Protein Science (1997), 6:175–184. Cambridge University Press. Printed in the USA.

Heteronuclear (^{1}H, ^{13}C, ^{15}N) NMR assignments and secondary structure of the basic region-helix-loop-helix domain of E47

ROBERT FAIRMAN,[1] RITA K. BERAN-STEED,[2] AND TRACY M. HANDEL[3]

The DuPont Merck Pharmaceutical Company, Experimental Station, P.O. Box 80500, Wilmington, Delaware 19880-0500

(RECEIVED July 29, 1996; ACCEPTED October 16, 1996)

Abstract

E47 is an immunoglobulin enhancer DNA-binding protein that contains a basic region-helix-loop-helix (b/HLH) domain. This structural motif defines a class of transcription factors that are central to the developmental regulation of many tissues. Its function is to provide a dimerization interface through the formation of a parallel four-helix bundle, resulting in the juxtaposition of two basic DNA-recognition α-helices that control sequence-specific DNA-binding. In order to gain insight into the biophysical nature of b/HLH domains, we have initiated structural studies of the E47 homodimer by NMR. Sequence-specific resonance assignments have been obtained using a combination of heteronuclear double- and triple-resonance NMR experiments. The secondary structure was deduced from characteristic patterns of NOEs, $^{13}C_{\alpha/\beta}$ chemical shifts, and measurements of $^3J_{HNH\alpha}$ scalar couplings. Except for the basic region recognition helix, the secondary structural elements of the E47 homodimer are preserved in the absence DNA when compared with the co-crystal structure of E47 bound to DNA (Ellenberger T, Fass D, Arnaud M, Harrison SC, 1994, *Genes & Dev* 8:970–980). As expected, the DNA-binding helix is largely unstructured, but does show evidence of nascent helix formation. The HLH region of E47 is structured, but highly dynamic as judged by the rapid exchange of backbone hydrogen atoms and the relatively weak intensities of many of the NOEs defining the dimerization helices. This dynamic nature may be relevant to the ability of E47 both to homodimerize and to heterodimerize with MyoD, Id, and Tal1.

Keywords: E47; helix-loop-helix; homodimers; NMR; transcription factor

The basic region-helix-loop-helix (b/HLH) motif defines a class of transcription factors that are important regulators of many eukaryotic developmental pathways. The HLH region is an oligomerization domain capable of forming dimers, tetramers, and higher-order aggregates. Dimerization results in the pairing up of each HLH motif to form a four-helix bundle. Upon dimerization, this domain brings together two N-terminal arginine- and lysine-rich DNA-binding α-helices (the basic region), which promotes specific binding to the pseudo-symmetric CANNTG DNA sites found in the E-box control regions of many developmentally regulated genes.

Regulation of gene transcription results from both homo- and heterodimerization of heterologous HLH-containing proteins, which also contain other distal protein–protein interaction domains (cf. Jones, 1990; Lamb & McKnight, 1991; Lassar et al., 1991). For example, E47, a ubiquitously expressed protein identified originally as an immunoglobin enhancer-binding factor, forms homodimers as well as heterodimers with the muscle-specific differentiation factor, MyoD (Lassar et al., 1991; Kadesch, 1993). Likewise, the b/HLH/Z protein MAX can homodimerize or heterodimerize with the oncoprotein myc, which by itself does not normally oligomerize (Blackwood & Eisenman, 1991; Prendergast et al., 1991). It is clear that the functional relevance of homo- and heterodimer formation is to increase the range of DNA specificity and transcriptional control attainable from a finite number of pro-

Reprint requests to: Robert Fairman, Division of Macromolecular Structure, Bristol-Myers Squibb Pharmaceutical Research Institute, P.O. Box 4000, Princeton, New Jersey 08543-4000, e-mail: fairman@bms.com; or Tracy Handel, Department of Molecular and Cell Biology, University of California, Berkeley, Virus Laboratory, 229 Wendell M. Stanley Hall, Berkeley, California 94720, e-mail: handel@paradise1.berkeley.edu.

[1]Present address: Division of Macromolecular Structure, Bristol-Myers Squibb Pharmaceutical Research Institute, P.O. Box 4000, Princeton, New Jersey 08543-4000.

[2]Present address: Pharmacia Biotech Inc., P.O. Box 1327, Piscataway, New Jersey 08855-1327.

[3]Present address: Department of Molecular and Cell Biology, University of California, Berkeley, California 94720.

Abbreviations: PMSF, phenyl methyl sulfonyl fluoride; DTT, dithiothreitol; NOESY, NOE spectroscopy; TOCSY, total correlation spectroscopy; HMQC, heteronuclear multiple quantum coherence; HSQC, heteronuclear single-quantum coherence; CT, constant time; 1D, 2D, 3D, 4D, one-, two-, three-, and four-dimensional NMR; TPPI, time-proportional phase incrementation; $T_1\rho$, rotating frame relaxation time; T_2', the arithmetic mean of zero quantum and double quantum relaxation times; ^{1}H-^{2}D exchange, hydrogen deuterium exchange.

teins. However, the structural and chemical basis by which HLH domains discriminate their dimerization partners is still not understood completely.

An additional observation that is also not well understood is the ability of HLH proteins to form tetrameric and higher-order oligomeric structures. Several studies have demonstrated tetramer formation both in vitro (Starovasnik et al., 1992; Fairman et al., 1993; Ferré-D'Amaré et al., 1994) and in vivo (Farmer et al., 1992; Ferré-D'Amaré et al., 1994). In studies of the Id protein, an HLH protein that lacks a functional DNA-binding domain, it was proposed that the tetrameric form may act to inhibit DNA-binding (Fairman et al., 1993; cf. Phillips, 1994). A similar proposal was suggested by Fisher et al. (1991) in their studies of TFEB, a b/HLH/Z containing DNA-binding factor. Finally, under physiologically relevant concentrations, USF was shown to exist as a homotetramer that binds simultaneously to two independent DNA sites; these data, coupled with in vitro experiments, led to the intriguing proposal that this form of the protein may have a role in DNA looping (Ferré-D'Amaré et al., 1994). Less is known about larger oligomeric states, although a recent study reported the existence of MyoD micelles containing more than 100 monomer subunits, and several models were presented to describe the influence of a micellar depot of MyoD on homo- and heterodimerization and on DNA binding (Laue et al., 1995).

To gain insight into the function and assembly of these different HLH oligomers, structural studies of different forms of these proteins are necessary. Several models have been built predicting the tertiary fold of E47 and MyoD homodimers (Anthony-Cahill et al., 1992; Halazonetis & Kandil, 1992; Starovasnik et al., 1992; Vinson & Garcia, 1992; Gibson et al., 1993) using sequence alignments of the predicted amphipathic α-helices and the intervening, variable-length, loop region (Murre et al., 1989). The first experimental structural information for the b/HLH motif came from the co-crystal structure of MAX, which contains both a b/HLH domain and a leucine zipper (LZ), bound to its cognate DNA-binding site (Ferré-D'Amaré et al., 1993). This structure revealed that, upon dimerization, the b/HLH domain forms a left-handed parallel four-helix bundle, with the DNA-binding helix being an extension of helix 1 and the leucine zipper domain being an extension of helix 2. Subsequently, the co-crystal structures of a second b/HLH/Z protein, USF bound to DNA (Ferré-D'Amaré et al., 1994), and two other b/HLH proteins bound to DNA, MyoD (Ma et al., 1994) and E47 (Ellenberger et al., 1994), have been solved (cf. Wolberger, 1994).

We are also interested in structures of these proteins in the absence of DNA because of their relevance to the assembly of functional heterodimers and higher-order oligomers. We chose to work on the b/HLH domain of E47, a protein that regulates immunoglobulin gene transcription (Shen & Kadesch, 1995) and B-cell development (Bain et al., 1994). E47 can also heterodimerize with the tissue specific HLH protein MyoD to promote muscle differentiation, and with Tal1, a b/HLH proto-oncogene implicated in T-cell acute lymphoblastic leukemia (Hsu et al., 1994). Herein we present complete resonance assignments and the secondary structure of E47 based on double and triple resonance NMR experiments. This is a necessary first step in the determination of a solution structure that can be compared with the co-crystal structure of E47 bound to DNA (Ellenberger et al., 1994) and structures of E47 heterodimer complexes when they become available. We also describe data on the dynamic state of the E47 homodimer and interpret this in light of recent thermodynamic studies of protein–DNA complexes.

Results

Using circular dichroism, Anthony-Cahill and coworkers demonstrated that the HLH domain of the E47 homodimer is largely α-helical and the basic region is random coil in the absence of DNA (Anthony-Cahill et al., 1992). We therefore anticipated that 3D ^{15}N-separated experiments would be highly congested, and that 3D and possibly 4D double resonance experiments would be necessary for complete sequential assignments; this expectation was realized fully.

A 67-amino acid E47 fragment labeled uniformly with ^{15}N or ^{15}N/^{13}C was produced in *Escherichia coli* grown in minimal medium containing [^{15}N]$(NH_4)_2SO_4$ or [^{15}N]$(NH_4)_2SO_4$ and [$^{13}C_6$]*d*-glucose as the sole source of nitrogen and carbon. Conditions for optimal data collection were determined by recording ^{1}H-^{15}N HSQC spectra on ^{15}N-labeled E47 at a variety of pH, temperature, and salt conditions (Bax et al., 1990b; Norwood et al., 1990). Estimates of $T_1\rho$ were also made by collecting 1D ^{15}N-separated TOCSY spectra at different mixing times (Marion et al., 1989).

An HSQC spectrum of E47 collected under final conditions (20 mM DTT, pH 5.5, 30 °C) is shown in Figure 1. The spectrum shows reasonable dispersion in both the ^{1}H and ^{15}N dimensions. Several crosspeaks are aliased because the ^{15}N spectral width was reduced to 625.0 Hz in order to minimize the number of time increments necessary to achieve acceptable digital resolution in 3D and 4D spectra. Sixty-three of the expected 67 H_N-^{15}N crosspeaks are evident with only a few areas of serious overlap. The region of greatest difficulty is shown in expanded region of Figure 1, and, even in this highly congested region, a combination of several experiments (see below) allowed us to assign all resonances. Linewidths of well-resolved ^{15}N resonances are 9–10 Hz and H_N linewidths are 17–24 Hz. These linewidths are consistent with a homodimer that is perhaps somewhat elongated or slightly aggregated. The number of resonances in the HSQC spectrum confirm that E47 is a symmetric dimer. Analytical ultracentrifugation studies (Bishop et al., 1995; Laue et al., 1995) also suggest that the b-HLH domain of E47 is a homodimer, as measured in the range from 100 to 150 μM protein. Therefore, it is unlikely that E47 forms higher-order oligomeric states at 1.5–2.0 mM protein, the concentration at which our NMR experiments were done, because this would surely have been detected by the ultracentrifugation studies.

Sequential assignments

Initially, we tried using 3D ^{15}N-separated TOCSY HMQC, 3D ^{15}N-separated NOESY (Marion et al., 1989), and 3D-HNHB (Archer et al., 1991) experiments to assign the backbone and side-chain ^{1}H and ^{15}N nuclei. This proved to be extremely difficult. Magnetization transfer from side-chain to amide protons in the 3D ^{15}N-separated TOCSY HMQC was not very efficient, making spin system identification highly ambiguous. Assignment reliability was only somewhat improved by the HNHB experiment. Sequential connections were also difficult, even for the predicted helical regions, which should be relatively straightforward using ^1H_N-^1H_N NOEs. However, the ^1H_N-^1H_N NOEs were often very weak, which suggests that the α-helices are ill-defined or that the bundle is fairly dynamic; in either case, this would lead to more rapid exchange of the amide protons and bleaching from presaturation of the water resonance. Because assignments should be easiest in the helical region of the protein compared with the basic region and the loop, these results were not encouraging for obtaining complete assignments by double resonance experiments. The few residues in

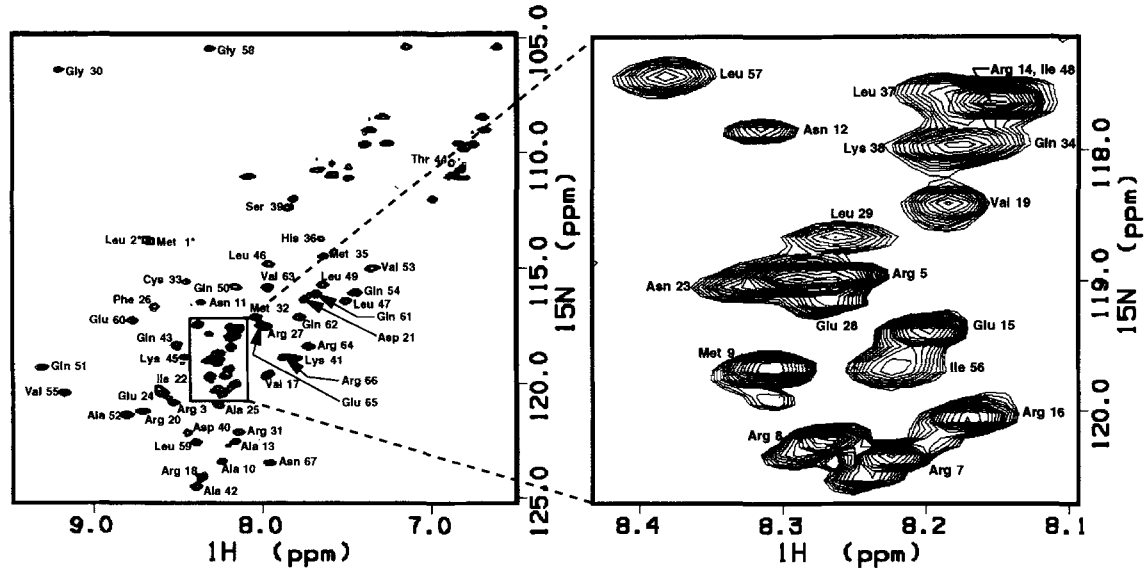

Fig. 1. 2D ^{1}H-^{15}N HSQC of E47 at 30 °C, pH 5.5. The spectral width in the ^{15}N dimension was 625.0 Hz. Aliased crosspeaks are displayed without sign discrimination. The region of greatest overlap is enlarged.

the basic region that were assignable had sparse NOE content and lacked $^{1}H_N$-$^{1}H_N$ NOEs, consistent with the expectation that this region of the protein is relatively unstructured. We therefore proceeded by collecting 4D $^{15}N/^{13}C$ hetero-nuclear NMR experiments to assign the protein via scalar connections.

Heteronuclear experiments for assignment of backbone nuclei (^{15}N, $^{1}H_N$, $^{13}C_\alpha$, $^{1}H_\alpha$) included 4D CT HCA(CO)NNH and 4D CT HCANNH experiments (Boucher et al., 1992a, 1992b) acquired on a uniformly labeled $^{15}N/^{13}C$ sample of E47 in 90% H_2O/10% D_2O. The HCANNH experiment provides intraresidue correlations between the ($^{1}H_\alpha$,$^{13}C_\alpha$) and (^{15}N, $^{1}H_N$) resonances of a given residue, as well as the interresidue correlation to the ($^{1}H_\alpha$,$^{13}C_\alpha$) chemical shifts of the preceding residue; for E47, this experiment was more sensitive than the related HNNCAHA experiment, which is not surprising because the protein is fairly small and appears to be somewhat dynamic. The HCA(CO)NNH experiment correlates the (^{15}N, $^{1}H_N$) resonances of a given residue with only the ($^{1}H_\alpha$,$^{13}C_\alpha$) of the preceding residue, providing independent confirmation of sequential connections. Side-chain ^{13}C and ^{1}H assignments were obtained via a 4D HCC(CO)NNH experiment (Clowes et al., 1993) that correlates ^{15}N-$^{1}H_N$ resonances of a given residue with ^{1}H-^{13}C chemical shifts of the entire preceding side chain. This experiment was especially useful because it takes advantage of the level of dispersion of $^{1}H_N$ and ^{15}N chemical shifts for resolving individual spin systems. For example, the ^{1}H and ^{13}C chemical shifts of several arginine side chains were almost completely degenerate, but assignable nonetheless because the chemical shifts of the C-terminal $^{1}H_N$ and ^{15}N resonances were well-resolved. This is illustrated in Figure 2, which shows separate 2D ^{1}H-^{13}C planes from the 4D HCC(CO)NNH experiment at the $^{1}H_N$-^{15}N coordinates of four different arginines.

Virtually all of the backbone and side-chain assignments were made with a combination of these three 4D experiments. Problematic regions were overcome by combining the information from these experiments with the 3D ^{15}N-separated NOESY, the 3D HNHB experiment, and a 3D ^{13}C-separated TOCSY (Bax et al., 1990a) collected on a sample of E47 in D_2O. For example, we observed weak or missing $C_\beta H_\beta$ crosspeaks of glutamic acid residues and $C_\gamma H_\gamma$ or $C_\delta H_\delta$ crosspeaks of some of the hydrophobic amino acids in the HCC(CO)NNH experiment. These missing assignments were provided by the HNHB experiment and the 3D ^{13}C-separated TOCSY. Together, the 3D ^{15}N-separated experiments and the ^{13}C-separated TOCSY also provided higher precision of the indirectly detected ^{1}H and ^{13}C side-chain chemical shifts relative to the 4D HCC(CO)NNH experiment. Table 1 is a list of assignments for 97% of the protein backbone chemical shifts and 79% of the side-chain chemical shifts.

Secondary structure

Short- and medium-range NOEs were obtained from the 3D ^{15}N-edited NOESY HMQC experiment and help define the secondary structure as illustrated in Figure 3. The presence of strong $H_\alpha H_N$ (i, i + 1) NOEs between Met 1 and Arg 20 is consistent with the expected random coil nature of the basic region in the absence of DNA (Fig. 3). Helix 1 extends from Asp 21 to Cys 33 as judged by the presence of $H_N H_N$ (i, i + 1) NOEs and weaker $H_\alpha H_N$ (i, i + 1) NOEs. Likewise, helix 2 extends from Thr 44 to Asn 67 and has several $H_\alpha H_N$ (i, i + 3) NOEs in addition to $H_N H_N$ (i, i + 1) NOEs; few $H_\alpha H_N$ (i, i + 3) NOEs are observed for helix 1, suggesting that helix 2 is the more stable of the two. The loop region from Gln 34 to Gln 43 is characterized by strong $H_\alpha H_N$ (i, i + 1) NOEs.

Confirmation of the secondary structure is obtained from the difference in C_α chemical shifts relative to random coil values (Spera & Bax, 1991) and from the three-bond coupling constant $^{3}JH_N H_\alpha$ (Billeter et al., 1992). These data are shown in Figure 4A and B. Helical regions of the protein are characterized by positive C_α chemical shifts from random coil values, whereas extended or β-sheet structures have negative shifts. The basic region (Met 1 to Arg 20) is dominated by slightly positive δC_α values. The strongly positive δC_α's from Ile 22 to Met 35 and Lys 45 to Asn 67 identify helix 1 and helix 2, respectively. The dip in the δC_α's in the middle of helix 1 may be due to ring current shifts from Phe 26; interestingly, the drop and subsequent rise have the periodicity of an

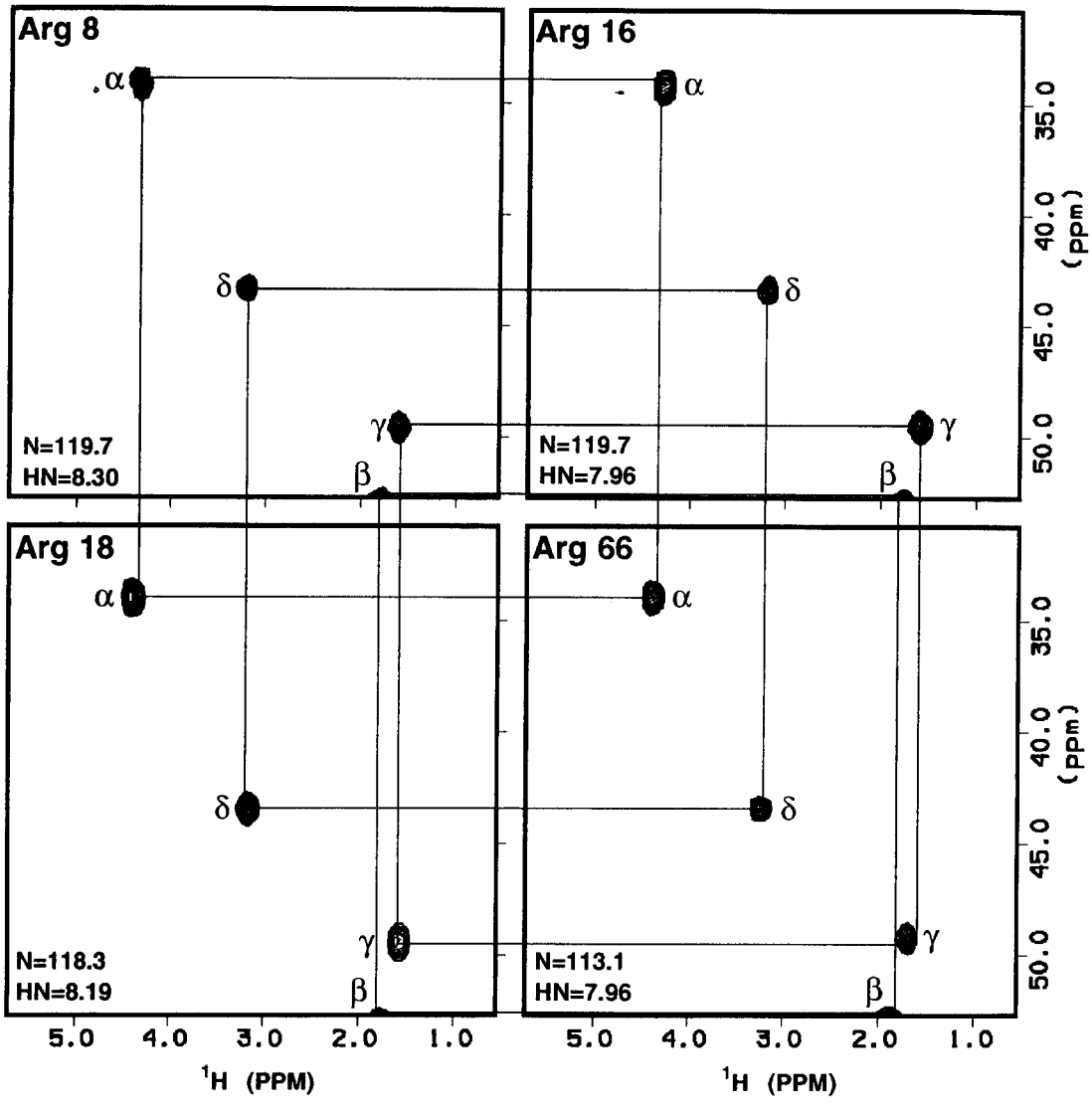

Fig. 2. 2D ^{1}H-^{13}C planes from the 4D HCC(CO)NNH displayed at the $^{1}H_{N}$-^{15}N chemical shifts of Arg 8, Arg 16, Arg 18, and Arg 66. Corresponding crosspeaks in all four planes have nearly identical ^{1}H and ^{13}C chemical shifts.

α-helix. Alternatively, the helix could be disrupted slightly by the presence of Gly 30, which is a helix breaking residue (Chou & Fasman, 1978). Negative shifts from Leu 37 to Gln 43 suggest that the loop region is in an extended conformation.

Further support for this assignment of regions of secondary structure comes from values of the $^{3}JH_{N}H_{\alpha}$ coupling constant (Fig. 4C). These were obtained from a series of J-modulated ^{1}H-^{15}N HSQC spectra (Billeter et al., 1992). Estimates of the error in the values of $^{3}JH_{N}H_{\alpha}$ are approximately ±0.2 Hz for nonoverlapped regions of the J-modulated HSQC spectra. These were made by evaluating the uncertainty in intensity measurements (10%) and propagating the uncertainty throughout the calculation. The uniformly lower J-values between Asp 21 and Cys 33 and Thr 41 and Glu 65 are consistent with an α-helical conformation, whereas between Gln 34 and Ala 42, the J-values are more compatible with an extended structure. Based on the J-values, the second half of helix 1 appears less ordered than the first half and the C terminus of helix 2 appears frayed.

Protein dynamics

An apparent relaxation time T_2', which is the arithmetic mean of zero and double quantum ^{1}H-^{15}N coherence relaxation times, can also be determined from the J-modulated ^{1}H-^{15}N HSQC experiment. The T_2' values provide qualitative insight into the dynamics of the protein backbone (Fig. 4C). The fitted T_2' values were in the range of 30–40 ms and are remarkably constant for the region between Asp 21 and Glu 65, including the predicted loop; this has implications for the orientation of the α-helices relative to one another. Only at the N-terminal and C-terminal extrema do the relaxation times increase, reflecting mobility of the basic region and fraying of the C-terminal end of helix 2.

Protein dynamics were characterized further using an ^{1}H-^{2}D exchange experiment. Regions of secondary structure typically have slowly exchanging amide protons due to the involvement of these protons in hydrogen bonds with carbonyl or side-chain oxygens. In the case of E47, amide protons were barely detectable

Table 1. *^{1}H, ^{15}N, and ^{13}C chemical shifts for E47 at pH 5.3 and 30 °C*

Residue	^{15}N (NH)	$^{13}C\alpha$ (Hα)	$^{13}C\beta$ (Hβ)	$^{13}C\gamma$ (Hγ)	Others
Met 1	113.7 (8.64)	54.9 (4.13)	28.2 (2.54)	30.6 (2.59)	
Leu 2	113.6 (8.68)	55.2 (4.42)	42.2 (1.58)	26.9 (1.60)	δ; 24.0,20.2 (0.87,0.86)
Arg 3	120.8 (8.53)	56.1 (4.32)	(1.88,1.78)		
Asp 4		57.7 (4.82)	41.2 (2.68)		
Arg 5	119.0 (8.25)	56.8 (4.20)	(2.04)		
Glu 6		57.0 (4.23)	30.6 (2.02)	36.3 (2.25)	
Arg 7	120.3 (8.22)	56.9 (4.22)	30.6 (1.82)	27.1 (1.61)	δ; 43.3 (3.20)
Arg 8	120.2 (8.26)	56.2 (4.33)	30.6 (1.77)	27.1 (1.61)	δ; 43.3 (3.19)
Met 9	119.7 (8.30)	55.4 (4.46)	27.2 (2.56)	29.6 (2.60)	
Ala 10	123.3 (8.24)	52.7 (4.31)	18.9 (1.36)		
Asn 11	116.4 (8.36)	53.4 (4.65)	38.7 (2.78)		
Asn 12	117.7 (8.27)	53.3 (4.68)	38.8 (2.79)		
Ala 13	122.6 (8.16)	53.1 (4.24)	19.1 (1.36)		
Arg 14	117.6 (8.16)	56.4 (4.25)	30.6 (1.80,1.51)	27.2 (1.63,1.26)	δ; 43.3 (3.20)
Glu 15	119.4 (8.20)	56.5 (4.26)	30.6 (2.04,1.94)	35.9 (2.24,1.84)	
Arg 16	120.1 (8.17)	56.4 (4.31)	30.6 (1.77)	27.4 (1.60)	δ; 43.4 (3.18)
Val 17	119.7 (7.96)	62.7 (4.02)	33.5 (2.03)	21.0,18.3 (0.87,0.89)	
Arg 18	124.0 (8.36)	56.1 (4.39)	30.6 (1.77)	27.3 (1.57)	δ; 43.5 (3.18)
Val 19	118.3 (8.19)	61.9 (4.17)	32.8 (2.08)	20.4,22.0 (0.95,0.86)	
Arg 20	121.2 (8.71)	57.1 (4.26)	30.6 (1.83)	27.1 (1.61)	δ; 43.3 (3.18)
Asp 21	116.2 (7.74)	53.3 (4.69)	42.6 (3.03,2.76)		
Ile 22	120.2 (8.62)	64.7 (3.81)	38.3 (2.16,1.94)	29.9 (1.69,1.34)	γ_m; 16.7 (1.10), δ; 13.2 (0.86)
Asn 23	119.0 (8.30)	56.7 (4.75)	37.6 (3.08,2.77)		
Glu 24	120.6 (8.62)	59.2 (4.12)	(2.18)	36.5 (2.56,2.34)	
Ala 25	121.0 (8.28)	55.1 (4.15)	18.7 (1.36)		
Phe 26	116.6 (8.63)	63.9 (4.13)	39.2 (3.25,3.11)		ring; (7.37)
Arg 27	117.5 (7.92)	59.7 (4.11)	30.6 (2.05,1.80)	27.1 (1.61)	δ; 43.3 (3.19)
Glu 28	119.2 (8.27)	56.9 (4.21)	29.8 (2.06)	35.2 (2.36)	
Leu 29	118.6 (8.27)	57.5 (4.21)	41.9 (1.85,1.38)	27.0 (1.44)	δ; 26.8,23.4 (0.64,0.75)
Gly 30	106.3 (9.17)	47.6 (4.00,3.76)			
Arg 31	122.1 (8.13)	59.5 (4.18)	30.1 (2.02)	27.5 (1.69)	δ; 43.3 (3.27)
Met 32	117.1 (8.04)	58.7 (4.31)	29.2 (2.19)	32.2 (2.86)	
Cys 33	115.5 (8.45)	64.5 (4.16)	31.8 (2.67)		
Gln 34	117.9 (8.15)	58.7 (4.21)	29.9 (2.14)	33.8 (2.45)	
Met 35	114.4 (7.65)	57.6 (4.19)	31.8 (2.12,2.02)	32.8 (2.66,2.55)	
His 36	113.7 (7.67)	54.9 (4.42)			
Leu 37	117.5 (8.18)	54.7 (4.41)	42.7 (1.56)	27.1 (1.56)	δ; 25.8,23.0 (0.76,0.79)
Lys 38	117.9 (8.18)	56.3 (4.13)	30.6 (1.82)	24.9 (1.29)	δ; 29.0 (1.61)
Ser 39	112.3 (7.84)	57.0 (4.60)	64.4 (3.71)		
Asp 40	122.1 (8.43)	53.4 (4.66)	40.8 (2.73,2.61)		
Lys 41	118.9 (7.81)	55.9 (4.18)	33.5 (1.62,1.75)	24.7 (1.37)	δ; 28.9 (1.61), ϵ; 42.2 (2.97)
Ala 42	124.4 (8.40)	52.6 (4.17)	18.2 (1.28)		
Gln 43	118.4 (8.50)	56.3 (4.26)	27.5 (2.23,1.98)	34.1 (2.44,2.25)	
Thr 44	110.1 (6.84)	59.1 (4.42)	(4.51)		
Lys 45	118.8 (8.42)	60.6 (3.39)	32.4 (1.20,1.14)	18.9 (1.43)	ϵ; 42.0 (2.71,2.63)
Leu 46	114.8 (7.96)	58.1 (3.85)	41.9 (1.53,1.36)	26.8 (1.33)	δ; 25.8,22.9 (0.81,0.67)
Leu 47	116.4 (7.50)	57.4 (4.23)	41.0 (1.38,0.62)	21.7 (1.39)	δ; 25.5,22.9 (0.84,0.82)
Ile 48	117.6 (8.15)	66.4 (3.70)	37.9 (2.07,1.89)	24.5 (1.00)	γ_m; 16.8 (0.95), δ; 14.4 (0.92)
Leu 49	115.6 (7.65)	58.0 (3.83)	42.4 (1.94,1.02)	26.6 (1.99)	δ; 27.1,22.9 (0.33,0.15)
Gln 50	115.7 (8.18)	59.2 (3.90)	29.1 (2.38,1.93)	34.2 (2.58,2.23)	
Gln 51	119.3 (9.29)	58.9 (3.96)	34.1 (2.18,2.08)	34.3 (2.78)	
Ala 52	121.5 (8.80)	55.6 (4.01)	18.4 (1.33)		
Val 53	114.9 (7.35)	66.9 (3.49)	32.2 (2.28)	23.2,20.9 (1.07,0.95)	
Gln 54	116.0 (7.44)	59.0 (3.93)	28.7 (2.27,2.23)	33.6 (2.44)	
Val 55	120.5 (9.18)	66.3 (3.63)	31.8 (2.18)	24.7,21.7 (1.10,0.54)	
Ile 56	119.7 (8.21)	67.1 (3.33)	37.6 (2.01,1.77)	27.5 (1.88)	γ_m; 19.0 (1.35), δ; 17.0 (0.61)
Leu 57	117.4 (8.37)	58.0 (4.08)	42.2 (1.85,1.42)	26.9 (2.18,1.86)	δ; 25.8,23.1 (0.79,0.86)
Gly 58	105.4 (8.29)	46.4 (3.92,3.98)			
Leu 59	122.6 (8.35)	57.9 (4.16)	43.4 (1.88,1.21)	27.3 (0.54)	δ; 23.6,22.9 (0.71,0.81)
Glu 60	117.2 (8.77)	59.5 (3.87)	29.5 (2.36,1.97)	36.5 (2.69,2.35)	
Gln 61	116.1 (7.68)	58.4 (4.06)	27.9 (2.25)	33.5 (2.45)	
Gln 62	117.1 (7.78)	58.7 (4.17)	29.8 (2.21)	34.5 (2.58,2.36)	
Val 63	115.7 (7.96)	65.0 (3.77)	58.5 (2.20)	23.0,21.9 (0.95,0.87)	
Arg 64	118.4 (7.71)	57.8 (4.22)	30.6 (1.89)	27.7 (1.70)	δ; 43.6 (3.23)
Glu 65	117.5 (8.00)	56.8 (4.26)	29.8 (2.05)	35.9 (2.38)	
Arg 66	118.9 (7.86)	56.0 (4.39)	30.6 (1.98,1.83)	27.1 (1.69)	δ; 43.4 (3.24)
Asn 67	123.4 (7.96)	54.7 (4.51)	41.0 (2.82,2.68)		

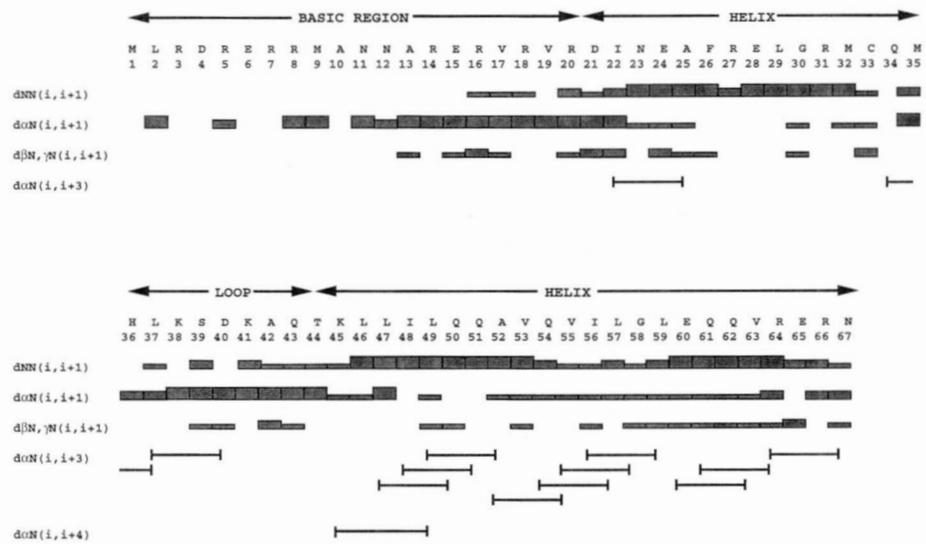

Fig. 3. Summary of short- and medium-range NOEs observed in the 3D ^{15}N-separated NOESY HMQC. Missing bars or lines represent connectivities that are not observed, with the exception of residues Asp 4 and Glu 6, for which we could not obtain resonance assignments.

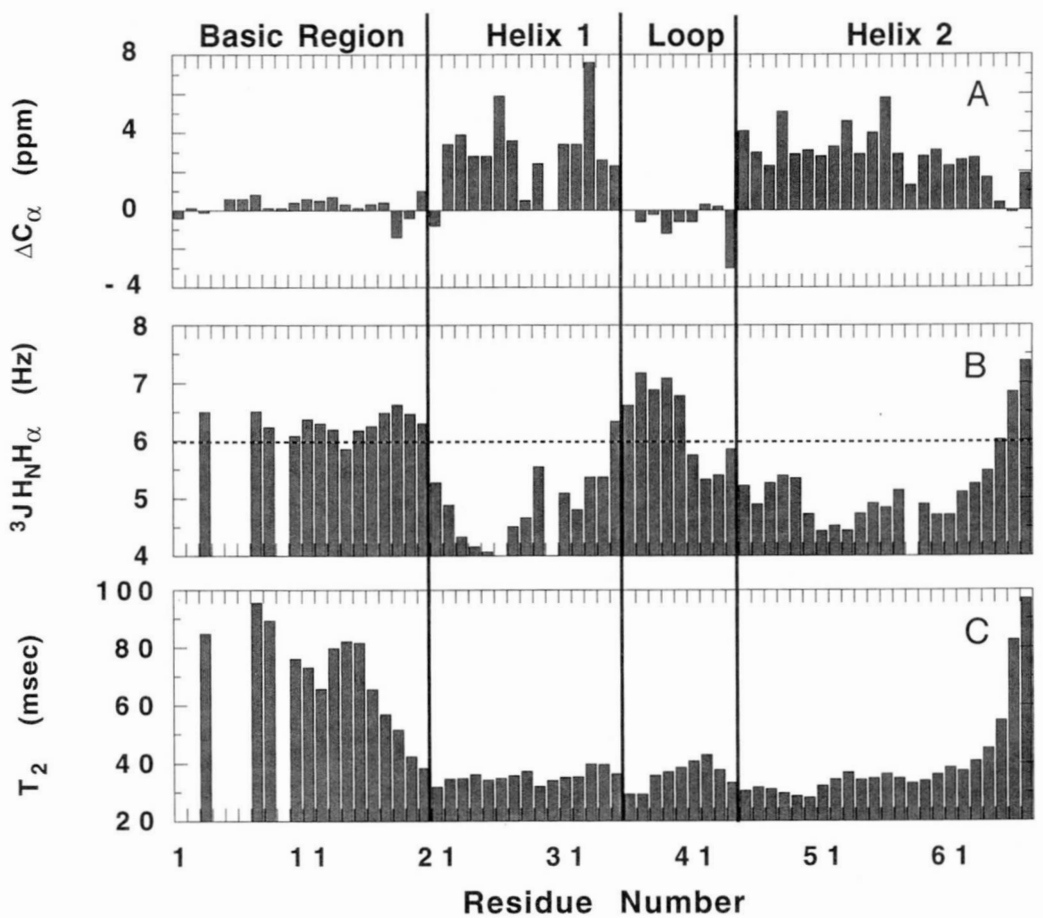

Fig. 4. **A:** Difference in Cα chemical shifts from random coil values (Spera & Bax, 1991; Wishart & Sykes, 1994). **B:** $^3JH_NH_\alpha$ coupling constants. **C:** T_2' relaxation times as a function of amino acid sequence (Billeter et al., 1992).

30 min following dissolution of the protein in D_2O. This is consistent with a relatively dynamic state of the α-helices, including motions that allow for hydrogen bond breakage and exchange with bulk solvent.

Discussion

A challenge in structure determination by NMR spectroscopy is that the protein of interest is well-behaved at high (1–5 mM) concentrations in aqueous solution. This is of special concern for self-associating systems, such as the HLH class of proteins. We undertook a biophysical study to understand the oligomerization states of several HLH-containing proteins (Fairman et al., 1993). These included the b/HLH domains of E47, MyoD, a muscle-specific protein, and the HLH domain from Id, an inhibitor of DNA binding. It became apparent that neither MyoD nor Id were viable candidates for structure determination because both of these protein domains formed tetramers at NMR concentrations. This is not simply an artifact of the elevated concentrations used in the NMR experiments; tetramer formation is a mechanism for effectively inhibiting the formation of an HLH dimer DNA complex. In the case of MyoD, a recent report indicates that it forms even higher-order aggregates, which also may have functional significance (Laue et al., 1995). Starovasnik and coworkers (1992) overcame the aggregation problem and stabilized the MyoD homodimer by crosslinking through disulfide bond formation using native cysteines in the HLH domain. Our studies revealed that E47 does not form tetramers or higher-order oligomers at high concentrations (cf. Fairman et al., 1993) and thus it became the obvious candidate for NMR assignment and structure determination.

Assignments of E47 required the use of 4D heteronuclear ^{15}N- and ^{13}C-separated experiments. This is because the b/HLH domain of E47 is predominantly α-helix and random coil, and is fairly dynamic, resulting in a high degree of spectral overlap. Using 4D methods, we have assigned >95% of all backbone and side-chain ^1H, ^{15}N, and ^{13}C resonances. These assignments are the first step in determining the solution structure of the homodimer and provide an additional contribution to the present database of chemical shifts.

Combined information from short- and medium-range NOEs (Fig. 3), C_α chemical shifts (Fig. 4A), and $^3JH_NH_\alpha$ coupling constant data (Fig. 4B) allowed us to define regions of secondary structure. The basic region of E47 is largely random coil, but the general trend of positive C_α chemical shifts suggests that this domain has limited helical propensity. This is reasonable given that the basic region binds to its cognate DNA-binding site as an α-helix. Helix 1 of the HLH domain is confined to 15 residues from Asp 21 through Met 35. This α-helix is not well-defined as judged by the lack of H_NH_N $(i, i + 3)$ and $(i, i + 4)$ NOEs (Fig. 3). It is not clear whether this is caused by dynamics or deviations from ideal α-helix formation; the presence of a glycine in the C-terminal half of helix 1 may play a role here. In contrast, helix 2, which encompasses 23 residues from Lys 45 to Asn 67, is defined significantly better; this may be due in part to the fact that it is almost twice as long as helix 1 and significantly more hydrophobic. The loop region has a pattern of NOEs (Fig. 3) that is consistent with an extended backbone structure. The similarity of T_2' values between the loop residues and the α-helices is consistent with a parallel orientation of helix 1 and 2, because this orientation requires the loop to form an extended structure, spanning one end of the bundle to the other. By contrast, if the α-helices adopted an antiparallel orientation, the loop is predicted to be more flexible, and this would be reflected in increased values of T_2' for the loop relative to the α-helices.

We compared our secondary structure assignment with that determined from the co-crystal structure of E47 bound to DNA (Ellenberger et al., 1994) and found excellent overall agreement. Our protein construct spans residues Leu 332 (Leu 2 in our numbering system) with an N-terminal Met to Asn 397 (Asn 67), whereas the E47 used in the co-crystal structure spans residues Arg 335 to Gln 392, but these differences are unlikely to be important structurally. As expected, the biggest difference is in the basic region of E47, which is largely unstructured, as suggested by earlier work (Anthony-Cahill et al., 1992). Residues 1–20 show evidence of nascent helix, but significant helical structure does not begin until Asp 21. This may be a consequence of the amino-terminal valine residues, which have a low α-helix propensity. Additionally, in the absence of DNA, the arginine residues may further destabilize helix formation, because their positive charges can interact unfavorably with the amino-terminal end of the α-helix macrodipole and with each other. The carboxy-terminal end of helix 1 extends to Met 35, similar to that in the co-crystal structure of E47; in the crystal structure, the very next residue, His 36, forms the only intersubunit salt bridge to Glu 60 at the C terminus of helix 2′. The loop region is comprised of nine residues in an extended conformation; the relative rigidity of the loop is consistent with the constraint of joining the ends of helix 1 and helix 2 across the four-helix bundle in a manner similar to that found in the co-crystal structure. The amino-terminal end of helix 2 is punctuated with a threonine residue, which is often found as an N-cap amino acid (Richardson & Richardson, 1988) and fulfills this function in the co-crystal structure. The apparent helix fraying observed at the C-terminal end of helix 2 is not unusual because this is often seen in coiled coil domains; however, it may also reflect imprecise termination of the polypeptide chain with respect to the position of this helix in our construct.

An unexpected finding was the high degree of dynamics of the HLH domain, which is manifest by rapid hydrogen–deuterium exchange and relatively weak H_N-H_N NOEs. The stability of the E47 homodimer is 4.5 μM (Fairman et al., 1993), representing a free energy of 7.2 kcal/mol (the amount of unfolded monomer at 1.5 mM protein is <4% based on this stability). Thus, it is unlikely that the observed rapid exchange can be ascribed to a low thermodynamic stability with the rate proceeding through an Ex2 mechanism, and more likely reflects local fluctuations with exchange following an Ex1 mechanism. Within minutes, all amide protons are exchanged to deuterium, whereas for most well-folded proteins, at least a core of amide protons typically remain resistant to exchange. On the other hand, this observation is not inconsistent with recent thermodynamic investigations of protein–DNA complex formation that demonstrate the importance of large negative heat capacity changes (Burley, 1994; Spolar & Record, 1994). The negative heat capacity change has been attributed to burial of nonpolar surface area and displacement of cations on binding. However, the magnitude of the change is frequently too large to be accounted for by rigid body association of protein and DNA, and thus a paradigm for site-specific recognition has been proposed that includes conformational changes or "induced fit" of the protein or nucleic acid (Spolar & Record, 1994). This provides a potentially important mechanism for modulating specificity and stability because the entropy loss upon complexation must be compensated with favorable free energy changes from interactions between and within the protein and DNA. Clearly, the basic region

undergoes a large conformational change from a predominantly random coil to a highly structured α-helix upon DNA binding. However, our data also suggest that a more subtle level of flexibility exists within the HLH domain as well, and this may also have a role in site-specific recognition of DNA and/or discrimination of dimerization partners.

The information about dynamics taken from our NMR experiments, when combined with the sequence information and details from the co-crystal structure of E47 with DNA, affords us the opportunity to better understand basic principles of helix-packing interactions. In the crystal structure, stabilizing contacts between the two subunits are made by hydrophobic residues at the *a* and *d* position of helix 1, and the *a*, *d*, *e*, and *g* positions of helix 2 and 2′. The only intersubunit salt bridge is made between His 36 of helix 1 and Glu 60 of helix 2′ (Ellenberger et al., 1994). Additional hydrogen bonds are evident between the loop and the helices of each HLH subunit, but are not involved directly in intersubunit stability. The emphasis of the role of hydrophobic residues on structure is somewhat reminiscent of that seen in de novo-designed 4-helix bundles (Regan & Degrado, 1988) and coiled coils, whose assembly is driven primarily by hydrophobic interactions. These proteins tend to be extremely stable, but lack structural uniqueness, as evidenced by a variety of experimental determinants, including rapid H/D exchange (Handel et al., 1993). The introduction of buried or interfacial salt bridges or polar groups in these designed proteins imparts specificity to the fold and greater resistance to exchange, despite the fact that stability is generally sacrificed (O'Shea et al., 1991; Handel et al., 1993; Lumb & Kim, 1995; Raleigh et al., 1995). In the case of the E47 homodimer, three acidic residues (Asp 21, Glu 24, and Glu 28) lie along one face of helix 1 close to helix 2′, whereas basic residues (R20, R27, and R31) lie on the face nearest helix 2. As for the designed protein, α4, these residues form salt bridges along the α-helix, but not between the α-helices, and therefore do not contribute to interactions that stabilize the orientation of the α-helices relative to one another (Ellenberger et al., 1994). However, based on mutational studies (Shirataka et al., 1993), it has been proposed that these charged residues participate in important intersubunit salt bridges that mediate heterodimerization with other HLH proteins, such as MyoD, and thus, the relative flexibility of the E47/MyoD heterodimer would be less than that of the E47 homodimer. Whether this is true remains to be seen; however, it does seem plausible that in the same way that a balance between unfavorable entropic changes versus favorable binding interactions contribute to the specificity and stability of protein–DNA interactions, similar rules may be operable in specifying protein–protein interactions in the HLH family of transcription factors.

Materials and methods

Cloning and protein preparation

A 198-base pair DNA fragment encoding a 66-amino acid sequence that encompasses the basic and HLH domains of E47 (residues 332–397 from the native sequence) was subcloned from pE47s (Sun & Baltimore, 1991) using standard PCR methods (Saiki et al., 1988). *Nde* I and *Bam*H I sites were engineered at the 5′ and 3′ ends of this DNA fragment for insertion into pET-24a (Novagen) for overexpression utilizing the T7 expression system (Studier & Moffatt, 1986); this construct was verified by DNA sequencing. Insertion of the sequence for the b/HLH domain of E47 into pET-24a introduced a methionine at the amino terminus, resulting in a sequence 67 amino acids in length.

The vector containing the E47 construct was transformed into the BL21 (DE3) *E. coli* strain for overexpression (Moffatt & Studier, 1987). The E47 fragment was overexpressed in minimal medium for incorporation of ^{15}N and/or ^{13}C using $[^{15}N](NH_4)_2SO_4$ (Isotec, Inc.) and $[^{13}C_6]$d-glucose (Cambridge Isotope Laboratories). Cultures were grown in a Bioflo III batch/continuous fermentor (New Brunswick Scientific) at 37 °C. The cell paste from the growths was resuspended in 50 mM HEPES, pH 7.6, 1 mM EDTA, 0.1 M KCl, 5 mM DTT, 1 mM $MgCl_2$, 0.1% Triton X-100 buffer using a 25-mL Wheaton hand homogenizer. The cells were subjected to three freeze/thaw cycles and incubated on ice for 30 min with the addition of 6 mg lysozyme to assure complete lysis. The lysate was treated with 750 μg each of deoxyribonuclease I (Boehringer-Mannheim) and ribonuclease A (Sigma). The cell debris was pelleted by centrifugation and the supernatant was diluted 1:2 with column buffer (50 mM HEPES, pH 7.6, 0.4 mM NaCl, 1 mM DTT, 1 mM EDTA, 1 mM PMSF) and fractionated on an S-Sepharose (Pharmacia) column using a 0.4–2.0 M NaCl gradient. The appropriate fractions were purified further by water–acetonitrile reverse phase chromatography using a Vydac C18 semi-preparative column attached to a Rainin HPLC. The E47 protein fragment was >95% pure as judged by analytical HPLC. The identity of the purified material was confirmed by electrospray mass spectrometry. Yields from the different preparations were 4–10 mg of protein per liter of cell growth.

The E47 protein fragment was characterized structurally by CD spectrometry (Aviv 62DS CD spectropolarimeter). Analytical reverse-phase HPLC (Hewlett-Packard HP1090) was used to probe the extent of oxidation of the lone cysteine in the E47 sequence. The NMR samples contained 1.5–2.0 mM protein in 20 mM deuterated DTT and 10% D_2O. Samples were left unbuffered because of problems with solubility. Samples were sparged and blanketed with argon in the NMR tube to prevent oxidation of the free cysteine. Under these conditions they were stable for several months when stored at 4 °C.

NMR methods: Experiments

NMR experiments were performed with 1.5–2.0 mM samples of E47 at 30 °C on a Bruker AMX 600 spectrometer equipped with external fourth channel class A′ amplifiers. Water suppression was achieved with weak presaturation and/or scrambling pulses (Messerle et al., 1989) because gradient equipment was unavailable at the time. Quadrature detection in indirect dimensions was achieved by States-TPPI except for certain ^{13}C dimensions where axials are better placed at the center of the spectrum rather than at the edge. Initial timings for evolution periods were delayed by a full dwell to improve baseline characteristics (linear phase correction = −360 degrees) or by half a dwell when sign discrimination of aliased crosspeaks was desired (linear phase correction = −180 degrees). Alternatively, refocusing delays were incorporated to eliminate the need for any phase correction. Nitrogen and carbon chemical shifts were referenced with respect to external standards as described previously (Live et al., 1984) using liquid ammonia and tetramethylsilane as chemical shift standards. Additional details of the data collection for all NMR experiments are summarized in Table 2.

The ^1H-^2D exchange experiment was done by hydrating with a D_2O buffer, protonated lyophilized ^{15}N-labeled E47 (pH 5.5). The

Table 2. *NMR data acquisition parameters for all experiments*

Experiment	Dims[b]	Spectral width (Hz)	Carrier (ppm)	Acq time (ms)	Complex points	Mix time (ms)	Transients	References
2D ^{15}N-^{1}H HSQC	^{15}N-t1	625.0	117.8	204.8	128		16	Bax et al. (1990b)
	^{1}H-t2	8,064.52	4.72	63.5	512			Norwood et al. (1990)
2D JH_NH_α modulated HSQC	^{5}N-t1	625.0	117.8	51.2	32		256	Billeter et al. (1992)
	^{1}H-t2	8,064.52	4.72	63.5				
3D^{15}N-NOESY HMQC	^{1}H-t1	8,064.52	4.72	12.4	100	125	32	Marion et al. (1989)
3D ^{15}N-TOCSY HMQC	^{15}N-t2	625.0	117.8	20.0	16	52		
	^{1}H-t3	8,064.52	4.72	63.5	512			
3D ^{15}N-HNHB	^{1}H-t1[a]	5,000.0	4.72	12.8	64		32	Archer et al. (1991)
	^{15}N-t2	625.0	117.8	20.0	16			
	^{1}H-t3	8,064.52	4.72	63.5	512			
3D ^{15}N-HNHA	^{15}N-t1	625.0	117.8	40.0	32		256	Billeter et al. (1992)
	^{1}H-t2	8,064.52	4.72	63.5	512			
3D ^{13}C-TOCSY HMQC	^{1}H-t1	6,250.0	4.72	20.5	128	17m	16	Bax et al. (1990a)
	^{13}C-t2[a]	4,000.0	46.5	8.0	32			
	^{1}H-t3	6,250.0	4.72	40.0	256			
4D HCANNH	^{1}H-t1	1,000.0	4.72	16.0	16		16	Boucher et al. (1992a, 1992b)
	^{13}C-t2	2,500.0	59.6	6.4	16			
	^{15}N-t3	625.0	117.8	10.0	8			
	^{1}H-t4	8,064.52	4.72	63.5	512			
4D HCA(CO)NNH	^{1}H-t1	1,000.0	4.72	16.0	16		16	Boucher et al. (1992a, 1992b)
	^{13}C-t2	2,500.0	59.6	6.4	16			
	^{15}N-t3	625.0	117.8	10.0	8			
	^{1}H-t4	8,064.52	4.72	63.5	512			
4D HCC(CO)NNH	^{1}H-t1[a]	3,125.0	4.72	5.1	16		16	Clowes et al. (1993)
	^{13}C-t2	3,333.3		3.0	10			
	^{15}N-t3	625.0	117.8	20.0	16			
	^{1}H-t4	8,064.52	4.72	63.5	512			

[a]Quadrature detection achieved with States rather than States-TPPI.
[b]Dimension.

pH was adjusted rapidly to 5.5 and the first slice of an ^{1}H-^{15}N HSQC recorded. No further experiments were collected because virtually all of the amide protons were gone within the 30 min or less it took to prepare the sample and collect the data.

NMR methods: Data processing and analysis

Data were treated in the directly detected dimension with a convolution difference function to reduce the water resonance, a shifted sine or squared sine bell apodization function, and conventional Fourier transformation using the program Felix1.1 (Hare Research). For 3D experiments, the indirectly detected dimensions were processed with a 2D maximum entropy algorithm (Laue et al., 1986). Four-dimensional experiments were processed with 2D maximum entropy for the two most underdigitized dimensions, and linearly extended in the third indirect dimension using the linear prediction/extension algorithm in Felix2.1 (Biosym Technologies), or simply windowed and processed with a conventional FT. All processing was done on a Silicon Graphics R4000 Indigo workstation.

For chemical shift assignments, spectra were analyzed manually in Felix1.1 using in-house written macros. Intensities from the $^{3}JH_NH_\alpha$-modulated HSQC spectra were also analyzed in Felix1.1, but the values of T_2' and the $^{3}J_{HNH\alpha}$ coupling constants were determined by a least-squares fit of the intensities of the crosspeaks, in a similar fashion to Billeter et al. (1992).

Acknowledgments

We thank Dr. Xiao-Hong Sun for the pE47s vector containing the full-length E47 gene and Dr. Harold Weintraub for the pRK171a vector. We also thank Steve Brenner, Peter Domaille, and Bill DeGrado for many helpful discussions on all aspects of this work. T.M.H. is a PEW Scholar in the Biomedical Sciences; support for T.M.H. was provided in part by the PEW Charitable Trusts, grant 93-05252, and by the University of California Cancer Research Coordinating Committee funds. R.F. acknowledges support from NIH grant GM14321.

References

Anthony-Cahill SJ, Benfield PA, Fairman R, Wasserman ZR, Brenner SL, Altenbach C, Hubbell WL, Stafford WF III, DeGrado WF. 1992. Molecular characterization of helix-loop-helix peptides. *Science 255*:979–983.

Archer SJ, Ikura M, Torchia DA, Bax A. 1991. An alternative 3D-NMR technique for correlating backbone N-15 with side chain H-beta resonances in larger proteins. *J Magn Reson 95*:636–641.

Bain G, Maandag ECR, Izon DJ, Amsen D, Kruisbeek AM, Weintraub BC, Krop I, Schlissel MS, Feeney AJ, van Roon M, van der Valk M, te Riele HPJ, Berns A, Murre C. 1994. E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements. *Cell 79*:885–892.

Bax A, Clore GM, Gronenborn AM. 1990a. 1H-1H correlation via isotropic mixing of 13C magnetization, a new 3-dimensional approach for assigning 1H and 13C spectra of 13C-enriched proteins. *J Magn Reson 88*:425–431.

Bax A, Ikura M, Kay LE, Torchia DA, Tschudin R. 1990b. Comparison of different modes of 2-dimensional reverse-correlation NMR for the study of proteins. *J Magn Reson 86*:304–318.

Billeter M, Neri D, Otting G, Qian Q, Wüthrich K. 1992. Precise vicinal coupling constants $^3JH_NH_\alpha$ in proteins from nonlinear fits of J-modulated ^{15}N-^1H COSY experiments. *J Biomol NMR 2*:257–274.

Bishop P, Jones C, Ghosh I, Chmielewski J. 1995. Synthesis of the basic-helix-loop-helix region of the immunoglobulin enhancer binding protein E47 and evaluation of its structural and DNA binding properties. *Int J Pept Prot Res 46*:149–154.

Blackwood E, Eisenman RN. 1991. MAX—A helix-loop-helix protein that forms a sequence specific DNA-binding complex with myc. *Science 251*: 1211–1217.

Boucher W, Laue ED, Campbell-Burk S, Domaille PJ. 1992a. 4-Dimensional heteronuclear triple resonance NMR methods for the assignment of backbone nuclei in proteins. *J Am Chem Soc 114*:2262–2264.

Boucher W, Laue ED, Campbell-Burk SL, Domaille PJ. 1992b. Improved 4D-NMR experiments for the assignment of backbone nuclei in C-13/N-15 labelled proteins. *J Biomol NMR 2*:631–637.

Burley SK. 1994. Plus ça change, plus c'est la même chose. *Nature Struct Biol 1*:207–208.

Chou PY, Fasman GD. 1978. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv Enzymol 47*:45–148.

Clowes RT, Boucher W, Hardman CH, Domaille PJ, Laue ED. 1993. A 4D HC-C(CO)NNH experiment for the correlation of aliphatic side chain and backbone resonances in C-13/N-15 labelled proteins. *J Biomol NMR 3*:349–354.

Ellenberger T, Fass D, Arnaud M, Harrison SC. 1994. Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. *Genes & Dev 8*:970–980.

Fairman R, Beran-Steed RK, Anthony-Cahill SJ, Lear JD, Stafford WF, DeGrado WF, Benfield PA, Brenner SL. 1993. Multiple oligomeric states regulate the DNA-binding of helix-loop-helix peptides. *Proc Natl Acad Sci USA 90*:10429–10433.

Farmer K, Catala F, Wright WE. 1992. Alternative multimeric structures affect myogenin DNA binding activity. *J Biol Chem 267*:5631–5636.

Ferré-D'Amaré AR, Pognonec P, Roeder RG, Burley SK. 1994. Structure and function of the b/HLH/Z domain of USF. *EMBO J 13*:180–189.

Ferré-D'Amaré AR, Prendergast GC, Ziff EB, Burley SK. 1993. Recognition by MAX of its cognate DNA through a dimeric b/HLH/Z domain. *Nature 363*:38–45.

Fisher DE, Carr CS, Parent LA, Sharp PA. 1991. TFEB has DNA-binding and oligomerization properties of a unique helix-loop-helix/leucine-zipper family. *Genes & Dev 5*:2342–2352.

Gibson TJ, Thompson JD, Abagyan RA. 1993. Proposed structure for the DNA-binding domain of the helix-loop-helix family of eukaryotic gene regulatory proteins. *Protein Eng 6*:41–50.

Halazonetis TD, Kandil AN. 1992. Predicted structural similarities of the DNA binding domains of c-Myc and endonuclease Eco RI. *Science 255*:464–466.

Handel TM, Williams SA, DeGrado WF. 1993. Metal ion-dependent modulation of the dynamics of a designed protein. *Science 261*:879–885.

Hsu HL, Wadman I, Tsan JT, Baer R. 1994. Positive and negative transcriptional control by the TAL1 helix-loop-helix protein. *Proc Natl Acad Sci USA 91*:5947–5951.

Jones N. 1990. Transcriptional regulation by dimerization: Two sides to an incestuous relationship. *Cell 61*:9–11.

Kadesch T. 1993. Consequences of heteromeric interactions among helix-loop-helix proteins. *Cell Growth Differ 4*:49–55.

Lamb P, McKnight SL. 1991. Diversity and specificity in transcriptional regulation: The benefits of heterotypic dimerization. *Trends Biochem Sci 16*:417–422.

Lassar AB, Davis RL, Wright WE, Kadesch T, Murre C, Voronova A, Baltimore D, Weintraub H. 1991. Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins in vivo. *Cell 66*:305–315.

Laue ED, Mayger MR, Skilling J, Staunton J. 1986. Reconstruction of phase sensitive two-dimensional NMR spectra by maximum entropy. *J Magn Reson 68*:14–29.

Laue TM, Starovasnik MA, Weintraub H, Sun XH, Snider L, Klevitt R. 1995. MyoD forms micelles which can dissociate to form heterodimers with E47: Implications of micellization on function. *Proc Natl Acad Sci USA 92*:11824–11828.

Live DH, Davis DG, Agosta WC, Cowburn D. 1984. Long-range hydrogen bond mediated effects in peptides: 15N NMR study of gramacidin S in water and organic solvents. *J Am Chem Soc 106*:1934–1941.

Lumb KJ, Kim PS. 1995. A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled coil. *Biochemistry 34*:8642–8648.

Ma PCM, Rould MA, Weintraub H, Pabo CO. 1994. Crystal structure of MyoD b/HLH domain-DNA complex: Perspectives on DNA recognition and implications for transcriptional activation. *Cell 77*:451–459.

Marion D, Driscoll PC, Kay LE, Wingfield PT, Bax A, Gronenborn AM, Clore GM. 1989. Overcoming the overlap problem in the assignment of H-1 NMR spectra of larger proteins by use of 3-dimensional heteronuclear 1H-15N Hartmann–Hahn multiple quantum coherence and nuclear Overhauser multiple quantum coherence spectroscopy—Application to interleukin-1-beta. *Biochemistry 28*:6150–6156.

Messerle BA, Wider G, Otting G, Weber C, Wüthrich K. 1989. Solvent suppression using a spin lock in 2D and 3D NMR spectroscopy with H_2O solutions. *J Magn Reson 85*:608–613.

Moffatt BA, Studier FW. 1987. T7 lysozyme inhibits transcription by T7 RNA polymerase. *Cell 49*:221–227.

Murre C, McCaw PS, Baltimore D. 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, *daughterless*, MyoD, and *myc* proteins. *Cell 56*:777–783.

Norwood TJ, Boyd J, Heritage J, Soffe N, Campbell ID. 1990. Comparison of techniques for 1H-detected heteronuclear 1H-15N spectroscopy *J Magn Reson 87*:488–501.

O'Shea EK, Klemm JD, Kim PS, Alber T. 1991. X-ray structure of the GCN4 leucine zipper, a 2-stranded, parallel coiled coil. *Science 254*:539–544.

Phillips SEV. 1994. Built by association: Structure and function of helix-loop-helix DNA-binding proteins. *Structure 2*:1–4.

Prendergast GC, Lawe D, Ziff E. 1991. Association of MYN, the murine homolog of MAX, with C-MYC stimulates methylation sensitive DNA-binding and ras cotransformation. *Cell 65*:395–407.

Raleigh DP, Betz SF, Degrado WF. 1995. A de novo designed protein mimics the native state of natural proteins. *J Am Chem Soc 117*:7558–7559.

Regan L, DeGrado WF. 1988. Characterization of a helical protein designed from first principles. *Science 241*:976–978.

Richardson JS, Richardson DC. 1988. Amino acid preferences for specific locations at the ends of α-helices. *Science 240*:1648–1652.

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science 239*:487–491.

Shen CP, Kadesch T. 1995. B-cell-specific DNA binding by an E47 homodimer. *Mol Cell Biol 15*:4518–4524.

Shirataka M, Friedman FK, Wei Q, Paterson BM. 1993. Dimerization specificity of myogenic helix-loop-helix DNA-binding factors directed by nonconserved hydrophilic residues. *Genes & Dev 7*:2456–2470.

Spera S, Bax A. 1991. Empirical correlation between protein backbone conformation and C-alpha and C-beta C-13 nuclear magnetic resonance chemical shifts. *J Am Chem Soc 113*:5490–5492.

Spolar RS, Record TR. 1994. Coupling of local folding to site-specific binding of proteins to DNA. *Science 263*:777–783.

Starovasnik MA, Blackwell TK, Laue TM, Weintraub H, Klevit RE. 1992. Folding topology of the disulfide-bonded dimeric DNA-binding domain of the myogenic determination factor MyoD. *Biochemistry 31*:9891–9903.

Studier FW, Moffatt BA. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol 189*:113–130.

Sun XH, Baltimore D. 1991. An inhibitory domain of E12 transcription factor prevents DNA binding in E12 homodimers but not in E12 heterodimers. *Cell 64*:459–470.

Vinson CR, Garcia C. 1992. Molecular model for DNA recognition by the family of basic-helix-loop-helix-zipper proteins. *New Biol 4*:396–403.

Wishart DS, Sykes BD. 1994. The 13C chemical-shift index: A simple method for the identification of protein secondary structure using 13C chemical-shift data. *J Biomol NMR 4*:171–180.

Wolberger C. 1994. b/HLH without the zip. *Nature Struct Biol 1*:413–416.