Proton and Nitrogen NMR Sequence-Specific Assignments and Secondary Structure Determination of the *Bacillus subtilis* SPO1-Encoded Transcription Factor 1†

Xin Jia,§ Joseph M. Reisman,§ Victor L. Hsu,§,# E. Peter Geiduschek,⊥ Joseph Parello,‖ and David R. Kearns*,§

Department of Chemistry, Department of Biology, and Center for Molecular Genetics, University of California at San Diego, La Jolla, California 92093, Unité de Recherche Associée au CNRS, No. 1111, Faculté de Pharmacie, Montpellier, France, and La Jolla Cancer Research Foundation, La Jolla, California 92037

Received March 11, 1994; Revised Manuscript Received May 16, 1994®

ABSTRACT: Sequence-specific ^1H and ^{15}N NMR[1] assignments are reported for the transcription factor 1 (TF1), a 22-kDa type II DNA-binding protein (DBPII) that consists of two 99-residue monomers. An assignment strategy is employed that uses six complementary selectively deuterium-labeled TF1 variants and an uniformly ^{15}N-labeled TF1 variant. Two-dimensional and three-dimensional homonuclear and heteronuclear NMR correlated spectra are analyzed and yield nearly complete assignments for the ^1H and ^{15}N resonances. Discrete protein secondary structure domains are also defined; in each monomer, three α-helices, an antiparallel β-sheet, and an antiparallel β-ribbon are identified. Analyses of two dimers formed from two distinct selectively deuteriated monomers serve to identify a number of interproton contacts as either intermonomeric or intramonomeric. An analysis of amide proton exchange reveals that the carboxy-terminal α-helix is less stable than the other two α-helices in each monomer. A previously proposed working structural model of the TF1 dimer [Geiduschek et al. (1990) *J. Struct. Biol. 104,* 84–90], based on the crystal structure of a highly homologous DBPII, the *Bacillus stearothermophilus*-encoded HU protein, is generally supported by our results. Several departures from this model, however, are noted. Most notably, the carboxy-terminal tail of TF1 adopts an α-helical conformation with a backbone distortion at Lys93.

The type II DNA-binding proteins (DBPII) are a family of basic, dimeric proteins found in all bacteria (Pettijohn, 1988). *E. coli* encodes two such proteins, HU[1] (Kellenberger, 1988) and the integration host factor (IHF) (Craig & Nash, 1986), that exhibit vastly different DNA-binding characteristics. The HU protein, a heterodimer of HU_A and HU_B subunits, that can also exist as homodimer, exhibits no sequence specificity in DNA-binding in any of its forms (Dri et al., 1991). IHF exists as a heterodimer (Yang & Nash, 1989; Flamm & Weisberg, 1985) that displays sequence specificity in DNA binding (Friedman, 1988). The *B. subtilis* bacteriophage SPO1 also encodes a DBPII. This homodimeric 22-kDa protein, transcription factor 1 (TF1), binds preferentially to DNA containing 5-(hydroxymethyl)-2′-deoxyuridine (hmUra) in place of thymine and to specific sites in the hmUra-containing SPO1 genome with high affinity (Greene & Geiduschek, 1985).

A highly resolved structure is known for only one DBPII, the *B. stearothermophilus*-encoded HU protein (*Bst.* HU) (White et al., 1989; Tanaka et al., 1984). Analysis of the *Bst.* HU and the other DBPII primary sequences suggests that the overall structural organization of *Bst.* HU may serve as a reasonable model for all the DBPII (Geiduschek et al., 1990). The *Bst.* HU crystal structure contains two concave β-ribbon arms, which have been suggested to play a role in the recognition of DNA by the DBPII. The DBPII, therefore, may exemplify an interesting DNA-binding structural motif: the β-ribbon motif (Kim, 1992). This motif has been proposed to be one in which contacts are formed between the protein and nucleic acid bases in the major groove (Somers & Phillips, 1992) or minor groove (Barry & Hawley, 1991; Nash & Granston, 1991) of bound *B*-form DNA.

Of all the known DBPII, only TF1 and IHF exhibit sequence specificity in DNA binding. This functional difference has been explained by the presence of short carboxy-terminal extensions (with respect to the 90-residue HU proteins) in IHF and TF1. IHF, with monomers of 93 and 98 residues, exhibits some sequence specificity in DNA binding (Yang & Nash, 1989) while TF1, with 99-residue monomers, preferentially binds to DNA containing hydroxymethyluracil (hmUra) and to certain sites in hmUra-containing DNA (Greene & Geiduschek, 1985). While the structure of the TF1 nine-amino-acid tail is not known and cannot be determined by analogy to the *Bst.* HU structure, its analysis may be important to understanding the DNA-binding motif employed by both TF1 and IHF. For this reason and to establish whether the *Bst.* HU crystal structure serves as a good model for TF1 in solution, we have used nuclear magnetic resonance spectroscopy techniques to determine the structure of TF1 in solution.

† This work is supported by the National Institute of General Medical Sciences (GM40635 to D.R.K. and GM39418 to E.P.G.) and the PICS (141 CNRS-MAE, France; J.P.). We are grateful to the National Institutes of Health (1S01-RR03342) and to the National Science Foundation (BB586-12359) for providing resources for a 500-MHz NMR instrument. One of us (J.M.R.) is indebted to the French Ministry of Foreign Affairs for a research fellowship.

* To whom correspondence should be addressed.

§ Department of Chemistry.

Current address: Department of Biochemistry and Biophysics, Oregon State University, ALS 2011, Corvallis, OR 97331.

⊥ Department of Biology and Center for Molecular Genetics.

‖ Unité de Recherche Associée au CNRS, No. 1111, Faculté de Pharmacie, Montpellier, France, and La Jolla Cancer Research Foundation, La Jolla, CA 92037.

® Abstract published in *Advance ACS Abstracts,* July 1, 1994.

[1] Abbreviations: NMR, nuclear magnetic resonance; 2D, two-dimensional; 3D, three-dimensional; *Bst.* HU, the *B. stearothermophilus*-encoded HU protein; IHF, the *E. coli*-encoded integration host factor; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; TF1, the *B. subtilis* SPO1-encoded transcription factor 1; TOCSY, homonuclear total correlation spectroscopy; HMQC, heteronuclear multiple-quantum correlation; HSQC, heteronuclear single-quantum correlation, TPPI, time-proportional phase incrementation; TSP-d_4, 3-(trimethylsilyl)[2,2,3,3-$^2H_4$]propionic acid.

0006-2960/94/0433-8842$04.50/0

Table 1: Isotopic Composition of Selectively Deuteriated TF1

variant name[b]	fully protiated residues	partially protiated residues	protiation[a] min %	protiation[a] max %
FGIY	Phe, Gly, Ile, Tyr	Ala, Glu, Gln, Ser, Asp, Asn, Val, Thr	16	58
FVSY	Phe, Val, Ser, Tyr	Ala, Gly, Glu, Gln, Asp, Asn, Thr, Ser	22	53
LAS	Leu, Ala, Ser	Gly, Glu, Gln, Asp, Asn, Thr, Tyr, Ser	25	50
VIFY	Val, Ile, Phe, Tyr	Ala, Gly, Glu, Gln, Asp, Asn, Tyr, Ser	23	51
Phe*	Phe-H3,H5,H6	Ala, Gly, Glu, Gln, Asp, Asn, Tyr, Ser	2	30
TE[c]	Thr, Glu, Gln, Asn	Ala, Lys, Gly, Ser, Asp, Tyr	15	52
PRAG[c]	Pro, Arg, Ala, Gln, Asn	Gly, Glu, Asp, Ser, Tyr	28	35
KMP[c]	Lys, Met, Pro, Gln, Asn	Ala, Gly, Glu, Asp, Ser, Tyr	35	54

[a] Values indicate percentages of the nonexchangeable protons in TF1, where min % is calculated from the total number of protons for the residues in the second column and max % is from those in the second column plus the third column. [b] The nomenclature is based on names of protiated residues. [c] These proteins were produced in the auxotrophic *E. coli* strain, DL39(pTF1X).

We have employed both two-dimensional (2D) and three-dimensional (3D) NMR techniques to approach the structure determination of TF1. Because the resonance assignment of TF1 in solution presents a number of experimental difficulties (due primarily to the protein's relatively high molecular weight), a strategy of using a set of complementary selectively deuteriated TF1 variants has been employed. This strategy, absent the use of 3D NMR techniques, has proven quite successful in providing the ^{1}H NMR assignment of the *trp* repressor, a homodimeric protein with a molecular weight of 25 kDa (Arrowsmith et al., 1990). Our previous studies also revealed that this strategy should yield similar success in the analysis of TF1 (Reisman et al., 1991, 1993). In this paper we present nearly complete ^{1}H and ^{15}N NMR resonance assignments for TF1 which are based upon NMR analyses of six ^{2}H-labeled TF1 variants and an uniformly ^{15}N-labeled TF1 variant using 2D and 3D, homonuclear and heteronuclear techniques. While a further use of the results presented will be to make possible a high-resolution structure determination of TF1, the present analyses yield information regarding the dominant secondary structural features, the global structure, and relative stability of the various structural domains of TF1 in solution.

MATERIALS AND METHODS

Protein Preparation. Eight selectively deuteriated TF1 variants were synthesized in either *E. coli* MG1655 or in the auxotroph DL39. Both strains contained the pTF1X TF1-overproducing plasmid (Sayre, 1989; Härd et al., 1989b). Table 1 lists and names these selectively deuteriated TF1 variants along with the amino acids expected to be fully protiated or partially protiated in each variant. The Phe* variant contained as its sole added protiated residue, [2,4-$^2H_2$]phenylalanine, synthesized according to a procedure to be discussed elsewhere (Dauzonne & Parello, in preparation). Procedures for the purification of TF1 from MG1655(pTF1X), have been described previously (Reisman et al., 1993; Härd et al., 1989a). When similar procedures were used to obtain protein from DL39(pTF1X), we used a medium described by Muchmore et al. (1989), containing appropriate unlabeled amino acids for growth and labeled amino acids for over expression. An uniformly ^{15}N-labeled TF1 sample was prepared from MG1655(TF1X) as described above, using ^{15}N-labeled amino acids as well as ^{15}N-labeled ammonium chloride as nitrogen sources.

Protein concentrations were determined by UV absorbance, using an extinction coefficient of 1200 M^{-1} cm^{-1} at 280 nm (Sayre & Geiduschek, 1990). Light-scattering contributions were eliminated by monitoring absorbance at 330 nm. Between 20 and 25 mg of TF1 were obtained from 6 L of MG1655/TF1X culture and 3 L of DL39/TF1X culture.

"Heterodimeric" TF1 samples were prepared by adding two differently deuteriated TF1 variants in a 1:1 ratio and then allowing the mixture to equilibrate at room temperature for at least 24 hours.

NMR Spectroscopy. All 2D NMR spectra and one heteronuclear 3D spectrum, ^{1}H–^{15}N-NOESY-HMQC, were acquired on a Bruker AMX-500 spectrometer equipped with an X32 Aspect-3000 computer and either a fixed-frequency ^{1}H probe (homonuclear experiments) or a two-channel NMR interface and a double-resonance ^{1}H/^{15}N reverse probe (heteronuclear experiments). The homonuclear 3D spectrum, ^{1}H-TOCSY-NOESY, was acquired on a Bruker AMX-600 spectrometer equipped with an X32 Aspect-3000 computer and a fixed-frequency ^{1}H probe. NMR samples consisted of 2.0-4.5 mM TF1 monomer in 400 mM NaCl/100 mM sodium phosphate (pH 6.8). 2D spectra were acquired with samples in 90% H_2O/10% D_2O and in 99.99% D_2O solutions. 3D spectra were acquired with samples in 95% H_2O/5% D_2O solutions. All spectra were acquired at 30 °C.

Two-dimensional NOESY spectra were acquired with mixing times of 75, 100, 150, and 200 ms. Solvent suppression for NOESY spectra acquired in H_2O solutions was achieved by using a jump–return observe pulse (Plateau & Gueron, 1982) combined with low-power presaturation during both relaxation and mixing time delays.

For 2D TOCSY experiments (Bax et al., 1985), a spin-locked mixing time of 40 ms and a radio frequency (rf) field strength of 12 kHz were used. The residual H_2O signal was purged by high power trim pulses of 2.5 ms at the beginning and end of each spin–lock pulse and by low-power presaturation during the relaxation delay.

Two-dimensional ^{1}H–^{15}N-HMQC experiments employed a jump–return water suppression pulse sequence but no low-power presaturation of the residual H_2O signal (Sklenar & Bax, 1987). Two-dimensional ^{1}H–^{15}N-HSQC experiments (Tate et al., 1991), however, used low-power rf presaturation during the relaxation delay. Both 2D heteronuclear experiments employed signal aliasing in the ^{15}N dimension (Archer et al., 1992) and heteronuclear GARP decoupling (Shaka et al., 1985). Each 2D spectrum was acquired as a 2048 × 500 (t2 × t1) time-domain matrix and then processed to yield a 1024 × 1024 real matrix in the frequency domain.

A homonuclear 3D ^{1}H-TOCSY-NOESY experiment (Oshkinat et al., 1989) with a NOESY mixing time of 100 ms and a TOCSY mixing time of 35 ms was used to analyze a fully protiated TF1 sample (7.0 mM) in H_2O. Time-domain data, consisting of a 512 × 192 × 176 (t3 × t2 × t1) complex matrix in which each t2 × t1 vector was signal-averaged over eight

scans, were used to generate the 512 × 256 × 256 real frequency-domain spectrum.

A heteronuclear 3D ^1H–^{15}N-NOESY-HMQC experiment (Kay et al., 1989) with a NOESY mixing time of 100 ms was used to analyze 4.5 mM uniformly ^{15}N-labeled TF1 in a H_2O solution. Time-domain data, consisting of a 512(^1H) × 44-(^{15}N) × 192(^1H) (t3 × t2 × t1) matrix in which each t2 × t1 vector was signal-averaged over 32 scans, were used to generate the 512 × 64 × 256 frequency-domain spectra. In all 3D NMR experiments, presaturation was employed to suppress the residual H_2O resonance.

For all experiments, TPPI was used for quadrature detection in all indirectly detected dimensions (Marion & Wüthrich, 1983), and spectral widths of 6097.56 Hz and 2032.52 Hz were observed in the ^1H and ^{15}N dimensions, respectively. All ^1H-NMR chemical shifts are reported downfield from external TSP-d_4. All ^{15}N-NMR chemical shifts are referenced with respect to an external 1 M NH_4Cl standard and reported relative to the liquid ammonia reference frequency (Srinivasan & Lichter, 1977). All NMR data were processed and analyzed using Felix2.05 software (Biosym) on an SGI Indigo workstation.

The following procedure was followed to determine which amide protons were in slow chemical exchange with solvent deuterons: a 2D HMQC experiment was recorded in H_2O solution, and the NMR sample was then dried under a stream of N_2 and redissolved in 99.99% D_2O. A 2D ^1H–^{15}N HMQC spectrum was acquired 30 min after the sample was redissolved. This time-domain spectrum, consisting of a 2048(^1H) × 100-(^{15}N) matrix and obtained over 65 min, was used to generate a 1024(^1H) × 512(^{15}N) frequency-domain spectrum. Another 2D HMQC spectrum was acquired after the sample was maintained at room temperature in D_2O for 5 days.

RESULTS

Identification of the ^1H Spin Systems. Crosspeak intensities in some multidimensional NMR experiments (i.e., COSY and TOCSY) are proportional to the homonuclear scalar coupling, and thus, these experiments are of only limited aid in identifying amino acid spin systems in large biomolecules (Arrowsmith et al., 1990). Identification is especially difficult in large biomolecules for those spin systems with three or more *J*-coupled (Rosen et al., 1991) or weakly coupled proton pairs. Therefore, although TOCSY spectra can often be used for the identification of less complex side chains, a NOESY-based approach to spin system identification is often favored over the more traditional *J*-correlated experiments. In our application of this approach, NOESY spectra of selectively deuteriated protein variants were used to identify the ^1H spin systems and to obtain sequence-specific assignment. Furthermore, because spin system identification requires a high degree of internal confirmation from complementary sequence-specific assignment procedures, analysis of 2D NOESY spectra was combined with an analysis of 2D and 3D ^1H–^{15}N correlated spectra, as well as with an analysis of the 3D homonuclear TOCSY-NOESY spectrum. Analyses of the NOESY spectra with mixing times of 150 ms, coupled with an analysis of the two 3D spectra, proved to be very successful in obtaining the complete assignment of the TF1 ^1H NMR spectrum.

Of the eight leucine spin systems in each TF1 monomer, only seven could be readily identified in the NOESY spectrum of the LAS variant. Intraresidue NOEs from the NH to two or more side-chain protons and from the αH to other side-chain protons allowed the facile identification of all the leucine side chains except for Leu44. Only one intraresidue NOE from Leu44-NH to its side-chain protons, Leu44-δCH_3, was found. This significant upfield shifted resonance (at 0.34 ppm) was correlated to Leu44-γH and -βH, but correlations to Leu44-αH were missing.

The eight threonine spin systems were identified by the characteristically intense NOE correlation between each γCH_3 resonance and both the αH and βH resonances in the NOESY spectrum of the TE variant. Differentiation of αH and βH required an analysis of TOCSY data, in which the γH–βH crosspeak should be much more intense than the γH–αH crosspeak, owing to the stronger γCH_3–βH *J*-coupling. While spectral overlap between αH and βH resonances complicated the assignment of Thr4 and Thr17, NOEs between Thr4-αH, -βH and Glu5-NH allowed the identification of the Thr4 spin system. The Thr17-γH resonance was tentatively assigned based on its strong NOESY correlation to both αH and βH resonances, which form a strongly *J*-coupled proton pair with chemical shifts typical for threonine αH and βH (at 4.04 and 4.26 ppm, respectively).

Assignments of the 15 glutamic acid and glutamine residues were made largely from an analysis of the TE variant even though these residues were highly protiated in all the TF1 variants. The Glu29 spin system was the most difficult to identify due to an unusually large downfield shift of its βH while its γH was shifted upfield. The Glu29 assignments were ultimately verified as part of a series of sequential NH_i–NH_{i+1} and αH_i–NH_{i+1} NOEs, which trace the protein backbone from Val21-NH through Asp40-NH. Sequential NOEs correlating the Glu29 side-chain protons to Lys30-NH and the Phe28 side-chain protons to Glu29-NH were also clearly observed and served to verify the Glu29 assignments.

The five isoleucine residues in each TF1 monomer were identified through analyses of the NOESY spectrum of the FGIY variant. While overlap between Ile71-NH and Ile7-NH resonances limited the utility of 2D techniques, the heteronuclear 3D spectrum allowed these spin systems to be distinguished easily via the differences in their ^{15}N chemical shifts. This strategy was also used to identify one of the eight valine residues (Val19). For the remaining seven valine residues, 2D analysis of the FVSY and VIFY variants proved to be sufficient. The assignments for all isoleucine and valine resonances were confirmed with 2D spectral anslysis of the VIFY variant.

As there are only two arginine residues in each TF1 monomer, their identification in the NOESY spectrum of the PRAG variant was straightforward. The 3D-TOCSY-NOESY spectrum of fully protiated TF1 served to verify these assignments.

Both the PRAG and KMP variants were used to identify the four proline spin systems. In the NOESY spectra, both Pro52 and Pro73 displayed NOE correlations from δ, δ′H to β, β′H and γ, γ′H. However, Pro81 offered only a weak and incomplete NOE connectivity pattern, and Pro63 provided no such pattern. Moreover, while strong NOE correlations were observed between 53NH and 52αH, between 74NH and 73αH, and between 82NH and 81αH, no clear evidence was found for a similar correlation to Pro63. For all four prolines, however, strong NOEs were observed from δ, δ′H to the αH's of the preceding residues (Lys51-αH–Pro52-δ, δ′H, Asn62-αH–Pro63-δ, δ′H, Ala72-αH–Pro73-δ, δ′-H, and Lys80-αH–Pro81-δ, δ′H). This not only serves to confirm the assignments but also indicates that all four proline residues of TF1 are in the *trans* conformation.

There are 13 lysine residues each TF1 monomer. To identify their highly overlapped resonances, we relied on the 3D-TOCSY-NOESY spectrum. Because the TOCSY crosspeaks correlating Lys-δHs to εHs in the 3D spectrum were strong

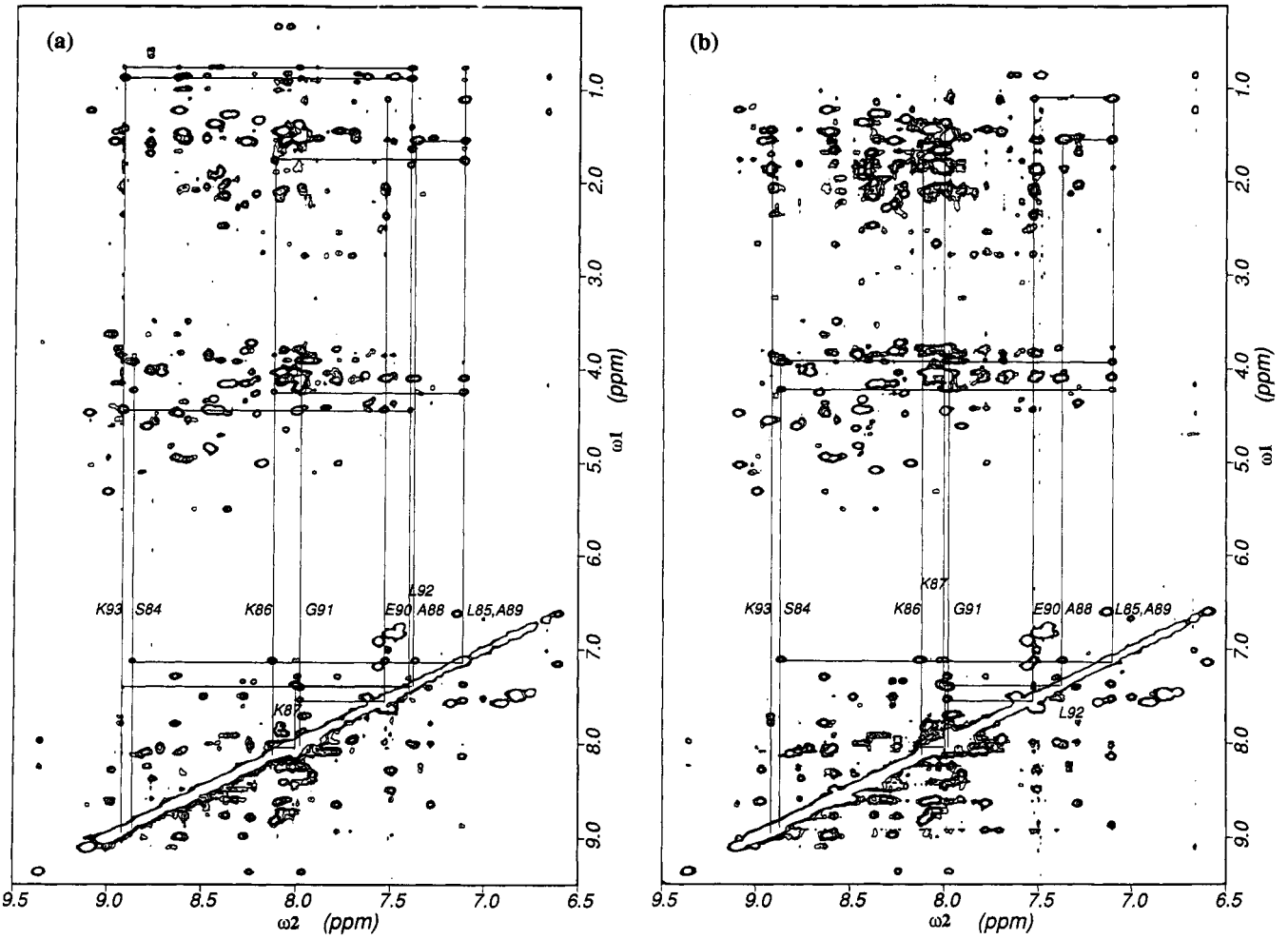

FIGURE 1: Expanded regions from the NOESY spectra (τ_m = 150 ms) of two TF1 variants, (a) LAS and (b) KMP, in H_2O buffer. Sequential NH_i–NH_{i+1} connectivities linking Ser84 to Lys93 are shown. In b, the connectivity from Leu92-NH to Lys93-NH is not seen. Intraresidue connectivities from NH to side-chain protons are indicated with vertical lines; selected interresidue connectivities are also indicated by horizontal lines.

and the characteristic chemical shift of Lys-ϵHs (at about 2.9-3.0 ppm) was readily identified, these crosspeaks were used as initial indicators of the lysine spin systems. Spin systems were then traced from ϵHs to δHs and through αH resonances in both TOCSY and NOESY planes of the 3D spectrum. All lysines, except for Lys3, were assigned in this manner, and the assignments were verified through an analysis of the heteronuclear 3D spectrum and the 2D NOESY and 2D TOCSY spectra of the KMP variant. Met24, one of two methionine residues in TF1, was also assigned through an analysis of the NOESY spectrum of the KMP variant. No resonances corresponding to the amino-terminal methionine were identified. The inherent flexibility and mobility of the amino terminus of TF1 may explain why Lys3 and Met1 resonances were not observed.

The protiated forms of the remainder of the amino acids (Ala, Asp, Asn, Gly, and Ser) were observed, to some degree, in every TF1 variant. Identification of these spin systems was based upon an analysis of all the 2D NOESY spectra. Each resonance from these residues (with the exception of Asn2) was assigned unambiguously; the assignments were verified through a series of self-consistent, intra- and interresidue NOESY crosspeak patterns. In the case of Asn2, three medium-intensity NOESY crosspeaks that appeared in almost every deuteriated variant NOESY spectrum correlated the Gly46-αH resonance to resonances at 8.37, 3.53, and 2.98 ppm. When the spin system identification was nearly complete, we assigned these resonances as Asn2-NH, αH, and -βH, respectively. Because no independent sequential connectivities can confirm this assignment, we consider this a tentative assignment. For the remainder of the Ala, Asp, and Asn spin systems, αH-βH TOCSY correlations were used to further verify the assignments.

While the five aromatic spin systems have been assigned previously (Reisman et al., 1993), the present assignment reveals that the previously published assignments for Phe28 and Phe47 were transposed. The segment of residues from 21 to 40, including Phe28, displays a continuous set of sequential NH-NH and αH-NH NOE connectivities, and each residue within this segment was unambiguously assigned. In contrast, the assignment for Phe47 was essentially derived from long-range contacts. Complete sets of sequential NOE connectivities were observed for Tyr94 and Phe96. The spin system of Phe61 was readily identified by the characteristic relaxation properties of its aromatic protons (Reisman et al., 1993).

Finally we need to mention that uncertainties among β, γ, and δ protons within one spin system, of some of the Leu, Arg, Lys, Glu, Gln, and Pro residues, do exist in their assignments. These uncertainties were caused by severe line broadening in COSY spectra and using only one TOCSY mixing time. In such situations, the assignments of these protons were based on their typical chemical shift values in random coil.

Sequence-Specific Assignment of the ^{1}H NMR Spectrum. Our use of selectively deuteriated variants placed severe limitations on our ability to identify sequential αH_i–NH_{i+1}

correlations. Although a number of αH resonances could be seen for even the most highly deuteriated variants, the intensity of resonances corresponding to the deuteriated residues was often quite weak and exhibited no sequential crosspeaks. It is well-known, however, that relatively strong NOEs can be observed for NH_i–NH_{i+1} connectivities for residues in an α-helical conformation (Wüthrich, 1986). For this reason, we relied heavily on sequential NH_i–NH_{i+1} correlations in regions we judged likely to be α-helical. For domains which do not adopt an α-helical structure, we relied on a combination of αH_i–NH_{i+1} and NH_i–NH_{i+1} correlations to produce a sequential assignment of the protein backbone. While tracing the backbone in this manner, we correlated the aliphatic spin systems of each residue to its corresponding NH resonance. Figure 1 provides examples of this assignment procedure as applied to the α-helix 3 domain of TF1 for the LAS(a) and KMP(b) variants. Sequential NH_i–NH_{i+1} connectivities for residues 84–93 are shown in the NH–NH and in the NH–aliphatic spectral regions; several intraaliphatic contacts (not shown) also verify the proximity of each sequential pair of residues.

Although its utility has not been emphasized in the literature, the ^{1}H homonuclear 3D NMR spectrum was quite useful in our assignment of the TF1 ^{1}H NMR resonances. About 20% of the ^{1}H resonances were assigned from a single 3D TOCSY-NOESY spectrum of fully protiated TF1, independent of selectively deuteriated 2D spectra. Figure 2 shows how the sequence-specific assignments were traced from residue 97 to residue 94 in the homonuclear 3D spectrum. The left panels of Figure 2 present ω3 cross-sections for the NH resonance of residues Asp96, Glu95, and Tyr94. The peaks on the "diagonals" (ω1=ω2) represent single-transfer NOE correlations to the NH and, when spectral over-lap is not too severe, indicate clear through-space contacts between proximate residues. For example, correlations can be traced from 96-NH to 96-αH, β, β′H, 95-NH, and 95-β, β′H in the ω3 cross-section of 96-NH. In the same panel, single-transfer TOCSY correlations from 96-NH to 96-αH and from 96-αH to 96-βH are identified by horizontal lines. The right panels of Figure 2 present ω3 cross-sections for the α proton resonances of the same residues. The largely redundant information obtained from these cross-sectional spectra were used to confirm earlier assignments and, in some cases, to extend assignments to other protons in amino acid side chains. Analyses could proceed from the other correlated cross-sectional resonances (i.e., βH, γH, etc.) as well. Analysis of this 3D spectrum, along with analysis of the much better resolved 2D spectra, allowed us to trace the TF1 backbone from Glu5 through Asp13, from Ser20 through Lys41, and from Gly82 through the carboxy-terminal Lys99.

Tracing the backbone from Val42 through Pro81, however, proved to be less straightforward. Yielding fewer sequential NOEs and more complex long-range NOESY interconnectivity patterns than the remainder of the protein, analysis of this segment required a more piecemeal approach. Spin systems were linked with as few as two sequential NOEs and occasionally without any αH_i–NH_{i+1} correlation.

Similarly, sequential assignments for residues Thr14–Val19 were difficult to establish. This difficulty arose because the NH resonances of these residues all cluster near 8.0 ppm (with the exception of Gln18-NH). Only a careful analysis of the 2D spectra and, in some cases, the use of resolution enhancement techniques such as the application of sharp apodization functions to the time-domain data, zero-filling, or linear prediction in the t2 dimension prior to Fourier transformation, led to the assignments presented in Table 2. In sum, the ^{1}H

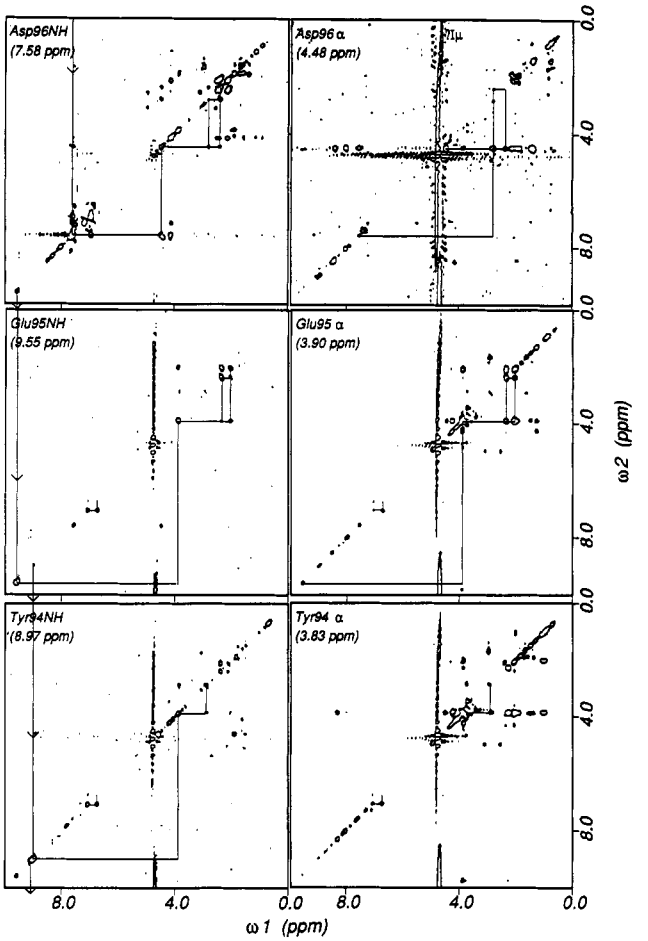

FIGURE 2: Two-dimensional cross-sections in ω3 from the homonuclear 3D TOCSY-NOESY spectrum of fully protiated TF1 in H_2O buffer. Cross-sections at the NH (left) and at the αH (right) resonance frequencies for Asp96, Glu95, and Tyr94 are shown. Intraresidue TOCSY connectivities (within individual panels) are linked by horizontal and vertical lines with sequential NOE connectivities (between adjacent panels) are traced with arrows. NOE connectivities also appear along the diagonals of the individual panels.

assignment of TF1 was largely based on analysis of the homonuclear 2D and 3D spectra. In only a select few cases was the heteronuclear 3D HMQC-NOESY required to establish a ^{1}H assignment. A complete list of proton chemical shift assignments is presented in Table 2.

Sequence-Specific Assignment of the ^{15}N Spectrum. We encountered difficulties in attempting to distinguish interresidue from intraresidue ^{1}H NOESY crosspeaks in analysis of the heteronuclear 3D NOESY-HMQC spectrum. For this reason, the ^{15}N assignments may be viewed as being determined from the previously assigned ^{1}H resonances. We were also able to use this spectrum as an aid in verifying ^{1}H NMR assignments.

Figure 3 contains a typical ^{1}H–^{15}N HSQC spectrum of TF1 in which 92 of 94 expected backbone NH resonances (99 residues less four proline residues and the amino-terminal residue) are correlated to ^{15}N resonances. Only correlations that correspond to Lys3 and to Gly60 are not observed. Several side-chain NH_2 residues from glutamine and asparagine are also observed near ω(^{15}N) = 116 ppm in the HSQC spectrum. The two most upfield-shifted ^{15}N resonances (indicated by boxes in Figure 3) are "aliased" into the displayed spectrum. Their actual ^{15}N chemical shifts, 89.80 and 89.36 ppm, have been assigned to Lys87-εNH_2 and Lys93-εNH_2 side-chain nitrogens. While most of the backbone ^{15}N assignments were determined from the 3D NOESY-HMQC spectrum, a few of

Table 2: ^{15}N and ^{1}H Chemical Shifts (ppm) of Transcription Factor 1[a]

	^{15}N	NH	αH	β, β′H	γ, γ′H	δ, δ′H	other protons
Met1	*b*						
Asn2[c]	122.27	8.37	3.53	2.98			
Lys3	*b*						
Thr4	118.36	7.92	3.85	4.24	1.26		
Glu5	125.30	8.32	3.93	1.98	2.29		
					2.53		
Leu6	127.55	8.59	4.35	2.05	1.54	0.86	
						1.00	
Ile7	123.69	8.49	3.45	1.93	0.92	0.70	0.89 (γCH$_3$)
				1.85			
Lys8	121.67	7.49	3.84	1.86	1.55	1.66	2.91 (εH)
Ala9	125.93	8.28	4.16	1.57			
Ile10	124.22	8.97	3.95	1.91	0.97	0.64	0.78 (γCH$_3$)
Ala11	129.82	8.61	3.83	1.46			
Gln12	119.62	8.10	4.04	2.09	2.42		7.57/6.84 (CONH$_2$)
				2.16	2.52		
Asp13	121.48	7.86	4.41	2.78			
Thr14	123.96	8.02	3.94	4.03	1.28		
Glu15	121.55	8.07	3.64	2.00	2.48		
				2.07			
Leu16	120.48	7.99	4.09	*b*	1.63	0.76	
						0.87	
Thr17	125.26	7.97	4.04	4.26	1.28[c]		
Gln18	123.04	8.38	4.17	2.00	2.45		7.57/6.89 (CONH$_2$)
					2.18		
Val19	121.54	7.97	3.82	1.97	0.94		
					1.05		
Ser20	122.66	7.79	4.04	3.80			
Val21	126.39	8.07	3.78	2.09	0.87		
					1.03		
Ser22	118.95	8.72	4.04	3.98			
Lys23	125.27	8.03	4.04	1.80	1.44	1.65	2.92 (εH)
Met24	123.92	8.24	3.83	2.24	*b*		
				2.31			
Leu25	124.41	8.78	4.01	1.44	1.69	0.58	
				1.58		0.65	
Ala26	126.62	8.09	4.12	1.43			
Ser27	118.99	7.80	4.02	3.81			
				3.90			
Phe28	124.58	8.93	4.17	3.23			7.02 (H2,6), 6.97 (H3.5),
				3.42			6.84 (H4)
Glu29	124.45	7.70	4.10	1.58	2.78		
Lys30	123.23	7.97	3.82	1.80	1.45	1.64	2.94 (εH)
				1.88			
Ile31	122.53	9.36	3.69	1.70	1.31	0.66	0.74 (γCH$_3$)
					1.42		
Thr32	123.65	8.25	3.72	4.05	1.08		
Thr33	119.12	7.96	3.59	4.21	1.16		
Glu34	123.98	8.41	3.92	1.90	2.30		
					2.49		
Thr35	126.99	8.00	3.78	4.10	1.11		
Val36	126.66	7.88	4.15	1.93	0.65		
					0.90		
Ala37	124.08	7.99	4.20	1.52			
Lys38	118.95	7.29	4.37	2.01	1.52	1.61	2.94 (εH)
						1.67	
Gly39	113.40	8.64	3.63				4.42 (α′)
Asp40	126.33	7.79	5.00	2.53			
				2.75			
Lys41	117.48	8.18	5.04	1.90	1.43	1.59	2.91 (eH)
Val42	124.84	9.09	4.56	1.94	0.86		
					1.04		
Gln43	124.72	8.52	4.94	1.98	2.12		7.43/6.83 (CONH$_2$)
				2.10	2.22		
Leu44	125.13	8.11	*b*	1.06	0.67	0.34	
						0.39	
Thr45	112.23	8.07	4.28	4.65	1.40		
Gly46	124.58	8.92	4.27				5.51 (α′)
Phe47	125.23	8.82	5.12	2.80			7.13 (H2,6), 6.85 (H3,5),
				2.93			6.34 (H4)
Leu48	128.87	8.63	4.78	1.91	1.67	0.94	
				2.12		1.01	
Asn49	123.82	8.05	5.32	2.66			7.14/6.60 (CONH$_2$)
Ile50	125.96	9.00	4.70	1.95	1.45	0.63	0.70 (γCH$_3$)
Lys51	127.28	9.03	5.11	1.78	1.20	1.58	2.83 (εH)
						1.42	
Pro52	–	–	5.09	1.17	1.84	3.58	
				1.56	1.90	3.72	
Val53	120.12	8.37	4.49	1.90	0.73		
					0.80		
Ala54	130.88	8.64	4.47	1.23			
Arg55	126.92	9.10	4.58	1.72	1.49	3.05	
					1.54		
Gln56	114.54	8.48	4.43	1.90	2.06		7.57/6.93 (CONH$_2$)
				1.96	2.15		
Ala57	127.68	8.39	4.16	1.27			
Arg58	122.93	8.35	4.56	1.69	1.49	3.11	
				1.77	1.54		
Lys59	124.61	8.47	4.63	1.83	1.39	1.56	2.94 (εH)
						1.67	
Gly60	*b*	7.69	3.91				4.17 (α′)
Phe61	124.64	8.46	4.89	2.81			7.04 (H2,6), 7.25 (H3,5),
							7.18 (H4)
Asp62	128.44	8.45	4.89	2.52			7.56/7.18 (CONH$_2$)
				3.06			
Pro63	–	–	*b*	2.03	1.82	3.70	
					1.90	3.78	
Gln64	120.45	8.22	4.25	2.13	2.37		
Thr65	111.98	7.34	4.29	4.33	1.08		
Gln66	121.08	8.27	3.81	2.22	2.24		7.44/6.75 (CONH$_2$)
Glu67	122.27	7.60	4.33	1.77	2.22		
				1.94			
Ala68	125.01	8.00	4.45	1.39			
Leu69	124.58	8.07	4.66	1.37	1.53	0.83	
						0.91	
Glu70	125.78	8.42	4.44	1.87	2.06		
				1.94	2.15		
Ile71	129.40	8.50	4.09	1.70	1.32	0.78	0.86 (γCH$_3$)
Ala72	133.36	8.21	4.47	1.32			
Pro73	–	–	4.25	1.68	1.28	3.44	
				1.70	1.33	3.67	
Ser74	119.76	8.66	4.83	3.87			
				3.93			
Val75	130.25	8.76	4.57	1.91	0.49		
					0.76		
Gly76	122.04	8.58	3.50				4.54 (αε)
Val77	129.07	8.60	4.40	1.61	0.62		
					0.73		
Ser78	130.58	8.97	4.60	3.63			
Val79	117.31	8.79	4.97	1.80	0.61		
Lys80	130.10	8.58	4.97	1.79	1.35	1.65	2.94 (εH)
						1.50	
Pro81	–	–	4.61	2.09	1.78	3.76	
				2.11	1.85	3.95	
Gly82	111.98	7.92	3.24				4.34 (α′)
Glu83	122.27	8.91	3.86	2.01	2.27		
				2.08	2.35		
Ser84	117.44	8.86	4.24	3.92			
Leu85	125.37	7.12	4.25	1.54	1.63	0.76	
				1.76		0.90	
Lys86	120.72	8.12	3.77	1.81	1.37	1.60	2.92 (εH)
					1.41	1.71	
Lys87	121.54	8.02	4.04	1.88	1.69	1.46	2.96 (εH)
Ala88	125.07	7.37	4.10	1.55			
Ala89	119.76	7.11	4.09	1.11			
Glu90	121.01	7.53	4.11	2.05	2.37		
				2.12	2.48		
Gly91	109.33	7.98	3.78				4.10 (α′)
Leu92	125.07	7.41	4.43	1.42	1.63	0.76	
				1.82		0.88	
Lys93	125.21	8.91	4.54	1.86	1.44	1.71	3.04 (εH)
				1.92	1.56		
Tyr94	130.55	8.96	3.80	2.80			7.01 (H2,6), 6.67 (H3,5)
				2.90			
Glu95	119.93	9.54	3.85	1.96	2.30		
				2.03	2.35		
Asp96	121.28	7.54	4.44	2.35			
				2.78			
Phe97	121.11	7.62	4.38	2.49			7.29 (H2,6), 7.41 (H3,5),
				3.13			7.31 (H4)
Ala98	128.83	7.49	4.07	0.86			
Lys99	129.49	7.65	4.15	1.81	1.35	1.65	2.96 (εH)

[a] ±0.04 ppm, at 30 °C, 400 mM sodium chloride/100 mM sodium phosphate buffer, pH 6.8. Uncertainties among β, γ, and δ protons within one spin system, of some of the Leu, Arg, Lys, Glu, Gln, and Pro residues, do exist in their assignments (see text). [b] Unassigned. [c] Tentative assignment.

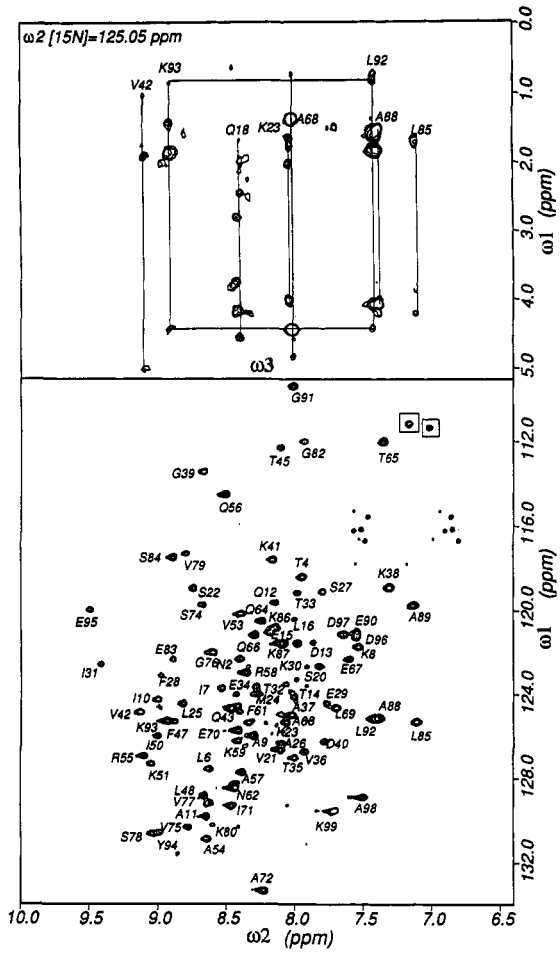

FIGURE 3: An HSQC spectrum (lower panel) and a 2D cross-section in $\omega 2$ from the heteronuclear 3D NOESY-HMQC spectrum (upper panel) of uniformly ^{15}N-labeled TF1 in H_2O buffer. The vertical lines indicate intraresidue and interresidue NOE connectivities while the horizontal lines indicate NOE connectivities from Lys93-NH to Leu92 side-chain protons. The two boxed peaks in the HSQC spectrum are aliased in the ^{15}N dimension.

the assignments (those corresponding to Leu16, Gln18, Gly46, Phe47, Lys59, and Glu70) were based on a comparison of NH chemical shifts in the 2D HSQC and in the 2D NOESY spectra. The ^{15}N chemical shift assignments are listed along with the ^1H assignments in Table 2.

Differentiation of Intermonomeric and Intramonomeric NOESY Correlations. Because TF1 is a homodimer, only one monomer is represented by the NMR spectrum. While the 2-fold symmetry proved advantageous in many aspects of the spectral analysis, making the spectra far less complex than that of a nonsymmetric 22-kDa protein, it did complicate our attempts to distinguish between those NOESY crosspeaks that represent intermonomeric contacts and those that represent intramonomeric contacts. Therefore, we used the selectively deuterated heterodimer method (Arrowsmith et al., 1990) to reliably differentiate intermonomeric and intramonomeric corrlations. In our application of this strategy, two "heterodimeric" TF1 variants, Phe*-LAS and VIFY-TE, were examined. By comparing NOESY crosspeak intensities observed in spectra of the selectively deuterated homodimers to the intensities observed in spectra of the selectivity deuteriated heterodimers, we were able to distinguish between intermonomeric NOE contacts and intramonomeric NOE contacts. Because TF1 establishes a rapid monomer–dimer equilibrium in the absence of a suitable DNA ligand (Andera et al., 1994), we treated each hybridized sample as a mixture

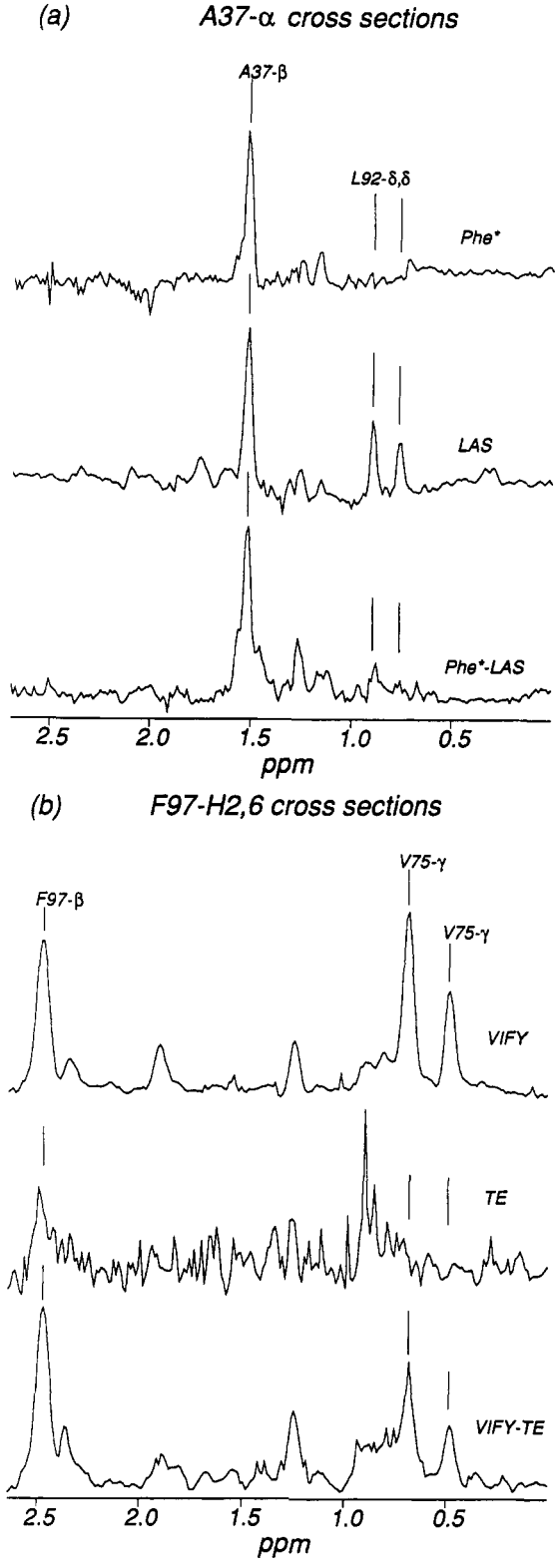

FIGURE 4: One-dimensional slices of NOESY spectra in D_2O buffer. (a) Slices at $\omega 1$ = A37-αH (top, Phe*; middle, LAS; and bottom, mixed Phe*-LAS) are shown. (b) Slices at $\omega 1$ = F97-H2,6 (top, VIFY; middle, TE; and bottom, mixed VIFY-TE) are shown. Vertical lines indicate positions at which NOE correlations with the specified resonances would be detected.

of the two homodimers and the heterodimer in a 1:1:2 molar ratio. Two examples of our analyses are shown in Figure 4.

In Figure 4a, $\omega 2$ cross-sections at the Ala37-αH resonance for three TF1 variants (Phe*, LAS, and Phe*-LAS) are shown; each spectrum exhibits an NOE correlation between the Ala37-αH and Ala37-βH resonances. Because of the higher level

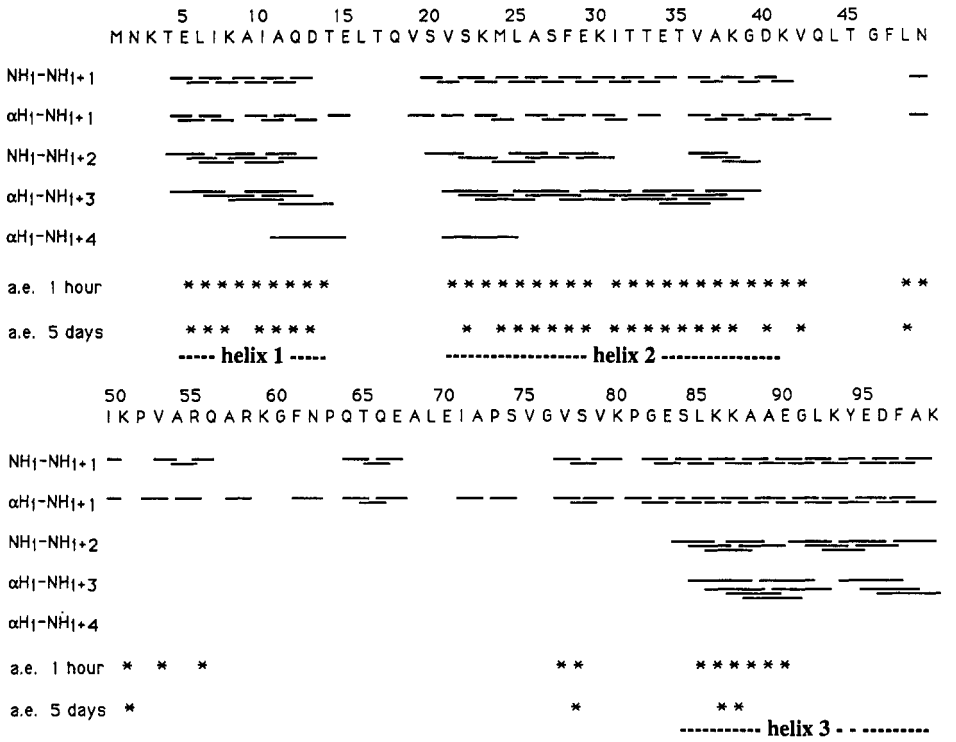

FIGURE 5: A summary of sequential NOE correlations for TF1. No information regarding the relative intensities of the correlations is provided. Those NH resonances that could be detected after exchange (a.e.) for 1 h and for 5 days in D_2O buffer are marked with asterisks.

of amino acid-specific deuteriation in the Phe* variant, this crosspeak has a lower absolute intensity in the Phe* cross-section. However, because we are concerned with interresidue correlations involving Ala37-αH, the three cross-sections are normalized against the intensity of this intraresidue crosspeak. In the LAS cross-section, representative of a protein in which all Ala and Leu side chains are protiated, the two upfield crosspeaks have been assigned as NOEs correlating Ala37-αH to Leu92-δH and δ′H. For the hybridized Phe*-LAS, these two crosspeaks exhibit dramatically reduced intensities. On the basis of the 1:1:2 molar ratio of LAS, Phe*, and Phe*-LAS dimers, a 0.25:1 intensity ratio between the crosspeaks in the Phe*-LAS and LAS would be expected if the Ala37-Leu92 contact were intermonomeric, while a 0.75:1 intensity ratio would be expected if the contact were intramonomeric. Integration of these crosspeaks showed a strong agreement to the predicted intensity ratio for an intermonomeric contact, indicating that the Ala37 residue from each monomer is proximate to the Leu92 residue from its paired monomer.

From the spectra in Figure 4b, a similar analysis for the hybridized variant VIFY-TE led to the conclusion that the correlations between Phe97-H2,6 and Val75-γ, γ′H are also intermonomeric. Other NOESY crosspeaks were determined as arising from either intermonomeric or intramonomeric contacts in this manner, and these contacts are listed in Table 3. However, because it would be quite costly and time consuming to conduct a similar analysis for every NOE correlation that we have observed, we have limited our analysis to those correlations which we believe to lie at or near the dimeric interface and to those correlations involving resonances that are not severely overlapped with other resonances. While our choice of correlations is based on earlier modeling studies of TF1 (Geiduschek et al., 1990), these experiments do provide independent methods to determine whether specific secondary structure domains of one monomer interact with a specific domain of the paired monomer.

Analysis of TF1 Secondary Structure. A summary of sequential NH-NH and αH-NH NOE correlations and the slowly exchanging amide protons of TF1 is shown in Figure 5. For the segments spanning residues 4–13 and 21–40, these correlations are quite evident, and the amide protons in these segments also exhibit slow exchange with solvent deuterons indicating well-structured α-helical domains. For the segment spanning residues 84–99, most of backbone amide protons exhibit relatively rapid exchange with solvent deuterons, although the NH-NH connectivity patterns indicate a similar structure to that observed in segments 4–13 and 21–40. In the domain spanning residues 41–83, no sequential αH_i–NH_{i+3} NOEs indicative of an α-helical structure were found. Furthermore, the number of slowly exchanging amide protons in this region is far less than the number seen in the α-helical segments.

Three α-helical domains (Figure 5) are evident in the TF1 structure. All of them are characterized by a well-defined pattern of sequential NOEs and in addition, helices 1 and 2 exhibit slow backbone amide proton exchange. Helix 3 begins at Leu85 and appears to continue through the carboxy-terminus of the protein, Lys99. If the sequential NOEs between αH_i and NH_{i+3} are analyzed (Wüthrich, 1986), however, a distortion, or "kink", can be detected between Leu92 and Lys93. This was initially indicated by the presence of a relatively weak Leu92-NH to Lys93-NH NOE. Furthermore,

Table 3: Summary of Inter- and Intramonomeric NOEs from NMR Experiments

hybridized variants	NOESY crosspeaks	inter/intramonomeric NOEs
Phe*-LAS	Ala37-αH–Leu92-δ,δ′H	inter
	Phe28-H3,5–Phe47-H2,6	inter
	Ala57-β–Ala72-αH	intra
	Phe47-αH–Leu85-δ,δ′H	intra
	Leu44-δH–Phe47-αH	intra
VIFY-TE	Thr33-αH–Ala88-γH	inter
	Tyr94-H2,6–Val75-γH	inter
	Phe97-H2,6–Val75-γ,γ′H	inter
	Phe97-H3,5–Val75-γ,γ′H	inter
	Val53-γ,γ′H–Ser78-βH	intra

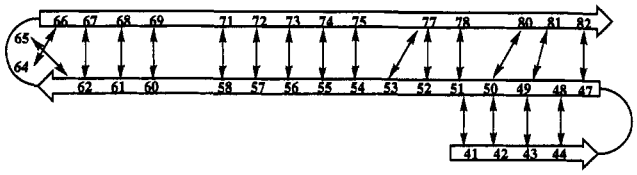

FIGURE 6: Proposed topology for the β-sheet domain and the β-ribbon domain of TF1.

the absence of sequential NOEs correlating Gly91-αH to Tyr94-NH, Leu92-αH to Glu95-NH, and Lys93-αH to Glu96-NH confirmed that this region of the protein does not adopt a standard α-helical structure. Analysis of backbone amide proton exchange over 1 h also serves to divide helix 3 into two characteristic segments. In the Leu85-Glu90 segment, amide proton exchange occur over a period of hours and in the other segment (Gly91–Lys99), amide proton exchange occurs over a period of minutes.

The segment from Lys41 through Gly82 is characterized by weak αH_i–NH_{i+1} and NH_i–NH_{i+1} NOE correlations and by relatively strong long-range NH–αH and αH–αH NOE correlations. These correlation patterns are indicative of antiparallel β-sheet or β-ribbon secondary structural units. The experimentally-derived topology of this segment is shown in Figure 6. The lines connecting each residue to others represent crosspeaks observed in 2D NOESY spectra. The conclusion that this region adopts a series of β structures was also confirmed by an analysis of the chemical shift values of the αH resonances. Wishart and co-workers (1991) have demonstrated that the αH resonances for domains that adopt β structures are typically shifted downfield from their "random coil" chemical shifts by approximately 0.4 ppm.

The largest looped region in TF1 is located from Glu15 through Val19, the protein segment that joins helix 1 to helix 2. The absence of sequential NOEs linking the residues in this region further suggested that it does not adopt a standard secondary structure. Also, no assignments were obtained for resonances of either Met1 or Lys3 and only a tentative assignment was made for Asn2. While this result may suggest that the amino terminus of TF1 is poorly structured (yielding no interresidue correlation to the remainder of the molecule), long-range NOE correlations between Asn2 and Gly46 side chains suggest that the amino terminus may be anchored against the body of the protein. Further study of this region is required to make a definitive statement in regard to the conformational stability or the structure of this region.

The chemical shift values of the αHs were analyzed according to the observations of Wishart and co-workers (1991). The resulting distribution of chemical shifts relative to each residue's "random coil" chemical shift, $\Delta\delta$ indicated that, in TF1, the $\Delta\delta$ values for the vast majority of residues corresponded to upfield shifted resonances for the α-helices and downfield shifted resonances for the β-sheet and β-ribbon regions, a phenomenon consistent with previous studies (Wishart et al., 1991). Of special interest, however, are the $\Delta\delta$ values of Lys93-αH, located on helix 3, and of Gln66-αH, located near the apex of the β-ribbon domain. The Lys93-αH resonance is not shifted upfield of its "random coil" value, consistent with the location of Lys93 in an α-helical domain, but nearly 0.3 ppm downfield. In contrast, the Gln66-αH resonance is displaced upfield from its random coil value, although Gln66 appears to be part of the β-ribbon. These chemical shift effects may be indicative of unusual backbone conformations.

The TF1 Global Fold. A set of interresidue NOE correlations for TF1 is presented below the diagonal in Figure 7. Strong sets of long-range correlations link the Thr4-Asp13 segment to the Ser20-Asp40 segment, Thr35-Asp40 to Asp40-Ile50, Ile50-Asp62 to Pro63-Ser78, Thr35-Val42 to Leu85-Leu92, and Val53-Ser78 to Lys87-Ala98. Figure 7 also presents, above the diagonal, those interresidue, interproton distances of less than 4.0 Å that appear in our earlier model based on the crystal structure of the *Bst.* HU protein, which has only 90 amino acid residues per monomer. While it is clear that the global folding pattern of TF1 resembles that of *Bst.* HU, differences do exist. In particular, the relative orientations of helix 2 and the β-sheet domain near Leu48 seem to differ, as do the intramonomeric interactions between helices 1 and 2 at Val21 and Leu25. Furthermore, and perhaps more importantly, the TF1 helix 3 domain exhibits a number of interactions that are not seen in the *Bst.* HU. We find a number of intermonomeric contacts that link helix 3 to the β-ribbon arm and to the carboxy-terminal region of helix 2 of the paired monomer.

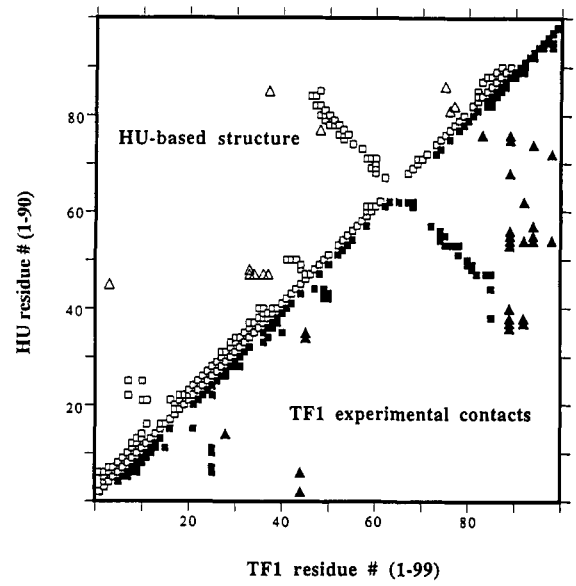

FIGURE 7: Through-space correlations between any two residues of TF1 (below diagonal; observed NOE contacts) and of the *Bst.* HU-based model (above diagonal; interproton distances of less than 4.0 Å). α-Helical characteristics (i to i+3 contacts) are seen for both TF1 and the model over residues 4–40 and 84–90. An extended β-ribbon is seen for both at residues 50–84. Intramonomer and intermonomer contacts are represented by squares and triangles, respectively.

DISCUSSION

The structure determination for relatively large proteins (with molecular weights in excess of 20 kDa) by NMR techniques is often quite challenging (Wagner, 1993). However, the use of NMR techniques is nonetheless attractive if the structure cannot be determined by crystallographic techniques, as has been the case for TF1. Multidimensional NMR techniques coupled with ^{15}N and/or ^{13}C labeling have proven quite promising and productive (Clore & Gronenborn, 1991; Bax & Grzesiek, 1993), and techniques that rely on deuterium labeling do offer significant advantages (LeMaster & Richards, 1988; Arrowsmith et al., 1990). In our current study, selective deuterium labeling proved to be an effective approach to the problem of assigning the proton NMR spectrum of this 22-kDa protein. We have assigned the proton resonances by analyzing six deuteriated TF1 variants (the VIFY and Phe* variants were used exclusively to differentiate intramonomeric from intermonomeric contacts) with a combination of 2D and 3D, homonuclear and heteronuclear NMR

techniques. We were also able to obtain ^{15}N assignments for 92 of the 94 backbone amide protons, relying primarily on the proton assignment and 3D heteronuclear techniques. Thus, the limitations introduced by spectral overlap and spin diffusion were overcome with an integrated analysis of several selectively ^{2}H-labeled variants.

It should, however, be pointed out that biosynthetic ^{2}H labeling did present an undesired effect: in each variant, more residues were protiated than would be predicted if it were assumed that only the protiated residues supplied to the system would be protiated in the purified TF1. Generally, this means that the experimental percent protiation tended to reflect the tabulated "max %" values more than they reflected the "min %" values (Table 1), indicating that there was a high degree of biosynthetic interconversion of protiated amino acids, even in the auxotroph DL39(pTF1X). While this effect could potentially undermine the analytic value of ^{2}H-labeling schemes in the study of relatively large proteins, it did not severely hamper our analysis of TF1, and in some cases it actually facilitated the sequential assignment of the protein backbone.

Previously, we proposed a model for TF1 that was based largely on the crystal structure of the *Bst.* HU protein (Geiduschek et al., 1990; Reisman et al., 1993). In this model, helices 1 and 2 from both monomers form a rigid "scaffold" for the dimeric core which is, in turn, covered by two separate three-stranded β-sheets. From each of these β-sheets, a β-ribbon extends out from the core. The present study suggests that TF1 adopts a folding pattern similar to that of the HU-based model of TF1. In particular, the existence of large segments of helices 1 and 2 from that model is evidenced by the αH_i–NH_{i+3} NOE connectivity patterns in TF1 (Figure 5). Furthermore, the short-distance NOE connectivities in both the β-sheet and the β-ribbon domains indicate that these domains adopt the specific antiparallel β-structures predicted by the *Bst.* Hu-based model of TF1 (Figure 6).

On close examination, however, the NMR data reveal details of the TF1 structure that are not consistent with the earlier model. On the basis of the NMR data, α-helix 1 appears to be shorter by two residues and the three amino-terminal residues (Met1, Asn2, and Lys3) appear not to adopt a helical conformation. Helices 1 and 2 comprise the most rigid domains of the protein, while the β-sheet domain is less stable, as indicated by its predominantly rapidly exchanging backbone NH protons (Figure 5) and its unusually weak sequential NOEs. The domain that exhibits the fewest and least intense NOESY connectivities and the most rapidly exchanging backbone NHs is the β-ribbon; these observations suggest that this is the most flexible structural domain in TF1. This domain also exhibited that highest *R* values in the *Bst. HU* crystal structure (White et al., 1989; Tanaka et al., 1984).

Most interestingly, this study confirms that the nine-amino-acid tail of TF1 adopts an α-helical conformation, as had been previously suggested (Geiduschek et al., 1990). This helix (labeled helix 3), however, displays NMR features different from those observed for α-helices 1 and 2 in the dimeric core. Primarily, the backbone of helix 3 is less stable than those of the other helices, as evidenced by its relatively rapid NH exchange (Figure 5). Secondly, the numerous NOE contacts observed between helix 3 and the β-ribbon arm domain suggest that helix 3 cannot adopt a standard α-helical conformation *and* represent a continuation of the carboxy-terminal α-helix observed in the *Bst.* HU crystal structure. Thirdly, this observation is supported by the unusual chemical shift of Lys93-αH, which suggests that the protein backbone near this residue may be "kinked" as observed in helix D of Parvalbumin (Padilla et al., 1988). The distortion in TF1 may be required to promote contacts between helix 3 and the β-ribbon arm. Furthermore, a number of studies have revealed that helix 3 is, in fact, essential for DNA binding (Sayre & Geiduschek, 1988; Andera & Geiduschek, 1994). The unusual structural characteristics of helix 3, along with its essential biological role, make it an ideal focus for further more detailed structural studies.

The β-ribbon domain, long proposed to function as a DNA-binding arm for the DBPII (Tanaka et al., 1984; White et al., 1989), is and will be the most difficult region of TF1 in which to assign a specific structure. While this difficulty may be a result of inherent flexibility, the analysis will ultimately be limited by the domain's relatively few interresidue NOE contacts. We have, however, uncovered several important structural features for this region of the protein. We have determined that the prolyl amide bond of Pro63 adopts an *trans* conformation, a structural feature that could not be determined from analogy to the *Bst.* HU crystal structure (White et al., 1989). Also, because the Gln66-αH resonance is shifted upfield by 0.5 ppm from its "random-coil" value (Wishart et al., 1991), we propose that a sharp backbone distortion may exist at this residue and that this residue may lie at the extreme of the β-ribbon arm, a region that was unresolved in the *Bst. HU* crystal structure.

From the NMR data that are presented here, we have obtained over 1200 NOE-derived interproton distance constraints, 37 hydrogen-bonding constraints, and 51 dihedral angle constraints for each monomer. These data, along with the further analyses of intermonomeric NOESY contacts through the double half-filter technique (Otting & Wüthrich, 1989), will provide the basis for a series of restrained molecular dynamics calculations and for the determination of a more refined structure for TF1 in solution. That work will be presented in a forthcoming paper.

ACKNOWLEDGMENT

We thank Dr. André Padilla for providing us with his expert advice and help in carrying out the homonuclear 3D experiment. We also thank Dr. Ladislav Andera for his outstanding interest in the progress of this work and for his help and advice in the preparation of ^{2}H-labeled TF1 variants from the auxotroph DL39 and Christian Lecou for supplying other isotope-labeled TF1 variants. Finally, we thank Dr. Daniel Dauzonne (Institut Curie, Paris, France) for providing a sample of DL-[2,4-$^2H_2$]phenylalanine.

REFERENCES

Andera, L., & Geiduschek, E. P. (1994) *J. Bacteriol. 176*, 1364–1373.

Andera, L., Schneider, G. J., & Geiduschek, E. P. (1994) *Biochimie* in press.

Archer, S. J., Baldisseri, D. M., & Torchia, D. A. (1992) *J. Magn. Reson. 97*, 602–606.

Arrowsmith, C. H., Pachter, R., Altman, R. B., Iyer, S. B., & Jardetzky, O. (1990) *Biochemistry 29*, 6332–6341.

Barry, S., & Hawley, D. K. (1991) *Cell 67*, 1231–1240.

Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355–360.

Bax, A., & Grzesiek, S. (1993) *Acc. Chem. Res. 26*, 131–138.

Clore, G. M., & Gronenborn, A. M. (1991) *Science 252*, 1390–1399.

Craig, N. L., & Nash, H. A. (1986) *Cell 39*, 707–716.

Dri, A. M., Rouvière-Yaniv, J., & Moreau, P. L. (1991) *J. Bacteriol. 173*, 2852–2863.

Flamm, E. L., & Weisberg, R. A. (1985) *J. Mol. Biol. 183*, 117–128.
Friedman, D. I. (1988) *Cell 55*, 545–554.
Geiduschek, E. P., Schneider, G. J., & Sayre, M. H. (1990) *J. Struct. Biol. 104*, 84–90.
Härd, T., Hsu, V. L., Sayre, M. H., Geiduschek, E. P., Appelt, K., & Kearns, D. R. (1989a) *Biochemistry 28*, 396–406.
Härd, T., Sayre, M. H., Geiduschek, E. P., & Kearns, D. E. (1989b) *Biochemistry 28*, 2813–2819.
Kay, L. E., Marion, D., & Bax, A. (1989) *J. Magn. Reson. 84*, 72–84.
Kellenberger, E. (1988) *Biophys. Chem. 29*, 51–62.
Kim, S.-H. (1992) *Science 255*, 1217–1218.
LeMaster, D. M., & Richards, F. M. (1988) *Biochemistry 27*, 142–150.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.
Muchmore, D. C., McIntosh, L. P., Russell, C. B., Anderson, D. E., & Dahlquist, F. W. (1989) *Methods Enzymol. 177*, 44–73.
Nash, H. A., & Granston, A. E. (1991) *Cell 67*, 1037–1038.
Oshkinat, H., Cieslar, C., Gronenborn, A. M., & Clore, G. M. (1989) *J. Magn. Reson. 81*, 2120–220.
Otting, G., & Wüthrich, K. (1989) *J. Magn. Reson. 85*, 586–594.
Padilla, A., Cavé, A., & Parello, J. (1988) *J. Mol. Biol. 204*, 995–1017.
Pettijohn, D. E. (1988) *J. Biol. Chem. 263*, 12793–12796.
Plateau, P., & Gueron, M. (1982) *J. Am. Chem. Soc. 104*, 7310–7311.
Reisman, J. M., Jariel-Encontre, I., Hsu, V. L., Parello, J., Geiduschek, E. P., & Kearns, D. R. (1991) *J. Am. Chem. Soc. 113*, 2787–2789.
Reisman, J. M., Hsu, V. L., Jariel-Encontre, I., Lecou, C., Sayre, M. H., Kearns, D. R., & Parello, J. (1993) *Eur. J. Biochem. 213*, 865–873.
Rosen, M. K., Michnick, S. W., Karplus, M., & Schreiber, S. L. (1991) *Biochemistry 30*, 4774–4789.
Sayre, M. H. (1989) Ph.D. Thesis, University of California, San Diego.
Sayre, M. H., & Geiduschek, E. P. (1988) *J. Virol. 62*, 3455–3462.
Sayre, M. H., & Geiduschek, E. P. (1990) *J. Mol. Biol. 216*, 819–833.
Shaka, A. J., Barker, P. B., & Freeman, R. (1985) *J. Magn. Reson. 64*, 547–552.
Sklenar, V., & Bax, A. (1987) *J. Magn. Reson. 74*, 469–479.
Somers, W. S., & Phillips, S. E. V. (1992) *Nature 359*, 387–393.
Srinivasan, P. R., & Lichter, R. L. (1977) *J. Magn. Reson. 28*, 227–234.
Tanaka, I., Appelt, K., Dijk, J., White, S. W., & Wilson, K. S. (1984) *Nature 310*, 376-381.
Tate, S., Masui, Y., & Inagaki, F. (1991) *J. Magn. Reson. 94*, 394–400.
Vuister, G. W., Boelens, R., Padilla, A., Kleywegt, G. J., & Kaptein, R. (1990) *Biochemistry 29*, 1829–1839.
Wagner, G. (1993) *J. Biomol. NMR 3*, 275–385.
White, S. W., Appelt, K., Wilson, K. S., & Tanaka, I. (1989) *Proteins: Struct. Funct. Genet. 5*, 281–288.
Wishart, D. S., Sykes, B. D., & Richards, F. M. (1991) *J. Mol. Biol. 222*, 311–333.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Yang, C. C., & Nash, H. A. (1989) *Cell 57*, 869–880.
Zhao, D., Arrowsmith, C. H., Jia, X., & Jardetzky, O. (1993) *J. Mol. Biol. 229*, 735–746.