^{1}H Resonance Assignment and Secondary Structure Determination of the Dimerization Domain of Transcription Factor LFB1

Annalisa Pastore,*,‡ Raffaele De Francesco,‡,§ Gaetano Barbato,∥ Maria Antonietta Castiglione Morelli,∥ Andrea Motta,⊥ and Riccardo Cortese‡,§

European Molecular Biology Laboratory, Meyerhofstrasse 1, Postfach 10.2209, D-6900 Heidelberg, Federal Republic of Germany, Istituto di Chimica, via Mezzocannone 4, I-80100 Napoli, Italy, and Istituto per la Chimica delle Molecole di Interesse Biologico, CNR, I-80072 Arco Felice, Italy

Received June 29, 1990; Revised Manuscript Received September 6, 1990

ABSTRACT: We have started the structure determination of the dimerization domain of LFB1 in solution by nuclear magnetic resonance in order to elucidate the way that the LFB1 protein dimerizes and then interacts with DNA. A 32 amino acid peptide was synthesized, and full assignment of the NMR resonances in acidic solution was achieved. The secondary structure determination is presented here. Three structurally distinct regions can be distinguished. The N-terminal region from residues 1 to 6 is extended. Two helical regions span from residues 7 to 18 and from 23 to 32. The absence of dipolar effects involving residues more than four positions apart in the sequence excludes the possibilities both of a four-helix bundle formed by two hairpins and of an antiparallel dimer; the domain must therefore be arranged as a parallel dimer formed by kinked monomers. This structural solution presents important differences from the leucine zipper-type structure observed in other transcriptional activators. Although further studies are still necessary to determine the 3D structure of the peptide, we can exclude the possibility of a coiled-coil structure.

Transcriptional activator proteins are often constructed from smaller structurally independent domains of size suitable for nuclear magnetic resonance (NMR)[1] studies. NMR techniques have already been applied successfully to the structure determination of small proteins and of independent subdomains of larger proteins (Wuethrich, 1986).

In the preceding paper, we reported a characterization of the thermodynamic properties of a 32-residue peptide containing the amino acid sequence of the N-terminal region of the LFB1 protein (De Francesco et al., 1991). The stability of this peptide is comparable with that of other much larger globular proteins and of the dimeric Arc repressor (Ahmad & Bigelow, 1982; Pace, 1975; Bowie & Sauer, 1989). Together with the evidence arising from mutagenesis studies (Nicosia et al., 1990) we concluded, therefore, that this domain folds as an independent module and represents the dimerization domain of the full protein.

We present here an NMR study of the same synthetic peptide (B1-Dim). Full NMR spectrum assignment was achieved as an essential prerequisite for further studies. From it we could determine the secondary structure of the peptide.

Materials and Methods

The peptide was synthesized and purified by Dr. H. Gausepohl of the Chemical Instrumentation Program at the European Molecular Biology Laboratory.

The NMR measurements were carried out using 4–7 mM samples in 90% H_2O/10% D_2O solutions. Unless otherwise specified, the pH was not adjusted after dissolving the peptide in solution (pH 2.7). Preliminary NMR studies were carried out to find conditions for optimal resolution. After an initial screening to determine the effect of the temperature, all of the 2D spectra were acquired at 26 °C. No influence from the ionic strength was observed up to 150 mM (physiological) sodium chloride concentration.

The NMR spectra were recorded on a 500-MHz AM-Bruker spectrometer. Some of the spectra were also recorded at 500 and 600 MHz at the Varian and Bruker companies in Darmstadt and Karlsruhe, respectively. 2D spectra were recorded in the phase-sensitive mode (Marion & Wuethrich, 1986) with and without water proton frequency irradiation. A rather substantial loss in sensitivity was observed when using water saturation, so the 1-1 pulse sequence was used in all of the following experiments. HOHAHA spectra (Bax & Davis, 1985) were measured by using the MLEV-16 composite pulse cycle (Rance, 1987) for the generation of the spin-lock field of $\gamma B_2 = 10 - 12$ kHz with mixing times between 35 and 120 ms. Mixing times of 50, 100, 150, 200, and 300 ms were used in the NOESY experiments (Jeener et al., 1979; Macura et al., 1981). Data were processed on a Bruker X32 station using the UXNMR program for data acquired on Bruker spectrometers. The AURELIA program was then used for displaying and plotting spectra. Varian data were processed by the Darmstadt branch of the company.

Slowly exchanging protons were identified by first lyophilizing a fully protonated sample in H_2O and subsequently recording a NOESY spectrum of this sample immediately after dissolution in D_2O. After a period of 24 h, nearly all of the amide protons were fully exchanged; 60% of the resonances were still observable in the NOESY spectrum.

Results

Protein Stability. Preliminary 1D studies were carried out to obtain information about the protein stability for comparison with circular dichroism data (De Francesco et al., 1991). No significant changes of the NMR spectrum appearance were observed by varying the temperature in the range 26–66 °C.

* To whom correspondence should be addressed.
‡ European Molecular Biology Laboratory.
§ Present address: IRBM, Istituto di Ricerche di Biologia Molecolare, 00040 Pomezia, Italy.
∥ Istituto di Chimica.
⊥ Istituto per la Chimica delle Molecole di Interesse Biologico.

[1] Abbreviations: NMR, nuclear magnetic resonance; DQF-COSY, double quantum filter correlation; HOHAHA, homonuclear Hartman-Hahn; NOESY, 2D nuclear Overhauser effect spectroscopy; 2D and 3D, two and three dimensional; the one-letter amino acid code is used so that M1 means methionine-1 etc.

0006-2960/91/0430-148$02.50/0

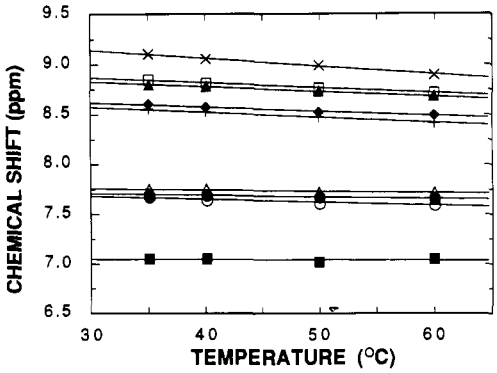

FIGURE 1: Temperature dependence of the amide proton chemical shifts of B1-Dim.

This agrees well with the circular dichroism results. From them, we can expect thermal denaturation to occur at the peptide concentrations necessary for NMR studies, at an apparent melting temperature higher than 90 °C. This makes it impossible to follow complete denaturation by NMR.

Further information about the protein stability can be obtained from the 1D spectrum by looking at the amide protons. Labile protons can be used as a probe for conformational stability. Experiments as a function of pH and temperature and the proton/deuterium exchange rates show that B1-Dim amide protons are remarkably stable. In a peptide, amide protons are not detectable at physiological pH unless otherwise stabilized, because the NMR peaks are broadened by slow exchange with the solvent. More than 50% of the B1-Dim amide protons are still perfectly observable at 36 °C and pH 7.

Similarly, the resistance of an amide proton to exchange with deuterons in D_2O solution is an indication of hydrogen bonding and/or low accessibility to the solvent. In the absence of tertiary structure, peptides of a similar size show almost instantaneous exchange. The proton/deuterium exchange rate of almost all of the amide protons is slow (vide infra) so that 60% of the resonances can still be detected 18 h after dissolution.

Figure 1 shows a plot of some of the amide proton chemical shifts (chosen among the most isolated resonances) vs temperature. Temperature coefficients are very small for all the amide resonances as expected for residues either hydrogen bonded and/or buried from the solvent.

NMR Assignment. Proton resonances in B1-Dim are generally broader than expected for a typical peptide of the same size. Only one very sharp peak of the methyl group of M1 at 2.08 ppm could be distinguished clearly from the other resonances. The first residue is predictably flexible, with the side chain in free rotation.

The interpretation of the spectra was complicated by the presence of several identical amino acid types (e.g., the palindromic region ELLAALLE) and by the absence of aromatic residues which usually bring about dispersion of the chemical shifts. Sequence-specific proton NMR assignment was obtained by combining the strategies suggested by Englander and Wand (1987) and by Wuethrich (1986).

Sequential assignment was completed by concatenation of HN–HN and Hα–HN sequential NOESY connectivities. Unfortunately, several interruptions in Hα–HN sequential connectivities were caused by peak overlaps. As a starting point for sequential assignment, unique residues such as V2, T10, and the two lysines K4 and K23 were considered. HOHAHA spectra, recorded at two mixing times (60 and 120 ms), were used to detect direct single and relayed scalar connectivities and, together with DQF-COSY and NOESY spectra, to identify amino acid spin systems. The long-range connectivities between the ε-protons of lysines and their α-protons observed in the HOHAHA spectra were very useful in connecting the side chains of these residues to their amide protons. The list of the assignments is shown as Table I.

A visual comparison of the HN–aliphatic region of the NOESY and the HOHAHA experiments shows the side chains of the leucine residues to be rigid, and locked in one conformation (Figure 2). This behavior, already observed in the GCN4 leucine zipper peptide (Saudek et al., 1990), seems to be a common feature of residues locked in dimeric interfaces. Very rigid groups are completely invisible in HOHAHA experiments (mixing times from 35 to 120 ms) but strongly present in NOESY experiments. All of the leucine δ-protons display the same behavior, except for that in L21. Only for this residue, in fact, it is possible to observe relay peaks from the HN up to the δ-protons in the HOHAHA experiment. Conversely, no NOESY peaks other than the intramolecular ones involve the same amide proton (L21 is identified by an arrow in Figure 2).

Identification of the Secondary Structure. Clear indications of well-defined secondary structure are already evident in the 1D spectra. B1-Dim resonances are noticeably shifted from random-coil positions, suggesting the presence of tertiary structure.

The chemical shift of the α resonances is sensitive to the dihedral angle ψ between the α-proton and the carbonyl. Plots of the difference between the chemical shifts of the α-protons and those in a random coil of the same sequence vs the position along the peptide sequence are a simple and effective way to relate chemical shifts to features of secondary structure (Clayden & Williams, 1982; Pastore & Saudek, 1990). After data smoothing, helices always show a positive shift from the random-coil values, whereas β sheets present negative shifts. Kinks in helices can also be observed as breaks in the positive behavior. For comparison, Figure 3A,B shows plots for B1-Dim and for the dimerization domain of GCN4 (leucine zipper domain; Oas et al., 1990; Saudek et al., 1990). In GCN4, a long uninterrupted helix can be seen, but a break in the helix is detectable for B1-Dim. A similar profile can be observed from a Chou–Fasman secondary structure prediction (Figure 3C; Chou & Fasman, 1977).

$J_{HN-H\alpha}$ coupling constants could not be measured from 1D experiments because the peaks were too broad and there was not enough digital resolution. From the DQF-COSY, they are thought to be smaller than 5 Hz along the whole chain.

2D NOESY experiments confirm a definite 3D structure and the presence of helical conformation involving at least 21 residues in the chain (other connectivities might be hidden under the diagonal). The different successions of several intense sequential NOESY cross-peaks between amide protons are indicative of regions with helical secondary structure. It is possible to define chains of up to 12 sequential residues linked by HN–HN NOEs (see Figure 4). The chains terminate at frequency degeneracies of the backbone amide protons.

A survey of the observed NOE is shown in Figure 5. Connectivities typical of an extended structure [very intense Hα–HN(i,i+1) peaks and rather weak Hα–HN(i,j) peaks] were detected between residues 1 and 6. However, this region is not flexible. A number of interresidue NOEs and the unusual difference in chemical shift of the β and β′ protons of S6 confirm that this region is structured. In NOESY experiments with long mixing times (200 and 300 ms), HN–

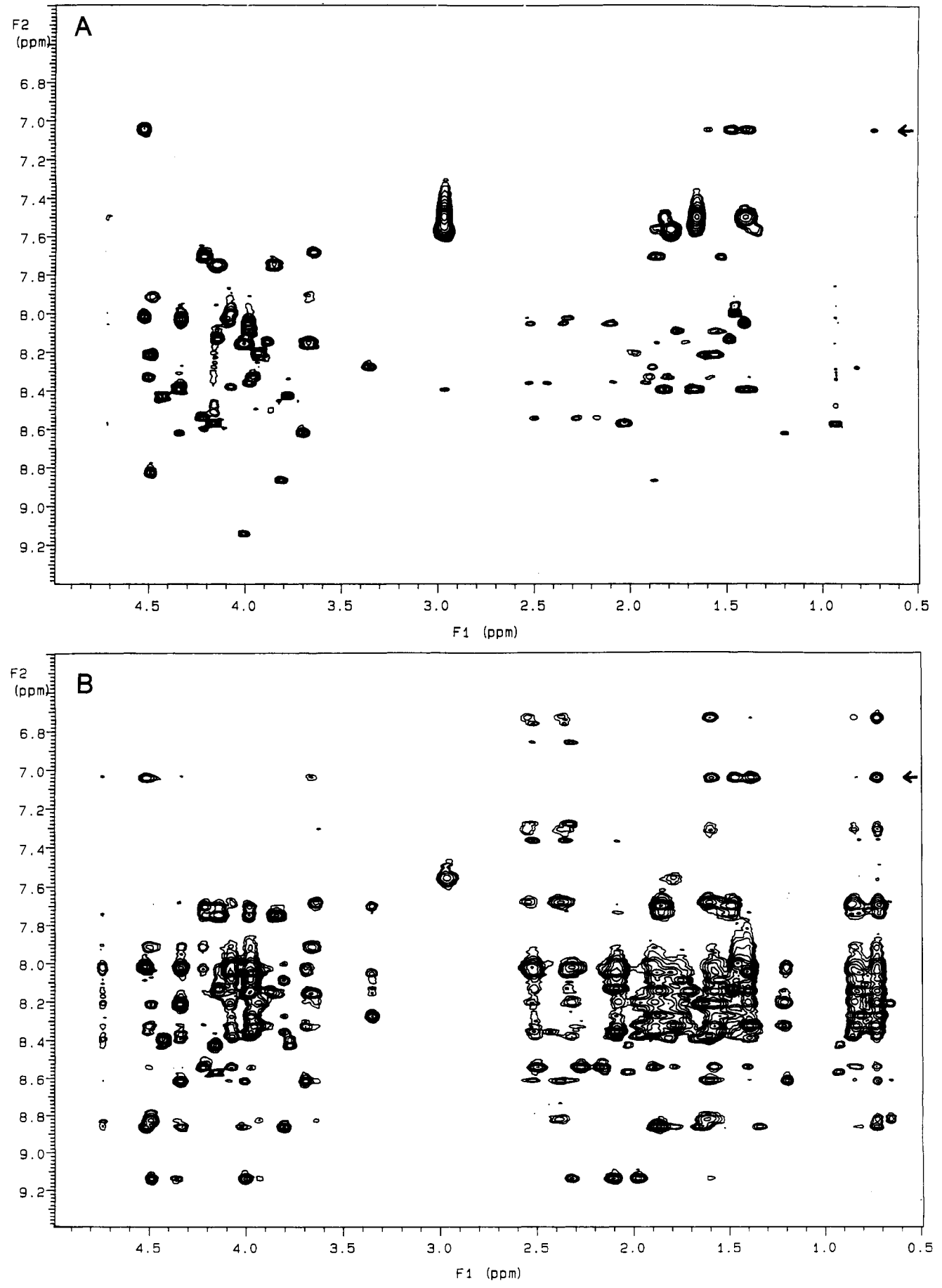

FIGURE 2: Comparison between the HN–aliphatic region of the 60-ms mixing time HOHAHA spectrum of B1-Dim (A) and the corresponding region of a NOESY spectrum with 150-ms mixing time (B). The two spectra were recorded at 26 °C on a 500-MHz Varian spectrometer. An arrow identifies L21 (see text).

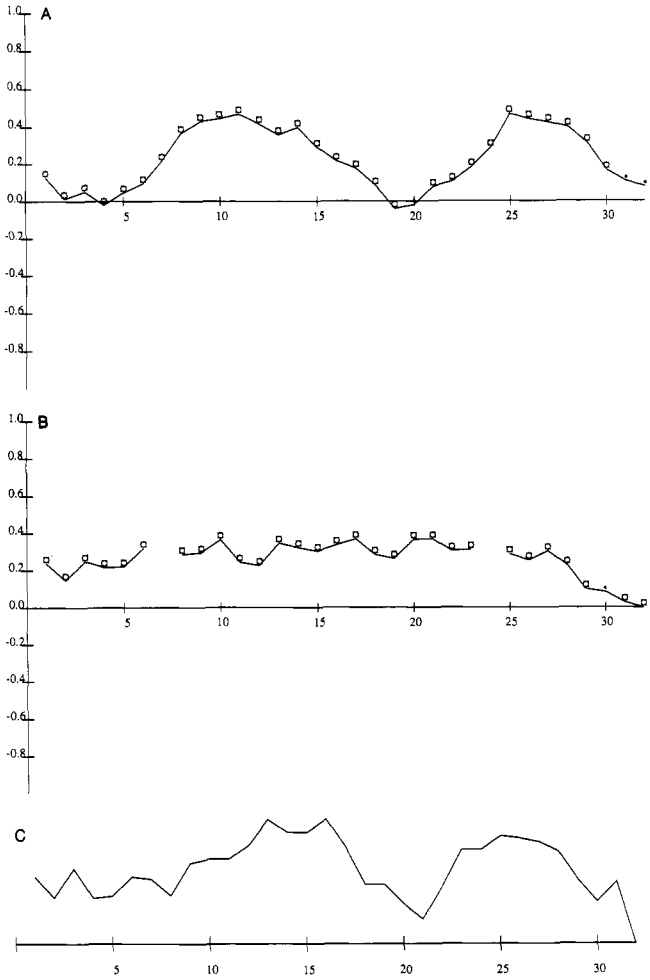

FIGURE 3: Plot with the secondary chemical shift vs the amino acid sequence of B1-Dim (A) and the leucine zipper region of GCN4 (B). The profile for B1-Dim is very similar to the one obtained by a Chou–Fasman secondary structure prediction (C). Data smoothing with a window of two points was used in all of the plots.

HN(i,i+1) connectivities appear between residues 4 and 7. We can estimate these distances aroun 3·6 Å from the appearance at these mixing times of some of the Hα–HN(i,i+1) connectivities in the rest of the chain (typically around 3·6 Å for a helical structure and shorter otherwise). An HN–HN(i,i+3) between S6 and Q9 shows the N-terminal region to be kinked.

The combined information from the sequential NOE connectivities HN–HN and Hβ–HN and the medium-range connectivities Hα–HN(i,i+3) and Hα–Hβ(i,i+3) identifies a helix extending approximately from residues 7 to 18 and from 23 to 32. Supporting evidence for a helix in these two regions comes from amide proton exchange rates (see Figure 5). The absence of HN–HN(i,i+2) effects along the entire chain not even at long mixing times excludes the possibility of a 3-10 helix. No sign of opening of the C-terminal region could be observed, this region being quite rigidly defined.

The absence of Hβ–HN and Hα–Hβ(i,i+3) NOE effects for the residues in the region 18–22 indicates an opening of the helix with the formation of a kink. HN–HN(i,i+1) connectivities between S19–G20 and L21–S22 can still be observed, but they are much weaker and they are observable only beyond 150-ms mixing times. In this region, the amide proton exchange rates are high (see Figure 5). We looked for contacts between residues distant along the sequence. The observation of any of these connectivities would suggest the presence of a hairpin structure and therefore of a four-helix bundle in the dimer. It could also be consistent with a

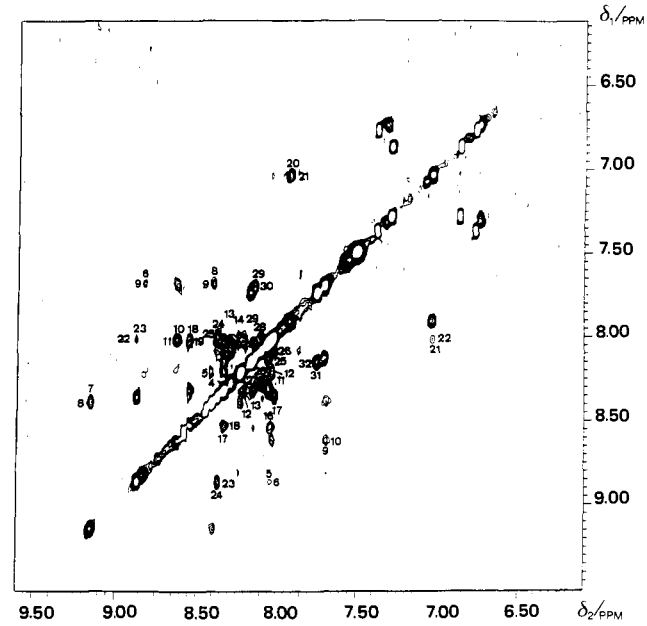

FIGURE 4: Amide region of a NOESY spectrum of B1-Dim with 200-ms mixing time. The NOEs are labeled along the appropriate axis with the sequence positions of the residues involved.

head-to-tail arrangement of the two monomers, each of them having a kinked helix. Four-helix bundles are extremely stable and, if present, could account for the unusual stability of the peptide.

No connectivities involving residues more than four positions apart in the sequence were observed. This suggests a parallel packing of the monomers.

Discussion

Isolated α-helices are seldom stable in aqueous solutions (Wright et al., 1988). When observed, they are usually stabilized by hydrophobic packing from the formation of dimers or higher aggregates. Hairpins and coiled-coil conformations are the typical structural solutions of the potential instability.

Originally, the discovery of the LFB1 dimerization domain was based on the observation of sequence homology with a stretch of the rod portion of myosin. In addition, the amphipathic character of the peptide and the presence of bulky hydrophobic residues repeated every seven residues were somewhat reminiscent of the leucine zipper motif (Landschulz et al., 1988). A rather natural consequence of these two considerations led us to suggest the peptide to be folded in a coiled-coil conformation. However, our present results exclude the myosin-like/coiled-coil model.

We can conclude that in the N-terminus of LFB1, helix–helix packing through the four heptads is allowed by a kink around the region between 19 and 22 which precludes a coiled-coil structure. Previous experiments on the LFB1 protein showed that the single-point mutation of residues V2, L21, or L30 does not affect DNA binding (Tomei and De Francesco, unpublished results). Conversely, a mutation at position L5, L16, or L26 severely impairs the protein activity. These data agree well with our results, where L5, L16, and L26 are necessary to stabilize the dimer. Terminal positions and the L21 involved in the kink region play a less important role.

Our result might also be compared with the leucine zipper structure which plays a similar role as a dimerization domain in other transcriptional activators. In leucine zipper-type domains, a long uninterrupted helix is observed, and a coiled-coil structure has been inferred from NMR studies (Oas

Table I: Chemical Shifts (ppm) of the Assigned ^{1}H NMR Lines of B1-Dim at 26 °C

	HN	HαC	HβC	HγC	HδC	HεC	others
1 Met	NA	4.18	2.14	2.55			
2 Val	8.57	4.15	2.04	0.94			
3 Ser	8.44	4.43	3.78				
4 Lys	8.40	4.36	1.81, 1.68	1.43, 1.38	1.64	2.97	7.52
5 Leu	8.22	4.50	1.57	1.62	0.72, 0.66		
6 Ser	8.82	4.48	4.35, 3.94				
7 Gln	9.15	4.04	2.10, 1.98	2.31			7.27, 6.84
8 Leu	8.38	4.09	1.60	1.48	0.83, 0.71		
9 Gln	7.68	3.63	2.35, 1.60	2.53			7.66, 6.70
10 Thr	8.61	3.70	4.37	1.21			
11 Glu	8.04	4.08	2.32, 2.10	2.52			
12 Leu	8.20	3.94	1.97, 1.90	1.22	0.83, 0.72		
13 Leu	8.32	3.97	1.88, 1.81	1.63	0.86, 0.75		
14 Ala	8.03	4.00	1.42				
15 Ala	8.04	4.08	1.42				
16 Leu	8.15	3.86	1.86, 1.72	1.57	0.86, 0.75		
17 Leu	8.32	4.52	1.90, 1.77	1.58	0.86, 0.75		
18 Glu	8.53	4.22	2.26, 2.18	2.51			
19 Ser	8.01	4.30	4.12, 4.00				
20 Gly	7.90	4.50, 3.68					
21 Leu	7.02	4.51	1.47, 1.39	1.57	0.73		
22 Ser	8.02	4.52	4.30, 4.03				
23 Lys	8.85	3.82	1.86	1.34	1.76, 1.66	2.97	7.52
24 Glu	8.36	3.97	2.07, 1.92	2.53, 2.45			
25 Ala	8.00	4.07	1.50				
26 Leu	8.08	3.99	1.74, 1.57		0.80, 0.71		
27 Ile	8.27	3.37	1.87	1.87, 0.84	0.73		
28 Gln	8.05	3.99	2.10	2.34, 2.50			7.35, 6.75
29 Ala	8.13	4.15	1.48				
30 Leu	7.70	4.20	1.86	1.52	0.88, 0.72		
31 Gly	7.73	4.15, 3.85					
32 Gly	8.12	4.03, 3.68					

FIGURE 5: Survey of the NOEs observed for B1-Dim that fall into the seven major classes shown. For the first three classes, the heights of the bars indicate the intensities of the NOEs. Slow exchanging amide protons are labeled.

et al., 1990; Saudek et al., 1990).

Since the discovery of the leucine zipper motif, a large number of sequences containing a repeat of bulky hydrophobic residues has been recognized, and new putative leucine zipper-type structures have been suggested. Our conclusions show that, although large hydrophobic residues play a fundamental role in helix packing, coiled-coils are not the only structural solution to the problem of dimerization.

Other considerations, other than the occurrence of hydrophobic repeats, might help instead in evaluating a sequence. Coiled-coils of myosin and mysoin-like proteins are characterized by the presence of hydrophobic amino acids (mainly leucines, valines, or alanines; Cohen & Parry, 1990) at positions *a* and *d* when the protein sequence is displayed as an array of contiguous heptad repeats. Conversely, positions *e* and *g* are often occupied by charged amino acids that may be important for interhelical stabilization by ion pairing. In B1-Dim, we find that the rule for the occurrence of hydrophobic amino acids at positions *a* and *d* is respected, but we do not find charged residues in positions *e* and *g* in an arrangement suitable for ion pairing in two helices packed in a parallel alignment. Furthermore, serines occur at positions *a* and *d* in the third heptad repeat of the dimerization domain of LFB1, whereas they have the lowest probability of occurring at those positions in myosin-related coiled-coils (Cohen & Parry, 1990).

Calculations based on a distance geometry algorithm are now being undertaken to translate the proton–proton distances observed into the 3D structure of the monomer.

We know that the LFB1 dimerization domain is present in solution as a dimer from the thermodynamic stability of B1-Dim and from biochemical experiments both on the peptide and on the whole protein. Although we cannot exclude a priori the presence of higher order species at the NMR concentra-

tions (several orders of magnitude higher than those used in both CD and biochemical assays), only one main species could be detected by NMR under the experimental conditions used. Therefore, from the absence of intermolecular effects, we infer the presence of a *symmetrical parallel* dimer.

A large number of DNA binding proteins function as dimers, which is undoubtedly related to the need to recognize palindromic stretches of DNA. From the experimental NMR point of view, dimeric proteins pose specific problems which are not always easy to overcome. Resonances become broader for large molecules. In cases like the leucine zipper and LFB1, dimerization is obtained by packing bulky residues, usually leucines and valines. Similar residues repeated along the chain result in only slight dispersion of the chemical shift. More importantly, the necessity of distinguishing between dipolar effects within each monomer or across the dimer is particularly difficult when the two molecules or the interactive subregions are related by dyad symmetry. In this case, already discussed in some detail for the Arc repressor and for the leucine zipper (Breg et al., 1989; Oas et al., 1990; Saudek et al., 1990), no theoretical solution is known. Experimentally, future studies of selectively labeled B1-Dim should help in overcoming the problem.

ACKNOWLEDGMENTS

We thank Dr. H. Gausepohl for synthesizing the peptide, Dr. L. Tomei for making available to us unpublished results, and Drs. Spraul and E. Hoffman from Bruker (Karlsruhe) and Varian (Darmstadt), respectively, for performing some of the 2D experiments. We also thank Drs. D. J. Thomas and A. M. Lesk for critical reading of the manuscript.

REFERENCES

Ahmad, F., & Bigelow, C. C. (1982) *J. Biol. Chem. 257*, 12935.
Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 455.
Bowie, J. U., & Sauer, R. T. (1989) *Biochemistry 28*, 7139.
Breg, J. N., Boelens, R., George, A. V. E., & Kaptein, R. (1989) *Biochemistry 28*, 9826.
Chou, P. Y., & Fasman, G. D. (1977) *J. Mol. Biol. 115*, 135.
Clayden, N. J., & Williams, R. J. P. (1982) *J. Magn. Reson. 49*, 383.
Cohen, C., & Parry, D. A. D. (1990) *Proteins: Struct., Funct., Genet. 7*, 1.
De Francesco, R., Pastore, A., Vecchio, G., & Cortese, R. (1991) *Biochemistry* (preceding paper in this issue).
Englander, S. W., & Wand, A. J. (1987) *Biochemistry 26*, 5953.
Jeener, J., Meier, B. H., Bachman, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546.
Macura, S., Huang, Y., Suter, D., & Ernst, R. R. (1981) *J. Magn. Reson. 43*, 259.
Marion, D., & Wuethrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967.
Nicosia, A., Monaci, P., Tomei, L., De Francesco, R., Nuzzo, M., Stunnenberg, H., & Cortese, R. (1990) *Cell 61*, 1225.
Oas, T. G., McIntoch, L. P., O'Scea, E. K., Dahlquist, F. W., & Kim, P. S. (1990) *Biochemistry 29*, 2891.
Pace, N. C. (1975) *CRC Crit. Rev. Biochem. 3*, 1.
Pastore, A., & Saudek, V. (1990) *J. Magn. Reson. 89*.
Rance, M. (1987) *J. Magn. Reson. 74*, 557.
Saudek, V., Pastore, A., Castiglione Morelli, M. A., Frank, R., Gausepohl, H., Gibson, T., Weih, F., & Roesch, P. (1990) *Protein Eng. 4*, 3.
Williams, R. J. P. (1987) *Carlsberg Res. Commun. 52*, 1.
Wright, P. E., Dyson, H. J., & Lerner, R. A. (1988) *Biochemistry 27*, 7168.
Wuethrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.