Crystal structure of transcription factor E47: E-box recognition by a basic region helix–loop–helix dimer

Tom Ellenberger,1,3,4 Deborah Fass,1 Martha Arnaud,1 and Stephen C. Harrison1,2

1Department of Biochemistry and Molecular Biology, Harvard University; 2The Howard Hughes Medical Institute, Cambridge, Massachusetts 02138 USA
Crystal structure of transcription factor E47: E-box recognition by a basic region helix–loop–helix dimer

Tom Ellenberger,1,3,4 Deborah Fass,1 Martha Arnaud,1 and Stephen C. Harrison1,2

1Department of Biochemistry and Molecular Biology, Harvard University; 2The Howard Hughes Medical Institute, Cambridge, Massachusetts 02138 USA

A large group of transcription factors regulating cell growth and differentiation share a dimeric α-helical DNA-binding domain termed the basic region helix–loop–helix (bHLH). bHLH proteins associate as homodimers and heterodimers having distinctive DNA-binding activities and transcriptional activities that are central to the regulated differentiation of a number of tissues. Some of the bHLH residues specifying these activities have been identified, but a full understanding of their function has awaited further structural information. We report here the crystal structure of the transcription factor E47 bHLH domain bound to DNA. The bHLH of E47 is a parallel, four-helix bundle with structural features that distinguish it from the bHLH–zipper protein Max. The E47 dimer makes nonequivalent contacts to each half of the -CACCTG- binding site. Sequence discrimination at the center of the E box may result from interaction with both the DNA bases and the phosphodiester backbone.

[Key Words: E47; E-protein; bHLH; E box]

Received February 2, 1994; revised version accepted March 10, 1994.

The basic region helix–loop–helix (bHLH) proteins E47, E12, HEB, and da (the daughterless gene product) are expressed in a variety of cell types and bind to related DNA sites known as E boxes [Caudy et al. 1988; Murre et al. 1989a; Henthorn et al. 1990; Hu et al. 1992]. These widely distributed E proteins dimerize with the tissue-restricted bHLH proteins MyoD, SC [achaete-scute T4], and tal-1/SCL, promoting their DNA-binding activities [Murre et al. 1989b; Cabrera and Alonso 1991; Hsu et al. 1991; Sun and Baltimore 1991; Aplan et al. 1992; Hu et al. 1992]. Heterodimers of E proteins and the tissue-specific bHLH factors have roles in Drosophila neural development and sex determination, vertebrate muscle development, and the differentiation of myeloid and erythroid cells [Cabrera and Alonso 1991; Lassar et al. 1991; for review, see Weintraub et al. 1991; Kadesch 1992; Murre 1992]. The positively acting E proteins are inhibited by the HLH proteins Id and emc, which sequester them as inactive heterodimers [Benezra et al. 1990; Garrell and Modolell 1990; Sun et al. 1991; Wilson et al. 1991; Jen et al. 1992; Pesce and Benezra 1993]. Therefore, the DNA-binding and transcriptional activities of bHLH proteins depend on their dimerization preferences.

Residues within the HLH segment specify the dimerization activities of bHLH proteins [Murre et al. 1989b; Davis et al. 1990; Voronova and Baltimore 1990; Shirakata et al. 1993]. The crystal structure of a Max/DNA complex [Ferré-D'Amare et al. 1993] established that the HLH is a parallel, four-helix bundle with conserved hydrophobic residues at the core of this globular domain. In addition to the Max, the HLH has a carboxy-terminal coiled-coil (termed the zipper, Fig. 1) that is required for efficient dimerization of Max and other bHLH–zipper proteins [Dang et al. 1989; Gregor et al. 1990; Hu et al. 1990; Blackwood and Eisenman 1991; Beckmann and Kadesch 1991; Fisher et al. 1991; Pendergast et al. 1991; Davis and Halazonetis 1992; Kato et al. 1992; Sirito et al. 1992, Ayer et al. 1993]. E proteins lacking a zipper form stable dimers with other "zipperless" bHLH proteins, but they do not dimerize with bHLH–zipper proteins [Murre et al. 1989b; Beckmann and Kadesch 1991; Blackwood and Eisenman 1991; Fisher et al. 1991]. These findings suggest that the HLH segments of bHLH and bHLH–zipper proteins have distinct structural features that may account for their different dimerization strengths and preferences.

Basic region residues determine the DNA-binding activities of bHLH proteins [Murre et al. 1989b; Davis et al. 1990; Voronova and Baltimore 1990]. The binding sites of bHLH and bHLH–zipper proteins define the conserved E-box sequence motif, -CAXXTG-, in which the central 2 bp are specified by each protein [Blackwell and Weintraub 1990; Alex et al. 1992; Dang et al. 1992; Halazonet-
tis and Kandil 1992, Van Antwerp et al. 1992, Blackwell et al. 1993). DNA-binding specificity at the center of the E box is markedly influenced by a single basic region residue that is arginine in Max, Myc, and other bHLH–zipper proteins. In the Max homodimer/DNA crystal structure, this arginine directly contacts the central base pairs of the E box in a symmetric manner (Ferré-D’Amaré et al. 1993). The analogous residue of most bHLH proteins is smaller and nonpolar (V347 of E47; Fig. 1). A mechanism for binding site selection by the E proteins and other bHLH proteins lacking arginine at this position is not readily discerned from the Max/DNA structure.

We report here the crystal structure of the transcription factor E47 bHLH dimer complexed with DNA. The bHLH domain of E47 is a parallel, four-helix bundle with two protruding basic region helices. E47 contacts the major groove of its binding site in a manner similar to the Max bHLH–zipper dimer (Ferré-D’Amaré et al. 1993), although E47 lacks a zipper segment, and helix 1 of its HLH segment is one turn longer than helix 1 of Max. This carboxy-terminal extension of helix 1 creates additional dimer contacts between E47 subunits. A network of hydrogen bonds bridging helices 1 and 2 with the loop further stabilizes the E47 bHLH domain. E47 and Max contact the conserved outer base pairs of the E box in a similar manner. However, E47 contacts the central base pairs in an asymmetric fashion using a conserved arginine that interacts with the DNA backbone in the Max/DNA complex. The crystal structures of the E47 and Max bHLH/DNA complexes provide a basis for understanding the diverse dimerization and DNA-binding properties of these and other bHLH domains.

**Results**

Crystals of the E47 bHLH complexed with a variety of oligonucleotides were grown in hanging drops by vapor diffusion. However, most of these crystals were small and showed a marked lattice disorder in a plane perpendicular to the helical axes of the DNA. Larger, well-ordered crystals of the bHLH peptide complexed with the 11-mer DNA shown in Figure 1 were obtained under similar conditions. These crystals diffract isotropically to spacings of >2.5 Å resolution. The crystal structure of the E47/DNA complex was determined by multiple isomorphous replacement, and it has been refined at 2.8 Å resolution, as described in the Materials and methods. 2Fo–Fc electron density calculated from the final model is shown in Figure 2.

**Structure of the E47 dimer**

The E47 bHLH domain symmetrically positions two α-helices in the major groove of the DNA-binding site (Fig. 3). The dyad axis of the bHLH dimer is normal to the DNA helical axis and is nearly coincident with the pseudodyad of the -CACCTG- DNA site. Each subunit lies in the major groove over its half-site in a similar orientation. The E47 dimer is stabilized by van der Waals interactions and hydrogen bonds between HLH helices 1 and 2. Hydrophobic residues are located at residue positions g, a, d, and e in helix 2 of the HLH (Fig 1). Residues a and d pack in a side-by-side manner against residues a’ and d’ of the opposing helix 2’. Near the carboxyl terminus of the dimer, this packing interaction resembles a coiled-coil, but helices 2 and 2’ diverge toward their amino termini to accommodate helix 1 (Fig. 4). This divergence changes the geometry of side chain packing so that conserved alamines at positions d and d’ (A382; Fig. 4) are in a head-to-head arrangement. Helix 1 of each subunit lies in the cleft formed by the divergence of helices 2 and 2’. Nonpolar residues a and d of helix 1 interdigitate with residues g, a, d, and e of helices 2 and 2’, completing the four-helix bundle. The buried surface area of the E47 HLH dimer (1721 Å²; calculated with a 1.4 Å radius probe) is nearly the same as that buried in the leucine zipper of the GCN4 basic region leucine zipper dimer (1745 Å²; Ellenberger et al. 1992).

In comparison to Max, E47 has one additional turn at the carboxyl terminus of helix 1 of the HLH (Fig. 5; Ferré-D’Amaré et al. 1993). This extra helix turn increases the dimer contact surface of E47 and it allows for an additional polar contact between E47 subunits. Histidine 366, located near the carboxyl terminus of helix 1, forms a salt bridge with E390 of helix 2’ (Fig. 5). This histidine/glutamic acid pair is present in all of the E proteins [E47, E12, HEB, and da; Fig. 1], and these dimerize efficiently without a zipper [Murre et al. 1989b; Hu et al. 1992]. The shorter helix 1 of Max terminates in a proline, and the succeeding residues are part of the loop joining helix 1 to helix 2. The E47 loop segment is an extended strand of 6 residues that follows a relatively straight path from the carboxyl terminus of helix 1 to the amino terminus of helix 2. Within the loop, Q373 participates in hydrogen bonds with the carbonyl oxygen of G360 [helix 1] and the side chains of Q364 [helix 2] and Q381 [helix 2; Fig. 6]. This network of hydrogen bonds stabilizes the conformation of the loop as well as the orientation of helices 1 and 2 within each subunit. Polar residues on the surface of the loop contact symmetry-related molecules in the crystal.

The amino terminus of HLH helix 2 is defined by T374, which forms a hydrogen bond with the amide nitrogen of leucine 377 in a typical “N-cap” arrangement (Fig. 7). The axis of helix 2 points toward phosphate 3 of the DNA half-site opposite that engaged by helix 1. The amide nitrogen of K375, the second residue in helix 2, donates a hydrogen bond to this phosphate (Fig. 7). The close apposition of helix 2, and the DNA extends the footprint of each subunit in a 5° direction across the dyad of the binding site. The side chain of K375 drapes across the face of a conserved helix 1 phenylalanine (F356) and donates hydrogen bonds to N353 of helix 1 and to the phosphate of adenine 2 in the adjacent half-site (Fig. 7). These three hydrogen bonds from K375 anchor helix 2 to the DNA, and they constrain the position of helix 1. Lysine-375 is highly conserved in the bHLH protein family (Fig. 1); mutation to alanine eliminates both dimer-
Figure 1. (A) An alignment of bHLH and bHLH–zipper protein sequences highlighting the conserved hydrophobic residues at positions a and d of HLH helices 1 and 2, and the neighboring e and g residues of helix 2. These residues form the core of the HLH four-helix bundle and they are conserved among bHLH and bHLH–zipper proteins. Residues facing the major groove in the E47/DNA complex include N342, E345, R346, R348, and V349. With the exception of E345, these residues vary in the bHLH and bHLH–zipper proteins, and they are likely to account for the distinct DNA-binding preferences of these proteins. The length of helix 1 and the length and composition of the loop are different in the E47 and Max bHLH segments because of a proline at the carboxyl terminus of Max helix 1 and a tyrosine (at the position of E47 V385) that stacks on the carboxyl terminus of helix 1 in Max. [B] The 11-mer oligonucleotide in the E47/DNA complex crystals contains the preferred binding site -CACCTG- (Blackwell and Weintraub 1990) and is numbered from the center of the binding site.

Figure 2. Stereo view of the final 2Fo–Fo electron density around E47 of helix 1. Electron density is contoured at 1 σ above the mean. Solid representations of the E47/DNA complex shown in Figs. 2–7 were prepared using the program SECTOR (Evans 1990).

Interaction of E47 with DNA

bHLH proteins show distinct sequence preferences for bases at the center of the -CACXTG- E-box consensus. E47 binds tightly to -CACCTG-, a site with nonequivalent -CAC and -CAG half-sites (Fig. 1; Blackwell and Weintraub 1990). In the crystal structure, residues within the basic region, the loop, and the amino-terminal end of helix 2 contact this sequence (Figs. 8 and 9). These protein segments form an extensive DNA contact surface of 1052 Å². The basic region lies in the major groove over one half-site, and helix 2 of the same subunit 1 nearest to helix 2' of the apposing subunit. A stripe of basic residues [R350, R357, R361] is located on the other side of helix 1, next to helix 2 of the same subunit. Several intrahelical hydrogen bonds link these acidic and basic residues of helix 1. Although they do not form intersubunit salt bridges in the E47 dimer, these charged surface residues may be important for heterodimer formation, as has been proposed from mutational analyses of other bHLH proteins (Shirakata et al. 1993).
The central cytosine–guanine base pairs of the -CACCTG- site are contacted in an asymmetric manner by the E47 dimer. The E47 subunit on the -CAC half-site inserts R346 into the major groove, where it donates a hydrogen bond to N7 of guanine 1L' (Fig. 8). A hydrogen bond between Ne of R346 and DNA phosphate 1R further fixes the orientation of the R346 side chain. R346 of the protein subunit on the -CAG half-site does not make the analogous base contact, because cytosine 1R' lacks a suitable hydrogen bond acceptor. Arginine-346 of this subunit instead contacts the adjacent DNA phosphates of nucleotides 1L and 1R' (Fig. 9). The protein does not directly contact the bases of positions 1R and 1R'. DNA phosphates interact with basic and polar residues from the basic region, helix 1, and the amino-terminal end of helix 2. Similar DNA backbone contacts are made by both E47 subunits, except at the center of the binding site (Fig. 9). These are asymmetric because of the different orientations of R346 in each subunit of the E47 dimer. The register of N342 is shifted by 1 base so that it contacts phosphate 2R' instead of 1R' (the latter contact would be symmetric with the one made by N342 of the -CAC- subunit).

E47 makes a limited number of interactions with the base pairs flanking the -CACCTG- sequence. Arginine-344 of the -CAC monomer donates a hydrogen bond to N7 of adenine 5R. Arginine-344 of the -CAG subunit instead contacts the phosphate at position 5L (Fig. 9). This asymmetry may result from a slight distortion at the end of the DNA helix near base 5L to accommodate the end-to-end packing of DNA in the crystals.

Discussion

The HLH dimerization motif is a compact four-helix bundle that positions two basic region helices for interaction with DNA. Conserved hydrophobic residues at the interior of the E47 HLH form an extensive van der Waals surface and provide much of the favorable dimer interactions in the HLH (Fig. 4). Substitution of polar or charged residues at some of these interior positions eliminates the dimerization and DNA-binding activities of E47 (Voronova and Baltimore 1990) and other bHLH proteins (Winter et al. 1992; Shiratake et al. 1993). The amino terminus of the HLH is positioned over the major groove of the binding site, and here a conserved phenylalanine in helix 1 (E356) stacks across the dimer interface (Figs. 4 and 7). The aliphatic side chain of K375 lies across the ring of E356, excluding it from solvent. This lysine–phenylalanine interaction is a prominent feature of the amino-terminal end of the HLH, and these residues are conserved in the bHLH family (Fig. 1). Lysine-375 also contributes two phosphate contacts to the complex with DNA (Figs. 7 and 9).

Helices 1 and 2 of the Max bHLH–zipper include many of the same interior-facing residues as the E47 HLH (Fig. 1). The helix crossing angles in the Max HLH (Ferré-D'Amaré et al. 1993) and the E47 HLH (Figs. 3 and 4) are similar, and the positions of main chain atoms deviate by <1.0 Å (mean square difference (RMSD) = 0.7 Å).
Figure 4. The side chains of conserved hydrophobic residues face the interior of the E47 HLH. Residues a and d of the helix 2 heptad repeat pack in a side-by-side manner against residues a' and d' of the opposite subunit. Helices 2 and 2' diverge at their amino termini, creating a cleft that accepts helix 1. Helix 1 residues a and d insert into the cleft and interdigitate with residues a, d, e, and g of helices 2 and 2'.

Å² in the Max and E47 basic region, helix 1, and helix 2 segments. Max has a coiled-coil zipper segment extending from the carboxyl terminus of its HLH (Ferré-D’Amare et al. 1993). The zipper constrains the orientations of helices 2 and 2' so that helix 2 and the zipper form a continuous, smoothly bent α-helix. A similar packing arrangement is present at the carboxyl termini of helices 2 and 2' in the E47 dimer lacking a zipper.

The segment joining helices 1 and 2 of the E47 HLH is an extended strand consisting mainly of polar residues. The parallel packing arrangement of HLH helices 1 and 2 is stabilized by hydrogen bonds that connect the loop with helices 1 and 2 and thus constrain their orientations. The side chains of Q364 (helix 1), Q373 (loop), and Q381 participate in this hydrogen bond network (Fig. 6). This triad of glutamines is a unique feature of the E proteins (E47, E12, HEB, and da, Fig. 1), and it may contribute to the marked stability of E47 dimers in comparison to those of other bHLH proteins, such as MyoD (Anthony-Cahill et al. 1992). The E47 loop is short and it does not contact the DNA. However, the loops of other bHLH proteins are longer (Fig. 1) and may interact with their binding sites.

Structural differences are evident at the junction between helix 1 and the loop of E47 and Max (Fig. 5). Helix 1 of E47 is 3 residues (one turn) longer than helix 1 of Max, which terminates in proline (Fig. 1). Proline occupies this position in the HLH–zipper proteins Myc, Max, USF, TFE3, and TFE6, and the HLH proteins PHO4, tal/1(SCL, Twist, and Id (Fig. 1). It is likely that this proline defines a subclass of HLH proteins with a helix 1 similar to that of Max. A tyrosine at position g of helix 2 caps the carboxyl terminus of Max helix 1. Tyrosine-73 protrudes from the surface of Max helix 2 and packs against the proline at the carboxy-terminal end of helix 1 (Fig. 5). Valine 385, at this position of E47 helix 2 [Fig. 1], is smaller and it allows for the increased length of helix 1 in the E47 dimer. The additional turn in helix 1 of E47 permits H366 of helix 1 and E390 of helix 2' to form a salt bridge across the dimer interface (Fig. 5). The amino acid sequence of the helix 1/loop junction is conserved among the E proteins E47, E12, HEB, and Da (Fig. 1), and it is likely that the architecture of their HLH segments is similar to that of E47. The myogenic bHLH proteins MyoD, Myf4 (myogenin), and Myf6 (herculin) lack the helix 1-breaking proline of Max, but they have a helix 2 tyrosine that would interfere with the extension of helix 1 (Fig. 1). The conformation of the helix 1/loop junction
in a MyoD/DNA complex does resemble that of the Max dimer (P. Ma, H. Weintraub, and C. Pabo, pers. comm.). The E47 bHLH domain is a stable dimer in solution, based on its elution profile by gel filtration chromatography and its constant α-helicity at peptide concentrations ranging from 0.5 μM to 0.5 μM (the lower limit for reliable circular dichroism measurements; data not shown). In contrast, the MyoD bHLH dimer unfolds with an apparent equilibrium constant of ~5 μM [Anthony-Cahill et al. 1992]. Several unique features of the E47 HLH structure may contribute to the enhanced stability of the E47 dimer. The extension of helix 1 increases the buried surface of the HLH four-helix bundle and provides for a salt bridge between subunits. Salt bridges have been implicated in the dimerization strengths and preferences of bHLH proteins [Shirakata et al. 1993], bHLH–zipper proteins [Amati et al. 1993; Ferré-D’Amaré et al. 1993], and basic leucine zipper [bZIP] proteins [Schuermann et al. 1991; O’Shea et al. 1992, 1993; Hu et al. 1993; Vinson et al. 1993]. Although H366 and E390 form the only salt bridge between subunits of the E47 dimer, other charged surface residues may contribute to the stability of E47/bHLH heterodimers. Additional stability is gained from the triad of glutamines connecting the E47 loop to helices 1 and 2 through a series of hydrogen bonds.

Residues at the helix 1/loop junction have also been implicated in the unique dimerization properties of the Id protein. Id is an HLH protein that dimerizes with bHLH proteins, yet lacks a basic region and does not bind to DNA [Benezra et al. 1990]. Id does not dimerize with itself but forms very stable heterodimers with the E proteins [Sun et al. 1991; Pesce and Benezra 1993]. The dimerization specificity of Id can be conferred on E47 or E12 by swapping 4 residues from the helix 1/loop junction of Id [E47 residues Q364–L367]. These substitutions include the addition of the helix 1-breaking proline and the loss of the histidine [H366] participating in the interhelical salt bridge. One of the substituted residues, Q364, belongs to the triad of glutamines stabilizing the E47 loop. These amino acid substitutions are likely to alter the junction of helix 1 and the loop, and this effect may account for the altered dimerization properties of the mutant.

The DNA-binding sites of bHLH proteins define an E-box consensus sequence, -CAXXTG-. The conserved cytosine–adenine base step in each half of the E-box is contacted by an invariant glutamic acid (E345) in both the E47/DNA complex (Figs. 8 and 9) and the Max/DNA complex [Ferré-D’Amaré et al. 1993]. In both structures, an invariant arginine (R348 of E47) fixes the orientation of the E345 side chain by contact with it and the DNA backbone [Fig. 9]. This conserved glutamic acid/arginine
Figure 8. Stereo view of DNA contacts made by the subunit on the -CAC half-site. The side chain of E345 accepts hydrogen bonds from N4 of cytosine 3L and N6 of adenine 2L while contacting the 5-methyl group of thymine 2L'. R346 contacts phosphate 1R' and donates a hydrogen bond to N7 of guanine 1L'. N342 also stabilizes the orientation of the R346 side chain. V349 is in van der Waals contact with phosphate 1R' but does not contact a base. Similar contacts are made by the opposite subunit of the E47 dimer, except that R346 does not contact a DNA base and instead interacts with phosphates 1L and 1R'.

The guanine in the E-box consensus is contacted by the basic region residue at position a, 4 amino acids amino-terminal to E345 [Figs. 1 and 9]. Asparagine-341 donates a hydrogen bond to N7 of guanine 3L' in the E47/DNA complex [Fig. 9]. Histidine-28 of Max makes the analogous DNA contact [Ferré-D'Amaré et al. 1993].

bHLH proteins show distinct base preferences at the center of their binding sites. The bHLH-zipper proteins Myc, Max, USF, and TFE3 prefer CACGTTG sites, whereas MyoD and MyoD/E47 heterodimers prefer CAGCTG [Blackwell and Weintraub 1990; Carr and Sharp 1990; Gregor et al. 1990; Halazonetis and Kandel 1991; Alex et al. 1992, Blackwell et al. 1993, Dang et al. 1992; Van Antwerp et al. 1992]. These palindromic sites consist of -CAG or -CAC half-sites, respectively. Proteins of either specificity group are distinguished by the identity of a residue located near the junction of the basic region and helix 1 [V349 of E47] at a position 4 residues carboxy-terminal to the conserved glutamic acid (E345; Fig. 1). The arginine at this position of the Max/DNA complex donates a hydrogen bond to guanine 1' and thereby specifies a purine at this position of the binding site [Ferré-D'Amaré et al. 1993].

Figure 9. [A] Schematic diagram of the E47 subunit that contacts the -CAC half-site. [B] Summary of the DNA base and phosphate contacts in the E47/DNA complex. The 'CAC' and 'CAG' protein subunits are named according to the half-site contacted by the basic region of each monomer. Helix 2 of each subunit "crosses over" the DNA dyad and contacts the phosphodiester backbone of the opposite half-site with the amide nitrogen and side chain of K375 [see Fig. 7].
The corresponding residue of E47, V349, does not contact a base [Fig. 8] and its role in specifying base pair preferences at the center of the binding site is not readily apparent from the structure of the complex. Instead, R346 contacts guanine at the center of the E47-binding site and, therefore, may influence the DNA-binding preference of E47. Valine-349 may affect DNA-binding activity simply by allowing access to the major groove so that R346 can contact N7 of a purine. The close apposition of V349 and phosphate 1' on both halves of the binding site may also influence the orientation of the E47 dimer on DNA. A complete understanding of the mechanism of sequence discrimination at the center of the E-box awaits further mutational and structural studies.

E47 prefers the nonpalindromic sequence -CACCTG- (Blackwell and Weintraub 1990). However, the subunits of the E47 dimer are oriented symmetrically about the pseudosymmetry of this DNA site, and their basic region helices are positioned similarly in the major groove. Asymmetry is evident in the contacts made to the central base pairs of the binding site. The different orientations of the R346 side chain in each half of the E47 dimer create additional asymmetries in the DNA backbone contacts made by neighboring residues. Asparagine-342 of the subunit on the -CAC- half-site contacts phosphate of 1L', whereas N342 of the subunit on the -CAG- half-site contacts phosphate 2R'. The 1-base offset in this phosphate contact is probably attributable to the interaction of R346 with phosphate 1R' [Fig. 9].

The specification of base pairs flanking the E box is less stringent. E47 and E47/MyoD dimers discriminate against thymine at position 4L [Blackwell and Weintraub 1990, Sun and Baltimore 1991]. This discrimination may result from the potential steric clash between a thymine methyl at this position and R344, which contacts a nearby phosphate in the E47/DNA complex. Myc [Halonetzis and Kandil 1991] and PHO4 [Fisher and Goding 1992] also discriminate against a thymine at position 4 of either half-site.

E47 has a central role in promoting the DNA-binding activities of tissue-specific bHLH transcription factors. The unique features of the E47 bHLH domain may strengthen its dimerization contacts to these other bHLH proteins as well. The demonstration that residues in the helix 1/loop junction of Id are key effectors of the interaction of Id with the E proteins [Pesce and Benezra 1993] further underscores the importance of this segment of the HLH domain. Such localized structural differences among bHLH domains are likely to account for their distinct dimerization and DNA-binding properties.

Materials and methods

Expression and purification of the E47 bHLH domain

A minimal fragment of the E47 protein encompassing the bHLH region that binds the E-box DNA with high affinity was isolated by Exo III deletion mutagenesis and in vitro expression [Arnaud 1992]. This fragment, spanning E47 residues R335 to Q392 [Fig. 1], comigrates during SDS–gel electrophoresis with proteolytically stable fragments of the E47 DNA-binding domain, and it was chosen for crystallization with the pRSET [Invitrogen, Inc.] T7 expression vector in Escherichia coli BL21(DE3) pLysS [Studier and Moffatt 1986]. The protein was purified by chromatography of the crude E. coli lysate on sulfopropyl Sepharose [S Sepharose, Pharmacia, Inc.], followed by ammonium sulfate precipitation and gel filtration chromatography. Ten to 20 mg of pure protein was obtained from each liter of E. coli culture. Circular dichroism spectra of the purified peptide were collected using an AVIV 62DS spectropolarimeter. Protein samples ranging in concentration from 0.5 mM to 0.5 uM were prepared in 0.02 M ammonium acetate buffer/0.5 mM dithiothreitol and assayed at 22°C.

Crystallization of an E47/DNA complex

Crystallization of the E47 bHLH fragment was performed by vapor diffusion in the presence of stoichiometric amounts of synthetic oligonucleotides ranging from 9 to 22 bp in length. The best crystals were obtained with 11-mer DNA shown in Figure 1, from drops containing 0.5 mM protein and 0.7 mM DNA in 0.05 M NaCl, 0.05 mM CaCl2, 0.02 M MES [pH 6.5], 4% PEG 4000, and 1 mM spermine HCl. These crystals grew as clustered plates to dimensions of ~0.2 x 0.1 x 1.0 mm over the course of 7–10 days. They belong to the space group C2 and have unit cell parameters: a = 63.7 Å, b = 35.3 Å, c = 92.0 Å and β = 95.5°. The DNA stacks end to end along the a-b diagonal of the unit cell, and there is one protein/DNA complex per asymmetric unit.

X-ray data collection and structure determination

Crystals of the E47/DNA complex were flash-frozen in a stream of nitrogen after equilibration in buffer containing 0.05 M CaCl2, 0.03 M NaCl, 0.05 M MES [pH 6.5], 8% PEG 3000, 0.5 mM spermine HCl, and 30% glycerol. Diffraction data were collected on a Siemens–Xentronic multilayer wire detector mounted on an Elliott GX13 X-ray generator. Reflected X-ray intensities were integrated with the program XDS [Kabsch 1988], and the individual data sets scaled with SCXSCALE [W. Kabsch]. Subsequent data analysis were done with the CCP4 suite of programs [v.2.1.4, 1993, Daresbury Laboratory, Warrington, UK]. Heavy atom derivatives were generated by substitution of thymine with 5-iodouridine at position 7R and substitution of cytosine with 5-iodocytosine at position 5R' (see Fig. 1b for numbering of bases). Crystals of the iodinated DNA/protein were grown and harvested as described for crystals of the native complex. Native intensity data were phased with three derivatives that contained iodine substitutions at one or both of these positions [Table 1]. Introduction of iodinated bases at any of the other thymine or cytosine positions in the DNA sequence either prevented crystal growth or resulted in severe crystal lattice disorder. The iodine sites in the crystals were located by Patterson methods and refined using the program MAPHARE (CCP4 program suite, 1993).

Electron density maps calculated to 4.0 Å resolution using multiple isomorphic replacement (MIR) phase information showed electron density for most of the DNA backbone and the protein α-helices. However, the maps were of marginal quality because of the limited number of heavy atom sites and problems with the correlated refinement of heavy atom sites common to several derivatives. We therefore sought to improve the MIR phase information by a combination of density modification and restrained refinement of heavy atom parameters, using methods similar in principle to those used by Rouxd et al [1989]. A model of straight B-form DNA was positioned into density using the coordinates of the two iodine heavy atom sites as a
Table 1. X-ray data collection and refinement

<table>
<thead>
<tr>
<th></th>
<th>Native</th>
<th>I1</th>
<th>I2</th>
<th>I3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution</td>
<td>2.7</td>
<td>2.7</td>
<td>2.7</td>
<td>2.7</td>
</tr>
<tr>
<td>Reflections</td>
<td>16,520</td>
<td>11,329</td>
<td>5739</td>
<td>10,085</td>
</tr>
<tr>
<td>Unique</td>
<td>5,422</td>
<td>4,234</td>
<td>3379</td>
<td>5,152</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>97</td>
<td>75</td>
<td>68</td>
<td>91</td>
</tr>
<tr>
<td>R_{sym}</td>
<td>0.087</td>
<td>0.083</td>
<td>0.099</td>
<td>0.077</td>
</tr>
<tr>
<td>R_{inc}</td>
<td>0.145</td>
<td>0.124</td>
<td>0.112</td>
<td></td>
</tr>
<tr>
<td>R_{F_{o}-E_{c}}</td>
<td>1.1</td>
<td>0.8</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>Figure of merit</td>
<td>0.44</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Refinement statistics

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution range</td>
<td>7.0-2.8 Å</td>
</tr>
<tr>
<td>Number of reflections</td>
<td>4540</td>
</tr>
<tr>
<td>Number of nonhydrogen atoms</td>
<td>1437</td>
</tr>
<tr>
<td>Number of solvent molecules</td>
<td>24</td>
</tr>
<tr>
<td>R_{sym}</td>
<td>0.23</td>
</tr>
<tr>
<td>R_{free}</td>
<td>0.32</td>
</tr>
<tr>
<td>RMS deviation of bond lengths</td>
<td>0.016 Å</td>
</tr>
<tr>
<td>RMS deviation of bond angles</td>
<td>2.9°</td>
</tr>
</tbody>
</table>

---

*aR_{sym} = \sum_h |I_n(h) - I_{n,0}|/\sum_h I_n(h), where I_n is the mean intensity of n observations of reflection h.

*bR_{inc}, the mean isomorphous difference, = \sum |F_{o}-F_{c}|/\sum F_{o}, where F_{o} and F_{c} are the structure factor amplitudes of the native and derivative data, respectively.

"Phasing power, P_{h}/E = ([F_{c,calc}]^2/|F_{o,obs}| - |F_{c,calc}|^2)^{1/2}."

---

guide. The rotational orientation around the axis defined by these iodine sites was fixed such that the pseudo twofold axis of the DNA model was aligned with a noncrystallographic twofold axis that had been located in a self-rotation correlation search (The AMoRe programs, Navaza 1987). Next, polyalanine α-helices were positioned in the electron density corresponding to the basic region and HLH helices 1 and 2. A generous envelope was created around the partial model of the protein/DNA complex [4 Å radius around all atoms], and the quality of the electron density maps was then improved by cycles of density modification and repeated refinement of heavy atom parameters. Phases calculated from the enveloped density were modified by several cycles of solvent flattening and histogram matching using the program SQUEASH (Zhang and Main 1990; Zhang 1993). Electron density maps calculated from these modified phases showed that one of the polyalanine helices in the model was not centered in density. This helix was repositioned, and the density modification procedure was repeated with a 4 Å envelope around the new model. Phases calculated from the modified density were used for a restrained refinement of heavy atom parameters. MIR phases calculated from these new heavy atom parameters were then input for another round of solvent flattening and histogram matching. Phases converged after three cycles of density modification and heavy atom parameter refinement.

Electron density maps calculated from the final, modified MIR phases showed clear density for amino acid side chains not visible in the initial MIR maps. At this stage it was apparent that the DNA model needed adjustment. The model was rebuilt, and protein side chains were added where appropriate. The model was initially refined as a single rigid group using a constrained–restrained least-squares procedure (Sussman et al. 1977). The partially refined model was divided into progressively smaller rigid groups as rigid body refinement was extended from 7.0 to 3.5 Å resolution. A subset of reflections were reserved prior to positional refinement so that a free R-factor could be used to gauge the progress of the refinement (Brünger 1992a). Positional refinement using X-PLOR (Brünger 1992b) was followed by manual rebuilding of the model. Once the refinement had extended to data beyond 3.0 Å resolution, simulated annealing was used in conjunction with positional refinement and atomic temperature factors were applied to the model. The accuracy of the model was checked periodically with omit maps calculated after a constrained omit-simulated annealing refinement procedure (Brünger 1992b). The current model has been refined at 2.8 Å resolution, and it includes 24 solvent molecules [Table 1].

Acknowledgments

We thank Xiao-Hong Sun and David Baltimore for providing initial samples of the E47 protein and for the expression vectors used in the construction of a minimal E47 DNA-binding domain, Adrian Ferré-D’Amaré and Stephen Burley for the coordinates of the Max/DNA complex, and Phillip Ma, Hal Weintraub, and Carl Fabo for freely exchanging ideas and coordinates of the MyoD/DNA complex prior to publication. We also thank Keith Blackwell, Steven Gamblin, Ronen Marmorstein, Mark Rouéd, and members of the Harrison laboratory for stimulating discussions and suggestions. This work was supported in part by a grant from the Lucille P. Markey Charitable Trust. T.E. is a Lucille P. Markey Scholar. Coordinates of the E47/DNA complex will be deposited in the Protein Data Bank. They may also be obtained from the authors.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Alex, R., O. Sözeri, S. Meyer, and R. Dildrop. 1992. Determination of the DNA sequence recognized by the bHLH-zip do-


——. 1992b. X-PLOR version 3.1A system for X-ray crystallography and NMR. Yale University Press, New Haven, CT.


Gregor, P.D., M. Sawadogo, and R.G. Roeder. 1990. The adenovirus major late transcription factor USF is a member of the helix-loop-helix group of regulatory proteins and binds to DNA as a dimer. **Genes & Dev.** **4:** 1730–1740.


Henthorn, P., M. Kiledjian, and T. Kadesch. 1990. Two distinct transcription factors that bind the immunoglobulin enhancer μE5/κE2 motif. **Science** **247:** 467–470.


