Crystal Structure of E47−NeuroD1/Beta2 bHLH Domain−DNA Complex: Heterodimer Selectivity and DNA Recognition[†,‡]

Antonella Longo,[§,||,⊥] Gerald P. Guanga,[§,⊥] and Robert B. Rose*[,§]

Department of Molecular and Structural Biochemistry, North Carolina State University, 128 Polk Hall, Raleigh, North Carolina 27695-7622, and Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831

Received August 1, 2007; Revised Manuscript Received October 25, 2007

ABSTRACT: The ubiquitous class I basic helix−loop−helix (bHLH) factor E47 forms heterodimers with multiple tissue specific class II bHLH proteins to regulate distinct differentiation pathways. In order to define how class I− class II heterodimer partners are selected, we determined the crystal structure of the E47−NeuroD1-bHLH dimer in complex with the insulin promoter E-box sequence. Purification of the bHLH domain of E47−NeuroD1 indicates that E47 heterodimers are stable in solution. The interactions between E47 and NeuroD1 in the heterodimer are comparable to the interactions between E47 monomers in the homodimer, including hydrogen bonding, buried hydrophobic surface, and packing interactions. This is consistent with a model in which E47−NeuroD1 heterodimers are favored due to the instability of NeuroD1 homodimers. Although E47−NeuroD1 is oriented uniquely on the E-box sequence (CATCTG) within the promoter of the insulin gene, no direct contacts are observed with the central base pairs within this E-box sequence. We propose that concerted domain motions allow E47 to form specific base contacts in solution. NeuroD1 is restrained from adopting the same base contacts by an additional phosphate backbone interaction by the neurogenic-specific residue His115. Orienting E47−NeuroD1 on promoters may foster protein−protein contacts essential to initiate transcription.

The basic helix−loop−helix (bHLH[1]) domain is a dimerization and DNA-binding module identified in 125 human transcription factors (*1*). Selection of dimerization partners has important consequences for cell proliferation and differentiation (*2*). A number of bHLH classes (class I to VII) have been characterized according to tissue distribution, dimerization capabilities, and DNA-binding specificities (*2*, *3*). Class I and II factors dimerize exclusively through the HLH domain, while other classes include secondary dimerization domains, such as leucine zipper (ZIP) (*4*) and Per-Arnt-Sim (PAS) domains (*5*). The role of the ZIP secondary dimerization domain in selecting dimerization partners has been demonstrated (*6*, *7*). The basis for dimer selection by the HLH domain is less clear.

The relative stability of bHLH homo- and heterodimers allows class I and II factors to fulfill distinct functional roles. The class I factors E47, E12, ITF2, and HEB are ubiquitously expressed and form homodimers. E47 homodimers regulate B lymphocytes development (*8*). Class II factors are tissue-specific and generally require class I factors as heterodimerization partners. They include the neurogenic factors NeuroD1, -2, neurogenin1, -2, and, -3, and Math1, -2, and -3 (*9*, *10*); the myogenic factors MyoD, Myf5, myogenin, and MRF4 (*11*, *12*); the heart morphogenic factors dHand and eHand (*13*); and the pancreogenic factors Beta2 (identical to NeuroD1) and neurogenin3 (*2*, *14*)

Our work has focused on heterodimer formation between E47 and NeuroD1/Beta2 (NeuroD1 hereafter) (*15*, *16*). NeuroD1 regulates endocrine-related genes during pancreas development and insulin and glucagon in mature α- and *β*-cells (*9*). NeuroD1 also regulates neuronal differentiation (*15*, *17*) and the pituitary adrenocorticotropic hormone (ACTH) precursor proopiomelanocortin (POMC) (*18*). Mutations in the human *neuroD1* gene lead to ataxia, deafness (*19*), and a dominant inherited form of diabetes called maturity onset diabetes of the young type-6 (MODY6) (*20*, *21*). Mouse knockouts of *neuroD1* failed to develop mature islets and developed diabetes, though the nervous system appeared to develop normally (*22*). Disruption of the *e2a*

[†] Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38. This work was supported by an American Diabetes Association Junior Faculty Award 7-03-JF-34 (to R.B.R) and a Joint Faculty Agreement between UT-Battelle/Oak Ridge National Laboratory and North Carolina State University (to A.L.).

[‡] Coordinates and structure factors for the E47−NeuroD1−DNA complex were deposited in the Protein Data Bank under entry code 2QL2.

* To whom correspondence should be addressed. Phone: (919) 513-4191. Fax: (919) 515-2047. E-mail: bob_rose@ncsu.edu.

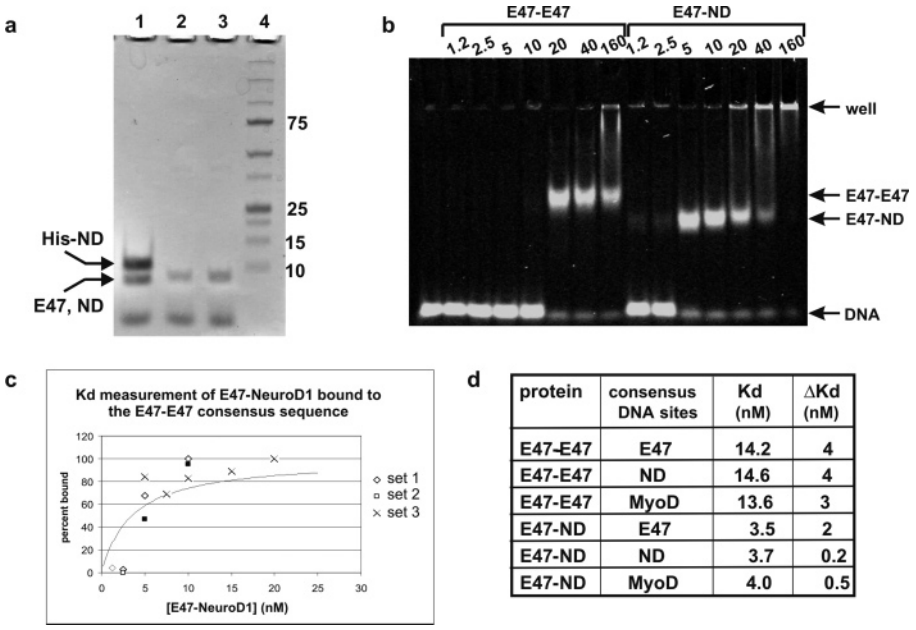
[§] North Carolina State University.

[||] Oak Ridge National Laboratory.

[⊥] These authors contributed equally to this work. R.B.R. and A.L. designed the research and wrote the paper; A.L. carried out the cloning, initial expression, and purification; G.P.G. determined conditions for heterodimerization, optimized and scaled-up purification for crystallization, obtained and optimized crystals, and collected data; A.L. indexed the data, solved the structure by MR and built the crystallographic model; G.P.G. performed the EMSA experiments.

[1] Abbreviations: BETA2, *β*-cell E-box transactivator 2; bHLH, basic helix−loop−helix; bp, base pair; bps, base pairs; b-ZIP, basic-leucine zipper; EMSA, electrophoretic mobility shift assay; MODY-6, maturity-onset diabetes of the young type-6; NeuroD1 or ND, neurogenic differentiation factor 1; PAS, Per-Arnt-Sim; Pdx1, pancreatic and duodenal homeobox-1; POMC, proopiomelanocortin; rmsd, root-mean-square deviation; VDW, van der Waals; ZIP, leucine zipper.

10.1021/bi701527r CCC: $40.75
Published on Web 12/11/2007



d

protein	consensus DNA sites	Kd (nM)	ΔKd (nM)
E47-E47	E47	14.2	4
E47-E47	ND	14.6	4
E47-E47	MyoD	13.6	3
E47-ND	E47	3.5	2
E47-ND	ND	3.7	0.2
E47-ND	MyoD	4.0	0.5

FIGURE 1: Purification and DNA binding. (a) Coomassie Blue-stained SDS−PAGE gel showing the purified E47−NeuroD1 bHLH heterodimer. Lane 1, bHLH domains of E47−NeuroD1 prior to His-tag removal from NeuroD1 (labeled ND). NeuroD1 can be distinguished from E47. Lanes 2 and 3, bHLHs of E47 and NeuroD1 after His-tag removal from NeuroD1. NeuroD1 and E47 are now undistinguishable. Lane 4, molecular mass markers with bands labeled in kDa (Biorad Precision Plus). (b) EMSA of E47−E47 and E47−NeuroD1 bound to the NeuroD1-specific DNA sequence. Both E47 and NeuroD1 aggregate in the well at higher concentrations. (c) Graph showing the final fit for E47−NeuroD1 binding to the E47−E47 consensus DNA sequence. The rmsd error of the fit to the data is ±20%. (d) Table showing the calculated K_d values and standard deviations (ΔK_d) for E47−NeuroD1 and E47−E47 bound to the E47 consensus sequence (GAA **CACCTG** GCT), the E47−NeuroD1-specific site (GGC **CATCTG** GTC), and the MyoD consensus sequence (CGA **CAGCTG** TTC).

gene (which encodes the two alternatively spliced products E47 and E12) did not affect pancreas development or insulin production, indicating that other members of class I bHLH factors can substitute for E47 (*23*, *24*).

bHLH factors bind a pseudo-palindromic DNA binding consensus sequence, CANNTG, referred to as the E-box (*25*). Binding preference for the central two bases and two flanking bases depends on the bHLH dimer, with each monomer specifically recognizing a half-site. E47−NeuroD1 binds a CATCTG E-box on the insulin and POMC promoters. The E1 box is conserved in all mammalian insulin promoters, suggesting the importance of this sequence for DNA binding specificity (*26*).

In order to characterize selective heterodimer formation and DNA binding specificity, we determined the crystal structure of the E47−NeuroD1 bHLH heterodimer in complex with the rat insulin I promoter E-box to 2.5 Å resolution. A stable E47−NeuroD1 bHLH heterodimer was formed by coexpressing the two proteins. Intermonomer interactions observed in the structure are comparable to those found in the E47 homodimer. This is consistent with the interpretation that E47−NeuroD1 heterodimers are thermodynamically favored in solution due to the instability of NeuroD1−NeuroD1 homodimers. In the crystal structure, E47−NeuroD1 binds in a unique orientation to the CATCTG binding site. Surprisingly, in the structure there are no direct contacts to the central two bases of the E-box. From the comparison with the class I E47−E47 (*27*) and the class II MyoD−MyoD (*28*) homodimer structures, we propose domain motions that in solution would allow E47 to form specific base contacts. The distribution of phosphate backbone contacts is key to orienting the heterodimer on the DNA.

EXPERIMENTAL PROCEDURES

Protein Cloning, Expression, and Purification. Vectors containing the genes for the mouse *neuroD1* and *e47* (pBAT12.mNeuroD1 and pBAT14.shPAN1, respectively) were a generous gift of the German lab (*29*). Residues 102−160 of NeuroD1 were subcloned into a pET24b vector (Promega) modified in our laboratory to encode an N-terminal hexahistidine-tag followed by a thrombin cleavage site (referred to as pET24b-6H). Residues 544−603 and 544−648 of E47 were subcloned into the pET24b-6H and the pCDF-1b (Promega) vectors. The pCDF-1b constructs lacked the His-tag. pET24b-6H−NeuroD1(102−160) and pCDF-1b−E47(544−603) (for crystallization) or pCDF-1b−E47(544−648) (for EMSA) were cotransformed into the *Escherichia coli* strain BL21(*DE3*)Star (Invitrogen). Expression was induced with 1 mM IPTG and the cells grown for 16 h at 22 °C. Cell pellets were lysed in 50 mM HEPES, 500 mM NaCl, 10% glycerol, 12.5 mM imidazole, and 10 mM β-mercaptoethanol. The heterodimer was purified with HIS-Select nickel affinity gel (Sigma) according to the manufacturer's instructions. Two bands were clearly visible on a NEXT-GEL 12.5% SDS polyacrylamide protein gel (AMRESCO), one for the 6His-tagged NeuroD1 and one for the untagged E47 (Figure 1a).

Formation of the Protein−DNA Complex for Crystallization. High-purity salt-free (HPSF) oligonucleotides for crystallization were purchased from MWG Biotech. Complementary oligos were heated to 90 °C and cooled to room temperature in a water bath. The protein was added to the annealed DNA (1:1.2 protein:DNA ratio) and dialyzed into a buffer containing 0.01 M HEPES (pH 7.6), 0.1 M NaCl, and 5% glycerol. The complex was concentrated with a centrifuge filter device (Millipore) and the 6His-tag was

a

	seq	b													H1															L		seq
		1	2	3	4	5	6	7	8	9	10	11	12	13	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	1	2	
E47 (with ND)	543	R	R	M	A	N	N	A	R	E	R	V	R	V	R	D	I	N	E	A	F	R	E	L	G	R	M	C	Q	L	H	L	K	574
NeuroD1	102	R	R	M	K	A	N	A	R	E	R	N	R	M	H	G	L	N	A	A	L	D	N	L	R	K	V	-	-	-	-	V	P	129
E47-E47	337	R	R	M	A	N	N	A	R	E	R	V	R	V	R	D	I	N	E	A	F	R	E	L	G	R	M	C	Q	M	H	L	K	368
MyoD-MyoD	110	R	R	K	A	A	T	M	R	E	R	R	R	L	S	K	V	N	E	A	F	E	T	L	K	R	S	-	-	-	-	T	S	137

	seq	L									H2																					seq
		3	4	5	6	7	8	9	10	11	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22		
E47 (with ND)	575	S	D	K	A	Q	T	-	-	-	K	L	L	I	L	Q	Q	A	V	Q	V	I	L	G	L	E	Q	Q	V	R	E	R		602
NeuroD1	130	C	Y	S	K	T	Q	K	L	S	K	I	E	T	L	R	L	A	K	N	Y	I	W	A	L	S	E	I	L	R	S	G		160
E47-E47	369	S	D	K	A	Q	T	-	-	-	K	L	L	I	L	Q	Q	A	V	Q	V	I	L	G	L	E	Q	Q	V	R	E	R		396
MyoD-MyoD	138	S	N	P	N	Q	R	L	P	-	K	V	E	I	L	R	N	A	I	R	Y	I	E	G	L	Q	A	L	L	R	D			166

b

	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	
5'-	T	A	G	G	C	**C**	**A**	**T**	**C**	**T**	**G**	G	T	C	C	T	- 3'
3'-	A	T	C	C	G	**G**	**T**	**A**	**G**	**A**	**C**	C	A	G	G	A	- 5'
	11*	10*	9*	8*	7*	6*	5*	4*	3*	2*	1*	-1*	-2*	-3*	-4*	-5*	

FIGURE 2: Numbering of protein and DNA sequences used in the text. (a) Alignment of the bHLH domains of E47, NeuroD1, and MyoD based on the secondary structure elements (shown above). For ease of comparison, residues in each sequence will be referred to as b1−b13 (basic helix), H1.1−H1.17 (helix 1), L1−L11 (loop), and H2.1−H2.22 (helix 2). The loop residues are not conserved, and are numbered consecutively in each sequence. The numbering of the sequences in the context of the full length proteins is shown in the "seq" columns. The numbering of E47 is different in the homodimer structure (*27*) and in the heterodimer with NeuroD1. Shaded residues are buried in the dimer interface. Note: H1.16 is a Leu in our structure and a Met in the E47 sequence reported in (*27*). Either Met or Leu is found in several E12/E47 isoforms in human E47, while in mouse only Leu is present. H1.16 is located on the surface of helix 1. MetH1.16 is ordered in the E47−E47 structure and LeuH1.16 is disordered in our structure. (b) Sequence of the 16-bps blunt-ended DNA fragment used for crystallization. To facilitate comparing the DNA interactions by the two monomers of the dimer, each strand is numbered starting from the initial C of the **CATCTG** E-box core (bold). As a result, bases on opposite strands that are paired have different numbers. The strand containing the CAT half-site, which binds E47, is not starred; the strand containing the CAG half-site, which binds NeuroD1, is starred.

removed by thrombin digestion overnight at room temperature. After removal of the His-tag, NeuroD1 is undistinguishable from E47 on the SDS−PAGE (Figure 1a). The complex was loaded on an S100 gel filtration column (GE Healthcare), and fractions were collected from a single peak, pooled, and concentrated to 5 mg/mL of protein for crystallization.

Crystallization. Initial crystallization conditions were identified by the sitting drop method using a PEG/Ion screen (Hampton Research). Twenty different oligonucleotide sequences of various lengths and end configuration (blunt-ended or single base overhang) were screened. The best crystals grew when a 16 base pair (bp) DNA fragment with blunt ends was used (Figure 2b). Crystals were grown in microbatch drops covered with Al's oil (Hampton Research). The final crystallization solution contained 0.1 mM imidazole (pH 7.6), 0.2 M ammonium citrate dibasic, and 22% PEG 4000. Crystals grew at 18 °C within 1 week. Crystals were harvested and flash frozen in liquid nitrogen directly from the drop.

Data Collection and Analysis. Data were collected at the SER-CAT 22-BM beam line at the Advanced Photon Source in Argonne, IL, with a MAR 225 detector. Data were indexed with MOSFLM (*30*) using the ELVES interface (*31*) and scaled with Scala (*32*, *33*). Structure factors were calculated with Truncate (*32*−*34*). The high-resolution cutoff of 2.5 Å was determined by requiring the signal-to-noise ratio ($\langle I \rangle / \langle \sigma(I) \rangle$) to be greater than 2 in the highest resolution bin. The space group is $P2_12_12$ with unit cell dimensions of $a = 60.3$ Å, $b = 169.5$ Å, and $c = 52.2$ Å and two complexes in the asymmetric unit (Table 1). The DNA stacks in the crystal forming a crystal contact, with adjacent DNA strands related through a translation. The end stacking is 5′ to 5′ instead of 5′ to 3′ because the 16 base pairs (bps) encompass 1.5 turns of DNA.

Structure Determination, Model Building, and Refinement. The structure was solved by molecular replacement using CNS (*35*). The search model was derived from the MyoD−MyoD−DNA crystallographic model (PDB accession code 1MDY) (*28*, *35*) with all nonconserved protein residues mutated to alanines and the DNA sequence unmodified. Two complexes were predicted in the asymmetric unit with 59% solvent (*31*). After the initial rotation and translation search at 15−3.5 Å, the correlation coefficient was 0.44. The second complex was located by rotation and translation searches with the position of the first solution fixed, increasing the correlation coefficient to 0.59. The NeuroD1 and E47 chains were unambiguously assigned utilizing density for unique aromatic residues. The rest of the model was built using O (*36*) and refined with CNS (*35*) to a final *R*-factor of 24.8% and R_{free} of 28.9% (Table 1). To validate the correct placement of the E47 and NeuroD1 monomers, we reduced the model to polyalanine chains with the loops and the extra turn of helix 1 of E47 deleted and refined it by simulated annealing to a maximum temperature of 2500 °C (*35*). The identity of E47 was determined unambiguously by $+3\sigma\ |F_o| - |F_c|$ density for the longer helix 1; the NeuroD1 monomer was identified by $+3\sigma\ |F_o| - |F_c|$ density for ProL2 and

Table 1: Data Collection and Refinement Statistics

Data Collection	
space group	$P2_12_12$
cell dimensions	
a, *b*, *c* (Å)	60.3, 169.5, 52.2
α, *β*, *γ* (deg)	90.0, 90.0, 90.0
resolution (Å)	56.8−2.5
R_{sym}[a] (%)	0.076 (0.77)[b]
$\langle I\rangle/\langle\sigma(I)\rangle$	13.7 (2.7)
completeness (%)	97.7 (94.8)
redundancy	7.2 (7.1)
Refinement	
resolution (Å)	56.8−2.5
no. reflections	18 783
R_{work}/R_{free} (%)	24.8/28.8
no. atoms	
protein	1756
DNA	1300
water	46
B-factors (Å^2)	
protein	52
DNA	64
water	46
rmsd	
bond lengths (Å)	0.006
bond angles (deg)	1.15
Ramachandran plot quality	
most favorite (%)	93
additional allowed (%)	6.2
disallowed (%)	0.5

[a] $R_{sym} = \sum|I - \langle I\rangle|/\sum\langle I\rangle$. [b] Values in parentheses are for the highest-resolution shell (2.64−2.50 Å).

TyrH2.11. The orientation of the DNA was determined by building the DNA sequence in both directions, refining by simulated annealing, and comparing the difference density around the central bases in the CA**TC**TG motif. In the wrong orientation, −3σ difference density was clearly visible around the central purine bases (Gua3*-Ade4*), and +3σ difference density appeared around the complementary pyrimidine bases (Thy3-Cyt4). These features were not present with the DNA built in the right orientation. The structure was validated using Procheck (*37*).

Structures were superimposed using either PyMOL (*38*) or O (*36*). Screw axes were determined in an Excel (Adobe) worksheet as described in ref *39*. Buried surface areas were calculated with CNS (*35*). Figures of structures were generated with PyMOL (*38*). van der Waals energy was calculated per residue using CHARMM (*40*). The overall van der Waals energy in the dimer interface was calculated by subtracting the van der Waals energy of each monomer from that of the dimer.

DNA Binding Assays. Electrophoretic mobility shift assays (EMSA) were performed using the purified E47(544−603)−6His-NeuroD1(102−160) heterodimer (MW = 15.8K) and the 6His-E47−6His-E47(544−648) homodimer (MW = 26.8K). The difference in molecular weight due to the extra 45 amino acids C-terminal of the E47 bHLH domain allowed us to distinguish the E47−E47 homodimer from the E47−NeuroD1 heterodimer in the EMSA. Synthetic 20 base oligonucleotides, including two to four bases upstream and downstream of the core E-boxes (CANNTG), were labeled with an infrared dye (IRDye 800 phosphoramidite) (LI-COR). The following sequences were used: (1) the E47−NeuroD1-specific E2-box sequence from the rat insulin I promoter, CGTAGGC**CATCTG**GTCCTCG (*41*); (2) the MyoD homodimer consensus sequence, CGTACGA**CAGCTG**TTCCTCG, similar to the sequence used in the MyoD−MyoD bHLH crystal structure (*25*, *28*); (3) the E47 homodimer consensus sequence, CGTAGAA**CACCTG**GCTCTCG, similar to the sequence in the E47−E47 bHLH crystal structure (*25*, *27*). Oligos were annealed with complementary unlabeled oligonucleotides (MWG), as described in the manufacturer's protocol (LI-COR). Labeled DNA (0.5 or 1 nM) was incubated with increasing concentrations (1−500 nM) of protein for 30 min on ice in EMSA binding buffer (10 mM HEPES (pH 7.2), 75 mM KCl, 2.5 mM $MgCl_2$, 0.1 mM EDTA, 10% glycerol, 2.5 mM DTT, 0.5% Tween-20). Protein samples were diluted to low salt in steps to maintain the solubility of E47−NeuroD1. Samples were separated by electrophoresis on a 6% native polyacrylamide gel using a Tris−glycine−EDTA buffer and analyzed with the Odyssey Infrared Imaging System (LI-COR Biosciences).

The apparent binding constants (K_d) were calculated by assuming association of the bHLH dimer with the DNA, with no dissociation of the dimer into monomers. Binding data were fit to a hyperbolic curve ($y = S/(K_d + S)$, y = fraction of DNA bound, S = protein concentration) using the Solver routine in Microsoft Excel (2002). The average K_d values and standard deviations were determined from three or four independent measurements, except for E47−NeuroD1 bound to the insulin promoter sequence, which was measured twice.

RESULTS AND DISCUSSION

Purification of the E47−NeuroD1 Heterodimer. Expression of the bHLH domains of E47 and NeuroD1 separately in *E. coli* resulted in soluble E47 homodimers and mostly insoluble NeuroD1 (data not shown). Refolding studies demonstrated formation of the E47−NeuroD1 heterodimer (data not shown). A soluble heterodimer was obtained by coexpressing the 6-His-tagged NeuroD1 and the untagged E47 bHLHs in *E. coli* (Figure 1a). The E47−NeuroD1 heterodimer was selected by affinity purification with Ni-NTA resin. The heterodimer was stable in solution at high salt concentrations (>300 mM NaCl) and aggregated at lower salt concentrations. In the presence of DNA, lowering the salt concentration promoted DNA binding instead of aggregation.

The stability of the E47−NeuroD1 bHLH heterodimer was unexpected, given a previous report showing the instability of E47 heterodimers with MyoD, another class II bHLH (*42*). Discrepancies between the DNA binding affinities of E47 homodimers and E47−MyoD heterodimers (in the low nanomolar range) with the measured dimerization equilibrium constants of E47−MyoD (in the micromolar range) and E47−E47 (in the nanomolar range) led to the proposal that bHLH factors bind DNA as monomers, and dimer formation is selected through DNA interactions (*42*−*46*). Given the stability of the E47−NeuroD1 heterodimer, an alternative explanation for the apparent high K_d for E47−MyoD dimerization is that the kinetics of folding of the class II monomers is slow and the K_d was measured before equilibrium was reached. The longer helix 1 of E47 might allow it to fold more quickly in solution than the class II bHLH monomers (see Figure 4). This would explain why extending helix 1 of MyoD to match that of E47 lowers the apparent dimerization constant (*42*).

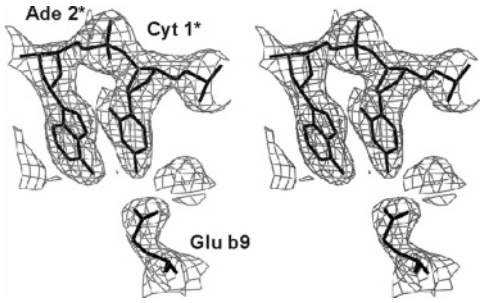


FIGURE 3: Final $2|F_o| - |F_c|$ density for Glub9 of NeuroD1 and for Cyt1*−Ade2* from the CAG half-site (wall-eyed stereo). Glub9 forms hydrogen bonds with both Cyt1* and Ade2*.

DNA Binding Specificity. We compared the DNA binding affinity of the E47−NeuroD1 heterodimer with that of the E47−E47 homodimer for three E-box sequences based on the E47−NeuroD1 insulin promoter site and the E47−E47 and MyoD−MyoD consensus sites (Figure 1b,c). The E47−NeuroD1 heterodimer bound DNA with 5-fold greater affinity than did the E47−E47 homodimer (Figure 1d). We measured no difference in binding specificity by E47−NeuroD1 or E47−E47 to the different E-box sequences.

Numbering. Throughout the paper we compare the E47−NeuroD1 heterodimer structure with the E47−E47 and MyoD−MyoD homodimer structures (*27*, *28*). In order to simplify our comparison, we adopt a universal numbering system as described in Atchley et al. (*47*). The protein sequence alignment in Figure 2a shows the corresponding numbering of the bHLH sequences in the original papers describing the structure. Both DNA strands of the CATCTG E-box are numbered starting from CA to facilitate comparing DNA contacts by each monomer; the CAG half-site is marked with stars (Figure 2b).

Overview of the Structure. The crystal structure of the E47−NeuroD1 bHLH heterodimer bound to a 16 bp blunt-ended DNA fragment was determined at 2.5 Å resolution (Table 1, Figure 3). The DNA used for crystallization encodes the E-box sequence CATCTG from the insulin promoter (*41*, *48*) (Figure 2b).

As expected, the topology of the heterodimer is identical to previously reported bHLH structures (Figure 4a,b). The HLH regions from the two monomers form a parallel, left-handed four-helix bundle, and the basic regions contact the major groove of the DNA.

The C-terminus of helix 1 is four amino acids longer in E47 than in NeuroD1 (Figure 4c). A longer helix 1 in NeuroD1 is blocked by the bulky side chain of TyrH2.11, which protrudes from helix 2 (Figure 4c) (*28*). ValH2.11 in E47 accommodates the extended helix 1 (Figure 4d) (*27*). The longer helix 1 has been observed only in the E47−E47 homodimer but not in other bHLH and bHLH-ZIP structures (*6*, *7*, *49*, *50*). The loop of E47 is disordered (residues 576−578 in one monomer and 578−579 in the other). In contrast, the loop of NeuroD1 is well ordered, except for the side chains of amino acids 133−135, and adopts a similar conformation in both complexes in the asymmetric unit, despite a crystal contact in one complex.

Dimerization Contacts. Two hydrogen bonds are formed between residues from E47 and NeuroD1 (Figure 4c). The hydrogen bond between GluH2.16 of E47 and TyrH2.11 of NeuroD1 may be a common feature of E47 heterodimers, since TyrH2.11 is conserved among all class II bHLH factors. The second hydrogen bond is a salt bridge between GluH1.9 of E47 and LysH2.9 of NeuroD1 and is less well conserved. Residue H2.9 of MyoD is an Ile, but ArgH2.10 might form the second hydrogen bond with GluH1.9 in the MyoD−E47 heterodimer. Other than these hydrogen bonds, the dimer interface consists primarily of hydrophobic interactions.

In order to explain the stability of the E47−NeuroD1 heterodimer in solution, we compared the HLH dimer interface of E47−NeuroD1 with that of E47−E47 and MyoD−MyoD homodimers (*27*, *28*). In the absence of a NeuroD1 homodimer structure, the MyoD homodimer provides an example of a class II homodimer. Like in the E47−NeuroD1 heterodimer, two hydrogen bonds are formed at the dimer interface of E47−E47 (Figure 4d). However, in the E47 homodimer the hydrogen bonds are symmetrically positioned, with GluH2.16 of helix 2 from one monomer contacting HisH1.17, the C-terminal residue of helix 1, from the other monomer (*27*). The shorter helix 1 in NeuroD1 precludes formation of this hydrogen bond in E47−NeuroD1. No hydrogen bonds are formed between monomers in the MyoD homodimer, a characteristic that may decrease the stability of class II homodimers (*28*).

In the E47−NeuroD1 structure, an area of 1870 Å^2 is buried in the hydrophobic core, similar to the 1738 Å^2 buried in the E47−E47 homodimer and greater than the 1670 Å^2 buried in the MyoD−MyoD homodimer. The larger buried surface in the E47−NeuroD1 heterodimer structure partly reflects the shorter E47 construct in the E47−E47 homodimer, which is missing residues H2.19−H2.22. Buried residues are similar in the three structures (Figure 2a). Using 25 cal/(mol Å^2) as a conservative estimate of the free energy derived by sequestering hydrophobic surface from water, the stability of the MyoD homodimer is about 5 kcal/mol less than that of the E47 dimers (*51*).

In addition to the hydrophobic effect, a significant contribution to protein stability results from packing interactions, as demonstrated by cavities generated through mutations of internal residues (*51*). In order to compare packing interactions of hetero- and homodimers, we calculated van der Waals interaction energies for the E47−NeuroD1 (74 kcal/mol), E47−E47 (82 kcal/mol), and MyoD−MyoD (64 kcal/mol) dimers (Table 2). The per residue van der Waals energies indicate that 14 residues contribute most of the packing energy of E47 in the E47 homodimer and E47−NeuroD1 heterodimer (Table 2, Figure 5). Most of these residues pack as well with NeuroD1 as with E47 (Table 2, columns 3 and 4).

In contrast, the interface of NeuroD1 in the E47−NeuroD1 heterodimer indicates many small differences in packing energy distributed over the whole surface (Table 2, column 6). Seven interface residues are conserved between NeuroD1 and E47 (AH1.6, LH1.10, LH2.5, KH2.1, AH2.6, IH2.12, and LH2.15). Loss of packing interactions due to the shorter helix 1 of NeuroD1 is compensated by van der Waals contact by TyrH2.11. ArgH2.6 and ValH1.13 of NeuroD1 (GlnH2.6 and MetH1.13 of E47) contribute favorable packing energies. Two smaller residues, LeuH1.7 and SerH2.16, pack less well than the larger E47 side chains (PheH1.7 and GluH2.16).

The differences between the NeuroD1 and E47 surfaces may explain the decreased stability of NeuroD1 homodimers. In particular, the smaller LeuH1.7 residues would generate

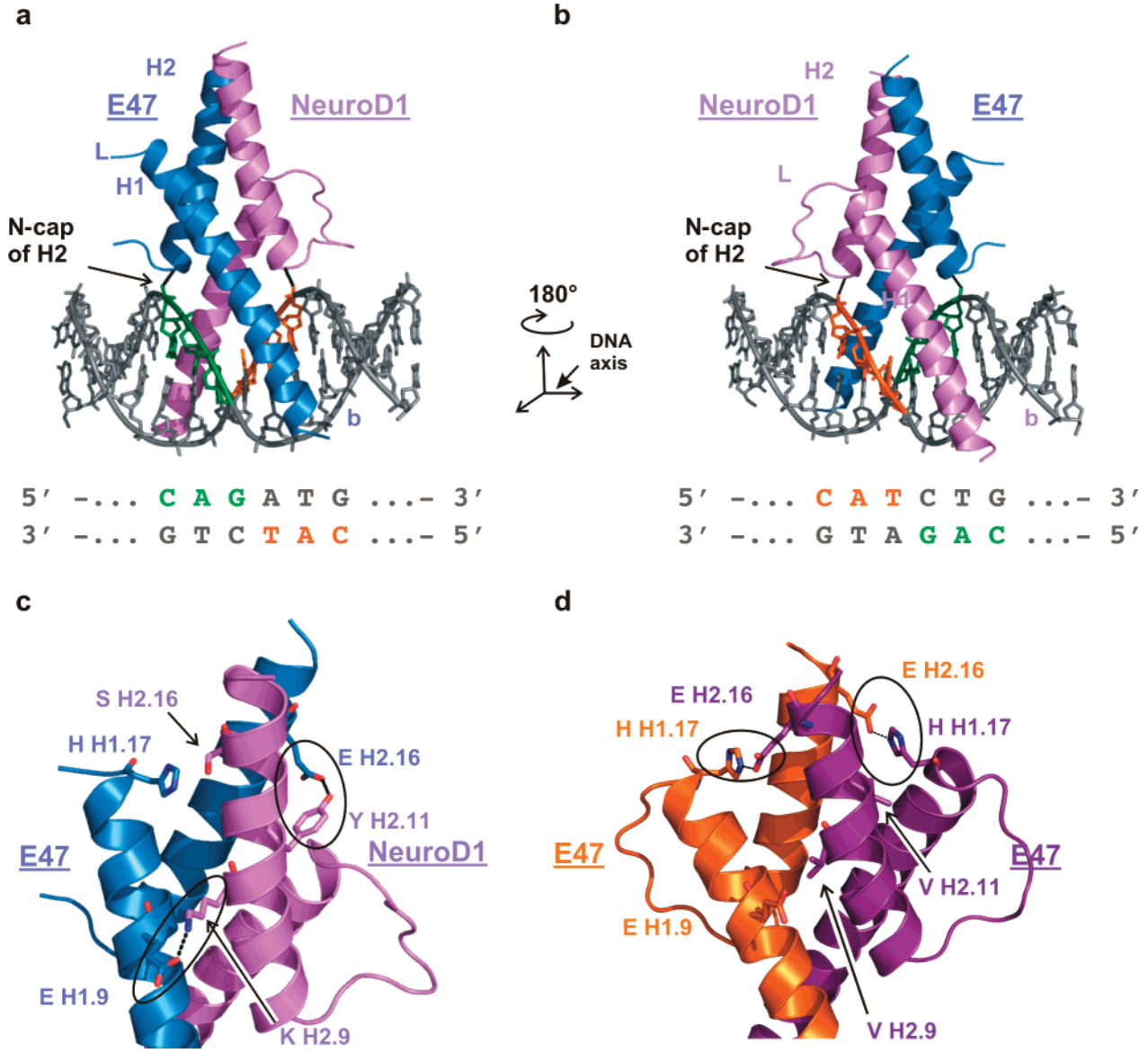


FIGURE 4: Overview of the E47−NeuroD1 heterodimer structure. (a and b) Ribbon drawing of the E47 (blue) and NeuroD1 (violet) bHLH domains bound to DNA. In panel b the structure is rotated 180° around the DNA 2-fold axis with respect to (a). E47 interacts through Glub9 with the CAT half-site (orange); NeuroD1 interacts through the same residue with the CAG half-site (green). In both half-sites, the amino terminal dipole of helix 2 is capped by the phosphate oxygen of Cys1/Cys1*. The basic (b), helix 1 (H1), loop (L), and helix 2 (H2) regions are labeled for E47 in panel a and for NeuroD1 in panel b. H1 is four amino acids shorter in NeuroD1. (c) Two intermonomer hydrogen bonds (circled) are asymmetrically positioned in the E47−NeuroD1 structure: between LysH2.9 of NeuroD1 and GluH1.9 of E47, and between TyrH2.11 of NeuroD1 and GluH2.16 of E47. TyrH2.11 of NeuroD1 also blocks H1 of NeuroD1 from extending one extra turn as found in E47. (d) Two intermonomer hydrogen bonds (circled) are symmetrically positioned in the E47−E47 homodimer: GluH2.16 contacts HisH1.17 in the extended H1 of E47. These hydrogen bonds are not present in the heterodimer because H2.16 is a Ser in NeuroD1 and H1 of NeuroD1 is shorter than H1 of E47, excluding residue HisH1.17. The homodimer cannot form the LysH2.9−GluH1.9 hydrogen bond because ValH2.9 in E47 excludes its formation.

a cavity across the dimer interface in the NeuroD1−NeuroD1 protein core, instead of packing with PheH1.7 of E47. The packing defect is not accounted for in the hydrophobic effect based on buried surface area, since LeuH1.7 would be buried in the NeuroD1 homodimer and E47−NeuroD1 heterodimer. Packing of SerH2.16 (GluH2.16 in E47) and TyrH2.11 with the shorter helix 1 of NeuroD1 may also result in a cavity, given the low van der Waals contact. In the MyoD homodimer structure, decreased packing can be attributed to smaller side chains in MyoD, for example, ValH1.3 (Ile in E47), ThrH1.9 (Glu in E47), SerH1.13 (Met in E47), and ValH2.2 (Leu in E47), and loss of contact by TyrH2.11. The helices in the MyoD homodimer pack at slightly different angles than in the E47 complexes (the crossing angle between the H2 helices differs by 5°), increasing the distance between residues at the C-terminus of helix 2 by as much as 1 Å, which might also reduce the packing interactions.

Overall, the hydrogen bonding and hydrophobic and packing interactions of the dimer interfaces favor formation of E47−NeuroD1 and E47−E47 dimers relative to NeuroD1 (or MyoD) homodimers. These interactions are consistent with a model in which heterodimerization is favored primarily because class II homodimers are disfavored. This mechanism was described for selection of the basic leucine zipper (b-ZIP) heterodimers between c-Fos and c-Jun in which the c-Fos homodimer is unstable (*52*).

Interaction with DNA. E47−NeuroD1 is oriented on the DNA with E47 binding the CAT half-site and NeuroD1

Table 2: Packing Energies between Monomers of bHLH Homo- and Heterodimers[a]

	E47	$\langle E_{VDW}\rangle$ in E47−E47 (kcal/mol)	E_{VDW} for E47 in E47−ND (kcal/mol)	ND	E_{VDW} for ND in E47−ND (kcal/mol)	MyoD	$\langle E_{VDW}\rangle$ in MyoD−MyoD (kcal/mol)
H1.2	Asp	−1.7	−2.0	Gly	−0.6	Lys	−1.8
H1.3	**Ile**	**−5.4**	**−5.8**	**Leu**	**−5.0**	Val	−4.0
H1.6	**Ala**	**−1.7**	**−3.3**	**Ala**	**−3.2**	Ala	−1.7
H1.7	**Phe**	**−5.2**	**−5.0**	Leu	−2.6	Phe	−4.1
H1.9	**Glu**	**−2.9**	**−4.2**	Asn	−1.7	Thr	−1.7
H1.10	**Leu**	**−6.0**	**−7.0**	**Leu**	**−5.3**	Leu	−5.6
H1.13	**Met**	**−3.7**	**+4.4**	**Val**	**−6.2**	Ser	−1.6
H1.17	**His**	**−3.2**	**−2.6**	–[b]	–[b]	–[b]	–[b]
H2.1	Lys	−2.5	−2.0	**Lys**	**−4.2**	Lys	−1.6
H2.2	**Leu**	**−4.1**	**−3.0**	**Ile**	**−4.3**	Val	−4.0
H2.5	**Leu**	**−8.6**	**−7.3**	**Leu**	**−8.6**	Leu	−5.4
H2.6	**Gln**	**−1.0**	**−1.3**	**Arg**	**−3.4**	Arg	−2.0
H2.8	**Ala**	**−2.6**	**−2.7**	**Ala**	**−2.8**	Ala	−1.0
H2.9	**Val**	**−5.9**	**−5.5**	Lys	+0.7	Ile	−5.6
H2.11	Val	−1.7	−1.9	**Tyr**	**−4.5**	Tyr	−3.1
H2.12	**Ile**	**−7.9**	**−7.2**	**Ile**	**−7.4**	Ile	−5.8
H2.13	Leu	−0.2	− 1.5	Trp	−1.3	Glu	−0.2
H2.15	**Leu**	**−5.6**	**−5.9**	**Leu**	**−6.5**	Leu	−3.9
H2.16	**Glu**	**−5.3**	**−4.9**	Ser	−1.7	Gln	−2.0
H2.18	Gln	–[b]	−1.1	Ile	−1.5	Leu	−0.8
H2.19	Val	–[b]	+0.3	Leu	+0.6	Leu	−3.1
partial totals		−75.2	−69.5		−69.5		−59.0
overall totals		−82.0	−74.0		−74.0		−64.0

[a] The calculated van der Waals (VDW) energies are shown for the residues contributing most to packing: $\langle E_{VDW}\rangle$, the average energy of both E47 monomers in the E47−E47 homodimer (column 3); E_{VDW}, the energy of E47 and NeuroD1 (abbreviated ND) in the E47−NeuroD1 heterodimer (column 4 and 6, respectively); $\langle E_{VDW}\rangle$, the average energy of both MyoD monomers in the MyoD−MyoD homodimer (column 8). The identity of the amino acids for each monomer is listed. Entries that contribute the most VDW energy in each dimer are in bold and are illustrated in Figure 5. The sum of each column (listed as the partial sum) is less than the total VDW contact for each dimer (listed as the overall totals) due to small contributions (arbitrarily set at 2 kcal/mol or less) from other residues that are not listed. [b] Not present in the structure because it is disordered or because helix 1 is shorter.

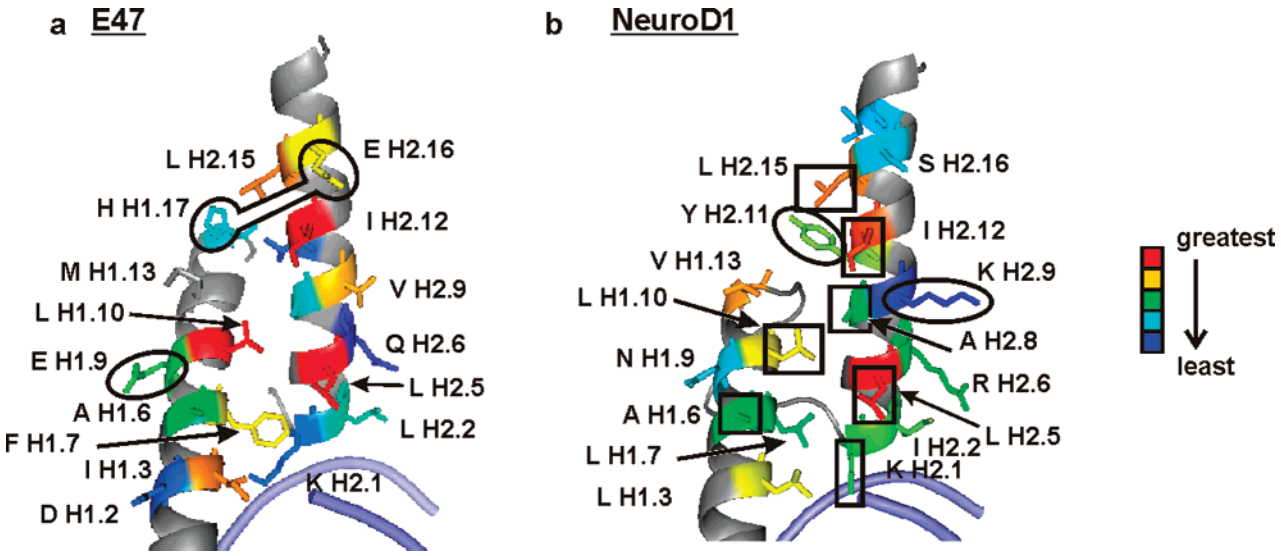


FIGURE 5: Packing interactions favor heterodimer complexes. (a and b) Residues of E47 in panel a and NeuroD1 in panel b that contribute most to packing in the E47−NeuroD1 interface are drawn as stick figures and colored by the per residue VDW contact (greatest in red to least in blue) (see Table 2 and color scale in this figure). Circled residues contribute to hydrogen bonds in the E4−E47 homodimer or E47−NeuroD1 heterodimer. (a) Most residues of E47 pack with similar energies in the E47−NeuroD1 heterodimer and E47−E47 homodimer, except MetH1.13, which clashes with LysH2.9 of NeuroD1. (b) Seven interface residues are identical in E47 and NeuroD1 (squared). The smaller SerH2.16 and LeuH1.7 side chains of NeuroD1 pack less well than the larger GluH2.16 and PheH1.7 side chains of E47. We propose that cavities formed in the NeuroD1−NeuroD1 interface by these residues will destabilize the homodimer.

binding the CAG half-site (Figure 6a,b). The universally conserved Glub9 of E47 and NeuroD1 form the only direct hydrogen bonds with the DNA bases in this structure. Glub9 of NeuroD1 contacts both Cyt1* and Ade2*. Glub9 of E47 contacts the DNA more loosely: in one of the two complexes in the asymmetric unit, it only contacts Cyt1 and is too far (4.1 Å) from Ade2 to form a hydrogen bond; in the other complex, the side chain of Glub9 is disordered. The position of Glub9 is stabilized by a salt bridge with Argb12; this interaction is strictly conserved in all bHLH structures. In both monomers, Glub9 and Asnb6 form van der Waals contacts with the base of Thy5/5*, and Argb2 contacts Cyt1/1* and Cyt−1/−1*. There are no direct contacts with the central bases of the E-box sequence.

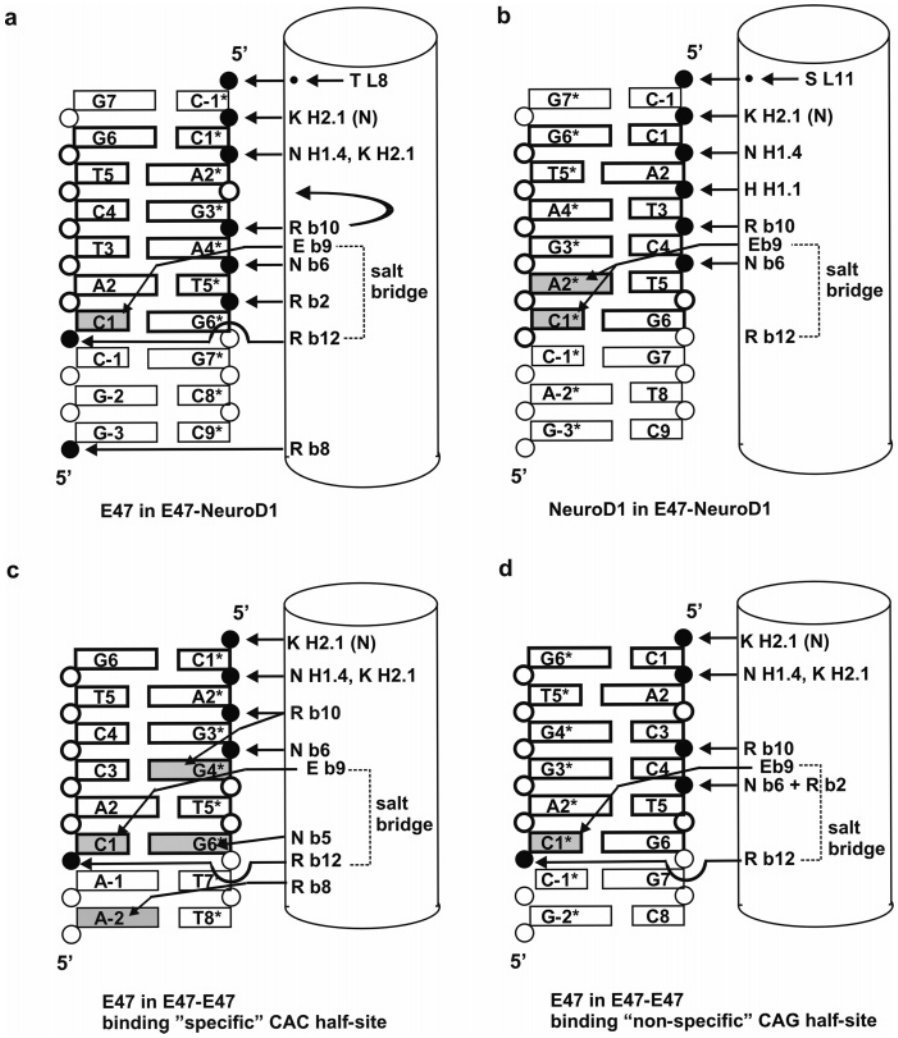


FIGURE 6: Comparison of protein−DNA contacts in the E47−NeuroD1 heterodimer and the E47−E47 homodimer. (a and b) Protein−DNA contacts for E47 (a) and NeuroD1 (b) in the E47−NeuroD1 heterodimer structure. Direct protein−DNA interactions as well as indirect interactions mediated through water molecules (dots) are shown. The sugar−phosphate backbone is represented as circles and bases as rectangles. The CATCTG E-box core is outlined in bold. Hydrogen bonds (within 3.5 Å) are represented as arrows to the backbone (black circles) or to the bases (shaded rectangles). Glub9 forms the only direct base contacts in the structure. HisH1.1 is unique to NeuroD1 and other neurogenic factors and contacts the phosphate of Thy3; E47 does not contact the corresponding phosphate in this structure. (c and d) Protein−DNA contacts in the E47−E47 homodimer structure. The E47 monomer contacting the CAC half-site (c) forms specific hydrogen bonds with Ade−2, Gua4*, and Gua6*. The other E47 monomer contacts the CAG half-site (d) only through Cys1*, as in E47−NeuroD1. In the specific configuration (c), a shift in the register of the phosphate backbone contacts allows Argb10 to contact the central purine base, Ade−2. We propose that E47 in the E47−NeuroD1 heterodimer can switch to contact this phosphate (a, curved arrow), as it does in the "specific" E47 conformation. To make this switch, E47 would lose a backbone contact by Argb2. Argb8 would shift from a backbone to a base contact, and two additional base contacts would be formed by Asnb5 and Argb10.

The remaining protein−DNA interactions are with the phosphate backbone. E47 forms a total of nine contacts with eight phosphate groups: five by residues from the basic helix (Argb2, Asnb6, Argb8, Argb10, and Argb12), one by a residue from helix 1 (AsnH1.4), two from helix 2 (LysH2.1), and a water-mediated contact by ThrL8 (Figure 6a). NeuroD1 forms six phosphate contacts: two by residues from the basic helix (Asnb6, Argb10), two from helix 1 (HisH1.1, AsnH1.4), one from the backbone nitrogen of LysH2.1, and a water-mediated contact by SerL11 (Figure 6b). HisH1.1 is conserved in all neurogenic bHLHs and its contact with the third phosphate group in the CANNTG sequence may be unique to this family.

In our structure, as well as in the E47−E47 and MyoD−MyoD structures, the phosphate of Cyt1/1* caps the N-terminus of helix 2 through the nitrogen of LysH2.1 (Figure 4a). The rotamer of LysH2.1 differs between E47 and NeuroD1 (Figure 6a,b). In E47, the aliphatic carbons of LysH2.1 pack against PheH1.7, orienting the Lys side chain to contact the phosphate backbone of the DNA. The position of LysH2.1 is further stabilized by a salt bridge with AsnH1.4 of E47. In NeuroD1, the H1.7 residue is a Leu. As a result, the rotamer of the LysH2.1 side chain is in the all-trans conformation and contacts two waters.

Global Positioning of the bHLH on the DNA. Blackwell and Weintraub proposed that preferential DNA binding by bHLH monomers is independent of the dimerization partner (*25*). They identified the consensus half-site sequence for E47 as CAC. The E47−E47 structure was determined with the homodimer bound to the E-box CACCTG, containing one consensus half-site (*27*). The two E47 monomers contact the two half-sites differently: the E47 monomer contacting the CAC half-site forms a specific contact with the central Gua4* base on the opposite strand and is referred to as being

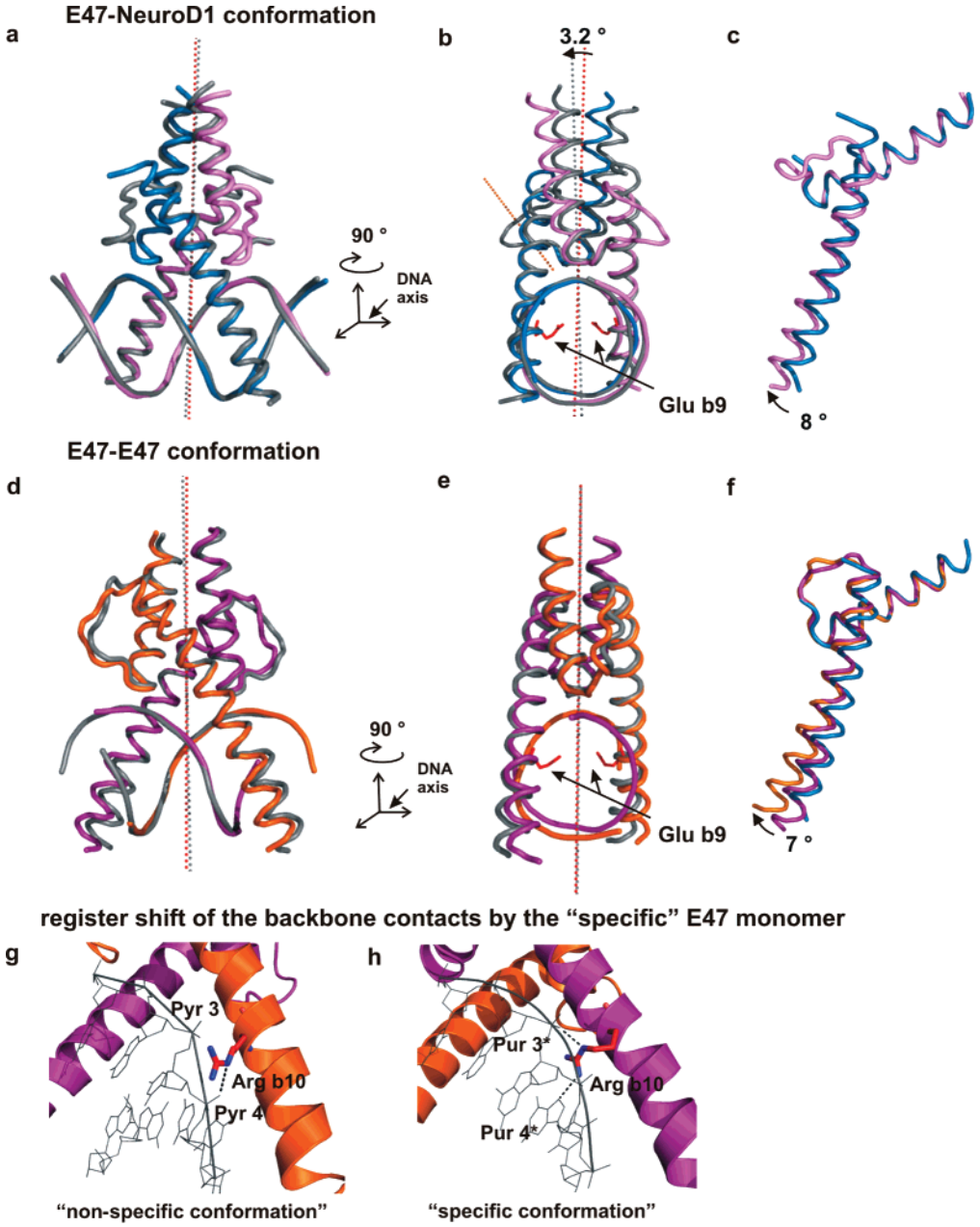


FIGURE 7: Subdomain motions of bHLH domains bound to DNA. (a−c) The E47−NeuroD1 dimer binds DNA with the HLH subdomain "tilted" and the H1−H2 hinge angle positioning the basic helices in the same "nonspecific" orientation. (a) The E47−NeuroD1 dimer (blue-violet) binds DNA asymmetrically. The DNA 2-fold axis (line of red dots) was defined by superimposing the CANNTG E-box sequence of the two DNA strands (rmsd 0.6 Å for 12 sugar C1′ atoms). The HLH of the rotated dimer (gray ribbon) does not overlap with the original dimer. The view is perpendicular to the DNA helical axis (not shown). (b) The HLH domain is tilted 3.2° relative to the DNA. The HLH 2-fold axis (line of black dots) was defined by superimposing helix 2 of both monomers (rmsd 0.3 Å for 34 Cα atoms from H2.1 to H2.17 of each monomer). Glub9 side chains, which contact the CA bases of the E-box, are shown. The view is down the DNA helical axis, rotated 90° from (a) around the DNA 2-fold axis (see coordinate system). (c) Different hinge angles between H1 and H2 of the E47 and NeuroD1 monomers. The hinge angle was calculated by first superimposing the H2 helices and determining the angle of rotation required to superimpose the basic and H1 helices (rmsd 0.3 Å for 46 Cα atoms from b3 to H1.11 of E47 and NeuroD1). The hinge angle of E47 (blue) is 8° wider than the hinge angle of NeuroD1 (violet). The axis of rotation of the hinge angle traverses the loop (orange axis in panel b). (d-f) The E47−E47 dimer binds DNA with the HLH subdomain "vertical" and the H1−H2 hinge angle positioning one monomer in the "specific" orientation and one in the "nonspecific" orientation. (d) The E47−E47 dimer (purple-orange) binds DNA with the HLH subdomain positioned almost symmetrically but with the basic helices rotated. The DNA 2-fold axis (line of red dots, rmsd 0.7 Å, defined as in panel a) is almost coincident with the HLH 2-fold axis (line of black dots, rmsd 0.26 Å, defined as in panel b). The HLH of the rotated dimer (gray ribbon) overlaps with the unrotated dimer, but the basic helices are oriented differently. The view is perpendicular to the DNA helical axis. (e) Same as in panel d, but the view is down the DNA helical axis, rotated 90° from panel a around the DNA 2-fold axis. (f) The hinge angle of the "specific" E47 monomer (orange) is 7° wider than the hinge angle of the "nonspecific" E47 monomer (purple) (rmsd 6 Å for 46 Cα atoms from b3 to H1.11 0.). (g) Close-up view showing the register shift by Argb10 in the "nonspecific" (left) and "specific" (right) conformations. In the "nonspecific" conformation, E47 binds the C_1*A_2*G_3* half-site, with a pyrimidine in position 4 opposite Gua3*; Argb10 contacts the phosphate at position 4; there are no protein contacts with the phosphate of Cyt3. In the "specific" conformation, E47 binds the $C_1A_2C_3$ half-site, with a purine in the 4* position opposite Cyt3. Argb10 shifts phosphate backbone contact to the 3* position, and contacts the Gua3* base directly.

in a "specific" conformation (Figure 6c); the E47 monomer contacts the C*A*G* half-site only through the base of Cyt1* and is in a "nonspecific" conformation (Figure 6d) (*27*). In the E47−NeuroD1 structure, E47 binds the CAT half-site, which encodes a pyrimidine in the third position of the half-site as in the CAC consensus sequence, but E47 is in the "nonspecific" conformation without any contact with the central base. In order to determine whether E47 in E47−NeuroD1 could recognize the CAT half-site similarly to the CAC half-site in E47−E47, we investigated differences between the "specific" and "nonspecific" conformations of the bHLH monomers.

We defined the global position of E47−NeuroD1 on the DNA by the angle of rotation of the 2-fold axis of the H2−H2 helices relative to the 2-fold axis of the E-box (Figure 7a,b). The H2−H2 axis of the dimer is tilted 3.2° from the DNA 2-fold axis. The hinge angle between H1 and H2 is 8° greater in E47 than in NeuroD1 (Figure 7c). Despite these conformational differences, the basic helices of E47 and NeuroD1 interact with the phosphate backbone of the DNA in the same register: Asnb6 and Argb10 of E47−NeuroD1 hydrogen bond with the phosphate oxygens of Thy5*/Thy5 and Ade4*/Cyt4, respectively (Figure 6a,b). HisH1.1 in NeuroD1 forms an additional phosphate backbone contact with Thy3, while E47 does not contact the corresponding Gua3*. In this register, the only direct base contacts are through Glub9 to the CA bases.

In the E47−E47 structure, the 2-fold axis of the H2−H2 helices is almost coincident with the DNA 2-fold axis (Figure 7d,e), but the hinge angle between H1 and H2 differs by 7° in the two monomers (Figure 7f). As a result, the basic helix of the "specific" E47 monomer shifts register in contacting the phosphate backbone of the DNA relative to the register in the E47−NeuroD1 heterodimer and the "nonspecific" E47 monomer (Figure 7 g,h). As noted in Ellenberger et al. (*27*), Asnb6 and Argb10 of the "specific" E47 monomer contact the backbone of Gua4* and Gua3*, respectively, instead of Thy5* and Cyt4* (Figure 6c,d). In this position, the basic helix of E47 forms three additional base contacts: Argb10 with the central base Gua4*, Asnb5 with Gua6*, and Argb8 with Ade−2.

Why does not E47 adopt the "specific" conformation in the E47−NeuroD1 structure? We propose that E47 in E47−NeuroD1 is prevented from adopting "specific" base contacts by crystal packing of the HLH domain. In solution, the HLH subdomain would be free to pivot around the phosphate backbone contacts with Cyt1 and Cyt1*, which cap the N-termini of both H2 helices. E47 could then shift from the "nonspecific" to the "specific" conformation, allowing Argb10 to hydrogen bond with N7 of Ade4*. In contrast, the position of the basic helix of NeuroD1 appears to be less flexible on the DNA due to the contact by HisH1.1 with the phosphate of Thy3. The position of NeuroD1 may also be stabilized by the closer contact by Glub9 with the base of Ade2* than in E47. Finally, the hydrogen-bond between H2 of NeuroD1 and H1 of E47 suggests that the basic helix of E47 is more closely associated with the position of the HLH domain than is the basic helix of NeuroD1.

It should be noted that E47−NeuroD1 binds the palindromic MyoD consensus site CAGCTG with similar affinity to the E47 consensus sequence in the absence of a CA(T/C) half-site (Figure 1d). Perhaps other interactions, for example Argb8 with Ade−2, increase the affinity with this sequence. Also, MyoD does not contain a residue comparable to HisH1.1 of NeuroD1 to orient an E47−MyoD heterodimer. Other residues in MyoD may determine the binding orientation of E47−MyoD.

The bZIP factors c-Fos and c-Jun demonstrate a similar relationship between the angle of the dimerization domain and the DNA contacts accessible to the basic helices (*53*). Protein−protein contacts between the zipper domain and other transcription factors can influence the angle of the basic helix of c-Jun and the resulting DNA interactions (*54*, *55*). Similarly orienting the E47−NeuroD1 heterodimer may be important for recruiting transcription factors or coactivators and therefore promoting formation of an active preinitiation complex. Protein−protein contacts between the HLH domain of E47−NeuroD1 and Pdx1 (*29*) or MafA (*56*, *57*) in pancreatic β-cells or Pitx1 and Tpit in pituitary corticotrophs (*18*, *58*) might alter the DNA contacts accessible to the basic helices.

CONCLUSIONS

The current study supports the notion that stable class I−class II heterodimers form in solution in the absence of DNA. E47 homo- and heterodimers are stabilized by two hydrogen bonds, burial of hydrophobic residues, and favorable packing interactions. Heterodimerization is favored primarily due to the relative instability of class II homodimers.

We propose that the E47−NeuroD1 heterodimer is oriented on the DNA through domain motions that allow specific DNA contacts by E47. According to this model, E47 preferentially binds CA(C/T) half-sites, allowing Argb10 to contact N7 of the purine base opposite the pyrimidine in position 3. NeuroD1 is prevented from forming this DNA contact by HisH1.1, which blocks the shift in phosphate backbone binding that would allow Argb10 to bind the DNA in the "specific" conformation. NeuroD1 does not contact the third base of its half-site, but the CA(G/A) half-site would disfavor binding of E47. Orienting the E47−NeuroD1 heterodimer may be important for recruiting other transcription factors or coactivators, promoting formation of an active preinitiation complex.

NMR studies have shown that the basic helices of bHLH domains fold in response to binding CANNTG E-box sequences (*59*). Further studies of bHLH mutations and complexes with interacting transcription factors are required to test how dynamic the bHLH domain is once it binds to the E-box in search of the minimal binding configuration.

ACKNOWLEDGMENT

We thank Michael German for the E47 and NeuroD1 clones, and Tom Ellenberger for providing the coordinates of the E47−E47 homodimer structure. We thank Helen Rho for cloning, expressing, and purifying the 6His-E47(544−648) homodimer for EMSA studies. We thank Carla Mattos for the use of CHARMM and Paul Swartz for assisting in calculating van der Waals energies. Data were collected at Southeast Regional Collaborative Access Team (SER-CAT) 22-BM beamline at the Advanced Photon Source, Argonne National Laboratory. Supporting institutions may be found at www.ser-cat.org/members.html.

REFERENCES

1. Ledent, V., Paquet, O., and Vervoort, M. (2002) Phylogenetic analysis of the human basic helix−loop−helix proteins, *Genome Biol. 3*, RESEARCH0030.
2. Massari, M. E., and Murre, C. (2000) Helix−loop−helix proteins: Regulators of transcription in eucaryotic organisms, *Mol. Cell. Biol. 20*, 429−440.
3. Murre, C., Bain, G., van Dijk, M. A., Engel, I., Furnari, B. A., Massari, M. E., Matthews, J. R., Quong, M. W., Rivera, R. R., and Stuiver, M. H. (1994) Structure and function of helix−loop−helix proteins, *Biochim. Biophys. Acta 1218*, 129−135.
4. Luscher, B., and Larsson, L. G. (1999) The basic region/helix−loop−helix/leucine zipper domain of Myc proto-oncoproteins: Function and regulation, *Oncogene 18*, 2955−2966.
5. Crews, S. T. (1998) Control of cell lineage-specific development and transcription by bHLH-PAS proteins, *Genes Dev. 12*, 607−620.
6. Nair, S. K., and Burley, S. K. (2003) X-ray structures of Myc-Max and Mad-Max recognizing DNA. Molecular bases of regulation by proto-oncogenic transcription factors, *Cell 112*, 193−205.
7. Ferre-D'Amare, A. R., Prendergast, G. C., Ziff, E. B., and Burley, S. K. (1993) Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain, *Nature 363*, 38−45.
8. Shen, C. P., and Kadesch, T. (1995) B-cell-specific DNA binding by an E47 homodimer, *Mol. Cell. Biol. 15*, 4518−4524.
9. Chu, K., Nemoz-Gaillard, E., and Tsai, M. J. (2001) BETA2 and pancreatic islet development, *Recent Prog. Horm. Res. 56*, 23−46.
10. Kageyama, R., Ishibashi, M., Takebayashi, K., and Tomita, K. (1997) bHLH transcription factors and mammalian neuronal differentiation, *Int. J. Biochem. Cell Biol. 29*, 1389−1399.
11. Pownall, M. E., Gustafsson, M. K., and Emerson, C. P., Jr. (2002) Myogenic regulatory factors and the specification of muscle progenitors in vertebrate embryos, *Annu. Rev. Cell Dev. Biol. 18*, 747−783.
12. Berkes, C. A., and Tapscott, S. J. (2005) MyoD and the transcriptional control of myogenesis, *Semin. Cell Dev. Biol. 16*, 585−595.
13. Srivastava, D., Cserjesi, P., and Olson, E. N. (1995) A subclass of bHLH proteins required for cardiac morphogenesis, *Science 270*, 1995−1999.
14. Wilson, M. E., Scheel, D., and German, M. S. (2003) Gene expression cascades in pancreatic development, *Mech. Dev. 120*, 65−80.
15. Lee, J. E., Hollenberg, S. M., Snider, L., Turner, D. L., Lipnick, N., and Weintraub, H. (1995) Conversion of Xenopus ectoderm into neurons by NeuroD, a basic helix−loop−helix protein, *Science 268*, 836−844.
16. Naya, F. J., Stellrecht, C. M., and Tsai, M. J. (1995) Tissue-specific regulation of the insulin gene by a novel basic helix−loop−helix transcription factor, *Genes Dev. 9*, 1009−1019.
17. Lee, J. E. (1997) Basic helix−loop−helix genes in neural development, *Curr. Opin. Neurobiol. 7*, 13−20.
18. Poulin, G., Turgeon, B., and Drouin, J. (1997) NeuroD1/beta2 contributes to cell-specific transcription of the proopiomelanocortin gene, *Mol. Cell. Biol. 17*, 6673−6682.
19. Chae, J. H., Stein, G. H., and Lee, J. E. (2004) NeuroD: The predicted and the surprising, *Mol. Cells 18*, 271−288.
20. Malecki, M. T., Jhala, U. S., Antonellis, A., Fields, L., Doria, A., Orban, T., Saad, M., Warram, J. H., Montminy, M., and Krolewski, A. S. (1999) Mutations in NEUROD1 are associated with the development of type 2 diabetes mellitus, *Nat. Genet. 23*, 323−328.
21. Liu, L., Furuta, H., Minami, A., Zheng, T., Jia, W., Nanjo, K., and Xiang, K. (2007) A novel mutation, Ser159Pro in the NeuroD1/BETA2 gene contributes to the development of diabetes in a Chinese potential MODY family, *Mol. Cell. Biochem.*
22. Naya, F. J., Huang, H. P., Qiu, Y., Mutoh, H., DeMayo, F. J., Leiter, A. B., and Tsai, M. J. (1997) Diabetes, defective pancreatic morphogenesis, and abnormal enteroendocrine differentiation in BETA2/neuroD-deficient mice, *Genes Dev. 11*, 2323−2334.
23. Itkin-Ansari, P., Bain, G., Beattie, G. M., Murre, C., Hayek, A., and Levine, F. (1996) E2A gene products are not required for insulin gene expression, *Endocrinology 137*, 3540−3543.
24. Sharma, A., Henderson, E., Gamer, L., Zhuang, Y., and Stein, R. (1997) Analysis of the role of E2A-encoded proteins in insulin gene transcription, *Mol. Endocrinol. 11*, 1608−1617.
25. Blackwell, T. K., and Weintraub, H. (1990) Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection, *Science 250*, 1104−1110.
26. Hay, C. W., and Docherty, K. (2006) Comparative analysis of insulin gene promoters: Implications for diabetes research, *Diabetes 55*, 3201−3213.
27. Ellenberger, T., Fass, D., Arnaud, M., and Harrison, S. C. (1994) Crystal structure of transcription factor E47: E-box recognition by a basic region helix−loop−helix dimer, *Genes Dev. 8*, 970−980.
28. Ma, P. C., Rould, M. A., Weintraub, H., and Pabo, C. O. (1994) Crystal structure of MyoD bHLH domain−DNA complex: Perspectives on DNA recognition and implications for transcriptional activation, *Cell 77*, 451−459.
29. Ohneda, K., Mirmira, R. G., Wang, J., Johnson, J. D., and German, M. S. (2000) The homeodomain of PDX-1 mediates multiple protein−protein interactions in the formation of a transcriptional activation complex on the insulin promoter, *Mol. Cell. Biol. 20*, 900−911.
30. Leslie, A., Brick, P., and Wonacott, A. (1992) Recent changes to the MOSFLM package for processing film and image plate data, *Joint CCP4/ESF-EAMCB Newslett. Protein Crystallogr*.
31. Holton, T., and Alber, T. (2003) Automated protein crystal structure determination using ELVES, *Proc. Natl. Acad. Sci. U.S.A. 101*, 1537−1542.
32. Collaborative Computation Project (1994) The CCP4 suite: Programs for protein crystallography, *Acta Crystallogr. D50*, 760−763.
33. Evans, P. (1993) Data reduction, in *Proceedings of CCP4 Study Weekend on Data Collection & Processing*, pp 114−122.
34. French, G., and Wilson, K. (1978) On the treatment of negative intensity observations, *Acta Crystallogr. A34*, 517.
35. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr. D Biol. Crystallogr. 54*, 905−921.
36. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models, *Acta Crystallogr A 47*, 110−119.
37. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283−291.
38. DeLano, W. L. (2002) The PyMOL Molecular Graphics System, on World Wide Web at http://www.pymol.org.
39. Gerstein, M., Anderson, B. F., Norris, G. E., Baker, E. N., Lesk, A. M., and Chothia, C. (1993) Domain closure in lactoferrin. Two hinges produce a see-saw motion between alternative close-packed interfaces, *J. Mol. Biol. 234*, 357−372.
40. MacKerell, A. D., Jr., Brooks, B., Brooks, C. L., III, Nilsson, L., Roux, N. B., Won, Y., and Karplus, M. (Eds.) (1998) *CHARMM: The energy function and its parametrization with an overview of the program*, Vol. 1, John Wiley & Sons, Chichester.
41. German, M. S., Moss, L. G., Wang, J., and Rutter, W. J. (1992) The insulin and islet amyloid polypeptide genes contain similar cell-specific promoter elements that bind identical beta-cell nuclear complexes, *Mol. Cell. Biol. 12*, 1777−1788.
42. Wendt, H., Thomas, R. M., and Ellenberger, T. (1998) DNA-mediated folding and assembly of MyoD−E47 heterodimers, *J. Biol. Chem. 273*, 5735−5743.
43. Maleki, S. J., Royer, C. A., and Hurlburt, B. K. (2002) Analysis of the DNA-binding properties of MyoD, myogenin, and E12 by fluorescence anisotropy, *Biochemistry 41*, 10888−10894.
44. Maleki, S. J., Royer, C. A., and Hurlburt, B. K. (1997) MyoD−E12 heterodimers and MyoD−MyoD homodimers are equally stable, *Biochemistry 36*, 6762−6767.
45. Sun, X. H., and 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.
46. Sun, X. H., and Baltimore, D. (1991) Correction: An inhibitory domain of E12 transcription factor prevents DNA binding in E12 homodimers but not in E12 heterodimers, *Cell 66*, 423.
47. Atchley, W. R., Terhalle, W., and Dress, A. (1999) Positional dependence, cliques, and predictive motifs in the bHLH protein domain, *J. Mol. Evol. 48*, 501−516.

48. Docherty, H. M., Hay, C. W., Ferguson, L. A., Barrow, J., Durward, E., and Docherty, K. (2005) Relative contribution of PDX-1, MafA and E47/beta2 to the regulation of the human insulin promoter, *Biochem. J. 389*, 813−820.
49. Ferre-D'Amare, A. R., Pognonec, P., Roeder, R. G., and Burley, S. K. (1994) Structure and function of the b/HLH/Z domain of USF, *EMBO J. 13*, 180−189.
50. Shimizu, T., Toumoto, A., Ihara, K., Shimizu, M., Kyogoku, Y., Ogawa, N., Oshima, Y., and Hakoshima, T. (1997) Crystal structure of PHO4 bHLH domain−DNA complex: Flanking base recognition, *EMBO J. 16*, 4689−4697.
51. Matthews, B. W. (1996) Structural and genetic analysis of the folding and function of T4 lysozyme, *FASEB J. 10*, 35−41.
52. O'Shea, E. K., Rutkowski, R., and Kim, P. S. (1992) Mechanism of specificity in the Fos−Jun oncoprotein heterodimer, *Cell 68*, 699−708.
53. Glover, J. N., and Harrison, S. C. (1995) Crystal structure of the heterodimeric bZIP transcription factor c-Fos−c-Jun bound to DNA, *Nature 373*, 257−261.
54. Chen, L., Glover, J. N., Hogan, P. G., Rao, A., and Harrison, S. C. (1998) Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA, *Nature 392*, 42−48.
55. Panne, D., Maniatis, T., and Harrison, S. C. (2004) Crystal structure of ATF-2/c-Jun and IRF-3 bound to the interferon-beta enhancer, *EMBO J. 23*, 4384−4393.
56. Aramata, S., Han, S. I., Yasuda, K., and Kataoka, K. (2005) Synergistic activation of the insulin gene promoter by the beta-cell enriched transcription factors MafA, Beta2, and Pdx1, *Biochim. Biophys. Acta 1730*, 41−46.
57. Zhao, L., Guo, M., Matsuoka, T. A., Hagman, D. K., Parazzoli, S. D., Poitout, V., and Stein, R. (2005) The islet beta cell-enriched MafA activator is a key regulator of insulin gene transcription, *J. Biol. Chem. 280*, 11887−11894.
58. Poulin, G., Lebel, M., Chamberland, M., Paradis, F. W., and Drouin, J. (2000) Specific protein-protein interaction between basic helix−loop−helix transcription factors and homeoproteins of the Pitx family, *Mol. Cell. Biol. 20*, 4826−4837.
59. Sauve, S., Naud, J. F., and Lavigne, P. (2007) The mechanism of discrimination between cognate and non-specific DNA by dimeric b/HLH/LZ transcription factors, *J. Mol. Biol. 365*, 1163−1175.

BI701527R