

YOUNG INVESTIGATOR AWARD LECTURE

Structures of larger proteins, protein–ligand and protein–DNA complexes by multidimensional 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, Maryland 20892

(RECEIVED December 22, 1993; ACCEPTED January 7, 1994)

Abstract

The recent development of a whole panoply of multidimensional heteronuclear-edited and -filtered NMR experiments has revolutionized the field of protein structure determination by NMR, making it possible to extend the methodology from the 10-kDa limit of conventional 2-dimensional NMR to systems up to potentially 35–40 kDa. The basic strategy for solving 3-dimensional structures of larger proteins and protein–ligand complexes in solution using 3- and 4-dimensional NMR spectroscopy is summarized, and the power of these methods is illustrated using 3 examples: interleukin-1 β , the complex of calmodulin with a target peptide, and the specific complex of the transcription factor GATA-1 with its cognate DNA target site.

Keywords: calmodulin; GATA-1; heteronuclear NMR; interleukin-1 β ; multidimensional NMR; protein–DNA complexes; protein–peptide complexes; proteins; solution structure

The last few years have seen a quantum jump both in the size and accuracy of protein structures that can be determined by NMR (Clare & 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 (Clare & Gronenborn, 1991b). This is attributable to the development of 3- and 4-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 Clare & Gronenborn, 1991a, 1991c, 1991d; Bax & Grzesiek, 1993, for reviews). In this short review, we summarize some of these developments and illustrate their application to the structure determination of interleukin-1 β (Clare et al., 1991b), 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., 1993a).

General strategy for the determination of 3-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, which can be used to identify protons separated by less 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 Figure 1. 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 3-bond scalar couplings

Reprint requests to: G. Marius Clare, Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892; e-mail: clare@vger.niddk.nih.gov.

*This paper is based on the lecture presented by Dr. Clare at the Seventh Annual Symposium of The Protein Society in San Diego on July 26, 1993, as one of the co-recipients of the Young Investigator Award sponsored by the DuPont Merck Pharmaceutical Company. The other recipient was Dr. Ad Bax.

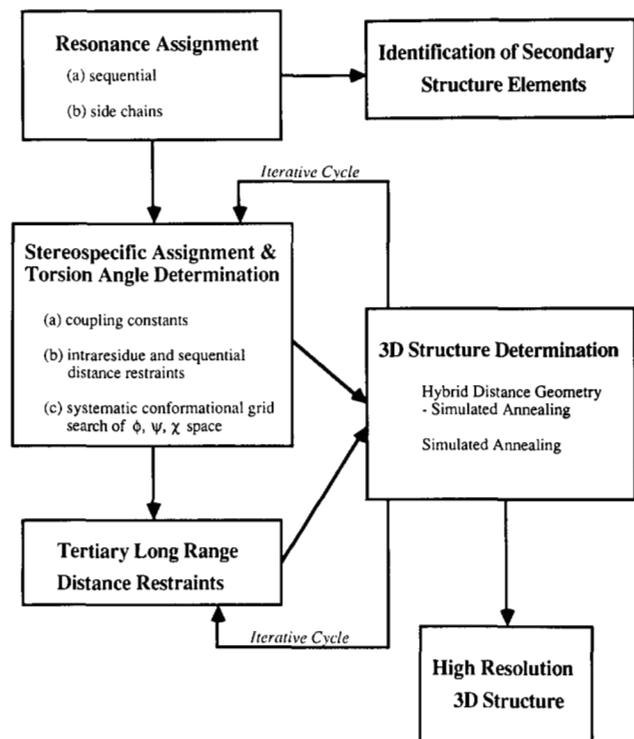


Fig. 1. Summary of general strategy employed to solve 3D structures of macromolecules by NMR.

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 3D 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 & 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 that 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 discriminate between these 2 possibilities. Thus, if protons A and C are significantly greater than 5 Å apart, and 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 & Wüthrich, 1985; Clore & Gronenborn, 1991a, 1991d; Havel, 1991; Clore et al., 1993). This progression in coordinate precision is illustrated in Figure 2, which shows 4 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.

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 $\text{C}^\alpha\text{H}(i)$ - $\text{NH}(i+1, 2, 3, 4)$, $\text{NH}(i)$ - $\text{NH}(i\pm 2)$, and $\text{C}^\alpha\text{H}(i)$ - $\text{C}^\beta\text{H}(i+3)$ (Wüthrich, 1986; Clore & 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 Figure 3 (Fesik & Zuiderweg, 1988, 1990; Marion et al., 1989; 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 & Wagner, 1989, 1990; Ikura et al., 1990; Clore & Gronenborn, 1991c; Bax & 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 pulse field gradients to eliminate undesired coherence transfer pathways (Bax & Pochapsky, 1992), as opposed to selecting desired coherence transfer pathways (Hurd & John, 1991; Vuister et al., 1991), it is now possible to employ only 2-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 h. In most cases, however, signal-to-noise requirements necessitate 1-3 days measuring time depending on the experiment.

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^αH , and C^βH protons (Hyberts et al., 1987; Wagner 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 $^3J_{\text{HN}\alpha}$ and $^3J_{\alpha\beta}$ coupling constants (which are related to ϕ and χ_1 , respectively), and the intraresidue and sequential ($i\pm 1$) interresidue NOEs involving the NH, C^αH , and C^βH protons (Güntert et al., 1989; Nilges et al., 1990). This information can be supplemented by the measurement of heteronuclear $^3J_{\text{NH}\beta}$ and $^3J_{\text{COH}\beta}$ 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 $^3J_{\text{C}\gamma\text{CO}}$, $^3J_{\text{NC}\gamma}$ couplings (Vuister et al., 1994), as well as on the basis of the pattern of intraresidue NOEs involving the NH, C^αH , and C^γH protons (Zuiderweg et al., 1985). Finally, stereospecific assignments of leucine methyl groups can be made on the basis of heteronuclear $^3J_{\text{C}\delta\text{C}\alpha}$ and $^3J_{\text{C}\delta\text{H}\beta}$ 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).

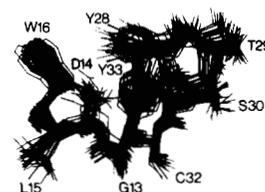
1st Generation

~ 7 restraints per residues
 rmsd: 1.5Å for backbone atoms
 2.0Å for all atoms
 example: puorhionin



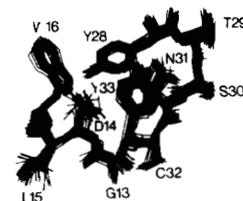
2nd Generation

~ 10 restraints per residue
 rmsd: 0.9Å for backbone atoms
 1.2Å for all atoms
 example: BDS-I



3rd Generation

~ 13 restraints per residue
 rmsd: 0.7Å for backbone atoms
 0.9Å for all atoms
 example: BDS-I



4th Generation

~ 16 restraints per residue
 rmsd: 0.4Å for backbone atoms
 0.9Å for all atoms,
 ≤ 0.5Å for ordered side chains
 example: Interleukin-8

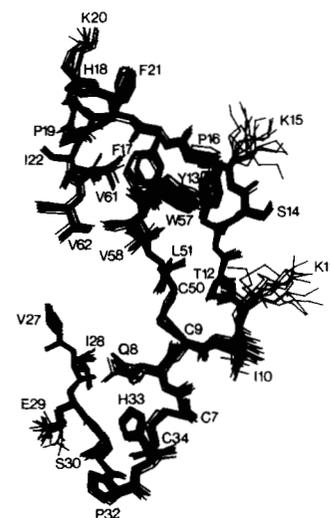
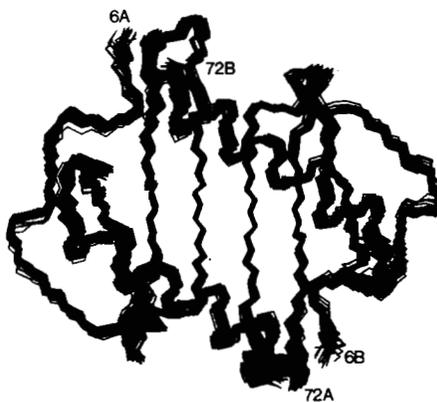


Fig. 2. Illustration of the progressive improvement in the precision and accuracy of NMR structure determinations with increasing number of experimental restraints. All the structures have been calculated using the hybrid distance geometry-simulated annealing method, and in each case the NOE-derived interproton distance restraints have been grouped into 3 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.

Assignment of through-space proton-proton interactions with a protein

Although the panoply of 3D heteronuclear experiments is sufficient for the purposes of spectral assignment, yet further in-

creases in resolution are required for the reliable identification of NOE through-space interactions. This can be achieved by extending the dimensionality still further to 4 dimensions (Kay et al., 1990). This is illustrated in Figure 4. Consider a simple 2D spectrum demonstrating 11 cross peaks from aliphatic res-

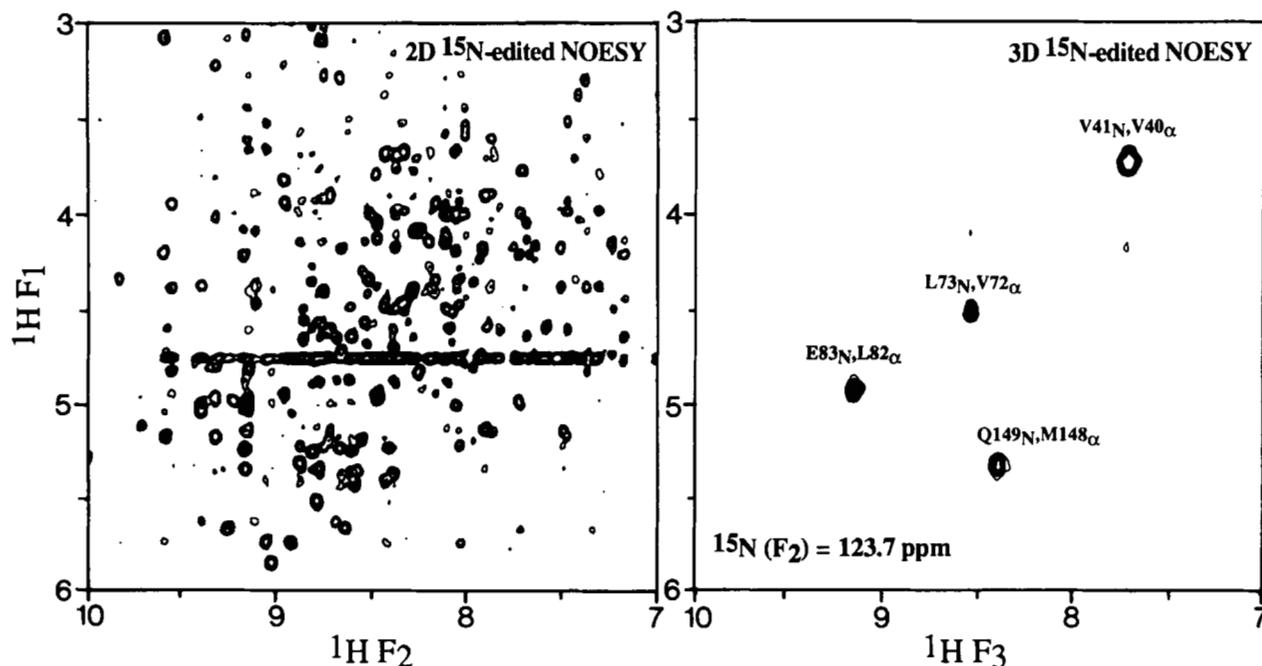


Fig. 3. Comparison of the NH-C α H/C β H region of a 2D ^{15}N -edited NOESY spectrum with that of a single plane taken from the 3D ^{15}N -edited NOESY spectrum, illustrating the increase in spectral resolution afforded by increasing the dimensionality from 2 to 3.

onances to a single NH resonance position. In the 2D spectrum it is impossible to ascertain whether this involves 1 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 3 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 ^1H - ^1H NOE interaction is specified by 4 chemical shift coordinates, the 2 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 Figure 5.

Because the number of NOE interactions present in each 2D plane of a 4D $^{13}\text{C}/^{15}\text{N}$ - or $^{13}\text{C}/^{13}\text{C}$ -edited NOESY spectrum is so small, the inherent resolution in a 4D spectrum is 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 ^1H , ^{15}N , and ^{13}C assignments are obtained, analysis of 4D $^{15}\text{N}/^{13}\text{C}$ - (Kay et al., 1990) and $^{13}\text{C}/^{13}\text{C}$ - (Clare et al., 1991a; Zuiderweg et al., 1991; Vuister et al., 1993) edited NOE spectra should permit the automated assignment of almost all NOE interactions.

Application of 3D and 4D NMR to protein structure determination of larger proteins: the structure of interleukin-1 β

Although 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 that it was originally designed to tackle, namely the determination of high resolution 3D 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-1 β (IL-1 β), a cytokine of 153 residues and molecular weight 17.4 kDa, which plays a key role in the immune and inflammatory responses (see Kinemage 1; Clare et al., 1991b). At the time IL-1 β 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-1 β still represents one of the most highly refined and precise structures for proteins of this size solved 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 ^1H assignment for more than about 30% of the residues of IL-1 β (Driscoll et al., 1990a). Thus, any further progress could only be made by resorting to higher dimension-

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 ^a
¹⁵ N-edited HOHAHA	C ^α H(<i>i</i>)- ¹⁵ N(<i>i</i>)-NH(<i>i</i>)	³ J _{HNα}
HNHA	C ^β H(<i>i</i>)- ¹⁵ N(<i>i</i>)-NH(<i>i</i>)	³ J _{HNα} and ³ J _{αβ}
H(CA)NH	C ^α H(<i>i</i>)- ¹⁵ N(<i>i</i>)-NH(<i>i</i>)	³ J _{HNα}
HNCA	C ^α H(<i>i</i>)- ¹⁵ N(<i>i</i>)-NH(<i>i</i>)	¹ J _{NCα}
	C ^α H(<i>i</i> - 1)- ¹⁵ N(<i>i</i>)-NH(<i>i</i>)	² J _{NCα}
	¹³ C ^α (<i>i</i>)- ¹⁵ N(<i>i</i>)-NH(<i>i</i>)	¹ J _{NCα}
	¹³ C ^α (<i>i</i> - 1)- ¹⁵ N(<i>i</i>)-NH(<i>i</i>)	² J _{NCα}
HN(CO)CA	¹³ C ^α (<i>i</i> - 1)- ¹⁵ N(<i>i</i>)-NH(<i>i</i>)	¹ J _{NCO} and ¹ J _{CαCO}
HNCO	¹³ CO(<i>i</i> - 1)- ¹⁵ N(<i>i</i>)-NH(<i>i</i>)	¹ J _{NCO}
HCACO	C ^α H(<i>i</i>)- ¹³ C ^α (<i>i</i>)- ¹³ CO(<i>i</i>)	¹ J _{CαCO}
HCA(CO)N	C ^α H(<i>i</i>)- ¹³ C ^α (<i>i</i>)- ¹⁵ N(<i>i</i> + 1)	¹ J _{CαCO} and ¹ J _{NCO}
CBCA(CO)NH	¹³ C ^β (<i>i</i> - 1)/ ¹³ C ^α (<i>i</i> - 1)- ¹⁵ N(<i>i</i>)-NH(<i>i</i>)	¹ J _{CαCO} , ¹ J _{NCO} , and ¹ J _{CC}
CBCANH	¹³ C ^β (<i>i</i>)/ ¹³ C ^α (<i>i</i>)- ¹⁵ N(<i>i</i>)-NH(<i>i</i>)	¹ J _{NCα} and ¹ J _{CC}
	¹³ C ^β (<i>i</i> - 1)/ ¹³ C ^α (<i>i</i> - 1)- ¹⁵ N(<i>i</i>)-NH(<i>i</i>)	² J _{NCα} and ¹ J _{CC}
HBHA(CO)NH	C ^β H(<i>i</i> - 1)/C ^α H(<i>i</i> - 1)- ¹⁵ N(<i>i</i>)-NH(<i>i</i>)	¹ J _{CαCO} , ¹ J _{NCO} , and ¹ J _{CC}
HBHANH	C ^β H(<i>i</i>)/C ^α H(<i>i</i>) - ¹⁵ N(<i>i</i>)-NH(<i>i</i>)	¹ J _{NCα} and ¹ J _{CC}
	C ^β H(<i>i</i> - 1)/C ^α H(<i>i</i> - 1)- ¹⁵ N(<i>i</i>)-NH(<i>i</i>)	² J _{NCα} and ¹ J _{CC}
C(CO)NH	¹³ C ^j (<i>i</i> - 1)- ¹⁵ N(<i>i</i>)-NH(<i>i</i>)	¹ J _{CαCO} , ¹ J _{NCO} , and ¹ J _{CC}
H(CCO)NH	H ^j (<i>i</i> - 1)- ¹⁵ N(<i>i</i>)-NH(<i>i</i>)	¹ J _{CαCO} , ¹ J _{NCO} , and ¹ J _{CC}
HCCH-COSY	H ^j - ¹³ C ^j - ¹³ C ^{j±1} -H ^{j±1}	¹ J _{CC}
HCCH-TOCSY	H ^j - ¹³ C ^j ... ¹³ C ^{j±n} -H ^{j±n}	¹ J _{CC}

^a In addition to the couplings indicated, all the experiments make use of the ¹J_{CH} (~140 Hz) and/or ¹J_{NH} (~95 Hz) couplings. The values of the couplings employed are as follows: ³J_{HNα} ~ 3-10 Hz, ¹J_{CC} ~ 35 Hz, ¹J_{CαCO} ~ 55 Hz, ¹J_{NCO} ~ 15 Hz, ¹J_{NCα} ~ 11 Hz, ²J_{NCα} ~ 7 Hz.

ality heteronuclear NMR. The initial step involved the complete assignment of the ¹H, ¹⁵N, and ¹³C resonances of the backbone and side chains using many of the double and triple resonance 3D experiments listed in Table 1 (Clore et al., 1990a; Driscoll et al., 1990a, 1990b). 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 3D systematic grid search of φ, ψ, χ₁ 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^αH, and C^βH protons (Clore et al., 1990c). 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 ¹⁵N-separated rotating frame Overhauser (ROE) spectrum, which permits one to distinguish specific protein-water NOE interactions from chemical exchange with bulk solvent (Clore et al., 1990b). 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 3 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., 1988). A key aspect of the overall strategy lies in the use of an iterative approach whereby the experimental data are 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 IL-1β comprised a total of 3,146 approximate and loose experimental restraints made up of 2,780 distance and 366 torsion angle restraints (Clore et al., 1991b). This represents an average of ~21 experimental restraints per residue. If one takes into account that interresidue NOEs affect 2 residues, whereas 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 Figure 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

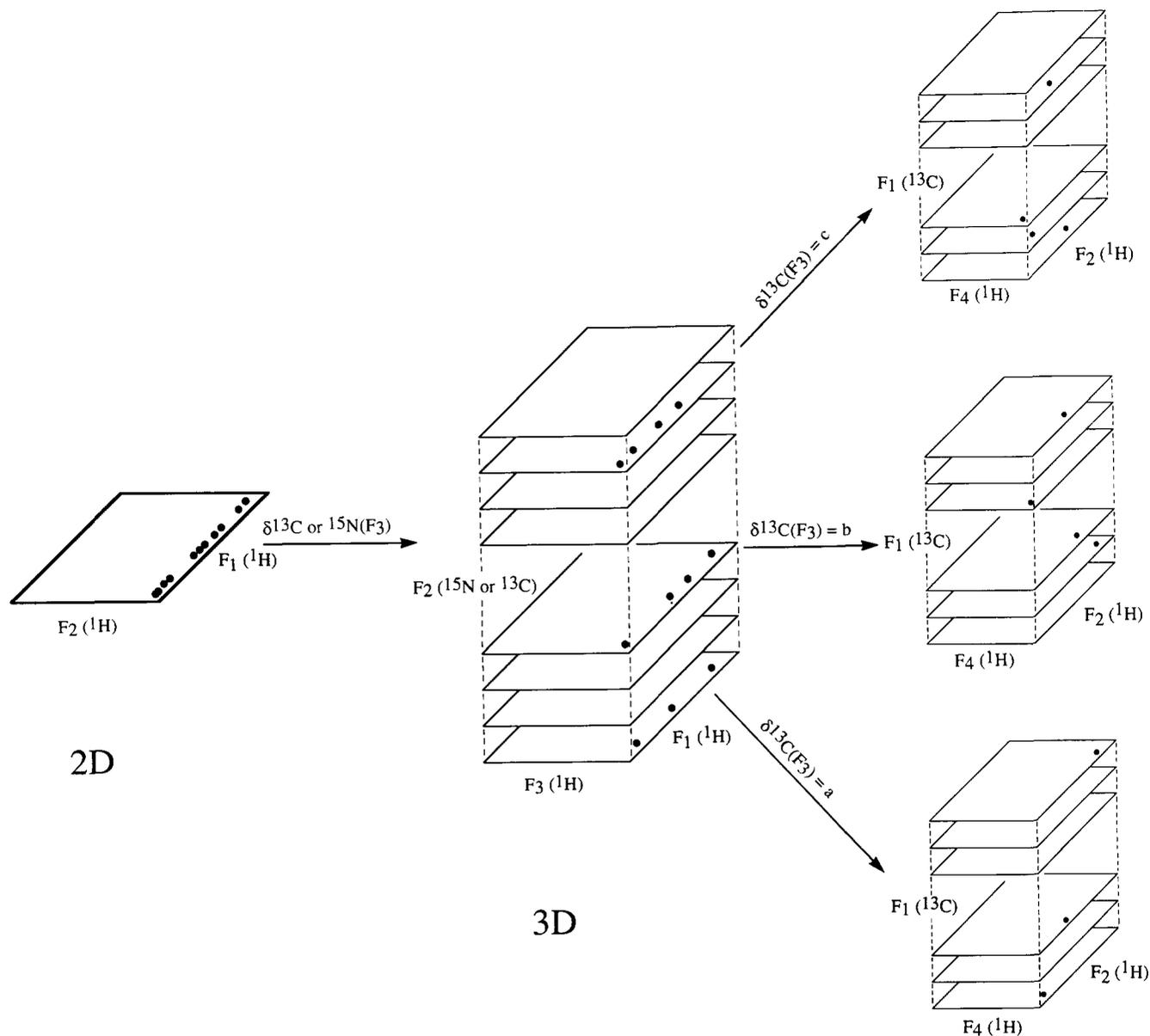


Fig. 4. Schematic illustration of the progression and relationship between 2D, 3D, and 4D heteronuclear-edited NMR spectroscopy.

positions is 0.4 Å for the backbone atoms, 0.8 Å for all atoms, and 0.5 Å for side chains with $\leq 40\%$ of their surface (relative to that in a tripeptide Gly-X-Gly) accessible to solvent (Clare et al., 1991b).

The structure of IL-1 β itself resembles a tetrahedron and displays 3-fold internal pseudosymmetry (Kinemage 1). There are 12 β -strands arranged in an exclusively antiparallel β -structure, and 6 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 6 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 Figure 6C. Water molecules occupy very similar positions in all 3 topological units, as well as at the interface of the 3 units, and are involved in

bridging backbone hydrogen bonds. Thus, in the case of the topological unit shown in Figure 6C, the water molecule labeled W5 accepts a hydrogen bond from the NH of Phe-112 in strand IX and donates 2 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 Figure 6D. Because of the high resolution of the IL-1 β 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 3 distinct sites involved in the binding of IL-1 β to its cell surface receptor (Clare et al., 1991b).

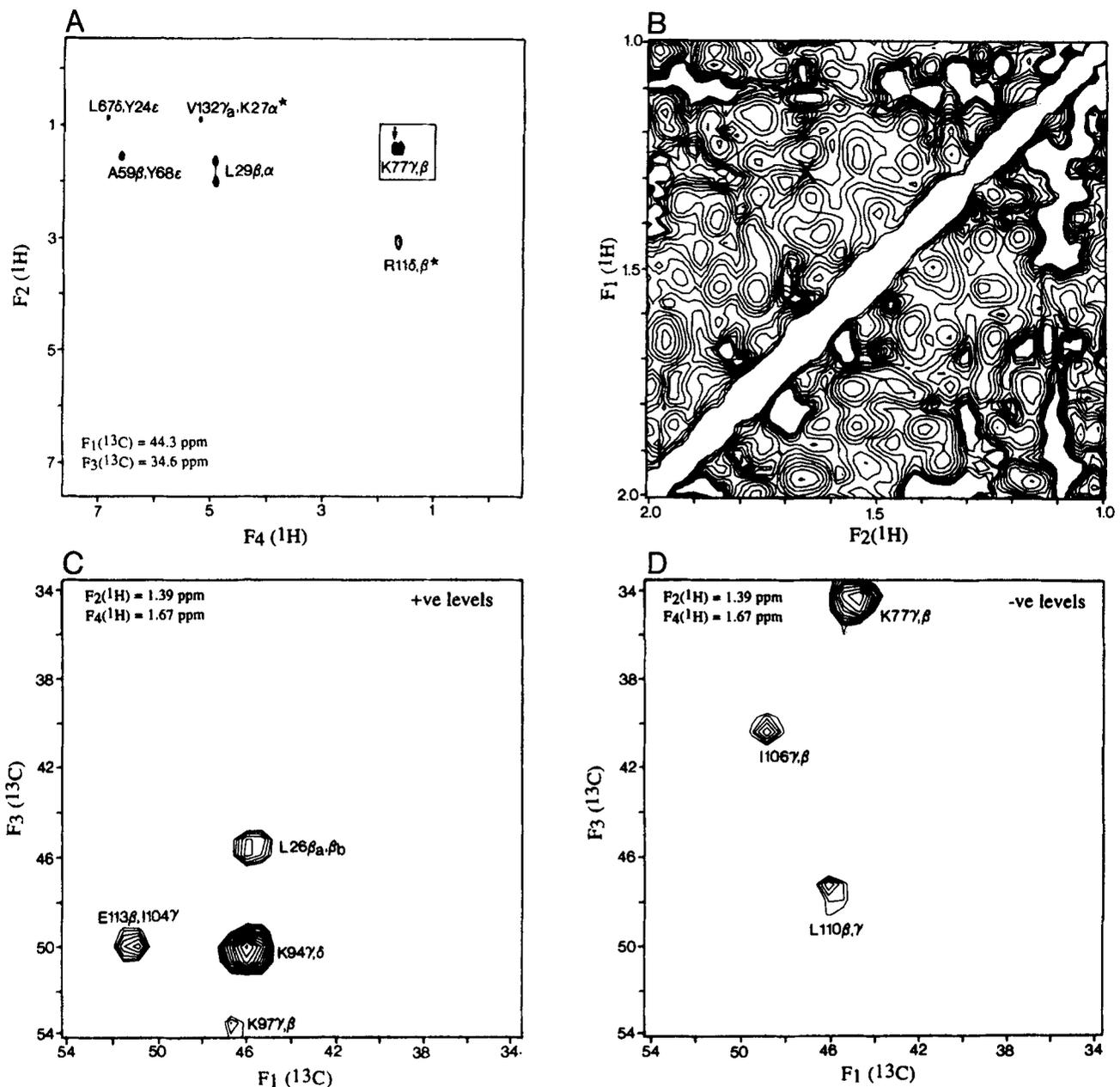


Fig. 5. Example of the increase in spectral resolution afforded by 4D $^{13}\text{C}/^{13}\text{C}$ -edited NOE spectroscopy, illustrated with interleukin-1 β . **A:** $^1\text{H}(F_2)$ - $^1\text{H}(F_4)$ plane of the 4D spectrum at $\delta^{13}\text{C}(F_1) = 44.3$ ppm and $\delta^{13}\text{C}(F_3) = 34.6$ ppm; the region between 1 and 2 ppm is boxed in and the arrow indicates the position of the Lys-77 C^γH - C^βH NOE cross peak. **B:** 2D ^1H - ^1H NOE spectrum between 1 and 2 ppm; the X marks the chemical shift position of the Lys-77 C^γH - C^βH NOE cross peak seen in A. **C, D:** Positive and negative contours in the $^{13}\text{C}(F_1)$ - $^{13}\text{C}(F_3)$ plane of the 4D spectrum at the ^1H chemical shift coordinates, $\delta^1\text{H}(F_2) = 1.39$ ppm and $\delta^1\text{H}(F_4) = 1.67$ ppm, corresponding to the Lys-77 C^γH - C^βH 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 $x \pm n\text{SW}$ 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.

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. IL-1 β offers an ideal system for comparing the results of NMR and X-ray crystallography as, in addition to the solution structure, there are 3 independently solved X-ray structures at 2 Å resolution of the same crystal form (Fin-

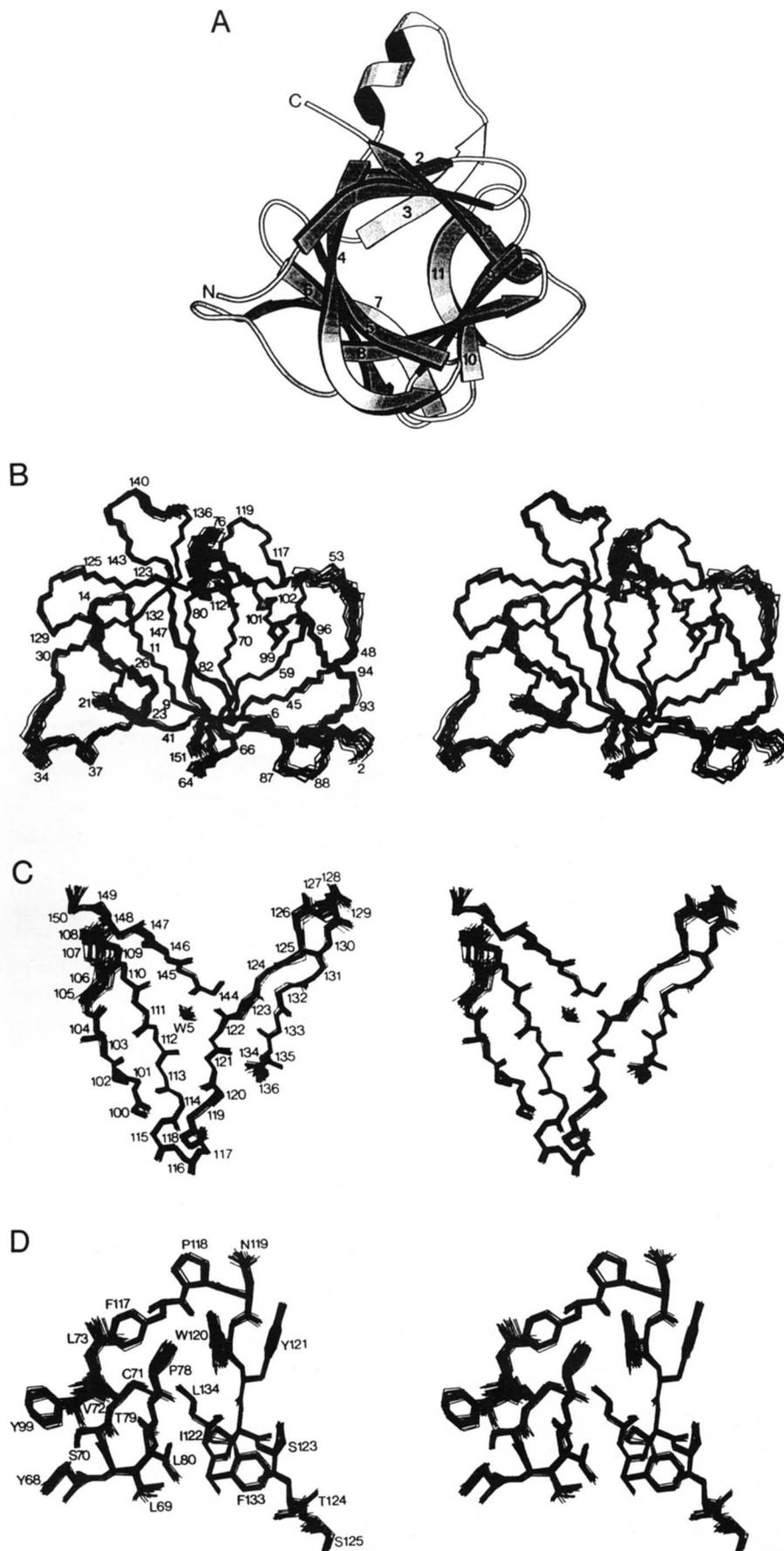


Fig. 6. Solution structure of interleukin-1 β determined by 3D and 4D heteronuclear NMR spectroscopy. **A:** Ribbon diagram of the polypeptide fold. **B:** Superposition of the backbone (N, C α , C) atoms of 32 simulated annealing structures calculated from the experimental NMR data. **C:** Superposition of the backbone (N, C α , C, O) atoms of one of the 3 repeating topological units, illustrating the position of tightly bound water at the interface of the 3 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).

zel 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 & Gronenborn, 1991b). Interestingly, however, the atomic RMS distribution of the 32 calculated solution structures about their mean coordinate positions (~ 0.4 Å for the backbone atoms, ~ 0.8 Å for all atoms, and ~ 0.5 Å for all atoms of internal residues) is approximately the same as the atomic RMS differences between the 3 X-ray structures, indicating that the positional errors in the atomic coordinates determined by the 2 methods are similar (Clore & 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 2 methods, it is not immediately apparent that these discrepancies reflect genuine differences between the solution and X-ray 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.

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 & Bax, 1992; Ikura et al., 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 complex of calmodulin with a target peptide from skeletal muscle myosin light chain kinase (see Kinemage 2; 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 (see Kinemage 3; Omichinski et al., 1993a).

Structure of the calmodulin-target peptide complex

Calmodulin (CaM) is a ubiquitous Ca^{2+} binding protein of 148 residues that is involved in a wide range of cellular Ca^{2+} -dependent signaling pathways, thereby regulating the activity of a large number of proteins (Cohen & Klee, 1988). The crystal

Table 2. Summary of heteronuclear-filtered and -edited NOE experiments used to study protein-ligand complexes comprising a uniformly $^{15}\text{N}/^{13}\text{C}$ labeled protein and an unlabeled ligand

Type of contact	Connectivity
A. Intramolecular protein contacts	
4D $^{13}\text{C}/^{13}\text{C}$ -edited NOE in D_2O	$\text{H}(j)-^{13}\text{C}(j) \text{---} \text{H}(i)-^{13}\text{C}(i)$
4D $^{15}\text{N}/^{13}\text{C}$ -edited NOE in H_2O	$\text{H}(j)-^{15}\text{N}(j) \text{---} \text{H}(i)-^{13}\text{C}(i)$
3D $^{15}\text{N}/^{15}\text{N}$ -edited NOE in H_2O	$\text{H}(j)-^{15}\text{N}(j) \text{---} \text{H}(i)-^{15}\text{N}(i)$
B. Intramolecular ligand contacts	
2D $^{12}\text{C}, ^{14}\text{N}(\text{F}_1)/^{12}\text{C}, ^{14}\text{N}(\text{F}_2)$ -filtered NOE in H_2O^a	$\text{H}(j)-^{12}\text{C}(j) \text{---} \text{H}(i)-^{12}\text{C}(i)$ $\text{H}(j)-^{14}\text{N}(j) \text{---} \text{H}(i)-^{12}\text{C}(i)$ $\text{H}(j)-^{12}\text{C}(j) \text{---} \text{H}(i)-^{14}\text{N}(i)$ $\text{H}(j)-^{14}\text{N}(j) \text{---} \text{H}(i)-^{14}\text{N}(i)$
2D $^{12}\text{C}(\text{F}_1)/^{12}\text{C}(\text{F}_2)$ -filtered NOE in D_2O^a	$\text{H}(j)-^{12}\text{C}(j) \text{---} \text{H}(i)-^{12}\text{C}(i)$
C. Intermolecular protein-ligand contacts	
3D ^{15}N -edited(F_1)/ $^{14}\text{N}, ^{12}\text{C}(\text{F}_3)$ -filtered NOE in H_2O	$\text{H}(j)-^{15}\text{N}(j) \text{---} \text{H}(i)-^{12}\text{C}(i)$ $\text{H}(j)-^{15}\text{N}(j) \text{---} \text{H}(i)-^{14}\text{N}(i)$
3D ^{13}C -edited(F_1)/ $^{12}\text{C}(\text{F}_3)$ -filtered NOE in D_2O	$\text{H}(j)-^{13}\text{C}(j) \text{---} \text{H}(i)-^{12}\text{C}(i)$

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

structure of Ca^{2+} -CaM had been solved a number of years ago (Babu et al., 1985). It is a dumbbell-shaped molecule with an overall length of $\sim 65 \text{ \AA}$ consisting of 2 globular domains, each of which contains 2 Ca^{2+} binding sites of the helix-loop-helix type, connected by a long, solvent-exposed, rigid central helix some 8 turns in length (residues 66–92). In solution, on the other hand, ^1H - ^{15}N NMR relaxation measurements have demonstrated unambiguously that the central helix is disrupted near its midpoint 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 2 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 CaM binding domain of skeletal muscle myosin light

chain kinase (Kinemage 2). The solution structure was determined on the basis of 1,995 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 2 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 \AA for the backbone atoms and 1.4 \AA for all atoms. Thus this structure represents a second generation structure in the classification (Clare & Gronenborn, 1991a). A stereo view showing a best fit superposition of the 24 calculated structures is shown in Figure 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 to 81, flanked by 2 helices (residues 65–73

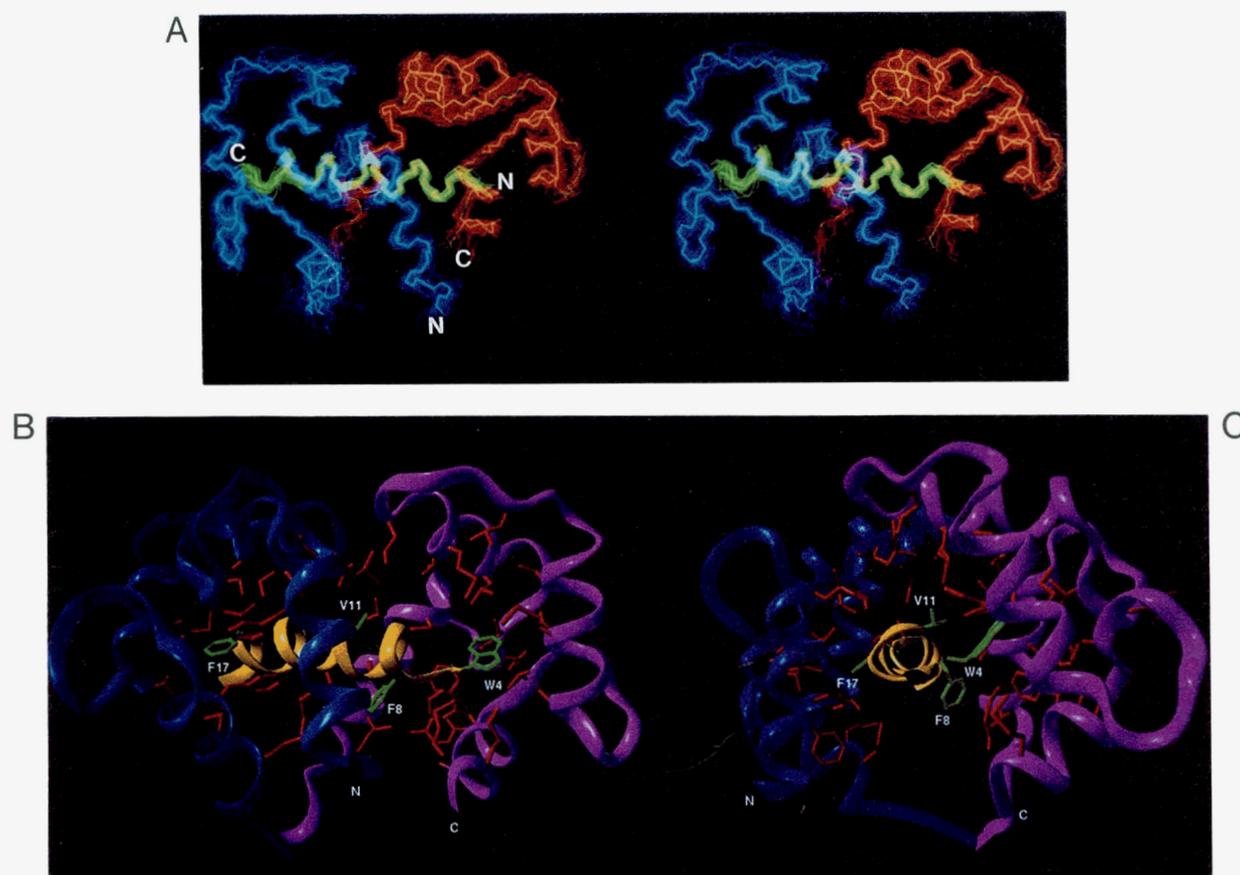


Fig. 7. Solution structure of the Ca^{2+} -CaM-M13 peptide complex determined by 3D and 4D heteronuclear NMR spectroscopy. **A:** Superposition of the backbone (N, C^α , C) atoms of 24 simulated annealing structures calculated from the experimental NMR data; the N- and C-terminal domains of calmodulin are shown in blue and red, respectively, and the M13 peptide is in green; the restrained regularized average structure is highlighted. **B, 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 VISIP (de Castro & Edelstein, 1992). The coordinates are from Ikura et al. (1992) (PDB accession code 1BBM).

and 83–93), thereby enabling the 2 domains to come together gripping the peptide rather like 2 hands capturing a rope. The hydrophobic channel formed by the 2 domains is complementary in shape to that of the peptide helix. This is clearly illustrated by the schematic ribbon drawings shown in Figure 7B and C, which also serve to highlight the approximate 2-fold pseudosymmetry of the complex. Thus, whereas the 2 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 2-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 residues (3–21) of M13 that interact with CaM, whereas 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 1,848 and 1,477 \AA^2 , respectively, which corresponds to a decrease in the calculated solvation free energy of folding (Eisenberg & McLaglan, 1986) of 18 and 20 kcal·mol⁻¹, respectively. This large decrease in solvation free energy would account for the very tight binding ($K_{\text{ass}} \sim 10^9 \text{ M}^{-1}$) 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 \AA^2 compared to an accessible surface area of 3,123 \AA^2 for a random coil and 2,250 \AA^2 for a helix. Thus, over 80% of the surface of the peptide in contact with CaM is buried.

In the view shown in Figure 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; the floor is formed by the flexible loop (residues 74–82) connecting the 2 domains and by helix VIII (residues 138–146) of the C-terminal domain. The front of the channel in Figure 7B and the left wall of the channel in Figure 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 Figure 7B and the right wall of the channel in Figure 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 2 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, whereas 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 \AA . The helical M13 peptide passes through the center of the ellipsoid at an angle of $\sim 45^\circ$ to its long axis. By way of contrast the approximate dimensions of the Ca^{2+} -CaM X-ray structure are 65 × 30 × 30 \AA (Babu et al., 1985). In addition, the calculated radius of gyration for Ca^{2+} -CaM-M13 is $\sim 17 \text{ \AA}$ which is completely consistent with the decrease in the radius of gyration from $\sim 21 \text{ \AA}$ to $\sim 16 \text{ \AA}$ 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 Figure 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 that are unusually abundant in CaM, in particular 4 methionines in the C-terminal domain (Met-109, Met-124, Met-144, and Met-145) and 3 methionines in the N-terminal domain (Met-36, Met-51, and Met-71). Because methionine is an unbranched hydrophobic residue extending over 4 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 & 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 Figure 8. Glu-11 and Glu-14 in helix I are within 5 \AA 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-3 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 2 domains. Similarly, the structure of the complex explains the finding that as many as 4 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., 1991a), as the long flexible loop connecting the 2 domains can readily be shortened without causing any alteration in the structure (cf. Fig. 7). The observation from photoaffinity labeling studies that the 2 domains of CaM interact simultaneously with opposite ends of the peptide such that residue 4 of the peptide (numbering for M13) can be crosslinked to Met-124 or Met-144 of the C-terminal domain and that residue 13 of the peptide can be crosslinked 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, whereas the C-terminal half of the peptide interacts predominantly with the N-terminal domain (Figs. 7, 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 & 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 crosslinking of residues 3 and 146 of CaM, mu-

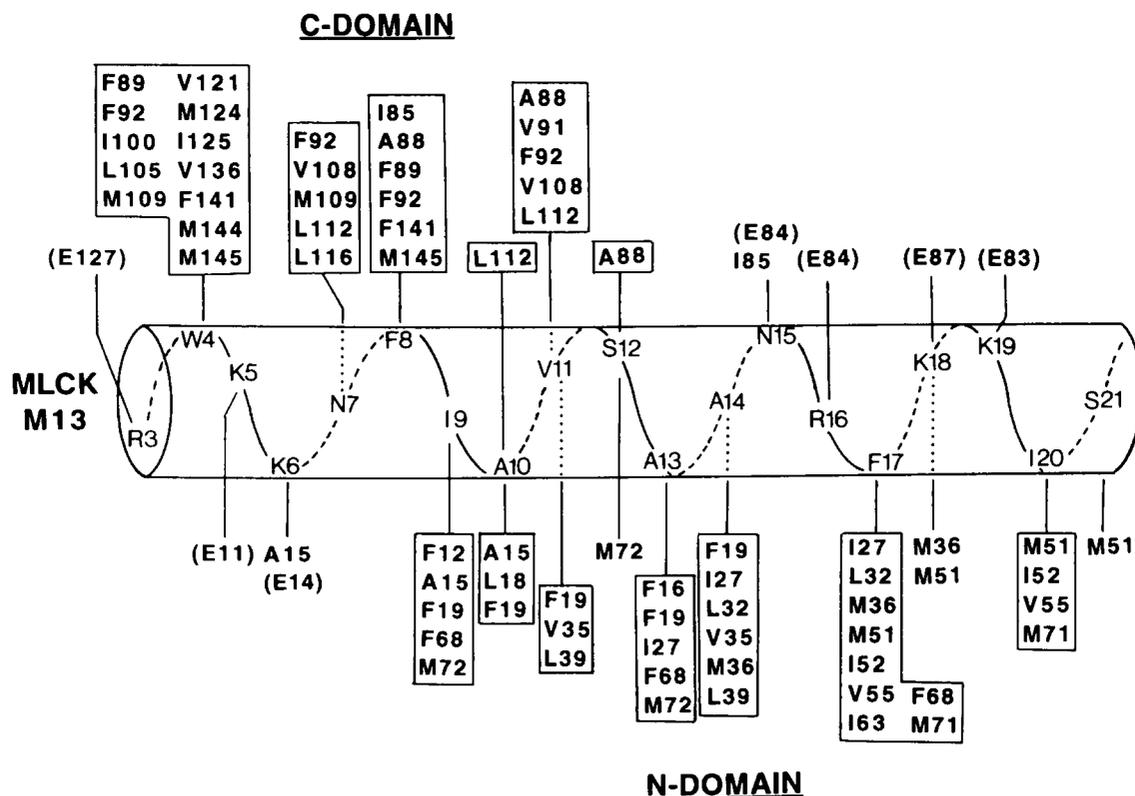


Fig. 8. Summary of residue pairs for which intermolecular NOEs between CaM and M13 are observed. CaM residues involved in hydrophobic interactions are boxed. Also included are potential electrostatic interactions between negatively charged Glu residues of CaM (shown in parentheses) and positively charged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

tated 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 & Kretsinger, 1988). Thus, although the C α carbons of residues 3 and 146 are 37 Å apart in the X-ray structure of 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 crosslinking 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 & Klee, 1988). A comparison of these sequences reveals little homology. Nevertheless, many of the very tightly binding peptides ($K_{ass} \geq 5 \times 10^7 \text{ M}^{-1}$) 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 Figure 9. In the case of M13, these 2 residues are Trp-4 and Phe-17, which are exclusively in contact with the C- and N-terminal domains of CaM, respectively (Figs. 7, 8). Given that these 2 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 Figure 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 equiva-

lent to Phe-8, which interacts with the C-terminal domain, and Val-11, which interacts with both domains (cf. Figs. 7, 8). In addition, there are no acidic residues present that 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 2 hydrophobic residues at the N- and C-termini (Fig. 9) and have approximately the same equilibrium association constant ($K_{ass} \sim 1-3 \times 10^9 \text{ M}^{-1}$) 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 dumbbell shape of Ca $^{2+}$ -CaM. This is exactly what has been observed by small angle X-ray scattering using 2 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., 1991b). 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 dumbbell 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

	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	
Ca Pump 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 L	
Mastaporan X		I N W K	G I A A	M A K K L	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 domain of CaM		C	C	C	N	
				N		

Fig. 9. Alignment of tightly binding ($K_{ass} > 5 \times 10^7 \text{ M}^{-1}$) 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.

peptide is sequestered into a hydrophobic channel formed by the 2 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 2 long chain hydrophobic or aromatic residues separated by 12 residues in order to anchor the peptide to the 2 domains of CaM (Fig. 7). By analogy, the rope (i.e., the CaM binding domain of the target) has to be long enough and have 2 knots at each end for the 2 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 & Krebs, 1987; Guerini & 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 2 discontinuous CaM binding sites that 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 (Trehwella 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 & Kennedy, 1987), the CaM binding sequences do not have the same spacing of hydrophobic residues seen in M13 and the other sequences listed in Figure 13, and, in addition, CaM kinase II is not susceptible to proteolysis in the absence of phosphorylation (Kwiatkowski & King, 1989), suggesting that the mode of binding is different again. Thus, in all likelihood, the complex of Ca^{2+} -CaM with the M13 peptide from skeletal muscle myosin light chain kinase represents one of a range of Ca^{2+} -CaM binding modes achieving CaM-target protein interactions in an efficient and elegant manner.

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 & 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 2 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 & 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 ($K_{ass} \sim 1.2 \times 10^8 \text{ M}^{-1}$) (Omichinski et al., 1993b). 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-bp oligonucleotide containing the target sequence AGATAA, by means of multidimensional heteronuclear filtered and separated NMR spectroscopy (see Kinemage 3; Omichinski et al., 1993a).

The structure calculations were based on a total of 1,772 experimental NMR restraints, including 117 intermolecular interproton distance restraints between the protein and the DNA. A stereo view 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 Figure 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 bp 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.

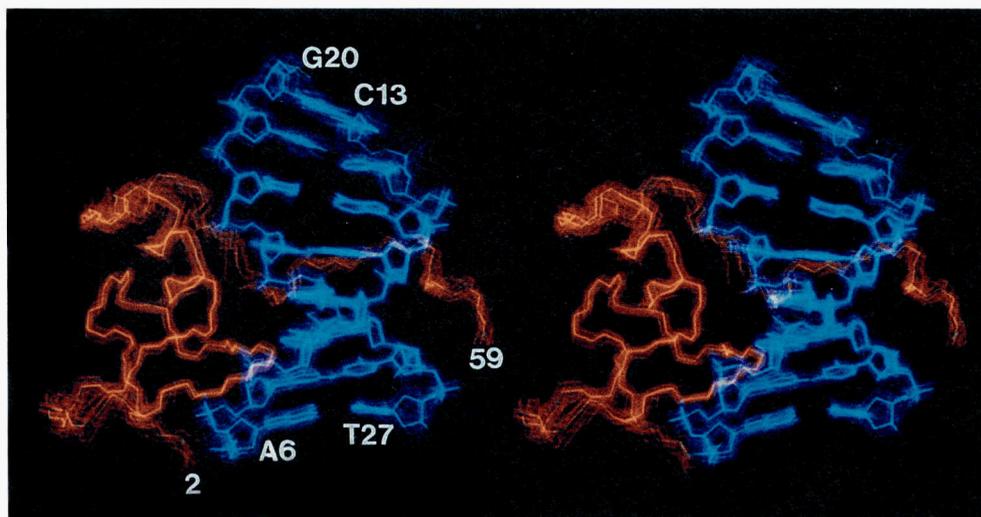


Fig. 10. Stereo view showing a superposition of the 30 simulated annealing structures of the specific complex of the DNA binding domain of cGATA-1 with DNA calculated on the basis of the experimental NMR data derived from 3D and 4D heteronuclear NMR spectroscopy. The backbone (N, C α , C) atoms of cGATA-1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

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 bp, 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 divided into 2 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 Figure 11A. The core starts out with a turn (residues 2–5), followed by 2 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, and 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 2 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 conserved between the 2 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 4 coordinating cysteines, are identical. The loop between strands β 2 and β 3 has 3 deletions, and the turn between strands β 3 and β 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 2 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 2 strands and 2 helices, whereas 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 Figure 12A and B. The conformation of the oligonucleotide is B-type. The helix and the loop connecting strands β 2 and β 3 (which is located directly beneath the helix) are located in the major groove, whereas 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 $\sim 40^\circ$ to the base planes of the DNA (Fig. 12A), whereas 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 Figure 12C and D, respectively, and a schematic representation of all the contacts is provided in Figure 13. The cGATA-1 DNA binding domain makes specific contacts with 8 bases, 7 in the 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-

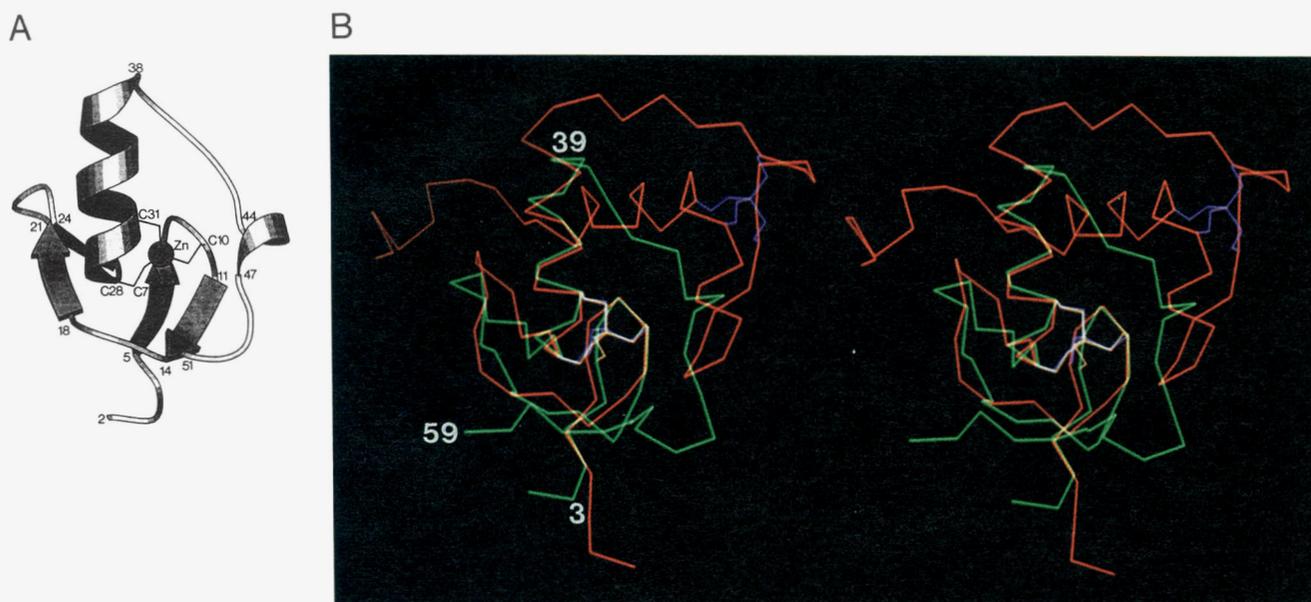


Fig. 11. A: Schematic ribbon drawing of the core of the cGATA-1 DNA binding domain. **B:** Superposition of the C α atoms of the cGATA-1 (green) and glucocorticoid receptor (red) DNA binding domains. The zinc and coordinating cysteines are shown in yellow for cGATA-1 and in purple for the glucocorticoid 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 deviation 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. (1993a) (PDB accession code 1GAT).

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 3 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 δ 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 1,127 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 & McLaglan, 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 sugar 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 3 classes of zinc-containing DNA binding domains whose structures have previously been solved (Luisi et al., 1991; Pavletich & Pabo, 1991, 1993; Mamorstein et al., 1992; Fairall et al., 1993; Schwabe 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 that 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 4 Cys residues that coordinate zinc, there is no significant sequence identity between these regions of the 2 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

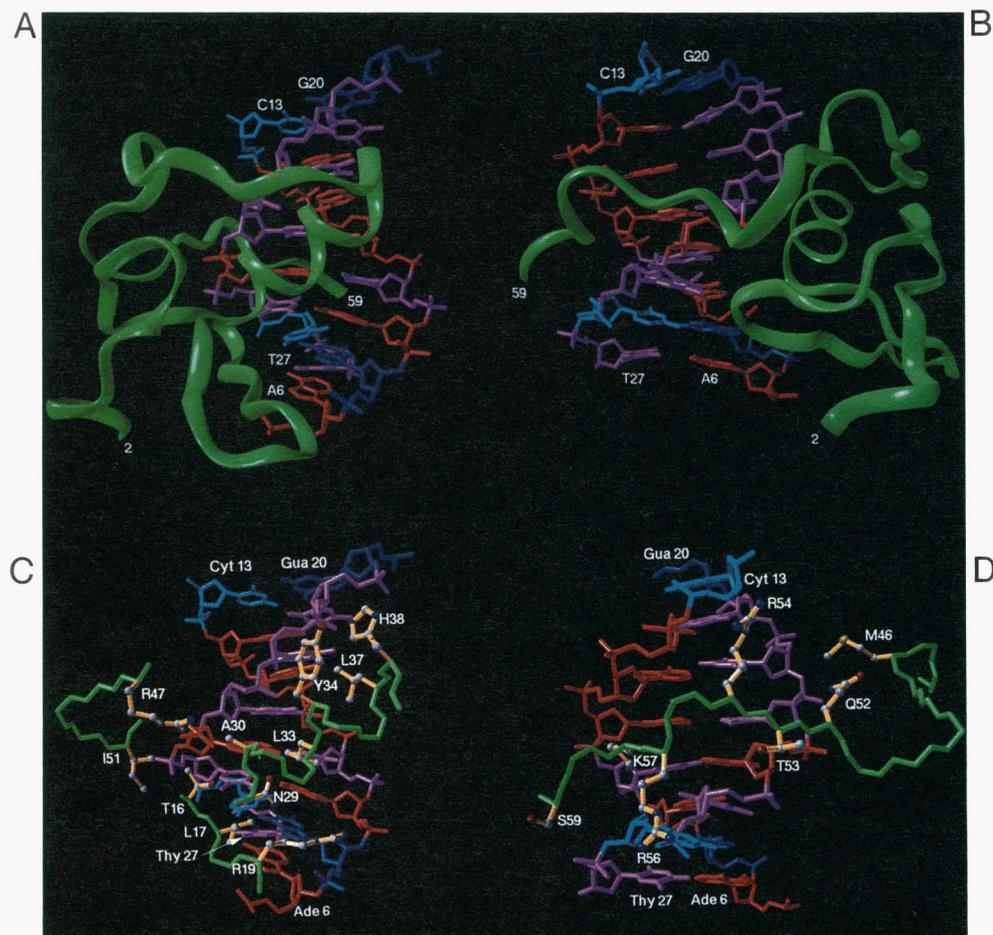


Fig. 12. **A, B:** Schematic ribbon drawings illustrating the interactions of cGATA-1 with DNA. **C, D:** Side-chain interactions between cGATA-1 and the DNA in the major and minor grooves, respectively. The protein backbone is shown in green and the protein side chains in yellow; the color code for 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 & Edelstein, 1992). The coordinates of the cGATA-1-DNA complex are from Omichinski et al. (1993a) (PDB accession code 1GAT).

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).

Concluding remarks

From the examples presented in this review it should be clear that the recent development of a whole range of highly sensitive multidimensional 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 multidimensional heteronuclear NMR spectroscopy. In addition to the 3 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 multidimensional heteronuclear experiments coupled with semi-automated assignment procedures will result in many more NMR structures of such larger proteins and protein-ligand complexes.

Acknowledgments

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,

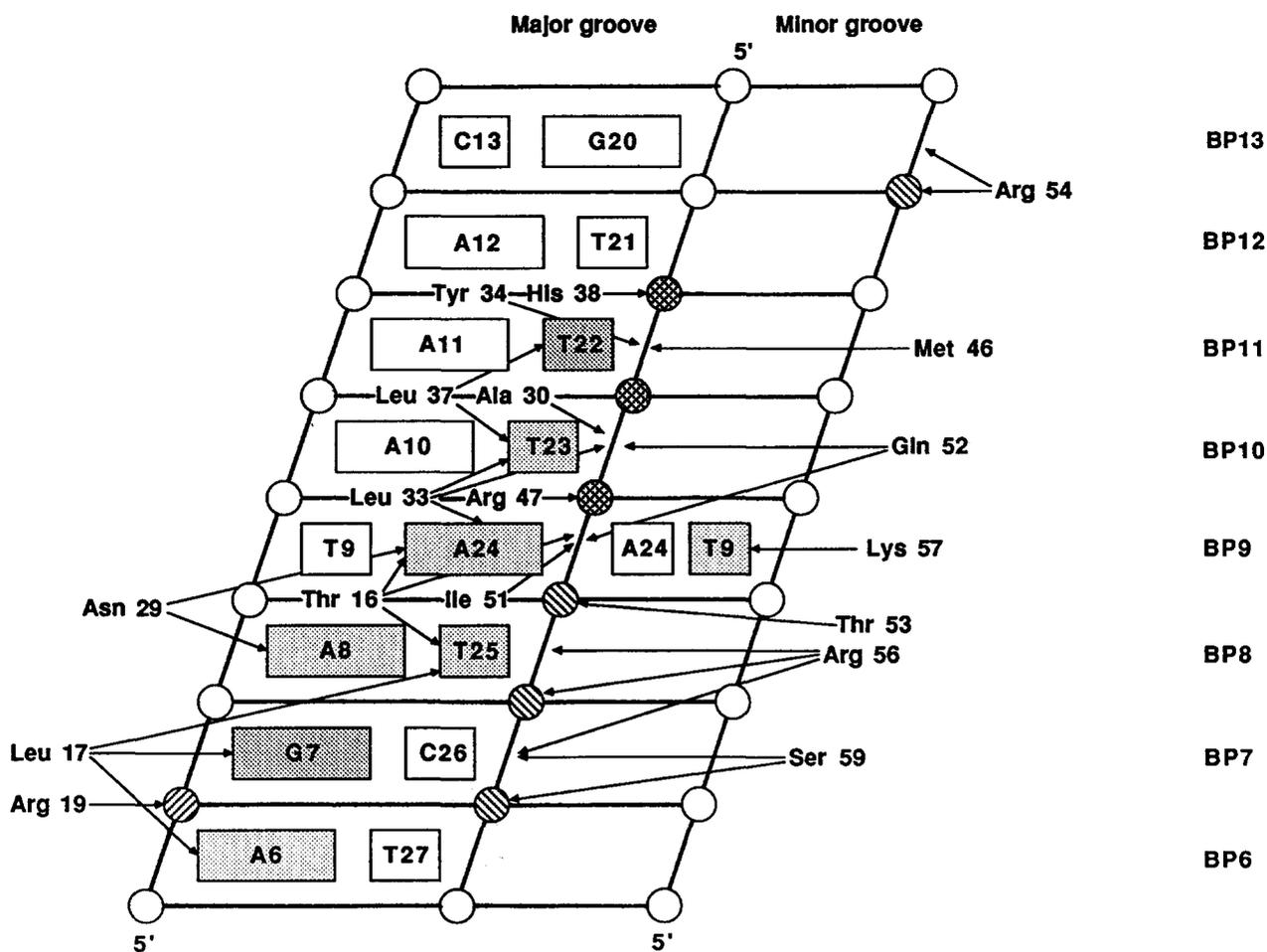


Fig. 13. Schematic diagram summarizing the contacts between cGATA-1 and DNA. The DNA is represented as a cylindrical projection. Bases interacting with the protein are shaded; phosphates are represented as circles; circles with hatches directed toward the right and left indicate sites of interaction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

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

- Babu YS, Sack JS, Greenhough TJ, Bugg CE, Means AR, Cook WJ. 1985. Three dimensional structure of calmodulin. *Nature* 315:37-40.
- Barbato G, Ikura M, Kay L, Pastor RW, 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, Grzesiek S. 1993. Methodological advances in protein NMR. *Acc Chem Res* 26:131-138.
- Bax A, Pochapsky SS. 1992. Optimized recording of heteronuclear multidimensional NMR spectra using pulse field gradients. *J Magn Reson* 99:638-643.
- Bennett MK, Kennedy MB. 1987. Deduced primary structure of the β subunit of brain type II Ca^{2+} /calmodulin dependent protein kinase determined by molecular cloning. *Proc Natl Acad Sci USA* 84:1794-1798.
- Blumenthal DK, Krebs EG. 1987. Preparation and properties of the calmodulin binding domain of skeletal muscle myosin light chain kinase. *Methods Enzymol* 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 AT. 1992. The free R value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature (Lond)* 355:472-474.
- Charbonneau H, Kumar S, Novack JP, Blumenthal DK, Griffin PR, Shabanowitz J, Hunt DF, Beavo JA, Walsh KA. 1991. Evidence for domain organization within the 61-kDa calmodulin-dependent cyclic nucleotide phosphodiesterase from bovine brain. *Biochemistry* 30:7931-7940.
- Clore GM, Bax A, Driscoll PC, Wingfield PT, Gronenborn AM. 1990a. Assignment of the side chain ^1H and ^{13}C resonances of interleukin- 1β using double and triple resonance heteronuclear three-dimensional NMR spectroscopy. *Biochemistry* 29:8172-8184.
- Clore GM, Bax A, Wingfield PT, Gronenborn AM. 1990b. Identification and localization of bound internal water in the solution structure of interleukin- 1β by heteronuclear three-dimensional ^1H rotating frame Overhauser ^{15}N - ^1H multiple quantum coherence NMR spectroscopy. *Biochemistry* 29:5671-5676.
- Clore GM, Driscoll PC, Wingfield PT, Gronenborn AM. 1990c. Low resolution structure of interleukin- 1β in solution derived from ^1H - ^{15}N heteronuclear three-dimensional NMR spectroscopy. *J Mol Biol* 214:811-817.
- Clore GM, Gronenborn AM. 1987. Determination of three-dimensional structures of proteins in solution by nuclear magnetic resonance spectroscopy. *Protein Eng* 1:275-288.
- Clore GM, Gronenborn AM. 1989. Determination of three-dimensional structures of proteins and nucleic acids in solution by nuclear magnetic resonance spectroscopy. *CRC Crit Rev Biochem Mol Biol* 24:479-564.
- Clore GM, Gronenborn AM. 1991a. Structures of larger proteins in solution: Three- and four-dimensional heteronuclear NMR spectroscopy. *Science* 252:1390-1399.

- Clare GM, Gronenborn AM. 1991b. Comparison of the solution nuclear magnetic resonance and X-ray crystal structures of human recombinant interleukin- β . *J Mol Biol* 221:47-53.
- Clare GM, Gronenborn AM. 1991c. Applications of three- and four-dimensional heteronuclear NMR spectroscopy to protein structure determination. *Progr Nucl Magn Reson Spectrosc* 23:43-92.
- Clare GM, Gronenborn AM. 1991d. Two, three and four dimensional NMR methods for obtaining larger and more precise three-dimensional structures of proteins in solution. *Annu Rev Biophys Chem* 20:29-63.
- Clare GM, Kay LE, Bax A, Gronenborn AM. 1991a. Four dimensional $^{13}\text{C}/^{13}\text{C}$ -edited nuclear Overhauser enhancement spectroscopy of a protein in solution: Application to interleukin- β . *Biochemistry* 30:12-18.
- Clare GM, Robien MA, Gronenborn AM. 1993. Exploring the limits of precision and accuracy of protein structures determined by nuclear magnetic resonance spectroscopy. *J Mol Biol* 231:82-102.
- Clare GM, Wingfield PT, Gronenborn AM. 1991b. High resolution three-dimensional structure of interleukin- β in solution by three and four dimensional nuclear magnetic resonance spectroscopy. *Biochemistry* 30:2315-2323.
- Cohen P, Klee CB. 1988. *Molecular aspects of cellular recognition, vol 5*. New York: Elsevier.
- Cox JA, Comte M, Fitton JE, DeGrado WF. 1985. The interaction of calmodulin with amphiphilic peptides. *J Biol Chem* 260:2527-2534.
- Dasgupta M, Honeycutt T, Blumenthal DK. 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, Edelstein S. 1992. *VISP 1.0 user's guide*. Geneva: University of Geneva.
- Driscoll PC, Clare GM, Marion D, Wingfield PT, Gronenborn AM. 1990a. Complete resonance assignment for the polypeptide backbone of interleukin- β using three-dimensional heteronuclear NMR spectroscopy. *Biochemistry* 29:3542-3556.
- Driscoll PC, Gronenborn AM, Wingfield PT, Clare GM. 1990b. Determination of the secondary structure and molecular topology of interleukin- β using two- and three-dimensional heteronuclear ^{15}N - ^1H NMR spectroscopy. *Biochemistry* 29:4468-4482.
- Dyson HJ, Gippert GP, Case DA, Holmgren A, Wright PE. 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, McLaglan AD. 1986. Solvation energy in protein folding and binding. *Nature* 319:199-203.
- Ernst RR, Bodenhausen G, Wokaun A. 1987. *Principles of nuclear magnetic resonance in one and two dimensions*. Oxford, UK: Clarendon Press.
- Evans T, Felsenfeld G. 1989. The erythroid-specific transcription factor eryf1: A new finger protein. *Cell* 58:877-885.
- Fairall L, Schwabe JWR, Chapman L, Finch JT, 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 WJ, Gippert GP, Reizer J, Saier MH, Wright PE. 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 SW, Zuiderweg ERP. 1988. Heteronuclear three-dimensional NMR spectroscopy: A strategy for the simplification of homonuclear two-dimensional NMR spectra. *J Magn Reson* 78:588-593.
- Fesik SW, Zuiderweg ERP. 1990. Heteronuclear three-dimensional NMR spectroscopy of isotopically labelled biological macromolecules. *Q Rev Biophys* 23:97-131.
- Finzel BC, Clancy LL, Holland DR, Muchmore SW, Watenpaugh KD, Einspahr HM. 1989. Crystal structure of recombinant human interleukin- β at 2.0 Å resolution. *J Mol Biol* 209:779-791.
- Forman-Kay JD, Clare GM, Wingfield PT, Gronenborn AM. 1991. The high resolution three-dimensional structure of reduced recombinant human thioredoxin in solution. *Biochemistry* 30:2685-2698.
- Guerini D, Klee CB. 1991. Structural diversity of calcineurin Ca^{2+} -calmodulin stimulated phosphatases. *Adv Protein Phosphatases* 6:391-410.
- Güntert P, Braun W, Billeter M, 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, Gould H. 1991. Structure and promoter activity of the gene for the erythroid transcription factor GATA-1. *Proc Natl Acad Sci USA* 88:3004-3008.
- Havel TF. 1991. An evaluation of computational strategies for use in the determination of protein structure from distance constraints obtained by nuclear magnetic resonance. *Progr Biophys Mol Biol* 56:43-78.
- Havel TF, Kuntz ID, Crippen GM. 1983. Theory and practice of distance geometry. *Bull Math Biol* 45:665-720.
- Havel TF, 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 Mol Biol* 182:281-294.
- Heidorn DB, Seeger PA, Rokop SE, Blumenthal DK, Means AR, Crespi H, Trehwella 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 RE, John BK. 1991. Gradient-enhanced proton-detected heteronuclear multiple-quantum coherence spectroscopy. *J Magn Reson* 91:648-653.
- Hyberts SG, Märki W, 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, 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, Clare GM, Gronenborn AM, Zhu G, Klee CB, Bax A. 1992. Solution structure of a calmodulin-target peptide complex by multidimensional NMR. *Science* 256:632-638.
- Ikura M, Kay LE, 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 LE, Krinks M, 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.
- Kataoka M, Head JF, Persechini A, Kretsinger RH, Engelman DM. 1991a. 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.
- Kataoka M, Head JF, Vorherr T, Krebs J, Carafoli E. 1991b. 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.
- Kay LE, Clare GM, Bax A, Gronenborn AM. 1990. Four-dimensional heteronuclear triple resonance NMR spectroscopy of interleukin- β in solution. *Science* 249:411-414.
- Kraulis PJ. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* 24:946-950.
- Kwiatkowski AP, King MM. 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.
- Luisi BF, Xu WX, Otwinowski Z, Freedman LP, Yamamoto KR, Sigler PG. 1991. Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* 352:497-505.
- Lukas TJ, Burgess WH, Predergast FG, Lau W, Watterson DM. 1986. Calmodulin binding domains: Characterization of a phosphorylation and calmodulin binding site from myosin light chain kinase. *Biochemistry* 25:1458-1464.
- Mamorstein R, Carey M, Ptashne M, Harrison SC. 1992. DNA recognition by GAL4: Structure of a protein-DNA complex. *Nature* 356:408-414.
- Marion D, Driscoll PC, Kay LE, Wingfield PT, Bax A, Gronenborn AM, Clare GM. 1989. Overcoming the overlap problem in the assignment of ^1H -NMR spectra of larger proteins using three-dimensional heteronuclear ^1H - ^{15}N Hartmann-Hahn and nuclear Overhauser-multiple quantum coherence spectroscopy. Application to interleukin- β . *Biochemistry* 29:6150-6156.
- Martin D, Orkin S. 1986. Transcriptional activation and DNA binding by the erythroid factor GF-1/NF-E1/Eryf 1. *Genes Dev* 4:1886-1989.
- Montelione GT, 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 GT, Wagner G. 1990. Conformation independent sequential NMR connections in isotope-enriched polypeptides by ^1H - ^{13}C - ^{15}N triple resonance experiments. *J Magn Reson* 87:183-188.
- Nilges M, Clare GM, Gronenborn AM. 1988. Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Lett* 229:317-324.
- Nilges M, Clare GM, Gronenborn AM. 1990. ^1H -NMR stereospecific assignments by conformational database searches. *Biopolymers* 29:813-822.
- Omichinski JG, Clare GM, Schaad O, Felsenfeld G, Trainor C, Appella E, Stahl SJ, Gronenborn AM. 1993a. NMR structure of a specific DNA complex of the Zn-containing DNA binding domain of GATA-1. *Science* 261:438-446.
- Omichinski JG, Trainor C, Evans T, Gronenborn AM, Clare GM, Felsenfeld G. 1993b. A small single-finger peptide from the erythroid factor

- GATA-1 binds specifically to DNA as a zinc or iron complex. *Proc Natl Acad Sci USA* 90:1676-1680.
- O'Neil KT, DeGrado WF. 1990. How calmodulin binds its targets: Sequence independent recognition of amphiphilic α -helices. *Trends Biochem Sci* 15:59-64.
- O'Neil KT, Erickson-Viitanen S, DeGrado WF. 1989. Photolabeling of calmodulin with basic, amphiphilic α -helical peptides containing *p*-benzoylphenylalanine. *J Biol Chem* 264:14571-14578.
- Orkin SH. 1992. GATA-binding transcription factors in hematopoietic cells. *Blood* 80:575-581.
- Oschkinat H, Griesinger C, Kraulis PJ, Sørensen OW, Ernst RR, Gronenborn AM, Clore GM. 1988. Three-dimensional NMR spectroscopy of a protein in solution. *Nature (Lond)* 332:374-376.
- Pavletich NP, Pabo CO. 1991. Zinc finger-DNA recognition: Crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* 252:809-817.
- Pavletich NP, Pabo CO. 1993. Crystal structure of a five-finger GLI-DNA complex: New perspectives on zinc fingers. *Science* 261:1701-1707.
- Persechini A, Blumenthal DK, Jarrett HW, Klee CB, Hardy DO, Kretsinger RH. 1989. The effects of deletions in the central helix of calmodulin on enzyme activation and peptide binding. *J Biol Chem* 264:8052-8058.
- Persechini A, Kretsinger RH. 1988. The central helix of calmodulin functions as a flexible tether. *J Biol Chem* 263:12175-12178.
- Powers R, Garrett DS, March CJ, Frieden EA, Gronenborn AM, Clore GM. 1992. Three-dimensional structure of interleukin-4 by multi-dimensional heteronuclear magnetic resonance spectroscopy. *Science* 256:1673-1677.
- Powers R, Garrett DS, March CJ, Frieden EA, Gronenborn AM, Clore GM. 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 JP, Schär HP, Grütter MG. 1989. Crystallographic refinement of interleukin-1 β at 2.0 Å resolution. *Proc Natl Acad Sci USA* 86:9667-9671.
- Schwabe JWR, Chapman L, Finch JT, 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.
- Shaanan B, Gronenborn AM, Cogen GH, Gilliland GL, Veerapandian B, Davies DR, Clore GM. 1992. Combining experimental information from crystal and solution studies: Joint X-ray and NMR refinement. *Science* 257:961-964.
- Smith LJ, Redfield C, Boyd J, Lawrence GMP, Edwards RG, Smith RAG, Dobson CM. 1992. Human interleukin-4: The solution structure of a four-helix bundle protein. *J Mol Biol* 224:899-904.
- Spera S, Ikura M, 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 Biomol NMR* 1:155-165.
- Theriault Y, Logan TM, Meadows R, Yu L, Olejniczak ET, Holzman T, Simmer RL, Fesik TM. 1993. Solution structure of the cyclosporin A/cyclophilin complex by NMR. *Nature* 361:88-91.
- Trewhella J, Blumenthal DK, Rokop SE, Seeger PA. 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 GL, Raag R, Svensson AL, Masui Y, Hirai Y, Poulos TL. 1992. Functional implications of interleukin-1 β based on the three dimensional structure. *Proteins Struct Funct Genet* 12:10-24.
- Vuister GW, Boelens R, Kaptein R, Hurd RE, John BK, Van Zijl PCM. 1991. Gradient enhanced HMQC and HSQC spectroscopy: Applications to ¹⁵N-labeled Mnr repressor. *J Am Chem Soc* 113:9688-9690.
- Vuister GW, Clore GM, Gronenborn AM, Powers R, Garrett DS, Tschudin R, Bax A. 1993. Increased resolution and improved quality in four-dimensional ¹³C/¹³C separated HMQC-NOE-HMQC spectra using pulse field gradients. *J Magn Reson Ser B* 101:210-213.
- Vuister GW, Grzesiek S, Delaglio F, Wang AC, Tschudin R, Zhu G, Bax A. 1994. Measurement of homo- and heteronuclear J couplings from quantitative J correlation. *Methods Enzymol*. Forthcoming.
- Wagner G, Braun W, Havel TF, Schauman T, Go N, 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 Mol Biol* 196:611-639.
- Wüthrich K. 1986. *NMR of proteins and nucleic acids*. New York: Wiley.
- Zuiderweg ERP, Boelens R, Kaptein R. 1985. Stereospecific assignment of ¹H-NMR methyl lines and conformation of valyl residues in the lac repressor headpiece. *Biopolymers* 24:601-611.
- Zuiderweg ERP, Petros AM, Fesik SW, Olejniczak ET. 1991. Four-dimensional [¹³C, ¹H, ¹³C, ¹H] HMQC-NOE-HMQC NMR spectroscopy: Resolving tertiary NOE distance restraints in spectra of larger proteins. *J Am Chem Soc* 113:370-372.