

Structural Basis for DNA Recognition by the Basic Region Leucine Zipper Transcription Factor CCAAT/Enhancer-binding Protein α^*

Received for publication, January 14, 2003
Published, JBC Papers in Press, February 10, 2003, DOI 10.1074/jbc.M300417200

Maria Miller^{‡§}, Jon D. Shuman[¶], Thomas Sebastian[¶], Zbigniew Dauter^{||}, and Peter F. Johnson^{¶**}

From the [‡]Protein Structure Section, Macromolecular Crystallography Laboratory, and the [¶]Regulation of Cell Growth Laboratory, National Cancer Institute-Frederick, Frederick, Maryland 21702-1201 and ^{||}Synchrotron Radiation Research Section, Macromolecular Crystallography Laboratory, National Cancer Institute, and National Synchrotron Light Source, Brookhaven National Laboratory, Upton, New York 11973

CCAAT/enhancer-binding proteins (C/EBPs) are basic region leucine zipper (bZIP) transcription factors that regulate cell differentiation, growth, survival, and inflammation. To understand the molecular basis of DNA recognition by the C/EBP family we determined the x-ray structure of a C/EBP α bZIP polypeptide bound to its cognate DNA site (A⁻⁵T⁻⁴T⁻³G⁻²C⁻¹G¹C²A³A⁴T⁵) and characterized several basic region mutants. Binding specificity is provided by interactions of basic region residues Arg²⁸⁹, Asn²⁹², Ala²⁹⁵, Val²⁹⁶, Ser²⁹⁹, and Arg³⁰⁰ with DNA bases. A striking feature of the C/EBP α protein-DNA interface that distinguishes it from known bZIP-DNA complexes is the central role of Arg²⁸⁹, which is hydrogen-bonded to base A³, phosphate, Asn²⁹² (invariant in bZIPs), and Asn²⁹³. The conformation of Arg²⁸⁹ is also restricted by Tyr²⁸⁵. In accordance with the structural model, mutation of Arg²⁸⁹ or a pair of its interacting partners (Tyr²⁸⁵ and Asn²⁹³) abolished C/EBP α binding activity. Val²⁹⁶ (Ala in most other bZIPs) contributes to C/EBP α specificity by discriminating against purines at position -3 and imposing steric restraints on the invariant Arg³⁰⁰. Mutating Val²⁹⁶ to Ala strongly enhanced C/EBP α binding to cAMP response element (CRE) sites while retaining affinity for C/EBP sites. Thus, Arg²⁸⁹ is essential for formation of the complementary protein-DNA interface, whereas Val²⁹⁶ functions primarily to restrict interactions with related sequences such as CRE sites rather than specifying binding to C/EBP sites. Our studies also help to explain the phenotypes of mice carrying targeted mutations in the C/EBP α bZIP region.

The transcription factor C/EBP α ¹ is the founding member of the bZIP class of DNA-binding proteins (1). C/EBP α was orig-

inally identified as a DNA binding activity in rat liver nuclear extracts (2–4) and has since been shown to regulate terminal differentiation of several cell types including adipocytes (5–7) and neutrophilic granulocytes (reviewed in Ref. 8). Consistent with these functions, mice carrying a homozygous deletion of the C/EBP α gene die at birth due to energy imbalance caused by impaired glycogen storage in the liver and lack lipid accumulation in their adipose tissues (9). C/EBP α -deficient embryos also display a complete absence of mature neutrophils and increased numbers of immature myeloid progenitor cells (10). Furthermore, mutations in the C/EBP α gene have been found in a subclass of human myeloid leukemias (11, 12). It has been proposed that loss of C/EBP α can contribute to the development of myeloid neoplasias by impairing the normal program of cellular differentiation and mitotic arrest (11, 12).

bZIP proteins function as transcriptional regulators in most or all eukaryotes (13) and can be arranged into several subfamilies, each recognizing a unique palindromic DNA motif (13) (Fig. 1). The canonical bZIP DNA-binding domain consists of a basic region juxtaposed to a sequence of heptad leucine repeats (the leucine zipper) (1). bZIP basic regions show a high degree of sequence similarity and contain two invariant residues, Asn and Arg (Fig. 1). The bZIP-DNA complex consists of two α -helices lying perpendicular to the DNA, associated in a coiled-coil structure with each basic region contacting a half-site in the DNA major groove (1, 14). A unique aspect of DNA recognition by bZIP proteins is that the helical fold of the basic region, which creates a surface complementary to the target DNA sequence, is induced upon association with the DNA ligand (15–17).

To date, x-ray structures of bZIP domain peptides bound to DNA have been determined for GCN4 (bound either to an AP-1 site (18) or to a CRE site (19, 20)), the Jun/Fos heterodimer (21), CREB (22), and PAP1 (23). These structures show that bZIPs recognize specific DNA sites through base contacts made by five residues within the basic region motif characteristic for each subfamily. These five positions are well conserved among all bZIP proteins and include the invariant Asn and Arg residues (Fig. 1). The invariant Asn contacts bases C³ and T⁻⁴ in the complexes of GCN4, Fos/Jun, and CREB with DNA (G¹T²C³A⁴) (18, 20–22), whereas in PAP1 bound to a PAR site (G¹T²A³A⁴) this residue adopts a different conformation and makes a direct contact with base A⁴ (23). Thus, the bZIP structures have revealed functional variability of conserved residues in DNA recognition.

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 1NWQ) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

[§] To whom correspondence for structure-related issues should be addressed: Macromolecular Crystallography Laboratory, NCI-Frederick, Frederick, MD 21702-1201. Tel.: 301-846-5342; Fax: 301-846-7101; E-mail: millerm@ncifcrf.gov.

^{**} To whom correspondence for all other issues should be addressed: Regulation of Cell Growth Laboratory, NCI-Frederick, Frederick, MD 21702-1201. Tel.: 301-846-1627; Fax: 301-846-5991; E-mail: johnsopf@ncifcrf.gov.

¹ The abbreviations used are: C/EBP, CCAAT/enhancer-binding protein; bZIP, basic region leucine zipper; CRE, cyclic AMP response element; CREB, cAMP-response element-binding protein; r.m.s.d., root

mean square deviation; wt, wild type; EMSA, electrophoretic mobility shift assay; MES, 4-morpholineethanesulfonic acid; GCN, general control of amino acid biosynthesis; AP-1, activating protein 1; PAR, proline and acidic amino acid-rich; E2F, E2 promoter binding factor.

Protein	Basic Region	Binding Site	Family
CEBP α	YRVRRERNNIAVRRKSRDKAKQRNVET 290 300 310	ATTGCGCAAT	C/EBP
CREB	REVRLMKNRREAARECRHKKKEYDKCL	ATGACGTCAT	CREB/ATF
GCN4	AALKRRARNTAARRSRARKLQRMKQL		
GCN4	AALKRRARNTAARRSRARKLQRMKQL		
c-FOS	RRIRRRERNNKMAAAKCRNRRRELTDTL	ATGAGTCAT	AP-1
c-JUN	AERKRMRNRRIAASKCRKRKLERIARL		
PAP1	SSKRKAQNRAAQRFRKRKEDHLKAL	GTTACGTAAC	PAP
DBP	YWSRRERNNAAAKRSRDARRLKENQI	ATTACGTAAT	PAR

FIG. 1. Basic region amino acid sequences and DNA binding sites for representatives of bZIP subfamilies. Shown here are proteins for which crystal structures have been determined (except D box-binding protein (DBP)). The invariant Asn and Arg are depicted in red, conserved basic residues in blue, and the first residue of the leucine zipper in green. The core sequences of the DNA half-sites are underlined.

Here we present the crystal structure of a C/EBP α bZIP polypeptide bound to its cognate DNA site. The structure reveals that two residues (Tyr²⁸⁵ and Arg²⁸⁹) located amino-terminally to the signature bZIP basic region motif are integral components of the protein-DNA interface. The functional significance of these residues was confirmed by analysis of substitution mutants. Arg²⁸⁹ is conserved in AP-1 proteins but adopts a very different conformation; therefore, its importance for DNA binding in C/EBP proteins could not have been predicted from structures of other bZIP proteins nor was it indicated by previous mutagenesis studies. Finally, our data clarify the key role of Val²⁹⁶ in determining C/EBP binding specificity. This study thus provides the first detailed description of DNA recognition by a C/EBP family member.

EXPERIMENTAL PROCEDURES

Protein Production and Crystallization—A protein containing residues 281–340 of rat C/EBP α was over-produced in *Escherichia coli* and purified as described previously (15). Cloning appended the sequence Met-Gly-Ser to the amino terminus of the polypeptide; the terminal Met residue was removed in *E. coli*, resulting in a 62-amino acid product (Fig. 2A). DNA oligomers (Yale Keck Laboratory) were gel-purified.

The crystallization strategy included trials with DNA duplexes of varying length and 5' overhanging bases. The best diffracting cocrystals were obtained with a 21-mer DNA duplex (Fig. 2B). Crystals were grown at 4 °C in hanging drops prepared by mixing equal volumes of the protein-DNA complex (0.75 mM bZIP monomer and 0.5 mM DNA duplex in 50 mM NaCl, 100 mM MES, pH 5.7) and the precipitant solution (100 mM MES, pH 5.7, 100 mM NaCl, 30 mM MgCl₂, 18% polyethylene glycol 400, 10% glycerol) and were harvested for flash freezing directly from the mother liquor. Crystals belong to the P2₁,2 space group with unit cell dimension $a = 140.89$ Å, $b = 53.09$ Å, and $c = 67.41$ Å. The solvent content of the crystals was estimated as 73% (24).

Crystallographic Procedures—X-ray data were collected at 100 K, 1.07-Å wavelength on beamline X9B, National Synchrotron Light Source, Brookhaven National Laboratory, using the ADSC CCD Quantum-4 detector, and were processed with HKL2000 (25) (Table I). The structure was solved by molecular replacement using the AMoRe package (26) with the truncated (residues 286 to 339 from chains A and C, -10 to +10 from chains B and D) polyalanine model derived from the CREB bZIP-DNA complex (22) (Protein Data Bank code: 1DH3). The asymmetric unit contains one protein-DNA complex. The DNA segments form a pseudo-continuous helix with base pair interactions between complementary 5'-overhanging bases from the adjacent complexes. The maximum likelihood target refinement was carried out with CNS_1.0 (27) initially against data from the 10–2.8-Å resolution shell. The solution was first refined as three rigid groups: the coiled-coiled region and the basic region of each monomer with its DNA half-site. Subsequent cycles of simulated annealing refinement and model rebuilding with O, version 8.0 (28) allowed localization of all missing residues (except the two derived from the expression vector) and fitting of most of the protein side chains and the DNA sequence. The orientation of the DNA backbone was determined based on the electron density corresponding to the asymmetric ends flanking the central palindromic sequence. Further refinement against all data with the correction for

TABLE I
Data collection and refinement statistics

Diffraction data statistics		
Cell dimensions (Å) in space group P2 ₁ ,2	140.89, 53.09, 67.41	
Resolution Å (last shell)	40.0–2.8 (2.9–2.8)	
No. of measurements: total/independent	222,356/12,415	
Completeness (last shell),%	95.6 (91.6)	
R_{merge} (last shell)	0.076 (0.415)	
I/δ_i (last shell)	19.2 (2.7)	
Wilson B-factor (Å ²) ^a	94.6	
Refinement statistics		
No. of reflections: working set/test set	11,785/630	
R/R_{free}	0.221/0.275	
No. of atoms: protein/DNA/water	1020/855/34	
r.m.s.d. of bonds (Å)/angles (°)	0.019/2.0	
Average B-factors (Å ²) ^b	92.0 (77.5)	
r.m.s.d. for bonded atoms (Å ²)	5.3	

^a Calculated with TRUNCATE from the CCP4 program suite.

^b Values in parentheses are for residues involved in protein-DNA interactions; i.e. the two recognition helices (residues 285–300) and the 10 central DNA base pairs.

bulk solvent was followed by restrained thermal parameter refinement. Water molecules were added based on peak heights ($\geq 3\sigma$) in the difference Fourier map and proper hydrogen bond distance. Several simulated annealing OMIT maps were calculated during the course of refinement to verify the model. The geometrical properties of the model were analyzed with the programs PROCHECK (29) and 3DNA (30). The refinement statistics are summarized in Table I. The mean B-factor of 91.6 Å² of the final model correlates well with an estimate of Wilson plot B-factors.

Mutagenesis and Reporter Assays—Specific amino acid changes were introduced into full-length rat C/EBP α or C/EBP β (in pcDNA3.1) with the QuikChange mutagenesis kit (Stratagene). For transactivation assays, 8×10^4 L929 cells were transfected with 1 μ g of 2 \times C/EBP Luc or 2 \times CRE Luc reporter plasmids and 25 ng of C/EBP α or C/EBP β expression plasmids using Fugene6 (Roche Molecular Biochemicals). A *Renilla* luciferase vector (pRLTK, Promega) was included to correct for transfection efficiency. After 48 h the cells were harvested, and lysates were prepared and assayed for reporter expression using the Dual Luciferase System (Promega). Firefly luciferase activity was normalized to *Renilla* luciferase, and the ratio for reporter alone was set to 1. Three experiments were averaged, and the data were graphed as the mean \pm S.E.

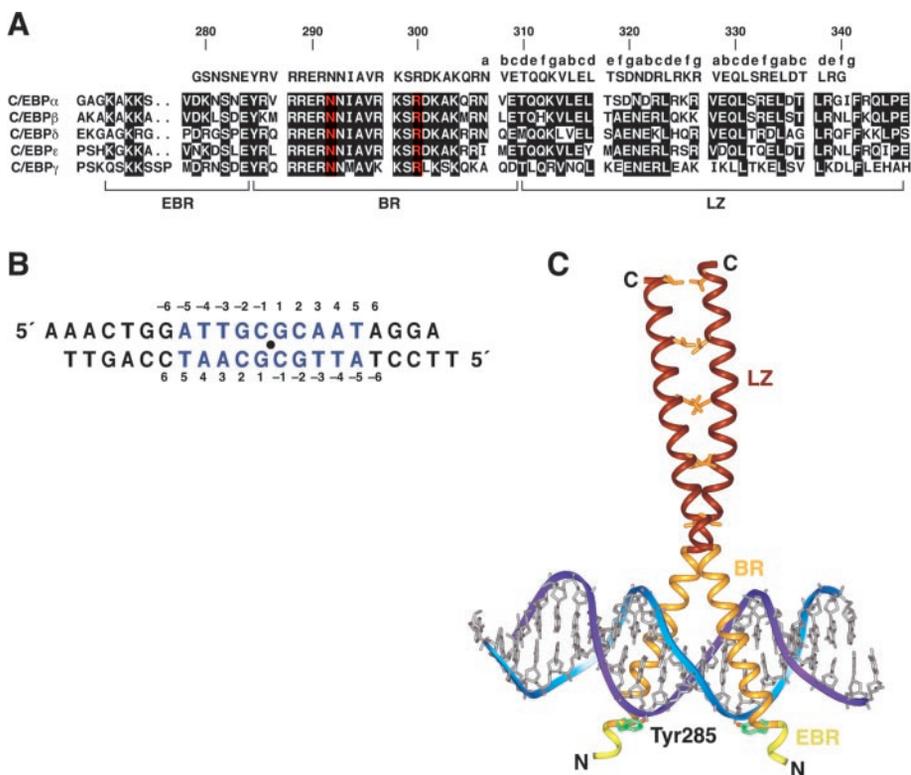
In Vitro Protein Expression and EMSA—Proteins were expressed using the TnT Quick transcription/translation system (Promega). For Western blot analysis, 1 μ l of each protein extract was separated on 12% SDS-PAGE, transferred to nitrocellulose, and probed using C/EBP α or C/EBP β antibodies (C-14 and C-19, respectively; Santa Cruz Biotechnology). Secondary goat anti rabbit-horseradish peroxidase was used to visualize antigen-antibody with chemiluminescence (ECL, Pierce). DNA binding reactions contained 20 mM Hepes, pH 7.5, 5% Ficoll, 1 mM EDTA, 200 mM NaCl, 1 mM dithiothreitol, 0.01% Nonidet P-40, 400 ng/ μ l bovine serum albumin, and 40 ng/ μ l poly(dI-dC). The following probes (binding sites indicated in bold type) were used for EMSA assays:

C/EBP probe: 5'-GATCCATATCCCTG**ATTGCGCAAT**AGGCTCAAAA
GTATAGGGACT**TAACCGCTT**ATCCGAGTTTCTAG-5'
CRE probe: 5'-GATCCATATCCCTG**GATGCGTC**ATAGGCTCAAAA
GTATAGGGACT**ACTGCGAT**ATCCGAGTTTCTAG-5'
PAR probe: 5'-GATCCATATCCCTG**ATTACGTA**ATAGGCTCAAAA
GTATAGGGACT**TAATGCATT**ATCCGAGTTTCTAG-5'

RESULTS

Overview of the Complex—We crystallized a C/EBP α bZIP polypeptide bound to a 21-mer DNA duplex and determined the structure at 2.8 Å resolution (Fig. 2). The complex consists of two polypeptides (residues 281–340) that form a dimer of α -helices associated with DNA in the well known fork-like structure (14, 18). Residues 285–300 from each subunit comprise the recognition helix located in the major groove of each DNA half-site. The DNA duplex adopts a nearly straight B-type helix with an average rise per nucleotide of 3.3 Å. Residues 281–284 from the “extended basic region” characteristic of the C/EBP family (Fig. 2) do not participate in any intramolecular or DNA contacts and are highly mobile.

FIG. 2. The C/EBP α bZIP-DNA complex. *A*, sequence alignment of the C/EBP family members. The C/EBP α peptide used for crystallization is shown at the top. The numbering refers to the complete rat C/EBP α protein sequence. Lowercase letters (top line) show the positions of leucine zipper residues as they would appear on a standard helical wheel representation of a coiled-coil dimer; "d" corresponds to the position of the leucine residues. Residues conserved within the family are shaded in black; Asn and Arg, present in all bZIPs, are in red. LZ, leucine zipper; BR, basic region; EBR, extended basic region. *B*, DNA duplex used for crystallization. The consensus C/EBP recognition site is indicated in blue, and the center of symmetry is shown as a solid circle. *C*, overall structure of the C/EBP α -DNA complex.



The first contact between the two protein subunits is made via electrostatic interactions between the side chains of their Asn³⁰⁷ residues. The C/EBP α leucine zipper contains two hydrophilic residues, Thr³¹⁰ and Asn³²¹, at the *d* and *a* positions of the heptad repeat (Fig. 2A), buried in the hydrophobic core of the coiled-coil dimer. The side chain hydroxyl oxygen of Thr³¹⁰ from each chain can potentially form a hydrogen bond to Gln³¹¹ from the other subunit (Gln^{311'}). Asn³²¹ is conserved among many bZIP transcription factors and, as observed in GCN4 (18, 19) and CREB (22), its side chains form an inter-helical hydrogen bond. The dimer is stabilized by two salt links formed by the interaction of Asp³²⁰ with Arg^{325'} and Glu³³⁴ with Arg^{339'}. Formation of equivalent salt bridges between the reciprocal pairs of subunits is prevented by crystal contact interactions.

The two protein chains are bent by crystal packing forces. The complete chains superimpose with an r.m.s.d. of 2.03 Å for 118 C α atom pairs, whereas for the leucine zippers the r.m.s.d. is 1.06 Å for superposition of 272 backbone atoms. The two basic regions immersed in the DNA major groove are very similar, the r.m.s.d. for superposition of the backbones of recognition helices together with all atoms of the corresponding DNA half-sites is 0.46 Å for 538 common atoms. Thus, the protein-DNA interface is not significantly affected by crystal packing.

Protein-DNA Interface—The boundaries of the C/EBP α -DNA interface are defined by Tyr²⁸⁵ and Arg³⁰⁰. Twelve amino acids from each recognition helix contact the DNA half-sites (Fig. 3A) essentially in a symmetrical manner. Eight of the side chains are potential proton donors in hydrogen bonds with phosphate oxygen atoms from the DNA backbone. Direct contacts with DNA bases are made by Arg²⁸⁹, Asn²⁹², Ala²⁹⁵, Val²⁹⁶, Ser²⁹⁹, and Arg³⁰⁰ (Fig. 3B). Details of the C/EBP α -DNA interface are shown in Figs. 4 and 5. The invariant Asn²⁹² is positioned to make side chain hydrogen bond interactions with the O-4 atom of base T⁻⁴ and the N-6 amino group of base A³. Adenine at position 3 is also specified by Arg²⁸⁹, which contacts the N-7 atom of A³. C β atoms of Ala²⁹⁵ and Ser²⁹⁹ as well as C γ_2 of Val²⁹⁶ make van der Waals contacts with the methyl group of

T⁻⁴. An apolar environment for the 5-methyl group of T⁻³ is formed by atoms from the side chain of Val²⁹⁶ and the aliphatic portion of Arg³⁰⁰ (Fig. 5A). The guanidinium group of Arg³⁰⁰ makes electrostatic interactions with G¹ and G⁻². Arg³⁰⁰ was modeled in one conformation in both subunits. However, weak electron density in this region of the second subunit (Arg^{300'}) indicates that Arg^{300'} is perhaps disordered with a low occupancy in this conformation. Asymmetric contacts of the invariant Arg were also observed in the structure of CREB bound to a symmetric CRE site (22).

The protein-DNA interface is stabilized by interactions between side chains of basic region amino acids. Particularly important in this respect is the buried side chain of Arg²⁸⁹, which can make bidentate hydrogen bonds with carboxylate oxygens from Asn²⁹² and Asn²⁹³. In addition, the aliphatic portion of the Arg²⁸⁹ side chain makes extensive van der Waals contacts with the phenolic ring of Tyr²⁸⁵, whereas Val²⁹⁶ contacts the C δ and C γ atoms of Arg³⁰⁰ (Figs. 4A and 5B).

The conformation and functional roles in DNA recognition of Asn²⁹², Ala²⁹⁵, and Ser²⁹⁹, residues that are conserved among bZIP proteins, are similar to those observed in GCN4 and CREB bound to CRE sites (Fig. 4B). In contrast, the ability of the invariant Arg³⁰⁰ residue to adopt the conformation seen in GCN4/CREB, where it contacts base G¹ and DNA phosphate oxygen, is precluded by the proximity of Val²⁹⁶ (Fig. 5B). In this manner, Val²⁹⁶ enhances the preference for G⁻² in the consensus CEBP binding site. However, the most striking difference between the C/EBP α and GCN4 structures is the conformation of Arg²⁸⁹ (Fig. 4). In AP-1 proteins, including Jun/Fos, this residue does not participate in specific base recognition (18, 19, 21, 22). Furthermore, the cluster of polar residues involving Arg²⁸⁹, Asn²⁹², and Asn²⁹³ pushes base C² away from the position occupied by T² in the GCN4 (Fig. 5C) and CREB (not shown) complexes and enforces movement of the DNA backbone away from the protein in this region. Because of this displacement, Val²⁹⁶ does not interact with base C² and would accept thymine at position 2 of the half-site (see below). In AP-1 and CREB proteins, Ala occupies the position analogous to

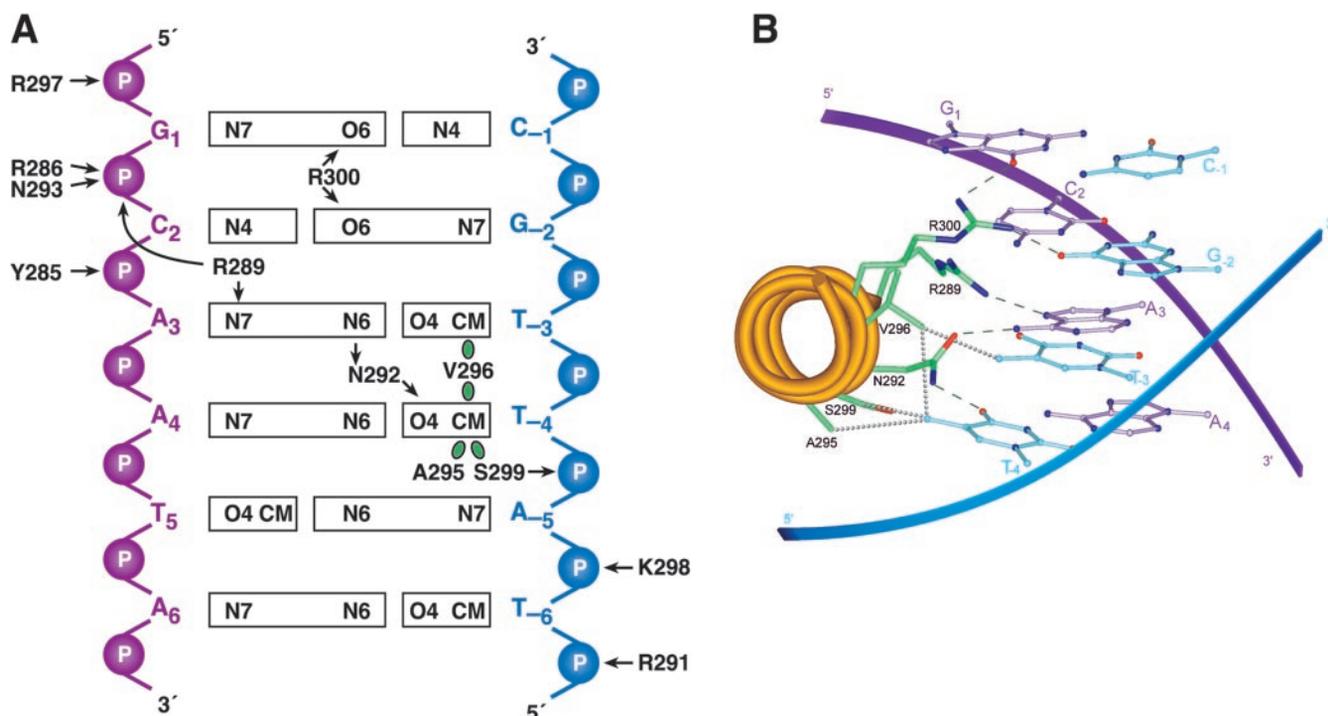
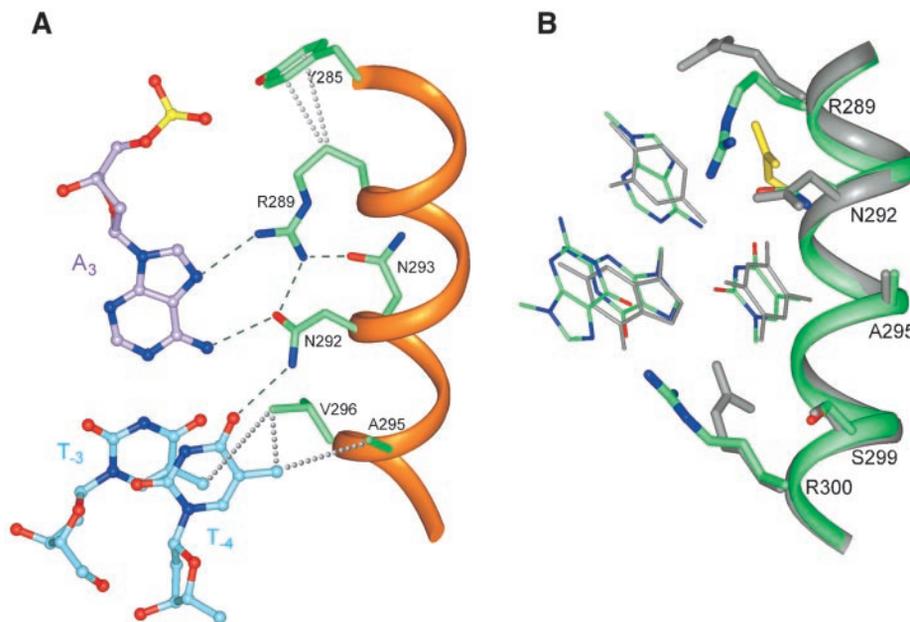


FIG. 3. **C/EBP α -DNA interactions.** *A*, schematic representation of interactions between residues from one basic region domain and a DNA half-site. Directions of possible hydrogen bonds are indicated by *black arrows*; hydrophobic interactions are shown as *ovals*. *B*, direct contacts between the side chains from the recognition helix and DNA bases. Selected side chains are represented as *sticks* and DNA as *balls-and-sticks*. Oxygen atoms are shown in *red*, nitrogen atoms in *blue*, and phosphate atoms in *yellow*. One DNA chain is *violet*, and the other is *blue*. Possible hydrogen bonds and van der Waals interactions are depicted as *dashed* and *dotted* lines, respectively. The DNA-sugar-phosphate backbone is represented as a *ribbon*.

FIG. 4. **Architecture of the protein surface complementary to the cognate DNA.** *A*, critical interactions in the C/EBP α protein-DNA interface. Protein side chains are represented as *sticks* and DNA as *balls-and-sticks*. Selected electrostatic and van der Waals interactions are depicted as *dashed* and *dotted* lines, respectively. *B*, comparison of conformations of the conserved side chains in the basic regions of C/EBP α (*green*) and GCN4 (*gray*). The side chain of the invariant Asn residue from the PAP1 structure is shown in *yellow*.



Val²⁹⁶ and specifies thymine at position 2 by hydrophobic contact with its methyl group (Fig. 5C).

Members of the PAP bZIP family recognize a half-site sequence (G¹T²A³A⁴) that differs from the consensus C/EBP site only at position 2. In the structure of PAP1 bound to its cognate DNA, the invariant Asn adopts a different conformation from that seen in C/EBP α and GCN4 (Fig. 4B) and does not contact base A³ directly (23). PAP1 recognizes the A³·T⁻³ base pair in a very different manner than C/EBP α because of dissimilarities in their signature basic region sequences (Fig. 1).

Basic Region Mutants—To test their functional roles, we mutated several basic region amino acids predicted from the

structure to be important for C/EBP DNA recognition (Fig. 6A). Tyr²⁸⁵, Arg²⁸⁹, and Val²⁹⁶ were converted to Ala and Asn²⁹³ was mutated to Arg, the corresponding residue in most AP-1 family members (Fig. 1). The mutations were introduced into full-length C/EBP α individually or in combination. Mutant proteins were first tested for their ability to activate transcription from a synthetic C/EBP-dependent reporter gene (2 \times C/EBP-luc) in L929 fibroblasts (Fig. 6B, *left panel*). Although wt C/EBP α stimulated 2 \times C/EBP-luc by \sim 15-fold, most of the basic region mutations severely reduced or eliminated transactivation. The activities of Y285A, N293R, and Y285A/N293R were 5–6-fold lower than wt, whereas R289A was essentially

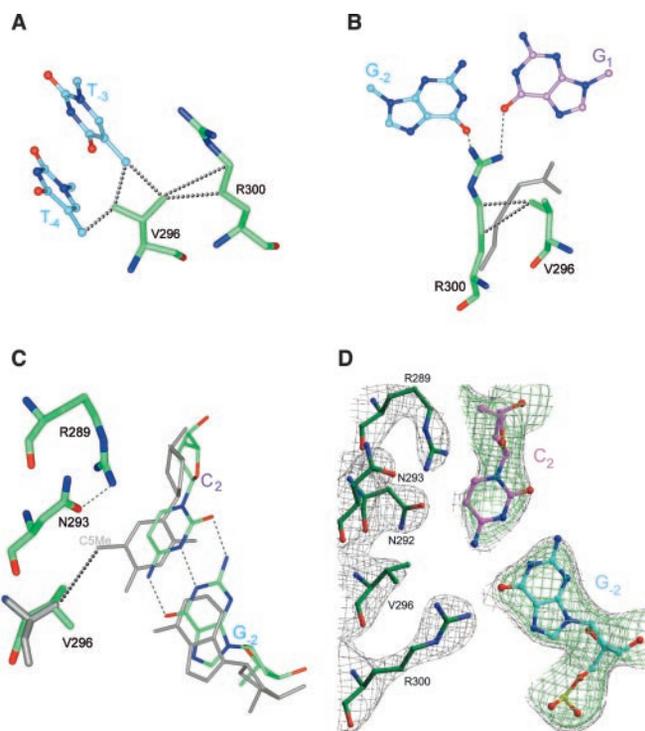


FIG. 5. **Details of the C/EBP α protein-DNA interface.** *A*, hydrophobic cluster involving two thymine moieties from the C/EBP site and side chains of Val²⁹⁶ and Arg³⁰⁰. *B*, interactions defining possible conformation of Arg³⁰⁰. The side chain of the homologous Arg residue from GCN4 (one observed conformation) is shown in gray. *C*, comparison of the protein environment of base 2 in C/EBP α (green) and GCN4 (gray) complexes. Hydrophobic interaction of the methyl group from T² and Ala occurring in GCN4 is marked by a dotted line. Note that C² in the C/EBP site does not interact with Val²⁹⁶, and its position is displaced relative to T² from the GCN4-CREB complex. *D*, electron density ($2F_o - F_c$) contoured at 1.0 σ (gray) and 1.8 σ (green; DNA only) is shown for important residues, with the coordinates superimposed.

inactive. Mutating the residue analogous to Tyr²⁸⁵ in C/EBP β (Y226A) also severely reduced transactivation (Fig. 6B), indicating that the Tyr residue has an important functional role within the C/EBP family. Interestingly, changing Val²⁹⁶ to Ala did not impair transactivation of the C/EBP reporter and in fact caused a modest increase in activity. Similarly, the Y285A/N293R/V296A mutant was ~4-fold more active than Y285A/N293R.

Several of the substitution mutants contain amino acids that occur at the equivalent positions in AP-1 or CREB proteins (see Fig. 1). Therefore, we tested their ability to transactivate a CRE-dependent reporter gene, 2 \times CRE-luc (Fig. 6B, right panel). wt C/EBP α weakly activated 2 \times CRE-luc (~3-fold), consistent with its low affinity for CRE sites. The Y285A, R289A, N293R, and Y285A/N293R mutants showed reduced activity. In marked contrast, V296A and Y285A/N293R/V296A activated the reporter by 9-fold and 7-fold, respectively. Thus, the V296A substitution significantly enhances the ability of C/EBP α to transactivate a CRE-driven promoter.

We next examined the binding specificities of recombinant C/EBP proteins using the electrophoretic mobility shift assay (EMSA) (Fig. 6C). wt C/EBP α (lane 2) bound efficiently to the C/EBP probe but interacted very weakly with the CRE. R289A (lane 4) did not bind appreciably to either probe. Interaction of C/EBP α Y285A (lane 3) and N293R (lane 5) with both probes was also significantly reduced, as was C/EBP β Y226A (lane 10). Y285A/N293R and Y285A/N293R/V296A bound poorly to the C/EBP site, whereas V296A activity was similar to wt C/EBP α . Strikingly, both mutants bearing the V296A substitution associated very efficiently with the CRE probe (lanes 6 and 8).

EMSA quantitation showed that the V296A substitution enhanced binding to the CRE site by 7–8-fold (Fig. 6C). Thus, Val²⁹⁶ inhibits the interaction of C/EBP α with CRE sites, whereas Ala strongly enhances binding to CREs but does not impair binding to C/EBP elements. We also examined affinity of the mutant proteins for a PAR site (Fig. 6C). Although the overall pattern of binding was similar to that for the C/EBP probe, the V296A mutant displayed increased binding to the PAR probe (1.4-fold greater than the wt protein). In addition, the Y285A/N293R and Y285A/N293R/V296A mutants maintained weak binding to the PAR site, in contrast to their severe effects on binding to C/EBP sites. Thus, with the exception of Y285A and R289A, the C/EBP α mutations were less detrimental to PAR site binding than to interaction with the C/EBP element.

DISCUSSION

Binding Specificity of C/EBP Proteins—The data presented here reveal the structural basis of C/EBP α binding to a consensus (*i.e.* highest affinity) C/EBP recognition element. However, C/EBP α frequently binds *in vivo* to promoters containing non-canonical C/EBP sites (see Ref. 31 for examples). Binding site selection experiments have established the preference of C/EBP family members for the sequence RTTGCGYAAY (where R and Y denote purines and pyrimidines, respectively); some tolerance for different bases at the first position of the half-site was also observed in these studies (32, 33). Allowed substitutions usually do not occur symmetrically, and the modified sites tend to be pseudo-palindromic.

The strict requirement for adenine at half-site positions 3 and 4 is clearly explained by the structure of the C/EBP α bound to consensus DNA site. The base pair A⁴·T⁻⁴ is defined by a very tight hydrophobic pocket for the methyl group of base T⁻⁴ and hydrogen bond of its carbonyl to Asn²⁹². Adenine, and not cytosine, is preferred at base 3 as a proton donor in hydrogen bond to the Asn²⁹² carboxylate oxygen, because its N-7 atom can interact with Arg²⁸⁹. More importantly, preference for its base pair partner (T⁻³) is dictated by the bulky side chain of Val²⁹⁶, which provides the proper environment for the T⁻³ methyl group and effectively discriminates against guanine at this position. On the other hand, base pairs G¹·C⁻¹ and C²·G⁻² are specified only by interactions of the guanine moieties with Arg³⁰⁰, in which the flexible side chain is partially exposed to solvent (Fig. 5B). To accommodate substitution of base pair C²·G⁻² with T²·A⁻², the guanidinium group of Arg³⁰⁰ must move away from DNA into the solvent channel between the two protein chains emerging from the major groove. Thymine can be accepted at position 2 without changes in the DNA backbone (Fig. 5C). This substitution will result in hydrophobic contact of the T² methyl group with Val²⁹⁶, possibly compensating for the loss of the interaction between G⁻² and Arg³⁰⁰. However, because of steric hindrance imposed by the Val²⁹⁶ side chain (see “Results”), movement of the Arg³⁰⁰ side chain is possible only toward the molecular dyad of the complex. Such a displacement of Arg³⁰⁰ on both subunits would place their charged groups too close to each other, which may explain why the T² substitution is usually tolerated on only one of the two half-sites.

Role of Val²⁹⁶ in DNA Recognition—The C/EBP α structure is apparently the first example in which Val plays a key role in DNA recognition. Replacing Val²⁹⁶ with Ala had little effect on the binding of C/EBP α to its consensus site (GCAAT), while greatly increasing affinity for CREB sites (GTCAT). An earlier study (34) found that a C/EBP β mutant containing the analogous Val \rightarrow Ala substitution displayed reduced binding to consensus C/EBP sites but retained its affinity for PAR sites (GTAAT) and concluded that Val is responsible for the relaxed sequence specificity (binding promiscuity) displayed by C/EBP

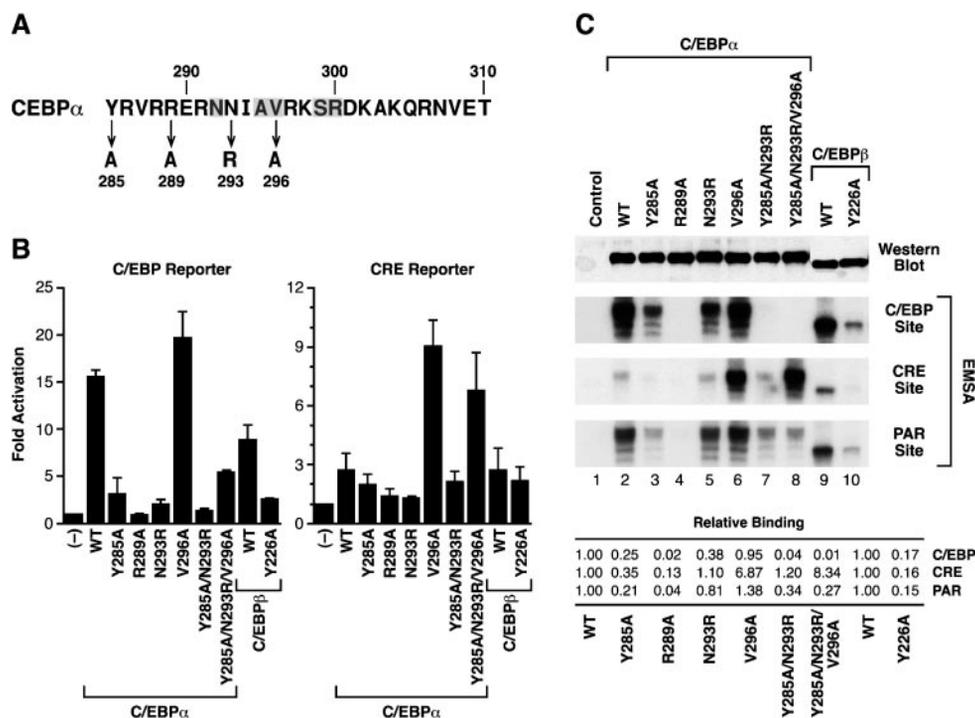


FIG. 6. Analysis of C/EBP basic region mutants. *A*, amino acid substitution mutations in the C/EBP α basic region. *B*, transactivation of C/EBP and CRE reporter genes. Reporter constructs were transfected into L929 cells, alone or with the indicated C/EBP expression vectors. Luciferase activity for the reporter alone (–) was set to 1. Western blotting demonstrated equivalent expression of all proteins (data not shown). *C*, DNA binding assays. *Upper panel*, Western blot analysis of *in vitro* translated proteins. *Lower panels*, EMSAs. The recombinant protein extracts were incubated with radiolabeled C/EBP (*top*), CRE (*middle*), or PAR (*bottom*) probes and the complexes resolved by electrophoresis. Radioactivity in the protein-DNA complexes was divided by the total cpm (bound + free probe). Binding activity is expressed relative to the wt C/EBP protein for each probe.

proteins. In contrast, we observe that replacing Val²⁹⁶ with Ala actually diminishes C/EBP α selectivity by increasing its affinity for CREB and PAR sites.

Our findings can be rationalized in structural terms. The V296A substitution releases restriction on the Arg³⁰⁰ guanidinium group, allowing it to occupy a position similar to that seen in GCN4 and CREB complexes on both subunits. This may account for the ability of V296A to recognize PAR sites. Simultaneously, Ala will tolerate guanine at position –3, permitting binding of the mutant protein to CREB sites. The V296A substitution reduces the area of hydrophobic interaction between the C/EBP site T^{–3} methyl group and its environment (Fig. 5A). However, the resulting loss of free energy of binding may be offset by increased entropy due to Arg³⁰⁰ gaining a second energetically favorable conformation and thereby maintaining affinity of the mutant protein for C/EBP sites. Val²⁹⁶ thus appears to affect C/EBP specificity by discriminating against guanine at position –3 and by restricting the conformation of Arg³⁰⁰.

Importance of Side Chain Interactions in the Protein-DNA Interface—The side chain conformations of Tyr²⁸⁵, Arg²⁸⁹, Asn²⁹², and Asn²⁹³, which form part of the protein surface conforming to the cognate DNA, depend on an extended network of interactions (Fig. 4A). The results of mutating Tyr²⁸⁵, Arg²⁸⁹, and Asn²⁹³ underscore the importance of these stabilizing interactions for DNA binding affinity. Mutation of Arg²⁸⁹ to Ala abolished C/EBP α binding to C/EBP, CRE, and PAR sites, demonstrating the critical role played by this residue in the formation of the complementary protein-DNA interface. The precise conformation of Arg²⁸⁹ is maintained by interactions with Tyr²⁸⁵ and Asn²⁹³, residues that are specific for the C/EBP family. In accordance with the structure, single mutations at either of these positions reduced C/EBP α binding to a consensus C/EBP site, whereas the double mutation Y285A/

N293R nearly abolished binding. On the other hand, the N293R substitution had very little effect on C/EBP α binding to CRE or PAR sites that have T at position 2, the methyl group of which can contact Val²⁹⁶ (Fig. 5C).

Biological Implications—Several groups recently postulated that C/EBP α can cause cell growth arrest by suppressing E2F-mediated transcription and have suggested that C/EBP α associates with and inhibits E2F in a DNA-binding independent manner (35–37). Porse *et al.* (36) proposed that C/EBP α repression of E2F function requires a trio of basic region residues, Tyr²⁸⁵, Ile²⁹⁴, and Arg²⁹⁷, which were predicted to face away from the DNA and thus could interact with proteins such as E2F. Mutation of these amino acids diminished the ability of C/EBP α to repress E2F-dependent reporter genes in transfected cells, and “knock-in” mice carrying either the Y285A or I294A/R297A mutations displayed defects in adipogenesis and granulopoiesis. It was concluded that C/EBP α promotes growth arrest and differentiation in part by making inhibitory interactions with E2F via residues Tyr²⁸⁵, Ile²⁹⁴, and Arg²⁹⁷.

Our findings provide an alternative interpretation for the phenotypes of mice carrying C/EBP α basic region mutations. Tyr²⁸⁵ contributes to a stable protein-DNA interface, and its replacement with Ala severely impairs C/EBP α binding to C/EBP sites in addition to transactivation of a C/EBP reporter gene (Fig. 6). Thus, the defects in mice carrying the Y285A mutation can be explained most simply by the reduced affinity of the mutant protein for C/EBP sites. Suckow *et al.* (38) also reported that an Ile → Glu substitution at the residue corresponding to Ile294 in GCN4 (see Fig. 1) conferred some C/EBP-like characteristics to GCN4, indicating that Ile294 may play a role in determining C/EBP α binding selectivity by influencing the conformation of neighboring side chains. In addition, the structure of the GCN4-CREB site complex (20) suggests that interaction between the analogous pair of GCN4 residues,

Glu²³⁷ and Arg²⁴⁰, is an important determinant of binding specificity by directing the Arg²⁴⁰ side chain to contact the phosphate groups of G¹ and C⁻². In C/EBP α Arg²⁹⁷ interacts only with the phosphate group of G¹. Thus, it is possible that changes in C/EBP α DNA-binding activity and/or specificity also account for the biological defects seen in mice carrying the I294A/R297A mutation. Clearly, further studies are necessary to enhance understanding of the mechanisms by which C/EBP α regulates cell growth arrest and differentiation. The availability of the C/EBP α bZIP structure should facilitate these investigations.

Concluding Remarks—In this paper we present the first detailed description of DNA recognition by a C/EBP family member. Tahirov *et al.* (39, 40) have reported structures of multiprotein complexes that include the bZIP domain of C/EBP β in association with DNA ligands containing non-consensus C/EBP sites; however, the interaction of C/EBP β with DNA was not described in these studies. We examined the C/EBP β structure (Protein Data Bank code: 1HJB) and found that it recognizes DNA in a manner consistent with our model. This is not surprising in view of the nearly identical basic region sequences of the C/EBP α and C/EBP β proteins.

The C/EBP α DNA-binding interface contains both similarities and dissimilarities with other bZIP structures. The distinct functional roles for conserved residues such as Arg²⁸⁹ underscore the notion that mechanisms of specific DNA recognition must be elucidated independently for each bZIP family.

Acknowledgments—We appreciate the encouragement and advice of A. Wlodawer and thank him and C. Vinson for critical reading of the manuscript. We acknowledge the help of colleagues in the Macromolecular Crystallography laboratory, particularly N. Nandhagopal. We also thank V. Pett for help with the initial crystallization experiments. Finally, we recognize the important contributions of S. McKnight and the late P. Sigler during early stages of this work.

REFERENCES

- Landschulz, W. H., Johnson, P. F., and McKnight, S. L. (1988) *Science* **240**, 1759–1764
- Graves, B. J., Johnson, P. F., and McKnight, S. L. (1986) *Cell* **44**, 565–576
- Johnson, P. F., Landschulz, W. H., Graves, B. J., and McKnight, S. L. (1987) *Genes Dev.* **1**, 133–146
- Landschulz, W. H., Johnson, P. F., Adashi, E. Y., Graves, B. J., and McKnight, S. L. (1988) *Genes Dev.* **2**, 786–800
- Lin, F. T., and Lane, M. D. (1992) *Genes Dev.* **6**, 533–544
- Freitag, S. O., Paielli, D. L., and Gilbert, J. D. (1994) *Genes Dev.* **8**, 1654–1663
- Freitag, S. O., and Geddes, T. J. (1992) *Science* **256**, 379–382
- Tenen, D. G., Hromas, R., Licht, J. D., and Zhang, D. E. (1997) *Blood* **90**, 489–519
- Wang, N. D., Finegold, M. J., Bradley, A., Ou, C. N., Abdelsayed, S. V., Wilde, M. D., Taylor, L. R., Wilson, D. R., and Darlington, G. J. (1995) *Science* **269**, 1108–1112
- Zhang, D. E., Zhang, P., Wang, N. D., Hetherington, C. J., Darlington, G. J., and Tenen, D. G. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 569–574
- Pabst, T., Mueller, B. U., Zhang, P., Radoska, H. S., Narravula, S., Schnittger, S., Behre, G., Hiddemann, W., and Tenen, D. G. (2001) *Nat. Genet.* **27**, 263–270
- Tenen, D. G. (2001) *Leukemia* **15**, 688–689
- Vinson, C., Myakishev, M., Acharya, A., Mir, A. A., Moll, J. R., and Bonovich, M. (2002) *Mol. Cell. Biol.* **22**, 6321–6335
- Vinson, C. R., Sigler, P. B., and McKnight, S. L. (1989) *Science* **246**, 911–916
- Shuman, J. D., Vinson, C. R., and McKnight, S. L. (1990) *Science* **249**, 771–774
- Patel, L., Abate, C., and Curran, T. (1990) *Nature* **347**, 572–575
- O'Neil, K. T., Shuman, J. D., Ampe, C., and DeGrado, W. F. (1991) *Biochemistry* **30**, 9030–9034
- Ellenberger, T. E., Brandl, C. J., Struhl, K., and Harrison, S. C. (1992) *Cell* **71**, 1223–1237
- Konig, P., and Richmond, T. J. (1993) *J. Mol. Biol.* **233**, 139–154
- Keller, W., Konig, P., and Richmond, T. J. (1995) *J. Mol. Biol.* **254**, 657–667
- Glover, J. N., and Harrison, S. C. (1995) *Nature* **373**, 257–261
- Schumacher, M. A., Goodman, R. H., and Brennan, R. G. (2000) *J. Biol. Chem.* **275**, 35242–35247
- Fujii, Y., Shimizu, T., Toda, T., Yanagida, M., and Hakoshima, T. (2000) *Nat. Struct. Biol.* **7**, 889–893
- Matthews, B. W. (1968) *J. Mol. Biol.* **33**, 491–497
- Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276**, 307–326
- Navaza, J. (1994) *Acta Crystallogr. Sect. A* **50**, 157–163
- Brünger, 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) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54**, 905–921
- Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. Sect. A* **47**, 110–119
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26**, 283–291
- Lu, X. J., Shakked, Z., and Olson, W. K. (2000) *J. Mol. Biol.* **300**, 819–840
- Johnson, P., and Williams, S. C. (1994) in *Liver Gene Expression* (Yaniv, M., and Tronche, F., eds) pp. 231–258, R. G. Landes Company, Austin, TX
- Osada, S., Yamamoto, H., Nishihara, T., and Imagawa, M. (1996) *J. Biol. Chem.* **271**, 3891–3896
- Johnson, P. F. (1993) *Mol. Cell. Biol.* **13**, 6919–6930
- Falvey, E., Marcacci, L., and Schibler, U. (1996) *Biol. Chem.* **377**, 797–809
- Slomiany, B. A., D'Arigo, K. L., Kelly, M. M., and Kurtz, D. T. (2000) *Mol. Cell. Biol.* **20**, 5986–5997
- Porse, B. T., Pedersen, T. A., Xu, X., Lindberg, B., Wewer, U. M., Friis-Hansen, L., and Nerlov, C. (2001) *Cell* **107**, 247–258
- Johansen, L. M., Iwama, A., Lodie, T. A., Sasaki, K., Felsner, D. W., Golub, T. R., and Tenen, D. G. (2001) *Mol. Cell. Biol.* **21**, 3789–3806
- Suckow, M., von Wilcken-Bergmann, B., and Muller-Hill, B. (1993) *EMBO J.* **12**, 1193–1200
- Tahirov, T. H., Inoue-Bungo, T., Morii, H., Fujikawa, A., Sasaki, M., Kimura, K., Shiina, M., Sato, K., Kumasaka, T., Yamamoto, M., Ishii, S., and Ogata, K. (2001) *Cell* **104**, 755–767
- Tahirov, T. H., Sato, K., Ichikawa-Iwata, E., Sasaki, M., Inoue-Bungo, T., Shiina, M., Kimura, K., Takata, S., Fujikawa, A., Morii, H., Kumasaka, T., Yamamoto, M., Ishii, S., and Ogata, K. (2002) *Cell* **108**, 57–70