Nucleic Acids Research, 1999, Vol. 27, No. 15 **3205–3212**

Specificity of DNA binding of the c-Myc/Max and ARNT/ARNT dimers at the CACGTG recognition site

Hollie I. Swanson* and Jun-Hua Yang

Department of Pharmacology, University of Kentucky, Lexington, KY 40536, USA

Received March 2, 1999; Revised May 26, 1999; Accepted June 4, 1999

ABSTRACT

Basic helix–loop–helix proteins that interact with the DNA recognition site CACGTG include the c-Myc/Max heterodimer and the ARNT (Ah receptor nuclear translocator) homodimer. We have utilized a PCR-based protocol to identify high affinity binding sites of either the c-Myc/Max or ARNT/ARNT dimers and analyzed the ability of these dimers to interact with their derived consensus sequences and activate genes. χ^2 analysis of the selected DNA recognition sites revealed that DNA binding of the ARNT homodimer is symmetric, resulting in the consensus sequence RTCACGTGAY. Gel shift analysis demonstrated that the flanking nucleotides play an important role in dictating DNA binding affinity of the ARNT homodimer. These flanking sequences also regulate the ability of ARNT to competitively displace the c-Myc/Max heterodimer from a CACGTG-containing sequence. However, transient transfection analyses in CV-1 cells revealed that ARNT and c-Myc/Max exhibited similar abilities to activate transcription through each other's consensus sequences. Taken together, these results indicate that although binding affinity of these dimers for the CACGTG core sequences may be differentially influenced by flanking nucleotides, transcriptional activity may also be determined by other factors, such as cellular concentrations of these proteins and their co-activators.

INTRODUCTION

Basic helix–loop–helix (bHLH) proteins are a group of transcription factors that regulate a wide variety of biological processes and include proteins such as c-Myc, Max, E47 and USF (1,2). The basic regions of these proteins contact their DNA recognition sites whereas the helix–loop–helix motifs are involved in dimerization. Additional regions that mediate dimerization and lie adjacent to the helix–loop–helix motifs are present in several classes of proteins within this group and are defined as either leucine zipper or PAS (3) domains.

Characteristics that define the bHLH proteins are that they: (i) often interact with the consensus sequence CANNTG; (ii) form homo- or heterodimeric pairs; (iii) may heterodimerize with multiple partners. For example, the c-Myc/Max heterodimer may interact with either the CATGTG or CACGTG recognition site (4). In addition to heterodimerizing with c-Myc, Max may form a homodimer as well as heterodimerizing with the Mad1–Mad4 proteins and Mnt (5). Each of the c-Myc, Mad and Max protein pairs interact with the DNA recognition sequence CACGTG with apparently similar affinities. The observation that a number of heterodimers, such as E47/E1A, c-Myc/Max and Mad/Max, interact with the CACGTG consensus site has initiated interest in determining whether the nucleotides that flank the CACGTG site play an important role in dictating the specificity and affinity with which these dimeric pairs interact with DNA. For example, it has been observed that the presence of a 5'-T or a 3'-A (e.g. TCACGTGA) inhibits DNA binding of the c-Myc/Max heterodimer, but not that of the Max/Max homodimer (6).

The bHLH/PAS protein family is an emerging group of proteins that is involved in regulating xenobiotic metabolism [the Ah receptor (AHR) and Ah receptor nuclear translocator (ARNT)] (7), the cellular response to low oxygen levels (ARNT and hypoxia inducible factor 1α) (8), neurogenesis [single-minded (SIM) and tango] (9,10) and circadian rhythms (period, Clock and BMAL) (11–14). In addition, several bHLH/PAS proteins act as transcriptional co-activators (e.g. Src1 and GRIP) (15,16). ARNT acts in a manner similar to that of Max in that it homodimerizes to interact with the CACGTG sequence (17,18) and is a common partner within the bHLH/PAS family that heterodimerizes with a number of proteins, including the AHR (19), hypoxia inducible factor 1α (8) and SIM (17). In contrast to that of the bHLH/LZ family, the DNA recognition elements of the bHLH/PAS family are specific for many of the partner pairs. For example, the CACGTG site is recognized by the ARNT homodimer (17,18) and the Clock/BMAL heterodimer (13,14) while the GCGTG site is specified by the AHR/ARNT heterodimer (17,20), the RCGTG site by the ARNT/HIFα heterodimer (21) and the ACGTG site by the ARNT/SIM heterodimer (17,22).

In an effort to understand the parameters that direct DNA binding of the bHLH/PAS proteins, we have studied the DNA binding specificities of the ARNT homodimer (as representative of the bHLH/PAS family) for preferences for the nucleotides that flank the CACGTG site and compared these preferences to that of the c-Myc/Max heterodimer (as representative of the bHLH/LZ family). Our results indicate that the nucleotides that flank the CACGTG site and are specified by the ARNT homodimer are distinct from that of the c-Myc heterodimer, play an important role in modulating DNA binding affinity of the ARNT homodimer and allow the ARNT homodimer to

*To whom correspondence should be addressed. Tel: +1 606 323 1463; Fax: +1 606 323 1981; Email: hswan@pop.uky.edu

competitively displace the c-Myc/Max heterodimer from the CACGTG sequence. However, the ARNT homodimer or the c-Myc/Max heterodimer induced similar levels of reporter activity when interacting with their own consensus or the consensus sequence specific for the other dimer.

These results indicate that differential gene regulation by protein pairs such as the ARNT homodimer and the c-Myc/Max heterodimer likely involves multiple mechanisms that include specific recognition of flanking nucleotides, in addition to variations in their cellular levels and that of their co-activators and co-repressors.

MATERIALS AND METHODS

Oligonucleotides

HIS 8, TCGAGCTCGGTCACGTGACATGCCCAGC; HIS 9, TCGAGCTGGGCATGTCACGTGACCGAGC; HIS 45, GGAAGCTTACGCAGTCACGC; HIS 46, GCGCTCGAGTCCATTGCA; HIS 53, GCACTAGTACCATGAGCGATAACGATGACATCG; HIS 54, CCAAGCTTAGCTGGCCTCCATCCG; HIS 71, CGACGCGTACCATGCCCCTCAACGTTAGCTTC; HIS 72, GCACGCGTTTACGCACAAGAGTTCCGTAGCTG; HIS 101, GGAAGCTTACGCAGTCACGCNNNNNNNCACGTGNNNNNNNTGCAATGGACTCGAGCGC; HIS 108, TCGAGCCTGGGGGCATTGATTGACATAC; HIS 109, TCGAGGTATGTCAATGAATGCCCCCAGC; HIS 128, GATCTTCGGGAGGTCACGTGATTGTGGC; HIS 129, TCGAGCCACAATCACGTGACCTCCCGAA; HIS 138, GATCTTCAGTTCAACACGTGTCATGGGC; HIS 139, CTGAGCCCATGACACGTGTTGAACTGAA; HIS 161, GATCTTCAGTTGACCACGTGGTCTGGGC; HIS 162, CTGAGCCCAGACCACGTGGTCAACTGAA; HIS 163, GATCTTCGGGAGGTAGATCTATTGTGGA; HIS 164, GATCTCCACAATTCTAGAACCTCCCGAA; HIS 165, GATCTTCAGTTCAAAGATCTTCATGGGA; HIS 166, GATCTCCCATGAAGATCTTTGAACTGAA.

Plasmids and antibodies

phuARNT was obtained from Dr Christopher Bradfield (University of Wisconsin, Madison, WI), pMycC92/pGex2T from Dr Robert Eisenman (University of Washington, Seattle WA), pSVMax from Dr Nissam Hay (University of Chicago, Chicago, IL) and CMV-Myc from Dr Richard Pestell (Northwestern University, Chicago, IL). The ARNT antibody was obtained from Dr Richard Pollenz (Medical University of South Carolina, Charleston, SC). The c-Myc antibody (catalog no. sc-788) was purchased from Santa Cruz Biotechnology (Santa Cruz, CA) whereas the Max antibody (catalog no. 06-525) was purchased from Upstate Biotechnology (Lake Placid, NY). The glutathione *S*-transferase antibody was purchased from Pharmacia (Piscataway, NJ) and the non-specific rabbit immunoglobulins were purchased from Sigma (St Louis, MO).

Plasmid construction

Standard reaction mixtures for all PCR experiments were: 10 mM Tris–HCl, 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% gelatin, 200 µM each deoxyribonucleotide triphosphate and 2.5 U *Pfu* polymerase in a total volume of 100 µl. The PCR reactions were generally performed using annealing temperatures that were 4°C below the calculated T_m of the primers. The amplified products were purified following agarose gel electrophoresis (0.8%) and electroelution and subcloned using standard molecular biology procedures. Sequencing was performed using the dideoxy chain termination method (23).

The Max expression construct was generated following amplification using pSVMax as the template, HIS 53 and HIS 54 as the primers and subcloning into the PCR 3.1 vector (Invitrogen, Carlsbad, CA). The luciferase reporter vectors (c-Myc/Max Con1, c-Myc/Max Con2, c-Myc/Max mut, ARNT Con and ARNT mut) were generated by subcloning one copy each of the annealed oligonucleotides HIS 156/157, HIS 161/162, HIS 165/166, HIS 158/159 and HIS 163/164, respectively, into the pGL3 promoter vector (Promega, Madison, WI).

Protein expression

In vitro expression of Max was performed using rabbit reticulocyte lysates (Promega, Madison, WI) as described previously (24). For verification of protein expression, the translation reactions were performed in the presence of [^{35}S]methionine and the products were analyzed by SDS–PAGE. The c-Myc–GST fusion (pMycC92/pGex2T) protein was generated and purified from *Escherichia coli* as follows. One hour following the addition of isopropyl β-D-thiogalactoside (1 mM) the cells were subjected to centrifugation at 6000 *g* for 10 min. The pellet was washed and resuspended in phosphate-buffered saline with 1% Triton X-100. The cells were lysed by sonication for 10 s and the supernatant was recovered following centrifugation at 12 000 *g* for 15 min. A 50% slurry of glutathione–agarose beads (Pharmacia) was added to the supernatant and the mixture rotated for 1 h at 4°C. The beads were collected by centrifugation at 1000 *g* for 10 s, washed twice in PD buffer (50 mM Tris, pH 8.0, 0.1 M KCl, 0.14 M NaCl, 0.5% NP-40 and 10% glycerol) and resuspended in 1.5 ml of PD buffer. To elute the c-Myc–GST fusion protein from the agarose beads, PD buffer containing 50 mM reduced glutathione was added, the mixture was rotated for 20 min at room temperature and the supernatant was removed following centrifugation (1000 *g* for 10 s). Baculovirus expression and purification of the histidine-tagged ARNT was performed as described previously (25). The eluant containing ARNT was dialyzed overnight in MENG buffer (25 mM MOPS, pH 7.5, 1 mM EDTA, 0.02% NaN_3, 20% glycerol, 1 µg/µl leupeptin, 1 mM phenylmethylsulfonyl fluoride). The purified c-Myc and ARNT proteins were quantitated spectrophotometrically.

Gel shift analysis

The DNA probes were radiolabeled with [γ-^{32}P]ATP by end-labeling with T4 polynucleotide kinase (26). The experiments that analyzed DNA binding of the ARNT homodimer were performed as follows. For each gel shift reaction, ~1 pmol of purified ARNT protein was incubated with 1 µl of unprogrammed rabbit reticulocyte lysate for 30 min at 30°C. Non-specific competitor [poly(dI·dC), 200 ng] was added and the KCl concentration was adjusted to 100 mM. After incubating the mixture for 10 min at room temperature, the probe was added (100 000 c.p.m., ~0.5 ng) and the mixture was incubated for an additional 10 min at room temperature. When indicated, either the ARNT or non-specific antibodies were added following the addition of probe and the samples were incubated at room temperature for 10 min. The samples were subjected to 4% acrylamide non-denaturing gel electrophoresis using 0.5× TBE (45 mM Tris base, 45 mM boric acid, 1 mM

EDTA, pH 8.0) as the running buffer (26). Analysis of c-Myc/Max DNA binding was performed as follows. Approximately 20 pmol of the c-Myc–GST fusion protein was incubated with 1 μl of the reaction mixture containing the *in vitro* expressed Max protein for 30 min at 30°C. Nineteen microliters of binding buffer [7.1 mM HEPES, pH 7.0, 3.6 mM $MgCl_2$, 100 mM KCl, 5.7% glycerol, 0.03% NP-40, 2 μg salmon sperm DNA and 100 000 c.p.m. (~0.5 ng) of the probe] were added and the reaction was incubated at room temperature for 20 min. When indicated, the c-Myc, Max, GST or non-specific antibodies were added following addition of the probe and the mixture was incubated for an additional 10 min at room temperature. The samples were subjected to 4% acrylamide non-denaturing gel electrophoresis using 22.5 mM Tris, pH 7.0, and 0.5 mM EDTA as the running buffer. The following annealed oligonucleotides were used either as probes or as competitor DNA: HIS 108/109 (non-specific); HIS 156/157 (c-Myc/Max Con1); HIS 158/159 (ARNT Con); HIS 161/162 (c-Myc/Max Con2); HIS 163/164 (ARNT mut); HIS 165/166 (c-Myc/Max mut).

DNA selection and amplification

The DNA binding site selection and amplification analysis was performed essentially as described previously (17). To generate the oligonucleotide pool, 10 ng of HIS 101 was annealed to a 5-fold molar excess of HIS 45 and the complementary strand was generated following incubation with the Klenow fragment of DNA polymerase for 1 h at 37°C. Approximately 10 ng of the double-stranded oligonucleotide was incubated with mixtures containing either the c-Myc and Max or ARNT proteins and subjected to gel shift analysis. For the first two rounds of selection, the electrophoresis was terminated when the bromophenol dye had migrated 1.5 cm, the upper 1 cm of the gel was excised and the DNA was eluted. The PCR was performed using the eluted oligonucleotide pool as the template and HIS 45 and HIS 46 as the primers. For rounds 3 and 4 (for the ARNT homodimer) or rounds 3–5 (for the c-Myc/Max heterodimer) the oligonucleotide pool was ^{32}P-labeled using PCR. The primers HIS 45 and HIS 46 were first end-labeled using T4 kinase, added to the PCR reactions and amplification was allowed to proceed for 10 cycles. Use of a synthetic double-stranded oligonucleotide containing the CACGTG sequence (HIS 8/9) as a probe and migration marker, as well as the appropriate antibodies, allowed identification of the ARNT- and c-Myc/Max-containing complexes. After a discrete protein/DNA complex could be detected, the oligonucleotide pool was amplified, extracted with phenol:chloroform:isoamyl alcohol, precipitated and subcloned into the pGem-T vector (Promega, Madison, WI). Individual clones were sequenced using the dideoxy chain termination method (23).

Competition curves

To determine the relative affinity of the consensus oligonucleotides, we performed competitive gel shift assays as follows. Increasing concentrations of the indicated competitor DNA was added to the incubation mixture prior to addition of the ^{32}P-labeled probe. Following electrophoresis, the specific protein/DNA complexes were quantitated using phosphorimager analysis. The competitive displacement curves, EC_{50} values and statistical analyses were determined using analysis of one-site competition curves by the GraphPad Prism software (San Diego, CA). The values represent at least two experiments performed in duplicate.

Cell culture and transient transfection analysis

CV-1 cells were maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum, 100 U/ml penicillin and 100 μg/ml streptomycin at 37°C in a humidified 5% CO_2 atmosphere. All transient transfections were performed using the calcium phosphate method (27). The CV-1 cells were co-tranfected with an expression vector (either c-Myc or ARNT) and the appropriate reporter vector (c-Myc/Max Con1, c-Myc/Max Con2, c-Myc/Max mut, ARNT Con or ARNT mut). The cells were harvested in 400 μl lysis buffer (0.25 mM Tris–HCl, pH 7.8) and soluble extracts were prepared following three cycles of freeze/thaw. The cellular pellets were removed upon centrifugation (16 000 *g* at 4°C for 10 min) and the supernatants stored at –80°C until needed for further analysis.

β-Galactosidase and luciferase assays

Aliquots of the soluble extracts (10 μl) were incubated for 30 min to 4 h at 37°C in assay buffer (60 mM Na_2HPO_4, 2 mM $MgCl_2$, 100 mM β-mercaptoethanol, 100 mM β-D-galactopyranosidase, total volume 300 μl). Following the addition of 500 μl of 1 M Na_2CO_3, the absorbance at 420 nm was determined. To determine luciferase activity, a 150 μl aliquot of soluble extract was added to the luciferase assay buffer (0.1 M K_2HPO_4, 0.015 M $MgSO_4$ and 5 mM ATP). Luciferase values were determined following the addition of 0.5 mg of D-luciferin.

Statistical analysis

To determine whether the frequencies at each nucleotide position were statistically different from that expected from random occurrence, the χ^2 goodness of fit test was used (28). One-way ANOVA and the *t*-test for differences among several means was performed to analyze the luciferase values.

RESULTS

Preferences of the ARNT homodimer and the c-Myc/Max heterodimer for nucleotides that flank the CACGTG sequence

The goal of this study was to determine whether a bHLH/PAS protein, such as ARNT, displays preferences for nucleotides that flank the CACGTG sequence that may significantly impact on its DNA binding affinity and gene activation as compared to that of a bHLH/LZ protein, such as c-Myc or Max. To this end, we first employed a PCR-based DNA site selection and amplification protocol that has been used previously to identify high affinity DNA binding sites (6,17,29). A pool of oligonucleotides that contained the CACGTG site and were flanked by seven random nucleotides (Fig. 1A) was selected by the ARNT homodimer and the selected sites were amplified by PCR. After four rounds of selection and amplification, 21 individual oligonucleotides that were selected were cloned and sequenced. Given that the CACGTG core sequence is a palindrome, the appropriate orientation of the sequences was difficult to ascertain. Thus, each sequence was analyzed as a half-site with respect to CAC, as described previously (6). For example, the first sequence was analyzed as two half-sites, AATCGATCAC and TCTCGAACAC. The frequency of each nucleotide was calculated and subjected to χ^2 analysis

A)

N	N	N	N	N	N	N	CACGTG	N	N	N	N	N	N	N
A	A	T	C	G	A	T		T	T	C	G	A	G	A
G	G	C	T	T	G	T		C	G	G	G	T	T	G
A	T	G	A	G	G	T		A	T	G	G	T	G	G
G	G	T	A	A	G	T		A	T	T	A	T	G	C
A	A	C	A	G	A	T		C	A	G	G	C	T	G
T	G	G	A	C	T	A		A	T	T	T	G	T	G
G	T	A	G	T	G	T		A	C	C	A	G	G	G
G	C	A	G	T	G	T		A	G	G	T	T	T	T
G	T	C	T	G	G	T		A	A	G	T	T	G	G
G	G	T	G	T	G	G		T	T	A	G	T	G	G
A	C	C	T	G	G	T		T	T	T	G	C	C	T
A	A	T	A	T	G	G		C	T	A	G	T	G	G
G	G	G	T	G	G	T		A	G	T	T	T	G	C
G	A	G	G	A	G	T		C	T	T	T	T	T	G
T	A	G	C	G	G	G		A	G	T	T	A	G	T
G	G	G	T	G	G	T		A	T	T	A	C	C	G
C	G	G	C	A	G	T		C	T	A	G	T	T	G
G	G	T	A	G	G	T		T	T	A	G	G	T	G
C	C	G	G	G	G	T		A	G	G	T	T	C	G
T	G	T	G	T	G	T		T	T	C	T	G	T	G
G	A	T	G	G	G	T		A	T	C	C	T	G	G

B)

	−7	−6	−5	−4	−3	−2	−1	C	A	C
A	7	14	14	14	10	15	6			
C	16	13	8	12	7	5	0			
G	14	12	11	8	15	19	8			
T	5	3	9	8	10	3	28			

C) CONSENSUS

R T C A C G T G A Y

Figure 1. Analysis of the optimum DNA binding sequence of the ARNT homodimer as determined by the site affinity and amplification assay. (**A**) Twenty-one oligonucleotides that represent high affinity binding sites of the ARNT homodimer were selected from a pool of oligonucleotides containing seven random nucleotides that flank the CACGTG recognition site using the gel shift assay followed by amplification using PCR. After four rounds of selection, the oligonucleotides were subcloned into the pGem-T vector and sequenced. (**B**) The representation of each nucleotide at the indicated positions are expressed as percentages and were evaluated by χ^2 goodness of fit analysis. Those that occurred with a greater than expected frequency (>25%) at $P < 0.01$ are underlined. (**C**) The consensus ARNT DNA binding sequence that was derived from the analysis of nucleotides presented in (B).

A)

N	N	N	N	N	N	N	C	A	C	G	T	G	N	N	N	N	N	N	N
A	G	A	G	G	G	T							C	T	A	T	G	G	G
T	G	G	A	T	A	T							T	G	T	C	A	T	G
T	C	A	G	T	A	C							T	C	A	A	C	G	G
C	T	T	G	G	A	T							G	G	C	T	G	T	G
G	G	T	G	C	A	A							A	C	C	T	G	G	G
A	T	T	T	G	C	A							G	G	A	T	T	T	A
C	G	G	T	C	G	T							A	C	T	C	T	T	G
A	A	G	T	C	T	C							C	T	A	T	G	G	G
T	A	A	C	T	C	T							G	T	T	C	T	C	C
T	G	G	G	T	G	G							T	C	T	G	T	T	G
A	A	T	C	A	C	C							G	T	A	T	G	G	C
A	G	T	T	C	G	G							A	A	C	C	T	G	G
A	G	G	G	T	G	T							A	T	T	C	A	T	A
A	G	C	G	T	A	A							T	A	G	T	T	T	C
A	T	A	T	C	A	C							T	C	C	G	T	A	C
T	C	C	G	T	G	A							G	T	A	C	T	T	G
A	G	G	G	T	G	A							G	T	T	A	G	T	A
C	A	G	A	C	T	G							A	T	C	G	T	C	C
A	G	C	T	G	G	A							G	C	T	G	G	A	C

B)

	−7	−6	−5	−4	−3	−2	−1	C	A	C
A	10	13	13	9	8	14	11			
C	13	8	10	6	7	6	11			
G	7	12	8	15	9	14	5			
T	8	5	7	8	14	4	11			

C) CONSENSUS

R H C A C G T G D Y

Figure 2. Analysis of the optimum DNA binding sequence of the c-Myc/Max heterodimer as determined by the site affinity and amplification assay. (**A**) Nineteen oligonucleotides that represent high affinity binding sites of the c-Myc/Max heterodimer were selected from a pool of oligonucleotides containing seven random nucleotides that flank the CACGTG recognition site using the gel shift assay followed by amplification using PCR. After five rounds of selection, the oligonucleotides were subcloned into the pGem-T vector and sequenced. (**B**) The representation of each nucleotide at the indicated positions are expressed as percentages and were evaluated by χ^2 goodness of fit analysis. Those that occurred with a greater than expected frequency (>25%) at $P < 0.01$ are underlined. (**C**) The consensus c-Myc/Max DNA binding sequence was derived from the analysis of nucleotides presented in (B).

(Fig. 1B). Nucleotides that occurred at a frequency greater than that expected were observed for two of the possible 7 nt (underlined). From this analysis, it is apparent that the interaction of the ARNT/ARNT homodimer with its recognition sequence is symmetric and that the consensus is RTCACGTGAY (Fig. 1C).

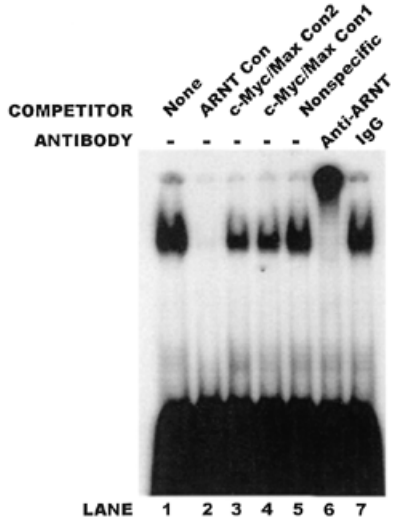
Similar analysis was performed to determine the flanking sequences that are preferred by the c-Myc/Max heterodimer. Nineteen oligonucleotides were selected and sequenced (Fig. 2A). Evaluation of the nucleotide frequency by χ^2 analysis demonstrated that the c-Myc/Max heterodimer exhibited few preferences for the flanking nucleotides (Fig. 2B), resulting in a consensus sequence RHCACGTGDY (Fig. 2C). These results are in contrast to those reported previously which found strong preferences in the –1, –2 and –3 positions, resulting in a consensus GACCACGTGGTC (6). This discrepancy may be due to several differences in methodology. First, the results presented in Figure 2 are representative of only five rounds of selection, whereas those reported by Solomon *et al*. (6) were the result of eight rounds of selection. Increased rounds of selection would theoretically increase the selection of very high affinity binding sites. Second, the composition of the random oligonuclotides that served as the starting material differed. As shown in Figure 2, only the 7 nt that flanked the CACGTG site were random, whereas that used by Solomon *et al*. was composed of 26 random nucleotides that did not include the CACGTG site. We suggest that the presence of the invariant CACGTG site in the oligonucleotides used in this study allowed for the selection of flanking sequences that could compensate for the presence of unfavorable nucleotides at the –1 position (i.e. a T at –1). For example, it has been previously shown that the presence of a G at the –2 position and a C at the +2 position allows high affinity binding of the c-Myc/Max heterodimer despite the presence of the unfavorable T at the –1 position (30).

The ability of the ARNT homodimer and the c-Myc/Max heterodimer to specifically interact with their respective derived consensus sequences was determined using gel shift

A

ARNT Con	GGGAGGT CACGTG ATTGTGG
ARNT mut	GGGAGGT AGATCT ATTGTGG
c-Myc/Max Con1	AGTTGAC CACGTG GTCTGGG
c-Myc/Max Con2	AGTTCAA CACGTG TCATGGG
c-Myc/Max mut	AGTTCAA AGATCT TCATGGG
Nonspecific	GGGGCAT TGATTG ACATACC

B



C

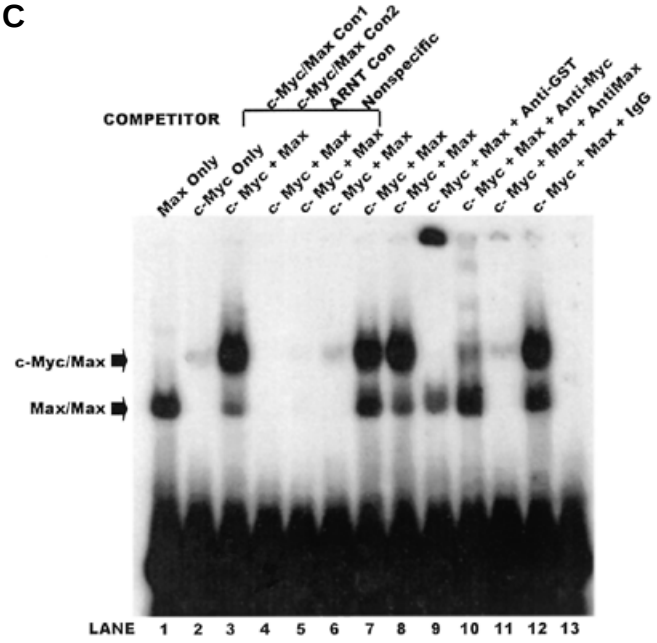


Figure 3. Gel shift analysis of the ARNT homodimer and c-Myc/Max heterodimer using the derived consensus sequences as probes. (**A**) The oligonucleotides that were used as either probes or competitor DNA. (**B**) Gel shift analysis of the ARNT homodimer using ARNT Con as a probe in the absence (lane 1) or presence of the indicated competitor oligonucleotides (lanes 2–5). In lanes 6 and 7, 0.4 μg of either the ARNT antibody or of the non-specific IgG, respectively, were added and the mixture was incubated for 10 min at room temperature prior to gel electrophoresis. (**C**) Gel shift analysis of the c-Myc/Max heterodimer using the derived consensus c-Myc/Max Con1 as probe. The gel shift reactions contained the following: lane 1, Max alone; lane 2, c-Myc alone; lane 3, c-Myc + Max; lanes 4–7, c-Myc + Max with the indicated competitor oligonucleotides; lanes 9–12, c-Myc + Max with the addition of 0.4 μg of the indicated antibodies; lane 13, probe alone. The gel shift assays were performed as described in Materials and Methods.

analysis (Fig. 3). The oligonucleotides used for the radiolabeled probes and competitors are depicted in Figure 3A. The c-Myc/Max Con1 sequence is the consensus sequence derived by Solomon *et al*. (6) whereas the c-Myc/Max Con2 sequence is that shown in Figure 2 and varies at the –3, –1, +1, +2 and +3 positions. The interaction between the ARNT homodimer and its consensus sequence is depicted in Figure 3B. Addition of a 50-fold excess of an oligonucleotide bearing the ARNT consensus sequence eliminated DNA binding of the ARNT homodimer (Fig. 3B, lanes 1 and 2). In contrast, a 50-fold molar excess of either the c-Myc/Max consensus diminished, but did not abolish, the ARNT DNA binding complex (Fig. 3B, lanes 3 and 4). Specificity of the interaction between the ARNT homodimer and the ARNT consensus sequence is illustrated by the inability of an oligonucleotide that lacks the CACGTG binding site to alter migration of the ARNT DNA binding complex and the ability of the ARNT antibody, but not a non-specific IgG, to supershift the ARNT-containing complex (Fig. 3B, lanes 5–7).

Analysis of the interaction between the c-Myc/Max heterodimer and its derived consensus sequence (c-Myc/Max Con2) by the gel shift assay is shown in Figure 3C. Although both the Max homodimer and c-Myc/Max heterodimer interact with c-Myc/Max Con2, the DNA binding complex of the c-Myc/Max heterodimer is distinguished from that of the Max homodimer by its slower rate of migration (Fig. 3C, lanes 1–3). The addition of excess unlabeled oligonucleotides that contained the CACGTG core but differed in the composition of their flanking sequences (TGAC*CACGTG*GTCT, TCAA*CACGTG*TCCAT, or AGGT*CACGTG*ATTG) effectively eliminated DNA binding of the c-Myc/Max complex (Fig. 3C, lanes 4–6). In contrast, the addition of excess unlabeled oligonucleotide that contained the mutated core, TGATTG, did not alter the ability of the c-Myc/Max heterodimer to interact with the radiolabeled probe (Fig. 3C, lane 7). Further evidence that the upper migrating complex is the DNA-bound form of the c-Myc/Max heterodimer is indicated by the ability of either the anti-GST or anti-c-Myc IgGs, but not the anti-Max IgGs, to alter formation of the complex (Fig. 3C, lanes 8–13). In addition, the anti-Max IgGs abolished formation of both the upper and lower migrating complexes. These results indicate that the upper migrating complex represents that of the c-Myc/Max heterodimer whereas the lower migrating complex is that of the Max/Max homodimer.

To determine the relative DNA binding affinities of the ARNT and c-Myc/Max dimers for their derived consensus sequences, competitive gel shift analyses were performed (Fig. 4). As shown in Figure 4A, the relative affinity of the ARNT homodimer for its consensus sequence (ARNT Con) is represented by an EC_{50} value of 1.2×10^{-8} and is significantly higher than that for either of the c-Myc/Max consensus sequences ($EC_{50} = 1.6 \times 10^{-7}$ and 8.5×10^{-8}, respectively). In contrast, the relative affinity of the c-Myc/Max complex for any of the consensus DNA binding sites was virtually indistinguishable (Fig. 4B). The ability of the ARNT homodimer to interact with the ARNT consensus sequence with an affinity that is higher than its affinity for the c-Myc/Max consensus sequence illustrates the fact that the flanking nucleotides that are preferred by the ARNT homodimer are significantly different from those preferred by the c-Myc/Max heterodimer.

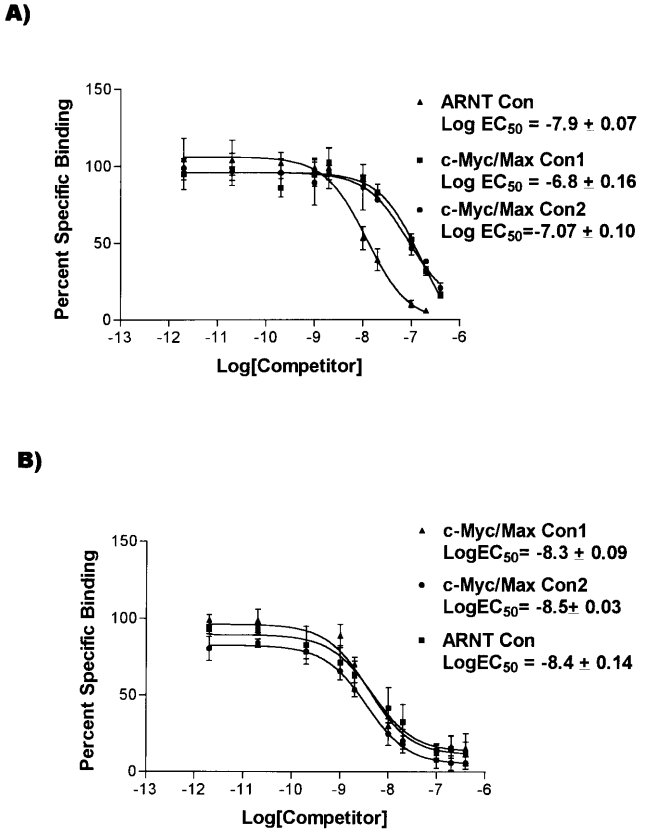


Figure 4. Relative affinity of either the ARNT homodimer (**A**) or the c-Myc/Max heterodimer (**B**) for the derived consensus binding sites. Gel shift analyses of either the ARNT homodimer or the c-Myc/Max heterodimer were performed as described in Figure 3B and C, respectively, in the presence of increasing concentrations of the indicated unlabeled oligonucleotides. The specifically bound complexes were quantitated using phosphorimager analysis. The competitive displacement curves, EC_{50} values and statistical analyses were generated using analysis of one-site competition curves by GraphPad Prism software (San Diego, CA). The values represent at least two experiments performed in duplicate.

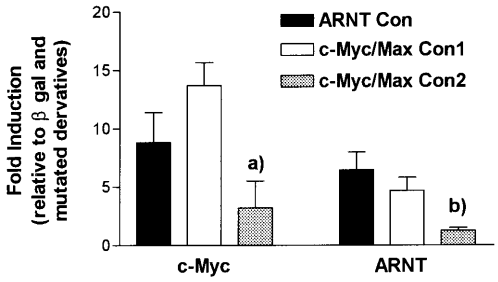


Figure 5. Comparison of the ability of the ARNT homodimer and the c-Myc/Max heterodimer to activate the luciferase reporter gene regulated by either the ARNT Con, c-Myc/Max Con1 or c-Myc/Max Con2 sequences. CV-1 cells were transiently transfected with either the ARNT or c-Myc expression vectors, the β-galactose expression vector and luciferase reporter constructs regulated by either ARNT Con (black bars), ARNT mut, c-Myc/Max Con1 (white bars), c-Myc/Max Con2 (gray bars) or c-Myc/Max mut sequences using the calcium phosphate method as described in Materials and Methods. Fold induction refers to the luciferase/β-galactose values obtained from the reporter containing the consensus sequence divided by the luciferase/β-galactose obtained from the reporter that contained the respective mutated sequence (i.e. either the ARNT mut- or c-Myc/Max mut-containing construct). The values represent three experiments performed in duplicate ± SE. $^{a}P < 0.05$ compared with the ARNT Con or c-Myc/Max Con1 group co-transfected with c-Myc. $^{b}P < 0.05$ compared with the ARNT Con or c-Myc/Max Con1 group co-transfected with ARNT.

Effect of flanking nucleotides on the ability of either the ARNT homodimer or c-Myc/Max heterodimer to activate genes

To determine whether the competitive gel shift analysis shown in Figure 4 accurately predicted the ability of either the ARNT homodimer or the c-Myc/Max heterodimer to activate gene transcription regulated by the derived consensus sequences, we performed transient transfection assays. CV-1 cells were transiently co-transfected with either the ARNT or c-Myc expression vector and luciferase reporter vectors containing either the ARNT Con, ARNT mut, c-Myc/Max Con1, c-Myc/Max Con2 or c-Myc/Max mut sequences. While the gel shift assays demonstrated that ARNT preferentially interacted with its consensus sequence (ARNT Con), gene activation mediated by this interaction was significantly greater than that mediated by its interaction with the c-Myc/Max Con2 sequence, but not the c-Myc/Max Con1 sequence (Fig. 5). Similarly, interaction of the c-Myc/Max heterodimer with the c-Myc/Max Con1 sequence resulted in reporter gene activity that was significantly greater than that mediated by the c-Myc/Max Con2 sequence, but was not significantly different from that of the ARNT Con sequence. Thus, although the ARNT homodimer and the c-Myc/Max heterodimer differentially interact with CACGTG sequences that vary in the context of the flanking sequences (Figs 1–4) and contain the CACGTG core (ARNT Con versus c-Myc/Max Con1), this discrimination does not result in significant differences in their ability to activate genes (Fig. 5).

Interestingly, the c-Myc/Max complex is able to discriminate between recognition sites that vary in the flanking nucleotides when gene activation (Fig. 5), but not DNA binding (Fig. 4), is used as the end-point. This apparent discrepancy raises several possibilities, including: (i) DNA binding of the c-Myc–GST fusion protein may not accurately predict DNA binding of the c-Myc protein expressed *in vivo*; (ii) additional proteins that are not present in the gel shift reactions may play a role in recognition of the flanking sequences.

The ability of the ARNT homodimer to displace the c-Myc/Max heterodimer from the CACGTG recognition site is dependent on the context of the flanking sequences

To determine whether the ARNT homodimer may displace the c-Myc/Max heterodimer from CACGTG-containing recognition sites, we performed gel shift analysis using constant amounts of c-Myc and Max while increasing the concentrations of ARNT. As shown in Figure 6A, excess molar amounts of ARNT failed to displace the c-Myc/Max heterodimer from interacting with the c-Myc/Max Con2 sequence. In contrast, the ARNT homodimer competitively displaced the c-Myc/Max heterodimer from its interaction with the ARNT Con sequence (Fig. 6B), indicating that the ARNT homodimer may competitively displace the c-Myc/Max heterodimer only from CACGTG sequences that contain its preferred flanking nucleotides.

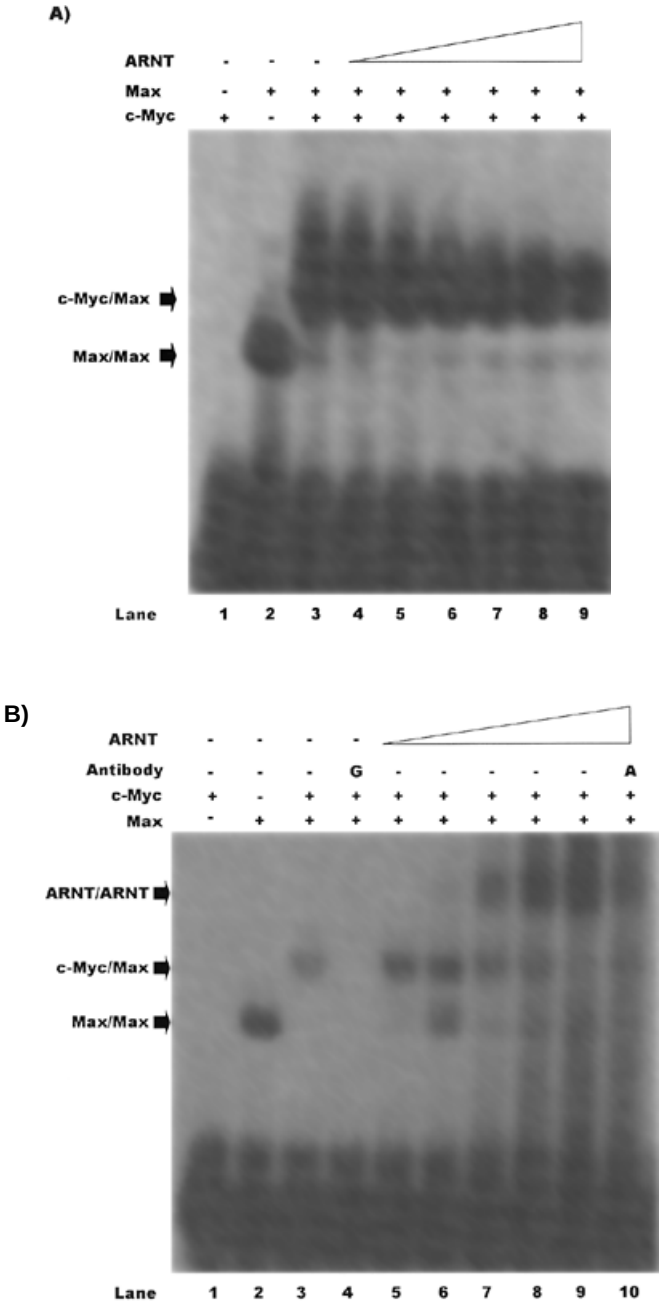


Figure 6. The ability of the ARNT homodimer to displace the c-Myc/Max heterodimer from the CACGTG sequence is dependent upon the context of the flanking sequences. Gel shift analysis was performed using constant amounts of the c-Myc/Max heterodimer in the presence of increasing concentrations of ARNT as described in Materials and Methods. (**A**) Gel shift analysis using c-Myc/Max Con2 as probe. Lane 1, 20 pmol of c-Myc with 1 µl unprogrammed reticulocyte lysate; lane 2, 1 µl of *in vitro* expressed Max alone; lanes 3–9, 20 pmol of c-Myc + 1 µl Max in the absence (lane 3) or presence (lanes 4–9) of increasing concentrations of ARNT (1.3–26 pmol). (**B**) Gel shift analysis using ARNT Con as the probe. Lane 1, 40 pmol of c-Myc and 1 µl of unprogrammed reticulocyte lysate; lane 2, 1 µl of *in vitro* expressed Max; lane 3, 40 pmol of c-Myc and 1 µl of *in vitro* expressed Max alone; lane 4, c-Myc and Max + the GST antibody; lanes 5–9, 40 pmol of c-Myc + 1 µl of Max in the presence of increasing concentrations of ARNT (1.3–26 pmol); lane 10, 40 pmol of c-Myc + 1 µl of Max + 6.5 pmol ARNT + the ARNT antibody. G, the anti-GST IgGs; A, the anti-ARNT IgGs.

DISCUSSION

In an effort to understand how transcription factors that represent distinct protein families interact with apparently identical core DNA recognition sequences yet regulate disparate signaling pathways, we have examined the role of flanking nucleotides in directing these proteins to their appropriate gene targets. The protein families examined were the bHLH/LZ family, represented by the c-Myc/Max heterodimer, and the bHLH/PAS family, represented by the ARNT homodimer. Since the critical amino acids within the basic regions of c-Myc, Max and ARNT that contact the core CACGTG site are very similar (17,31), we questioned whether additional amino acid/nucleotide contacts provided by flanking nucleotides may play a role in the DNA binding specificities and recognition of these proteins.

c-Myc, Max and ARNT have been assigned to Class B of the bHLH proteins based on their DNA recognition half-sites and the amino acids that lie within their basic regions that contact these sites (17). The amino acids that contact the CAC half-site have been defined by crystallization and site-directed mutagenesis as E-R-R-R/Q-R (17,32,33). Crystallization studies of USF, E47, PHO4 and Max predict that the glutamate residues of c-Myc and ARNT interact with the CA of the CAC half-site, while the terminal arginine residue (E-R-R-R/Q-R) contacts the C of the CAC half-site (33). The amino acids that are involved in contacts with nucleotides that flank the CACGTG site are localized outside the I-E-R-R-R-R motif, yet their utilization is protein specific. For example, within the PHO4 protein, an arginine and a histidine residue make contacts with the neighboring guanine residues (CACGTGGG) (34). Although these residues are conserved within other bHLH class B proteins, such as E47, USF and Max, they do not appear to play important roles in contacting flanking nucleotides (33,35,36). In contrast, amino acids contained within the loop of the helix–loop–helix motif of Max appear to make limited contacts with neighboring nucleotides (33). From the crystallization data generated thus far, it appears that the overall conformation of the DNA binding protein may dictate the specific contacts made with nucleotides that flank the CACGTG sequence. Thus, the chosen partner of a particular protein may induce subtle changes in the DNA binding form of that protein resulting in distinct preferences for flanking nucleotides for each protein pair. The data in the present study indicate that amino acids that lie outside the I-E-R-R-R-R motif of ARNT may play a role in specifying contacts with nucleotides that flank the CACGTG site.

Several mechanisms that dictate specificity of regulation at the CACGTG recognition site have been proposed. First, nucleotides that flank the CACGTG site have been shown to play an important role in discriminating between different protein pairs for DNA binding. As shown in Table 1, the preferences for flanking nucleotides of proteins that interact with the CACGTG recognition site (c-Myc, Max, USF and ARNT) are distinct. The importance of preferences for flanking nucleotides in regulating endogenous genes has been demonstrated using the promoter elements of the *cad* (carbamoyl phosphate synthase/aspartate carbamoyltransferase/dihydroorotase) gene (37). Activation of the cad promoter was observed when the CACGTG element was flanked by nucleotides that allowed a high affinity interaction between the CACGTG site and the c-Myc/Max heterodimer, but not the USF homodimer, during the

S phase of the cell cycle. Similarly, our data (Fig. 6) indicate that the ARNT homodimer may be capable of displacing protein pairs, such as the c-Myc/Max heterodimer, from the CACGTG sequence only when the flanking nucleotides conform to those preferred by the ARNT homodimer. Second, the distance of the CACGTG element from the transcription start site of the regulated gene appears to play an important role. For example, the c-Myc/Max complex, but not USF, transactivates the rat prothymosin-α gene when the CACGTG site is at its endogenous position (1.6 kb downstream of the transcription start site) while activation by USF is achieved only when the CACGTG site is placed <400 bp from the transcription start site (38). Thus, multiple mechanisms are involved in dictating specificity of gene regulation by a number of bHLH proteins.

Table 1. Comparison of flanking nucleotide preferences between Class B bHLH proteins

Protein pair	Consensus sequence	Reference
Max/Max	RANCACGTGNTY	6
c-Myc/Max	-ATCACGTGGG-	6
USF/USF	GACCACGTGGTC	30
ARNT/ARNT	RTCACGTGAY	Figure 1

The work presented in this study has shown that nucleotides that flank the CACGTG sequence play an important role in facilitating high affinity DNA binding of the ARNT homodimer and that the ARNT homodimer can differentially interact with sites that vary in flanking nucleotide composition (Figs 1–4). However, transient transfection studies failed to demonstrate a significant difference in the ability of c-Myc and ARNT to discriminately activate transcription through their consensus sequences (Fig. 5). These results indicate that in addition to DNA binding affinity, other factors play an important role. For example, the selectivity of the protein/DNA interactions may be compromised by the high protein levels of either c-Myc or ARNT that are present in the transient transfection assays. In addition, the co-activators and co-repressors that modulate transcriptional activity of the c-Myc/Max or ARNT dimers may be distinct and may vary in their functional activity. In summary, we have shown that the bHLH/PAS protein ARNT preferentially interacts with flanking sequences that are distinct from that of the c-Myc/Max heterodimer and that although flanking sequences play an important role in determining specificity of gene regulation, other factors such as distance of the enhancer from the promoter and relative proteins levels within the cell may also be critical.

ACKNOWLEDGEMENTS

This work was supported by the University of Kentucky Research Fund (grant 847) and the National Institutes of Health (grant ES 08088).

REFERENCES

1. Phillips,S.E.V. (1994) *Curr. Biol.*, **2**, 1–4.
2. Ellenberg,T. (1994) *Curr. Opin. Struct. Biol.*, **4**, 12–21.
3. Huang,Z.J., Edery,I. and Rosbash,M. (1993) *Nature*, **364**, 259–262.
4. Blackwell,T.K., Huang,J., Ma,A., Kretzner,L., Alt,F.W., Eisenman,R.N. and Weintraub,H. (1993) *Mol. Cell. Biol.*, **13**, 5216–5224.
5. Sommer,A., Bousset,K., Kremmer,E., Austen,M. and Luscher,B. (1998) *J. Biol. Chem.*, **273**, 6632–6642.
6. Solomon,D.L.C., Amati,B. and Land,H. (1993) *Nucleic Acids Res.*, **21**, 5372–5376.
7. Swanson H.I. and Bradfield,C.A. (1993) *Pharmacogenetics*, **3**, 213–230.
8. Wang,G.L., Jiang,B.-H., Rue,E.A. and Semenza,G. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 5510–5514.
9. Muralidhar,M.G., Callahan,C.A. and Thomas,J.B. (1993) *Mech. Dev.*, **41**, 129–138.
10. Sonnenfeld,M., Ward,M., Nystrom,G., Mosher,J., Stahl,S. and Crews,S. (1997) *Development*, **124**, 4583–4594.
11. Zeng,H., Hardin,P.E. and Rosbash,M. (1994) *EMBO J.*, **13**, 3590–3598.
12. King,D.P., Zhao,Y., Sangoram,A.M., Wilsbacher,L.D., Tanaka,M., Antoch,M.P., Steeves,T.D.L., Vitaterna,M.H., Kornhauser,J.M., Lowrey,P.L., Turek,F.W. and Takahashi,J.S. (1997) *Cell*, **89**, 641–653.
13. Gekakis,N., Stakni,D., Nguyen,H.B., Davis,F.C., Welsbacher,L.D., King,D.P., Takashashi,J.S. and Weitz,C.J. (1998) *Science*, **280**, 1564–1569.
14. Hogenesch,J.B., Gu,Y.-Z., Jain,S. and Bradfield,C.A. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 5474–5479.
15. Yao,T.-P., Ku,G., Zhou,N., Scully,R. and Livingston,D.M. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 10626–10631.
16. Hong,H., Koh,K., Garabedian,M.J. and Stallcup,M.R. (1997) *Mol. Cell. Biol.*, **17**, 2735–2744.
17. Swanson,H.I., Chan,W.K. and Bradfield,C.A. (1995) *J. Biol. Chem.*, **270**, 26292–26302.
18. Sogawa,K., Nakano,R., Kobayashi,A., Kikuchi,Y., Ohe,N., Matshushita,N. and Fuji-Kuriyama,Y. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 1936–1940.
19. Reyes,H., Reisz-Porszasz,S. and Hankinson,O. (1992) *Science*, **256**, 1193–1195.
20. Denison,M.S., Fisher,J.M. and Whitlock,J.P.,Jr, (1989) *J. Biol. Chem.*, **264**, 16478–16482.
21. Semenza,G.L., Jiang,B.-H., Leung,S.W., Passantino,R., Concordet,J.-P., Maire,P. and Giallongo,A. (1996) *J. Biol. Chem.*, **271**, 32529–32537.
22. Wharton,K.A.,Jr, Franks,R.G., Kasai,Y. and Crews,S.T. (1994) *Development*, **120**, 3563–3569.
23. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.
24. Dolwick,K.M., Swanson,H.I. and Bradfield,C.A. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 8566–8570.
25. Chan,W.K., Chu,R., Jain,S., Reddy,J.K. and Bradfield,C.A. (1994) *J. Biol. Chem.*, **269**, 26464–26471.
26. Garabedian,M.J., LaBaer,J., Liu,W.-H. and Thomas,J.R. (1993) In Hames,B.D. and Higgins,S.J. (eds), *Gene Transcription: A Practical Approach*. Oxford University Press, Oxford, UK, pp. 243–293.
27. Docherty,K. and Clark,A.R. (1993) In Hames,B.D. and Higgins,S.J. (eds), *Gene Transcription: A Practical Approach*. Oxford University Press, Oxford, UK, pp. 71–73.
28. Milton,J.S. and Tsokos,J.O. (1983) In Corrigan,J.J. and Amar,J.S. (eds), *Statistical Methods in the Biological and Health Sciences*. McGraw-Hill, New York, NY, pp. 33–37.
29. Blackwell,T.K., Kretzner,L., Blackwood,E.M., Eisenman,R.N. and Weintraub,H. (1990) *Science*, **250**, 1149–1151.
30. Bendall,A.J. and Molloy,P.L. (1994) *Nucleic Acids Res.*, **22**, 2801–2810.
31. Swanson,H.I. and Yang,J.-H. (1996) *J. Biol. Chem.*, **271**, 31657–31665.
32. Halazonetis,T. and Kandil,A.N. (1998) *Science*, **255**, 464–466.
33. Ferre-D'Amare',A.R., Prendergast,G.C., Ziff,E.B. and Burley,S.K. (1993) *Nature*, **363**, 38–44.
34. Shimizu,T., Toumoto,A., Ihara,K., Shimizu,M., Kyogoku,Y., Ogawa,N., Oshima,Y. and Hakoshima,T. (1997) *EMBO J.*, **16**, 4689–4697.
35. Ellenberger,T., Fass,D., Arnadu,M. and Harrison,S.C. (1994) *Genes Dev.*, **8**, 970–980.
36. Ferre-D'Amare',A.R., Pognonec,P., Roeder,R.G. and Burley,S.K. (1994) *EMBO J.*, **13**, 180–189.
37. Boyd,K.E. and Farnham,P.J. (1997) *Mol. Cell. Biol.*, **17**, 2529–2537.
38. Desbarats,L., Gaubatz,S. and Eilers,M. (1996) *Genes Dev.*, **10**, 447–460.