Critical Reviews in Biochemistry and Molecular Biology, 30(5):351–385 (1995)

Structures of Protein Complexes by Multidimensional Heteronuclear Magnetic Resonance Spectroscopy

Angela M. Gronenborn and G. Marius Clore

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892–0520

ABSTRACT: With the advent of multidimensional heteronuclear-edited and -filtered NMR experiments, the field of three-dimensional structure determination by NMR has again increased in scope, making it possible to move the technology beyond the approximately 10 kDa limit inherent to conventional two-dimensional NMR to systems up to potentially 35 to 40 kDa. This article outlines the basic strategies for solving three-dimensional structures of larger systems, in particular, protein complexes and multimeric proteins using three- and four-dimensional NMR spectroscopy, summarizes the key experiments, and illustrates the power of these methods using several examples of protein-DNA, protein-peptide complexes, and oligomeric proteins from the authors' laboratories.

KEY WORDS: multidimensional NMR, protein complexes, DNA complexes, solution strucutre.

I. INTRODUCTION

The size of macromolecular structures that can be solved by nuclear magnetic resonance (NMR) has been dramatically increased over the last few years (Clore and Gronenborn, 1991a). Equally importantly, this advance has been accompanied by significant improvements in the accuracy with which such structures can be determined. Thus, it is now possible to determine the structures of proteins in the 15- to 25-kDa range at a resolution comparable to 2- to 2.5-Å-resolution crystal structures (Clore and Gronenborn, 1991b). The key methodological developments responsible for these advances comprise 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 (for reviews, see Clore and Gronenborn, 1991a, c, d, 1994; Bax and Grzesiek, 1993) in conjunction with improved refinement techniques incorporating accurate coupling constants (Garrett et al., 1994) and ^{13}C and ^{1}H chemical shifts (Kuszewski et al., 1995a, b). In this review, we summarize some of these developments and illustrate their application with a number of examples from our laboratory dealing with protein-peptide, protein-DNA, and protein-protein (in the form of oligomers) complexes, specifically, a complex of calmodulin with a target peptide from skeletal muscle myosin light-chain kinase (Ikura et al., 1992), a complex of human thioredoxin and a NFκB peptide (Qin et al., 1995), complexes of the DNA binding domains of the transcription factor GATA-1 (Omichinski et al., 1993a) and SRY (Werner et al., 1995a) with their cognate DNA target sites, and the tetramerization domain of the tumor suppressor p53 (Clore et al., 1994, 1995a, b). These examples are not intended to provide a comprehensive review of all the complexes that have been solved by NMR but to provide the reader with a feel for the type and complexity of systems that can be tackled, the quality of the structures that can be obtained, and the sort of biological information that can be gleaned.

II. GENERAL STRATEGY FOR THE STRUCTURE DETERMINATION OF COMPLEXES BY NMR

The main source of geometric information used in protein structure determination lies in the

1040-9238/95/$.50

nuclear Overhauser effect (NOE), 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 (at short mixing times) 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 approximate interproton distance restraints derived from NOE measurements can be highly conformationally restrictive, particularly when they involve residues that are far apart in the sequence but close together in space (Clore et al., 1993a).

The power of NMR over other spectroscopic techniques results from the fact that every proton gives rise to an individual resonance in the spectrum that 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 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 (3D) structures on the basis of the experimental NMR restraints using one or more of a number of algorithms such as distance geometry and/or simulated annealing (for reviews of the various methodologies, see Braun, 1987; Clore and Gronenborn, 1989; Havel, 1991). 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 possible to employ iterative methods to resolve the vast majority of ambiguities. Consider, for example, an NOE cross peak that 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 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.

If 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). 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 (Ikura and Bax, 1992; Ikura et al., 1992; Otting and Wüthrich, 1990; Fesik et al., 1991). These experiments are summarized in Table 1 and have been employed successfully in a number of laboratories for a range of systems. Examples of protein-drug complexes are cyclophylin-cyclosporin (Thériault et al., 1993; Spitzfaden et al., 1994) and FK506 binding protein-ascomycin (Meadows et al., 1993). The cyclophylin-cyclosporin studies employed uniformly ^{13}C and ^{15}N-labeled protein as well as ^{13}C-labeled drug. The structures of several protein-peptide complexes have also been determined by NMR. The first example was the structure of calmodulin bound to its target peptide from skeletal muscle myosin light-chain kinase (Ikura et al., 1992). Since then, a number of other protein-peptide complexes have been determined. These include several SH2 (Pascal et al., 1994; Xu et al., 1995)- and SH3-peptide complexes (Wittekind et al., 1994; Terasawa et al., 1994; Goudreau et al., 1994; Yu et al., 1994; Feng et al., 1994), and mixed disulfide intermediates of glutaredoxin and glutathione (Bushweller et al., 1994) and human thioredoxin and its target peptide from the transcription factor NFκB (Qin et al., 1995). The structures of protein-DNA complexes have been the focus of several laboratories, and to date a number have been determined by NMR: the specific complex of the DNA-binding domain of the transcription factor GATA-1 with its target site (Omichinski et al., 1993a), the lac repressor headpiece-operator complex (Chuprina et al., 1993),

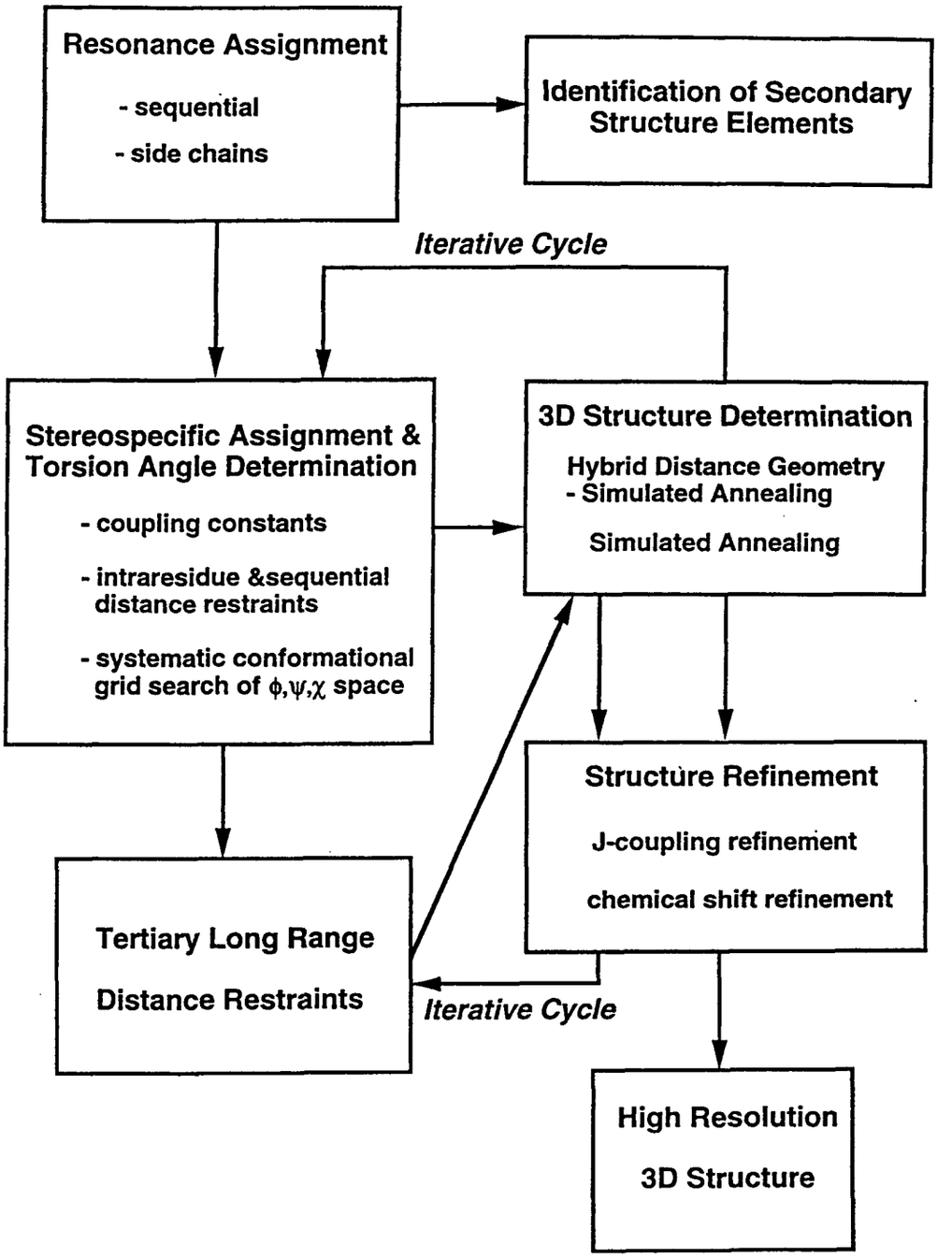


FIGURE 1. Summary of the general strategy employed to solve 3D structures of macromolecules by NMR.

the antenennapedia homeodomain-DNA complex (Billeter et al., 1993), the minimal DNA-binding domain of the protooncogene c-myb complexed with DNA (Ogata et al., 1994), the trp repressor-operator complex (Zhang et al., 1994), and the structure of the DNA binding domain of the human testis-determining factor SRY complexed with a promoter fragment (Werner et al., 1995a).

Oligomeric proteins represent complexes between identical subunits. The first dimer to be

TABLE 1
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
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)
Intramolecular Ligand Contacts	
2D $^{12}C,^{14}N(F_1)/^{12}C,^{14}N(F_2)$-filtered NOE in H_2O[a]	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[a]	H(j)-^{12}C(j)-------H(i)-^{12}C(i)
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)

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

solved by NMR was the chemokine interleukin-8 (Clore et al., 1989, 1990). Since that time, a number of other homodimeric systems have been solved, including the Arc repressor (Breg et al., 1990; Bonvin et al., 1994), the gene 5 protein from M13 (Folkers et al., 1994); the chemokines hMIP-1β (Lodi et al., 1994), GRO/MGSA (Fairbrother et al., 1994; Kim et al., 1994), RANTES (Skelton et al., 1995), the Mnt repressor (Burgering et al., 1994), and the DNA-binding domain of HIV-1 integrase (Lodi et al., 1995, and Eijkelenboom et al., 1995). More recently, the methodology has been extended to a tetramer, namely, the oligomerization domain of the tumor suppressor p53 (Clore et al., 1994, 1995a, b; Lee et al., 1994).

III. THREE- AND FOUR-DIMENSIONAL NMR

A. Sequential 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±2), 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; Vuister et al., 1988). In some cases, it is still possible to apply the same sequential assignment strategy by making use of 3D heteronuclear (^{15}N or ^{13}C)-edited experiments to increase the spectral resolution (Marion et al., 1989; Zuiderweg and Fesik, 1989; Driscoll et al., 1990). Frequently, 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, 1990; Ikura et al., 1990; Clore and Gronenborn, 1991c; Bax and Grzesiek, 1993). The double and triple reso-

nance experiments that we currently use together with the correlations that they demonstrate are summarized in Table 2. 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 h. In most cases, however, signal-to-noise requirements necessitate 1 to 3 d of measuring time, depending on the experiment.

B. 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 restraints, involves the application of a conformational grid search of φ, ψ, χ_i space on the basis of the homonuclear $^3J_{HN\alpha}$ and $^3J_{\alpha\beta}$ coupling constants (which are related to φ and χ_1, respectively), and the intraresidue and sequential interresidue NOEs involving the NH, $C^\alpha H$, and $C^\beta H$ protons (Günter et al., 1989; Nilges et al., 1990). This information can be supplemented by the measurement of heteronuclear $^3J_{NH\beta}$ and $^3J_{COH\beta}$ couplings, which are also related to $\chi 1$ (Bax et al., 1994). Stereospecific assignment of valine methyl groups can be made on the basis of $^3J_{C\gamma CO}$, $^3J_{NC\gamma}$ couplings (Bax 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

TABLE 2
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	$C^\alpha H(i)$-$^{15}N(i)$-NH(i)	$^3J_{HN\alpha}$
	$C^\beta H(i)$-$^{15}N(i)$-NH(i)	$^3J_{HN\alpha}$ and $^3J_{\alpha\beta}$
HNHA	$C^\alpha HH(i)$-$^{15}N(i)$-NH(i)	$^3J_{HN\alpha}$
H(CA)NH	$C^\alpha H(i)$-$^{15}N(i)$-NH(i)	$^1J_{NC\alpha}$
	$C^\alpha H(i-1)$-$^{15}N(i)$-NH(i)	$^2J_{NC\alpha}$
HNCA	$^{13}C^\alpha(i)$-$^{15}N(i)$-NH(i)	$^1J_{NC\alpha}$
	$^{13}C^\alpha(i-1)$-$^{15}N(i)$-NH(i)	$^2J_{NC\alpha}$
HN(CO)CA	$^{13}C^\alpha(i-1)$-$^{15}N(i)$-NH(i)	$^1J_{NCO}$ and $^1J_{C\alpha CO}$
HNCO	$^{13}CO(i-1)$-$^{15}N(i)$-NH(i)	$^1J_{NCO}$
HCACO	$C^\alpha H(i)$-$^{13}C^\alpha(i)$-$^{13}CO(i)$	$^1J_{C\alpha CO}$
HCA(CO)N	$C^\alpha H(i)$-$^{13}C^\alpha(i)$-$^{15}N(i+1)$	$^1J_{C\alpha CO}$ and $^1J_{NCO}$
CBCA(CO)NH	$^{13}C^\beta(i-1)/^{13}C^\alpha(i-1)$-$^{15}N(i)$-NH(i)	$^1J_{C\alpha CO}$, $^1J_{NCO}$, and $^1J_{CC}$
CBCANH	$^{13}C^\beta(i)/^{13}C^\alpha(i)$-$^{15}N(i)$-NH(i)	$^1J_{NC\alpha}$ and $^1J_{CC}$
	$^{13}C^\beta(i-1)/^{13}C^\alpha(i-1)$-$^{15}N(i)$-NH(i)	$^2J_{NC\alpha}$ and $^1J_{CC}$
HBHA(CO)NH	$C^\beta H(i-1)/C^\alpha H(i-1)$-$^{15}N(i)$-NH(i)	$^1J_{C\alpha CO}$, $^1J_{NCO}$, and $^1J_{CC}$
HBHANH	$C^\beta H(i)/C^\alpha H(i)$-$^{15}N(i)$-NH(i)	$^1J_{NC\alpha}$ and $^1J_{CC}$
	$C^\beta H(i-1)/C^\alpha H(i-1)$-$^{15}N(i)$-NH(i)	$^2J_{NC\alpha}$ and $^1J_{CC}$
C(CO)NH	$^{13}Cj(i-1)$-$^{15}N(i)$-NH(i)	$^1J_{C\alpha CO}$, $^1J_{NCO}$, and $^1J_{CC}$
H(CCO)NH	$Hj(i-1)$-$^{15}N(i)$-NH(i)	$^1J_{C\alpha CO}$, $^1J_{NCO}$, and $^1J_{CC}$
HCCH-COSY[b]	Hj-^{13}Cj-$^{13}Cj^{\pm1}$-$Hj^{\pm1}$	$^1J_{CC}$
HCCH-TOCSY	Hj-^{13}Cj. . .$^{13}Cj^{\pm n}$-$Hj^{\pm n}$	$^1J_{CC}$

[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: $^3J_{HN\alpha}$, ~3 to 10 Hz; $^1J_{CC}$, ~35 Hz; $^1J_{C\alpha CO}$, ~55 Hz; $^1J_{NCO}$, ~15 Hz; $^1J_{NC\alpha}$, ~11 Hz; $^2J_{NC\alpha}$, ~7 Hz.

[b] COSY, correlation spectroscopy.

(Zuiderweg et al., 1985). Finally, stereospecific assignments of leucine methyl groups can be made on the basis of heteronuclear $^3J_{C\delta C\alpha}$ and $^3J_{C\delta H\beta}$ couplings (Bax et al., 1994) in combination with the pattern of intraresidue NOEs, provided that the stereospecific assignment of the β-methylene protons and the χ1 rotamer have been determined previously (Powers et al., 1993). A summary of the heteronuclear quantitative J correlation experiments that we currently employ is provided in Table 3.

C. Assignment of Through-Space Proton-Proton Interactions with a Protein

While the panoply of 3D heteronuclear experiments is sufficient for the purposes of spectral assignment, 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 Plate 1.* Consider a simple 2D spectrum demonstrating 11 cross peaks from aliphatic resonances to a single proton resonance position. In the 2D spectrum, it is impossible to ascertain whether this destination resonance involves one proton or many. Extending the spectrum to 3D by separating the NOE interactions according to the chemical shift of the heavy atom (^{15}N or ^{13}C) attached to each proton reveals that three individual protons are involved. The identity of the originating aliphatic protons, however, is only specified by their proton chemical shifts. Because the extent of spectral overlap in the aliphatic region of the spectrum is considerable, additional editing is necessary. This is achieved by adding a further dimension such that 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 of the originating protons. In this manner, each ^1H-^1H 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.

Because the number of NOE interactions present in each 2D plane of a 4D $^{13}C/^{15}N$- or $^{13}C/^{13}C$-edited NOE 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 a virtual lack of resonance overlap and good sensitivity can be ob-

TABLE 3
Experiments Currently Used in Our Laboratory to Determine Three-Bond Coupling Constants by Quantitative J Correlation Spectroscopy

Experiment	Three-bond coupling	Torsion angle
3D HNHA	$^3J_{HN\alpha}$	φ
2D ^{13}C-{^{15}N} -spin-echo difference CT-HSQC[a]	$^3J_{C\gamma N}$	χ1 of Thr and Val
2D ^{13}C-{^{13}CO} -spin-echo difference CT-HSQC	$^3J_{C\gamma CO}$	χ1 of Thr and Val
3D HN(CO)HB	$^3J_{COH\beta}$	χ1
3D HNHB	$^3J_{NH\beta}$	χ1
3D HACAHB	$^3J_{\alpha\beta}$	χ1
2D or 3D ^1H-detected long-range C-C COSY	$^3J_{CC}$	χ2 of Leu and Ile χ3 of Met
3D ^1H-detected [^{13}C-^1H] long-range COSY	$^3J_{CH}$	χ2 of Leu and Ile χ3 of Met

Note: Details of all experiments are provided in the review by Bax et al. (1994), with the exception of the 3D HACAHB experiment (Grzesiek et al., 1995).

[a] CT-HSQC, constant time heteronuclear single quantum coherence.

* Plates 1 through 12 appear after page 378.

tained 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., 1991a; Zuiderweg et al., 1991; Vuister et al., 1993)-edited NOE spectra should permit the assignment of almost all NOE interactions in a relatively straightforward manner. The first successful application of these methods to the structure determination of a protein greater than 15 kDa was achieved in 1991 with the determination of the solution structure of interleukin-1β, a protein of 17 kDa, and 153 residues (Clore et al., 1991b).

IV. PRECISION AND ACCURACY OF NMR STRUCTURES

The quality of an NMR protein structure determination increases as the number of restraints increases (Havel and Wüthrich, 1985; Clore and Gronenborn, 1991a, d; Havel, 1991; Clore et al., 1993a). Irrespective of the algorithm employed, any structure determination by NMR seeks to find the global minimum region of a target function F given by:

$$F = F_{cov} + F_{vdw} + F_{NMR} \quad (1)$$

where F_{cov}, F_{vdw}, and F_{NMR} are terms representing the covalent geometry, the nonbonded contacts, and the experimental NMR restraints, respectively. Systematic bias arising from the different algorithms used to calculate the structures may be introduced via the first two terms, F_{cov} and F_{vdw}, in Equation 1.

As the values of bond lengths, bond angles, planes, and chirality are known to very high accuracy, it is clear that the deviations from idealized geometry, as represented by the term F_{cov}, should be kept very small. Acceptable deviations are ≤0.006 Å for bonds, ≤2° for angles, and ≤0.5° for improper torsions (i.e., planarity and chirality). In algorithms operating in torsion angle space, the covalent geometry is fixed so that the deviations from idealized geometry are zero. In contrast, in algorithms operating in Cartesian coordinate space, it is essential to use appropriate values of the force constants to ensure that the deviations from idealized geometry fall within these ranges.

The second term, F_{vdw}, representing the nonbonded contacts, is associated with considerably more uncertainty than the covalent geometry. Given the numerous ways to represent F_{vdw} (e.g., a simple van der Waals repulsion term or a complete empirical energy function, including a van der Waals Lennard-Jones 6-12 potential, an electrostatic potential, and a hydrogen bonding potential), it is evident that variability is introduced via F_{vdw}. It is therefore essential to ensure that the calculated structures display good nonbonded contacts. This can be assessed, for example, by calculating the Lennard-Jones energy even if this has not been employed in the target function. We previously investigated the uncertainties in the representation of the van der Waals term, in particular variations in the van der Waals radii, and found that they introduce an uncertainty of 0.2 to 0.3 Å in the accuracy of the coordinates (Clore et al., 1993a).

Because uncertainties associated with idealized geometry and van der Waals representation will introduce an error of at most 0.3 Å, it follows that a major determinant of accuracy resides in the number and quality of the experimental NMR restraints that enter into the third term, F_{NMR}, in Equation 1.

While a high-resolution, carefully refined X-ray structure of a given protein may not be identical to the "true" solution structure, it is likely to be reasonably close in many instances, as evidenced, for example, by the excellent agreement (≤1 Hz rms deviation) in a number of cases between the experimentally determined values of the $^{3}J_{HN\alpha}$ coupling constants in solution and their corresponding calculated values from crystal structures (Bartik et al., 1993; Vuister and Bax, 1993; Garrett et al., 1994). It is therefore instructive to examine the dependence of the backbone rms difference between NMR and X-ray structures on the precision of the NMR structures. This is shown in Figure 2 for 14 proteins for which both NMR and X-ray structures are available and that are representative of some of the different algorithms and programs, as well as their implementations, used in NMR protein structure determination. (Note that not all algorithms or NMR structure laboratories are represented because we have only included NMR structures for which there exists a corresponding X-ray structure and that have a

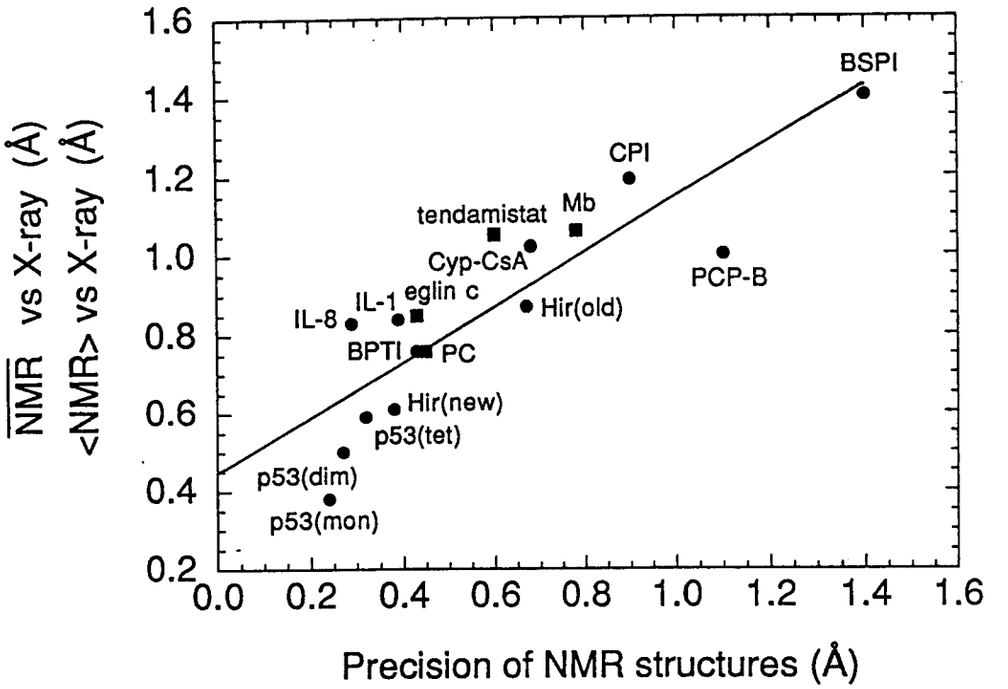


FIGURE 2. Correlation between backbone precision of NMR structures and their agreement with X-ray structures. Where the backbone rms difference between the average NMR coordinates (NMR) and the corresponding X-ray structures is available, the values are represented as circles. When only the average backbone rms difference between an ensemble of NMR structures (⟨NMR⟩) and the corresponding X-ray structure is quoted in the literature, squares are used. The straight line represents a linear fit to the data with a slope of 0.70, an intercept of 0.45 Å, and a correlation coefficient of 0.9. p53(mon), p53(dim), and p53(tet) are the monomer, dimer, and tetramer, respectively, of the p53 oligomerization domain (Clore et al., 1995b); IL-8, interleukin-8 monomer (Clore et al., 1990; Clore and Gronenborn, 1991e); Hir (new), highly refined structure of hirudin (Szyperski et al., 1992); IL-1, interleukin-1β (Clore et al., 1991b; Clore and Gronenborn, 1991b); BPTI, bovine pancreatic trypsin inhibitor (Berndt et al., 1992); eglin c (Hyberts et al., 1992); PC, French bean plastocyanin (Moore et al., 1991); tendamistat (Billeter et al., 1989); Hir (old), hirudin (Folkers et al., 1989; Clore and Gronenborn, 1991f); Cyp-CsA, cyclophilin A/cyclosporin A complex (Spitzfaden et al., 1994); Mb, carbonmonoxy myoglobin (helices plus heme; Osapay et al., 1994); CPI, potato carboxypeptidase inhibitor (Clore et al., 1987a); PCP-B, procarboxypeptidase B (Billeter et al., 1992); BSPI, barley serine proteinase inhibitor 2 (Clore et al., 1987b). The values given exclude conformationally disordered regions as described in the papers describing the structure determinations. Note that the NMR structures of IL-8 (Clore et al., 1990) and Hir(old) (Folkers et al., 1989) were obtained prior to the corresponding X-ray structures, and that the NMR structure of tendamistat was obtained independently of and at the same time as the X-ray structure.

precision of 1.5 Å or better.) An obvious linear relationship is evident. In addition, in cases where both low- and high-precision NMR structures are available for the same protein, the high-precision structure is significantly closer to the X-ray structure than the low-precision one (e.g., the two hirudin structures [Folkers et al., 1989; Szyperski et al., 1992] and the structure of barley serine proteinase inhibitor-2 and the related eglin c [Clore et al., 1987a; Hyberts et al., 1992]). The data can readily be fit to a straight line with a correlation coefficient of 0.9 and a limiting rms difference between NMR and X-ray structures of about 0.45 Å. Moreover, all the monomeric NMR structures with a precision of better than 0.5 Å are less than 0.85 Å away from the corresponding crystal structures. Given the fact that the coordinate errors in 1.5- to 2-Å-resolution X-ray structures are

about 0.2 to 0.3 Å (Luzzati et al., 1952; Clore and Gronenborn, 1991b), these data provide strong empirical evidence that an accuracy of 0.4 to 0.8 Å in the backbone coordinates can be obtained using current NMR methodology.

As noted by Zhao and Jardetzky (1994), errors in the experimental NMR restraints, which are mainly in the form of interproton distance restraints, can lead to a significant loss of accuracy. In their model calculations on crambin, they found that structures calculated from artificial data sets, comprising an average of 12 and 19 interproton distance restraints per residue and incorporating random errors in 20% of the distance restraints within a range of ±3 Å, displayed a reduction in accuracy of 0.8 and 1.1 Å, respectively. Calculations using the larger data set with either 50 or 100% of the restraints having random errors within a ±1-Å range led to reductions in accuracy of about 0.6 Å. If such errors were realistic for state-of-the-art NMR structure determinations, their conclusion that the best one could hope for was an accuracy within the 1- to 2-Å range would be correct. When appropriate care, however, is taken in the analysis of both the data and the structures generated at successive stages of the iterative refinement process, the number of errors in the experimental restraints, as well their magnitude, is considerably lower.

Errors in the interproton distance restraints can arise from two sources: (1) misassignments and (2) errors in distance estimates. Errors due to misassignments are a common occurrence in low-resolution NMR structures, and one can only urge the experimenter to exercise extreme care to avoid these. Systematic errors in distance estimates may be introduced in attempts to obtain precise distance restraints. For example, interactive relaxation matrix analysis of the NOE intensities (Borgias and James, 1990; Borgias et al., 1990; Post et al., 1990) and direct refinement against the NOE intensities (Yip and Case, 1991; Nilges et al., 1991), while accounting for spin diffusion, can result in systematic errors from several sources, such as the presence of internal motions (not only on the picosecond time scale but also on the nanosecond to millisecond time scales), insufficient time for complete relaxation back to equilibrium to occur between successive scans, and differential efficiency of magnetization transfer between protons and their attached heteronuclei in multidimensional heteronuclear NOE experiments (Clore et al., 1993a). For these reasons, it is probably prudent at the present time, at least in the case of protein structure determinations, to convert the NOE intensities into loose approximate interproton distance restraints (e.g., 1.8 to 2.7, 1.8 to 3.3, 1.8 to 5.0, and, if appropriate, 1.8 to 6.0 Å for strong, medium, weak, and very weak NOEs, respectively) with the lower bounds always given by the sum of the van der Waals radii of two protons and with additional corrections to the upper bounds in the case of methyl groups and nonstereospecifically assigned protons (Williamson et al., 1985; Clore et al., 1986). These distance ranges are sufficiently generous to take into account all untoward effects in the conversion of NOE intensities into distances (Wüthrich, 1986; Clore and Gronenborn, 1989; Clore et al., 1993a). Using this approach, the only point at which errors in the interproton distance restraints will be introduced resides at the boundary of two distance ranges.

In the case of experimental structures calculated with an incomplete set of restraints (i.e., comprising less than 90% of the structurally useful NOEs), there is no doubt that errors, arising from misassignments as well as from the incorrect classification of NOEs into the various loose approximate distance ranges, will occur, resulting in less accurate structures. This is due to the fact that until a significant degree of redundancy is present in the experimental restraints, such errors can often be readily accommodated without unduly compromising the agreement with either the experimental NMR restraints or the restraints for covalent geometry and nonbonded contacts. However, once 90% of the structurally useful NOEs have been assigned, and incorporated into the restraints set corresponding typically to an average of 15 restraints per residue with greater than 60% of the NOEs involving unique proton pairs, two sensitive and complementary techniques can be easily employed to identify and correct such errors.

The first method involves an analysis of the distribution of restraints violations in the ensemble of calculated structures. If a given restraint is systematically violated in more than, say, 20% of the calculated structures, even by as little as 0.1 Å, it is highly likely that it should either be reclassi-

fied into the next looser category (i.e., strong to medium, medium to weak) or that errors in NOE assignments are present (Clore et al., 1993a).

The second approach employs complete cross-validation in order to assess the completeness of the experimental restraints and the degree to which each distance restraint can be predicted by the remaining ones (Brünger et al., 1993). Typically, this involves calculating a series of simulated annealing calculations in which the restraints are randomly partitioned into a test set comprising about 10% of the data and a reference set. Only the reference set is incorporated into the target function, and each calculation is carried out with a different test and reference set pair, thereby permitting one to fully explore the constraining power of the experimental restraints. The average agreement with all the test sets as well as the atomic rms shift following complete cross-validation then provides a good indicator of accuracy.

Finally, a further check on the correctness of the structures is provided by verifying that all short interproton distances (e.g., less than 3.5 Å) predicted by the structures are in fact observed in the NOE spectra. Indeed, this procedure forms the basis of the iterative refinement process, as the structures at each successive stage of refinement are used to predict all short interproton distance contacts, which are then searched for in the NOE spectra (Kraulis et al., 1989; Forman-Kay et al., 1991; Clore and Gronenborn, 1991a; Robien et al., 1992). In general, the vast majority of interproton distances less than 3.5 Å, and certainly all those less than 2.5 Å, should be observed. Exceptions can occasionally occur if the linewidths of the corresponding resonances are severely broadened due to some sort of intermediate chemical exchange process on the chemical shift scale (due, for example, to multiple conformations or microheterogeneity), resulting in severe attenuation of the NOE cross peaks.

V. STRUCTURE DETERMINATION OF PROTEIN-PEPTIDE COMPLEXES

A. Structure of the Calmodulin-Target Peptide Complex

Calmodulin (CaM) is an 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 and Klee, 1988). The crystal structure of Ca^{2+}-CaM was solved a number of years ago (Babu et al., 1985). It is a dumbbell-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 to 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 midpoint, with residues 78 to 81 adopting an essentially unstructured "random coil" conformation that 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 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 to 602 of the calmodulin-binding domain of skeletal muscle myosin light-chain kinase (Ikura et al., 1992; Clore et al., 1993b). The solution structure, a ribbon diagram of which is shown in Plate 2, 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 to 5) and C (residues 147 and 148) termini of CaM, the tether connecting the two domains of CaM (residues 74 to 82), and the N (residues 1 and 2) and C (residues 22 to 26) termini 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 (that is the precision of the coordinates) for the rest of the structure (i.e., residues 6 to 73 and 83 to 146 of CaM and residues 3 to 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). Subsequent determination of the X-ray structure of CaM with a related

peptide from smooth muscle myosin light-chain kinase (Meador et al., 1992) revealed essentially an identical conformation of the complex in the crystal.

The major conformational change in Ca^{2+}-CaM that occurs after binding M13 involves an extension of the flexible tether (residues 78 to 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 two helices (residues 65 to 73 and 83 to 93), thereby enabling the two domains to come together, gripping the peptide rather like two hands holding 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 drawing shown in Plate 2, which also serves to highlight the approximate twofold pseudosymmetry 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 twofold axis. 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. 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 to 21) of M13 that interact with CaM, while the N (residues 1 and 2) and C (residues 22 to 26) termini of the peptide, which do not interact with CaM, remain disordered.

The Ca^{2+}-CaM-M13 complex is stabilized by numerous hydrophobic interactions, supplemented by some electrostatic interactions, which are summarized in Figure 3. Particularly striking are the

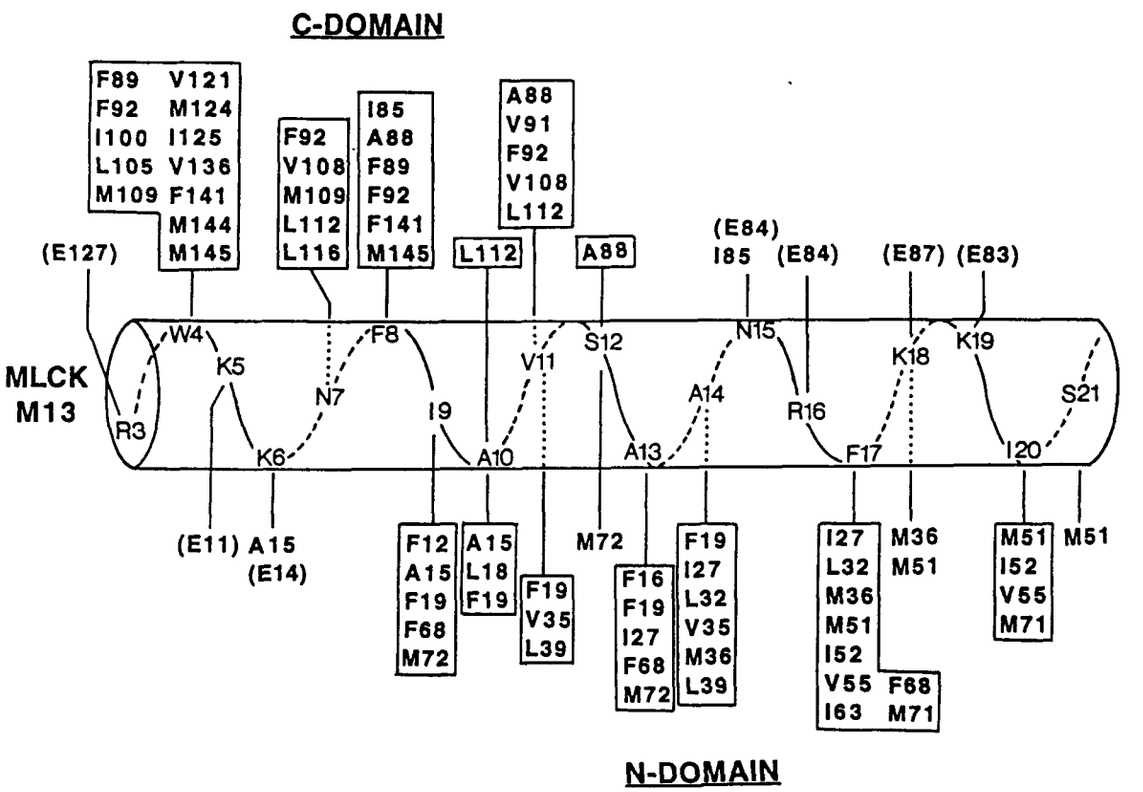


FIGURE 3. Summary of the residue pairs for which intermolecular NOEs between CaM and M13 are observed. CaM residues involved in hydrophobic interactions are boxed. Also included are potential electrostatic interactions between negatively charged Glu residues of CaM (shown in parentheses) and positively charged Lys and Arg residues of M13. (Adapted from Ikura et al., 1992.)

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- and N-terminal hydrophobic patches of CaM, respectively (Plate 2). These interactions also involve a large number of methionine residues that are unusually abundant in CaM, in 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}, and 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).

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 to 21 of M13). 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. 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.

B. Structure of Human Thioredoxin in a Mixed Disulfide Intermediate Complex with its Target Peptide from the Transcription Factor NFκB

Recently, it has been shown that human thioredoxin, a 12-kDa cellular redox protein, is responsible for activating the cellular transcription factor NFκB by reducing a disulfide bond involving Cys62 of the p50 subunit (Matthews et al., 1992; Hayashi et al., 1993; Mitomo et al., 1994). Using multidimensional heteronuclear-edited and -filtered NMR spectroscopy, we solved the solution structure of a complex of human thioredoxin and a 13-residue peptide comprising residues 56 to 68 of the p50 subunit (Qin et al., 1995). This structure represents a kinetically stable, mixed disulfide intermediate along the reaction pathway.

The proposed mechanism of disulfide bond reduction by thioredoxin involves nucleophilic attack by the highly reactive Cys32 thiolate anion (pK_a ~6.3 in hTRX; Forman-Kay et al., 1992) of the reduced species on a disulfide-containing substrate, producing a mixed disulfide intermediate, followed by attack of the thiolate of Cys35 (pK_a ~7.5 to 8.6 in hTRX; Forman-Kay et al., 1992) to yield a reduced substrate and oxidized TRX (Kallis and Holmgren, 1980). We therefore trapped a kinetically stable, mixed disulfide intermediate of hTRX and the NFκB peptide using the quadruple Cys → Ala mutant (C35A, C62A, C69A, and C73A) of hTRX, which retains as the sole cysteine the reactive Cys32. The three cysteine mutations at positions 62, 69, and 73 were previously introduced in our study of the structures of reduced and oxidized hTRX to circumvent any problems arising from intermolecular disulfide bond formation on oxidation (Qin et al., 1994). The mixed disulfide intermediate was formed by reacting the mutated hTRX with the NFκB peptide at pH 8, and the complex was purified by HPLC. A similar strategy was employed to determine the structure of a mixed disulfide intermediate of glutaredoxin and glutathione (Bushweller et al., 1994).

The structure was solved by means of multidimensional double and triple resonance heteronuclear-filtered and -edited NMR spectroscopy making use of the NFκB peptide at natural isotopic abundance (i.e., ^{12}C and ^{14}N), and uniformly (>95%) ^{15}N- and $^{15}N/^{13}C$-labeled hTRX. An example of the quality of the data is shown in Figure 4, which illustrates strips taken from the 3D ^{13}C-edited (F_2)/^{12}C(F_3)-filtered NOE spectrum that provide short interproton distance contacts specifically between ^{13}C-attached protons of hTRX and ^{12}C-attached protons of the unlabeled NFκB peptide.

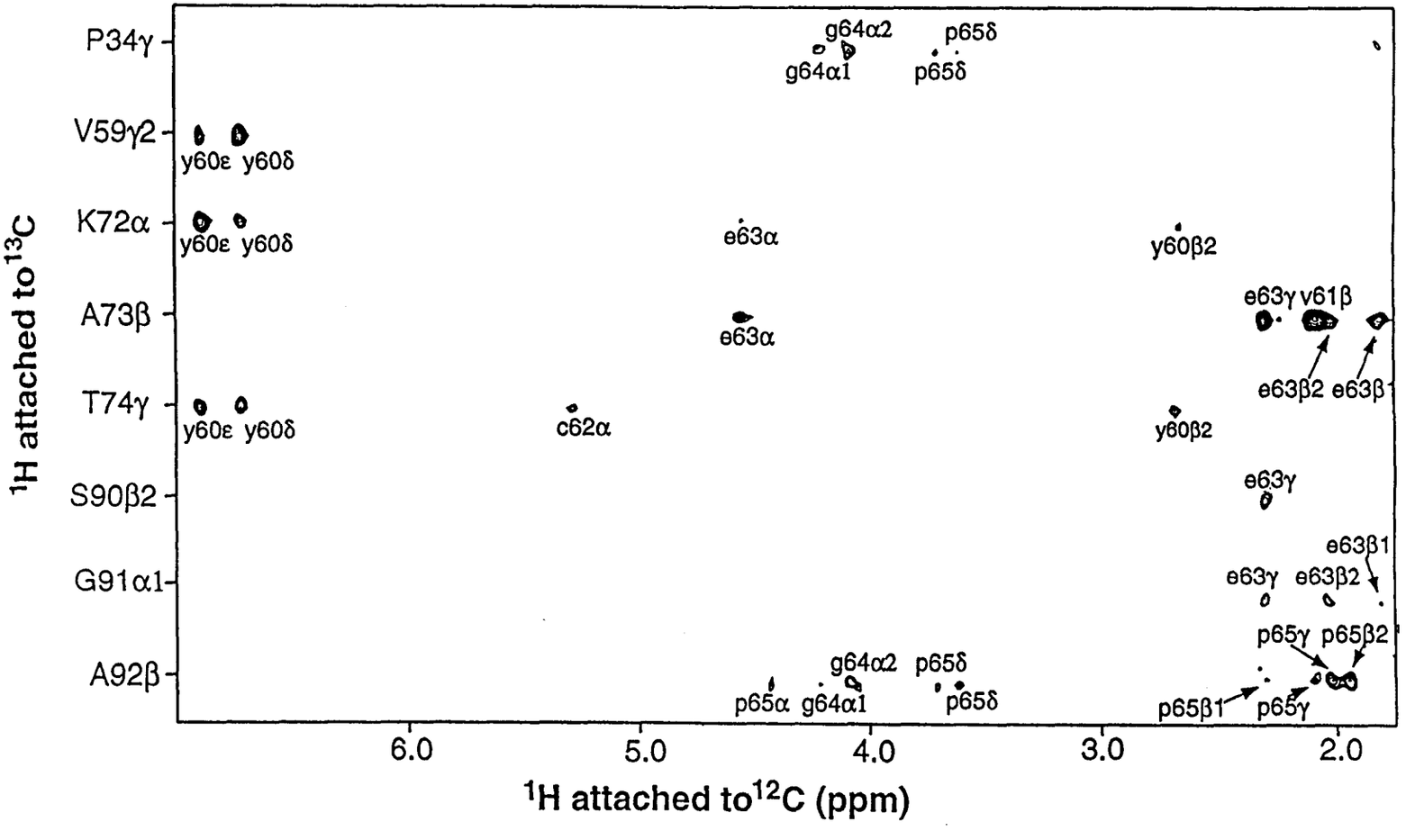


FIGURE 4. Composite of ^{13}C-H strips taken from the 120-ms mixing time, 600-MHz 3D ^{13}C(F_2)-edited/^{12}C(F_3)-filtered NOE spectrum of the hTRX-NFκB peptide complex, illustrating NOEs between protons of the unlabeled peptide (attached to ^{12}C along the F_3 axis) and protons of labeled hTRX (attached to ^{13}C). Residues of hTRX and the NFκB peptide are indicated by upper- and lower-case letters, respectively. (From Qin, J., Clore, G. M., Kennedy, W. M. P., Huth, J. R., and Gronenborn, A. M., *Structure*, 3, 289, 1995. With permission.)

The structure determination was based on 3169 experimental NMR restraints. A stereoview of a best-fit superposition of the backbone atoms and ordered side chains for the final 60 simulated annealing structures is shown in Plate 3. The restraints comprised 2582 approximate distance restraints, as follows: 2357 intramolecular hTRX interproton distance restraints (subdivided into 690 intraresidue restraints and 503 sequential ($|i-j|=1$), 437 short-range ($1 < |i-j| \leq 5$), and 727 long-range ($|i-j| > 5$) interresidue restraints), 115 intramolecular NFκB peptide interproton distance restraints (subdivided into 64 intraresidue, 38 sequential and 13 medium-range interresidue restraints), 74 intermolecular hTRX-NFκB interproton distance restraints, and 36 distance restraints for 18 hydrogen bonds in hTRX. These were supplemented by 300 torsion angle restraints (104 ϕ, 76 ψ. 78 $\chi 1$, and 31 $\chi 2$ restraints for hTRX, and 7 $\chi 1$ and 4 $\chi 2$ restraints for the NFκB peptide), 88 $^{3}J_{HN\alpha}$ coupling constant restraints, and 102 $^{13}C^{\alpha}$ and 97 $^{13}C^{\beta}$ chemical shift restraints. The N (residue 56)- and C (residue 68)-terminal residues of the NFκB peptide, which are not in contact with hTRX, are disordered. The remainder of the complex is very well defined, with a precision of 0.21 ± 0.02 Å for the backbone atoms, 0.57 ± 0.03 Å for all atoms, and 0.34 ± 0.03 Å for all ordered atoms.

The overall structure of hTRX in the complex is similar to that of oxidized and reduced hTRX (Qin et al., 1994) and is characterized by a five-stranded β-sheet (residues 1 to 5, 22 to 28, 53 to 59, 75 to 81, and 84 to 91) arranged in a –2x, +1x, –2, –1 topology and surrounded by four α-helices (residues 7 to 17, 33 to 49, 62 to 70, and 94 to 105). The backbone atomic rms difference between the mean coordinates of the hTRX-NFκB complex and oxidized hTRX, the hTRX-NFκB complex and reduced hTRX, and reduced and oxidized hTRX are 0.64, 0.83, and 0.86 Å, respectively, and the corresponding values for all atoms are 0.78, 1.0, and 1.0 Å, respectively. As the precision of the solution structures of reduced and oxidized hTRX (Qin et al., 1994) is comparable to that of the complex with the NFκB peptide, the differences between all the structures, albeit small, are significant. Thus, the overall structure of hTRX in the complex is slightly closer to that of the oxidized species than to the reduced one.

To facilitate discussion, we denote the residue numbering of the NFκB peptide as follows: S_0 represents the disulfide-bonded cysteine, and negative and positive numbers indicate residues N and C terminal to this cysteine, respectively. Residues S_{-5}(Arg57) to S_{+3}(Pro65) are in contact with hTRX, binding in a crescent-like conformation to a boot-shaped cleft on the surface of hTRX, 8 to 10 Å wide and 5 to 6 Å deep, with the shank and foot of the boot approximately 18 and 15 Å in length, respectively (Plate 4). The angle between the long axes of the boot and shank is about 110°. In the views shown in Plate 4, the borders of the cleft are formed by the N terminus of helix α2 and the C terminus of helix α4 that are located at the top of hTRX, the active site loop (residues 31 to 36) and the C terminus of strand β3 on the left-hand side, and by helix α3, the loop connecting helix α3 and strand β4, and strand β5 on the right-hand side. The S_{-5}(Arg57) to S_{-3}(Arg59) residues are located in the boot of the cleft, while the S_{-2}(Tyr60) to S_{+3}(Pro65) residues are located in the shank of the cleft. The side chain of the S_{-3}(Arg59) residue points directly into the heel of the cleft. The chain of the NFκB peptide follows the cleft such that there is a smooth 110° bend centered around the S_{-2}(Tyr60) residue. The S_{+3}(Pro65) residue abuts the uppermost point of the cleft whose boundaries force the chain of the NFκB peptide to reverse direction, resulting in a turn centered around the S_{+3}(Pro65) and S_{+4}(Ser66) residues, accompanied by a backbone carbonyl-to-amide hydrogen bond between the S_{+2}(Gly64) and S_{+5}(His67) residues. As a result, the polypeptide backbone of the NFκB peptide is defined up to the Cα carbon of the S_{+5} residue; the side chain of this residue makes no contact with hTRX and is disordered.

The disulfide bridge between Cys32 of hTRX and S_0(Cys62) has a right-handed *trans-trans*-gauche(+)-*trans*-gauche(–) conformation with a C^{α}-C^{α} separation of 6.8 Å, similar to that of the disulfides that span the β-barrels in immunoglobulins (Richardson, 1981). In addition to the covalent linkage afforded by the disulfide bridge, the complex between the NFκB peptide and hTRX is stabilized by numerous noncovalent interac-

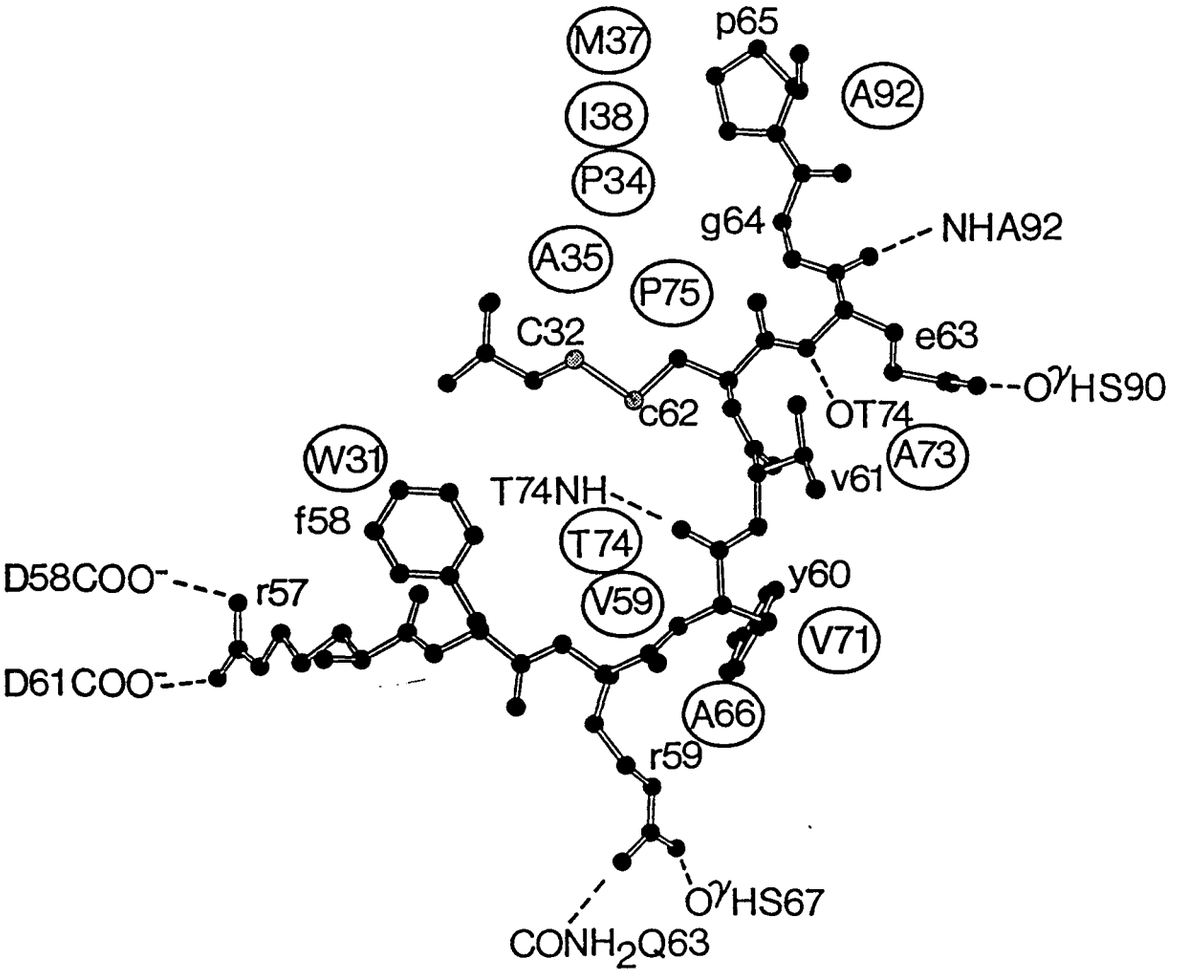


FIGURE 5. Schematic representation showing the interactions between the NFκB peptide and hTRX. Residues of hTRX involved in hydrophobic interactions with the peptide are circled, and the dashed lines indicate hydrogen bonds, salt bridges, or electrostatic interactions. Only residues 57 to 65 of the NFκB peptide, which are in contact with hTRX, are shown, and the residues of the peptide and hTRX are depicted by upper- and lower-case letters, respectively. (From Qin, J., Clore, G. M., Kennedy, W. M. P., Huth, J. R., and Gronenborn, A. M., *Structure*, 3, 289, 1995. With permission.)

tions that confer specificity to substrate binding. These are illustrated in detail in Figure 5.

There are three backbone hydrogen bonds between the NFκB peptide and hTRX in the immediate vicinity of the S_0(Cys62) residue. In particular, the backbone carbonyl of the S_{-2}(Tyr60) residue is hydrogen bonded to the backbone amide of Thr74, and the backbone amide and carbonyl of S_{+1}(Glu63) are hydrogen bonded to the backbone carbonyl of Thr74 and the backbone amide of Ala92, respectively. There are five electrostatic interactions involving three side chains of the NFκB peptide. The guanidinium group of the S_{-5}(Arg57) residue forms a salt bridge with the side chain carboxylates of Asp58 and Asp61, the guanidinium group of the S_{-3}(Arg59) residue is hydrogen bonded to the side chain carbonyl of Gln63 and the hydroxyl group of Ser67, and the carboxylate of the S_{+1}(Glu63) residue forms a hydrogen bond with the hydroxyl group of Ser90. It should be noted that, on the basis of pH titration data, the carboxylate groups of Asp58 and Asp61

are negatively charged at the pH of 4.4 employed in this study. In addition, it seems likely that the hydroxyl group of the S_{-2}(Tyr60) residue is hydrogen bonded to the backbone carbonyl of Phe27 via a bridging water molecule. Finally, there is an extensive set of hydrophobic interactions. The aromatic ring of the S_{-4}(Phe58) residue is stacked against the six-membered ring of Trp31; the aliphatic portion of the side chain of the S_{-3}(Arg59) residue is in contact with the methyl groups of Val59 and Ala66; the aromatic ring of the S_{-2}(Tyr60) residue is in contact with the methyl groups of Val59, Ala66, Val71, and Thr74; the β-methine of the S_{-1}(Val61) residue and the β- and γ-methylene groups of the S_{+1}(Glu63) residue are in contact with Ala73; the disulfide bridge comprising the S_0(Cys62) residue is in contact with the methyl group of Ala35 and the side chains of Pro34 and Pro75; the S_{+2}(Gly64) residue is in contact with Pro34; and, finally, the S_{+5}(Pro65) residue is in contact with Pro34, the Cδ methyl and Cγ methylene groups of Ile38, the Cβ and Cγ methylene groups of Met37, and the methyl group of Ala92.

Following an attack on the disulfide bridge between Cys32 and S_0(Cys62) by the Cys35 thiolate anion (which, from the position of Ala35 in the present structure, would be positioned directly on top of the disulfide bridge) and the subsequent formation of an intramolecular Cys32-Cys35 disulfide bond, it is clear that the affinity of the peptide would have to be much reduced to permit its release. We suggest that this is accomplished as follows. Formation of the intramolecular disulfide bridge will necessarily alter the positioning of the S_0(Cys62) residue to avoid steric clash. This in turn is likely to disrupt the backbone hydrogen bonds involving the S_{-1} and S_{+1} residues, thereby reducing the affinity of the peptide for oxidized hTRX. In this regard it is interesting to note that the aperture of the cleft formed by the active site and the two opposing loops (residues 74 to 76 and 90 to 92) as measured by the C^{α}-C^{α} distances between Cys/Ala35, Pro75, and Ala92 is 10 to 20% wider in the oxidized protein than in the complex. We should also note that in the intact system, NFκB itself may undergo a conformational change upon reduction that facilitates its release.

V. STRUCTURE DETERMINATION OF PROTEIN-DNA COMPLEXES

A. Structure of the Specific Complex of the Transcription Factor GATA-1 with DNA

The erythroid-specific transcription factor GATA-1 is responsible for regulation of the 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 to 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} ~1.2×10^8 M^{-1}) (Omichinski et al., 1993b). In order to understand the mechanism of specific DNA recognition by GATA-1, the solution structure of the specific complex of a 66-residue fragment (residues 158 to 223) comprising the DNA-binding domain of chicken GATA-1 (cGATA-1) with a 16-base pair oligonucleotide containing the target sequence AGATAA was determined by means of multidimensional heteronuclear-filtered and -separated NMR spectroscopy (Omichinski et al., 1993a).

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. The N (residue 1) and C (residues 60 to 66) termini of

the protein are disordered. Base pairs 6 to 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 five and last three 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 five to six steps with any great degree of precision and accuracy. The global conformation of the central eight 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 precision of the coordinates for the complex proper (i.e., residues 2 to 59 of the protein and base pairs 6 to 13 of the DNA) is 0.70 ± 0.13 and 1.13 ± 0.08 Å for protein backbone plus DNA and all protein atoms plus DNA, respectively.

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

A schematic ribbon drawing of the core is presented in Figure 6. The core starts out with a turn (residues 2 to 5), followed by two short irregular antiparallel β-sheets, a helix (residues 28 to 38), and a long loop (residues 39 to 51) that includes a helical turn (residues 44 to 47) as well as an Ω-like loop (residues 47 to 51). β-strands 1 (residues 5 to 7) and 2 (residues 11 to 14) form the first β-sheet, while β-strands 3 (residues 18 to 21) and 4 (residues 24 to 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. This was an unexpected finding given that apart from the four Cys residues that coordinate the zinc atom, only one residue (Lys36 in the cGATA-1 DNA-binding domain and Lys465 in the glucocorticoid receptor) is conserved between the two proteins.

The overall topology and structural organization of the complex is shown in Plates 5A and B. The conformation of the oligonucleotide is B-type. The helix and the loop connecting strands β2 and β3 (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 ~40° to the base planes of the DNA (Plate 5A), while the C-terminal tail is approximately parallel to the base planes (Plate 5B).

Views of side-chain contacts with the DNA in the major and minor grooves are shown in Plates 5C and D, respectively. The cGATA-1 DNA-binding domain makes specific contacts with eight bases, seven in the major groove (A6, G7, A8, T25, A24, T23, and T22) and one 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, Leu17 interacts with A6, G7, and T25, Thr16 with A24 and T25, Leu33 with A24 and T23, and Leu37 with T23 and T22. This accounts for the predominance of thymidines in the DNA target site. Indeed, there are only three hydrogen-bonding interactions: between the side chain of Asn29 and the N6 atoms of A24 and A8 in the major groove, and between the NζH3+ of Lys57 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. This latter

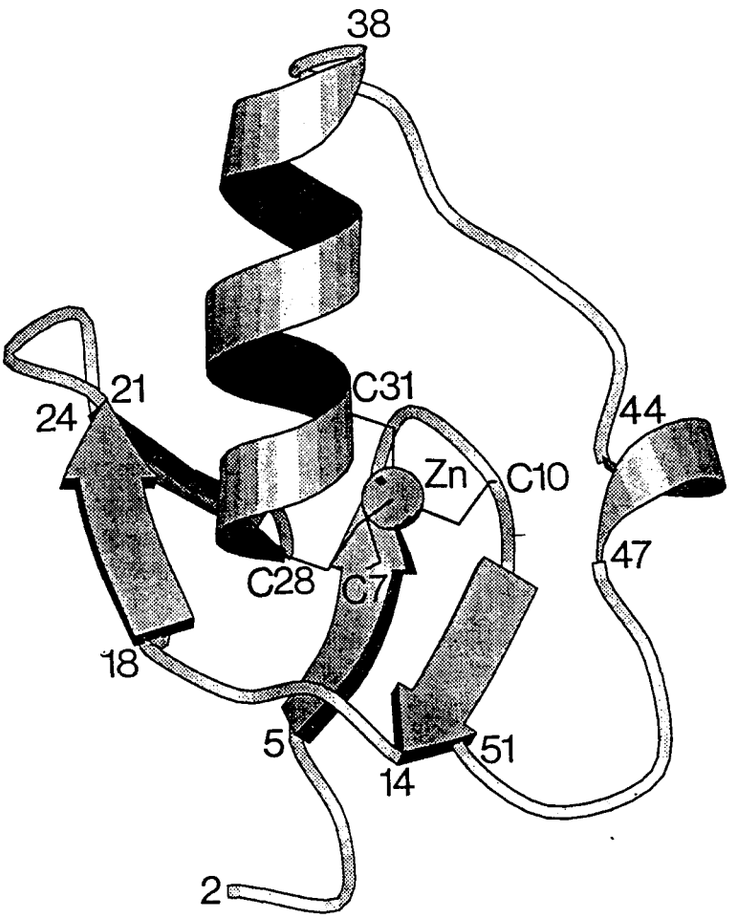


FIGURE 6. Schematic ribbon drawing of the core of the cGATA-1 DNA-binding domain. (From Omichinski, J. G., Clore, G. M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S. J., and Gronenborn, A. M., *Science,* 261, 438, 1993a. With permission.)

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 (G20 to T27). Salt bridges and/or hydrogen bonds with the phosphates of G7, A24, and T22 are made by Arg19, Arg47, and His38, respectively, in the major groove, and with the phosphates of C13, T25, C26, and T27 by Arg54, Thr53, Arg56, and Ser59, respectively, in the minor groove. The interactions of Arg54 and Arg56 above and below the polypeptide chain span the full length of the target site and are probably responsible for bending 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 Tyr34, Leu33 and Ala30, and Ile51 and Thr16, respectively. In the minor groove, hydrophobic sugar DNA-protein interactions are made by C13 with the aliphatic portion of the side chain of Arg54, T23 and T24 with Gln52, T25 and C26 with the aliphatic portion of the side chain of Arg56, and C26 with Ser59. In addition, there is a hydrogen bond between the side-chain amide of Gln52 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 been solved previously (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 (eight base pairs, of which only a contiguous stretch of six 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 like a hand gripping a rope. The structure of the cGATA-1 DNA-binding domain reveals a modular design. The fold of residues 3 to 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 to 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).

B. Structure of the Human SRY-DNA Complex

Male sex determination in mammals is directed by the genetic information encoded on the Y chromosome, leading to specialization of the embryonic gonads into testes. A key component is the protein encoded by the human testis-determining gene, SRY (sex-determining region Y), mutations in which are responsible for 15% of male-to-female sex reversal, that is, 46X,Y females (Goodfellow and Lovell-Badge, 1993; Gustafson and Donahoe, 1994; Haqq et al., 1994). The SRY protein is a transcriptional activator of the Müllerian inhibiting substance (MIS) gene, whose product is responsible for the regression of the female Müllerian ducts (the precursor of the uterus, fallopian tubes, and upper vagina) in male embryos.

The human SRY (hSRY) gene codes for a protein of 203 residues that comprises three domains: an N-terminal domain, a central 77-residue DNA-binding domain consisting of a single HMG (high-mobility group) box, and a C-terminal domain (Sinclair et al., 1990). The HMG box is an approximately 80-residue domain that mediates the minor groove DNA binding of a large family of eukaryotic proteins known as the HMG-1/2 family (Laudet et al., 1993). In order to establish the atomic and molecular basis of SRY-dependent 46X,Y sex reversal, we have determined the solution structure of a specific complex of the DNA-binding domain of hSRY (hereafter referred to as hSRY-HMG) with a DNA octamer comprising its specific target site in the MIS promoter (Werner et al., 1995a).

The structure calculations, using simulated annealing, were based on a total of 1805 experimental NMR restraints, including 75 intermolecular contacts, some of which are illustrated in the spectra shown in Figure 7. A best-fit superposition of the final 35 simulated annealing structures is shown in Plate 6. Residues 1 to 3 and 75 to 78 at the N- and C-termini, respectively, are disordered in solution. Excluding these residues, the precision of the coordinates is 0.59 ± 0.05 Å for the protein backbone (N, Cα, C, and O atoms) plus the DNA, and 0.96 ± 0.06 Å for all protein atoms plus DNA.

The overall topology of hSRY-HMG is best described as a twisted letter L or boomerang with the long and short arms approximately 28 and 22 Å in length, respectively (Plate 7A). The structure is composed of irregular N (residues 3 to 9)- and C (residues 71 to 74)-terminal strands that lie directly opposite each other, although they are not hydrogen bonded, and three helices (residues 10

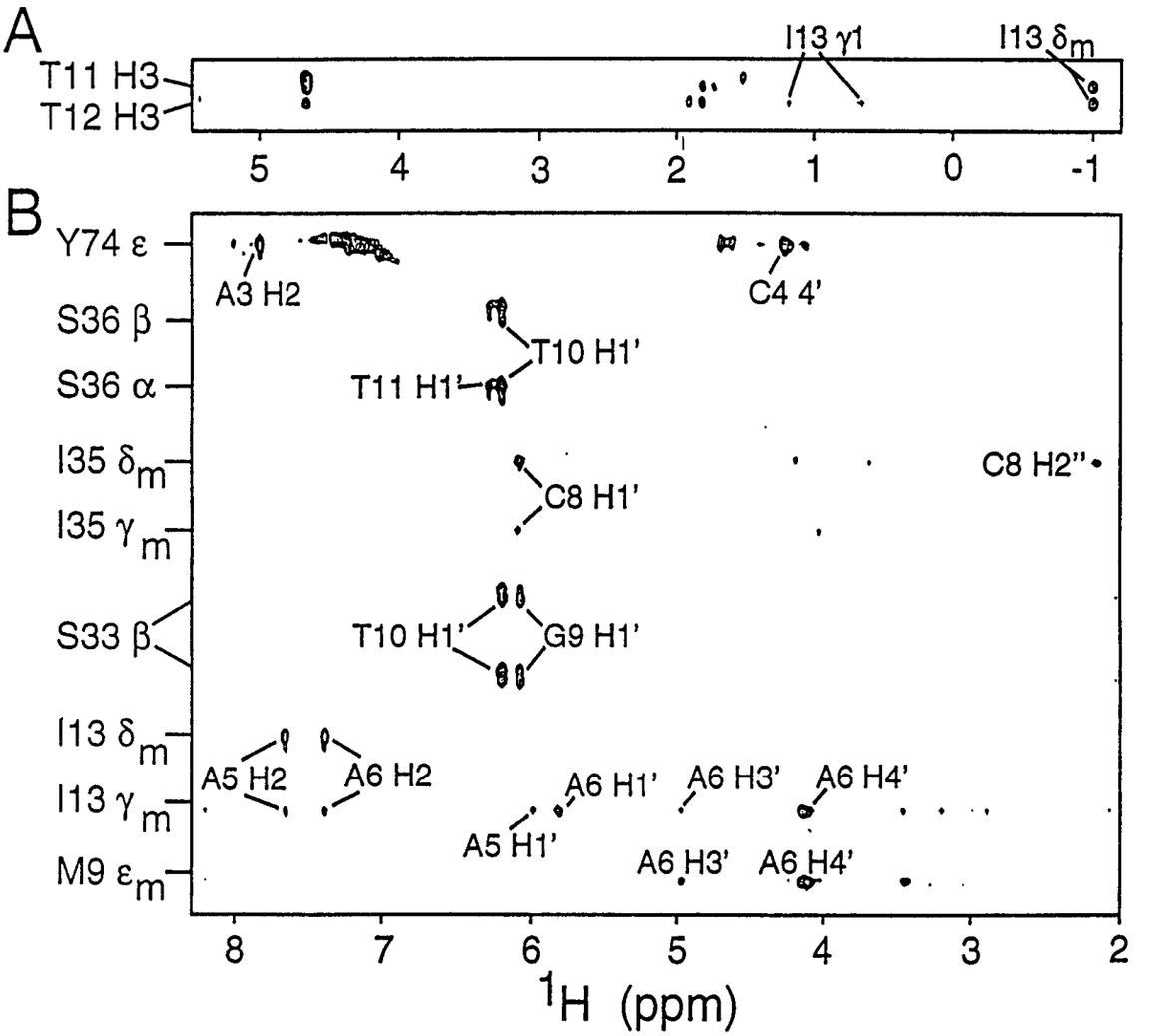


FIGURE 7. Experimental observation of contacts between hSRY-HMG and DNA. (A) Portion of the 150-ms 2D 1-1 NOE spectrum illustrating NOEs between imino protons of the DNA and protein protons (the unlabeled peaks arise from NOEs within the DNA). (B) A composite of ^{13}C-^{1}H strips taken from the 155-ms 3D $^{13}C(F_2)$-separated/$^{12}C(F_3)$-filtered NOE spectrum, illustrating NOES between protons of the protein (attached to ^{13}C) and protons of the DNA (attached to ^{12}C). (From Werner, M. H., Bianchi, M. E., Gronenborn, A. M., and Clore, G. M., *Biochemistry*, 34, 11998, 1995. With permission.)

to 26, 33 to 45, and 47 to 68). Helices 1 and 2 are connected by a six-residue loop, helices 2 and 3 by a very short turn. Finally, helix 3 leads into the C-terminal strand via a turn centered around Pro70 and Asn71. Helix 1 has a slightly bent appearance, due to a distortion in the backbone at Asp 18. Helix 3 exhibits an approximately 50° kink at the junction of Pro[53] and Phe[54]. The long arm of the L is formed by helix 3 and the N-terminal strand (residues 3 to 9), while the short arm of the L is formed by helices 1 and 2. Helix 1 is oriented at ~120° to helix 2 and ~130 to 140° to helix 3. Helices 2 and 3 are approximately orthogonal to each other. The orientation of the long and short arms is maintained by the packing of several side chains located on both helix 1 and helix 3. The top of the long arm is held in place by a hydrophobic cluster between the N and C termini.

The overall structure of hSRY-HMG is similar to that of the previously determined NMR

structures of two nonspecific DNA-binding domains of the HMG-1/2 family, the rat HMG-1 B box and the drosophila HMG-D box (Weir et al., 1993; Read et al., 1993; Jones et al., 1994), both of which have 20 to 25% sequence identity with hSRY-HMG. The regions corresponding to the N- and C-terminal strands of hSRY-HMG, however, are disordered in both HMG-1 B and HMG-D. This is probably due to the replacement of Val5 in hSRY-HMG by Pro in HMG-1 B and HMG-D, resulting in substantial structural instability at the N and C termini.

After binding to hSRY-HMG, the DNA undergoes a profound structural change from B-type DNA in the free state to an underwound form that is distinct from either B- or A-type DNA (Plates 7 and 8). The sugar pucker appears primarily A-like, although the general appearance of the DNA is intermediate between the B and A type. The DNA helix is severely underwound, with an average local interbase-pair helical twist of 26 ± 6°, ranging from a low value of ~19° between base pairs 1 and 2 and base pairs 7 and 8, to a high value of ~34° between base pairs 4 and 5. As a result, the minor groove is shallow and significantly expanded, with a width of 9.4 ± 0.6 Å (Plate 8A) compared with 4.0 Å in B-DNA and 6.2 Å in A-DNA (Saenger, 1984). Concomitantly, the major groove is substantially compressed (Plate 8B). The sugar puckers range from O1′-endo to C3′-endo for 14 of 16 deoxyribose rings typical of A-DNA, with the exception of A5 and G9, which have C1′-exo to C2′-endo conformations, typical of B-DNA. The local interbase-pair rise has an average value of 4.1 ± 0.3 Å, which is larger than that for either B-DNA (3.6 Å) or A-DNA (3.4 Å). The local interbase-pair slide has an average of –0.21 ± 0.35 Å, which is comparable to that of B-DNA (–0.76 Å) but much smaller than that for A-DNA (–2.06 Å). Finally, the sequential phosphate-phosphate distances have an average value of 6.3 ± 0.4 Å, which is closer to that of B-DNA (6.5 Å) than A-DNA (5.6 Å). This distance, however, is significantly reduced to a value of 5.4 Å between base pairs 5 and 6, which is the site of partial intercalation by Ile13.

The DNA in the complex is bent by ~70 to 80° in the direction of the major groove (Plate 8). This is principally accomplished through large, positive local interbase-pair roll angles for six of the seven base steps (ranging in value from 10.5° between base pairs 7 and 8, up to 35° between base pairs 2 and 3), in conjunction with the helical unwinding described above. This contrasts with values of 10.8 and –3.6° for the local interbase-pair roll angles in classic A- and B-DNA, respectively (Saenger, 1984).

Comparison with the structure of the DNA in the complex with TATA-binding protein (TBP; Kim et al., 1993a, b), which also binds in the minor groove, reveals numerous features in common. These include the A-like sugar puckers except at the sites of intercalation where they are B-like, the large positive roll angles, the helix unwinding and concomitant widening of the minor groove, the increase in local helix rise relative to A- and B-DNA, and the B-like local slide. There are some interesting differences, however. For example, the helix is ~20% more underwound and the roll angles are on average ~70% larger in the TBP complex than in the hSRY-HMG complex, while the minor groove is ~10% wider in the hSRY-HMG complex than in the TBP complex.

The general topology and structural organization of the complex is depicted in Plates 7 and 8. The DNA is located in the concave surface of the L-shaped hSRY-HMG domain, and binding occurs exclusively in the minor groove of the DNA. In the view shown in Plates 7C and D, it can be seen that the binding surface is formed by helix 1 bounded at the bottom by a ridge comprising helix 2 and at the top by a ridge comprising the N- and C-terminal strands. In contrast, the recognition of the minor groove by TBP is provided by the concave surface of a ten-stranded antiparallel β-sheet (Kim et al., 1993a, b). Thus, the structural scaffold of the DNA-binding surface is entirely different in hSRY-HMG and TBP. The conformation of the DNA, which is severely distorted from that of classic B-DNA, follows the contours of the concave binding surface perfectly (Plate 7). The orientation of the DNA with respect to the protein can be determined unambiguously on the basis of a qualitative interpretation of a few intermolecular NOEs: in particular, the NOEs from Ser33, Ile35, and Ser36 to the sugars of C8, G9, and T10; the NOEs from the side chain of Ile13 to the imino protons of T11 and T12, the H2 protons of A5 and A6, and the sugar protons of A5 and A6; and the NOEs from the aromatic ring

of Tyr74 to the H2 proton of A3 and the sugar of C4 (Figure 7).

The two principal areas of bending occur between base pairs 5 and 6 as a result of the partial intercalation of Ile13, and between base pairs 2 and 3, as the DNA is pushed away from the body of the protein by the ridge formed by Lys73 and Tyr74 (Plates 7C and D). Widening of the minor groove appears to be mediated by five residues that form a T-shaped wedge in direct contact with the central base pairs of the DNA octamer (Plate 8C). Specifically, Met9, Phe12, Ile13, and Trp43 form a hydrophobic wedge across base pairs 5 and 6, anchored to base pairs 4 and 5 by electrostatic interactions involving Asn10. The residues constituting the central portion of the wedge (Phe12 and Ile13) and the stem of the T (Asn10) bind to the DNA bases, while the residues at the wings of the wedge (Met9 and Trp43) bind to the DNA sugar-phosphate backbone and pry open the minor groove (Plate 7C). This T-shaped wedge on the surface of hSRY-HMG can be considered to directly induce and stabilize the helix unwinding of the DNA. In addition, hSRY-HMG is anchored at the two ends of the DNA by Tyr74 at base pair 3, and Ser33, Ile35, and Ser36 at base pairs 7 and 8 (Plate 8C).

A comparison of the sequences from a variety of specific and nonspecific HMG domains (Laudet et al., 1993) indicates that, in general, residues that interact with the sugar-phosphate backbone tend to be either conserved or substituted conservatively (Figure 8). The distinction between the two subclasses is found in residues that contact the bases and in certain core-packing amino acids.

In the case of the specific DNA-binding domains, the packing of Val5, Tyr69, and Tyr72 at the N and C termini appears to be critical for the correct presentation of the DNA-binding surface made up of the side chains of Arg4, Lys73, and Tyr74 (Plate 9C). The sequence-specific HMG domains have either Val or Ile at position 5. Substitution by the longer Leu side chain in hSRY leads to a substantial reduction in specific DNA binding affinity (Harley et al., 1992). In the nonspecific HMG domains, the shorter Pro side chain is found predominantly at position 5. Thus, both a shorter and a longer side chain at this position either destabilizes the hydrophobic packing in this region of the molecule or distorts the surface topology of the binding surface at the N and C termini sufficiently to perturb specific DNA binding. The importance of the N and C termini is further supported by the results on a chimeric HMG domain composed of the N and C termini of TCF-1 (a specific HMG domain) and helices 1,2, and a portion of helix 3 of HMG1-B (a nonspecific HMG domain) that retained the ability to both bend and bind DNA specifically (Read et al., 1994). Similar arguments can be made for the stabilization of the twisted L-shape by the hydrophobic core between the stem and base of the HMG domain. A key residue in this regard is Phe55 in hSRY-HMG, which is either retained or substituted conservatively by a Tyr or Val in the specific HMG domains. In the nonspecific domains, on the other hand, Phe55 is most often substituted by a Glu, perhaps destabilizing the hydrophobic core and thereby leading to less efficient unwinding of the DNA target. Indeed, the twist between the stem and base of the L reinforces the helical unwinding induced by the partial intercalation of Ile13 by presenting the proper complementary surface to the DNA. Hence, destabilization of the hydrophobic core could, in part, disrupt the surface complementarity, leading to a more weakly bound protein and a concomitant loss of DNA sequence discrimination.

In addition, the nonspecific HMG domains lack a number of critical residues that make base-specific contacts in the structure of the hSRY-HMG complex. Asn10 is conserved in all the specific HMG domains and is involved in hydrogen bonding or electrostatic interactions with the bases of C4, A5, and G13 (Plate 9B). In the nonspecific HMG domains, Asn10 is substituted by Ser, Thr, or Arg. The side chains of Ser and Thr would be too short to make appropriate contact, while the longer Arg side chain is likely to introduce steric clash. A similar argument can be made for Ser36, which hydrogen bonds with the base of T10. In the specific HMG domains, position 36 is occupied by Ser or Asn. In the nonspecific HMG domains, position 36 tends to be occupied by Ala, Val, or Lys. The impact of this change, however, is likely to be weaker than that for Asn10, as a few nonspecific domains retain Ser or Thr at this position.

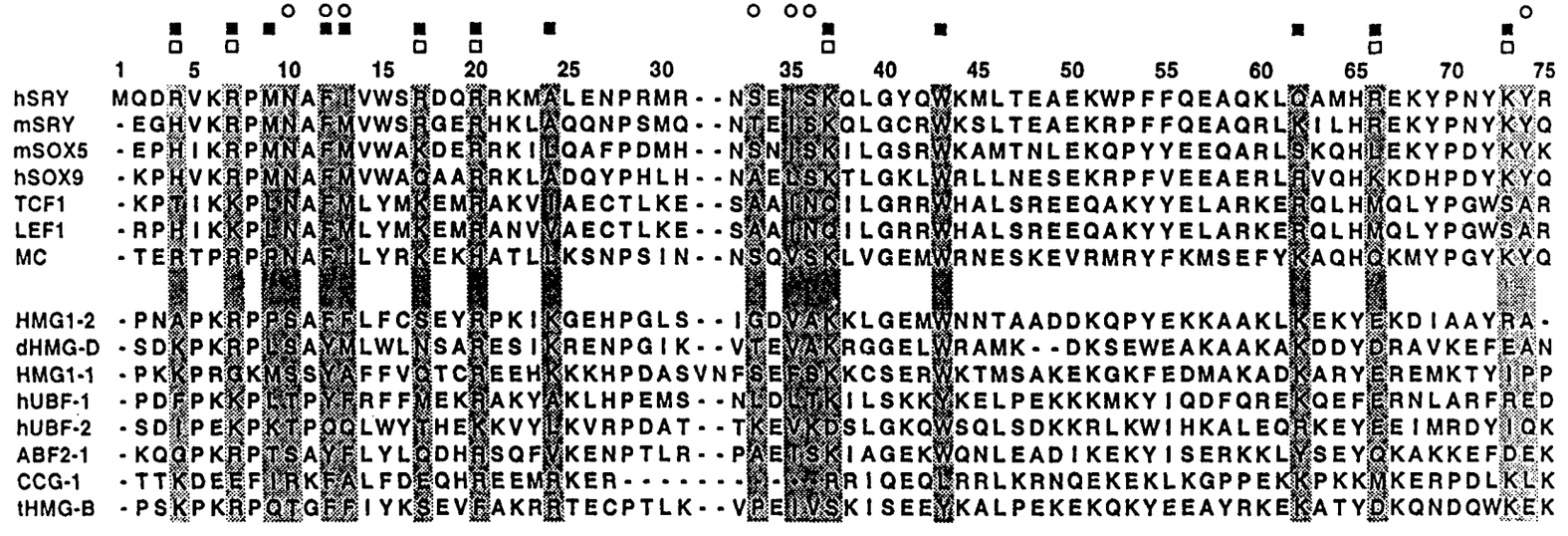


FIGURE 8. Alignment of the hSRY-HMG sequence with other specific and nonspecific DNA-binding HMG domains. Residues in hSRY that contact the bases, sugars, and phosphates are indicated by open circles, solid squares, and open squares, respectively, at the top of the figure. All residues that make contact with DNA in hSRY-HMG are shaded in each sequence. The first set of sequences (hSRY, mouse SRY [mSRY], murine SRY Box 5 [mSOX5], human SRY BOX 9 [hSOX9], human T cell factor 1 [TCF1], human lymphoid enhancer-binding factor 1 [LEF1], and the yeast mating-type gene product [MC]) are representative of specific DNA binders. The second set of sequences (human high-mobility-group protein 1 box 2 [HMG1-2], drosophila HMG-d [dHMG-D], human high-mobility-group protein 1 box 1 [HMG1-1], human upstream-binding factors 1 and 2 [hUBF-1, hUBF-2], yeast ARS-binding factor 2 box 1 [ABF2-1], product of the human cell cycle gene 1 [CCG-1], and tetrahymena HMG-B [tHMG-B]) display little or no sequence specificity.

The structure of the complex of hSRY-HMG with DNA also provides an explanation for the different sequence specificities of the SOX subfamily (which includes hSRY, mSRY, and the SOX cluster of HMG domains) vs. the LEF/TCF-1 subfamily. From a functional viewpoint, the former comprise proteins that are involved in sex differentiation, whereas the latter consists of transcription factors found specifically in lymphocytes. The DNA target sites for the SOX subfamily have a strong preference for A at the position of base 3, while those for the LEF/TCF-1 family display a strong preference for T at bases 2 and 3 (Giese et al., 1992). The main amino acid sequence difference between the two subfamilies is located at positions 73 and 74, which are Lys-Tyr in the SOX subfamily and Ser-Ala in the LEF/TCF-1 subfamily (Figure 8). In the hSRY-HMG complex, Lys73 forms a phosphate contact to C16, while Tyr74 is packed orthogonally to the bases of A3 and T14 and may form a hydrogen bond with the O2 atom of T14 (Plate 9C). The latter is precluded when T14 is replaced by A in the LEF/TCF-1 target site. Conversely, substitution of Tyr74 by Ala in the LEF/TCF-1 subfamily removes a bulky hydrophobic group capable of extensive interactions with the bases. The substitution of Lys73 by a Ser in LEF/TCF-1, however, permits an alternative hydrogen bond between the hydroxyl group of Ser73 and the O2 atom of T2 (or C2 in the MIS promoter sequence). This interpretation is reinforced by the observation that LEF-1 and TCF-1 can bind with significant affinity to SRY sequences, while the reverse is not true (Giese et al., 1992).

The amino acid sequences of SRY derived from different organisms diverge considerably outside the HMG domain (Goodfellow and Lovell-Badge, 1993). Further, all clinical mutations in hSRY resulting in phenotypic 46X,Y sex reversal, with the exception of only a single nonsense mutation, have been found to occur exclusively in the DNA-binding domain. These observations have led to the hypothesis that the primary influence of SRY on transcriptional regulation, and hence on gonadal differentiation, is largely a consequence of the structural effects induced by the protein at specific promoter targets *in vivo*. Indeed, SRY is best considered as an architectural control protein in transcriptional regulation (Tijan and Maniatis, 1994). An SRY-induced DNA bend could bring two proteins (such as a transcriptional enhancer and RNA polymerase) located at either side of the SRY-binding site into close contact, thereby permitting the assembly of a highly specific nucleoprotein transcription complex leading to the activation of transcription. In other cases, SRY could separate two proteins of the transcription complex by insertion of an unwanted bend, thereby acting as a repressor. The structure of the complex of hSRY-HMG with its target site in the MIS promoter therefore provides a framework for explaining the effects of clinical mutations at the atomic and molecular levels.

From a genetic standpoint, naturally occurring point mutations in hSRY can be classified into two types: inherited mutations with variable penetrance, and *de novo* mutations with full penetrance that are not shared between fathers and their 46X,Y daughters (Goodfellow and Lovell-Badge, 1993). Clearly, the functional effects of the former must be considerably less severe than those of the latter. From a structural viewpoint, the point mutations also fall into two categories: those that affect the packing of residues within the protein core and those that involve residues that directly contact the DNA.

Three inherited point mutations have been identified to date: Val5 → Leu, Phe54 → Ser, and Ile35 → Met. The first two mutations are best described as packing-defect mutants, while the third affects both side-chain packing within the protein and a protein-DNA contact (Plate 9A). The main effect of these mutations may be to accelerate the degradation of SRY within the cell, thereby providing a simple explanation for their variable penetrance.

There are five other packing-defect point mutants, four of which are *de novo* (Met23 → Thr, Leu46 → His, Ala58 → Thr, and Tyr72 → Cys), and the fourth (Lys51 → Ile) is of unknown origin because relatives are unavailable.

The remaining four *de novo* point mutations identified to date all involve residues that contact the DNA. The Arg7 → Gly mutant removes hydrophobic contacts with the sugar of C4 and a salt bridge to the phosphate of A5 (Plate 9B). Met9 makes extensive van der Waals contacts with the deoxyribose of A5 and A6 and with the O3′ atom of A6, which is located approximately at the hinge

point of the bend between base pairs 5 and 6 (Plate 9B). The introduction of a shorter Ile side chain at this position would disrupt these contacts. The Gly40 → Arg mutation is likely to introduce steric clash at two sites, thereby destabilizing the complex: first with the sugar-phosphate backbone of T12, and second with Trp43, thereby displacing a critical residue of the wedge that drives the helical unwinding of the DNA (Plate 8C). The conservation of Gly at position 40 in the sequence-specific HMG domains (Plate 8) supports the notion that the local packing in this region of the interface is critical for the formation of a stable complex. Finally, the Ile13 → Thr mutation would be expected to have one of the most profound effects on DNA binding of all the mutants, as the short, polar Thr would not be able to partially intercalate between base pairs 5 and 6 (Plate 9B), thereby removing one of the principal determinants of bending and helical unwinding in the complex. Indeed, the affinity of this mutant for DNA is reduced by almost two orders of magnitude relative to that of the wild type, and the mutant complex has a lifetime of less than 5 ms compared with greater than 200 ms for the wild-type complex (Haqq et al., 1994).

VI. STRUCTURE DETERMINATION OF OLIGOMERIC PROTEINS

Oligomeric proteins comprising identical subunits represent a special case of protein-protein complexes. In the case of homodimeric proteins, it is usually possible to sort out the intermolecular contacts in a relatively straightforward manner without recourse to experiments involving heterodimers, formed by a 1:1 mixture of uniformly $^{13}C/^{15}N$-labeled and unlabeled (natural isotopic abundance) subunits, provided the orientation at the dimer interface is antiparallel. Under these circumstances, there will inevitably be numerous NOEs that correspond to distances greater than, say, 8 Å in the monomer, which therefore must arise from intersubunit contacts. By making use of an iterative refinement procedure, it is then possible to ascertain most of the intersubunit contacts. Of course, there will be some NOEs that can have contributions from both intra- and intersubunit interactions, and these can be dealt with by making use of interproton distance restraints represented as $\langle \Sigma r^{-6} \rangle^{-1/6}$ sums to represent these NOEs (Nilges, 1993). This approach was first demonstrated successfully in the case of interleukin-8 (Clore et al., 1989, 1990). If the orientation of the two subunits at the interface, however, is parallel, or if the topology of the dimer is highly complex with intertwining of the polypeptide chains, it is absolutely essential to make use of mixed labeled and unlabeled subunits in order to unambiguously discriminate between intra- and intersubunit interactions (Burgering et al., 1993; Folkers et al., 1993). The same applies to higher-order multimers such as trimers and tetramers. In these cases, however, the problem is much more complex, as it is necessary not only to separate intra- from intersubunit interactions but also to distinguish between the various intersubunit pairwise combinations. Thus, for a tetramer, for example, it is essential to identify interactions within a dimer and between dimers. The latter cannot be done *a priori* and requires an iterative approach combined, where appropriate, with the use of $\langle \Sigma r^{-6} \rangle^{-1/6}$ sums to represent the ambiguous NOE restraints (Clore et al., 1994).

A. Structure of the Oligomerization Domain of the Tumor Suppressor p53

The tumor suppressor p53 has been found to be involved in approximately 50% of all human cancers (Harris, 1993). It is composed of an N-terminal transactivation domain, a central core DNA-binding domain, an oligomerization domain, and a C-terminal basic regulatory domain. The structure of the core domain complexed to DNA has been solved crystallographically (Cho et al., 1994). Similarly, the structure of the oligomerization domain has been solved by both NMR (Clore et al., 1994, 1995a, b; Lee et al., 1994) and X-ray crystallography (Jeffrey et al., 1995). All the structures of the oligomerization domain have the same overall topology comprising a dimer of dimers, oriented in an approximately orthogonal manner. Each dimer (comprising subunits A and C, and subunits B and D, using the original notation employed by Clore et al., 1994) consists of two antiparallel α-helices and an antiparallel β-sheet.

In the tetramer, the antiparallel β-sheets lie on opposing faces of the molecule, and the helices form an unusual four-helix bundle.

Because of the complexity of the system and the necessity to distinguish intra- from intersubunit NOEs, the structure determination was carried out on samples of uniformly labeled ($^{15}N/^{13}C$) protein and on samples of mixed heterotetramers comprising equal amounts of unlabeled and isotopically labeled protein, and involved the application of double- and triple-resonance NMR spectroscopy (Figure 9). The tetramer is completely symmetric, as evidenced by the observation of only a single set of resonances for each residue. Structural information in the form of interproton distance restraints was derived from isotope-edited and -filtered 3D and 4D NOE spectra (Clore et al., 1994, 1995a, b).

The most highly refined structures are based on 4472 experimental NMR restraints that comprise approximate interproton distance, hydrogen bonding, torsion angle, $^3J_{HN\alpha}$ coupling constant, and secondary $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shift restraints (Clore et al., 1995b). In addition, there are distance restraints involving four symmetrically related bound water molecules (i.e., one water molecule per subunit).

A best-fit superposition of the backbone atoms and ordered side chains is shown in Plate 10. The first few residues at the N terminus (residues 319 to 323) and the last C-terminal ones (residues 357 to 360) are disordered in solution, as evidenced by the observation of only intraresidue and sequential NOEs for these residues and by low-order parameters derived from ^{15}N relaxation measurements (Clubb et al., 1995). The remainder of the tetramer (comprising residues 324 to 356 of each subunit) is very well defined, with a precision of 0.38 ± 0.09 Å for the backbone atoms, 0.84 ± 0.06 Å for all atoms, and 0.41 ± 0.08 Å for all atoms that do not exhibit conformational disorder. The backbone atomic rms difference for residues 326 to 356 between the most refined NMR structure and the X-ray structure is 0.59 Å for the tetramer, 0.50 Å for the AC dimer, and 0.38 Å for the monomer.

Ribbon diagrams illustrating five different views of the solution structure of the p53 oligomerization domain are shown in Plate 11. The structure is a dimer of dimers. Each monomer is composed of a turn (Asp324-Gly325), a β-strand (residues Glu326-Arg333), a turn (Gly334), and an α-helix (Arg335-Gly356).

The contacts within the primary AC dimer (composed of an antiparallel β-sheet and two antiparallel α-helices oriented at an angle of ~155°) are extensive: 1397 Å^2 of accessible surface area per monomer (involving 22 residues per monomer) are buried after dimerization of subunits A and C. This interaction yields a solvation free energy of dimerization (Eisenberg and McLaghlan, 1986) of –14.7 kcal/mol per monomer. The extent of the contacts between subunits A and C is illustrated by the stereoview shown in Plate 12A. The β-sheet is stabilized by eight backbone hydrogen bonds. A hydrophobic core is formed by the packing of the β-sheet with the overlying α-helices and involves Phe328, Leu330, Ile332, Phe338, Phe341, and Arg342. Hydrophobic helix-helix contacts extend the hydrophobic core and comprise the interactions of Met340(A) with Leu348(C), Phe341(A) with Leu344(C) and Leu348(C), Leu344(A) with Phe341(C) and Leu344(C), and Leu348(A) with Arg337(C) and Phe341(C). In addition, there are a number of side-chain hydrogen-bonding interactions between subunits A and C. In particular, there is a salt bridge between the carboxylate of Asp352(A) and guanidinium group of Arg337(C), and a water-bridged hydrogen-bonding cluster involving the backbone amide of Arg333(A), the side-chain carboxyamide of Asn345(C), and the carboxylate of Glu349(C). This water molecule was identified by NOE/ROEs from water to the backbone amide of Arg333, the side-chain amide of Asn345, and the $C\gamma mH_3$ methyl group of Ile332. Independent evidence for these electrostatic interactions is afforded by pH titration studies. In particular, the chemical shifts of the side-chain Cγ atom of Asp352 and the NεH of Arg337 show complex titration behavior, with one pK_a <3, and the chemical shifts of the backbone amide group of Arg333, the side-chain amino group of Asn345, and the Cδ atom of Glu349 titrate with a pK_a ~3.6. The pK_a of the carboxylate of Asp352 is substantially lower than that for a free Asp (pK_a ~3.7 to 3.9), indicating a strong salt bridge interaction with its positively charged partner Arg337.

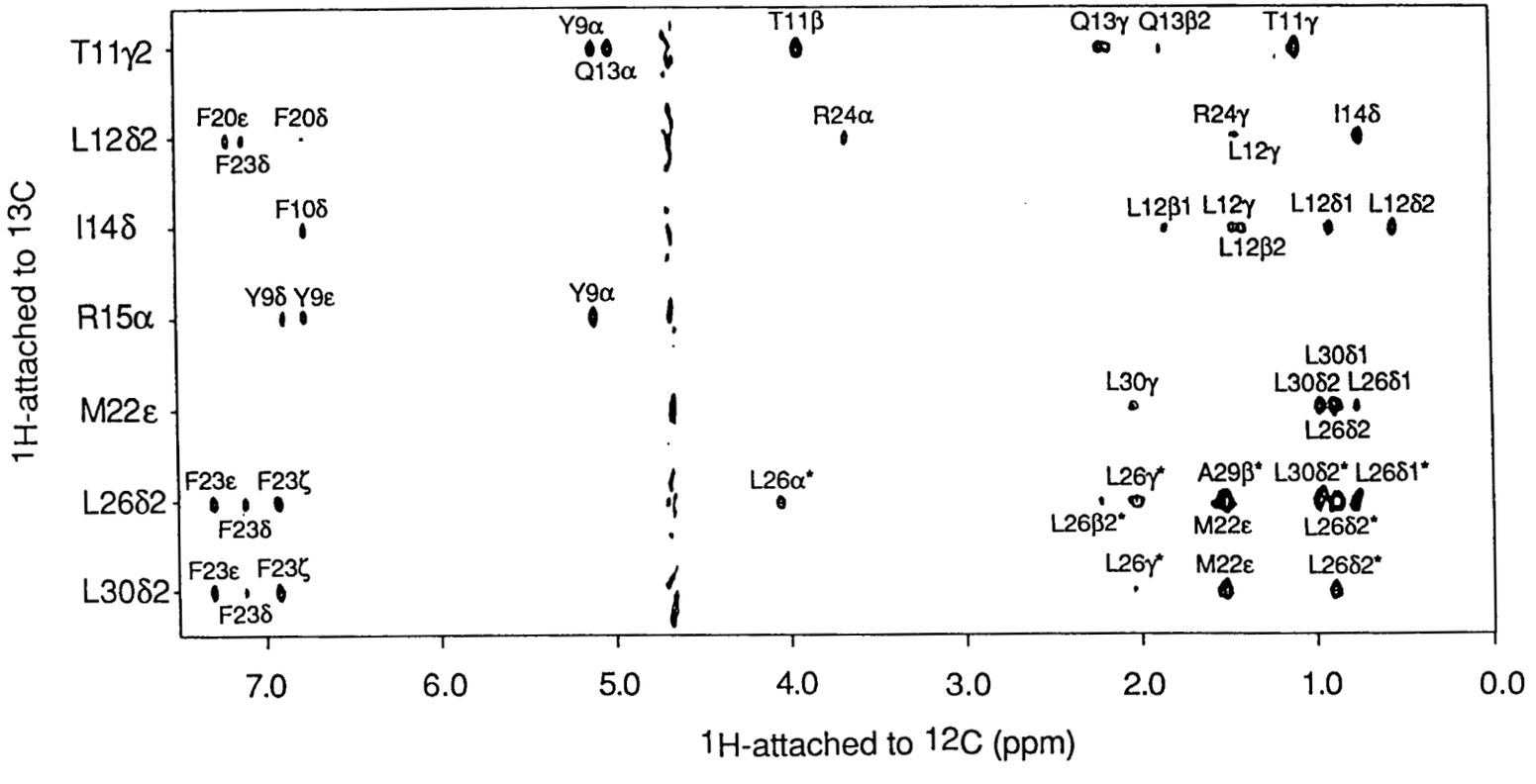


FIGURE 9. Composite of ^{13}C-^{1}H strips taken from the 120-ms mixing time 3D $^{13}C(F_2)$-edited/$^{12}C(F_3)$-filtered NOE spectrum of the tetrameric oligomerization domain of p53 comprising a 1:1 mixture of unlabeled and $^{13}C/^{15}N$-labeled polypeptide, illustrating NOEs between protons of the unlabeled polypeptide (attached to ^{12}C along the F_3 axis) and protons of the labeled peptide (attached to ^{13}C). The NOEs annotated with asterisks involve interactions between subunits A and B (and the symmetrically equivalent subunits C and D), while the remainder involve interactions between subunits A and C (and the symmetrically equivalent subunits B and D). (From Clore, G. M., Omichinski, J. G., Sakaguchi, K., Zambrano, M., Sakamoto, H., Appella, E., and Gronenborn, A. M., *Science*, 265, 386, 1994. With permission.)

While the accessible surface area buried per monomer (572 Å^2) after tetramerization is about 2.5 times less than that buried after dimerization (1397 Å^2), the total accessible surface area buried after tetramerization (2290 Å^2) (i.e., after the association of the two primary dimers) is only slightly less than that buried after dimer formation (2793 Å^2). A total of ten residues, per monomer, are buried after tetramerization, six due to AB contacts alone, two due to AD contacts alone, and two due to both AB and AD contacts. The solvation free energy of tetramerization per monomer is –5.3 kcal/mol, which can be partitioned into –3.35 kcal/mol arising from the interaction of subunits A and B, and –1.95 kcal/mol arising from the interaction of subunits A and D. Thus, the total solvation free energy of tetramerization is –21.2 kcal/mol compared with a value of –29.3 kcal/mol for the total solvation free energy of dimerization. The total solvation free energy of folding for the tetramer is ~–140 kcal/mol, which is what one would expect for a protein of this size (Chiche et al., 1990).

Views of the AB and AD interfaces are shown in Plates 12B and C, respectively. The A and B helices are oriented in an approximately orthogonal manner at an angle of ~81°. Hydrophobic contacts occur between Leu344(A), Leu344(B), and Leu348(B), between Ala347(A), Ala347(B), Leu348(B), and the hydrophobic portion of Lys351(B), and between Leu350(A), Leu350(B), and the hydrophobic portion of Lys351(B) (Plate 12B). There is one intersubunit hydrogen bond between the $N\zeta H_3^+$ of Lys351(A) and the backbone carbonyl of Glu343(B) that can be identified unambiguously from the structures (Plate 12B). Helices A and D are also oriented approximately orthogonally but with an interhelical angle of ~–104°. The hydrophobic contacts between subunits A and D are entirely hydrophobic and involve Met340, Phe341, Leu344, and the hydrophobic portion of the side chain of Glu343 (Plate 12C).

The structure of the oligomerization domain of p53 provides a framework with which to interpret the effects of mutations within this domain that are associated with human cancers (Cariello et al., 1994). There are four point mutations, Leu330 → His, Gly334 → Val, Arg337 → Cys, and Glu349 → Asp, whose effects can be clearly explained in terms of the structure and destabilization of the primary dimer. The consequences of the other four point mutations, Asp324 → Glu, Gly325 → Val, Gln354 → Arg, and Gly356 → Trp, are not immediately apparent from the structure, and indeed they may simply represent incidental findings.

The Leu330 → His mutation substitutes a completely buried hydrophobic residue by a polar residue in the center of the hydrophobic core of the primary dimer (Plate 12A). The effects of the Gly334 → Val mutation can be understood in terms of the unusual conformation of Gly334, which has ϕ,ψ angles of 86°, 140°. A Val residue cannot accommodate a positive ϕ angle in the absence of severe steric clash, and any change in the ϕ angle at this position, together with any compensatory changes in the backbone torsion angles of the adjacent residue, will distort the angle between the strand and the helix and the consequent packing of the helices in the AC dimer. The Arg337 → Cys mutation removes a salt bridge between the guanidinium group of Arg337(A) and the carboxylate of Asp352(C) (Plate 10A). Finally, the Glu349 → Asp mutation would probably remove a water-bridged hydrogen-bonding interaction between the carboxylate of Glu349(A) and the backbone amide of Arg333(C) (Plate 10A), as the Asp side chain is likely to be too short to permit this hydrogen bond to occur.

VIII. PERSPECTIVES AND CONCLUDING REMARKS

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 20- to 30-kDa range are now amenable to detailed structural analysis in solution, and current methods can probably be extended to systems even up to 40 kDa provided they are well behaved from an NMR standpoint. The examples from our laboratory discussed in the previous sections provide a perspective on the power and breadth of current NMR methodology and its ability to shed light on complex structural problems in biological systems. Examples from other

laboratories are equally instructive and are given in the reference section.

Despite these advances, it should always be borne in mind that a number of key requirements must be satisfied to permit a successful structure determination of larger proteins and protein complexes by NMR: (1) the protein in hand must be soluble and should not aggregate up to concentrations of about 0.5 to 1 m*M*, (2) it must be stable at room temperature or slightly higher for considerable periods of time (particularly as it may take several months of measurement time to acquire all the necessary NMR data), (3) it should not exhibit significant conformational heterogeneity that could result in extensive line broadening, and, finally, (4) it must be amenable to uniform ^{15}N and ^{13}C labeling. At the present time, there are still only relatively few examples in the literature of proteins in the 15- to 25-kDa range that have been solved by NMR. Likewise, only a handful of protein complexes (with DNA or peptides) and oligomers have been determined to date using these methods. One can anticipate, however, that over the next few years, by the widespread use of multidimensional heteronuclear NMR experiments coupled with semiautomated assignment procedures, many more NMR structures of proteins and protein complexes will become available.

ACKNOWLEDGMENTS

We thank our numerous collaborators over the years who have participated in various aspects of the work described in this review, in particular 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. The work in the authors' laboratories 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, 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. E., 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.

Bartik, K., Dobson, C. M., and Redfield, C. 1993. ^1H-NMR analysis of turkey egg-white lysozyme and comparison with hen egg-white lysozyme. *Eur. J. Biochem.* **215:** 255–266.

Bax, A. and Pochapsky, S. S. 1992. Optimized recording of heteronuclear multi-dimensional NMR spectra using pulse field gradients. *J. Magn. Reson.* **99:** 638–643.

Bax, A. and Grzesiek, S. 1993. Methodological advances in protein NMR. *Acc. Chem. Res.* **26:** 131–138.

Bax, A., Vuister, G. W., Grzesiek, S., Delaglio, F., Wang, A. C., Tschudin, R. and Zhu, G. 1994. Measurement of homo- and heteronuclear J couplings from quantitative J correlation. *Methods Enzymol.* **239:** 79–105.

Berndt, K. D., Güntert, P., Orbons, L. P. M., and Wüthrich, K. 1992. Determination of a high-quality nuclear magnetic resonance solution structure of the bovine pancreatic trypsin inhibitor and comparison with three crystal structures. *J. Mol. Biol.* **227:** 757–775.

Billeter, M., Kline, A. D., Braun, W., Huber, R., and Wüthrich, K. 1989. Comparison of the high resolution structures of the α-amylase inhibitor tendamistat determined by nuclear magnetic resonance in solution and by X-ray diffraction in single crystals. *J. Mol. Biol.* **206:** 677–687.

Billeter, M., Vendrell, J., Wider, G., Aviles, F. X., Coll, M., Guasch, A., Huber, R., and Wüthrich, K. 1992. Comparison of the NMR solution structure with the X-ray crystal structure of the activation domain from procarboxypeptidase B. *J. Biomolec. NMR* **2:** 1–10.

Billeter, M., Qian, Y. Q., Otting, G., Müller, M., Gehring, W., and Wüthrich, K. 1993. Determination of the nuclear magnetic resonance solution structure of an Antennapedia homeodomain-DNA complex. *J. Mol. Biol.* **234:** 1084–1097.

Blumenthal, D. K. and Krebs, E. G. 1987. Preparation and properties of the calmodulin binding domain of skeletal muscle myosin light chain kinase. *Methods Enzymol.* **139:** 115–126.

Bonvin, A. M. J. J., Vis, H., Breg, J. N., Burgering, M. J. M., Boelens, R., and Kaptein, R. 1994. Nuclear magnetic resonance solution structure of the Arc repressor using relaxation matrix calculations. *J. Mol. Biol.* **236:** 328–341.

Borgias, B. A. and James, T. L. 1990. 1990. MARDIGRAS: a procedure for matrix analysis of relaxation for discerning geometry of aqueous structures. *J. Magn. Reson.* **87:** 475–487.

Borgias, B. A., Gochin, M., Kerwood, D. J., and James, T. L. 1990. Relaxation matrix analysis of 2D NMR data. *Prog. Nucl. Magn. Reson. Spectros.* **22:** 83–100.

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

Breg, J. N., van Opheusden, J. H. J., Burgering, M. J. M., Boelens, R., and Kaptein, R. 1990. Structure of arc repressor in solution: evidence for a family of β-sheet DNA-binding proteins. *Nature* **346:** 586–589.

Brünger, A. T., Clore, G. M., Gronenborn, A. M., Saffrich, R., and Nilges, M. 1993. Assessing the quality of solution nuclear magnetic resonance structures by complete cross-validation. *Science* **261:** 328–331.

Burgering, M. J. M., Boelens, R., Caffrey, M., Breg, J. N., and Kaptein, R. 1993. Observation of intersubunit nuclear Overhauser effects in a dimeric protein: application to the Arc repressor. *FEBS Lett.* **330:** 105–109.

Burgering, M. J. M., Boelens, R., Gilbert, D. E., Breg, J., Knight, K. L., Sauer, R. T., and Kaptein, R. 1994. Solution structure of the dimeric Mnt repressor (1-76). *Biochemistry* **33:** 15036–15045.

Bushweller, J. H., Billeter, M., Holmgren, A., and Wüthrich, K. 1994. The nuclear magnetic resonance solution structure of the mixed disulfide between *Escherichia coli* glutaredoxin (C14S) and glutathione. *J. Mol. Biol.* **235:** 1585–1597.

Cariello, N. F., Cui, L., Beroud, C., and Soussi, T. Database and software for the analysis of mutations in the human p53 gene. *Cancer Res.* **54:** 4545–4460.

Chiche, L., Gregoret, L. M., Coehn, F. E., and Kollman, P. A. 1990. Protein model structure evaluation using solvation free energy of folding. *Proc. Natl. Acad. Sci. U. S. A.* **87:** 3240–3243.

Cho, Y., Gorina, S., Jeffrey, P. D., and Pavletich, N. P. 1994. Crystal structure of a p53 tumor-suppressor-DNA complex: understanding tumorogenic mutations. *Science* **265:** 346–355.

Chuprina, V. P., Rullman, J. A. C., Lamerichs, R. M. N. J., van Boom, J. H., Boelens, R., and Kaptein, R. 1993. Structure of the complex of lac repressor headpiece and an 11 base-pair half-operator determined by nuclear magnetic resonance spectroscopy and restrained molecular dynamics. *J. Mol. Biol.* **234:** 446–462.

Clore, G. M. and Gronenborn, A. M. 1987. Determination of three-dimensional structures of proteins in solution by nuclear magnetic resonance spectroscopy. *Protein 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. Mol. Biol.* **24:** 479–564.

Clore, G. M., Appella, E., Yamada, M., Matsushima, K., and Gronenborn, A. M. 1989. Determination of the secondary structure of interleukin-8 by nuclear magnetic resonance spectroscopy. *J. Biol. Chem.* **264:** 18907–18911.

Clore, G. M., Appella, E., Yamada, M., Matsushima, K., and Gronenborn, A. M. 1990. The three-dimensional structure of interleukin-8 in solution. *Biochemistry* **29:** 1689–1696.

Clore, G. M. and Gronenborn, A. M. 1991a. Structures of larger proteins in solution: three- and four-dimensional heteronuclear NMR spectroscopy. *Science* **252:** 1390–1399.

Clore, G. M. and Gronenborn, A. M. 1991b. Comparison of the solution nuclear magnetic resonance and X-ray crystal structures of human recombinant interleukin-1β. *J. Mol. 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. *Annu. Rev. Biophys. Biophys. Chem.* **20:** 29–63.

Clore, G. M. and Gronenborn, A. M. 1991e. Comparison of the solution nuclear magnetic resonance and crystal structures of interleukin-8: possible implications for the mechanism of receptor binding. *J. Mol. Biol.* **217:** 611–620.

Clore, G. M. and Gronenborn, A. M. 1991f. Comparison of the NMR and x-ray structures of hirudin, in *Computational Aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance Spectroscopy,* (ed. Hoch, J. C.), pp. 57–66, NATO ASI Series, Plenum Press, New York.

Clore, G. M. and Gronenborn, A. M. 1994. Structures of larger proteins, protein-ligand and protein-DNA complexes by multidimensional heteronuclear NMR. *Protein Sci.* **3:** 372–390.

Clore, G. M., Nilges, M., Sukumaran, D. K., Brünger, A. T., Karplus, M., and Gronenborn, A. M. 1986. The three-dimensional structure of α1-purothionin in solution: combined use of nuclear magnetic resonance, distance geometry and restrained molecular dynamics. *EMBO J.* **5:** 2729–2735.

Clore, G. M., Gronenborn, A. M., Nilges, M., and Ryan, C. A. 1987a. The three-dimensional structure of potato carboxypeptidase inhibitor in solution: a study using nuclear magnetic resonance, distance geometry and restrained molecular dynamics. *Biochemistry* **26:** 8012–8023.

Clore, G. M., Gronenborn, A. M., Kjaer, M., and Poulsen, F. M. 1987b. The determination of the three-dimensional structure of barley serine proteinase inhibitor 2 by nuclear magnetic resonance, distance geometry and restrained molecular dynamics. *Protein Eng.* **1:** 305–311.

Clore, G. M., Kay, L. E., Bax, A., and Gronenborn, A. M. 1991a. Four-dimensional $^{13}C/^{13}C$-edited nuclear Overhauser enhancement spectroscopy of a protein in solution: application to interleukin-1β. *Biochemistry* **30:** 12–18.

Clore, G. M., Wingfield, P. T., and Gronenborn, A. M. 1991b. 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., Robien, M. A., and Gronenborn, A. M. 1993a. Exploring the limits of precision and accuracy of protein structures determined by nuclear magnetic resonance spectroscopy. *J. Mol. Biol.* **231:** 82–102.

Clore, G. M., Bax, A., Ikura, M., and Gronenborn, A. M. 1993b. Structure of calmodulin-target peptide complexes. *Curr. Opin. Struct. Biol.* **3:** 838–845.

Clore, G. M., Omichinski, J. G., Sakaguchi, K., Zambrano, N., Sakamoto, H., Appella, E., and Gronenborn, A. M. 1994. High-resolution solution structure of the oligomerization domain of p53 by multidimensional NMR. *Science* **265:** 386–391.

Clore, G. M., Omichinski, J. G., Sakaguchi, Z., Zambrano, N., Sakamoto, H., Appella, E., and Gronenborn, A. M. 1995a. Interhelical angles in the solution structure of the oligomerization domain of the tumour suppressor p53. *Science* **267:** 1515–1516.

Clore, G. M., Ernst, J., Clubb, R. T., Omichinski, J. G., Sakaguchi, K., Appella, E., and Gronenborn, A. M. 1995b. Refined solution structure of the oligomerization domain of the tumour suppressor p53. *Nature Struct. Biol.* **2:** 321–332.

Clubb, R. T., Omichinski, J. G., Sakaguchi, K., Appella, E., Gronenborn, A. M., and Clore, G. M. 1995. Backbone dynamics of the oligomerization domain of p53 determined from ^{15}N relaxation measurements. *Protein Sci.* **3:** 855–862.

Cohen, P. and Klee, C. B. 1988. *Molecular Aspects of Cellular Recognition.* Vol. 5, Elsevier, New York.

Driscoll, P. C., Clore, G. M., Marion, D., Wingfield, P. T., and Gronenborn, A. M. 1990. Complete resonance assignment for the polypeptide backbone of interleukin-1β using three-dimensional heteronuclear NMR spectroscopy. *Biochemistry* **29:** 3542–3556.

Eijkelenboom, A. P. A. M., Puras Lutzke, R. A., Boelens, R., Plasterk, R. H. A., Kaptein, R., and Hård, K. 1995. The DNA-binding domain of HIV-1 integrase has an SH3-like fold. *Nature Structural Biology* **2:** 807–810.

Eisenberg, D. and McLaghlan, A. D. 1986. Solvation energy in protein folding and binding. *Nature* **319:** 199–203.

Ernst, R. 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-DNA recognition. *Nature* **366:** 483–487.

Fairbrother, W. J., Reilly, D., Colby, T. J., Hesselgesser, J., and Horuk, R. 1994. The solution structure of melanoma growth stimulating activity. *J. Mol. Biol.* **242:** 252–270.

Feng, S., Chen, J. K., Yu, H., Simon, J. A., and Schreiber, S. L. 1994. Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions. *Science* **266:** 1241–1245.

Fesik, S. W., Gampe, R. T., Eaton, H. L., Gemmecker, G., Olejniczak, E. T., Neri, P., Holzman, T. F., Egan, D. A., Edalji, R., Simmer, R., Helfrich, R., Hochlowski, J., and Jackson, M. 1991. NMR studies of [U-^{13}C]cyclosporin A bound to cyclophilin: bound conformation and portions of cyclosporin involved in binding. *Biochemistry* **30:** 6754–6583.

Folkers, P. J. M., Clore, G. M., Driscoll, P. C., Dodt, J., Köhler, S., and Gronenborn, A. M. 1989. The solution structure of recombinant hirudin and the Lys47 → Glu mutant: a nuclear magnetic and hybrid distance geometry-dynamical simulated annealing study. *Biochemistry* **28:** 2601–2617.

Folkers, P. J. M., Folmer, R. H. A., Konings, R. N. H., and Hilbers, C. W. 1993. Overcoming the ambiguity problem encountered in the analysis of nuclear Overhauser magnetic resonance spectra of symmetric dimer proteins. *J. Am. Chem. Soc.* **115:** 3798–3799.

Folkers, P. J. M., Nilges, M., Folmer, R. H. A., Konnings, R. N. H., and Hilbers, C. W. 1994. The solution structure of the Tyr41 → His mutant of the single-stranded DNA binding protein encoded by Gene V of the filamentous bacteriophage M13. *J. Mol. Biol.* **236:** 229–246.

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.

Forman-Kay, J. D., Clore, G. M., and Gronenborn, A. M. 1992. Relationship between electrostatics and redox function in human thioredoxin: characterization of pH titration shifts using two-dimensional homo- and heteronuclear NMR. *Biochemistry* **31:** 3443–3452.

Giese, K., Cox, J., and Grosschedl, R. 1992. The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. *Cell* **69:** 185–195.

Garrett, D. S., Kuszewski, J., Hancock, T. J., Lodi, P. J., Vuister, G. W., Gronenborn, A. M., and Clore, G. M. 1994. The impact of direct refinement against three-bond HN-CαH coupling constants on protein structure determination by NMR. *J. Magn. Reson. Series B* **104:** 99–103.

Goodfellow, P. N. and Lovell-Badge, R. 1993. SRY and sex determination in mammals. *Annu. Rev. Genet.* **27:** 71–92.

Goudreau, N., Cornille, F., Duchesne, M., Tocque, B., Garbay, C., and Roques, B. P. 1994. NMR structure of the N-terminal SH3 domain of GRB2 and its complex with a proline-rich peptide from Sos. *Nature Struct. Biol.* **1:** 898–907.

Grzesiek, S., Kuboniwa, H., Hinck, A. P., and Bax, A. 1995. Multiple-quantum line narrowing for measurement of H^{α}-H^{β} J couplings in isotopically enriched proteins. *J. Am. Chem. Soc.* **117:** 5312–5315.

Guerini, D. and Klee, C. B. 1991. Structural diversity of calcineurin Ca^{2+}-calmodulin stimulated phosphatases. *Adv. Protein Phosphatases* **6:** 391–410.

Günter, 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.

Gustafson, M. L. and Donahoe, P. K. 1994. Male sex determination: current concepts of male sexual differentiation. *Annu. Rev. Med.* **45:** 505–524.

Haqq, C. M., King, C.-Y., Ukiyama, E., Falsafi, S., Haqq, T. N., Donahoe, P. K., and Weiss, M. A. 1994. Molecular basis of mammalian sexual determination: activation of Müllerian inhibiting substance gene expression by SRY. *Science* **266:** 1494–1500.

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. Natl. Acad. Sci. U. S. A.* **88:** 3005–3008.

Harley, V. R., Jackson, D. I., Hextall, P. J., Hawkins, J. R., Berkovitz, G. D., Sockanathan, S., Lovell-Badge, R., and Goodfellow, P. N. 1992. DNA binding activity of recombinant SRY from normal males and XY females. *Science* **255:** 453–456.

Harris, C. C. 1993. p53: at the crossroads of molecular carcinogenesis and risk assessment. *Science* **262:** 1980–1981.

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. Mol. 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 conformations in solution. *J. Mol. Biol.* **182:** 281–294.

Hayashi, T., Ueno, Y., and Okamoto, T. 1993. Oxidoreductive regulation of nuclear factor κB. *J. Biol. Chem.* **268:** 11380–11388.

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.

Hyberts, S. G., Goldberg, M. S., Havel, T. F., and Wagner, G. 1992. The solution structure of eglin c based on measurements of many NOEs and coupling constants and its comparison with X-ray structures. *Protein Sci.* **1:** 736–751.

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:** 2344–2440.

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

Ikura, M., Clore, G. M., Gronenborn, A. M., Zhu, G., Klee, C. B., and Bax, A. 1992. Solution structure of the calmodulin-target peptide complex by multi-dimensional NMR. *Science* **256:** 632–638.

Jeffrey, P. D., Gorina, S., and Paveletich, N. P. 1995. Crystal structure of the tetramerization domain of the p53 tumour suppressor at 1.7 Å. *Science* **267:** 1498–1502.

Jones, D. M. N., Searles, M. A., Shaw, G. L., Churchill, M. E. A., Ner, S. S., Keeler, J., Travers, A. A., and Neuhaus, D. 1994. The solution structure and dynamics of the DNA binding domain of HMG-D from *Drosophila melanogaster. Structure* **2:** 609–627.

Kallis, G. B. and Holmgren, A. 1980. Differential reactivity of the functional sulfhydryl groups of cysteine-32 and cysteine-35 present in the reduced form of thioredoxin from *Escherichia coli. J. Biol. Chem.* **255:** 10261–10265.

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.

Kim, Y., Geiger, J. H., Hahn, S., and Sigler, P. D. 1993a. Crystal structure of yeast TBP/TATA-box complex. *Nature* **365:** 512–520.

Kim, J. L., Nikolov, D. B., and Burley, S. K. 1993b. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365:** 520–527.

Kim, K.-S., Clark-Lewis, I., and Sykes, B. D. 1994. Solution structure of GRO/Melanoma growth stimulatory activity determined by ^{1}H NMR spectroscopy. *J. Biol. Chem.* **269:** 32909–32915.

Kraulis, P. J., Clore, G. M., Nilges, M., Jones, T. A., Petterson, G., Knowles, J., and Gronenborn, A. M. 1989. Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei:* a study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. *Biochemistry* **28:** 7241–7257.

Kuszewski, J., Qin, J., Gronenborn, A. M., and Clore, G. M. 1995. the impact of direct refinement against $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts on protein structure determination by NMR. *J. Magn. Reson. Series B* **106:** 92–96.

Kuszewski, J., Gronenborn, A. M., and Clore, G. M. 1995. The impact of direct refinement against proton chemical shifts in protein structure determination by NMR. *J. Magn. Reson. Series B* **107:** 293–297.

Laudet, V., Stehelin, D., and Clevers, H. 1993. Ancestry and diversity of the HMG box superfamily. *Nucl. Acids Res.* **21:** 2493–2501.

Lee, W., Harvey, T. S., Yin, Y., Yau, P., Litchfield, D., and Arrowsmith, C. H. 1994. Solution structure of the tetrameric minimum transforming domain of p53. *Nature Struct. Biol.* **1:** 877–890.

Lodi, P. J., Garrett, D. S., Kuszewski, J., Tsang, M. L.-S., Weatherbee, J. A., Leonard, W. J., Gronenborn, A. M., and Clore, G. M. 1994. High-resolution solution structure of the β chemokine hMIP-1β by multidimensional NMR. *Science* **263:** 1762–1767.

Lodi, P. J., Ernst, J. A., Kuszewski, J., Hickman, A. B., Engelman, A., Craigie, R., Clore, G. M., and Gronenborn, A. M. 1995. Solution structure of the DNA binding domain of HIV-1 integrase. *Biochemistry,* **34:** 9826–9833.

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.

Mamorstein, 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:** 1886–1889.

Matthews, J. R., Wakasugi, N., Virelizier, J. L., Yodoi, J., and Hay, R. T. 1992. Thioredoxin regulates the DNA binding activity of NFκB by reduction of a disulfide bond involving cysteine 62. *Nucleic Acids Res.* **20:** 3821–3830.

Meador, W. E., Means, A. R., and Qiocho, F. A. 1992. Target enzyme recognition by calmodulin: 2.4 Å structure of a calmodulin-peptide complex. *Science* **257:** 1251–1255.

Meadows, R. P., Nettesheim, D. G., Xu, R. X., Olejniczak, E. T., Petros, A. M., Holzman, T. F., Severin, J., Gubbins, E., Smith, H., and Fesik, S. W. 1993. *Biochemistry* **32:** 754–765.

Mitomo, K., Nakayama, K., Fujimoto, K., Sun, X., Seki, S., and Yamamoto, K. 1994. Two different cellular redox systems regulate the DNA-binding activity of the p50 subunit of NFκB in vitro. *Gene* **145:** 197–203.

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 NMR experiments. *J. Magn. Reson.* **87:** 183–188.

Moore, J. M., Lepre, C., Gippert, G. P., Chazin, W. J., Case, D. A., and Wright, P. E. 1991. High-resolution solution structure of reduced French bean plastocyanin and comparison with crystal structure of poplar plastocyanin. *J. Mol. Biol.* **221:** 533–555.

Nilges, M. 1993. A calculational strategy for the structure determination of symmetric dimers by ^{1}H NMR. *Proteins* **17:** 297–309.

Nilges, M., Clore, G. M., and Gronenborn, A. M. 1990. ^{1}H-NMR stereospecific assignments by conformational database searchers. *Biopolymers* **29:** 813–822.

Nilges, M., Habazettl, J., Brünger, A. T., and Holak, T. A. 1991. Relaxation matrix refinement of the solution structure of squash trypsin inhibitor. *J. Mol. Biol.* **291:** 499–510.

Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, S., and Nishimura, Y. 1994. Solution structure of a specific complex of the Myb DNA-binding domain with cooperative recognition helices. *Cell* **79:** 639–648.

Omichinski, J. G., Clore, G. M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S. J., and Gronenborn, A. M. 1993a. NMR structure of a specific DNA complex of the Zn-containing DNA binding domain of GATA-1. *Science* **261:** 438–446.

Omichinski, J. G., Trainor, C., Evans, T., Gronenborn, A. M., Clore, G. M., and 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. U. S. A.* **90:** 1676–1680.

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

Orkin, S. H. 1992. GATA-binding transcription factors in hematopoietic cells. *Blood* **80:** 575–581.

Ösapay, K., Theriault, Y., Wright, P. E., and Case, D. A. 1994. Solution structure of carbonmonoxy myoglobin determined from nuclear magnetic resonance distance and chemical shift constraints. *J. Mol. Biol.* **244:** 183–197.

Oschkinat, H., Griesinger, C., Kraulis, P. J., Sørensen, O. W., Ernst, R. R., Gronenborn, A. M., and Clore, G. M. 1988. Three-dimensional NMR spectroscopy of a protein in solution. *Nature* **332:** 374–376.

Otting, G. and Wüthrich, K. 1990. Heteronuclear filters in two-dimensional [^{1}H,^{1}H]-NMR spectroscopy: combined use with isotope labeling for studies of macromolecular conformation and intermolecular interactions. *Quart. Rev. Biophys.* **23:** 39–96.

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.

Post, C. B., Meadows, R. P., and Gorenstein, D. G. 1990. On the evaluation of interproton distances for three-dimensional structure determination by NMR using a relaxation rate matrix analysis. *J. Am. Chem. Soc.* **112:** 6796–6803.

Qin, J., Clore, G. M., and Gronenborn, A. M. 1994. The high-resolution three-dimensional solution structure of the oxidized and reduced states of human thioredoxin: delineation of conformation differences between the two redox states. *Structure* **2:** 503–522.

Qin, J., Clore, G. M., Kennedy, W. M. P., Huth, J. R., and Gronenborn, A. M. 1995. Solution structure of human thioredoxin in a mixed disulfide intermediate complex with its target peptide from the transcription factor NFκB. *Structure* **3:** 289–297.

Pascal, S. M., Singer, A. U., Gish, G., Tamazaki, T., Shoelson, S. E., Pawson, T., Kay, L. E., and Forman-Kay, J. D. 1994. Nuclear magnetic resonance structure of an SH2 domain of phospholipase C-1 complexed with a high-affinity binding peptide. *Cell* **77:** 461–472.

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.

Read, C. M., Cary, P. D., Crane-Robinson, C., Driscoll, P. C., and Norman, D. G. 1993. Solution structure of a DNA-binding domain from HMG-1. *Nucl. Acids Res.* **21:** 3427–3436.

Read, C. M., Cary, P. D., Preston, N. S., Lneicek-Allen, M., and Crane-Robinson, C. 1994. The DNA sequence specificity of HMG boxes lies in the minor wing of the structure. *EMBO J.* **13:** 5639–5646.

Richardson, J. S. 1981. The anatomy and taxonomy of protein structure. *Adv. Prot. Chem.* **34:** 167–330.

Robien, M. A., Clore, G. M., Omichinski, J. G., Perham, R. N., Appella, E., Sakaguchi, K., and Gronenborn, A. M. 1992. Three-dimensional solution structure of the E3-binding domain of the dihydrolipoamide succinyltransferase core from the 2-oxoglutarate dehydrogenase multienzyme complex of *Escherichia coli*. *Biochemistry* **31:** 3463–3471.

Saenger, W. 1984. *Principles of Nucleic Acid Structure.* Springer-Verlag, New York.

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:** 567–578.

Sinclair, A. H., Berta, P., Palmer, M. S., Hawkins, J. R., Griffiths, B. L., Smith, M. J., Foster, J. M., Frischauf, A.-M., Lovell-Badge, R., and Goodfellow, P. N. 1990. A gene from the human sex determining region encodes a protein with homology to a conserved DNA binding motif. *Nature* **346:** 240–244.

Skelton, N. J., Aspiras, F., Ogez, J., and Scall, T. J. 1995. Proton NMR assignments and solution conformation of RANTES, a chemokine of the C-C type. *Biochemistry* **34:** 5329–5342.

Spitzfaden, C., Braun, W., Wider, G., Widmer, H., and Wüthrich, K. 1994. Determination of the NMR solution structure of the cyclophilin A-cyclosporin complex. *J. Biomol. NMR* **4:** 463–482.

Szyperski, T., Güntert, P., Stone, S. R., and Wüthrich, K. 1992. Nuclear magnetic resonance solution structure of hirudin (1-51) and comparison with corresponding three-dimensional structures determined using the complete 65-residue hirudin polypeptide chain. *J. Mol. Biol.* **228:** 1193–1205.

Terasawa, H., Kohda, D., Hatanaka, H., Tsuchiya, S., Ogura, K., Nagata, K., Ishii, S., Mandiyan, U., Ullrich, A., Schlessinger, J., and Inagaki, F. 1994. Structure of the N-terminal SH3 domain of GRB2 complexed with a peptide from the guanine nucleotide releasing factor Sos. *Nature Struct. Biol.* **1:** 891–897.

Theriault, Y., Logan, T. M., Meadows, R., Yu, L., Olejniczak, E. T., Holzman, T., Simmer, R. L., and Fesik, S. W. 1993. Solution structure of the cyclosporin A/cyclophilin complex by NMR. *Nature* **361:** 88–91.

Tijan, R. and Maniatis, T. 1994. Transcriptional activation: a complex puzzle with few easy pieces. *Cell* **77:** 5–8.

Vuister, G. W. and Bax, A. 1993. Quantitative J correlation: a new approach for measuring homonuclear three-bond J($H^N H\alpha$) coupling constants in ^{15}N enriched proteins. *J. Am. Chem. Soc.* **115:** 7772–7777.

Vuister, G. W., Boeens, R., and Kaptein, R. 1988. Nonselective three-dimensional NMR spectroscopy: the 3D NOE-HOHAHA experiment. *J. Magn. Reson.* **80:** 176–185.

Vuister, G. W., Boelens, R., Kaptein, R., Hurd, R. E., Johnmn, B. K., and Van Zilj, P. C. M. 1991. Gradient enhanced HMQC and HSQC spectroscopy: applications to ^{15}N-labeled Mnt repressor. *J. Am. Chem. Soc.* **113:** 9688–9690.

Vuister, G. W., Clore, G. M., Gronenborn, A. M., Powers, R., Garrett, D. S., Tschudin, 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.

Wagner, G., Braun, W., Havel, T. F., Schaumann, 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. Mol. Biol.* **196:** 611–639.

Weir, H. M., Kraulis, P. J., Hill, C. S., Raine, A. R. C., Laue, E. D., and Thomas, J. O. 1993. Structure of the HMG box motif in the B-domain of HMG-1. *EMBO J.* **12:** 1311–1319.

Werner, M. H., Huth, J. R., Gronenborn, A. M., and Clore, G. M. 1995a. Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. *Cell* **81:** 705–714.

Werner, M. H., Bianchi, M. E., Gronenborn, A. M., and Clore, G. M. 1995b. NMR spectroscopic analysis of the DNA conformation induced by the human testis determining factor SRY. *Biochemistry* **34:** 11998–12004.

Wittekind, M., Mapelli, Garmer, B. T., Suen, K.-L., Goldfarb, V., Tsao, J., Lavoie, T., Barbacid, M., Meyers, C. A., and Mueller, L. 1994. The orientation of the fragments from Sos proteins bound to the N-terminal SH3 domain of Grb2 determined by NMR spectroscopy. *Biochemistry* **33:** 13531–13539.

Williamson, M. P., Havel, T. F., and Wüthrich, K. 1985. Solution conformation of proteinase inhibitor II from bull seminal plasma by ^1H nuclear magnetic resonance and distance geometry. *J. Mol. Biol.* **182:** 295–315.

Wüthrich, K. 1986. *NMR of Proteins and Nucleic Acids,* Wiley, New York.

Xu, R. X., Word, J. M., David, D. G., Rink, M. J., Willard, D. H., and Gampe, R. T. 1995. Solution structure of the human pp60c-src SH2 domain complexed with a phosphorylated tyrosine pentapeptide. *Biochemistry* **34:** 2107–2121.

Yip, P. and Case, D. A. 1991. A new method or refinement of macromolecular structures based on nuclear Overhauser effect spectra. *J. Magn. Reson.* **83:** 643–648.

Yu, H., Chen, J. K., Feng, S., Dalgarno, D. C., Brauer, A. W., and Schreiber, S. L. 1994. Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell* **76:** 933–945.

Zhang, H., Zhao, D., Revington, M., Lee, W., Jia, X., Arrowsmith, C., and Jardetzky, O. 1994. The solution structure of the trp repressor-operator DNA complex. *J. Mol. Biol.* **238:** 592–614.

Zhao, D. and Jardetzky, O. 1994. An assessment of the precision and accuracy of protein structures determined by NMR: dependence on distance errors. *J. Mol. Biol.* **239:** 601–607.

Zuiderweg, E. R. P. and Fesik, S. W. 1989. Heteronuclear thee-dimensional NMR spectroscopy of the inflammatory protein C5a. *Biochemistry* **28:** 2387–2391.

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 headpice. *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.

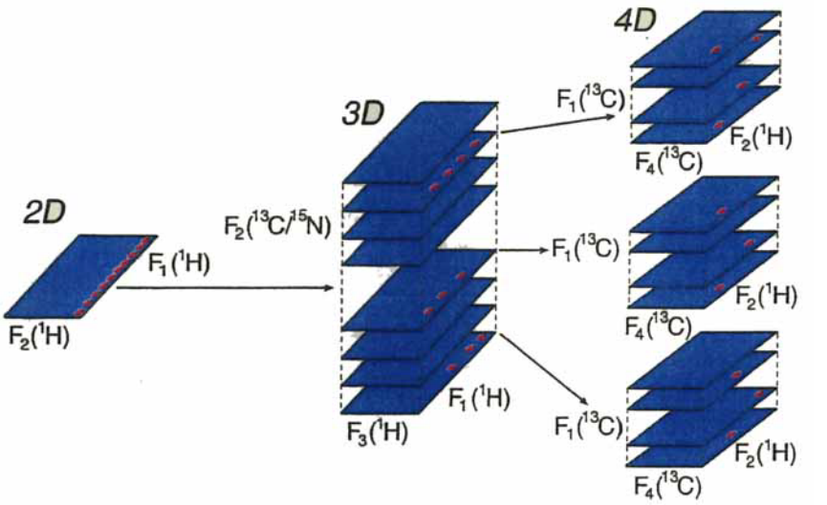


PLATE 1. Schematic diagram illustrating the effects of increasing the dimensionality on the spectral resolution of an NOE spectrum. In the 2D spectrum, the proton chemical shift of the destination resonances (along the F_2 axis) for all 11 cross-peaks is the same so that one cannot assess the number of destination protons involved. In the 3D spectrum, the cross-peaks appear in three planes, edited according to the shift of the heavy atom (^{15}N or ^{13}C) attached to the destination protons. The identity of the orginating protons, however, is only defined by their proton chemical shifts. Finally, in the 4D spectrum each cross-peak is characterized by four chemical shift coordinates, the proton chemical shifts of the two protons involved, and the chemical shifts of the heavy atoms to which they are attached.

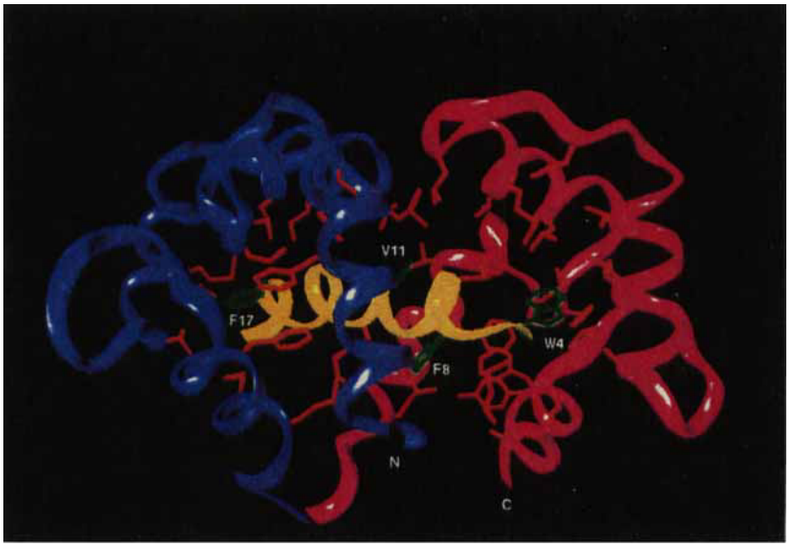


PLATE 2. Schematic ribbon drawing of the solution NMR structure of the Ca^{2+}-CaM-M13 complex with the amino-terminal domain in blue, the carboxy-terminal domain in purple, and the M13 peptide in yellow. The hydrophobic side chains of the amino- and carboxy-terminal domains of CaM are shown in red, and the Trp4, Phe8, Va111, and Phe17 side chains of the peptide are displayed in green.

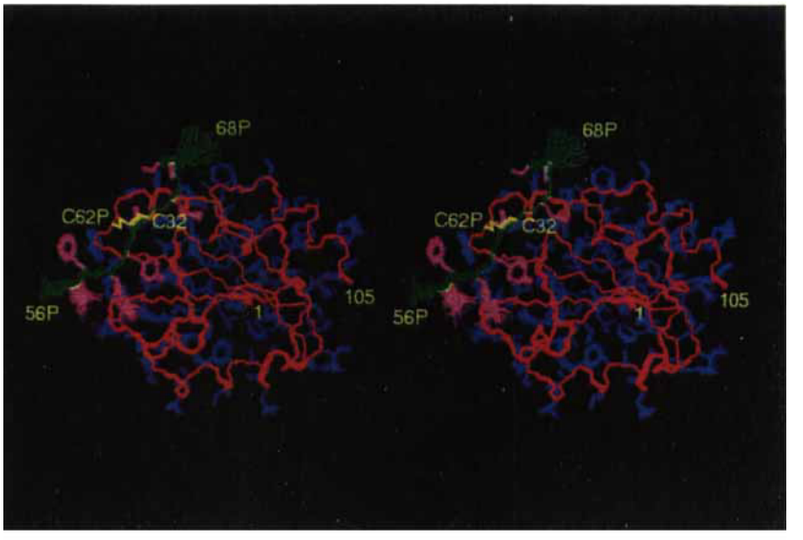
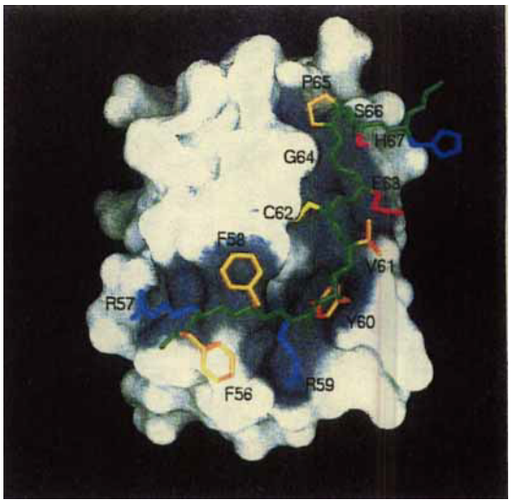


PLATE 3. Stereo view showing a best fit superposition of the backbone (N, Cα, C) atoms and ordered side chains of the 60 simulated annealing structures of the hTRX-NFκB complex. The backbone and side chains of hTRX are shown in red and blue, respectively; the backbone and side chains of the NFκB peptide are shown in green and magenta, respectively; and the disulfide bond between Cys32 of hTRX and Cys62 of the NFκB peptide in yellow. The letter P in the residue numbering scheme indicates the NFκB peptide. (From Qin et al., *Structure*, (1995) **3**: 289–297.)

PLATE 4. (A) View of the molecular surface of hTRX illustrating the cleft in which the NFκB peptide is located and (B) schematic representation showing the interactions between the NFκB peptide and hTRX. The degree of curvature of the molecular surface on (A) is color coded from white (convex) to dark grey (concave). Hence, the cleft is visualized as the contiguous boot-shaped grey region on the surface of hTRX. The backbone of the peptide is shown in green, and side chains are colored as follows; Phe, Tyr, Val, Pro, and Cys in orange; Arg and His in blue; Glu in red; and Ser in Magenta. Note that the side chains of Phe56 and His 67 of the bound NFκB peptide are disordered in solution.



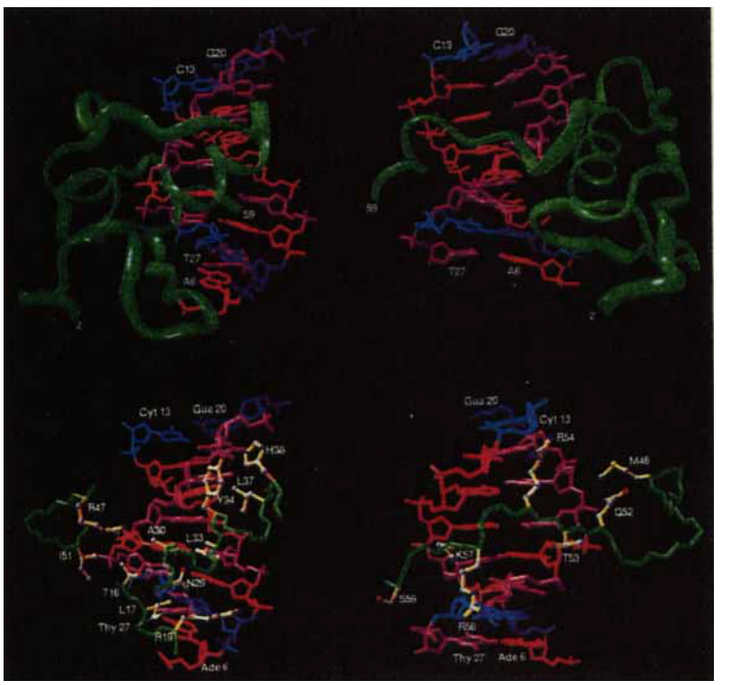


PLATE 5. (A) and (B) Schematic ribbon drawings illustrating interactions of cGATA-1 with DNA. (C) and (D) Side chain interactions between cGATA-1 code and the DNA in the major and minor grooves, respectively, The protein backbone is shown in green and the protein side chains in yellow; the color for the DNA bases is as follows: red for A, lilac for T, dark blue for G, and light blue for C. (From Omichinski et al., *Science* (1993a) **261**: 438–446.)

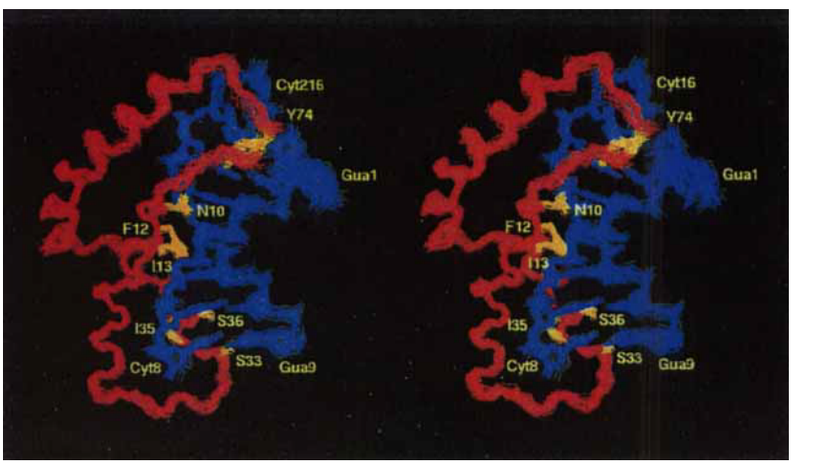


PLATE 6. Stereo view showing a best fit superposition of the 35 simulated annealing structures of the specific complex of hSRY-HMG with DNA. The backbone (N, Cα,C) atoms (residues 4–75) of hSRY-HMG are shown in red, side chains that contact the DNA bases in yellow, and all non-hydrogen atoms of the DNA in blue. (From Werner et al., *Cell* (1995a) 705–714.)

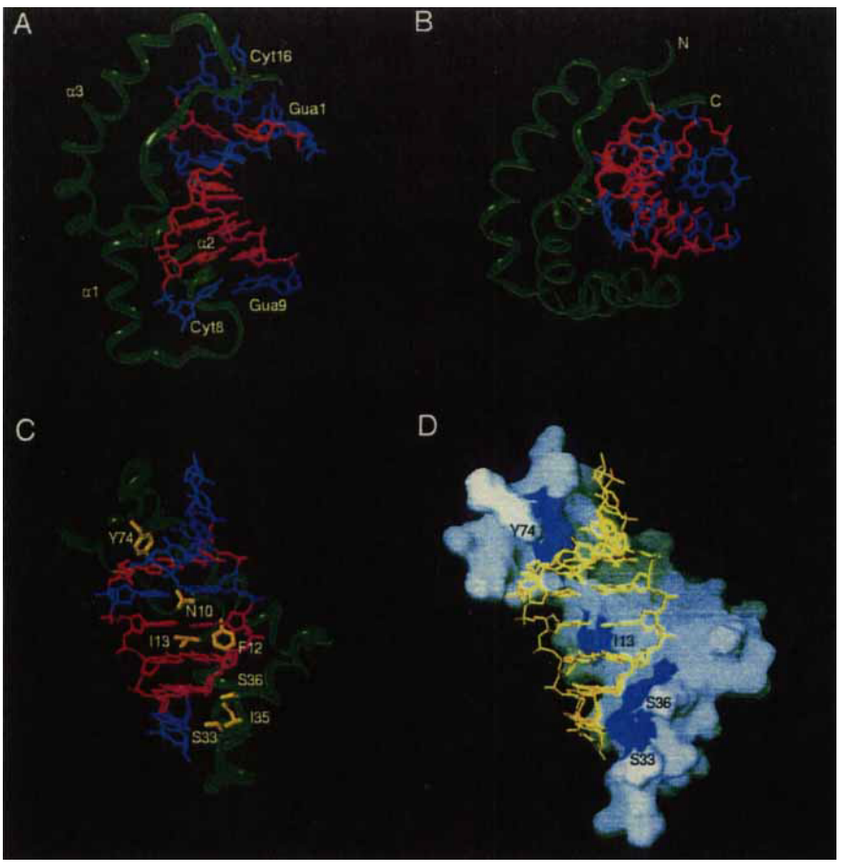


PLATE 7. (A,B,C) Three views illustrating the interaction of hSRY-HMG with DNA. The protein is shown as a schematic ribbon drawing in green, and the color coding used for the DNA bases is red for A, lilac for T, dark blue for G, and light blue for C. Side chains that contact the DNA bases are depicted in yellow in (C). (D) Same view as in (C) with the molecular surface of the protein shown in grey and the DNA atoms in yellow. The patches of blue on the protein surface indicate the location of the side chains of 4 of the 7 residue that interact with the DNA bases. The protrusion on the surface of the protein associated with the side chain of Ile13, which partially intercalates between base pairs 5 and 6 of the DNA, is clearly visible in (D). (From Werner et al., *Cell* (1995a) **81**: 705–714.)

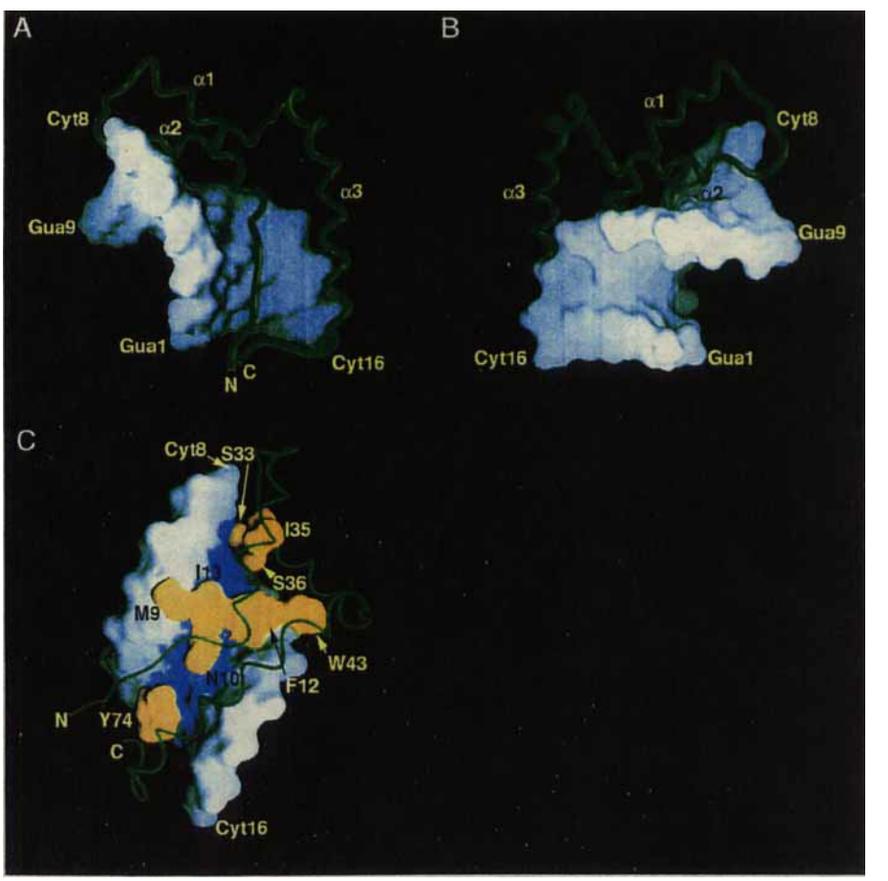


PLATE 8. Complex of hSRY-HMG with DNA viewed in the (A) minor and (B) major grooves of the DNA illustrating the bend in the DNA, the widening of the minor groove, and the compression of the major groove, with the molecular surface of the DNA shown in grey and a ribbon drawing of the protein backbone in green. (C) View of the complex in the minor groove showing the molecular surface of the DNA backbone (grey) and bases (blue), a ribbon drawing of the hSRY-HMG backbone (green), and the molecular surface of selected sidechain atoms (yellow), illustrating the T-shaped wedge that contacts basepairs 4 to 6 (Met9, Asn10, Phe12, Ile13, and Trp43) and the anchor points that contact the two ends of the DNA octamer (Tyr74 at basepair 3; and Ser33, Ile35 and Ser36 at basepairs 7 and 8). (From Werner et al., *Cell* (1995a) **81**:705–714.)

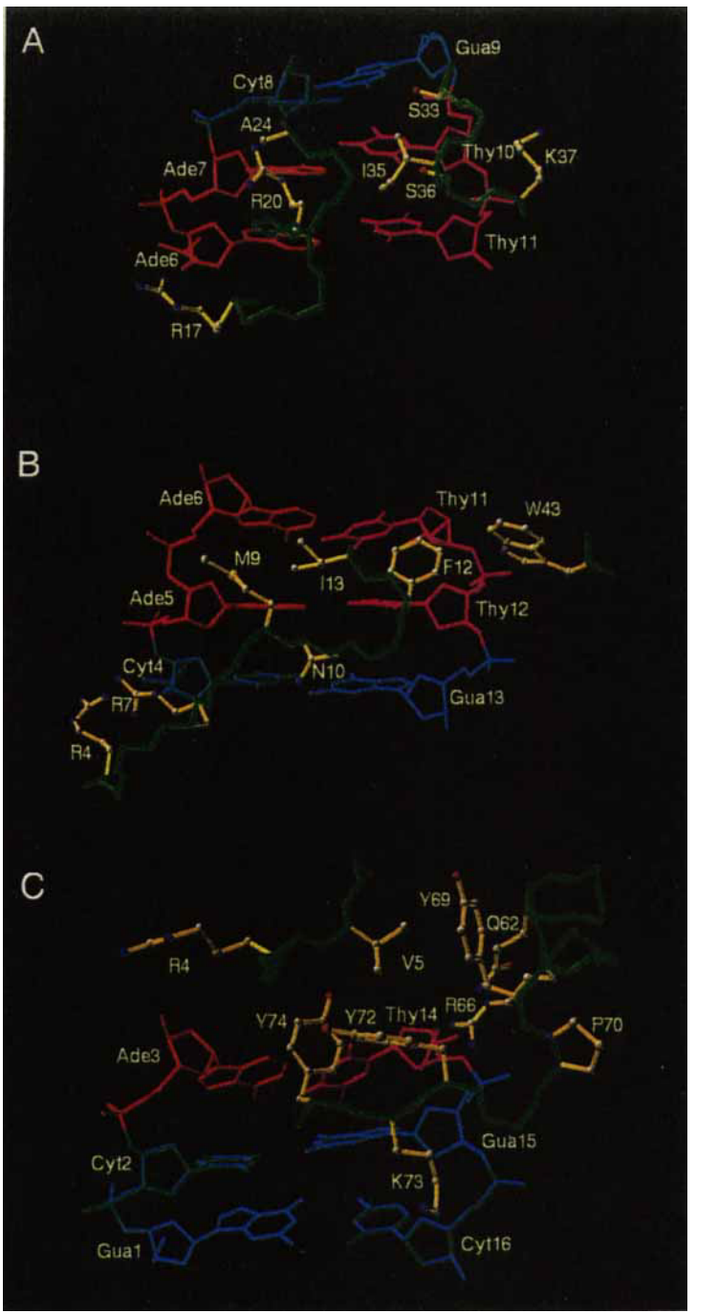


PLATE 9. Views of side chain interactions between hSRY-HMG and the DNA. All contacts are in the minor groove. The polypeptide backbone is shown in green, the side chain bonds in yellow, and the side chain carbon, oxygen, and nitrogen atoms in grey, red, and blue, respectively. The color coding for the DNA is the same as in Figure 12. (From Werner et al., *Cell* (1995a) **81**: 705–714.)

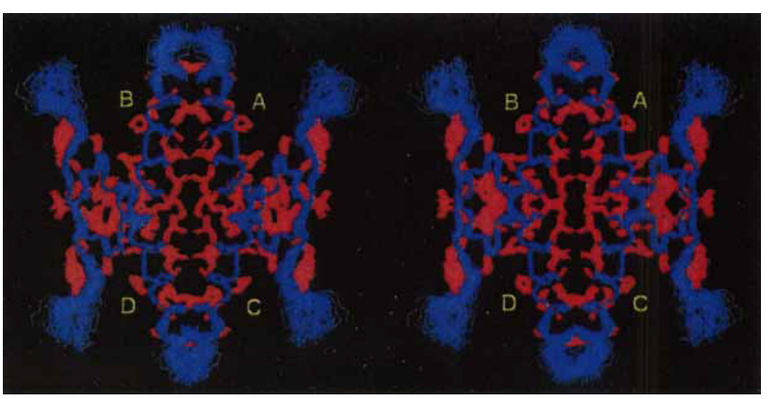
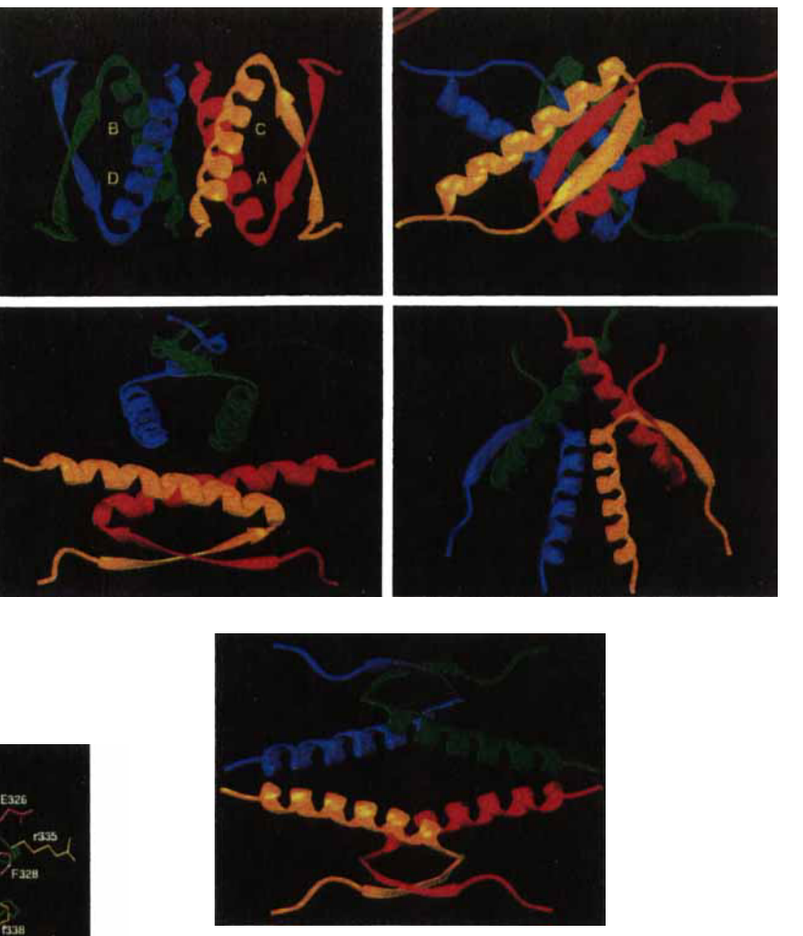


PLATE 10. Best fit superposition of the backbone (blue) and ordered side chains (red) of an ensemble of 76 simulated annealing structures of the p53 oligomerization domain. The backbone of residues 322–357 is displayed. (From Clore et al., *Nature Struct. Biol.* (1995b) **2**: 321–332.)

PLATE 11. Ribbon diagrams illustrating five different views of the solution structure of the p53 oligomerization domain (residues 321–358). The A, C, B, and D subunits are shown in red, yellow, green, and blue, respectively. (From Clore et al., 1995c.)



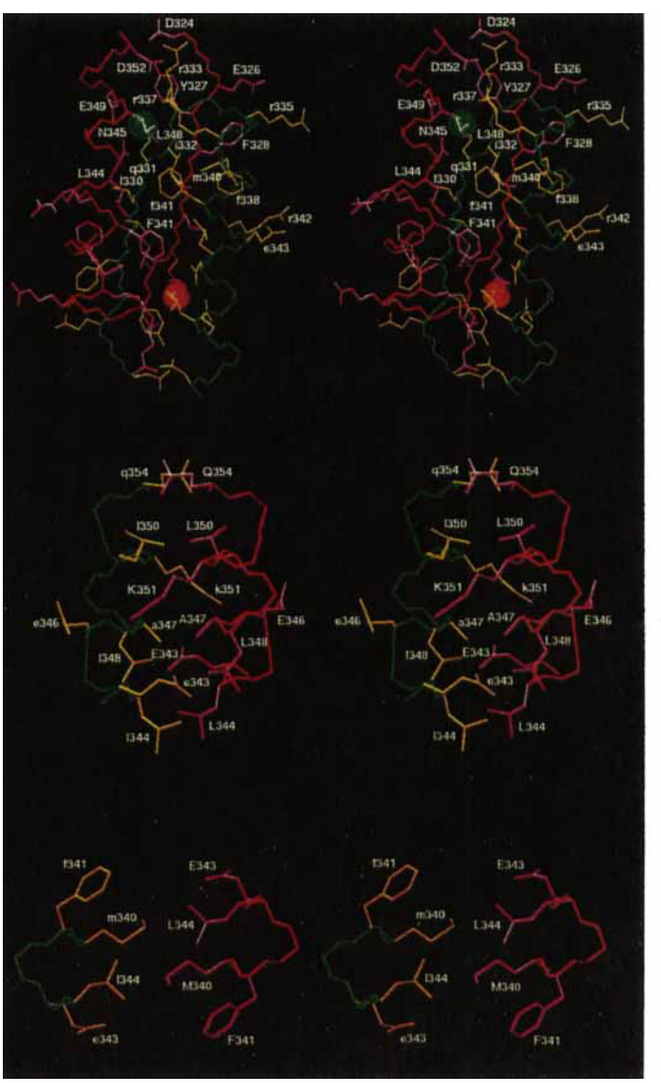


PLATE 12. Stereo views of the (A) AC, (B) AB, and (C) AD interfaces for the restrained regularized mean structure of the oligomerization domain of p53 derived from the ensemble of simulated annealing structures shown in Figure 17. The backbone (N, Cα, C atoms) and side chains of subunit A are shown in red and purple, respectively, with the residues labeled with capital letters; the backbone and side chains of the other subunit are shown in green and yellow respectively, with the residues labeled with small letters. In (A) the two symmetrically related water molecules within the primary dimer are shown as solid spheres. (From Clore et al., 1995c.)