Nucleic Acids Research, 1998, Vol. 26, No. 5 **1135–1143**

A survey of 178 NF-Y binding CCAAT boxes

Roberto Mantovani*

Dipartimento di Genetica e Biologia dei Microrganismi, Università di Milano, Via Celoria 26, 20133 Milano, Italy

Received October 23, 1997; Revised and Accepted December 2, 1997

ABSTRACT

The CCAAT box is one of the most common elements in eukaryotic promoters, found in the forward or reverse orientation. Among the various DNA binding proteins that interact with this sequence, only NF-Y (CBF, HAP2/3/4/5) has been shown to absolutely require all 5 nt. Analysis of a database with 178 *bona fide* NF-Y binding sites in 96 unrelated promoters confirms this need and points to specific additional flanking nucleotides (C, Pu, Pu on the 5′-side and C/G, A/G, G,A/C, G on the 3′-side) required for efficient binding. The frequency of CCAAT boxes appears to be relatively higher in TATA-less promoters, particularly in the reverse ATTGG orientation. In TATA-containing promoters the CCAAT box is preferentially located in the –80/–100 region (mean position –89) and is not found nearer to the Start site than –50. In TATA-less promoters it is usually closer to the +1 signal (at –66 on average) and is sometimes present in proximity to the Cap site. The consensus and location of NF-Y binding sites parallel almost perfectly a previous general statistical study on CCAAT boxes in 502 unrelated promoters. This is an indication that NF-Y is the major, if not the sole, CCAAT box recognizing protein and that it might serve different roles in TATA-containing and TATA-less promoters.

CCAAT BOXES AND CCAAT BOX BINDING PROTEINS

Regulation of transcription is a complex set of events controlled by DNA sequences positioned in proximity to the genes (promoters) and by elements acting at a distance (enhancers) (1). Promoters and enhancers that activate polymerase II transcribed mRNA genes are formed by a combinatorial puzzle of short sequences recognized by sequence-specific regulators. Some, such as the TATA, GC and CCAAT boxes, are encountered at extremely high frequency (2,3). The CCAAT box was one of the first elements identified (4,5). Later studies clearly established that such pentanucleotide sequences are present in a wide variety of vertebrate, yeast and plant promoters and are important for transcription. Performing a statistical analysis on a compilation of over 500 unrelated promoters Bucher established that the CCAAT pentanucleotide is present in ~30% of them. They are identified by highly preferred sequences on both the 5′ and 3′ flanking sides and are most frequently located in the –60/–100 region (3). However, from this study, it was not possible to identify the protein binding and activating ‘CCAAT consensus’. Over the last 15 years, with the parallel discovery of functionally important CCAAT boxes in different promoters, a plethora of CCAAT-interacting polypeptides have been detected, mainly by means of EMSA and footprinting assays. In many cases such activitities were purified and the corresponding genes cloned.

c/EBP (CCAAT/enhancer binding protein) was identified as the activator of two functionally important but apparently unrelated elements in the TK promoter and SV40 enhancer. Cloning of the genes revealed the presence of B-Zip dimerization and DNA binding domains (6,7). The binding sites of C/EBP are composed of palindromic repeats, occasionally containing a CCAAT pentanucleotide in the intervening sequence (8).

CTF/NF-I (CCAAT transcription factor) also binds as a dimer to viral and cellular promoters (9), recognizing a TGG(N)$_6$GCCAA sequence (10). A T after CCAA is sometimes present, but not strictly necessary, and binding requirements are centred on the two half palindromes, as confirmed by site selection, saturation mutagenesis and methylation interference (10,11).

Y Box factors, cloned by screening expression libraries with an MHC class II Y box oligo (12), were later shown to contain a nucleic acid interacting protein domain also shared by bacterial proteins (13). The binding specificity of such proteins is very large and includes single-stranded DNA, abasic DNA, CT-rich sequences and class III promoter elements. Moreover, they have also been involved in the control of translation (13).

CDP (CCAAT displacement protein) was identified as a binding activity recognizing a large piece of the sea urchin histone H2B and human γ-globin promoters, both encompassing two CCAAT boxes (14,15). The gene contains three repeats homologous to the *Drosophila* CUT homeodomain (16), each of which has a slightly different binding specificity, with the CCAAT sequence being necessary only for CR1 (16,17).

HSP-CBF was cloned by screening expression libraries with an HSP70 CCAAT oligo (18).

H1TF2A has been purified by affinity chromatography with a histone H1 CCAAT box. It is a multimeric protein and one of the genes, H1TF2A, has been cloned and shown to have some similarity to the Q-rich domain of NF-YA (19,20). The sequence specifities of the two latter factors are not well defined.

NF-Y (also called CBF, α-CP1 and CP1) was first identified as the activity binding to the MHC class II conserved Y box (21). Saturation mutagenensis studies performed in different laboratories clearly showed an almost absolute requirement for each of the CCAAT nucleotides (21–24). It was purified independently using

*Tel: +39 2 26605239; Fax: +39 2 2664551; Email: mantor@imiucca.csi.unimi.it

affinity columns containing the Y box, the α1(I) collagen CCAAT and the α-globin CCAAT (22,23,25), while conventional chromatography was used to purify the yeast complex, involved in activation of cytocrome genes (26). Recently a similar activity was purified from *Neurospora crassa* (27). NF-Y is a ubiquitous heteromeric protein composed of three subunits, NF-YA, NF-YB and NF-YC, all necessary for DNA binding (28). The mammalian and yeast genes have been cloned (25,28–37). NF-Y sequences are available from several other species: the NF-YA gene was cloned from *Schizosaccharomyces pombe*, *Brassica napus*, *Schistosoma mansoni* and sea urchin (38–41); NF-YB from *Kluyveromyces lactis*, *Aspergillus nidulans*, *Zea mays*, lamprey, *Xenopus* and chicken (31,42,43). Each of the three subunits displayed highly conserved domains. The NF-YA homology domain can be sharply divided into subunit association and DNA contacting subdomains (34,44–46). NF-YB and NF-YC tight association is a prerequisite for NF-YA binding and sequence-specific DNA interactions (28). Both NF-YB and NF-YC conserved domains contain putative histone fold motifs (47). This motif, common to all core histones, is composed of three α-helices separated by short loop/strand regions. Enabling histones to dimerize with companion subunits, this motif is responsible for formation of the histone octamer. Recent experiments on yeast HAP3 (34), CBF-A/NF-YB (48) and CBF-C/NF-YC (49) indicate that this 65 amino acid long motif is necessary for subunit interactions and DNA binding.

NF-Y CONSENSUS DERIVED FROM NATURAL PROMOTERS

Because of the apparent multiplicity of CCAAT binding proteins, a crucial question in understanding how CCAAT sequences activate transcription is which of the activators shown to recognize this box is actually operating on a given promoter. To this end, we and others developed NF-Y reagents, such as monoclonal and polyclonal antibodies and dominant negative vectors (46,50), which were employed by many laboratories in EMSA supershift experiments to unambiguously identify NF-Y as the DNA–protein complex generated with CCAAT containing fragments from different promoters.

From this large body of information a database comprising all promoters shown to contain a *bona fide* NF-Y binding site was organized. The following criteria were used. (i) EMSA competition experiments with the original Ea Y box oligo or with other *bona fide* high affinity NF-Y binding sites such as α1(I) collagen, albumin, α-globin, RSV or MLP (21,22,51). The albumin promoter was shown to function through a NF-Y binding CCAAT box (50); a trimeric complex was purified to homogeneity on α-globin CCAAT affinity columns; subunit composition and sequence specificities are indistinguishable from NF-Y (23); the RSV and MLP CCAAT proteins also shared these features (51,52). (ii) Direct supershift experiments with anti-NF-Y antibodies. (iii) Promoters of the same gene from different species, one of which harbours a NF-Y site, were included; in all cases so far tested [MHC class II, γ-globin, α1 (I) collagen, albumin, MDR1, topoisomerase IIα] this assumption was in fact formally proven. (iv) The list of yeast CCAAT-containing promoters is based upon dependence of the HAP2/3/4/5 complex.

A total of 178 *bona fide* NF-Y binding sites, 164 in promoters of higher eukaryotes (mainly human, rodent, chicken and *Xenopus*) is presented in Table 1. Information regarding the position of the CCAAT box with respect to the +1 signal, the orientation of the CCAAT sequence, the presence in the promoter of a recognizable TATA box, the role in transcriptional activation, the proximity of binding sites for other transcription factors and the tissue distribution of the gene are also presented. In general the vast majority of these CCAAT boxes have invariably been shown to significantly contribute to overall promoter strength and, indeed, sometimes to be strictly required for activity.

Inspection of sequence alignment of all these CCAAT boxes defines the consensus for NF-Y, as can be seen from Table 2. In general, although the number of lower eukaryote sequences is low, the NF-Y and HAP2/3/4/5 consensuses are very similar.

The CCAAT pentanucleotide

All five core nucleotides are almost invariably conserved. The rare exceptions concern position +2 (an adenine in the globin ρ enhancer and Apo-A-I and a G in the Ii promoter), position +3 (a T in CDC25 and Dpa and a G in Factor VIII) and position +5 (C in MVM P4 and yeast CYC1 and CYT1), while positions +1 and +4 are totally conserved. When measured, such as for Ii, MVM P4 and CYC1, the affinity of the variant CCAAT boxes was lower than for the intact sequence, highlighting the importance of all 5 nt.

The flanking sequences

At positions –1 and –2 there is a clear preference for purines: adenines are slightly more abundant at –2 and guanines at –1. Note that some of the high affinity NF-Y binding sites, such as α1(I) collagen and RSV, contain a C at position –2. At –3 adenines are under-represented (<10%), while C residues are more abundant (>40%). Indeed, a G→C mutation at this position in the albumin promoter increases both NF-Y binding and promoter activity. No obvious skewing is seen beyond this position. At the 3′-end guanines are well represented at positions +6/+7 and predominate at +8, but a clear preference (>50%) is given to C residues at +6 and A residues at +7. A C→A mutation at +6 and an A→C at +7 severely affects NF-Y binding to the Y box (21), while a G→C mutation at +6 of the albumin promoter increases NF-Y binding and transcriptional activity, as one would expect from the consensus. Position +10 shows several G residues and very few C residues. Finally, T residues are seldom found in close proximity to CCAAT, at positions –1/–2 and from +6 to +9. Overall, these data fit very well with methylation interference patterns (see 53). The optimal binding site encompasses 13 nt (3 nt at the 5′-end and 5 nt at the 3′-end) and thus is slightly over one turn of the double helix and it is devoid of any recognizable symmetry axis.

Further confirmation of this analysis comes from sequences that contain an intact CCAAT pentanucleotide yet bind NF-Y very inefficiently or not at all. Inspection of such sites (see Table 1) reinforces the importance of flanking nucleotides, both at the 5′- and 3′-ends: in β- and ε-globin, in the proximal site of gp91phox, in hamster topoisomerase IIα site V and in Fcγ receptor 1. The 5′-ends are in accordance with the consensus, except for an adenine in position –3 of the β-globin site. T residues are present at different positions between +6 and +8. The IL4 proximal, human topoisomerase IIα site V and carboxyesterase sites harbour T residues at –2 and at +7 and +6 respectively. The C residues at –2, –1 and +7 of the CD14 sites are probably responsible for its negligible affinity.

Comparison of the NF-Y consensus with the CCAAT Bucher consensus, statistically derived from random analysis of the most

Table 1.

NF-Y BINDING SITES

GENE		SEQUENCE	ORG	OR	POS	TATA	ACT	TF	COMP	AB	EXPR	REF
MHC II	Ea	TTTAACCAATCAGAAA	Mm	<	-54	-	Yes	RF-X	+	+	BlMo	56
	EaY'	CAGAACCAATCAGCAG	Mm	>	-1386	-	Yes	RF-X	+			57
	Dra	TTTGGCCAATCAGAAA	Hs	<	-70	+	Yes	RF-X	+	+		56
	DraY'	AAGAACCAATCAGTGT	Hs	>	-1300	+	ND	RF-X	+	+		58
	Aa	CAGAACCAATCAGAAA	Mm	<	-44	-	Yes	RF-X	+			59
	Dqa	TTTGGCCAATTAGAAA	Hs	<	-72	-	Yes	RF-X	+			60
	Dpa	ATTCACCTATCAGAGA	Hs	<	-59	-		RF-X	+			60
	Eb	GGGAGCCAATCAGCAT	Mm	<	-68	-	Yes	RF-X	+			59
	Drb	GAGAACCAATCAGCAT	Hs	<		-	ND	RF-X	+			60
	Ab	AGGAACCAATCAGCAT	Mm	<	-63	-	Yes	RF-X	+			59
	Dqb	AGGAACCAATCAGCAT	Hs	<	-75	-	Yes	RF-X	+			60
	Dpb	AAGAACCAATGGACAC	Hs	<	-80	-		RF-X	+			60
	Dpb	AAGAACCAATGGGCAT	Sp	<		-		RF-X	+			61
	B-Lb	GAGGGCCAATGAGCGG	Gg	<			ND	RF-X				
Ii	P	GTGGACGAATCAGATT	Hs	>	-49	-	Yes	Sp1	+	+	BlMo	62
	D	GGCAGCCAATGGGATC	Hs	>	-203	-	Yes	RF-X	+			
		GCCAGCCAATGGGATC	Mm	>	-200	-	Yes	RF-X	+			
Mig		ACCAGCCAATCAGAGA	Mm	>	-62	+	ND		+	+	Mo	63,165
GP91-phox		GTTGACCAATGATTAT	Hs	>	-122	+	Yes		+	+	Mo	64
CD10		CCCGACCAATGAGCGC	Hs	<	-125	-	Yes		+	+	TlBlGr	65
RAG-1		GATAGCCAATCACAGA	Mm	<	-95	+	Yes	Ebox	+	+	TlBl	66
IL4	D	CTGGGCCAATCAGCAC	Hs	<	-178	-	Yes		+	+	Tl	67
	P	TCAGACCAATAGGAAA	Hs	<	-110	-	Yes	NF-AT	+	+		
		TGGGGCCAATCAGCAC	Bt	<	-110	+						68
Thy-1		TTCAGCCAATCGGAGG	Mm	<	-79	-	Yes	SP1	+		Tl	69
Globin α		ACCAGCCAATGAGTAA	Mm	>	-87	+	Yes	a-IRP	+		Ery	23
		GCCAGCCAATGAGCGC	Hs	>	-70	+	ND	a-IRP				70
		GCCGGCCAATGAGCGG	Sp	>	-71	+	ND	a-IRP	+			
		CCAGGCCAATGATTAC	Xl	>	-90	+	ND					71
		CCAGTCCAATGGCTAC	Xt	>	-92	+	ND					71
	ζ	CCTGACCAATGGCCAC	Hs	>	-62	+	Yes	Gata	+	+		72
	γD	CTTGACCAATAGCCTT	Hs	>	-113	+	Yes		+	+		72
		CTTGACCAATAGGCTT	Gc	>	-120	+			+	+		73
		CTTGACCAATAGTCGT	Sp	>	-120	+			+	+		74
	γP	CTTGACCAATAGTCTT	Hs	>	-86	+	Yes		+	+		73
		CTTGACCAATAGCCTC	Gc	>	-90	+			+	+		73
		ACTGACCAATAGCCTC	Sp	>	-90	+						74
	ρ3'E	ACCAGCAAATGGCATT	Gg	>	+2018	+	Yes	YY1	+	+		75
Coll	α2(I)	CTCCACCAATGGGAGG	Mm	<	-82	+	Yes		+	+	BoSk	76
		CTCCACCAATGGGAGG	Hs	<	-80	+	Yes		+	+		77
	α1(I)	CCCAGCCAATCAGAGC	Mm	<	-96	+	Yes		+	+		78
Osteopontin		CTCCACCAATCAGCAC	Mm	<	-51	+	Yes	AP1	+	+	Bo	79
BSP		AGCAGCCAATCACGGT	Rn	<	-48	+	ND		+	+	Bo	80
Albumin		AGGAACCAATGAAATG	Rn	>	-81	+	Yes	cEBP	+	+	Li	81
		AGGAACCAATGAAATG	Mm	>	-86	+	Yes	cEBP				82
		GGCAGCCAATGAAATA	Hs	>	-81	+	Yes	cEBP	+			83
		GAAAACCAATATAGAG	Xl	>	-164	+						84
ApoA-I		CTGGGCAAATAGAGTC	Hs	<	-156	+	ND		+	+	Li	85,86
Aldolase B		ACGCGCCAATCAGAGT	Rn	>	-125	+	Yes	cEBP	+	+	Li	87
		ATGGGCCAATCAGAGG	Hs	>	-113	+						
		AGCAGCCAATCAGCTA	Gg	>	-119	+						
TAT	P	AGACCAATAAAGTT	Rn	>	-73		Yes		+		Li	88
	D	CTCAACCAATAGCACG	Rn	>	-285			HNF1	+			
γ-GT		ACGATCCAATCCTCTC	Rn	>	-107	-	ND	SP1	+		LiKi	89
SDH		GGCAGCCAATGAGGGC	Rn	<	-51	+	ND		+		Li	90
Fibronectin		CCGGGCCAATCGGGCG	Hs	>	-149	+	Yes	ATF	+	+	Li	91
Arg Lyase		TAGAACCAATTGGGAG	Rn	<	-81	+	Yes		+	+	Li	92
		CGCGCCAATAGGAGG	Hs	<	-91	-						
Factor VIII		AGTAACCGATAGGATT	Hs	<	-18	+	Yes	cEBP	+	+	LiSpLy	93
Factor X		CGGCTCCAATCAGGAG	Hs	>	-118	+	Yes	SP1	+	+	Li	94
MSP		GCCACCCAATCCCGTA	Hs	>	-26	-	Yes		+	+	LiLu	95

frequent sequences in 502 unrelated promoters, show a compelling degree of similarity. The purine preference at the 5′-end (A residues at –2 and G residues at –1), the high numbers of C residues at +6 and of A residues at +7, the equal presence of G residues at these positions, the notable absence of T residues at –2/–1/+6/+7, are all features that perfectly parallel the numbers observed for NF-Y binding CCAAT boxes (see Table 2). We note, however, two differences, a slight over-representation of T residues at –3 (C residues are more numerous in NF-Y binding sites) and a relative variance at +4, a position highly conserved for NF-Y. The latter discrepancy suggests that most, but not all, CCAAT sequences picked up in the Bucher study are actually NF-Y binding sites, since a minority of them, ~15–20%, contain nucleotides that are at odds with NF-Y binding. In agreement with this, the frequency of CCAAT-containing promoters as measured by Bucher (30%) is slightly higher than that we measured on a larger sample of 1200 promoters, evaluated at 25% (M.Pontoglio and R.Mantovani, unpublished results). These subtle differences

Table 1. *continued*

ALDH		TTCATCCAATCGTATC	Hs	>	-74	+	ND	OCT	+	+	Li	96
		CCCATCCAATCATATC	Mm	>	-70	+						
		GCCATCCAATCATATC	Rn	>	-94	+						
LPL		TATAGCCAATAGGTGA	Hs	>	-65	-	Yes	OCT	+	+	AdMy	97
		TATAGCCAATAGGTGA	Mm	>	-65	-						98
		GTGCGCCAATGGGTGT	Gg	>	-67	-						
ExoKII	D	CACAGCCAATCAGCGC	Rn	>	-84	+	Yes	ATF	+	+	AdHeSkm	99
	P	CGCAGCCAATGAGCGC	Rn	<	-141	+	Yes		+	+		
FAS	P	CCAGGCCAATGAGCGT	Rn	<	-97	+	Yes		+	+	AdLi	100
		CCAGGCCAATGAGCGT	Hs	<	-88	+						
		GCAGTCCAATGAGAGC	Gg	<	-90	+						
		GCAGTCCAATGAGAGC	Aa	<	-89	+						
	D	CGAGACCAATTGGACA	Rn	<	-502	+	Yes	RF-X	+	+		101
		CCGTGCCAATGCGGAG	Gg	<	-471	+						
		GGCACCCAATCAGGCG	Aa	<	-510	+						
TSP-1		TCCGGCCAATGGGCGG	Hs	<	-64	+	Yes		+	+	FiSm	102
FGF-4		GCCTGCCAATCAGGGC	Hs	<	-139	+	Yes		+	+	EcGl	103
		GCCTGCCAATCACCGC	Mm	<	-106	+						
α1-chim		GGTGGCCAATCTAATC	Hs	>	-52	-	Yes		+			104
Tr.Hydr		AACGGCCAATGGGCGC	Mm	<	-54		Yes		+	+	Br	105
NaKATPsea-3		TCTGGCCAATCAGGAG	Rn	<	-62	+	Yes	3SP1	+	+	HeSmBr	106
PDGFβ		CTTGGCCAATCAGAAT	Mm	>	-55	-	Yes		+	+	UInd	107
FerH		CCCGGCCAATCAGCGC	Hs	<	-55	+	Yes	SP1	+	+	UInd	108
MHC IA2 B8		GACACCCAATGGGAGT	Hs	<	-74	-	Yes		+	+	UInd	109
	Cw2Ld	GCCACCCAATGATAGT	Hs	<	-74	-						
	B7	AATCACCAATGGGAGT	Hs	<	-74	-						
MDR1		CCCAGCCAATCAGCCT	Hs	<	-77	-	Yes	cEBPSP1	+	+	UInd	110
		GGCAGCCAATCAGCCT	Mm	<	-65	+			+	+		111
CYP1A1		CTCTGCCAATCAAAGC	Hs	<	-780		Yes		+	+	UInd	112
c-JUN		GCGAGCCAATGGGAAG	Hs	>	-89	-		SP1AP1	+	+	UInd	113
Grp78		GTTCACCAATCGGAGG	Rn	>	-96	+	Yes		+	+	UInd	114
Hsp70	P	CTGAGCCAATCACCGA	Hs	<	-64	+	Yes	HSFSP1	+	+	UInd	115,53
		CTGGGCCAATCAGCGA	Mm	<	-64	+		HSFSP1				
		AAGGACCAATCCAGAC	Gg	<	-69	+		HSF				116
		ATTAGCCAATCAAGGC	Xl	<	-56	+	Yes					117
	D	CCTGCCCAATCAGAAG	Gg	<	-150	+						
		GTTAGCCAATCAGCAA	Xl	<	-140	+	Yes	HSF				
ADH2		GGCAGCCAATGGAAAG	Hs	<	-94	-	Yes		+	+	UInd	118
GPAT		ATCAGCCAATGAGCTC	Mm	<	-75	-	Yes	SREBP	+	+	UInd	119
FPP	D	ACTGGCCAATGAAAGG	Rn	<	-285	-	Yes		+	+	UInd	120
	P	TTCAGCCAATCAGCGA	Rn	>	-240	-						
HMG		GCCAACCAATAGCTGG	Rn	<	-181	+	Yes	SREBP	+	+	UIND	121
HSS		CCTGGCCAATCAGCGC	Hs	>	-126	-	Yes	SREBP	+	+	UInd	122
SREBP2		CTCAGCCAATGGGCGA	Hs	<	-99	-	Yes	SREBP	+	+	UInd	123
GST28		CACAGCCAATGAGGCA	Sm	>	-125	+	Yes		+	+	Uind	124
		AGTGACCAATAAAAAT	Sm	<	-143	+	Yes		+	+		
GHR		TTCCACCAATAGGGTT	Mm	>	-3500				+	+		125
CP2		GCCAACCAATCATGGC	Hs	<	-80	-	ND		+		U	126
		GCCAACCAATCAGGAC	Mm	<		-			+			
β-actin		CGCGGCCAATCAGCGT	Hs	>	-89	+	Yes	SRF	+	+	U	127
		CGGAGCCAATCAGCGG	Rn	>	-89	+						
		GGCAGCCAATCAGAGC	Gg	>	-92	+	Yes					128
TK	D	TGGGGCCAATCAGCGC	Hs	<	-69	-	Yes	E2FSP1	+	+	G1/S	129
		GCCGACCAATGCGGAG	Gg	<	-46	-						130
	P	GCTGGCCAATCACGAG	Hs	<	-37	-	Yes		+			129
		CCGAGCCAATCGCCGG	Gg	<	-12	-						
		ACCGACCAATGGCAGC	Ha	<	-40	-						
TopoIIα	I	ACCAGCCAATCCCTCA	Hs	<	-67	-	Yes	SP1	+	+	G2/M	131
		AGGAACCAATCACCGA	Ha	<	-63	-		SP1	+			133
		AGAAACCAATCACCGA	Mm	<	-42	-		SP1				134
	II	AAGAACCAATCGTAGC	Hs	<	-107	-	Yes		+			132
		AGTAACCAATCGTGGA	Ha	<	-103	-			+			133
		AGTAACCAATCGTAGA	Mm	<	-81	-						
	III	ATAAACCAATCAGGTT	Hs	<	-174	-	Yes		+			132
		ATGAACCAATTAGGTA	Ha	<	-173	-			+			133

notwithstanding, one can reasonably conclude that most of the CCAAT-containing promoters are indeed recognized by NF-Y.

I have also analysed the relative frequency of NF-Y sites with random DNA totalling twice as much DNA (a 86761 bp contig sequence from human Xq2.8) as the sum of the promoter sequences. Seven NF-Y consensus sites were detected (five in the CCAAT and two in the ATTGG orientation): this is in line with the theorical frequency of one site every 34096 bp (five sites expected in both orientations) and indicates that the NF-Y consensus is not over-represented in random DNA sequences. Thus NF-Y binding sites are extremely over-represented in promoter sequences as compared with intergenic DNA. Moreover, it should be noted that two of the sites pinpointed in this analysis were within 30 bp of each other, in close proximity to a transcribed region of the HMG0 gene (T.Vaccari and M.Bianchi, personal communication), possibly representing true promoter elements.

Table 1. *continued*

		ATGAACCAATTAGGTA	Mm	<	-133	-						
	IV	TCTGGCCAATGAGAAG	Hs	<	-248	-			+			132
		TCTGGCCAATGAGAAA	Ha	<	-257	-			+			133
		ATGGACCAATAGCAAT	Mm	<	-174	-						
cdc25	3	CATGGCCTATCGTTGG	Hs	<	-93	-	Yes		+	+	S/G2	135
	2	GTCAGCCAATCTCCGC	Hs	<	-60	-	Yes		+	+		
	1	AGTAACCTATCCCCGC	Hs	<	-29	-	Yes		+	+		
cdc2	1	ATTCACCAATCGGGTA	Hs	<	-44	-	Yes		+	+		136
		GTCCGCCAATCCGATT	Rn	<	-44	-						137
		ACCCACCAATGGAGCA	Cc	<	+13	-						138
	2	AGCAGCCAATCAGACG	Hs	<	-76	-	Yes		+	+		
		AGGAGCCAATCAGAGC	Rn	<	-76	-						
		AGCGACCAATGGGAGC	Cc	<	-21	-						
CyclA		ATAAACCAATGAGGGC	Mm	<	-3	-	Yes	ATFE2F	+	+	S/G2	139,140
		TAGGACCAATGAAAGC	Hs	<	-53	-						
CyclB1	D	AGCCGCCAATGGGAAG	Hs	>	-15	-	Yes		+	+	G2/M	141,142
	P	ACAGGCCAATAAGGAG	Hs	>	+18	-	Yes		+	+		
E2F1	P	CTCGGCCAATGGAAGC	Mm	>	-71	-	Yes	SP1E2F	+	+	G1/S	143
		GGCAGCCAATTGTGGC	Hs	>	-70	-						144
PLK		CGCGGCCAATCAGTGG	Hs	>	-37	-	Yes	SP1	+	+	G2/M	145
RRR2		ACCAACCAATCACAGA	Mm	>	-67	+	Yes		+	+	S	146
HisH2B	D	TTTAGCCAATCAGCTA	Sp	<	-131	+		OCT	+		Ts	14
	P	ATCTACCAATCAACGC	Sp	<	-99	+			+			
	D	ATTAACCAATCAGAAA	Rn	>	-128	+		Oct			Ts	
	P	ATTGTCCAATCATCTT	Rn	>	-71	+						
HisH3.3		GTTGACCAATCAACAG	Sp	>	-73	+			+		U	147
		CGCAGCCAATCAAGAG	Hs									
HBV S		CTCCACCAATCGGCAG		>	-45	-	Yes	SP1	+	+		148
MSV LTR		ACTAACCAATCAGTTC		<	-84	+	Yes	SP1	+	+	U	21
RSV LTR	P	TTCGTCCAATCCATGT		<	-67	+	Yes	SRF	+		U	52
	D	TTCCACCAATCGGCAG		<	-131	+	Yes	SRF	+			52
	ev2RAV0	CGCCACCAATGGGCAT		<	-95	+	ND		+			149
Ad EIIL II		GAAGACCAATCCCGCC		<	-74	+	ND	SP1USF	+		U	150
Ad ML		ATAAACCAATCACCT		>	-70	+	Yes	USF	+		U	51
CMV gpUL4		GGGACCCAATCACTGG		>	-96	+	Yes		+	+	U	151
HSV IE110K		TTCCCCCAATGGCCGC		<	-73		Yes		+	+		152
VZV ORF62		CTCGTCCAATCACTAC		>	-115		Yes		+	+		153
MVM P4		ACTGACCAACCATGTG		<	-97	-	Yes	USF	+	+		154

GENE	SEQUENCE	ORG	OR	POS	TATA	ACT	HAP2/3/4/5	REF
AmdS	GCCAGCCAATCACCAG	An	>	-137		Yes	+	155
TaaG2	ACCATCCAATTAGAAG	An	>	-310		Yes	+	155
GatA	TTCGACCAATAATATT	An	<	-134		Yes	+	155
TaA	ACCATCCAATTAGAAG	An	>	-310	+	Yes		155
CYC1UAS2	ATCCACCAACCAACGC	Sc	<	-206	+	Yes	+	156
CIT1	ATCCTCCAATAACACA	Sc	<	-290	+	Yes	+	157
COX6	ACGAGCCAATCAGGGC	Sc	>	-285		Yes	+	158
CYT1	CTCCACCAACCAAATC	Sc	<	-470		Yes	+	159
LPD1	TCTCGCCAATGAGGGA	Sc	<	-200		Yes	+	160
COX5a	ATCGTCCAATAGACGT	Sc	>	-170			+	161
HEM1	GCAGACCAATGAGCGA	Sc	<	-375		Yes	+	162
ASN1	ATCCACCAATCACACG	Sc	>	-355		Yes	+	163
NADPG1.Deh.	GGCGACCAATAAACAC	Nc	>	-1300			+	164
SOD2	CTGGACCAATAACACA	Sc	>	-220		Yes	+	165

CCAAT BOXES NOT BINDING NF-Y.

GENE	SEQUENCE	ORG	REF
Globin ε	CTTGACCAATGATTTT	Hs	72
β	TAAGGCCAATCTGCTC	Mm	21
	GTTGGCCAATCTACTC	Hs	
	CCTAGCCAATCAACAG	Xl	
GP91phoxP	ATTAGCCAATTTCTGA	Hs	63
IL4 P	AATTTCCAATGTAAAC	Hs	67
TopoIIαV	AACTGCCAATCTATTT	Hs	132
	ATAGACCAATAGTCTA	Ha	
FcγRec1	TTGAACCAATAGTCTA	Hs	166
Carboxyest.	CTCTCCCAATTAGAGG	Hs	166
CD14	ACCCCCCAATCCCCCT	Hs	166
HSV TK	ATTCGCCAATGACAAG	Hs	21

ARCHITECTURAL FEATURES OF NF-Y BINDING PROMOTERS

The first observation that can be made as to where the typical NF-Y binding site is positioned in regulatory regions is that it is rarely distant from the Start site: only the Ea and Dra Y′ boxes, the chicken ρ-globin 3′ enhancer and the CYP1A1, FAS and GHR CCAAT boxes are distant from proximal promoters. Indeed, the Ea/Dra and FAS genes also have NF-Y sites in their promoters. The CCAAT sequence can be found both in the direct and in the inverted orientation and it is present in both TATA-containing (such as the globins) and in TATA-less (such as MHC class II)

Table 2.

NF-Y CONSENSUS IN HIGHER EUKARYOTES

	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11
A	52	22	16	80	68	0	2	160	164	0	20	96	21	58	47	34
C	47	51	67	19	7	164	161	0	0	1	88	10	28	57	11	53
G	41	38	40	60	78	0	1	1	0	0	50	55	98	33	72	40
T	22	52	41	5	11	0	0	3	0	163	6	3	17	16	34	35

HAP2/3/5 CONSENSUS IN LOWER EUKARYOTES

	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11
A	7	1	1	4	6	0	0	14	14	0	4	13	3	8	4	5
C	2	6	9	5	0	14	14	0	0	2	6	0	5	4	3	3
G	2	0	3	5	4	0	0	0	0	0	2	1	5	2	5	4
T	3	7	1	0	4	0	0	0	0	12	2	0	1	2	2	2

NF-Y CONSENSUS

	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11
A	59	23	17	84	74	0	2	174	178	0	24	109	24	66	51	39
C	49	57	76	24	7	178	175	0	0	3	94	10	33	61	14	56
G	43	38	43	65	82	0	1	1	0	0	52	56	103	35	77	44
T	25	59	42	5	15	0	0	3	0	175	8	3	18	18	36	37

CCAAT CONSENSUS (Ref. 3)

	-5	-4	-3	-2	-1	1	2	3	4	5	6	7
A	55	32	25	102	51	0	0	175	119	17	23	116
C	55	52	47	1	6	173	174	0	8	0	90	6
G	12	43	24	70	99	1	0	0	21	15	59	52
T	52	48	79	2	19	1	1	0	27	143	3	1

Table 3.

Orientation	Number of sites
ATTGG tot.	99
CCAAT tot.	64

promoters. I therefore analyzed the position of *bona fide* NF-Y binding sites with respect to the transcriptional +1 signal, taking into account two parameters, the orientation of the CCAAT box and the presence of a TATA box. In higher eukaryotes the CCAAT box is present in the reverse ATTGG orientation in 60% of cases, considering both the overall number of sites (99 versus 64) or the most proximal sites only (73 versus 52) (see Tables 3 and 4).

I next verified how many of the TATA-containing and TATA-less promoters contain either a CCAAT or an ATTGG sequence and where, relative to +1, they are positioned. To derive these data I considered all promoters containing single NF-Y binding sites and the most proximal sites for those promoters in which multiple CCAAT are present. Table 4 indicates that 68 out of 119 promoters contain a TATA sequence. This figure of 57% represents a fair under-representation compared with the frequency of the TATA box in the overall promoter database as calculated in the Bucher study (79%), especially since some of the promoters

Table 4.

Orientation	TATA box	Number of promoters	Average position
ATTGG prox.	-	73	-74±32
CCAAT prox.	-	52	-86±38
	Yes	68	-89±29
	No	51	-66±39
CCAATprox	Yes	34	-93±25
CCAATprox	No	16	-72±55
ATTGGprox	Yes	34	-86±33
ATTGGprox	No	35	-63±29

containing a TATA-like sequence might actually work without it. In fact, my observations are based upon the presence of a TATA consensus in the –25/–30 region, but only for very few of these regions are functional data available, indicating that the TATA sequence is indeed required. In the case of the Ea promoter, for example, a TATA-like sequence binding TBP and TFIID is present at –25, but is functionally irrelevant (54,55).

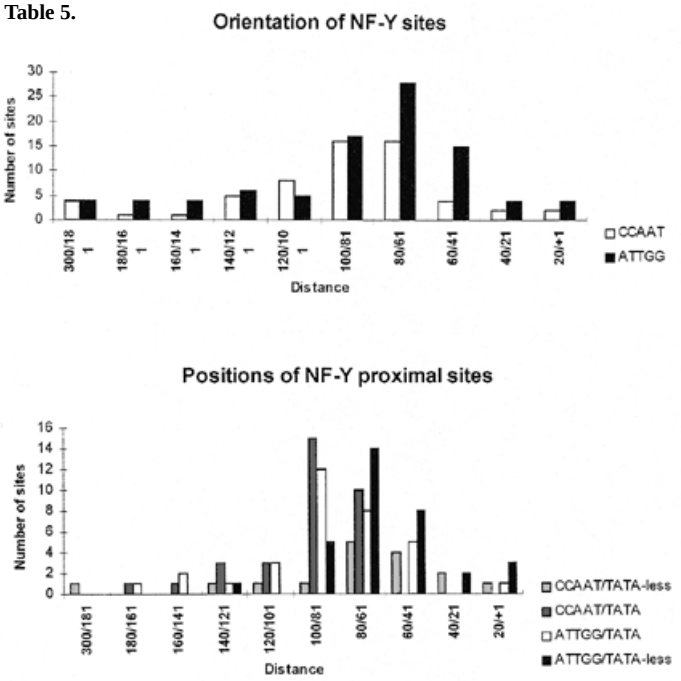
Comparison of the relative frequencies of CCAAT and ATTGG boxes in TATA-containing and TATA-less promoters shows clear skewing: TATA contains roughly equal numbers of CCAAT and ATTGG (40 and 43 respectively) for the total sites and 34 each if the analysis is limited to proximal sites (see Table 4). On the other hand, TATA-less promoters showed a significant difference in favour of ATTGG (53 versus 20 for total sites and 35 versus 16 considering proximal sites).

I then analyzed the relative distances of the CCAAT/ATTGG orientations from the Cap site, considering all NF-Y binding sites. Table 5 shows a peak of ATTGG in the –61/–80 area and many sites located between –41 and –60, while CCAAT is evenly distributed in more upstream regions, from –61 to –100. Note the relatively high number of ATTGG upstream sites beyond –120. This is largely due to a limited number of promoters with sequences from multiple species, such as topoisomerase IIα, containing several NF-Y sites. By limiting the analysis to proximal sites, the actual number of promoters in which each NF-Y/TATA arrangement is present can be more precisely calculated (Table 5): in the CCAAT–TATA combination the NF-Y sites have a peak between –80 and –100 (mean value 93 ± 25) and 74% of the sites are between –60 and –100; in most of the ATTGG–TATA-less promoters (62%) the NF-Y binding sites are located between –41 and –80 (mean value 63 ± 29). A similar situation is observed with CCAAT–TATA-less, whereas in ATTGG–TATA most of the NF-Y binding sites are in the –80/–100 region. Moreover, it is important to note that in the presence of a TATA box the NF-Y binding CCAAT box is never closer than –48 in the reverse ATTGG configuration or –62 in the

The list of NF-Y binding sites is as of September 1997.

GENE, the name of the gene is indicated as well as the CCAAT sequence in the promoter. P and D indicate proximal and distal sites respectively. ORG, abbreviated names of the different species: Hs, man; Rn, rat; Mm, mouse; Bt, bovine; Sp, rabbit; Gg, chicken; Xl, *Xenopus laevis*; Xt, *Xenopus tropicalis*; Gc, galago; Ha, hamster; Cc, quail; Aa, goose; Sm, *Schistosoma mansoni*; Sc, *Saccharomyces cerevisiae*; An, *Aspergillus nidulans*; Nc, *Neurospora crassa*. OR indicates the orientation of the NF-Y site: > is forward CCAAT; < is reverse ATTGG. POS is the CCAAT position with respect to the +1 signal, calculated taking into account the central +3 A. TATA indicates whether the promoter has a consensus TATA sequence in the –20/–30 region. ACT refers to a positive effect on promoter activity, as tested in functional assays either *in vitro* or *in vivo*. TF indicates the presence of a proven binding site for a transcription factor close to the NF-Y binding site. COMP indicates whether cross-competition data with *bona fide* NF-Y binding sites are available. AB indicates whether EMSA supershift experiments with anti-NF-Y antibodies were performed. EXPR refers to the tissue or cell type specificity of the gene activated by NF-Y: Bl, B lymphocytes; Tl, T lymphocytes; Mo, macrophages; Gr, granulocytes; Ery, erythroid cells; Bo, bone; Sk, skeletal muscle; Li, liver; Sp, spleen; Lu, lung; Ad, adipocytes; My, myoblasts; He, heart; Fi, fibroblast; Sm, smooth muscle; Ec, embryonal carcinoma; Br, brain; U, ubiquitous; Uind, ubiquitous and inducible; Ts, testis. REF is the reference number. HAP/2/3/4/5 indicates the dependence of the promoter from intact HAP genes.

Table 5.



forward CCAAT. It is usually positioned at ~60 nt from the TBP binding site, irrespective of orientation. In the absence of a recognizable TATA box, however, NF-Y sites are much closer to +1 and in some cases indeed overlap the transcriptional site. Among the several such examples we find the cell cycle-regulated genes cdc2, CDC25, cyclin A and cyclin B1, which seem to prefer the multiple CCAAT–TATA-less configuration.

In conclusion, NF-Y sites show a predominance in proximal promoter regions; the CCAAT/ATTGG position is far from being randomly distributed, both in terms of orientation and presence of a neighbouring TATA box. We take these data as yet another indication that NF-Y can serve multiple architectural roles in the functional organization of different classes of promoters and it is possible that, in the absence of TBP–TATA interactions in the –25 region, NF-Y functions as the pivotal factor in connecting upstream activators with the general transcription machinery, thus helping polymerase II to focus on the Start site(s).

ACKNOWLEDGEMENTS

I thank C.Liberati for help with the Tables, M.Pontoglio and M.Accarino for help with the computer analysis and M.Schwarzbauer, H.Zhu, G.Piaggio, J.Sodek, R.Kim, G.Marziali, T.Vaccari and M.Bianchi for sharing unpublished results. My work is supported by Milano Ricerche, MURST, CNR and Telethon (E582) grants. An NF-Y web site is accessible at http://www.csi.unimi.it/CCAAT-NFY

REFERENCES

1 Tjian,R. and Maniatis,T. (1994) *Cell*, **77**, 5–8.
2 Bucher,P. and Trifonov,E.N. (1988) *J. Biomol. Struct. Dynam.*, **5**, 1231–1236.
3 Bucher,P. (1990) *J. Mol. Biol.*, **212**, 563–578.
4 Benoist,C., O'Hare,K., Breathnach,R. and Chambon,P. (1980) *Nucleic Acids Res.*, **8**, 127–142.
5 Efstradiatis,A., Posakony,J.W., Maniatis,T., Lawn,R.M., O'Connel,C., Spritz,R.A., DeRiel,J.K., Forget,B.G., Weissman,S.M., Slightom,J.L., Blechl,A.E., Smithies,O., Baralle,F.E., Shoulders,C.C. and Proudfoot,N.J. (1980) *Cell*, **21**, 653–668.
6 Landshultz,W-H., Johnson,P.F., Adashi,E.Y., Graves,B.J. and McKnight,S.L. (1988) *Genes Dev.*, **2**, 786–800.
7 Umek,R.M., Friedman,A.D. and McKnight,S.L. (1991) *Science*, **251**, 288–292.
8 Osada,S., Yamamoto,H., Nishihara,T. and Imagawa,M. (1996) *J. Biol. Chem.*, **271**, 3891–3896.
9 Jones,K.A., Yamamoto,K.R. and Tjian,R. (1985) *Cell*, **42**, 559–572.
10 Zorbas,H., Rein,T., Krause,A., Hoffmann,K. and Winnaker,E.-L. (1992) *J. Biol. Chem.*, **267**, 8478–8484.
11 Osada,S., Daimon,S., Nishihara,T. And Imagawa,M. (1997) *FEBS Lett.*, **390**, 44–46.
12 Didier,D.K., Schiffenbauer,J., Woulfe,S.L., Zaceis,M. and Schwarz,B.D. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 283–287.
13 Wolffe,A.P., Tafuri,S., Ranjan,M. and Familari,M. (1992) *New Biologist*, **4**, 290–298.
14 Barberis,A., Superti-Furga,G. and Busslinger,M. (1987) *Cell*, **50**, 347–359.
15 Superti-Furga,G., Barberis,A., Schaffner,G. and Busslinger,M. (1988) *EMBO J.*, **7**, 3099–3107.
16 Neufeld,E.J., Skalnik,D.G., Lievens,P.M-J. and Orkin,S.H. (1992) *Nature Genet.*, **1**, 50–55.
17 Aufiero,B., Neufeld,E.J. and Orkin,S.H. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 7757–7761.
18 Lum,L.S.Y., Sultzman,L.A., Kaufman,R.J., Linzer,D.I.H. and Wu,B.J. (1990) *Mol. Cell. Biol.*, **10**, 6709–6717.
19 Gallinari,P., LaBella,F. and Heintz,N. (1989) *Mol. Cell. Biol.*, **9**, 1566–1575.
20 Martinelli,R. and Heintz,N. (1994) *Mol. Cell. Biol.*, **14**, 8322–8332.
21 Dorn,A., Bollekens,J., Staub,A., Benoist,C. and Mathis,D. (1987) *Cell*, **50**, 863–872.
22 Hatamochi,A., Golumbek,P., Van Schaftingen,E. and de Crombrugghe,B. (1987) *J. Biol. Chem.*, **263**, 5940–5947.
23 Kim,C.G. and Sheffery,M. (1990) *J. Biol. Chem.*, **265**, 13362–13369.
24 Hooft van Huijsduijnen,R., Bollekens,J., Dorn,A., Benoist,C. and Mathis,D. (1987) *Nucleic Acids Res.*, **15**, 7265–7272.
25 Hooft van Huijsduijnen,R., Li,X.Y., Black,D., Matthes,H., Benoist,C. and Mathis,D. (1990) *EMBO J.*, **9**, 3119–3127.
26 Hahn,S. and Guarente,L. (1988) *Science*, **240**, 317–321.
27 Chen,H. and Kinsey,J.A. (1995) *Mol. Gen. Genet.*, **249**, 301–308.
28 Sinha,S., Maity,S.N., Lu,J. and deCrombrugghe,B. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 1624–1628.
29 Maity,S., Vuorio,T. and deCrombrugghe,B. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 5378–5382.
30 Vuorio,T., Maity,S. and de Crombrugghe,B. (1990) *J. Biol. Chem.*, **265**, 22480–22486.
31 Li,X-Y., Mantovani,R., Hooft van Huijsduijnen,R., Andrè,I., Benoist,C. and Mathis,D. (1992) *Nucleic Acids Res.*, **20**, 1087–1091.
32 Pinkham,J.L., Olesen,S. and Guarente,L. (1987) *Mol. Cell. Biol.*, **7**, 578–585.
33 Hahn,S., Pinkham,J., Weil,R., Miller,R. and Guarente,L. (1988) *Mol. Cell. Biol.*, **8**, 655–663.
34 Xing,Y., Fikes,J.D. and Guarente,L. (1993) *EMBO J.*, **12**, 4647–4655.
35 McNabb,D.S., Xing,Y. and Guarente,L. (1995) *Genes Dev.*, **9**, 47–58.
36 Nakshatri,H., Bhat-Nakshatri,P. and Currie,R.A. (1996) *J. Biol. Chem.*, **271**, 28784–28791.
37 Bellorini,M., Zemzoumi,K., Farina,A., Berthelsen,J., Piaggio,G. Mantovani,R. (1997) *Gene*, **193**, 119–125.
38 Olesen,S., Fikes,J.D. and Guarente,L. (1991) *Mol. Cell. Biol.*, **11**, 611–619.
39 Albani,D. and Robert,L.S. (1995) *Gene*, **167**, 209–213.
40 Zemzoumi,K., Serra,E., Mantovani,R., Trolet,J., Capron,A. and Dissous,C. (1996) *Mol. Biochem. Parasitol.*, **77**, 161–172.
41 Li,X., Kalasapudis,S. and Childs,J. (1993) *Nucleic Acids Res.*, **21**, 4639.
42 Mulder,W., Scholten,H.J.M.I., deBoer,R.W. and Grivell,L.A. (1994) *Mol. Gen. Genet.*, **245**, 96–106.
43 Papagiannopoulos,P., Andrianopoulos,A., Sharp,J.A., Davis,M.A. and Hynes,M.J. (1996) *Mol. Gen. Genet.*, **251**, 412–421.
44 Maity,S.N. and deCrombrugghe,B. (1992) *J. Biol. Chem.*, **267**, 8286–8292.
45 Xing,Y., Zhang,S., Olesen,J.T., Rich,A. and Guarente,L. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 3009–3013.
46 Mantovani,R., Li,X.-Y, Pessara,U., Hooft van Huijsduijnen,R., Benoist,C. and Mathis,D. (1994) *J. Biol. Chem.*, **269**, 20340–20346.
47 Baxevanis,A.D., Arents,G., Moudrianakis,E.N. and Landsman,D. (1995) *Nucleic Acids Res.*, **23**, 2685–2691.

48 Sinha,S., Kim,I.-S., Sohn,K.Y., deCrombrugghe,B. and Maity,S.N. (1996) *Mol. Cell. Biol.*, **16**, 328–337.
49 Kim,I.-S., Sinha,S., deCrombrugghe,B. and Maity,S.N. (1996) *Mol. Cell. Biol.*, **16**, 4003–4013.
50 Mantovani,R., Pessara,U., Tronche,F., Li,X.Y., Knapp,A.M., Pasquali,J.L., Benoist,C. and Mathis,D. (1992) *EMBO J.*, **11**, 3315–3322.
51 Chodosh,L.A., Baldwin,A.S., Carthew,R.W. and Sharp,P. (1988) *Cell*, **53**, 11–25.
52 Faber,M. and Sealy,L. (1990) *J. Biol. Chem.*, **265**, 22243–22254.
53 Ronchi,A., Bellorini,M., Mongelli,N. and Mantovani,R. (1995) *Nucleic Acids Res.*, **23**, 4565–4572.
54 Viville,S., Jongeneel,V., Koch,W., Mantovani,R., Staub,A., Ruffenach,F., Benoist,C. and Mathis,D. (1991) *J. Immunol.*, **146**, 33211–3217.
55 Bellorini,M., Dantonel,J.C., Yoon,J.-B., Roeder,R.G., Tora,L. and Mantovani,R. (1996) *Mol. Cell. Biol.*, **16**, 503–512.
56 Benoist,C. and Mathis,D. (1990) *Annu. Rev. Immunol.*, **8**, 681–715.
57 Dorn,A., Fehling,H.J., Koch,W., LeMeur,M., Gerlinger,P., Benoist,C. and Mathis,D. (1989) *Mol. Cell. Biol.*, **8**, 3975–3987.
58 Feriotto,G., Mischiati,C., Bianchi,N., Rutigliano,C., Giacomini,P. and Gambari,R. (1995) *Nucleic Acids Res.*, **23**, 1671–1678.
59 Kern,M.J. and Woodward,J.G. (1991) *Mol. Cell. Biol.*, **11**, 578–581.
60 Turco,E., Manfras,B.J., Rudert,W.A and Trucco,M. (1990) *Immunogenetics*, **32**, 117–128.
61 Nittisombut,N. (1988) *Mol. Cell. Biol.*, **8**, 2034–2041.
62 Wright,K.L., Moore,T.L., Vilen,B.J., Brown,A.M. and Ting,J.P.-Y. (1995) *J. Biol. Chem.*, **270**, 20978–20986.
63 Wright,T.M. and Farber,J.M. (1991) *J. Exp. Med.*, **173**, 417–422.
64 Luo,W and Skalnik,D.G. (1996) *J. Biol. Chem.*, **271**, 18203–18210.
65 Ishimaru,F., Mari,B and Shipp,M.A. (1997) *Blood*, **89**, 4136–5145.
66 Brown,S.T., Miranda,G.A., Galic,Z., Hartman,I.Z., Lyon,C.J. and Aguilera,R.J. (1997) *J. Immunol.*, **158**, 5071–5074.
67 Li-Weber,M.L., Davydov,I.V., Krafft,H. and Krammer,P.H. (1994) *J. Immunol.*, **153**, 4122–4133.
68 Heussler,V.T., Eichhorn,M. and Dobbelaere,D.A.E. (1992) *Gene*, **114**, 273–278.
69 Spanopoulou,E., Giguere,V. and Grosveld,F. (1991) *Mol. Cell. Biol.*, **11**, 2216–2228.
70 Yost,S.E., Shewchuk,B. and Hardison,R. (1993) *Mol. Cell. Biol.*, **13**, 5439–5449.
71 Stalder,J., Wirthmuller,U., Beck,J., Gruber,A., Meyerhof,W., Knochel,W. and Weber,R. (1988) *J. Mol. Evol.*, **28**, 64–71.
72 Liberati,C., Ronchi,A., Lievens,P., Ottolenghi,S. and Mantovani,R. (1998) *J. Biol. Chem.*, submitted for publication.
73 Mantovani,R., Superti-Furga,G., Gilman,J. and Ottolenghi,S. (1989) *Nucleic Acids Res.*, **17**, 6681–6691.
74 Gumucio,D., Heilstedt-Williamson,H., Gray,T.A., Tarle,S.A., Shelton,D., Tagle,D.A., Slightom,J.L., Goodman,M. and Collins,F.S. (1992) *Mol. Cell. Biol.*, **12**, 4919–4929.
75 Wandersee,N.J., Ferris,R.C., Ginder,G.D. (1996) *Mol. Cell. Biol.*, **16**, 236–246.
76 Karsenty,G., Golumbek,P. and deCrombrugghe,B. (1988) *J. Biol. Chem.*, **263**, 13909–13915.
77 Collins,M., Leaner,V.D., Madikizela,M. and Parker,M.I. (1997) *Biochem. J.*, **322**, 199–206.
78 Karsenty,G. and deCrombrugghe,B. (1990) *J. Biol. Chem.*, **265**, 9934–9942.
79 Tezuka,K., Denhardt,D.T., Rodan,G.A. and Harada,S. (1996) *J. Biol. Chem.*, **271**, 22713–22717.
80 Sodek,J. and Kim,R. (1997) Personal communication.
81 Raymondjean,M., Cereghini,S. and Yaniv,M. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 757–761.
82 Wuarin,J., Muller,C. and Schibler,U. (1990) *J. Mol. Biol.*, **214**, 865–874.
83 Tronche,F., Rollier,A., Sourdive,D., Cereghini,S. and Yaniv,M. (1991) *J. Mol. Biol.*, **221**, 31–43.
84 Schorpp,M., Dobbeling,U., Wagner,U. and Ryffel,G.U. (1988) *J. Mol. Biol.*, **202**, 307–320.
85 Novak,EM. and Bydlowski,S.P. (1997) *Biochem. Biophys. Res. Commun.*, **231**, 140–143.
86 Papazafiri,P., Ogami,K., Nicosia,M., Monaci,P., Cladaras,C. and Zannis,V.I. (1991) *J. Biol. Chem.*, **266**, 5790–5797.
87 Raymondjean,M., Pichard,A.-L., Gregori,C., Ginot,F. and Kahn,A. (1991) *Nucleic Acids Res.*, **19**, 6145–6153.
88 Schweizer-Groyer,G., Groyer,A., Cadepond,F., Grange,T., Baulieu,E.-E. and Pictet,R. (1994) *Nucleic Acids Res.*, **22**, 1583–1592.
89 Brouillet,A., Darbouy,M., Okamoto,T., Chobert,M.-N., Lahuna,O., Garlatti,M., Goodspeed,D. and Laperche,Y. (1994) *J. Biol. Chem.*, **269**, 14878–14884.
90 Noda,C., Fukushima,C., Fujiwara,T., Matsuda,K., Kobune,Y. and Ichihara,A. (1994) *Biochim. Biophys. Acta*, **1217**, 163–173.
91 Alonso,C.R., Pesce,C.G. and Kornblihtt,A.R. (1996) *J. Biol. Chem.*, **271**, 22271–22279.
92 Matsubasa,T., Takiguchi,M., Matsuda,I. and Mori,M. (1994) *J. Biochem.*, **116**, 1044–1055.
93 McGlynn,L.K., Mueller,C.R., Begbie,M., Notley,C.R. and Lillicrap,D. (1996) *Mol. Cell. Biol.*, **16**, 1936–1945.
94 Hung,H.-L. and High,K.A. (1996) *J. Biol. Chem.*, **271**, 2323–2331.
95 Ueda,A. and Yoshimura,T. (1996) *J. Biol. Chem.*, **271**, 20265–20272.
96 Yanagawa,Y., Chen,J.C., Hsu,L.C. and Yoshida,A. (1995) *J. Biol. Chem.*, **270**, 17521–17527.
97 Currie,R.A. and Eckel,R.H. (1992) *Arch. Biochem. Biophys.*, **298**, 630–639.
98 Lu,S.-C. and Besdoun,A. (1993) *Biochim. Biophys. Acta*, **1216**, 375–384.
99 Osawa,H., Brooks Robey,R., Printz,R.L. and Granner,D.K. (1996) *J. Biol. Chem.*, **271**, 17296–17303.
100 Roder,K., Wolf,S.S., Beck,K.-F., Sickinger,S. and Schweizer,M. (1997) *Gene*, **184**, 21–26.
101 Roder,K., Wolf,S.S., Beck,K.-F. and Schweizer,M. (1997) *J. Biol. Chem.*, **272**, 21616–21624.
102 Framson,P. and Bornstein,P. (1993) *J. Biol. Chem.*, **268**, 4989–4996.
103 Hasan,S., Koda,T. and Kakinuma,M. (1994) *J. Biol. Chem.*, **269**, 25042–25048.
104 Dong,J.-M., Smith,P., Hall,C. and Lim,L. (1995) *Eur. J. Biochem.*, **227**, 636–646.
105 Reed,G.E., Kirchner,J.E. and Carr,L.G. (1995) *Brain Res.*, **682**, 1–12.
106 Murakami,Y., Ikeda,U., Shimada,K. and Kawakami,K. (1997) *Biochim. Biophys. Acta*, **1352**, 311–324.
107 Ishisaki,A., Murayama,T., Ballagi,A.E. and Funa,K. (1997) *Eur. J. Biochem.*, **246**, 142–146.
108 Marziali,G., Perrotti,E., Ilari,R., Testa,U., Coccia,E.U. and Battistini,A. (1997) *Mol. Cell. Biol.*, **17**, 1387–1395.
109 Schoneich,J., Lee,J.L., Manski,P., Sheffery,M. and Yang,S.Y. (1997) *J. Immunol.*, **158**, 4788–4796.
110 Yu,L., Wu,Q., Huang Yang,C.-P. and Horwitz,S.B. (1995) *Cell Growth Differentiat.*, **6**, 1505–1512.
111 Goldsmith,M.E., Madden,M.J., Morrow,C.S. and Cowan,K.H. (1993) *J. Biol. Chem.*, **268**, 5856–5860.
112 Boucher,P.D., Piechocki,M.P. and Hines,R. (1995) *Mol. Cell. Biol.*, **15**, 5144–5151.
113 Tommasi,S., Swiderski,P.M., Tu,Y., Kaplan,B.E. and Pfeiffer,G.P. (1996) *Biochemistry*, **35**, 15693–15703.
114 Roy,B. and Lee,A.S. (1995) *Mol. Cell. Biol.*, **15**, 2263–2274.
115 Greene,J.M., Larin,Z., Taylor,I.C.A., Prentice,H., Gwinn,K.A. and Kingston,R.E. (1987) *Mol. Cell. Biol.*, **7**, 3646–3655.
116 Hunt,C. and Calderwood,S. (1990) *Gene*, **87**, 199–204.
117 Bienz,M. and Pelham,H. (1987) *Cell*, **45**, 753–760.
118 Stewart,M.J., Dipple,K.M., Stewart,T.R. and Crabb,D.W. (1996) *Gene*, **173**, 155–161.
119 Ericsson,J., Jackson,S.M., Kim,J.B., Spiegelman,B.M. and Edwards,P.A. (1997) *J. Biol. Chem.*, **272**, 7298–7305.
120 Jackson,S.M., Ericsson,J., Orborne,T.F. and Edwards,P.A. (1995) *J. Biol. Chem.*, **270**, 21445–21448.
121 Osborne,T.F., Gil,G., Goldstein,J.L. and Brown,M.S. (1988) *J. Biol. Chem.*, **263**, 3380–3387.
122 Guan,G., Dai,P.-H., Osborne,T.F., Kim,J.B. and Sheckter,I. (1997) *J. Biol. Chem.*, **272**, 10295–10302.
123 Sato,R., Inoue,J., Kawabe,Y., Kodama,T., Takano,T. and Maeda,M. (1996) *J. Biol. Chem.*, **271**, 26461–26464.
124 Serra,E., Zemzoumi,K., Trolet,J., Capron,A. and Dissous,C. (1996) *Mol. Biochem. Parasitol.*, **83**, 69–80.
125 Schwarzbauer,M. (1997) Personal communication.
126 Swendemann,S.L., Spielholz,C., Jenkins,N.A. Gilbert,D.J., Copeland,N.G. and Sheffery,M. (1994) *J. Biol. Chem.*, **269**, 11663–11671.
127 Danilition,S.L., Frederickson,R.M., Taylor,C.Y. and Miyamoto,N.G. (1991) *Nucleic Acids Res.*, **19**, 6913–6922.
128 Quitschke,W.W., Lin,Z.-Y., DePonti-Zilli,L. and Paterson,B.M. (1989) *J. Biol. Chem.*, **264**, 9539–9546.
129 Chang,Z.F. and Liu,C.-J. (1994) *J. Biol. Chem.*, **269**, 17893–17898.
130 Arcot,S.S., Flemington,E.K. and Deininger,P.L. (1989) *J. Biol. Chem.*, **264**, 2343–2349.

131 Isaacs,R.J. Harris,A.L. and Hickson,I.D. (1996) *J. Biol. Chem*., **271**, 16741–16747.
132 Herzog,C.E. and Zwilling,L.A. (1997) *Biochem. Biophys. Res. Commun*., **232**, 608–612.
133 Ng,S.-W., Eder,J.P., Schnipper,L.E. and Chan,V.T.W. (1995) *J. Biol. Chem*., **270**, 25850–25858.
134 Adachi,N., Kobayashi,M. and Koyama,H. (1997) *Biochem. Biophys. Res. Commun*., **230**, 105–109.
135 Zwicker,J., Gross,C., Lucibello,F.C., Truss,M., Ehlert,F., Engeland,K. and Muller,R. (1995) *Nucleic Acids Res*., **23**, 3822–3830.
136 Zhu,H. (1997) Personal communication.
137 Shimizu,M., Ichikawa,E., Inoue,U., Nakamura,T., Nakajima,T., Nojima,H., Okoyama,H. and Oda,K. (1995) *Mol. Cell. Biol*., **15**, 2882–2892.
138 North,S., Espanel,X., Tavitian,B., Brun,G. and Gillet,G. (1996) *Cell Growth Differentiat*., **7**, 339–349.
139 Zwicker,J., Lucibello,F.C., Wolfraim,L.A., Gross,C., Truss,M., Engeland,K. and Muller,R. (1995) *EMBO J*., **14**, 4514–4522.
140 Huet,X., Rech,J., Plet,A., Vié,A. and Blanchard,J.M. (1996) *Mol. Cell. Biol*., **16**, 3789–3798.
141 Katula,K.S., Wright,K.L., Paul,H., Surnam,D.R., Nuckolls,F.J., Smith,J.F., Ting,J.P.-Y., Yates,J. and Cogswell,J.P. (1997) *Cell Growth Differentiat*., **8**, 811–820.
142 Farina,A., Tiainen,M., Manni,I., Cenciarelli,C., Bellorini,M., Mantovani,R., Sacchi,A. and Piaggio,G. (1997) *Mol. Cell. Biol*., submitted for publication.
143 van Ginkel,P.R., Hsiao,K.-M., Schjerven,H. and Farnham,P.J. (1997) *J. Biol. Chem*., **272**, 18367–18374.
144 Johnson,D.J., Ohtani,K. and Nevins,J.R. (1994) *Genes Dev*., **8**, 1514–1525.
145 Uchiumi,T., Longo,D.L. and Ferris,D.K. (1997) *J. Biol. Chem*., **272**, 9166–9174.
146 Filatov,D. and Thelander,L. (1995) *J. Biol. Chem*., **270**, 25239–25243.
147 DiLiberto,M., Lai,Z.-c., Fei,H. and Childs,G. (1989) *Genes Dev*., **3**, 973–985.
148 Lu,C.C. and Yen,T.S.B. (1996) *Virology*, **225**, 387–394.
149 Zachow,K.R. and Conklin,K.F. (1992) *J. Virol*., **66**, 1959–1970.
150 Goding,C.R., Temperley,S.M. and Fisher,F. (1987) *Nucleic Acids Res*., **15**, 7761–7780.
151 Huang,L., Malone,C.L. and Stinski,M. (1994) *J. Virol*., **68**, 2108–2117.
152 O'Rourke,D. and O'Hare,P. (1993) *J. Virol*., **67**, 7201–7214.
153 Moriuchi,M., Moriuchi,H. and Cohen,J.I. (1995) *Virology*, **214**, 256–258.
154 Gu,Z., Plaza,S., Perros,M., Cziepluch,C., Rommalaere,J. and Cornelis,J.J. (1995) *J. Virol*., **69**, 239–246.
155 Kato,M., Aoyama,A., Naruse,F., Kobayashi,T. and Tsukagoshi,N. (1997) *Mol. Gen. Genet*., **254**, 119–126.
156 Olesen,J., Hahn,S. and Guarente,L. (1987) *Cell*, **51**, 953–961.
157 Rosenkranz,M., Kell,C.S., Pennel,E.A. and Devenish,L.J. (1994) *Mol. Microbiol*., **13**, 119–131.
158 Trawick,J.D., Wright,R.M. and Poyton,R.O. (1989) *J. Biol. Chem*., **264**, 7005–7008.
159 Schneider,J.C. and Guarente,L. (1991) *Mol. Cell. Biol*., **11**, 4934–4942.
160 Bowman,S.B., Zaman,Z., Collison,L.P., Brown,A.J.P. and Dawes,I.W. (1992) *Mol. Gen. Genet*., **231**, 296–303.
161 Trueblood,C.E., Wright,R.M. and Poyton,R.O. (1988) *Mol. Cell. Biol*., **8**, 4537–4540.
162 Keng,T. and Guarente,L. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 9113–9117.
163 Dang,V.-D., Valens,M., Bolotin-Fukuhara,M. and Daignan-Fornier,B. (1996) *Mol. Microbiol*., **22**, 681–692.
164 Dang,V.-D., Bohn,C., Bolotin-Fukuhara,M. and Daignan-Fornier,B. (1996) *J. Bacteriol*., **178**, 1842–1849.
165 Flattery-O'Brien,J., Grant,C.M. and Dawes,I.W. (1997) *Mol. Microbiol*., **23**, 303–312.
166 Marziali,G. (1997) Personal communication.