to circumvent problems associated with chemical shift overlap and degeneracy on the one hand and large linewidths on the other (see Clore and Gronenborn, 1991a,c,d, and Bax and Grzesiek, 1993 for reviews). In this short review, we summarize some of these developments and illustrate their application to the structure determination of interleukin-1β (Clore *et al.*, 1991a), a complex of calmodulin with a target peptide (Ikura *et al.*, 1992) and a complex of the DNA binding domain of the transcription factor GATA-1 with its cognate DNA target site (Omichinski *et al.*, 1993).

II. GENERAL STRATEGY FOR THE DETERMINATION OF THREE-DIMENSIONAL STRUCTURES OF LARGER PROTEINS AND PROTEIN COMPLEXES BY NMR

The main source of geometric information used in protein structure determination lies in the nuclear Overhauser effect (NOE) which can be used to identify protons separated by less

* Adapted from the Young Investigator Award Lecture presented by G. M. Clore at the Seventh Annual Symposium of the Protein Society (San Diego, July, 1993). A version of this lecture is also being published in Protein Science (vol. 3, pp. 372–390, 1994).

than 5 Å (Ernst *et al.*, 1987). This distance limit arises from the fact that the NOE is proportional to the inverse sixth power of the distance between the protons. Hence the NOE intensity falls off very rapidly with increasing distance between proton pairs. Despite the short range nature of the observed interactions, the short approximate interproton distance restraints derived from the NOE measurements can be highly conformationally restrictive, particularly when they involve residues that are far apart in the sequence but close together in space.

The power of NMR over other spectroscopic techniques results from the fact that every proton gives rise to an individual resonance in the spectrum which can be resolved by higher dimensional (i.e. 2D, 3D and 4D) techniques. Bearing this in mind, the principles of structure determination by NMR can be summarized very simply by the scheme depicted in Fig. 1.

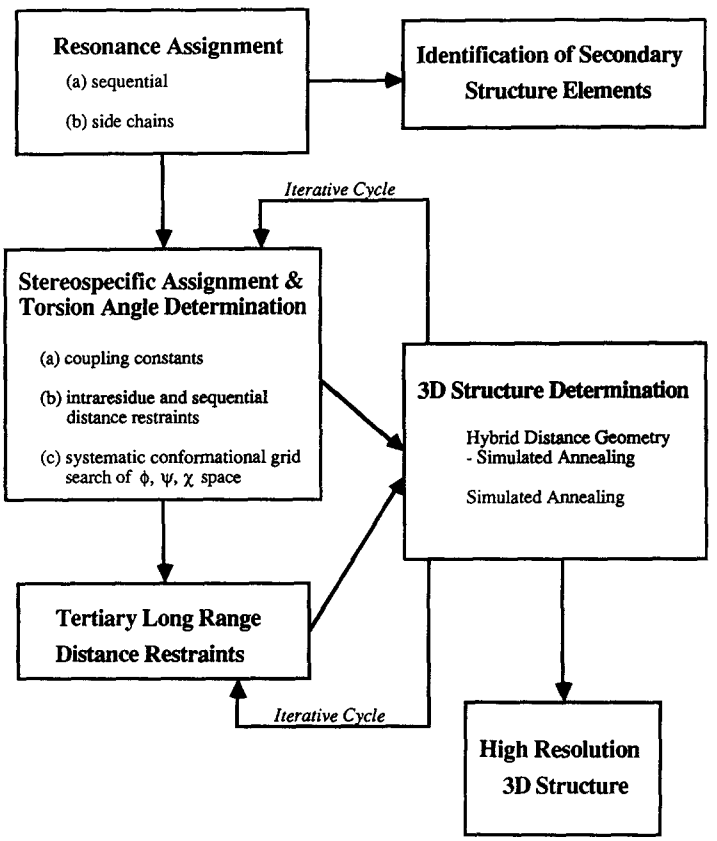


FIG. 1. Summary of general strategy employed to solve three-dimensional structures of macromolecules by NMR.

The first step is to obtain sequential resonance assignments using a combination of through-bond and through-space correlations; the second step is to obtain stereospecific assignments at chiral centers and torsion angle restraints using three-bond scalar couplings combined with intraresidue and sequential interresidue NOE data; the third step is to identify through-space connectivities between protons separated by less than 5 Å; and finally the fourth step involves calculating three-dimensional structures on the basis of the amassed interproton distance and torsion angle restraints using one or more of a number of algorithms (Havel *et al.*, 1983; Braun, 1987; Clore and Gronenborn, 1989) such as distance geometry and/or simulated annealing. It is not essential to assign all the NOEs initially. Indeed, many may be ambiguous and several possibilities may exist for their assignments. Once a low resolution structure, however, has been calculated from a subset of the NOE data

which can be interpreted unambiguously, it is then possible to employ iterative methods to resolve the vast majority of ambiguities. Consider for example an NOE cross peak which could be attributable to a through-space interaction between either protons A and B or between protons A and C. Once a low resolution structure is available it is usually possible to descriminate between these two possibilities. Thus, if protons A and C are significantly greater than 5 Å apart while protons A and B are less than 5 Å apart, it is clear that the cross peak must arise from an NOE between protons A and B.

The quality of an NMR protein structure determination increases as the number of restraints increase (Havel and Wüthrich, 1985; Clore and Gronenborn, 1991a,d; Havel, 1991; Clore *et al.*, 1993). This progression in coordinate precision is illustrated in Fig. 2 which shows four generations of structures ranging from the first generation which simply provides a picture of the polypeptide fold with little detail to the fourth generation which is broadly equivalent to a 2 Å resolution X-ray structure.

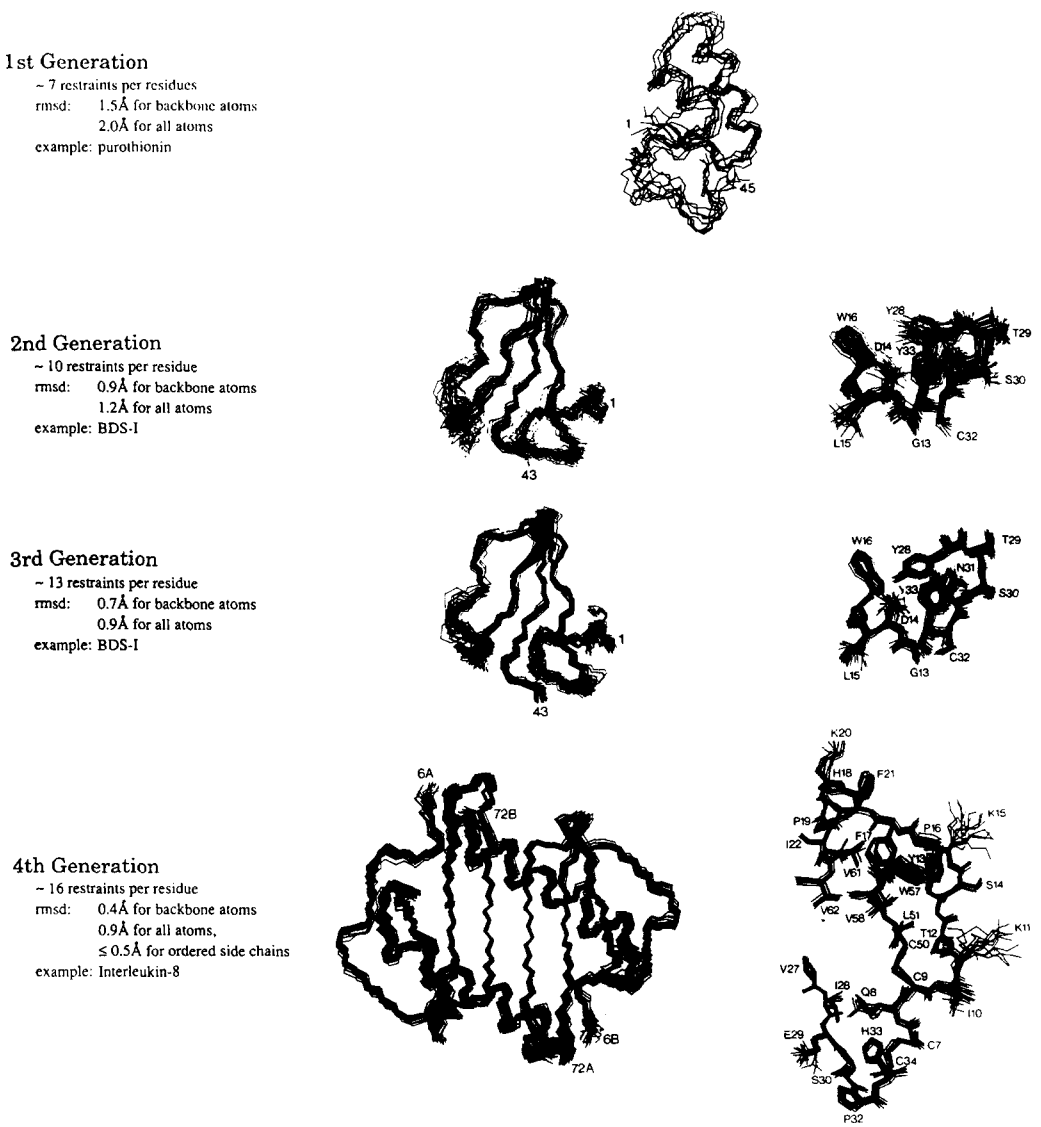


FIG. 2. Illustration of the progressive improvement in the precision and accuracy of NMR structure determinations with increasing number of experimental restraints. All the structures have been calculated using the hybrid distance geometry-simulated annealing method, and in each case the NOE-derived interproton distance restraints have been grouped into three broad ranges. 1.8–2.7 Å, 1.8–3.3 Å and 1.8–5.0 Å, corresponding to strong, medium and weak NOEs, respectively.

1. *Sequential Resonance Assignment*

Conventional sequential resonance assignment relies on 2D homonuclear ^1H–^1H correlation experiments to identify amino acid spin systems coupled with 2D ^1H–^1H NOE experiments to identify sequential connectivities along the backbone of the type $C^\alpha H(i)$–$NH(i+1,2,3,4)$, $NH(i)$–$NH(i\pm2)$ and $C^\alpha H(i)$–$C^\beta H(i+3)$ (Wüthrich, 1986; Clore and Gronenborn, 1987). This methodology has been successfully applied to proteins of less than 100 residues. For larger proteins, the spectral complexity is such that 2D experiments no longer suffice, and it is essential to increase the spectral resolution by increasing the dimensionality of the spectra (Oschkinat *et al.*, 1988). In some cases it is still possible to apply the same strategy by making use of 3D heteronuclear (^{15}N or ^{13}C) edited experiments to increase the spectral resolution, as illustrated in Fig. 3 (Marion *et al.*, 1989; Fesik and

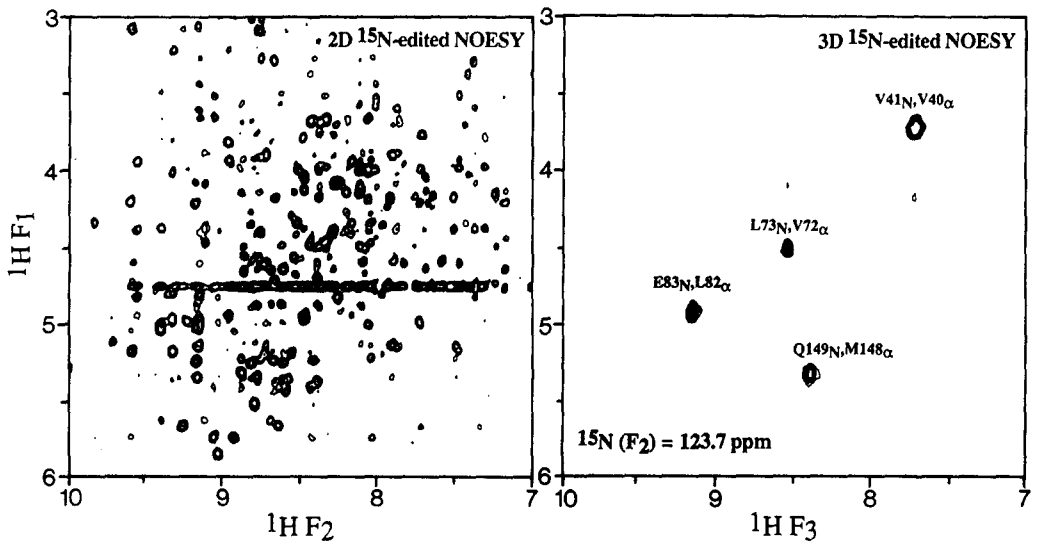


FIG. 3. Comparison of the NH–$C^\alpha H/C^\beta H$ region of a 2D ^{15}N-edited NOESY spectrum with that of a single plane taken from the 3D ^{15}N-edited NOESY spectrum, illustrating the increase in spectral resolution afforded by increasing the dimensionality from two to three.

Zuiderweg, 1988, 1990; Driscoll *et al.*, 1990a). In many cases, however, numerous ambiguities still remain and it is advisable to adopt a sequential assignment strategy based solely on well defined heteronuclear scalar couplings (Montelione and Wagner, 1989, 1990; Ikura *et al.*, 1990; Clore and Gronenborn, 1991c; Bax and Grzesiek, 1993). The double and triple resonance experiments that we currently use together with the correlations that they demonstrate, are summarized in Table 1. With the advent of pulsed field gradients to eliminate undesired coherence transfer pathways (Bax and Pochapsky, 1992), as opposed to selecting desired coherence transfer pathways (Hurd and John, 1991; Vuister *et al.*, 1991), it is now possible to employ only two step phase-cycles without any loss in sensitivity (other than that due to the reduction in measurement time) such that each 3D experiment can be recorded in as little as 7 hr. In most cases, however, signal-to-noise requirements necessitate 1–3 days measuring time depending on the experiment.

2. *Stereospecific Assignments and Torsion Angle Restraints*

It is often possible to obtain stereospecific assignments of β-methylene protons on the basis of a qualitative interpretation of the homonuclear $^3J_{\alpha\beta}$ coupling constants and the intraresidue NOE data involving the NH, $C^\alpha H$ and $C^\beta H$ protons (Wagner *et al.*, 1987; Hyberts *et al.*, 1987). A more rigorous approach, which also permits one to obtain ϕ, ψ, and χ_1 restraints as well involves the application of a conformational grid search of ϕ,ψ,χ_1 space on the basis of the homonuclear $^3J_{HN_\alpha}$ and $^3J_{\alpha\beta}$ coupling constants (which are related to ϕ and

TABLE 1. SUMMARY OF CORRELATIONS OBSERVED IN THE 3D DOUBLE AND TRIPLE RESONANCE EXPERIMENTS USED FOR SEQUENTIAL AND SIDE CHAIN ASSIGNMENTS IN OUR LABORATORY

Experiment	Correlation	J Coupling*
^{15}N-edited HOHAHA	$C^{\alpha}H(i)$–$^{15}N(i)$–$NH(i)$	$^{3}J_{HN\alpha}$
	$C^{\beta}H(i)$–$^{15}N(i)$–$NH(i)$	$^{3}J_{HN\alpha}$ and $^{3}J_{\alpha\beta}$
HNHA	$C^{\alpha}H(i)$–$^{15}N(i)$–$NH(i)$	$^{3}J_{HN\alpha}$
H(CA)NH	$C^{\alpha}H(i)$–$^{15}N(i)$–$NH(i)$	$^{1}J_{NC\alpha}$
	$C^{\alpha}H(i-1)$–$^{15}N(i)$–$NH(i)$	$^{2}J_{NC\alpha}$
HNCA	$^{13}C^{\alpha}(i)$–$^{15}N(i)$–$NH(i)$	$^{1}J_{NC\alpha}$
	$^{13}C^{\alpha}(i-1)$–$^{15}N(i)$–$NH(i)$	$^{2}J_{NC\alpha}$
HN(CO)CA	$^{13}C^{\alpha}(i-1)$–$^{15}N(i)$–$NH(i)$	$^{1}J_{NCO}$ and $^{1}J_{C\alpha CO}$
HNCO	$^{13}CO(i-1)$–$^{15}N(i)$–$NH(i)$	$^{1}J_{NCO}$
HCACO	$C^{\alpha}H(i)$–$^{13}C^{\alpha}(i)$–$^{13}CO(i)$	$^{1}J_{C\alpha CO}$
HCA(CO)N	$C^{\alpha}H(i)$–$^{13}C^{\alpha}(i)$–$^{15}N(i+1)$	$^{1}J_{C\alpha CO}$ and $^{1}J_{NCO}$
CBCA(CO)NH	$^{13}C^{\beta}(i-1)/^{13}C^{\alpha}(i-1)$–$^{15}N(i)$–$NH(i)$	$^{1}J_{C\alpha CO}$, $^{1}J_{NCO}$ and $^{1}J_{CC}$
CBCANH	$^{13}C^{\beta}(i)/^{13}C^{\alpha}(i)$–$^{15}N(i)$–$NH(i)$	$^{1}J_{NC\alpha}$ and $^{1}J_{CC}$
	$^{13}C^{\beta}(i-1)/^{13}C^{\alpha}(i-1)$–$^{15}N(i)$–$NH(i)$	$^{2}J_{NC\alpha}$ and $^{1}J_{CC}$
HBHA(CO)NH	$C^{\beta}H(i-1)/C^{\alpha}H(i-1)$–$^{15}N(i)$–$NH(i)$	$^{1}J_{C\alpha CO}$, $^{1}J_{NCO}$ and $^{1}J_{CC}$
HBHANH	$C^{\beta}H(i)/C^{\alpha}H(i)$–$^{15}N(i)$–$NH(i)$	$^{1}J_{NC\alpha}$ and $^{1}J_{CC}$
	$C^{\beta}H(i-1)/C^{\alpha}H(i-1)$–$^{15}N(i)$–$NH(i)$	$^{2}J_{NC\alpha}$ and $^{1}J_{CC}$
C(CO)NH	$^{13}C^{j}(i-1)$–$^{15}N(i)$–$NH(i)$	$^{1}J_{C\alpha CO}$, $^{1}J_{NCO}$ and $^{1}J_{CC}$
H(CCO)NH	$H^{j}(i-1)$–$^{15}N(i)$–$NH(i)$	$^{1}J_{C\alpha CO}$, $^{1}J_{NCO}$ and $^{1}J_{CC}$
HCCH-COSY	H^{j}–$^{13}C^{j}$–$^{13}C^{j\pm1}$–$H^{j\pm1}$	$^{1}J_{CC}$
HCCH-TOCSY	H^{j}–$^{13}C^{j}$–…–$^{13}C^{j\pm1}$–$H^{j\pm n}$	$^{1}J_{CC}$

*In addition to the couplings indicated, all the experiments make use of the $^{1}J_{CH}$ ($\sim$140 Hz) and/or $^{1}J_{NH}$ ($\sim$95 Hz) couplings. The values of the couplings employed are as follows: $^{3}J_{HN\alpha}$ $\sim$3–10 Hz, $^{1}J_{CC}$ $\sim$35 Hz, $^{1}J_{C\alpha CO}$ $\sim$55 Hz, $^{1}J_{NCO}$ $\sim$ 15 Hz, $^{1}J_{NC\alpha}$ $\sim$11 Hz, $^{2}J_{NC\alpha}$ $\sim$7 Hz.

χ_1, respectively) and the intraresidue and sequential ($i\pm1$) interresidue NOEs involving the NH, $C^{\alpha}H$ and $C^{\beta}H$ protons (Güntert *et al.*, 1989; Nilges *et al.*, 1990). This information can be supplemented by the measurement of heteronuclear $^{3}J_{NH\beta}$ and $^{3}J_{COH\beta}$ couplings which are also related to χ_1 (Vuister *et al.*, 1994). Stereospecific assignment of valine methyl groups can be made on the basis of $^{3}J_{C\gamma CO}$, $^{3}J_{NC\gamma}$ couplings (Vuister *et al.*, 1994), as well as on the basis of the pattern of intraresidue NOEs involving the NH, $C^{\alpha}H$ and $C^{\gamma}H$ protons (Zuiderweg *et al.*, 1985). Finally, stereospecific assignments of leucine methyl groups can be made on the basis of heteronuclear $^{3}J_{C\delta C\alpha}$ and $^{3}J_{C\delta H\beta}$ couplings (Vuister *et al.*, 1994) in combination with the pattern of intraresidue NOEs, provided that the stereospecific assignment of the β-methylene protons and the χ_1 rotamer have been previously determined (Powers *et al.*, 1993).

3. *Assignment of Through-Space Proton–Proton Interactions Within a Protein*

While the panoply of 3D heteronuclear experiments is sufficient for the purposes of spectral assignment, yet further increases in resolution are required for the reliable identification of NOE through-space interactions. This can be achieved by extending the dimensionality still further to four dimensions (Kay *et al.*, 1990). This is illustrated in Fig. 4. Consider a simple 2D spectrum demonstrating 11 cross peaks from aliphatic resonances to a single NH resonance position. In the 2D spectrum it is impossible to ascertain whether this involves one NH proton or many. Extending the spectrum to 3D by separating the NOE interactions according to the ^{15}N chemical shift of the nitrogen attached to each amide proton reveals that there are three NH protons involved. The identity of the originating aliphatic protons, however, is only specified by their proton chemical shifts. Yet the extent of spectral overlap in the aliphatic region of the spectrum vastly exceeds that in the amide region. This can be resolved by adding a further dimension in which each plane of the 3D spectrum now constitutes a cube in the 4D spectrum edited by the ^{13}C shift of the carbon atom attached to each aliphatic proton. In this manner, each ^{1}H–^{1}H NOE interaction is specified by four chemical shift coordinates, the two protons giving rise to the NOE and the heavy atoms to which they are attached. The resolving power of 4D heteronuclear-edited NOE spectroscopy is illustrated in Fig. 5.

Because the number of NOE interactions present in each 2D plane of a 4D ^{13}C/^{15}N or ^{13}C/^{13}C-edited NOESY spectrum is so small, the inherent resolution in a 4D spectrum is

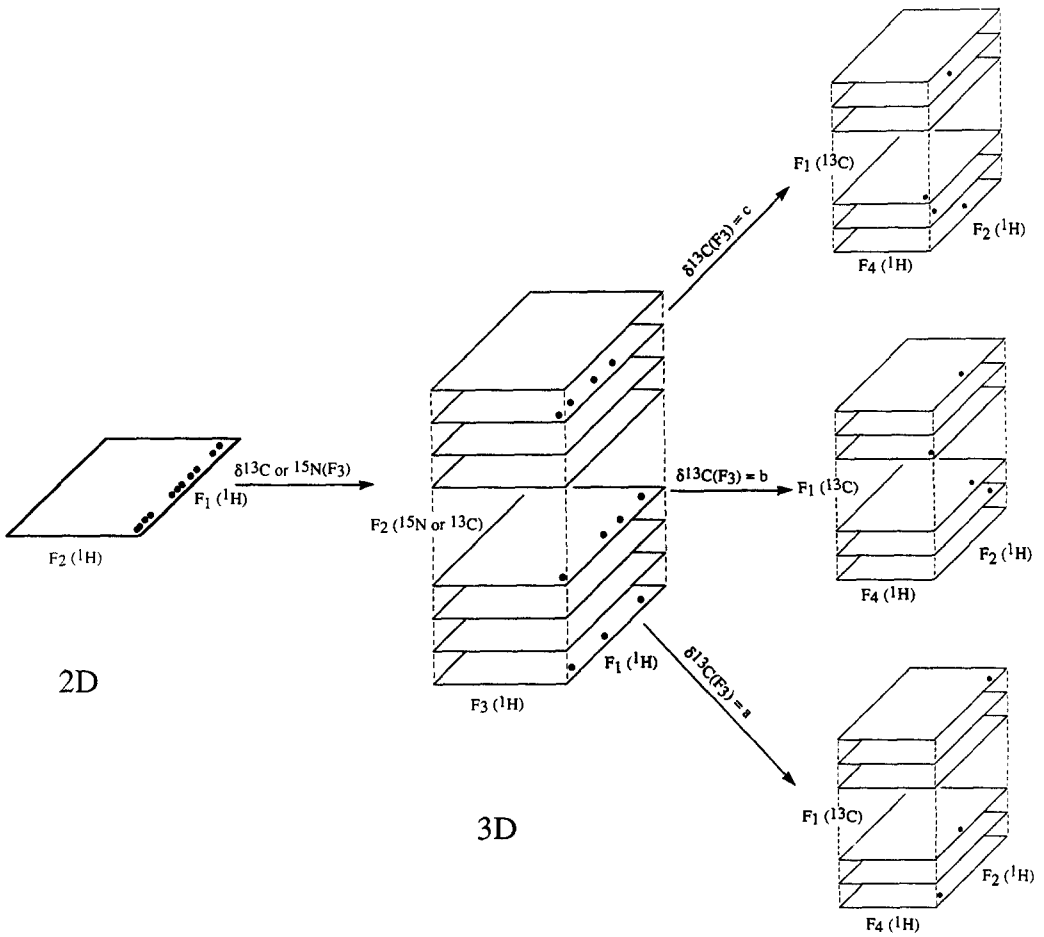


FIG. 4. Schematic illustration of the progression and relationship between 2D, 3D and 4D heteronuclear-edited NMR spectroscopy.

extremely high, despite the low level of digitization. Indeed, spectra with equivalent resolution can be recorded at magnetic field strengths considerably lower than 600 MHz, although this would obviously lead to a reduction in sensitivity. Further, it can be calculated that 4D spectra with virtual lack of resonance overlap and good sensitivity can be obtained on proteins with as many as 400 residues. Thus, once complete ^{1}H, ^{15}N and ^{13}C assignments are obtained, analysis of 4D $^{15}N/^{13}C$ (Kay *et al.*, 1990) and $^{13}C/^{13}C$ (Clore *et al.*, 1991b; Zuiderweg *et al.*, 1991; Vuister *et al.*, 1993) edited NOE spectra should permit the automated assignment of almost all NOE interactions.

III. APPLICATION OF 3D AND 4D NMR TO PROTEIN STRUCTURE DETERMINATION OF LARGER PROTEINS: THE STRUCTURE OF INTERLEUKIN-1β

While the potential of heteronuclear 3D and 4D NMR methods in resolving problems associated with both extensive resonance overlap and large linewidths is obvious, how does this new approach fare in practice? In this regard it should be borne in mind that resonance assignments are only a means to an end, and the true test of multidimensional NMR lies in examining its success in solving the problem for which it was originally designed to tackle, namely the determination of high resolution three-dimensional structures of larger proteins in solution.

The first successful demonstration of these new methods was the determination of the high resolution solution structure of interleukin-1β (IL-1β), a cytokine of 153 residues and molecular weight 17.4 kDa, which plays a key role in the immune and inflammatory

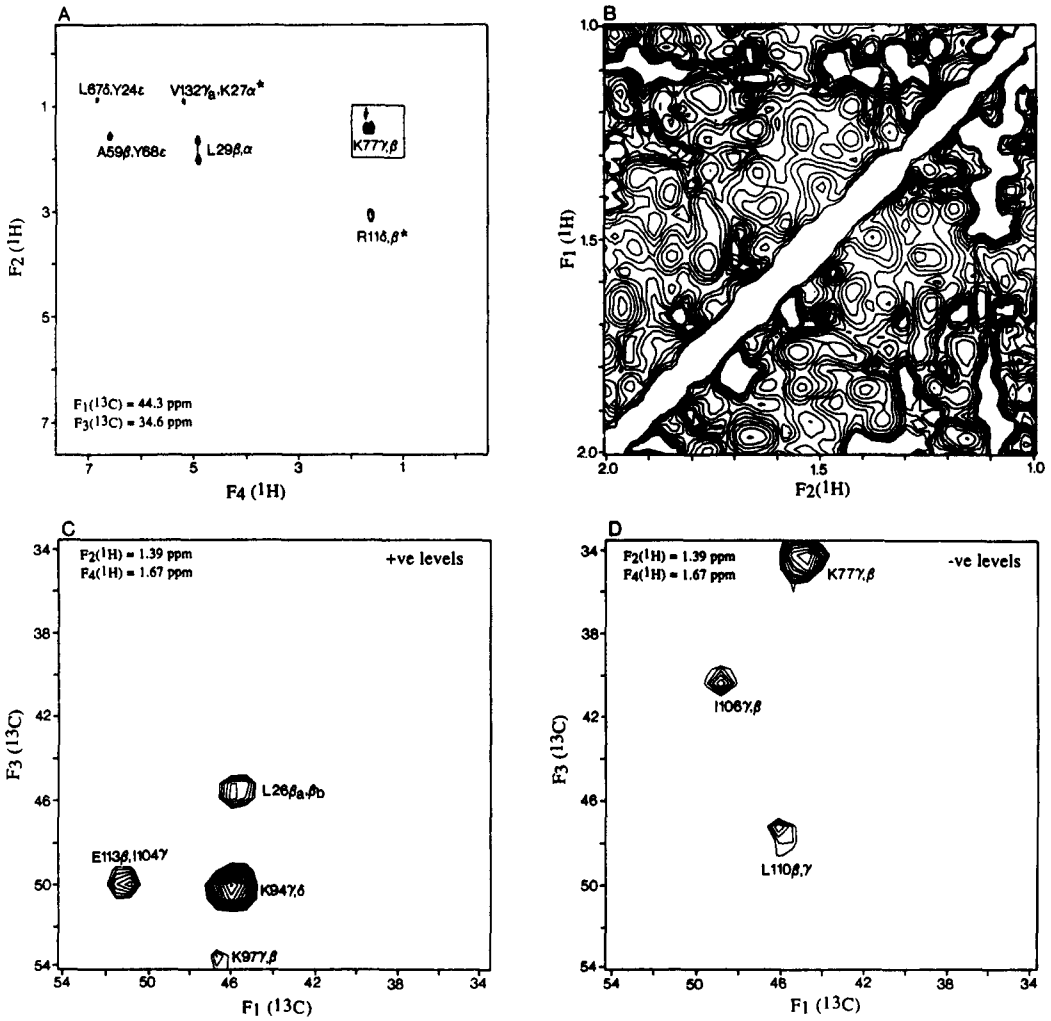


FIG. 5. Example of the increase in spectral resolution afforded by 4D $^{13}C/^{13}C$-edited NOE spectroscopy, illustrated with interleukin-1β. (A) $^{1}H(F_2)$–$^{1}H(F_4)$ plane of the 4D spectrum at $\delta^{13}C(F_1) = 44.3$ ppm and $\delta^{13}C(F_3) = 34.6$ ppm; the region between 1 and 2 ppm is boxed in and the arrow indicates the position of the Lys^{77} $C^{\gamma}H$–$C^{\beta}H$ NOE cross peak. (B) 2D ^{1}H–^{1}H NOE spectrum between 1 and 2 ppm; the X marks the chemical shift position of the Lys^{77} $C^{\gamma}H$–$C^{\beta}H$ NOE cross peak seen in (A). (C) and (D) positive and negative contours in the $^{13}C(F_2)$–$^{13}C(F_3)$ plane of the 4D spectrum at the ^{1}H chemical shift coordinates, $\delta^{1}H(F_2) = 1.39$ ppm and $\delta^{1}H(F_4) = 1.67$ ppm, corresponding to the Lys^{77} $C^{\gamma}H$–$C^{\beta}H$ NOE cross peak seen in (A) and the X mark shown in (B). Because extensive folding is employed the ^{13}C chemical shifts are given by $x \pm nSW$ where x is the ppm value listed in the figure, n an integer, and SW the spectral width (20.71 ppm). Peaks folded an even number of times are of opposite sign to those folded an odd number of times. All the peaks in (A) are positive except for the two indicated by an asterisk which are negative.

responses (Clore *et al.*, 1991a). At the time IL-1β was 50% larger, in terms of number of residues, than the previously largest protein structures solved by NMR, namely human (Forman-Kay *et al.*, 1991) and *E. coli* (Dyson *et al.*, 1990) thioredoxin which have 105 and 108 residues, respectively. Moreover, IL-1β still represents one of the most highly refined and precise structures for proteins of this size resolved by NMR.

Despite extensive analysis of 2D spectra obtained at different pH values and temperatures, as well as examination of 2D spectra of mutant proteins, it did not prove feasible to obtain unambiguous ^{1}H assignment for more than about 30% of the residues of interleukin-1β (Driscoll *et al.*, 1990a). Thus, any further progress could only be made by resorting to higher dimensionality heteronuclear NMR. The initial step involved the complete assignment of the ^{1}H, ^{15}N and ^{13}C resonances of the backbone and side chains using many of the double and triple resonance 3D experiments listed in Table 1 (Driscoll *et al.*, 1990a,b; Clore *et al.*,

1990a). In the second step backbone and side chain torsion angle restraints, as well as stereospecific assignments for β-methylene protons, were obtained by means of a three-dimensional systematic grid search of ϕ,ψ,χ_1 space (Nilges *et al.*, 1990). In the third step, approximate interproton distance restraints between nonadjacent residues were derived from analysis of 3D and 4D heteronuclear-edited NOE spectra. Analysis of the 3D heteronuclear-edited NOE spectra alone was sufficient to derive a low resolution structure on the basis of a small number of NOEs involving solely NH, $C^{\alpha}H$ and $C^{\beta}H$ protons (Clore *et al.*, 1990b). However, further progress using 3D NMR was severely hindered by the numerous ambiguities still present in these spectra, in particular for NOEs arising from the large number of aliphatic protons. Thus, the 4D heteronuclear-edited NOE spectra proved to be absolutely essential for the successful completion of this task. In addition, the proximity of backbone NH protons to bound structural water molecules was ascertained from a 3D ^{15}N-separated ROESY spectrum which permits one to distinguish specific protein–water NOE interactions from chemical exchange with bulk solvent (Clore *et al.*, 1990c). In this regard it should be emphasized that all the NOE data were interpreted in as conservative a manner as possible, and were simply classified into three distance ranges, 1.8–2.7 Å, 1.8–3.3 Å and 1.8–5.0 Å corresponding to strong, medium and weak intensity NOEs.

With an initial set of experimental restraints in hand, 3D structure calculations were initiated using the hybrid distance geometry–dynamical simulated annealing method (Nilges *et al.*, 1988a). A key aspect of the overall strategy lies in the use of an iterative approach whereby the experimental data is reexamined in the light of the initial set of calculated structures in order to resolve ambiguities in NOE assignments, to obtain more stereospecific assignments (e.g. the α-methylene protons of glycine and the methyl groups of valine and leucine) and torsion angle restraints, and to assign backbone hydrogen bonds associated with slowly exchanging NH protons as well as with bound water molecules. The iterative cycle comes to an end when all the experimental data have been interpreted.

The final experimental data set for interleukin-1β comprised a total of 3146 approximate and loose experimental restraints made up of 2780 distance and 366 torsion angle restraints (Clore *et al.*, 1991a). This represents an average of ~21 experimental restraints per residue. If one takes into account that interresidue NOEs affect two residues, while intraresidue NOE and torsion angle restraints only affect individual residues, the average number of restraints influencing the conformation of each residue is approximately 33. Superpositions of the backbone atoms and selected side chains for 32 independently calculated structures are shown in Figs 6B and D. All 32 structures satisfy the experimental restraints within their specified errors, display very small deviations from idealized covalent geometry, and have good nonbonded contacts. It can be seen that both the backbone as well as ordered side chains are exceptionally well-defined. Indeed, the atomic rms distribution about the mean coordinate positions is 0.4 Å for the backbone atoms, 0.8 Å for all atoms, and 0.5 Å for side chains with $\lesssim$40% of their surface (relative to that in a tripeptide Gly–X–Gly) accessible to solvent (Clore *et al.*, 1991a).

The structure of interleukin-1β itself resembles a tetrahedron and displays three-fold internal pseudo-symmetry. There are 12 β-strands arranged in an exclusively antiparallel β-structure, and six of the strands form a β-barrel (seen in the front of Fig. 6A) which is closed off at the back of the molecule by the other six strands. Each repeating topological unit is composed of 5 strands arranged in an antiparallel manner with respect to each other, and one of these units is shown in Fig. 6C. Water molecules occupy very similar positions in all three topological units, as well as at the interface of the three units, and are involved in bridging backbone hydrogen bonds. Thus, in the case of the topological unit shown in Fig. 6C, the water molecule labeled W5 accepts a hydrogen bond from the NH of Phe-112 in strand IX and donates two hydrogen bonds to the backbone carbonyls of Ile-122 in strand X and Thr-144 in strand XII. The packing of some internal residues with respect to one another, as well as the excellent definition of internal side chains is illustrated in Fig. 6D. Because of the high resolution of the interleukin-1β structure it was possible to analyze in detail side chain–side chain interactions involved in stabilizing the structure. In addition, examination of the structure in the light of mutational data permitted us to propose the

presence of three distinct sites involved in the binding of interleukin-1β to its cell surface receptor (Clore *et al.*, 1991a).

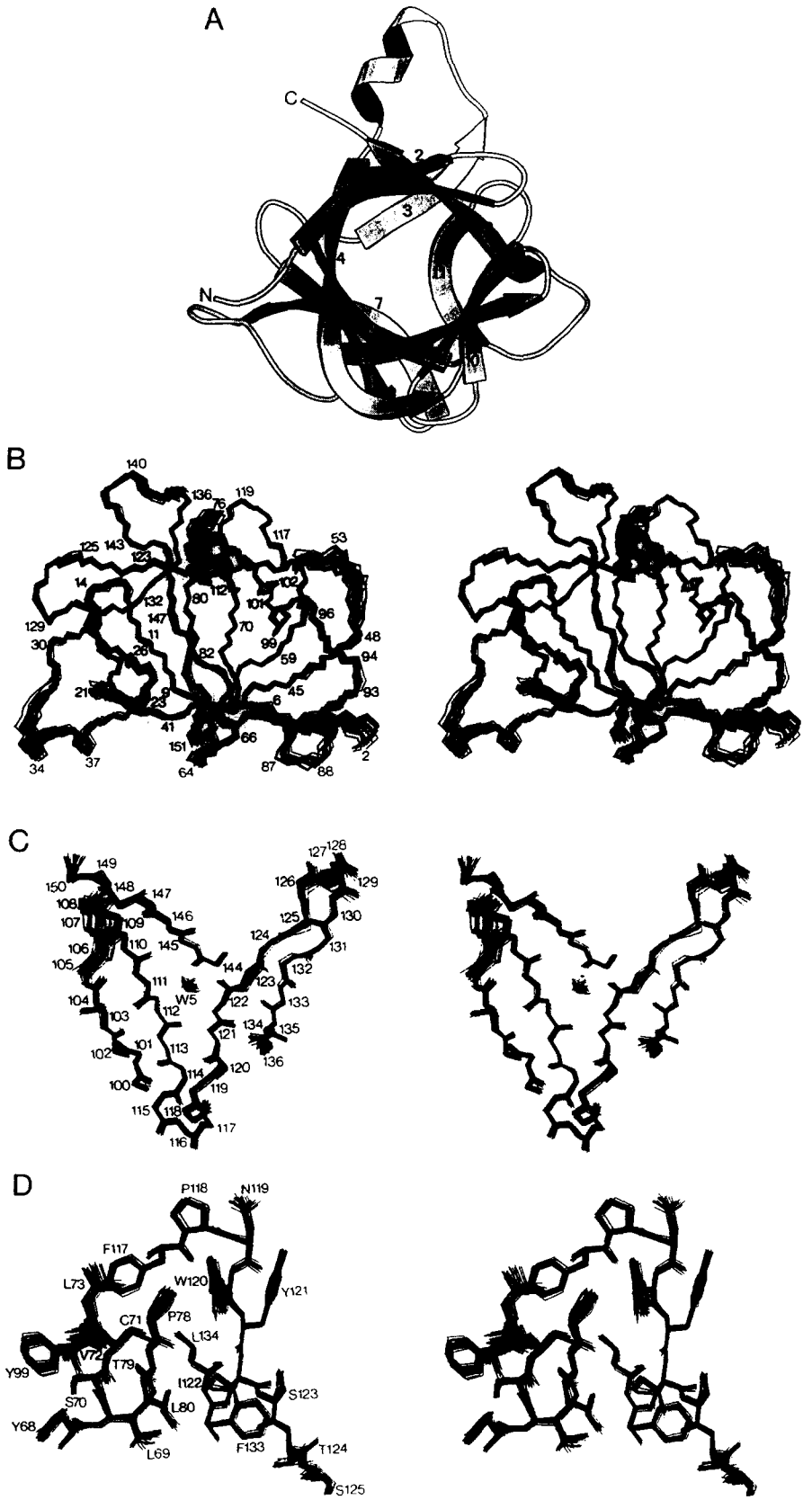
IV. COMBINING EXPERIMENTAL INFORMATION FROM CRYSTAL AND SOLUTION STUDIES: JOINT X-RAY AND NMR REFINEMENT

It is clear from the preceding discussion that NMR is a valid method, alongside X-ray crystallography, for determining high resolution structures of small to medium sized proteins of less than about 35 kDa. Interleukin-1β offers an ideal system for comparing the results of NMR and X-ray crystallography as, in addition to the solution structure, there are three independently solved X-ray structures at 2 Å resolution of the same crystal form (Finzel *et al.*, 1989; Priestle *et al.*, 1989; Veerapandian *et al.*, 1992). The backbone atomic rms difference between the NMR and the X-ray structures is about 1 Å with the largest differences being confined to some of the loops and turns connecting the β-strands (Clore and Gronenborn, 1991b). Interestingly, however, the atomic rms distribution of the 32 calculated solution structures about their mean coordinate positions ($\sim$0.4 Å for the backbone atoms, $\sim$0.8 Å for all atoms and $\sim$0.5 Å for all atoms of internal residues) is approximately the same as the atomic rms differences between the three X-ray structures, indicating that the positional errors in the atomic coordinates determined by the two methods are similar (Clore and Gronenborn, 1991b). Upon initial inspection, the X-ray structures appear to be incompatible with the NMR data, as manifested by a relatively large number of NOE and torsion angle violations and conversely, the NMR structure fits the X-ray data poorly with an R-factor of 40–50%. Because of the very different nature of the two methods, it is not immediately apparent that these discrepancies reflect genuine differences between the solution and crystal structures or whether they reflect differences in the computational procedures employed. To analyze this in more detail we have developed a new method of structure determination in which the NMR and X-ray data are combined and used simultaneously in the structure refinement (Shaanan *et al.*, 1992). Using this approach we have shown that a model can readily be generated from a joint NMR/X-ray refinement which is compatible with the data from both techniques. Thus, there are only minimal violations of the NMR restraints (NOEs and torsion angles), the value of the crystallographic R-factor is comparable to, if not better than that derived from refinement against the crystallographic data alone, and the deviations from idealized covalent geometry are small. In addition the R free (Brünger, 1992), for the model refined with the NMR and X-ray restraints, is smaller than that of the model obtained by conventional crystallographic refinement, indicating that the crystallographic phases obtained by the joint NMR/X-ray refinement are more accurate. Moreover, the few NMR observations that are still violated by the model serve as an indicator for genuine differences between the crystal and solution structures.

The implications of the joint NMR/X-ray refinement method to structural biology are of considerable significance. In particular, the full potential and future use of the method will be for structure determinations of multidomain proteins, for which only low resolution X-ray data for the entire protein are available but for which detailed structural information may be obtained by NMR on the individual domains. Using the joint X-ray/NMR refinement approach in such cases will open the way to the study of proteins which may otherwise never be structurally accessible by either of the two methods alone.

V. STRUCTURE DETERMINATION OF PROTEIN–PEPTIDE AND PROTEIN–DNA COMPLEXES

Providing the ligand (e.g. a peptide, an oligonucleotide, a drug, etc.) presents a relatively simple spectrum that can be assigned by 2D methods, the most convenient strategy for dealing with protein–ligand complexes involves one in which the protein is labeled with ^{15}N and ^{13}C and the ligand is unlabeled (i.e. at natural isotopic abundance) (Ikura and Bax, 1992). It is then possible to use a combination of heteronuclear filtering and editing to design experiments in which correlations involving only protein resonances, only ligand resonances, or only through-space interactions between ligand and protein are observed. These experiments are summarized in Table 2 and were first applied successfully to a


A
C
N
1
2
3
4
5
6
7
8
9
10
B
C
D
P118
N119
F117
L73
W120
Y121
C71
P78
L134
V72
T79
Y99
I122
S123
S70
L80
Y68
L69
F133
T124
S125

complex of calmodulin with a target peptide from skeletal muscle myosin light chain kinase (Ikura *et al.*, 1992), and subsequently to the specific complex of the DNA binding domain of the transcription factor GATA-1 with its cognate DNA target site (Omichinski *et al.*, 1993).

1. *The Structure of the Calmodulin–Target Peptide Complex*

Calmodulin (CaM) is a ubiquitous Ca^{2+} binding protein of 148 residues which is involved in a wide range of cellular Ca^{2+}-dependent signaling pathways, thereby regulating the activity of a large number of proteins (Cohen and Klee, 1988). The crystal structure of Ca^{2+}–CaM had been solved a number of years ago (Babu *et al.*, 1985). It is a dumb-bell shaped molecule with an overall length of ~65 Å consisting of two globular domains, each of which contains two Ca^{2+} binding sites of the helix–loop–helix type, connected by a long, solvent exposed, rigid central helix some eight turns in length (residues 66–92). In solution, on the other hand, ^{1}H–^{15}N NMR relaxation measurements have demonstrated unambiguously that the central helix is disrupted near its mid-point with residues 78–81 adopting an essentially unstructured "random coil" conformation which is so flexible that the N- and C-terminal domains of Ca^{2+}–CaM effectively tumble independently of each other (Barbato *et al.*, 1992). Thus, in solution, the so-called "central helix" is not a helix at all but is a "flexible tether" whose purpose is to keep the two domains in close proximity for binding to their target.

In order to understand the way in which Ca^{2+}–CaM recognizes its target sites, we set out to solve, in collaboration with Ad Bax, the solution structure of a complex of Ca^{2+}–CaM with a 26 residue peptide (known as M13) comprising residues 577–602 of the calmodulin binding domain of skeletal muscle myosin light chain kinase. The solution structure was determined on the basis of 1995 experimental NMR restraints including 133 interproton distance restraints between the peptide and the protein. The N- (residues 1–5) and C- (residues 147–148) termini of CaM, the tether connecting the two domains of CaM (residues 74–82), and the N- (residues 1–2) and C-termini (residues 22–26) of M13 were ill-defined by the NMR data and appear to be disordered in solution. The atomic rms distribution about the mean coordinate positions for the rest of the structure (i.e. residues 6–73 and 83–146 of CaM and residues 3–21 of M13) is 1.0 Å for the backbone atoms and 1.4 Å for all atoms. Thus this structure represents a second generation structure in the classification of Clore and Gronenborn (1991a). A stereoview showing a best fit superposition of the 24 calculated structures is shown in Fig. 7a.

The major conformational change in Ca^{2+}–CaM that occurs upon binding M13 involves an extension of the flexible tether (residues 78–81) in the middle of the "central" helix of the solution structure of free Ca^{2+}–CaM to a long flexible loop extending from residues 74–81, flanked by two helices (residues 65–73 and 83–93), thereby enabling the two domains to come together gripping the peptide rather like two hands capturing a rope. The hydrophobic channel formed by the two domains is complementary in shape to that of the peptide helix. This is clearly illustrated by the schematic ribbon drawings shown in Figs 7b and c which also serve to highlight the approximate two-fold pseudo-symmetry of the complex. Thus, whereas the two domains of CaM are arranged in an approximately orthogonal manner to each other in the crystal structure of Ca^{2+}–CaM (Babu *et al.*, 1985), in the Ca^{2+}–CaM–M13 complex they are almost symmetrically related by a 180° rotation about a two-fold axis. A large conformational change also occurs in the M13 peptide upon complexation from a random coil state to a well-defined helical conformation. Indeed, the helix involves all the

FIG. 6. (*opposite*) Solution structure of interleukin-1β determined by 3D and 4D heteronuclear NMR spectroscopy. (A) Ribbon diagram of the polypeptide fold. (B) Superposition of the backbone (N, C^{α}, C) atoms of 32 simulated annealing structures calculated from the experimental NMR data. (C) Superposition of the backbone (N, C^{α}, C, O) atoms of one of the three repeating topological units, illustrating the position of tightly bound water at the interface of the three central strands of the unit. (D) Superposition of all atoms (excluding protons) for selected side chains. The diagram in (A) was made with the program MOLSCRIPT (Kraulis, 1991). The coordinates are from Clore *et al.* (1991a). (PDB accession code 6I1B.)

TABLE 2. SUMMARY OF HETERONUCLEAR-FILTERED AND -EDITED NOE EXPERIMENTS USED TO STUDY PROTEIN–LIGAND COMPLEXES COMPRISING A UNIFORMLY $^{15}N/^{13}C$ LABELED PROTEIN AND AN UNLABELED LIGAND

Type of contact	Connectivity
A. Intramolecular protein contacts	
4D $^{13}C/^{13}C$-edited NOE in D_2O	$H(j)$–$^{13}C(j)$– – – – – – – – –$H(i)$–$^{13}C(i)$
4D $^{15}N/^{13}C$-edited NOE in H_2O	$H(j)$–$^{15}N(j)$– – – – – – – – –$H(i)$–$^{13}C(i)$
3D $^{15}N/^{15}N$-edited NOE in H_2O	$H(j)$–$^{15}N(j)$– – – – – – – – –$H(i)$–$^{15}N(i)$
B. Intramolecular ligand contacts	
2D $^{12}C,^{14}N(F_1)/^{12}C,^{14}N(F_2)$ filtered NOE in H_2O	$H(j)$–$^{12}C(j)$– – – – – – – – –$H(i)$–$^{12}C(i)$
	$H(j)$–$^{14}N(j)$– – – – – – – – –$H(i)$–$^{12}C(i)$
	$H(j)$–$^{12}C(j)$– – – – – – – – –$H(i)$–$^{14}N(i)$
	$H(j)$–$^{14}N(j)$– – – – – – – – –$H(i)$–$^{14}N(i)$
2D $^{12}C(F_1)/^{12}C(F_2)$ filtered NOE in D_2O*	$H(j)$–$^{12}C(j)$– – – – – – – – –$H(i)$–$^{12}C(i)$
C. Intermolecular protein–ligand contacts	
3D ^{15}N-edited$(F_1)/^{14}N,^{12}C(F_3)$ filtered NOE in H_2O	$H(j)$–$^{15}N(j)$– – – – – – – – –$H(i)$–$^{12}C(i)$
	$H(j)$–$^{15}N(j)$– – – – – – – – –$H(i)$–$^{14}N(i)$
3D ^{13}C-edited$(F_1)/^{12}C(F_3)$ filtered NOE in D_2O	$H(j)$–$^{13}C(j)$– – – – – – – – –$H(i)$–$^{12}C(i)$

*Similar heteronuclear filtered 2D correlation and Hartmann–Hahn spectra can also be recorded to assign the spin systems of the ligand.

residues (3–21) of M13 that interact with CaM, while the N- (residues 1–2) and C- (residues 22–26) termini of the peptide, which do not interact with CaM, remain disordered.

Upon complexation there is a decrease in the accessible surface area of CaM and M13 of 1848 and 1477 Å^2, respectively, which corresponds to a decrease in the calculated solvation free energy of folding (Eisenberg and McLaghlan, 1986) of 18 and 20 kcal·mol^{-1}, respectively. This large decrease in solvation free energy would account for the very tight binding ($K_{ass} \sim 10^9$ M^{-1}) of M13 to calmodulin. In addition, the accessible surface area of the portion of M13 (residues 3–21) in direct contact with CaM in the complex is only 494 Å^2 compared to an accessible surface area of 3123 Å^2 for a random coil and 2250 Å^2 for a helix. Thus, over 80% of the surface of the peptide in contact with CaM is buried.

In the view shown in Fig. 7b, the roof of the channel is formed by helices II (residues 29–38) and VI (residues 102–111) of the N- and C-terminal domains, respectively, which run antiparallel to each other; and the floor is formed by the flexible loop (residues 74–82) connecting the two domains and by helix VIII (residues 138–146) of the C-terminal domain. The front of the channel in Fig. 7b and the left wall of the channel in Fig. 7c is formed by helices I (residues 7–19) and IV (residues 65–73) and the mini-antiparallel β-sheet comprising residues 26–28 and 62–64), all from the N-terminal domain; the back of the channel in Fig. 7b and the right wall of the channel in Fig. 7c is formed by helices V (residues 83–93) and VIII (residues 138–146) and the mini-antiparallel β-sheet comprising residues 99–101 and 135–137, all from the C-terminal domain. The two domains of CaM are staggered with a small degree of overlap such that the hydrophobic face of the N-terminal domain mainly contacts the C-terminal half of the M13 peptide, while the C-terminal domain principally interacts with the N-terminal half of M13 (Fig. 7b).

The overall Ca^{2+}–CaM–M13 complex has a compact globular shape approximating to an ellipsoid with dimensions 47 × 32 × 30 Å. The helical M13 peptide passes through the center of the ellipsoid at an angle of ~45° to its long axis. By way of contrast the approximate dimensions of the Ca^{2+}–CaM X-ray structure are 65 × 30 × 30 Å (Babu *et al.*, 1985). In addition, the calculated radius of gyration for Ca^{2+}–CaM–M13 is ~17 Å which is completely consistent with the decrease in the radius of gyration from ~21 to ~16 Å observed by both small angle X-ray and neutron scattering upon complexation of Ca^{2+}–CaM with M13 (Heidorn *et al.*, 1989).

The Ca^{2+}–CaM–M13 complex is stabilized by numerous hydrophobic interactions which are summarized in Fig. 8. Particularly striking are the interactions of Trp-^{4} and Phe-^{17} of the peptide which serve to anchor the N- and C-terminal halves of M13 to the C-terminal and N-terminal hydrophobic patches of CaM, respectively (Fig. 7c). These interactions also involve a large number of methionine residues which are unusually abundant in CaM, in

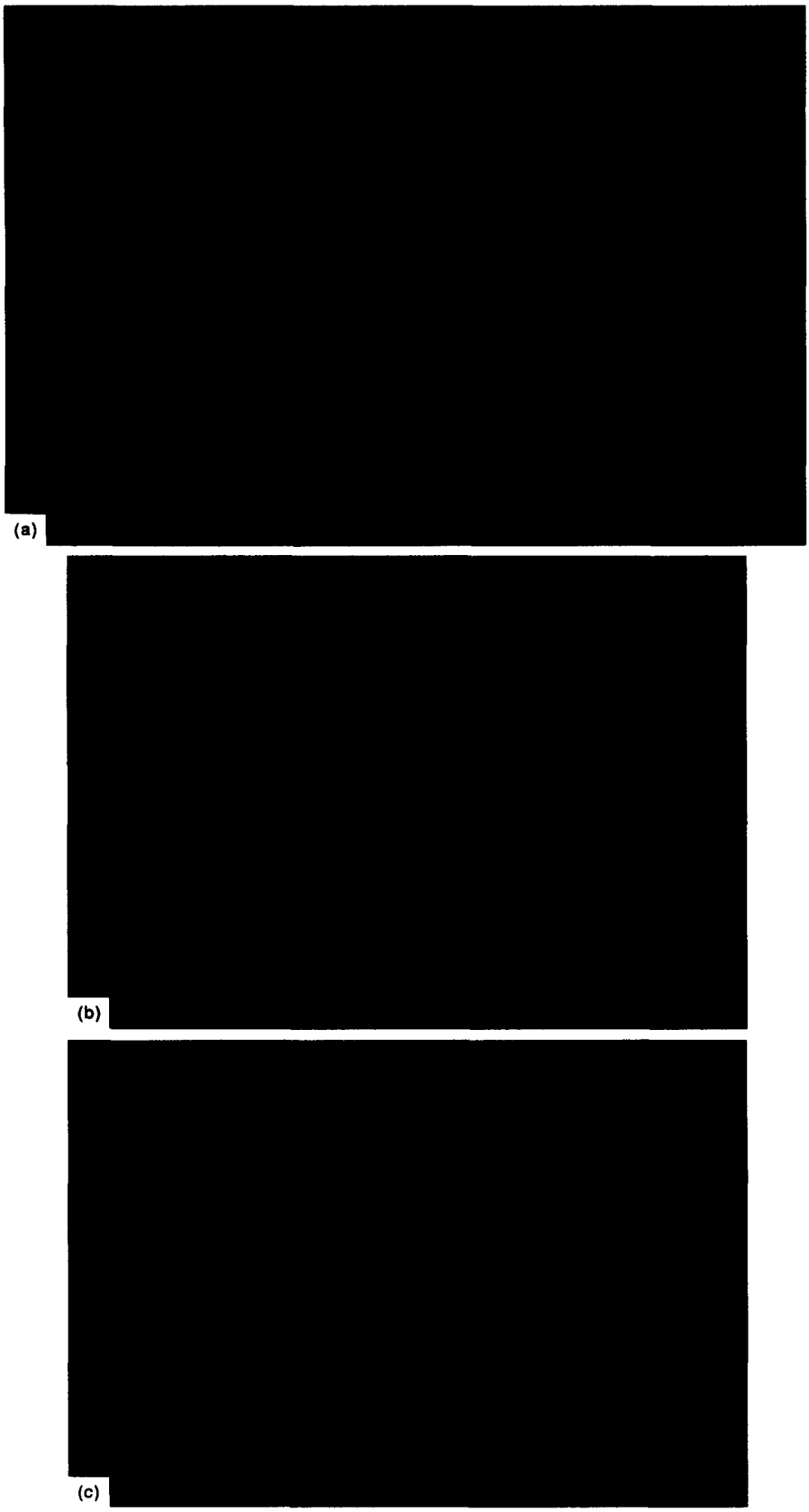


FIG. 7. Solution structure of the Ca^{2+}–CaM–M13 peptide complex determined by 3D and 4D heteronuclear NMR spectroscopy. (a) Superposition of the backbone (N, C^{α}, C) atoms of 24 simulated annealing structures calculated from the experimental NMR data; the N- and C-terminal domains of calmodulin are shown in blue and red, respectively, and the M13 peptide is in green; the restrained regularized average structure is highlighted. (b) and (c) Two orthogonal views of a schematic ribbon drawing representation of the structure with the N- and C-terminal domains of CaM in blue and purple, respectively, the M13 peptide in yellow, the hydrophobic side chains of the protein in red, and Trp^{4}, Phe^{8}, Val^{11} and Phe^{17} side chains of the peptide in green. The diagrams in (b) and (c) were generated with the program VISP (de Castro and Edelstein, 1992). The coordinates are from Ikura *et al.* (1992). (PDB accession code 1BBM.)

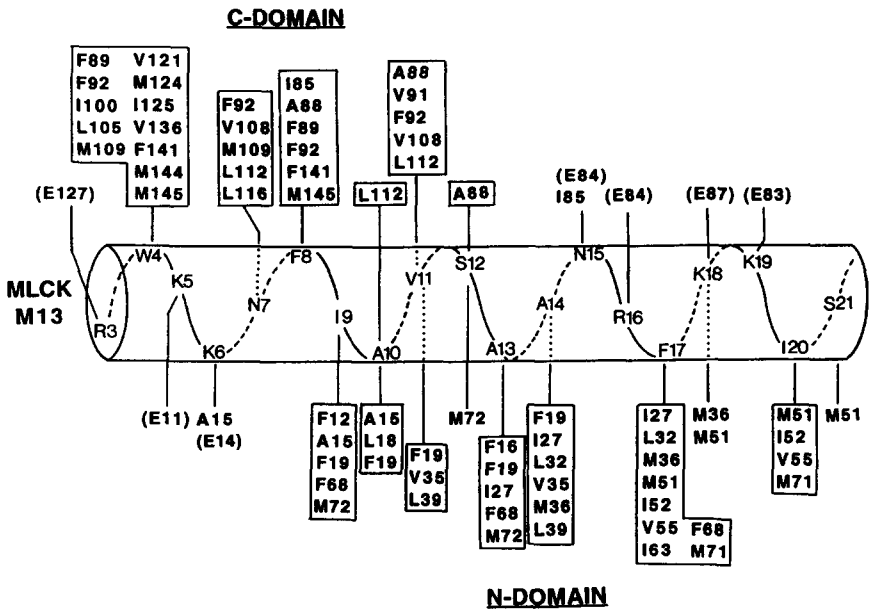


FIG. 8. Summary of residue pairs for which intermolecular NOEs between CaM and M13 are observed. CaM residues involved in hydrophobic interactions are boxed. Also included are potential electrostatic interactions between negatively charged Glu residues of CaM (shown in parentheses) and positively charged Lys and Arg residues of M13.

particular four methionines in the C-terminal domain (Met-^{109}, Met-^{124}, Met-^{144} and Met-^{145}) and three methionines in the N-terminal domain (Met-^{36}, Met-^{51} and Met-^{71}). As methionine is an unbranched hydrophobic residue extending over four heavy atoms (C^{β}, C^{γ}, S^{δ}, C^{ε}), the abundance of methionines can generate a hydrophobic surface whose detailed topology is readily adjusted by minor changes in side chain conformation, thereby providing a mechanism to accommodate and recognize different bound peptides (O'Neil and DeGrado, 1990).

In addition to hydrophobic interactions, there are a number of possible electrostatic interactions that can be deduced from the calculated NMR structures. Putative interactions exist between the Arg and Lys residues of M13 and the Glu residues of CaM, and these are also included in Fig. 8. Glu-^{11} and Glu-^{14} in helix I are within 5 Å of Lys-^{5} and Lys-^{6} of M13; Glu-^{83}, Glu-^{84} and Glu-^{87} in helix V of CaM are close to Lys-^{19}, Arg-^{16} and Lys-^{18}, respectively, of M13; and Glu-^{127} in helix VII of CaM is close to Arg-^{13} of M13.

The solution structure of the Ca^{2+}–CaM–M13 complex explains a number of interesting observations. Studies of backbone amide exchange behavior have shown that upon complexation with M13, the amide exchange rates of residues 75–79 are substantially increased (Spera *et al.*, 1991). Prior NMR studies on Ca^{2+}–CaM indicated that the long central helix is already disrupted near its middle (from Asp-^{78} to Ser-^{81}) in solution (Ikura *et al.*, 1991) and that large variations in the orientation of one domain relative to the other occur randomly with time (Barbato *et al.*, 1992). The further disruption of the central helix upon complexation seen in the structure of the complex is manifested by the increased amide exchange rates and supports the view of the central helix serving as a flexible linker between the two domains. Similarly, the structure of the complex explains the finding that as many as four residues can be deleted from the middle of the central helix without dramatically altering the stability or shape of the Ca^{2+}–CaM–M13 complex (Persechini *et al.*, 1989; Kataoka *et al.*, 1991), as the long flexible loop connecting the two domains can readily be shortened without causing any alteration in the structure (*cf.* Fig. 7). The observation from photoaffinity labeling studies that the two domains of CaM interact simultaneously with opposite ends of the peptide such that residue 4 of the peptide (numbering for M13) can be cross-linked to Met-^{124} or Met-^{144} of the C-terminal domain and that residue 13 of the

peptide can be cross-linked to Met-^{71} of the N-terminal domain (O'Neil *et al.*, 1989), is readily explained by the structural finding that the N-terminal half of the peptide interacts predominantly with the C-terminal domain while the C-terminal half of the peptide interacts predominantly with the N-terminal domain (Figs 7 and 8). The observation that at least 17 residues of the M13 peptide from either skeletal muscle or smooth muscle are necessary for high affinity binding (Lukas *et al.*, 1986; Blumenthal and Krebs, 1987) is readily explained by the intimate interactions of the C-terminal hydrophobic residue (i.e. Phe-^{17}) with the N-terminal domain of CaM by which the peptide is anchored. Finally, the structure accounts for experiments in which cross-linking of residues 3 and 146 of CaM, mutated to Cys, has no effect on the activation of myosin light chain kinase, even if the central helix is cleaved proteolytically at Lys-^{77} by trypsin (Persechini and Kretsinger, 1986). Thus, while the C^{α} carbons of the residues 3 and 146 are 37 Å apart in the X-ray structure Ca^{2+}–CaM, they are only ~20 Å apart in the solution structure of the Ca^{2+}–CaM–M13 complex, which is close enough to permit cross-linking to occur.

A large body of experimental data shows that CaM binds to numerous proteins whose binding domains exhibit a propensity for α-helix formation (Cohen and Klee, 1988). A comparison of these sequences reveals little homology. Nevertheless, many of the very tightly binding peptides ($K_{ass} \geq 5 \times 10^7$ M^{-1}) have the common property of containing either aromatic residues or long chain hydrophobic residues (Leu, Ile or Val) separated by 12 residues, as summarized in Fig. 9. In the case of M13, these two residues are Trp-^{4} and Phe-^{17}

```
                        1       5         10        15        20        25
SK-MLCK M13               K R R W K K N F I A V S A A N R F K K I S S S G A L M
SM-MLCK M13               R R K W Q K T G H A V R A I G R L S S S
Ca Pump C24W              Q I L W F R G L N R I Q T Q I R V V N A F R S S
        C20W    L R R G Q I L W F R G L N R I Q T Q I K
Calspermin          A R R K L K A A V K A V V A S S R L G S
Calcineurin       A R K E V I R W K I R A I G K M A R V S F V L
Mastaporan                  I N L K A L A A L A K K I L
Mastaporan X                I N W K G I A A M A K K L L
Mellitin          G I G A V L K V L T T G L P A L I S W I K R K R Q Q

Interacting                 C       C     C         N
domain of CaM                                 N
```

FIG. 9. Alignment of tightly binding ($K_{ass} > 5 \times 10^7$ M^{-1}) CaM binding sequences based on the structural role of Trp^4 and Phe^{17} in anchoring the M13 peptide to the C- and N-terminal domains of CaM, respectively.

which are exclusively in contact with the C- and N-terminal domains of CaM, respectively (Figs 7 and 8). Given that these two residues are involved in more hydrophobic interactions with CaM than any other residues of the peptide (*cf.* Fig. 8), it seems likely that this feature of the sequence can be used to align the CaM binding sequences listed in Fig. 9, thereby permitting one to predict their interaction with CaM. It is clear from this alignment that the pattern of hydrophobic and hydrophilic residues is in general comparable for the various peptides, suggesting that the mode of binding and the structure of the corresponding complexes with Ca^{2+}–CaM are also likely to be similar. For example, there is, in general, conservation of hydrophobic residues at the positions equivalent to Phe-^{8} which interacts with the C-terminal domain and Val-^{11} which interacts with both domains (*cf.* Figs 7 and 8). In addition, there are no acidic residues present which would result in unfavorable electrostatic interactions with the negatively charged Glu residues on the surface of CaM (*cf.* Fig. 7). The minimum length of peptide required for high affinity binding to Ca^{2+}–CaM is defined by the 14 residue mastaporans which comprise the two hydrophobic residues at the N- and C-termini (Fig. 9) and have approximately the same equilibrium association constant ($K_{ass} \sim 1$–3×10^9 M^{-1}) as M13 (Cox *et al.*, 1985). This structural alignment also predicts that a peptide stopping just short of the second hydrophobic residue of the pair (i.e. the residue equivalent to Phe-^{17}) would only bind to the C-terminal domain and that the

resulting complex would therefore retain the dumb-bell shape of Ca^{2+}–CaM. This is exactly what has been observed by small angle X-ray scattering using two synthetic peptides, C24W and C20W (Fig. 9), comprising different portions of the CaM binding domain of the plasma membrane Ca^{2+} pump (Kataoka *et al.*, 1991). The complex with the C24W peptide which corresponds to residues 1–24 of M13 and contains a Trp at position 4 and a Val at position 17, has a globular shape similar to that of Ca^{2+}–CaM–M13. The complex with the C20W peptide, on the other hand, which corresponds to residues −4 to 16 of M13 and therefore lacks the C-terminal hydrophobic residue of the pair, retains the dumb-bell shape of Ca^{2+}–CaM suggesting that the peptide only binds to the C-terminal domain.

Thus the solution structure of the complex of Ca^{2+}–CaM with M13 reveals an unusual binding mode in which the target peptide is sequestrated into a hydrophobic channel formed by the two domains of CaM with interactions involving 19 residues of the target peptide (i.e. residues 3–21 of M13). In addition, a key requirement appears to be the presence of two long chain hydrophobic or aromatic residues separated by 12 residues in order to anchor the peptide to the two domains of CaM (Fig. 7). By anology, the rope (i.e. the CaM binding domain of the target) has to be long enough and have two knots at each end for the two hands (i.e. domains) of CaM to grip it. This particular mode of binding is therefore only likely to occur if the CaM binding site is located either at an easily accessible C- or N-terminus or in a long exposed surface loop of the target protein. An example of the former is myosin light chain kinase and of the latter is calcineurin, and, in accordance with their location, the CaM binding sites are susceptible to proteolysis (Blumenthal and Krebs, 1987; Guerini and Klee, 1991). Clearly, other types of complexes between Ca^{2+}–CaM and its target proteins are possible given the inherent flexibility of the central helix. For example, in the case of the γ subunit of phosphorylase kinase, it appears that there are two discontinuous CaM binding sites which are capable of binding to Ca^{2+}–CaM simultaneously (Dasgupta *et al.*, 1989), and binding of a peptide derived from one of these sites causes elongation rather than contraction of Ca^{2+}–CaM (Trewhella *et al.*, 1990), indicating that the complex is of a quite different structural nature. Similarly, in the case of cyclic nucleotide phosphodiesterase (Charbonneau *et al.*, 1991) and CaM kinase II (Bennett and Kennedy, 1987), the CaM binding sequences do not have the same spacing of hydrophobic residues seen in M13 and the other sequences listed in Fig. 13, and, in addition, CaM kinase II is not susceptible to proteolysis in the absence of phosphorylation (Kwiatokowski *et al.*, 1989), suggesting that the mode of binding is different again. Thus, in all likelihood, the complexes of Ca^{2+}–CaM with target peptides from skeletal and smooth muscle myosin light chain kinase represent one of a range of Ca^{2+}–CaM binding modes achieving CaM–target protein interactions in an efficient and elegant manner.

2. *The Structure of the Specific Complex of the Transcription Factor GATA-1 with DNA*

The erythroid specific transcription factor GATA-1 is responsible for the regulation of transcription of erythroid-expressed genes and is an essential component required for the generation of the erythroid lineage (Orkin, 1992). GATA-1 binds specifically as a monomer to the asymmetric consensus target sequence (T/A) GATA (A/G) found in the *cis*-regulatory elements of all globin genes and most other erythroid specific genes that have been examined (Evans and Felsenfeld, 1989). GATA-1 was the first member of a family of proteins, which now includes regulatory proteins expressed in other cell lineages, characterized by their recognition of the GATA DNA sequence and by the presence of two metal binding regions of the form Cys–X–X–Cys–$(X)_{17}$–Cys–X–X–Cys separated by 29 residues. Mutation and deletion studies on GATA-1 have indicated that the N-terminal metal binding region is not required for specific DNA binding (Martin and Orkin, 1986), and studies with synthetic peptides have demonstrated conclusively that a 59 residue fragment (residues 158–216 of chicken GATA-1) comprising the C-terminal metal binding region complexed to zinc and 28 residues C-terminal to the last Cys constitutes the minimal unit required for specific binding ($K_{ass} \sim 1.2 \times 10^8$ M^{-1}) (Omichinski *et al.*, 1992). In order to understand the mechanism of specific DNA recognition by GATA-1 we set out to solve the solution structure of the specific complex of a 66 residue fragment (residues 158–223) comprising the DNA binding domain of

chicken GATA-1 (cGATA-1) with a 16 base pair oligonucleotide containing the target sequence AGATAA, by means of multidimensional heteronuclear filtered and separated NMR spectroscopy (Omichinski *et al.*, 1993).

The structure calculations were based on a total of 1772 experimental NMR restraints, including 117 intermolecular interproton distance restraints between the protein and the DNA. A stereoview of a best-fit superposition of 30 calculated structures (residues 2–59 of the protein and base pairs 6–13 of the DNA) is shown in Fig. 10. The N- (residue 1) and C- (residues 60–66) termini of the protein are disordered. Base pairs 6–13 of the DNA are in contact with the cGATA-1 DNA binding domain and are well defined both locally and globally. The orientation, however, of the first 5 and last 3 base pairs of the DNA, which are not in contact with the protein, is poorly defined with respect to the core of the complex, although the conformation of each of these bases at a local level is reasonably well-defined. This is due to the fact that, in addition to their approximate nature, the interproton distance restraints within the DNA are solely sequential. Hence, they are inadequate to ascertain the relative orientation of base pairs separated by more than 5–6 steps with any great degree of precision and accuracy. The global conformation of the central 8 base pairs, on the other hand, is determined not only by the restraints within the DNA, but more importantly by the large number of intermolecular interproton distance restraints between the protein and DNA. The atomic rms distribution of the 30 SA structures about the mean coordinate positions for the complex proper (i.e. residues 2–59 of the protein and base pairs 6–13 of the DNA) is 0.70 ± 0.13 Å and 1.13 ± 0.08 Å for protein backbone plus DNA and all protein atoms plus DNA, respectively.

The protein can be devided into two modules: the protein core which consists of residues 2–51 and contains the zinc coordination site, and an extended C-terminal tail (residues 52–59).

A schematic ribbon drawing of the core is presented in Fig. 11a. The core starts out with a turn (residues 2–5), followed by two short irregular antiparallel β-sheets, a helix (residues 28–38) and a long loop (residues 39–51) which includes a helical turn (residues 44–47), as well as an Ω-like loop (residues 47–51). β-Strands 1 (residues 5–7) and 2 (residues 11–14) form the first β-sheet, while β-strands 3 (residues 18–21) and 4 (residues 24–27) form the second β-sheet.

Part of the core of the cGATA-1 DNA binding domain is structurally similar to that of the N-terminal zinc containing module of the DNA binding domain of the glucocorticoid receptor (Luisi *et al.*, 1991). Thus the C^{α} atoms of 30 residues of these two proteins can be superimposed with an rms difference of only 1.4 Å (Fig. 11b). Apart from the 4 Cys residues that coordinate the zinc atom, only 1 residue (Lys^{36} in the cGATA-1 DNA binding domain and Lys^{465} in the glucocorticoid receptor) is conversed between the two proteins. The structural similarity extends from the N-terminus up to the end of the helix (residues 3–39 of the cGATA-1 DNA binding domain and residues 436–468 of the glucocorticoid receptor), and the Zn–S^{γ} geometry, as well as the side chain conformations of the four coordinating cysteines, are identical. The loop between strands $\beta 2$ and $\beta 3$ has 3 deletions, and the turn between strands $\beta 3$ and $\beta 4$ has 1 deletion in the glucocorticoid receptor with respect to cGATA-1. The topology and polypeptide trace following the carboxy end of the helix, however, are entirely different in the two proteins. Thus, in the DNA binding domain of the glucocorticoid receptor there is a second compact zinc containing module (residues 470–514) made up of two strands and two helices, while in the cGATA-1 DNA binding domain there is a long loop (residues 38–51) and extended strand (residues 52–59).

The overall topology and structural organization of the complex is shown in Figs 12a and b. The conformation of the oligonucleotide is B-type. The helix and the loop connecting strands $\beta 2$ and $\beta 3$ (which is located directly beneath the helix) are located in the major groove, while the C-terminal tail wraps around the DNA and lies in the minor groove, directly opposite the helix. The overall appearance is analogous to that of a right hand holding a rope, with the rope representing the DNA, the palm and fingers of the hand the core of the protein, and the thumb the C-terminal tail. It is this pincer like configuration of the protein that causes a small 10° kink in the DNA. The long axis of the helix lies at an angle of

FIG. 10. Stereoview showing a superposition of the 30 simulated annealing structures of the specific complex of the DNA binding domain of cGATA-1 with DNA calculated on the basis of the experimental NMR data derived from 3D and 4D heteronuclear NMR spectroscopy. The backbone (N, C^{α}, C) atoms of cGATA-1 are shown in red and all the nonhydrogen atoms of the DNA in blue. The restrained regularized mean structure of the complex is highlighted. The coordinates are from Omichinski *et al.* (1993). (PDB accession code 1GAT.)

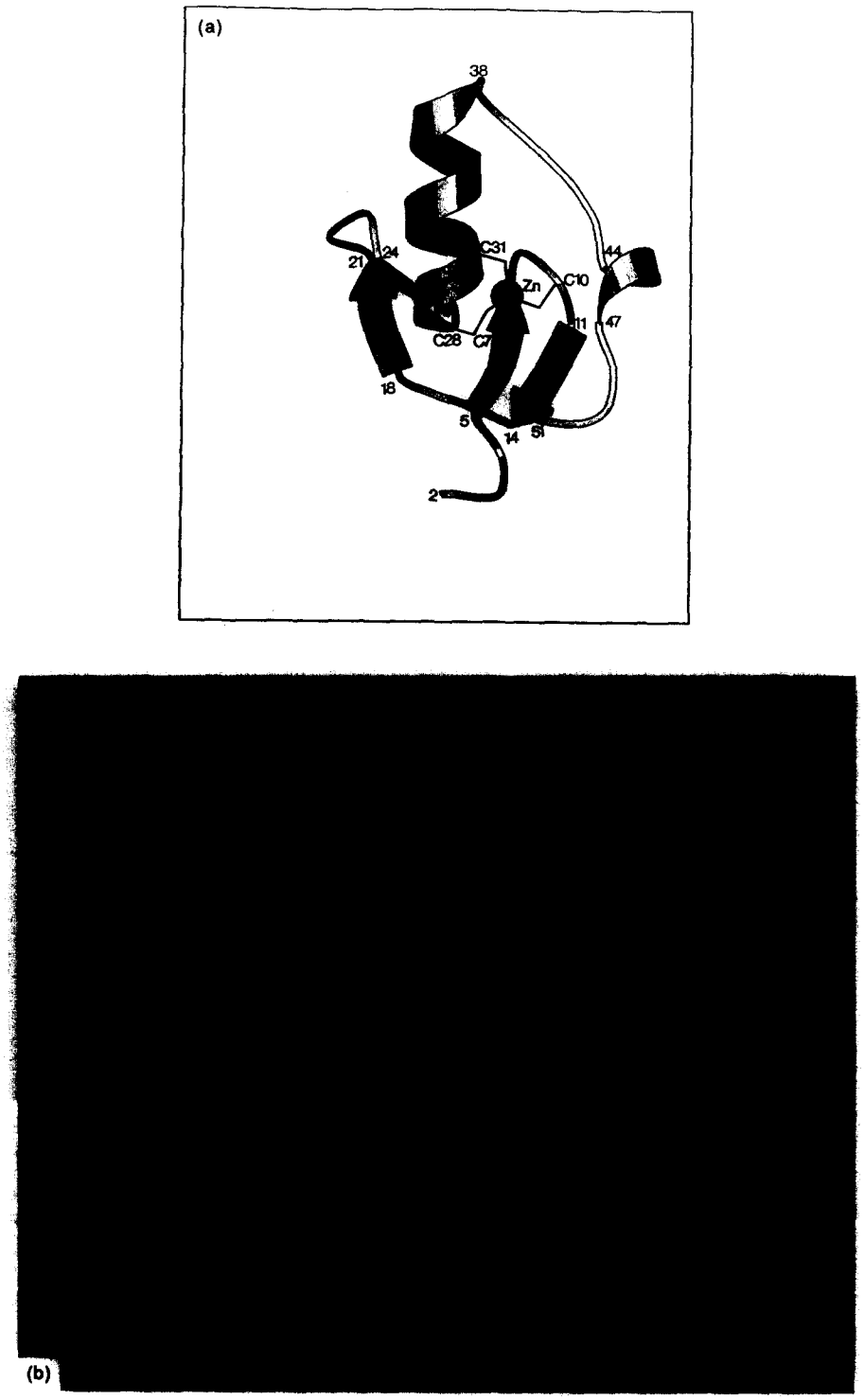


FIG. 11. (a) Schematic ribbon drawing of the core of the cGATA-1 DNA binding domain. (b) Superposition of the C^{α} atoms of the cGATA-1 (green) and glucocorticoid receptor (red) DNA binding domains. The zinc and coordinating cysteines are shown in yellow for cGATA-1 and in purple for the glucocorticoid receptor; the residues are labeled according to the numbering in cGATA-1. The alignment of cGATA-1 with the glucocorticoid receptor is as follows: residues 3–13, 18–21, 25–39 and 46 of cGATA-1 are superimposed on residues 436–446, 448–451, 454–468 and 490, respectively, of the glucocorticoid receptor with a C^{α} atomic rms of 1.4 Å. The diagram in (a) was made with the program MOLSCRIPT (Kraulis, 1991). The coordinates of the glucocorticoid receptor DNA binding domain shown in (b) is taken from Luisi *et al.* (1991). The cGATA-1 coordinates are from Omichinski *et al.* (1993). (PDB accession code 1GAT.)

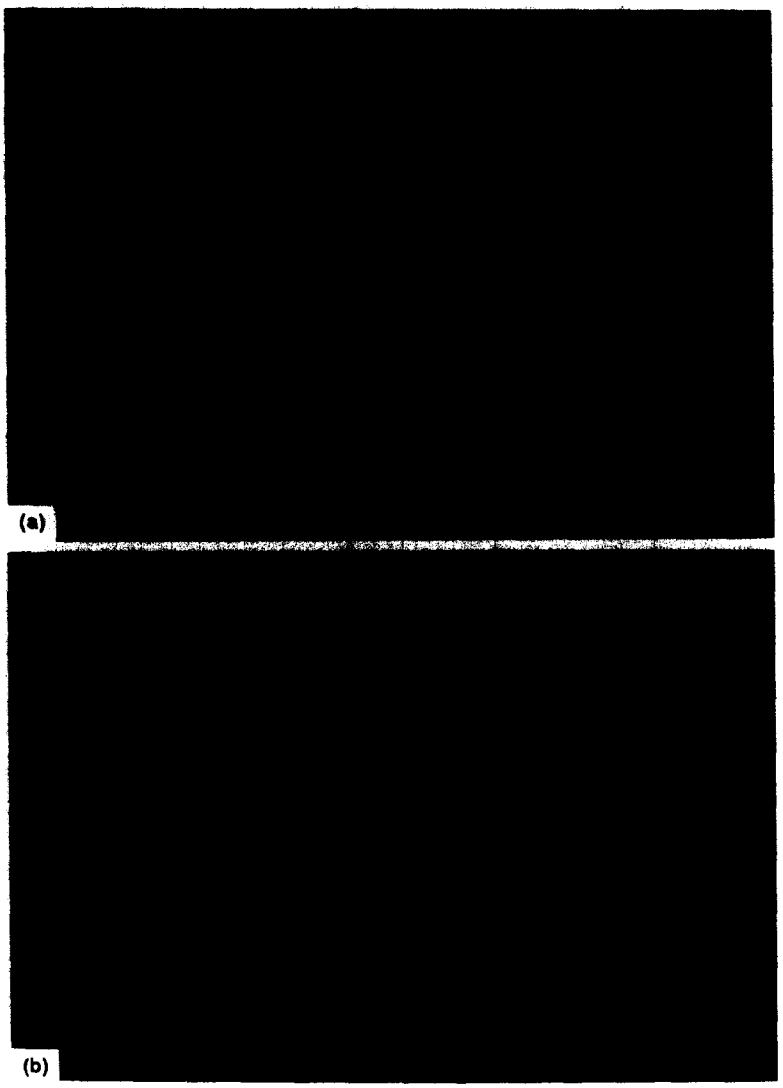


FIG. 12. (a) and (b). Schematic ribbon drawings illustrating the interactions of cGATA-1 with DNA.

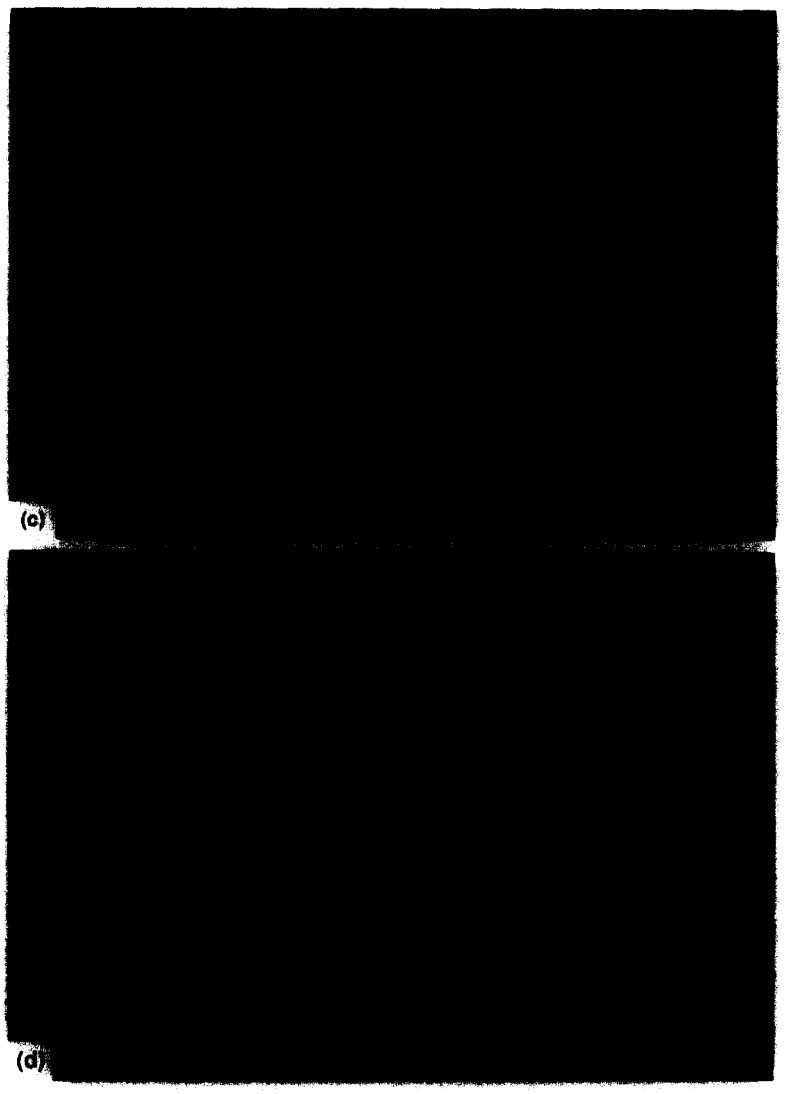


FIG. 12. (c) and (d) Side chain interactions between cGATA-1 and the DNA in the major and minor grooves, respectively. The protein backbone is shown in green and the protein side chains in yellow; the color code for the DNA bases is as follows: red for A, lilac for T, dark blue for G and light blue for C. The diagrams were made using the program VISP (de Castro and Edelstein, 1992). The coordinates of the cGATA-1–DNA complex are from Omichinski *et al.* (1993). (PDB accession code 1GAT.)

$\sim 40°$ to the base planes of the DNA (Fig. 12a), while the C-terminal tail is approximately parallel to the base planes (Fig. 12b).

Views of side chain contacts with the DNA in the major and minor grooves are shown in Figs 12c and d, respectively, while a schematic representation of all the contacts is provided in Fig. 13. The cGATA-1 DNA binding domain makes specific contacts with 8 bases, 7 in the

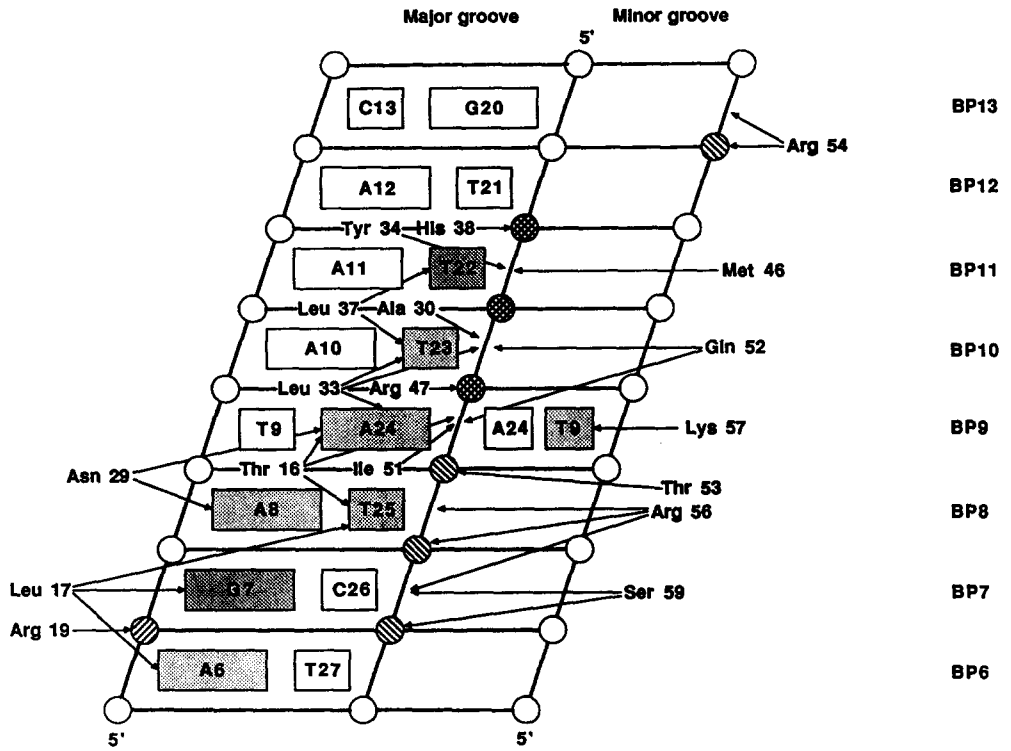


FIG. 13. Schematic diagram summarizing the contacts between cGATA-1 and DNA. The DNA is represented as a cylindrical projection. Bases interacting with the protein are shaded; phosphates are represented as circles; circles with hatches directed towards the right and left indicate sites of interaction with sugar or phosphate or both in the major and minor grooves, respectively.

major groove (A6, G7, A8, T25, A24, T23 and T22) and 1 in the minor groove (T9). All the base contacts in the major groove involve the helix and the loop connecting β-strands 2 and 3. In contrast to other DNA binding proteins, the majority of base contacts involve hydrophobic interactions. Thus, Leu^{17} interacts with A6, G7 and T25, Thr^{16} with A24 and T25, Leu^{33} with A24 and T23, and Leu^{37} with T23 and T22. This accounts for the predominance of thymidines in the DNA target site. Indeed, there are only three hydrogen bonding interactions: namely, between the side chain of Asn^{29} and the N6 atoms of A24 and A8 in the major groove; and between the $N\zeta H_3^+$ of Lys^{57} and the O2 atom of T9 in the minor groove. In this regard, it is interesting to note that there is a reduction of 1127 $Å^2$ in the surface accessible area of the cGATA-1 DNA binding domain in the presence of DNA (corresponding to a 20% decrease in the accessible surface), and a decrease in the calculated solvation free energy of folding (Eisenberg and McLaghlan, 1986) of 13 kcal.mol^{-1}. This latter effect can clearly make a sizeable contribution to the specific binding constant (K_{ass} $\sim 1.2 \times 10^8$ M^{-1}).

The remaining contacts involve the sugar–phosphate backbone, the majority of which are located on the second strand (that is G20 to T27). Salt bridges and/or hydrogen bonds with the phosphates of G7, A24 and T22 are made by Arg^{19}, Arg^{47} and His^{38}, respectively, in the major groove, and with the phosphates of C13, T25, C26 and T27 by Arg^{54}, Thr^{53}, Arg^{56} and Ser^{59}, respectively, in the minor groove. The interactions of Arg^{54} and Arg^{56} above and

below the polypeptide chain span the full length of the target site and are probably responsible for the bending of the DNA in the direction of the minor groove. Likewise, all the surgar contacts involve the second strand. In the major groove they are hydrophobic in nature, and involve contacts between the sugars of T22, T23 and A24 with Tyr^{34}, Leu^{33} and Ala^{30}, and Ile^{51} and Thr^{16}, respectively. In the minor groove, hydrophobic sugar DNA–protein interactions are made by C13 with the aliphatic portion of the side chain of Arg^{54}, T23 and T24 with Gln^{52}, T25 and C26 with the aliphatic portion of the side chain of Arg^{56}, and C26 with Ser^{59}. In addition, there is a hydrogen bond between the side chain amide of Gln^{52} and the sugar O3′ atom of T23.

The mode of specific DNA binding protein that is revealed in this structure is distinct from that observed for the other three classes of zinc containing DNA binding domains whose structures have previously been solved (Pavletich and Pabo, 1991, 1993; Luisi *et al.*, 1991; Mamorstein *et al.*, 1992; Schwabe *et al.*, 1993; Fairall *et al.*, 1993). Features specific to the complex with the DNA binding domain of cGATA-1 include the relatively small size of the DNA target site (8 base pairs of which only a contiguous stretch of 6 is involved in specific contacts), the monomeric nature of the complex in which only a *single* zinc binding module is required for specific binding, the predominance of hydrophobic interactions involved in specific base contacts in the major groove, the presence of a basic C-terminal tail which interacts with the DNA in the minor groove and constitutes a key component of specificity, and finally the pincer-like nature of the complex in which the core and tail subdomains are opposed and surround the DNA just like a hand gripping a rope. The structure of the cGATA-1 DNA binding domain reveals a modular design. The fold of residues 3–39 is similar to that of the N-terminal zinc binding module of the DNA binding domain of the glucocorticoid receptor, although, with the exception of the four Cys residues that coordinate zinc, there is no significant sequence identity between these regions of the two proteins. Residues 40–66 are part of a separate structural motif. In this regard it is interesting to note that, in addition to both zinc binding modules being encoded on separate exons in the cGATA-1 gene (exons 4 and 5), the next intron/exon boundary lies between amino acids 39 and 40 (current numbering scheme) of the DNA binding domain, thereby separating the C-terminal zinc binding domain from the basic tail (Hannon *et al.*, 1991).

3. *Hydration of the Specific Complex of the Transcription Factor GATA-1 with DNA*

A number of recent high resolution crystal structures of protein–DNA complexes have suggested that bound water may play an important role in the recognition process in the form of indirect read-out in which the bound water molecules serve to bridge hydrogen bonds between functional groups on the protein and the DNA bases (Otwinowski *et al.*, 1989; Lawson and Carey, 1993; Clark *et al.*, 1993). In addition a bound water molecule has been detected at the interface of the complex between the *Antp*(C39S) homeodomain and a 14-base pair duplex in solution using nuclear magnetic resonance spectroscopy (Qian *et al.*, 1993). In contrast to other protein–DNA complexes in which the majority of specific interactions involve hydrogen bonds between the protein and the DNA bases, the structure of the specific complex of GATA-1 with DNA, as indicated in the previous section, suggests that hydrophobic effects constitute a large proportion of the specific binding energy (Omichinski *et al.*, 1993). This would predict that water of hydration would be excluded from the interface between GATA-1 and the DNA bases in the major groove, but would be present at the interface between GATA-1 and the sugar–phosphate backbone, as well as at the solvent exposed surface of the protein. To test this hypothesis we carried out selective two-dimensional H_2O-NOE and H_2O-ROE ^1H–^{15}N and ^1H–^{13}C heteronuclear single quantum correlation experiments (Grzesiek and Bax, 1993a,b) to detect through-space (< 4 Å) contacts between bound water and protons of the protein, thereby identifying the location of bound water molecules in the specific complex of chicken GATA-1 with DNA (Clore *et al.*, 1994).

A number of water molecules could be detected between the protein and the phosphate backbone, as well as at the solvent exposed surface of the protein. However, no water molecules could be observed at the interface of the protein with the bases of the DNA. With

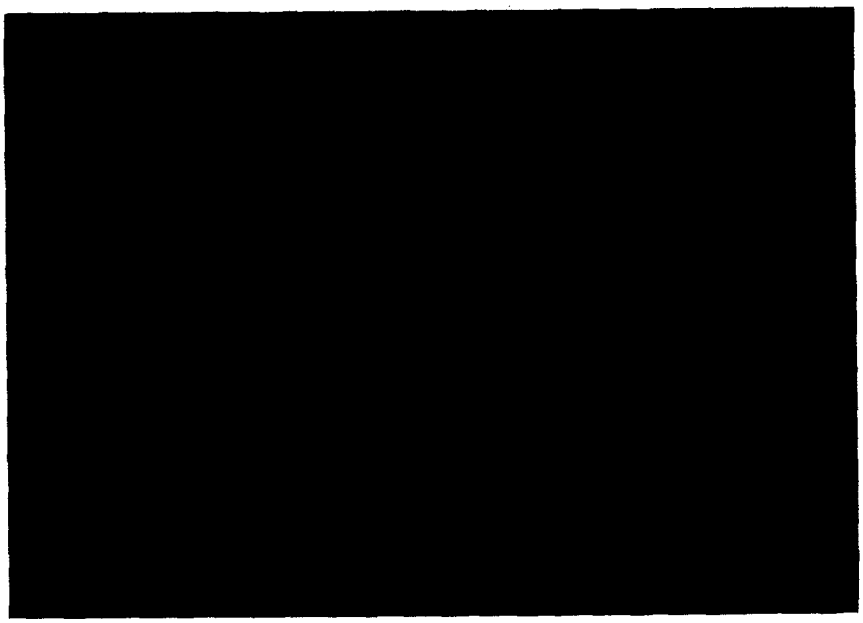


FIG. 14. Stereoview of the solution structure of the specific complex of GATA-1 and DNA, illustrating the location of groups in close proximity to bound water. The DNA is displayed in blue, the protein backbone in red, and the side chains in close proximity to bound water detected in the H_2O–ROE/NOE-HSQC experiments in green. The side chain of Leu^{33} which interacts directly with the bases of the DNA in the major groove and is not close to bound water is shown in orange. In the case of Leu^{17}, only the surface exposed $C^{\delta 2}H_3$ methyl group gives rise to a ROE with bound water; the $C^{\delta 1}H_3$ methyl group which interacts directly with the DNA bases is not close to bound water. The model is taken from the solution NMR structure of Omichinski *et al.* (1993) (PDB accession number 1GAT) and was generated with the program VISP (de Castro and Edelstein, 1992).

only one exception, the bound water molecules have a residency time greater than 200–300 psec. On the basis of the structure of the complex of GATA-1 with DNA (Omichinski *et al.*, 1993), we were able to unambiguously identify 17 methyl protons and 3 backbone NH protons in close proximity to bound water. The location of the protein residues near bound water are indicated in the structure of the complex of GATA-1 with DNA shown in Fig. 14.

An approximate estimate of the residency times of these bound water molecules can be obtained from the sign of the NOE (Ernst *et al.*, 1987; Otting *et al.*, 1991). If a water molecule is bound to a macromolecule with a correlation time in the spin-diffusion limit(i.e. greater than about 2 nsec), the NOE between water and a protein proton will be zero for a residency time of ~ 300 psec, while for residency times smaller and longer than 300 psec the NOE will be positive and negative, respectively. In contrast, the ROE remains positive for all residency and correlation times. The NH protons of Ala^{30}, Tyr^{34} and Tyr^{35} and the methyl groups of Ala^{3}, Ala^{30}, Met^{46} and Ile^{51} (γ_m and δ_m) exhibited negative NOEs to water, indicating a residency time greater than about 500 psec. The methyl protons of Met^{23} exhibited a positive NOE to water, indicative of a residency time of 100–300 psec. For the remaining surface methyl protons, the NOEs (at a mixing time of 60 msec) to water were too weak to observe and only ROEs were seen, indicating residency times in the range 200–500 psec.

All methyl groups that are exposed to solvent are associated with bound water molecules. These include Ala^{3}, Val^{6} (γ_1, γ_2), Leu^{17} (δ_2 only), Met^{23}, Val^{40} (γ_1, γ_2), Leu^{44} (δ_1, δ_2) and Val^{58} (γ_1, γ_2). Bound water molecules in contact with exposed hydrophobic surfaces are rarely seen in protein crystal structures as the water molecules are only well ordered (high occupancy and low thermal factor) if the protein donates anchor points in suitable hydrogen-bonding positions (Jeffrey and Saenger, 1991). In the absence of hydrogen bonding, low occupancies and high thermal B-factors render such water molecules unobservable (Jeffrey and Saenger, 1991). Indeed, there is only one crystallographic example where water surrounding an exposed hydrophobic group has been seen in the absence of hydrogen bonding, namely the water pentagons surrounding Leu^{18} in the 0.88 Å resolution crystal structure of crambin (Teeter, 1984). In contrast to the crystallographic case, bound water can be readily detected by NMR over a very wide range of residency times (with a lower limit of 50–100 psec) and is not eliminated by rapid independent rotation of a water pentagon around a rotating methyl group. Moreover, the NMR experiment is not dependent on uniform ordering in so far that the bound water could be in somewhat different positions in different molecules. In the X-ray diffraction experiment, on the other hand, bound water could only be detected if it occupied the same position in different protein molecules; that is to say there has to be uniform ordering of the bound water throughout the protein molecules of the crystal. Consequently, rapid rotation of a water pentagon about a methyl group, in the absence of discrete hopping, would result in smearing of the electron density beyond the limit of detectability.

The ubiquitous presence of bound water molecules close to exposed hydrophobic groups observed by NMR is completely consistent with the thermodynamic results of Privalov and Makhatadze (Makhatdze and Privalov, 1993; Privalov and Makhatadze, 1993), who showed that the hydration of methyl groups is associated with a large negative entropy ($-40.3\ J \cdot mol^{-1} \cdot K^{-1}$), a negative enthalpy ($-8.28\ kJ \cdot mol^{-1}$) and only a small positive Gibbs free energy ($3.72\ kJ \cdot mol^{-1}$). Indeed, the entropy of hydration of methyl groups is comparable to that of polar groups (-41 to $-50\ J \cdot mol^{-1} \cdot K^{-1}$), indicating that their ordering power for water is comparable.

All methyl groups that are in close proximity to the phosphodiester backbone but do not interact with the bases, are also associated with bound water molecules. In these cases, it is clear that the water is stabilized by hydrogen bonding interactions with either the phosphate group or the O5′ and O3′ oxygen atoms. Interestingly, similarly located water molecules have also been observed in crystal structure of DNA–protein complexes (Aggarwal *et al.*, 1988; Hegde *et al.*, 1992). The results on the specific complex of GATA-1 with DNA indicate the presence of a cluster of water molecules in the vicinity of the sugar–phosphate backbone of T22, T23, A24 and T25. It is important to note, however, that these interactions do not confer specificity. In the minor groove there is a bound water molecule(s) between the $C^{\varepsilon}H_3$

of Met[46], the sugar of T22 and the phosphate of T23. In the major groove 4 clusters of bound water molecules are observed: namely between the $C^{\delta 1}H_3$ methyl group of Leu[37] and the sugar and phosphate of T22; between the $C^{\delta 1}H_3$ of Leu[37], the backbone NH protons of Tyr[34] and Tyr[35], and the phosphate of T23; between the methyl group of Ala[30], the NH proton of Ala[30] and the phosphate of A24; and finally between the $C^{\gamma}H_3$ and $C^{\delta}H_3$ methyl groups of Ile[51], the sugar of A24 and the phosphate of T25. In all likelihood, the water molecule(s) associated with the NH protons Tyr[34] and Tyr[35], and with the NH proton of Ala[30], participate in bridging hydrogen bonds between the relevant amide group of the protein and the sugar–phosphate backbone of the DNA.

In contrast to the above methyl groups, no NOE/ROEs to water were observed for either of the two Leu33 methyl groups or the $C^{\delta 1}H_3$ methyl group of Leu[17], all of which are involved in hydrophobic interactions with the bases (Omichinski *et al.*, 1993). Thus, this indicates that water is excluded from the interface between the protein and the DNA bases in the major groove. In addition, complementary conventional 2D ^1H–^1H NOE experiments with ^{12}C filtering failed to detect any bound water close to the thymine methyl groups of the DNA, providing further evidence for the exclusion of water at the interface of the protein and the DNA bases. The absence of water at the interface between the protein and the DNA bases in the major groove lends further credence to the importance of hydrophobic interactions in stabilizing the specific interaction between chicken GATA-1 and DNA and indicates that hydrophobic interactions can play an important role in protein–DNA recognition and specificity.

VI. CONCLUDING REMARKS

From the examples presented in this review it should be clear that the recent development of a whole range of highly sensitive multi-dimensional heteronuclear-edited and -filtered NMR experiments has revolutionized the field of protein structure determination by NMR. Proteins and protein complexes in the 15–25 kDa range are now amenable to detailed structure analysis in solution. Moreover, the potential of the current methods can probably be extended to systems even up to 40 kDa providing that they are very well behaved from an NMR perspective. Nevertheless, despite these advances, it should always be borne in mind that there are a number of key requirements that have to be satisfied to permit a successful structure determination of larger proteins and protein complexes by NMR. The protein in hand must be soluble and should not aggregate up to concentrations of about 1 mM, it must be stable at room temperature or slightly higher for many weeks, it should not exhibit significant conformational heterogeneity that could result in extensive line broadening, and finally it must be amenable to uniform ^{15}N and ^{13}C labeling. At the present time there are only a few examples in the literature of proteins in the 15–25 kDa range that have been solved by multi-dimensional heteronuclear NMR spectroscopy. In addition to the three examples presented here, only the structures of interleukin-4 (Powers *et al.*, 1992, 1993; Smith *et al.*, 1992), glucose permease IIA [Fairbrother *et al.*, 1991) and the complex of cyclophilin with cyclosporin (Theriault *et al.*, 1993) have been published. It is hoped that over the next few years, the widespread use of these multi-dimensional heteronuclear experiments coupled with semi-automated assignment procedures will result in many more NMR structures of such larger proteins and protein–ligand complexes.

ACKNOWLEDGEMENTS

We thank our many colleagues, past and present, who have contributed to the work carried out in our laboratory. Above all, we thank Ad Bax with whom we have shared numerous stimulating discussions, fruitful experiments, and a continuous and most enjoyable collaboration in the best of scientific spirits. This work was supported in part by the AIDS Targeted Anti-viral Program of the Office of the Director of the National Institutes of Health.

REFERENCES

AGGARWAL, A. K., RODGERS, D. W., DROTTAR, M., PTASHNE, M. and HARRISON, S. C. (1988) Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. *Science* **242**, 899–907.

BABU, Y. S., SACK, J. S., GREENHOUGH, T. J., BUGG, C. E., MEANS, A. R. and COOK, W. J. (1985) Three dimensional structure of calmodulin. *Nature* **315,** 37–40.

BARBATO, G., IKURA, M., KAY, L., PASTOR, R. W. and BAX, A. (1992) Backbone dynamics of calmodulin studied by ^{15}N relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible. *Biochemistry* **31,** 5269–5278.

BAX, A. and GRZESIEK, S. (1993) Methodological advances in protein NMR. *Acc. Chem. Res.* **26,** 131–138.

BAX, A. and POCHAPSKY, S. S. (1992) Optimized recording of heteronuclear multi-dimensional NMR spectra using pulsed field gradients. *J. Magn. Reson.* **99,** 638–643.

BENNETT, M. K. and KENNEDY, M. B. (1987) Deduced primary structure of the β subunit of brain type II Ca^{2+}/calmodulin dependent protein kinase determined by molecular cloning. *Proc. natn. Acad. Sci. U.S.A.* **84,** 1794–1798.

BLUMENTHAL, D. K. and KREBS, E. G. (1987) Preparation and properties of the calmodulin binding domain of skeletal muscle myosin light chain kinase. *Meth. Enzyme.* **139,** 115–126.

BRAUN, W. (1987) Distance geometry and related methods for protein structure determination from NMR data. *Q. Rev. Biophys.* **19,** 115–157.

BRÜNGER, A. T. (1992) The free R value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355,** 472–474.

CHARBONNEAU, H., KUMAR, S., NOVACK, J. P., BLUMENTHAL, D. K., GRIFFIN, P. R., SHABANOWITZ, J., HUNT, D. F., BEAVO, J. A. and WALSH, K. A. (1991) Evidence for domain organization within the 61-kDa calmodulin-dependent cyclic nucleotide phosphodiesterase from bovine brain. *Biochemistry* **30,** 7931–7940.

CLARK, K. L., HALAY, E. D., LAI, E. and BURLEY, S. K. (1993) Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. (1993) *Nature* **364,** 412–420.

CLORE, G. M. and GRONENBORN, A. M. (1987) Determination of three-dimensional structures of proteins in solution by nuclear magnetic resonance spectroscopy. *Prot. Eng.* **1,** 275–288.

CLORE, G. M. and GRONENBORN, A. M. (1989) Determination of three-dimensional structures of proteins and nucleic acids in solution by nuclear magnetic resonance spectroscopy. *CRC Crit. Rev. Biochem. molec. Biol.* **24,** 479–564.

CLORE, G. M. and GRONENBORN, A. M. (1991a) Structures of larger proteins in solution: three- and four-dimensional heteronuclear NMR spectroscopy. *Science* **252,** 1309–1399.

CLORE, G. M. and GRONENBORN, A. M. (1991b) Comparison of the solution nuclear magnetic resonance and X-ray crystal structure of human recombinant interleukin-1β. *J. molec. Biol.* **221,** 47–53.

CLORE, G. M. and GRONENBORN, A. M. (1991c) Applications of three- and four-dimensional heteronuclear NMR spectroscopy to protein structure determination. *Prog. Nucl. Magn. Reson. Spectrosc.* **23,** 43–92.

CLORE, G. M. and GRONENBORN, A. M. (1991d) Two, three and four dimensional NMR methods for obtaining larger and more precise three-dimensional structures of proteins in solution. *A. Rev. Biophys. biophys. Chem.* **20,** 29–63.

CLORE, G. M., BAX, A., DRISCOLL, P. C., WINGFIELD, P. T. and GRONENBORN, A. M. (1990a) Assignment of the side chain ^{1}H and ^{13}C resonances of interleukin-1β using double and triple resonance heteronuclear three-dimensional NMR spectroscopy. *Biochemistry* **29,** 8172–8184.

CLORE, G. M., DRISCOLL, P. C., WINGFIELD, P. T. and GRONENBORN, A. M. (1990b) Low resolution structure of interleukin-1β in solution derived from ^{1}H–^{15}N heteronuclear three-dimensional NMR spectroscopy. *J. molec. Biol.* **214,** 881–817.

CLORE, G. M., BAX, A., WINGFIELD, P. T. and GRONENBORN, A. M. (1990c) Identification and localization of bound internal water in the solution structure of interleukin-1β by heteronuclear three-dimensional ^{1}H rotating frame Overhauser ^{15}N–^{1}H multiple quantum coherence NMR spectroscopy. *Biochemistry* **29,** 5671–5676.

CLORE, G. M., WINGFIELD, P. T. and GRONENBORN, A. M. (1991a) High resolution three-dimensional structure of interleukin-1β in solution by three and four dimensional nuclear magnetic resonance spectroscopy. *Biochemistry* **30,** 2315–2323.

CLORE, G. M., KAY, L. E., BAX, A. and GRONENBORN, A. M. (1991b) Four dimensional $^{13}C/^{13}C$-edited nuclear Overhauser enhancement spectroscopy of a protein in sulution: application to interleukin-1β. *Biochemistry* **30,** 12–18.

CLORE, G. M., ROBIEN, M. A. and GRONENBORN, A. M. (1993) Exploring the limits of precision and accuracy of protein structures determined by nuclear magnetic resonance spectroscopy. *J. molec. Biol.* **231,** 82–102.

CLORE, G. M., BAX, A., OMICHINSKI, J. G. and GRONENBORN, A. M. (1994) Localization of bound water in the solution structure of the erythroid transcription factor GATA-1 with DNA. *Structure* **2,** 89–94.

COHEN, P. and KLEE, C. B. (1988) *Molecular Aspects of Cellular Recognition*. Vol. 5. Elsevier, New York.

COX, J. A., COMTE, M., FITTON, J. E. and DEGRADO, W. F. (1985) The interaction of calmodulin with amphiphilic peptides. *J. biol. Chem.* **260,** 2527–2534.

DASGUPTA, M., HONEYCUTT, T. and BLUMENTHAL, D. K. (1989) The γ-subunit of skeletal muscle phosphorylase kinase contains two noncontiguous domains that act in concert to bind calmodulin. *J. biol. Chem.* **264,** 17156–17163.

DE CASTRO, E. and EDELSTEIN, S. (1992) *VISP* 1.0 *User's Guide*, University of Geneva.

DRISCOLL, P. C., CLORE, G. M., MARION, D., WINGFIELD, P. T. and GRONENBORN, A. M. (1990a) Complete resonance assignment for the polypeptide backbone of interleukin-1β using three-dimensional heteronuclear NMR spectroscopy. *Biochemistry* **29,** 3542–3556.

DRISCOLL, P. C., GRONENBORN, A. M., WINGFIELD, P. T. and CLORE, G. M. (1990b) Determination of the secondary structure and molecular topology of interleukin-1β using two- and three-dimensional heteronuclear ^{15}N–^{1}H NMR spectroscopy. *Biochemistry* **29,** 4468–4482.

DYSON, H. J., GIPPERT, G. P., CASE, D. A., HOLMGREN, A. and WRIGHT, P. E. (1990) Three-dimensional solution structure of the reduced form of *Escherichia coli* thioredoxin determined by nuclear magnetic resonance spectroscopy. *Biochemistry* **29,** 4129–4136.

EISENBERG, D. and MCLAGHLAN, A. D. (1986) Solvation energy in protein folding and binding. *Nature* **319,** 199–203.

ERNST, R., BODENHAUSEN, G. and WOKAUN, A. (1987) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon Press, Oxford.

EVANS, T. and FELSENFELD, G. (1989) The erythroid-specific transcription factor eryf1: a new finger protein. *Cell* **58**, 877–885.

FAIRALL, L., SCHWABE, J. W. R., CHAPMAN, L., FINCH, J. T. and RHODES, D. (1993) The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. *Nature* **366**, 483–487.

FAIRBROTHER, W. J., GIPPERT, G. P., REIZER, J., SAIER, M. H. and WRIGHT, P. E. (1992) Low resolution structure of the *Bacillus subtilis* glucose permease IIA domain derived from heteronuclear three-dimensional NMR spectroscopy. *FEBS Lett.* **296**, 148–152.

FESIK, S. W. and ZUIDERWEG, E. R. P. (1988) Heteronuclear three-dimensional NMR spectroscopy: a strategy for the simplification of homonuclear two-dimensional NMR spectra. *J. Magn. Reson.* **78**, 588–593.

FESIK, S. W. and ZUIDERWEG, E. R. P. (1990) Heteronuclear three-dimensional NMR spectroscopy of isotopically labelled biological macromolecules. *Q. Rev. Biophys.* **23**, 97–131.

FINZEL, B. C., CLANCY, L. L., HOLLAND, D. R., MUCHMORE, S. W., WATENPAUGH, K. D. and EINSPAHR, H. M. (1989) Crystal structure of recombinant human interleukin-1β at 2.0 Å resolution. *J. molec. Biol.* **209**, 779–791.

FORMAN-KAY, J. D., CLORE, G. M., WINGFIELD, P. T. and GRONENBORN, A. M. (1991) The high resolution three-dimensional structure of reduced recombinant human thioredoxin in solution. *Biochemistry* **30**, 2685–2698.

GRZESIEK, S. and BAX, A. (1993) Measurement of amide proton exchange rates and NOEs with water in $^{13}C/^{15}N$-enriched calcineurin B. *J. biol. molec. NMR* **3**, 627–638.

GRZESIEK, S. and BAX, A. (1994) The importance of not saturating H_2O in protein NMR. Application to sensitivity enhancement and NOE measurements. *J. Am. Chem. Soc.* in press.

GUERINI, D. and KLEE, C. B. (1991) Structural diversity of calcineurin Ca^{2+}–calmodulin stimulated phosphatases. *Adv. Prot. Phosphatases* **6**, 391–410.

GÜNTERT, P., BRAUN, W., BILLETER, M. and WÜTHRICH, K. (1989) Automated stereospecific ^1H assignments and their impact on the precision of protein structure determinations in solution. *J. Am. Chem. Soc.* **111**, 3997–4004.

HANNON, R., EVANS, T., FELSENFELD, G. and GOULD, H. (1991) Structure and promoter activity of the gene for the erythroid transcription factor GATA-1. *Proc. natn. Acad. Sci. U.S.A.* **88**, 3004–3008.

HARRISON, S. C. (1991) A structural taxonomy of DNA-binding domains. *Nature* **353**, 715–719.

HAVEL, T. F. (1991) An evaluation of computational strategies for use in the determination of protein structure from distance constraints obtained by nuclear magnetic resonance. *Prog. Biophys. molec. Biol.* **56**, 43–78.

HAVEL, T. F. and WÜTHRICH, K. (1985) An evaluation of the combined use of nuclear magnetic resonance and distance geometry for the determination of protein conformation in solution. *J. molec. Biol.* **182**, 281–294.

HAVEL, T. F., KUNTZ, I. D. and CRIPPEN, G. M. (1983) Theory and practice of distance geometry. *Bull. Math. Biol.* **45**, 665–720.

HEGDE, R. S., GROSSMAN, S. R., LAIMINS, L. A. and SIGLER, P. B. (1992) Crystal structure at 1.7 Å of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. *Nature* **359**, 505–512.

HEIDORN, D. B., SEEGER, P. A., ROBKOP, S. E., BLUMENTHAL, D. K., MEANS, A. R., CRESPI, H. and TREWHELLA, J. (1989) Changes in the structure of calmodulin induced by a peptide based on the calmodulin-binding domain of myosin light chain kinase. *Biochemistry* **28**, 6757–6764.

HURD, R. E. and JOHN, B. K. (1991) Gradient-enhanced proton-detected heteronuclear multiple-quantum coherence spectroscopy. *J. Magn. Reson.* **91**, 648–653.

HYBERTS, S. G., MÄRKI, W. and WAGNER, G. (1987) Stereospecific assignment of side chain protons and characterization of torsion angles in Eglin c. *Eur. J. Biochem.* **164**, 625–635.

IKURA, M. and BAX, A. (1992) Isotope filtered 2D NMR of a protein-peptide complex: study of a skeletal muscle myosin light chain kinase fragment bound to calmodulin. *J. Am. Chem. Soc.* **114**, 2433–2440.

IKURA, M., KAY, L. E. and BAX, A. (1990) A novel approach for sequential assignment of ^1H, ^{13}C and ^{15}N spectra of larger proteins: heteronuclear triple resonance NMR spectroscopy. Application to calmodulin. *Biochemistry* **29**, 4659–4667.

IKURA, M., KAY, L. E., KRINKS, M. and BAX, A. (1991) Triple-resonance multidimensional NMR study of calmodulin complexed with the binding domain of skeletal muscle myosin light-chain kinase: indication of a conformational change in the central helix. *Biochemistry* **30**, 5498–5504.

IKURA, M., CLORE, G. M., GRONENBORN, A. M., ZHU, G., KLEE, C. B. and BAX, A. (1992) Solution structure of a calmodulin–target peptide complex by multidimensional NMR. *Science* **256**, 632–638.

JEFFREY, G. A. and SAENGER, W. (1991) *Hydrogen Bonding in Biological Structures*. Springer, Berlin.

KATAOKA, M., HEAD, J. F., SEATON, B. A. and ENGELMAN, D. M., (1989) Mellitin binding causes a large calcium dependent conformational change in calmodulin. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6944–6948.

KATAOKA, M., HEAD, J. F., VORHERR, T., KREBS, J. and CARAFOLI, E. (1991) Small-angle X-ray scattering study of calmodulin bound to two peptides corresponding to parts of the calmodulin-binding domain of the plasma membrane Ca^{2+} pump. *Biochemistry* **30**, 6247–6251.

KATAOKA, M., HEAD, J. F., PERSECHINI, A., KRETSINGER, R. H. and ENGELMAN, D. M., (1991) Small-angle X-ray scattering studies of calmodulin mutants with deletions in the linker region of the central helix indicate that the linker region retains predominantly α-helical conformation. *Biochemistry* **30**, 1188–1192.

KAY, L. E., CLORE, G. M., BAX, A. and GRONENBORN, A. M. (1990) Four-dimensional heteronuclear triple resonance NMR spectroscopy of interleukin-1β in solution. *Science* **249**, 411–414.

KRAULIS, P. J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. appl. Crystallogr.* **24**, 946–950.

KWIATKOWSKI, A. P. and KING, M. M. (1989) Autophosphorylation of the type II calmodulin dependent protein kinase is essential for the formation of a proteolytic fragment with catalytic activity: implications for long-term synaptic potentiation. *Biochemistry* **28**, 5380–5385.

LAWSON, C. L. and CAREY, J. (1993) Tandem binding in cocrystals of a trp repressor/operator half-site complex. *Nature* **366**, 178–182.

LUKAS, T. J., BURGESS, W. H., PREDERGAST, F. G., LAU, W. and WATTERSON, D. M., (1986) Calmodulin binding domains: characterization of a phosphorylation and calmodulin binding site from myosin light chain kinase. *Biochemistry* **25,** 1458–1464.

LUISI, B. F., XU, W. X., OTWINOWSKI, Z., FREEDMAN, L. P., YAMAMOTO, K. R. and SIGLER, P. G. (1991) Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* **352,** 497–505.

MAKHATADZE, G. I. and PRIVALOV, P. L. (1993) Contribution of hydration to protein folding thermodynamics. I. The enthalpy of hydration. *J. molec. Biol.* **232,** 639–659.

MAMORTSTEIN, R., CAREY, M., PTASHNE, M. and HARRISON, S. C. (1992) DNA recognition by GAL4: structure of a protein–DNA complex. *Nature* **356,** 408–414.

MARION, D., DRISCOLL, P. C., KAY, L. E., WINGFIELD, P. T., BAX, A., GRONENBORN, A. M. and CLORE, G. M. (1989) Overcoming the overlap problem in the assignment of ^{1}H-NMR spectra of larger proteins using three-dimensional heteronuclear ^{1}H–^{15}N Hartmann–Hahn and nuclear Overhauser-multiple quantum coherence spectroscopy. Application to interleukin-1β. *Biochemistry* **29,** 6150–6156.

MARTIN, D. and ORKIN, S. (1986) Transcriptional activation and DNA binding by the erythroid factor GF-1/NF-E1/Eryf 1. *Genes Dev.* **4,** 1886–1989.

MONTELIONE, G. T. and WAGNER, G. (1989) Accurate measurement of homonuclear H^{N}–H^{α} coupling constants in polypeptides using heteronuclear 2D NMR experiments. *J. Am. Chem. Soc.* **111,** 5474–5475.

MONTELIONE, G. T. and WAGNER, G. (1990) Conformation independent sequential NMR connections in isotope-enriched polypeptides by ^{1}H–^{13}C–^{15}N triple resonance experiments. *J. Magn. Reson.* **87,** 183–188.

NILGES, M., CLORE, G. M. and GRONENBORN, A. M. (1988) Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance goemetry-dynamical simulated annealing calculations. *FEBS Lett.* **229,** 317–324.

NILGES, M., CLORE, G. M. and GRONENBORN, A. M. (1990) ^{1}H-NMR stereospecific assignments by conformational database searches. *Biopolymers* **29,** 813–822.

OMICHINSKI, J. G., TRAINOR, C., EVANS, T., GRONENBORN, A. M., CLORE, G. M. and FELSENFELD, G. (1993) A small single-'finger' peptide from the erythroid factor GATA-1 binds specifically to DNA as a zinc or iron complex. *Proc. natn. Acad. Sci. U.S.A.* **90,** 1676–1680.

OMICHINSKI, J. G., CLORE, G. M., SCHAAD, O., FELSENFELD, G., TRAINOR, C., APPELLA, E., STAHL, S. J. and GRONENBORN, A. M. (1993) NMR structure of a specific DNA complex of the Zn-containing DNA binding domain of GATA-1. *Science* **261,** 438–446.

O'NEIL, K. T. and DEGRADO, W. F. (1990) How calmodulin binds its targets: sequence independent recognition of amphiphilic α-helices. *Trends Biochem.* **15,** 59–64.

O'NEIL, K. T., ERICKSON-VIITANEN, S. and DEGRADO, W. F. (1989) Photolabeling of calmodulin with basic, amphiphilic α-helical peptides contaning p-benzoylphenylalanine. *J. biol. Chem.* **264,** 14571–14578.

ORKIN, S. H. (1992) GATA-binding transcription factors in hematopoietic cells. *Blood* **80,** 575–581.

OSCHKINAT, H., GRIESINGER, C., KRAULIS, P. J., SØRENSEN, O. W., ERNST, R. R., GRONENBORN, A. M. and CLORE, G. M. (1988) *Nature* **332,** 374–376.

OTTING, G., LIEPINSH, E. and WÜTHRICH, K. (1991) Protein hydration in aqueous solution. *Science* **254,** 974–980.

OTWINOWSKI, Z., SHEVITZ, R. W., ZHANG, R.-G., LAWSON, C. I., JOACHIMIAK, A., MARMORSTEIN, E. Q., LUISI, B. F. and SIGLER, P. B. (1988) Crystal structure of trp repressor/operator complex at atomic resolution. *Nature* **335,** 321–329.

PAVLETICH, N. P. and PABO, C. O. (1991) Zinc finger–DNA recognition: crystal structure of a Zif268–DNA complex at 2.1 Å. *Science* **252,** 809–817.

PAVLETICH, N. P. and PABO, C. O. (1993) Crystal structure of a five-finger GLI–DNA complex: new perspectives on zinc fingers. *Science* **261,** 1701–1707.

PERSECHINI, A. and KRETSINGER, R. H. (1988) The central helix of calmodulin functions as a flexible tether. *J. biol. Chem.* **263,** 12175–12178.

PERSECHINI, A., BLUMENTHAL, D. K., JARRETT, H. W., KLEE, C. B., HARDY, D. O. and KRETSINGER, R. H. (1989) The effects of deletions in the central helix of calmodulin on enzyme activation and peptide binding. *J. biol. Chem.* **264,** 8052–8058.

POWERS, R., GARRETT, D. S., MARCH, C. J., FRIEDEN, E. A., GRONENBORN, A. M. and CLORE, G. M. (1992) Three-dimensional structure of interleukin-4 by multi-dimensional heteronuclear magnetic resonance spectroscopy. *Science* **256,** 1673–1677.

POWERS, R., GARRETT, D. S., MARCH, C. J., FRIEDEN, E. A., GRONENBORN, A. M. and CLORE, G. M. (1993) The high resolution three-dimensional solution structure of human interleukin-4 determined by multi-dimensional heteronuclear magnetic resonance spectroscopy. *Biochemistry* **32,** 6744–6762.

PRIESTLE, J. P., SCHÄR, H. P. and GRÜTTER, M. G. (1989) Crystallographic refinement of interleukin-1β at 2.0 Å resolution. *Proc. natn. Acad. Sci. U.S.A.* **86,** 9667–9671.

PRIVALOV, P. I. and MAKHATADZE, G. I. (1993) Contribution of hydration to protein folding thermodynamics. II The entropy and Gibbs energy of hydration. *J. molec. Biol.* **232,** 660–679.

QIAN, Y. Q., OTTING, G. and WÜTHRICH, K. (1993) NMR detection of hydration water in the intermolecular interface of a protein–DNA complex. *J. Am. Chem. Soc.* **115,** 1189–1190.

SHAANAN, B., GRONENBORN, A. M., COGEN, G. H., GILLAND, G. L., VEERAPANDIAN, B., DAVIES, D. R. and CLORE, G. M. (1992) Combining experimental information from crystal and solution studies: joint X-ray and NMR refinement. *Science* **257,** 961–964.

SCHWABE, J. W. R., CHAPMAN, L., FINCH, J. T. and RHODES, D. (1993) The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements. *Cell* **75,** 567–568.

SMITH, L. J., REDFIELD, C., BOYD, J., LAWRENCE, G. M. P., EDWARDS, R. G., SMITH, R. A. G. and DOBSON, C. M. (1992) Human interleukin-4: the solution structure of a four-helix bundle protein. *J. molec. Biol.* **224,** 899–904.

SPERA, S., IKURA, M. and BAX, A. (1991) Measurement of the exchange rates of rapidly exchanging amide protons: application to the study of calmodulin and its complex with a myosin light chain kinase fragment. *J. biomolec. NMR* **1,** 155–165.

TEETER, M. M. (1984) Water structure of a hydrophobic protein at atomic resolution: pentagon rings of water molecules in crystals of crambin. *Proc. natn. Acad. Sci. U.S.A.* **81,** 6014–6018.

THERIAULT, Y., LOGAN, T. M., MEADOWS, R., YU, L., OLENJNICZSAK, E. T., HOLZMAN, T., SIMMER, R. L. and FESIK, T. M. (1993) Solution structure of the cyclosporin A/cyclophilin complex by NMR. *Nature* **361,** 88–91.

TREWHELLA, J., BLUMENTHAL, D. K., ROKOP, S. E. and SEEGER, P. A. (1990) Small-angle scattering studies show distinct conformations of calmodulin in its complex with two peptides based on the regulatory domain of the catalytic subunit of phosphorylase kinase. *Biochemistry* **29,** 9316–9324.

VEERAPANDIAN, B., GILLILAND, G. L., RAAG, R., SVENSSON, A. L., MASUI, Y., HIRAI, Y. and POULOS, T. L. (1991) Functional implications of interleukin-1β based on the three dimensional structure. *Prot. Struct. Funct. Genet.* **12,** 10–24.

VUISTER, G. W., BOELENS, R., KAPTEIN, R., HURD, R. E., JOHN, B. K. and VAN ZILJ, P. C. M. (1991) Gradient enhanced HMQC and HSQC spectroscopy: applications to ^{15}N-labeled Mnr repressor. *J. Am. Chem. Soc.* **113,** 9688–9690.

VUISTER, G. W., CLORE, G. M., GRONENBORN, A. M., POWERS, R., GARRETT, D. S., TSCHDUIN, R. and BAX, A. (1993) Increased resolution and improved quality in four-dimensional ^{13}C/^{13}C separated HMQC–NOE–HMQC spectra using pulse field gradients. *J. Magn. Reson. Series B* **101,** 210–213.

VUISTER, G. W., GRZESIEK, S., DELAGLIO, F., WANG, A. C., TSCHUDIN, R., ZHU, G. and BAX, A. (1994) Measurement of homo- and heteronuclear J couplings from quantitative J correlation. *Metho. Enzym.*, in press.

WAGNER, G., BRAUN, W., HAVEL, T. F., SCHAUMAN, T., GO, N. and WÜTHRICH, K. (1987) 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. molec. Biol.* **196,** 611–639.

WÜTHRICH, K. (1986) *NMR of Proteins and Nucleic Acids.* Wiley, New York.

ZUIDERWEG, E. R. P., BOELENS, R. and KAPTEIN, R. (1985) Stereospecific assignment of ^{1}H-NMR methyl lines and conformation of valyl residues in the lac repressor headpiece. *Biopolymers* **24,** 601–611.

ZUIDERWEG, E. R. P., PETROS, A. M., FESIK, S. W. and OLEJNICZAK, E. T. (1991) Four-dimensional [^{13}C, ^{1}H, ^{13}C, ^{1}H] HMQC–NOE–HMQC NMR spectroscopy: resolving tertiary NOE distance restraints in spectra of larger proteins. *J. Am. Chem. Soc.* **113,** 370–372.