Co-evolution from yeast to mouse: cDNA cloning of the two NF-Y (CP-1/CBF) subunits

Rob Hooft van Huijsduijnen*,1, Xiao Yan Li*,
Diane Black, Hans Matthes,
Christophe Benoist and Diane Mathis

Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Unité
184 de Biologie Moléculaire et de Génie Génétique de l'INSERM,
Institut de Chimie Biologique, Faculté de Médecine, 11, rue Humann,
67085 Strasbourg Cedex, France

*The first two authors both made major and equivalent contributions to
this work.
1Present address: Glaxo Institute for Molecular Biology SA, Route des
Acacias 46, 1211 Geneva 24, Switzerland

Communicated by D.Mathis

NF-Y is a CCAAT box binding protein critical for the
expression of diverse eukaryotic genes. We have purified
the A and B subunits of NF-Y and, using micro-
sequencing and PCR technology, have cloned the
corresponding cDNAs. NF-YA and NF-YB show stretches
of arresting sequence homology to the yeast transcrip-
tional activators HAP2 and HAP3. Unlike their yeast
counterparts, however, the two mouse subunits appear
necessary and sufficient for CCAAT binding activity. We
hypothesize that in the case of NF-Y, DNA binding and
dimerization both rely on residues within the homology
domains, which include rather suggestive 'half-finger'
motifs.
Key words: CCAAT box/cDNA cloning/NF-Y/PCR/tran-
scription factor

Introduction

The CCAAT box is a cardinal promoter element in several
eukaryotic protein coding genes. It usually resides ~80 bp
upstream from the RNA initiation site and may be present
in one or a few copies. A number of studies have
demonstrated the importance of CCAAT boxes to promoter
function, both for constitutively expressed genes and for
transcription units only active in terminally differentiated
cells (for reviews, see Maniatis et al., 1987; La Thangue
and Rigby, 1988). Their pleiotropic role includes the
regulation of transcription during development as well as
during the cell cycle (e.g. Collins et al., 1985; Gelinas
et al., 1985; Knight et al., 1987; van Wijnen et al., 1988).
Intriguingly, they also seem to be involved in controlling
transcriptional termination in adenovirus (Connelly and
Manley, 1989).

The identification of proteins that mediate CCAAT box
function has attracted considerable interest. The cloned
transcription factors NF-I and C/EBP have been shown to
recognize some CCAAT boxes (Santoro et al., 1988;
Landschulz et al., 1988; and references therein), but this
may only be because their true recognition sequences have
similarity to the CCAAT sequence (see Discussion). Several
other CCAAT box binding activities have also been

described. Among the best characterized are three factors
discovered in different laboratories, but which are perhaps
really the same entity: CBF, which has a high affinity for
the CCAAT sequence in and acts as a transcriptional
activator for the α2(I) collagen gene (Oikarinen et al., 1987;
Hatamochi et al., 1988; Maity et al., 1988); NF-Y, first
described as recognizing the Y box of major histocompati-
bility complex (MHC) class II genes (Dorn et al., 1987a,b;
Miwa et al., 1987; for review see Benoist and Mathis,
1990); and CP1, which binds to a CCAAT sequence in the
adenovirus major late promoter (Chodosh et al., 1988a).
These proteins have very similar binding specificities,
recognizing CCAAT boxes in a large number of other genes
including herpes simplex virus thymidine kinase, mouse
MHC class I, mouse albumin, and α and β globin (Dorn
et al., 1987b; Raymondjean et al., 1988; Chodosh et al.,
1988a). Indeed, an extensive compilation of promoter region
CCAAT boxes defines a consensus which is highly
reminiscent of the NF-Y binding site defined by saturation
mutagenesis (Bucher and Trifonov, 1988; Dorn et al.,
1987b). A number of less characterized CCAAT binding
factors may also be identical to NF-Y/CP1/CBF (Barberis
et al., 1987; Goding et al., 1987; Lichtsteiner et al., 1987;
Kim et al., 1988; van Wijnen et al., 1988; Mantovani et al.,
1989; others reviewed in Benoist and Mathis, 1990).

Most interestingly, Chodosh et al. (1988b) have shown
that CP1 is a heterodimer whose subunits are functionally
interchangeable with products of the Hap2 and Hap3 loci
in yeast. These loci control induction of the Cyc1 gene by
non-fermentable carbohydrates, and their products interact
with the controlling upstream activating sequence UAS2. The
CCAAT binding site for the HAP protein complex is
strikingly similar to a CCAAT box; mutation to CCAAT
actually enhances its binding (Oleson et al., 1987).

Because of its critical role in gene regulation and its curious
pleiotropic nature, it was imperative to elucidate the structure
of NF-Y. We have purified it to near-homogeneity, shown
that it is a two-chain molecule, sequenced peptides derived
from both chains, and cloned the corresponding cDNAs.
These encode proteins with regions of amazing homology
to HAP2 and HAP3, as well as stretches that resemble the
activation domains of other eukaryotic transcription factors.

Results

NF-Y purification
Our earlier studies indicated that NF-Y is a large multimer
of components of ~30 kd mol. wt (Hooft van Huijsduijnen
et al., 1987). That these components might be heterogeneous
was suggested by early purification attempts using ion-
exchange chromatography, wherein NF-Y subunits always
seemed to partition into different fractions (R.Hooft van
Huijsduijnen, X.Y.Li and A.Dorn, unpublished). We thus
adopted a strategy that employed some conventional
fractionation procedures (differential ammonium sulphate

precipitation, step elution from a heparin−agarose resin), but relied mainly on purification of the intact complex by affinity chromatography, as originally described by Kadonaga and Tjian (1986). We exploited NF-Y's specificity for several CCAAT motifs (Dorn *et al.*, 1987b) by passing extracts successively over affinity columns made with the MHC class II gene Y box and the albumin gene CCAAT box. This served to reduce the contamination by other proteins that fortuitously recognize CCAAT flanking sequences in one of the affinity matrices. Fractionation steps were monitored by the gel retardation assay and purity was assessed by SDS−PAGE.

As starting material, we chose nuclear extracts from the MHC class II-positive B cell lymphoma line CH27. Purification runs yielded ~2−4 μg of NF-Y from 200 mg of starting nuclear extract (from 3×10^{10} cells). Estimating a yield of 5−10%, this represents roughly a 5000-fold purification.

Figure 1 illustrates the last step of a successful purification run. Fractions were eluted from the albumin CCAAT box affinity column with increasing concentrations of KCl. As judged from the gel retardation assay shown in panel A, NF-Y binding activity eluted as a broad peak, centred around 0.9 M KC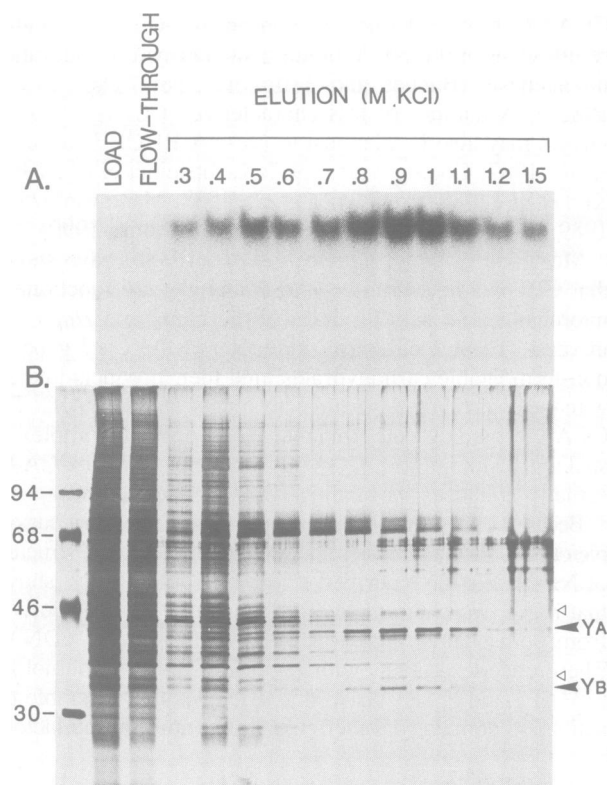l. In contrast, most of the proteins, visualized on the silver-stained gel in panel B, were not retained on the column or eluted in the first fractions. Two predominant bands did reproducibly appear in parallel with NF-Y activity: NF-YA and NF-YB of apparent mol. wt 40 and 32 kd, respectively. A few less intense bands also appeared in the NF-Y-containing fractions and may correspond to alternative forms or degradation products of NF-Y subunits (see below).

Several experiments were performed to verify that NF-YA and NF-YB actually partake in NF-Y activity. First, we verified that the fractions containing these two proteins gave rise to the same complex with the Y oligonucleotides as did unfractionated extract. Our criteria were electrophoretic mobility and methylation interference pattern (data not shown). Second, we applied partially purified material in parallel to two affinity columns—one made with the usual MHC class II Y box oligonucleotide, the other with an oligonucleotide carrying a three base mutation known to severely impede NF-Y binding (Figure 2A). The overall elution patterns were essentially identical, the only evident difference being that NF-YA and NF-YB were not retained on the mutant column. Third, we attempted to recover NF-Y binding activity from a size-fractionated SDS gel after renaturation (Figure 2B). None of the single fractions contained detectable activity (not shown, and − lanes in Figure 2B); however, binding was observed after mixing the 32 and 40−43 kd fractions. The former corresponded to the NF-YB band, while the 40−43 kd material probably included the NF-YA band and the minor one above it. The conclusion from these experiments was that NF-Y is composed of two different chains, one of which displays size

Fig. 1. Final steps in the purification of NF-Y. The figure displays data from the last affinity purification step in a successful run. Material was loaded onto a CCAAT affinity matrix, and retained proteins eluted at the indicated molarity of KCl. (A) Fractions were tested for activity as measured by the gel retardation assay using the 22mer MHC E_α gene Y-box oligonucleotide (Dorn *et al.*, 1987b). Only the retarded band is shown. (B) Silver-stained SDS−polyacrylamide gel loaded with the same fractions. Solid arrows indicate the NF-YA and B bands, open arrows point to minor species that co-eluted with NF-Y in some purification runs. The 60−65 kd bands present in the last five fractions are keratin contamination from the operator's fingers during the loading of the gel.

Fig. 2. Assessment of the heterodimeric character of NF-Y. (A) Semi-purified NF-Y was applied and eluted in parallel on affinity columns constructed with MHC class II Y-box oligonucleotides (left) or with oligonucleotides carrying a three base mutation in the Y-box (right). Fractions were analysed by SDS−PAGE and silver-staining. The position of the NF-Y bands in the left gel is indicated. M: marker proteins (sizes in kd); PC, FT: pre-column and flow-through fractions. (B) Reconstitution of NF-Y activity after SDS−PAGE separation. Protein was eluted from gel slices, renatured as described in Materials and methods and tested either alone or in combination with other samples for DNA binding in the gel retardation assay. Left panel: gel slice 10, corresponding to 32 kd protein tested alone (first lane) or mixed with other samples. Right panel: slice 4 (43 kd) tested similarly. The arrow indicates the position of the NF-Y/E_α−CCAAT oligonucleotide complex. The apparent mol. wt of protein in positive slices is indicated below the figure.

Table I. Peptide sequences

NF-YA	NF-YB
1. * * * * * * X G E G G R F F S P	1. X P T N I V Y K
	2. * * * * * * * * * * * * * X S F R E Q D I Y L P I A N V A R I M K
2. X F R V Q E L P L A	3. X R Y Q Q I S G V Q Q I Q F
	4. * * * X Y L Q K

Sequences obtained from peptides derived from NF-YA and NF-YB. X indicates the uncertain identity of the N-terminal amino acid. Asterisks denote homology to HAP2 (NF-YA) and HAP3 (NF-YB).

heterogeneity, either naturally or resulting from partial proteolysis during purification.

A similar heterodimeric structure was reported for CBF by Hatamochi *et al.* (1988), although the subunits appeared to have slightly different molecular weights. To confirm that CBF and NF-Y are identical, we performed a series of gel retardation and competition experiments. The data indicate that NF-Y and CBF from NIH 3T3 are indistinguishable, as judged by mobility on gels and by cross-competition with their respective binding templates (not shown).

Peptide sequences

NF-Y-containing fractions from several purifications were pooled, the material electrophoresed on an SDS−polyacrylamide gel and the separated proteins electroblotted onto a PVDF membrane. After staining, membrane fragments containing NF-YA and NF-YB were excised and incubated with the endoprotease Lys-C. Peptides released by this treatment were analysed and purified by reverse-phase HPLC (not shown); those which were in sufficient quantity and appeared sufficiently pure were subjected to Edman degradation microsequencing. The convincing sequences that we obtained are listed in Table I. We were immediately struck by the similarity between certain of the peptides and the HAP2 and HAP3 proteins from yeast (stars in Table I: Hahn *et al.*, 1988; Pinkham *et al.*, 1987). Given that the HAP2/3 subunits are functionally analogous to CP1 subunits (Chodosh *et al.*, 1988b) and that CP1 is probably identical to NF-Y, we took the sequence homology as strong evidence that the peptides indeed derive from NF-Y.

To verify this point, we produced anti-peptide antibodies. Two synthetic peptides were synthesized on the basis of the sequences in Table I (NF-YB peptides 2 and 3) and antibodies raised against them. As shown in Figure 3, serum against peptide 3 from NF-YB prevented the formation of NF-Y complex by the nuclear extract from CH27 cells. Instead, a more slowly migrating band appeared in the gel, presumably representing a complex composed of labelled DNA, NF-Y and bound antibodies. This complex was specific as its formation could be blocked by an excess of peptide 3 but not peptide 2 (Figure 3). Conversely, we have recently isolated a monoclonal antibody directed against peptide 2, which also blocks NF-Y binding to DNA (J.L.Pasquali, unpublished). Because of these results and of the aforementioned homologies to HAP2/3, we were confident that the peptide sequences derived from NF-Y and could be used as a basis for screening of cDNA clones.

NF-Y cDNA clones

After frustrating and fruitless failures to isolate NF-Y cDNA clones by screening libraries with degenerate oligonucleo-
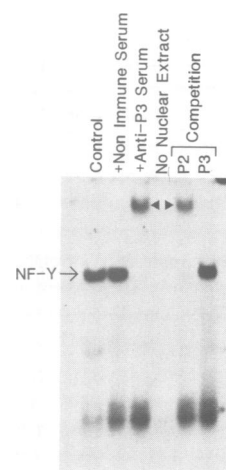
Fig. 3. Anti-peptide antibodies inhibit NF-Y activity. Gel retardation assays were performed with nuclear extracts from the B lymphoma line CH27 and the MHC Y-box oligonucleotide, after addition of non-immune mouse serum or of serum from mice immunized against 'peptide 3' from NF-YB (Table I; Materials and methods). Arrows point to the band which is further retarded in the presence of specific antibodies. In the two right-hand lanes, inhibition by the immune serum was prevented by the addition of an excess of peptide 3, but not of peptide 2.

tides, we turned to polymerase chain reaction (PCR) technology. Initial attempts following the direct approach described by Gonzalez *et al.* (1989) were also unsuccessful. We did, however, succeed in isolating NF-Y cDNA clones with the more elaborate strategies illustrated in Figure 4.

For NF-YB, we employed the approach of Gil *et al.* (1988) outlined in the right panel. The long sequence (19 amino acids) generated from peptide 2 allowed us to design short degenerate primers corresponding to the two extremities. Amplification between these two primers, using cloned cDNA as a template, yielded many fragments, one of which had the anticipated length of 53 bp. This fragment proved to encode the exact peptide 2 amino acid sequence, providing us with 25 bp of unambiguous nucleotide sequence [TATTTATCTTCCAATTGCAAATGTG, coding for the internal IYLPIANV peptide segment (Figure 4, right)]. From this information, non-degenerate probe was synthesized and cDNA clones were easily isolated.

For NF-YA, the short peptide sequences that were generated precluded the same approach. Instead, as outlined in the left panel, we relied on the observation that NF-YB has a long uninterrupted stretch of homology with HAP3 (see below). We hypothesized that this might also be true

A.

"Yeast Guessmers" G E G G R F F S P

K E S,F R E Q D I Y L P I A N V A R I M K

5'-TT$_C^T$AG$_G^A$CA$_G^A$CA$_G^A$CA - - - - - - ➤

◄ - - - - - CGITC$_C^T$TAGTACTT-5'

Polymerase chain reaction
Isolate fragments of defined length (HAP sequence)
Sequence (M&G)
Synthesize intervening oligonucleotide
Screen libraries

Polymerase chain reaction
Isolate 53bp fragment
Sequence (M&G)
Synthesize perfect central oligonucleotide
Screen libraries

B.

AUG   UAA

YA-EM56
YA-EM24
YA-EM13
YA-PB9
YA-PB7
YA-W22-1*
YA-W22-2*
YA-M12-1*
YA-M12-2*

AUG   UGA   An   An

YB-EM44
YB-EM38
YB-M12-11
YB-M12-21
YB-M12-61
YB-M12-62
YB-M12-63

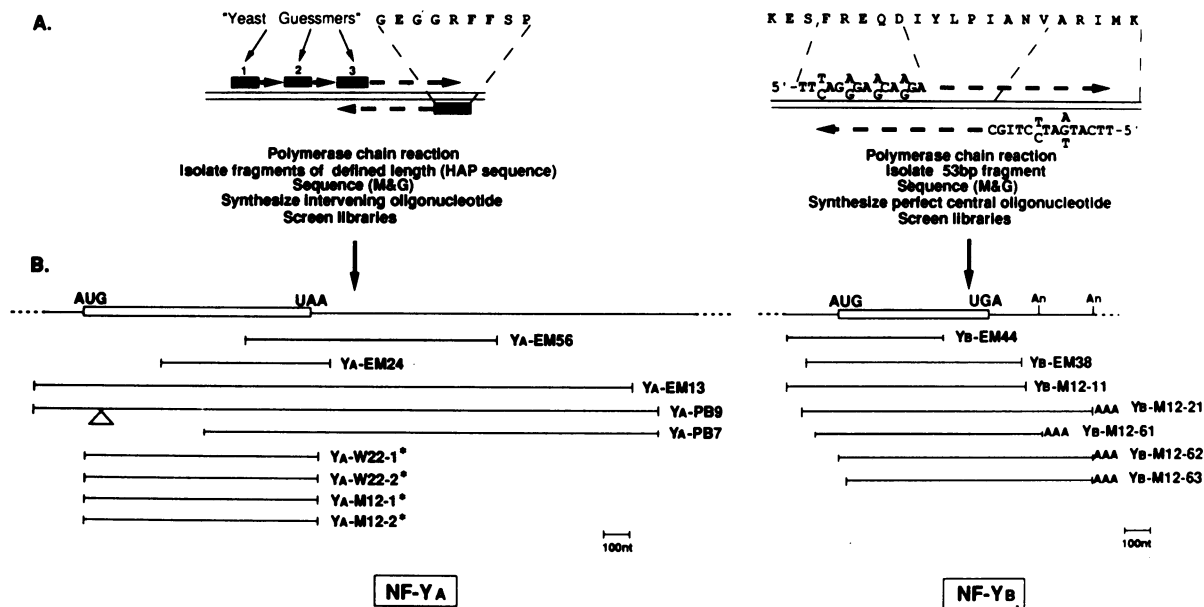100nt   100nt

NF-YA   NF-YB

Fig. 4. Cloning strategy and organization of cDNA clones for NF-YA and B. (A) Mixtures of oligonucleotides were synthesized based on the sequenced peptides as indicated, or (in the case of NF-YA) on a an anticipated similarity with the yeast *HAP2* sequence (yeast guessmers), using mammalian codon usage preferences (Lathe, 1985). These oligonucleotides, one set of which was end-labelled, were used in a polymerase chain reaction to amplify the intervening sequences, and the resulting material was sequenced. This information was used to synthesize perfect non-degenerate oligonucleotides that were used to screen cDNA libraries. (B) Orientation of the cDNA inserts with respect to the mRNA structure. The ORF is represented as a box delimited by the start and stop codons. Both mRNAs contain in-frame stop codons preceding the first AUG; no other AUGs in any frame are found in this region. 'An' indicates positions where polyadenylation occurs (AAA in the clones). The EM clones were derived from a 10 day old embryonic mouse cDNA library (B.Galliot), the PB and M12 series from cDNA libraries from P8 or M12.4.1 (pre-B or B lymphoma lines, respectively; (L.Leclerc, P.Dellabona); clones marked by an asterisk were obtained by PCR from M12 or WEHI-22 cDNA (B or T lymphomas, respectively).

of NF-YA and HAP2, and therefore the peptide 1 sequence, homologous to HAP2 (Table I), might be embedded in a longer block of homology. So we deduced PCR primer sequences from the amino acid sequence of peptide 1 and from the yeast amino acid sequence surrounding the peptide 1 homologous stretch ('yeast guessmers', Figure 4A). Amplification, using cloned cDNA as template, yielded a fragment of the expected size, as estimated from the HAP2 sequence. The sequence of this fragment revealed an open reading frame (ORF) encoding the peptide 1 sequence as well as other HAP2 related sequences. Once again, the unambiguous nucleotide sequence (120 bp between positions 902 and 1022, see Figure 5) permitted the ready isolation of NF-YA cDNA clones.

Schematic representations of the cDNA clones we obtained are found in Figure 4B. The clones derive from murine cDNA libraries of diverse origins (see legend to Figure 4). Sequence information was obtained from several independent clones, each analysed on both strands, and is presented in Figure 5. For NF-YA, the overlapping clones span a length of 2500 bp. A 5' untranslated region of at least 173 bp is followed by an ORF of 954 bp. One of the clones (PB9) contains an extra 81 bp of in-phase sequence, 75 bp within the ORF. This additional sequence almost certainly results from alternative splicing, since its 3' extremity includes a typical splice-acceptor motif. The protein encoded by this clone would have an extra 27 amino acids. Beyond the stop codon is what appears to be a long 3' untranslated region. The NF-YA clones do not include poly(A) or a poly-adenylation signal, indicating that the 3' untranslated sequences must continue beyond the internal *Eco*RI site which marks the end of our clones. For NF-YB, the clones span 1230 bp. There is an ORF of 621 bp, identical in all

cases. Several of the clones include poly(A) tails preceded by polyadenylation motifs. There are at least two distinct polyadenylation sites (Figure 4; the polyadenylation signals are underlined in Figure 5).

To verify that these cDNA clones did indeed correspond to NF-Y and that they covered its entire length, we performed *in vitro* transcription/translation. The coding regions of the NF-YA and NF-YB clones were introduced into plasmids carrying a T7 or T3 promoter. RNA was transcribed *in vitro*, the transcripts translated in a wheat-germ cell-free translation system and finally the products tested in gel retardation assays (Figure 6). Neither the NF-YA nor the NF-YB translation products could bind to the oligo-nucleotide probe to produce a retarded band, but together they produced a band that comigrated with NF-Y derived from a B cell lymphoma line. We are thus confident that the cDNA clones we have isolated encode both chains of NF-Y.

Protein sequences

The NF-YA and NF-YB clones code for proteins of 318 and 207 amino acids, respectively (Figure 5). We believe that these values represent the proteins in their entirety because the AUG codons are preceded by stop codons in all reading frames and because the same ORFs were found in a number of independent clones. The deduced molecular weights (23 and 34 kd) are lower than the apparent mol. wts estimated by SDS–PAGE (Figure 1). For NF-YB, this can be accounted for entirely by aberrant migration upon electro-phoresis, since protein produced by *in vitro* transcrip-tion/translation of the cDNA clones comigrates with native NF-YB from B cell extracts (not shown). For NF-YA, such aberrant migration can explain some but not all of the

NF-YA

```
...CTCGCAAAATGGCGGCGGCAGCGGCGGCTGGAGCCTCTGATTGGGTTTCGGGGTCCGGTACTGGAGCCAATCAGCGCGGGCAGCGAACCGGGGGAGCGAGGCACGGAGTCCCTACCTGGC

121 AGCTGCCTGGGATCTGTAGAGTGAAGCTTCAGGACTCTTAACGTGGCCGGGCCATGGAGCAGTATACGACAAACAGCAATAGTTCCACAGAGCAGATCGTGGTGCAGGCCGGCCAGATTC
                                                       M  E  Q  Y  T  T  N  S  N  S  S  T  E  Q  I  V  V  Q  A  G  Q  I

241 AGCAGCAGGTCCAGGGGCAGCCGTTAATGGTGCAAGTCAGTGGGAGGCCAGCTTATTACATCAACAGGCCAACCCATCATGGTACAGGCTGTGCCTGGTGGACAAGGCCAAACCATCATGC
     Q  Q  Q  V  Q  G  Q  P  L  M  V  Q  V  S  G  G  Q  L  I  T  S  T  G  Q  P  I  M  V  Q  A  V  P  G  G  Q  G  Q  T  I  M

361 AAGTACCTGTGTCTGGAACACAAGGTTTACAGCAGATACAGTTGGTGCCTCCTGGACAGATCCAGATTCAGGGTGGGCAGGCTGTGCAGGTGCAAGGCCAGCAGGGTCAAACCCAGCAGA
     Q  V  P  V  S  G  T  Q  G  L  Q  Q  I  Q  L  V  P  P  G  Q  I  Q  I  Q  G  G  Q  A  V  Q  V  Q  G  Q  Q  G  Q  T  Q  Q

481 TCATCATCCAACAGCCACAGACTGCAGTCACTGCTGGCCAGACTCAGACACAACAACAGATTGCTGTCCAGGGACAGCAAGTGGCCCAGACTGCTGAAGGACAGACTATTGTCTACCAAC
     I  I  I  Q  Q  P  Q  T  A  V  T  A  G  Q  T  Q  T  Q  Q  Q  I  A  V  Q  G  Q  Q  V  A  Q  T  A  E  G  Q  T  I  V  Y  Q

601 CAGTTAATGCAGATGGCACAATTCTCCAGCAAGTTACAGTCCCTGTTTCAGGCATGATCACCATCCCAGCAGCCAGTTTGGCAGGGGCACAGATCGTTCAGACAGGAGCCAATACCAACA
     P  V  N  A  D  G  T  I  L  Q  Q  V  T  V  P  V  S  G  M  I  T  I  P  A  A  S  L  A  G  A  Q  I  V  Q  T  G  A  N  T  N

721 CAACCAGCAGTGGACAAGGGACGGTCACTGTGACACTGCCTGTGGCAGGGAATGTGGTCAACTCAGGAGGAATGGTCATGATGGTACCAGGAGCTGGCTCTGTGCCTGCTATCCAAAGAA
     T  T  S  S  G  Q  G  T  V  T  V  T  L  P  V  A  G  N  V  V  N  S  G  G  M  V  M  M  V  P  G  A  G  S  V  P  A  I  Q  R

841 TCCCTTTACCTGGAGCAGAGATGCTGGAAGA|GAGCCCTGTATGTGAATGCCAAACAGTATCACCGCATCCTTAAGAGGAGACAAGCACGGGCTAAGCTAGAGGCAGAAGGGAAAATCC
     I  P  L  P  G  A  E  M  L  E  E|E  P  L  Y  V  N  A  K  Q  Y  H  R  I  L  K  R  R  Q  A  R  A  K  L  E  A  E  G  K  I

961 CAAAGGAACGAAGGAAATACCTCCATGAGTCTCGGCATCGGCACGCCATGGCACGGAAGCGTGGGGAAGGGGGCCGCTTCTTCTCTCCAAAAGAAAAGGACAGTCCTCACATGCAGGATC
     P  K  E  R  R  K  Y  L  H  E  S  R  H  R  H  A  M  A  R  K  R  G  E  G  G  R  F  F  S  F  K  E  K  D  S  P  H  M  Q  D

1081 CAAACCAAGCTGACGAAGAAGCCATGACACAGATCATCCGAGTTTCCTAACCACAGGAGGGTGGCAGAGCTGATCAGTCACGGTCCTCTCCACTGTTCTGGAAATGGACTGACTTCCAGT
     P  N  Q  A  D  E  E  A  M  T  Q  I  I  R  V  S  -

1201 GGGGCACTGACAGTCTCACTCTGCCCTTTCTCAGGACAGAAACTACTTAGCTCAGTATTACAGCTGCAATGATGGCTGACAAACTGAAGTGGCAAGCTTTTGTCGCACCCCTTCATTGAG

1321 GATCTGTTTCAGACTGTTGGACACTGACCTTTTGCAAATTTGGATGATTAAAAAGAGGTATTTGCCAGCCCAGCATTACAAACACAGCACTTATGTTGAGTGGTAAAGCCATCCAGTCAG

1441 CCATTGCAGTAGGGGAGAGCAGCCTTCTCAACCTGGTCCTCAGTCGAGGCAGCCACAGCCCCTGCCTCCCTGTGGGTTCTCAACAGCATCCATGCCCTGGTGGGGTGCCGGGCTGCCTGA

1561 TGCATAGACACTTGGCACTCAGGTGGCTCCCAAGCAAAGAAGCTTCCTGTTAGGACTTGGAGCAACAGACCAAGCTGCTGTCTTCTGACCTGTGCTGATAAGAATCTGTCTGCCCACATG

1681 CTAGGTTGGTGTTTACTTCAGCCAAGAAAGTACAAAAAGGATATTTGGACTCAGGATAAAAGGATAGATGGACCTAAAAGGTACATTTCTAGCTTTGACTGTGGTCTGCTCATTTCTATT

1801 AACAGCCATCTGGGTCATAGTTCATACTGTGGCCACTTTAGAGACAGTTTAGAGAGTAAAGAACAAAACTTCAAAACATCTGTGCTGAAGTGTTGAGGACATTCACAGTGGTTCTCCAAC

1921 AGAGGCGAACCTAGGGTATTCAATCATGGGCTTTATGACAGCAAGTAGCAAATGGAAGGATCTGTGCATTTCAATCTAGAGAGCAAGGGCCCTGCTGGCCCTAGTGCTGCTGTAGAGCGT

2041 GGTTCTCCAGTGGATGGATAGATACCAGGCATGACTCTTAGGCTGAAGAAACCAAGTCACATTTACATTGACACCTTCTGGCGTGCACTAAAAGAGAATCCTTATATTGCCTGAAGAAGC

2161 ACTGTTGAGCTAGCTGTCTTGGCTGCAGGCTGGAGCACTGCTGGAGCACAGAGTGGAGGAGAGAGATGTGCTTCATCCCTTGCCTGGGTGCTGACCCAGAGGTGAATGCTCTTCCCATGG

2281 CAGCTGATTCTCTGGTGCCAGTTTCCACTTGTTTTCTCCTTTCTCTGTAGGCCTTGCAGCAAGGTTTGCTCACTGCATCAGTGCTGTAGGGACCAACTACAGCCTGCAAGGATTCCTGCA

2401 GGGTCAGGAATTGGGCCTCAGAAATGTGTTTTCTGGACAAGCACTCTACCACTGAGCCACACACATACATACACACACACAC.....
```

NF-YB

```
...GCGGCCCGCGGCGGAGGGAGGGGAGGCCGAGTCCTGGAAGTGGAGCTCCGCGCTGGGACTGGTTCCTTCGCAGCCATTTTCTGTCCAACCAAACAGCCGATTGGAGACGGGAGCCAACCA

121 GGGCTGCATTGGAGGTTAAAATTATACAAGGATCCACCACCTTTTTGACAGGTATTTGAGAATACCACTGATACACAATACAGTATTTCATGACAATGGACGGCGACAGCTCCACCACAG
                                                                                        M  T  M  D  G  D  S  S  T  T

241 ATGCCTCCCAGCTAGGGATTTCTGCAGACTACATCGGAGGAAGTCATTATGTGATCCAGCCCCATGATGATACTGAGGACAGCATGAATGATCATGAAGATACAAATGGTTCAAAAGAAA
     D  A  S  Q  L  G  I  S  A  D  Y  I  G  G  S  H  Y  V  I  Q  P  H  D  D  T  E  D  S  M  N  D  H  E  D  T  N  G  S  K  E

361 GTTTCA|GAGAACAAGATATTTATCTTCCAATTGCAAATGTGGCTAGGATAATGAAAAATGCCATACCTCAAACAGGAAAGATCGCAAAAGATGCCAAAGAATGTGTTCAGGAATGTGTGA
     S  F|R  E  Q  D  I  Y  L  P  I  A  N  V  A  R  I  M  K  N  A  I  P  Q  T  G  K  I  A  K  D  A  K  E  C  V  Q  E  C  V

481 GTGAATTTATAAGCTTCATCACGTCGGAAGCAAGCGAAAGGTGTCATCAGGAGAAGCGGAAGACAATCAATGGGGAGGATATTCTGTTTGCCATGTCCACTCTAGGCTTTGACAGCTACG
     S  E  F  I  S  F  I  T  S  E  A  S  E  R  C  H  Q  E  K  R  K  T  I  N  G  E  D  I  L  F  A  M  S  T  L  G  F  D  S  Y

601 TGGAGCCTCTGAAACTGTACCTTCAGAAGTTCAG|GAGGCCATGAAAGGAGAGAAGGGTATCGGGGGAGCAGTCTCCGCCACAGATGGACTAAGCGAAGAGCTCACGGAAGAGGCATTCA
     V  E  P  L  K  L  Y  L  Q  K  F  R|E  A  M  K  G  E  K  G  I  G  G  A  V  S  A  T  D  G  L  S  E  E  L  T  E  E  A  F

721 CTAACCAGTTACCAGCTGGCTTAATAACTGCAGATGGTCAACAGCAAAACGTTATGGTTTACACAACGTCTTACCAACAGATTTCTGGTGTCCAGCAAATTCAGTTTTCATGATCGGAAG
     T  N  Q  L  P  A  G  L  I  T  A  D  G  Q  Q  Q  N  V  M  V  Y  T  T  S  Y  Q  Q  I  S  G  V  Q  Q  I  Q  F  S  -

841 GACGGACGGAACCGGGCTTGTGTGGAGAGAGTCTGATGGAGTCTGAAGCAGAGGCATGGGCGGAAGTGAGGGTGAAGCTGCCTCCTTGTACCATAAGTAGCTGTAATGTAGCTTCCTGAT

961 GCTTGACTAATTGAGGTGTTAATTCTGACTTGAGAATCTTTTTCATGAATGATTTTAAAAGAAAAAAAAATTGGATTTTAAAGGTATTAAAATATTTTTGTTTTGTACGAGAGTTTGTTG

1081 CTCTGTATGACGCCTGTATGCATTGTATATTGCAATTTATTACTGTCAGAGATTTGTAGACAGTTTCTTATTTTCATATTGAATCATGTTATTTGAATTCAAGTAACAGCTGGGTTAATT

1201 CATAATGTTTACCTTATAATATAGGGTAGAGTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

Fig. 5. Sequences of NF-YA and NF-YB. The sequences correspond to a composite from clones YA-EM56, YA-EM24, YA-EM13, YA-PB9, YA-M12-2 and YA-W22-1 (NF-YA; see Figure 4) and clones YB-EM44, YB-EM38, YB-M12-11 and YB-M12-62 (NF-YB), all of which were sequenced in their entirety. For the other clones, only the extremities were sequenced. Aside from the insertion in YA-PB9 (Figure 4), all sequences were identical. The HAP2 and HAP3 homologous regions are boxed. Broken lines mark acidic or glutamine-rich domains, and lines above the protein sequence indicate the positions of the peptide sequences originally obtained by microsequencing. The two polyadenylation signals in NF-YB are underlined.

discrepancy, hinting at post-translational modifications. Most of the peptide sequences from Table I were accounted for in the cDNA clones, with the exception of peptides NF-YA 2 and NF-YB 1. We assume that these must originate from contaminating proteins during the final stages of purification and sequencing.

Several structural features, highlighted in the diagrams of Figure 7, deserve special comment.

(i) Foremost are the striking sequence homologies between NF-YA and HAP2, and NF-YB and HAP3 (Figure 7). For NF-YA/HAP2, the region of homology is 57 amino acids long: 73% of the positions are identical and 7% have conservative replacements. For NF-YB/HAP3, the overall length is 90 amino acids; 66% of the residues are identical and 19% are conservative replacements. In both cases, identical positions tend to be grouped in clusters of 5−11 amino acids. The domains of mouse/yeast homology are colinear and their boundaries are sharp: NF-Y and HAP share no significant conservation of sequence in the flanking areas.

The homology regions are highly charged, with an equal number of acidic and basic residues for NF-YB but a strong preponderance of basic amino acids for NF-YA. These charges are well conserved between yeast and mouse.

(ii) Some areas of NF-Y are very reminiscent of activation domains found in other eukaryotic transcription factors (for review see Johnson and McKnight, 1989; Mitchell and Tjian, 1989). Particularly noteworthy are the glutamine-rich domains found in NF-YA (positions 18−30 and 72−132) and in NF-YB (at the C terminus) (Figures 5 and 7). These are characterized by a very high frequency of glutamine residues (up to 26/61 between positions 72 and 132 of NF-YA) and by an absence of charged residues. Interestingly, HAP2 also displays two glutamine-rich regions
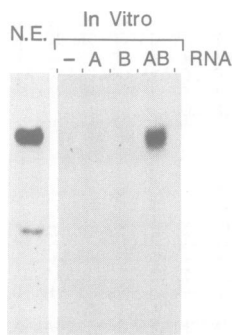
upstream of the homology domain but, other than the glutamines, there is no clear sequence similarity in this region.

(iii) NF-YB has cysteine residues in a conspicuous disposition: CX_3C (positions 85−89). NF-YA, on the other hand, has an HX_3H sequence (positions 271−275) (Figure 5). These motifs both occur in the yeast homology domains, and the cysteines and histidines are conserved between the two species. Such a disposition is, of course, highly reminiscent of the 'zinc fingers' found in a number of sequence specific DNA binding proteins ($CX_{2-4}CX_{1-2}\overset{H}{\underset{C}{}}X_{2-4}\overset{H}{\underset{C}{}}$; for refs, see Berg, 1986; Gibson *et al.*, 1988); more generally, it reminds one of metal binding domains in a variety of proteins ($CX_{2-4}C$ or $HX_{2-4}H$: for review and refs, see Berg, 1986; Frankel and Pabo, 1988).

Gene number and expression

To determine whether NF-Y might be a member of a multigene family, we hybridized coding region probes from the NF-YA and NF-YB cDNAs to Southern blots of digested genomic DNA from several species (Figure 8A). Very few bands were detected with either probe, indicating that NF-Y is not a member of a large family of closely related genes.

DNA binding activities similar or identical to NF-Y have been detected in a variety of cell types (Hooft van Huijsduijnen *et al.*, 1987). cDNA clones of identical sequence were derived from murine B, pre-B or T, lymphoma cells and 10 day embryos (Figure 5). To confirm the broad distribution of NF-Y, we hybridized cDNA probes to Northern blots of poly(A)$^+$ RNA from various sources (Figure 8B). As anticipated, NF-Y mRNA proved ubiquitous. The same bands were generated with RNA from all tissues, although band intensities did vary somewhat. The actively growing lymphoma lines had much higher levels of NF-YA and NF-YB mRNA than did mouse tissues composed largely of resting cells.

For NF-YA, a single RNA species was apparent at ~4.0 kb, a value in good agreement with the size of the largest cDNA clones (Figure 4). For NF-YB, two predominant species were detected: one of ~1.2 kb, which probably gave rise to our cDNA clones, and another of ~3.4 kb. We suspect that these large mRNAs result from alternative polyadenylation sites or splicing patterns in the 3' untranslated region, but have not yet been able to isolate the corresponding cDNA clones, even from libraries stringently selected for insert size.
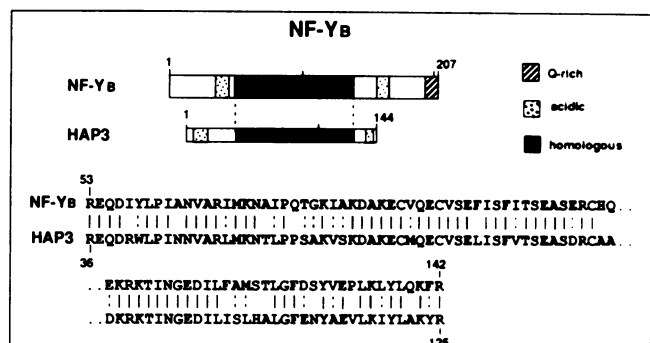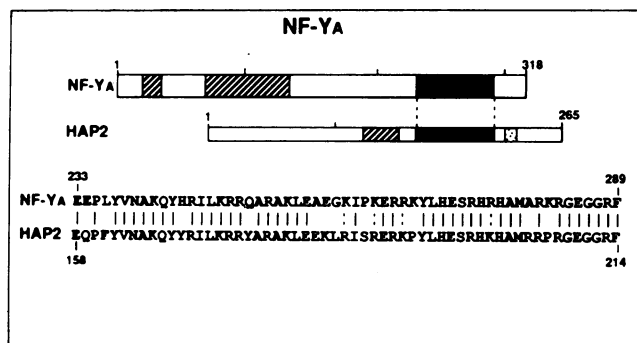

Fig. 6. The cDNA clones encode NF-Y, by *in vitro* transcription/translation. Gel retardation assays were performed with CH27 nuclear extract (N.E.), or with wheat-germ extract translation mixtures programmed by water (−), or by RNA transcribed *in vitro* from NF-YA (A) or NF-YB (B) cDNA clones, or a mixture thereof (AB).

Fig. 7. Comparison of NF-YA and -B with yeast HAP2 and HAP3. The hatched and dotted domains represent glutamine-rich and acidic stretches (>3/10 such residues). The black region shows the domains between NF-Y and HAP that display extensive sequence similarity, and their sequences are represented on the lower half of the figure. Bars denote identical residues, broken lines conserved positions.

Discussion

We have cloned and sequenced cDNAs coding for the A and B subunits of murine NF-Y. The identity of the clones is assured by two findings: first, anti-peptide antibodies were produced using the same sequences employed for cDNA clone isolation and these antibodies blocked NF-Y formation in CH27 nuclear extracts (Figure 3); second, the cDNA sequences were subcloned into the appropriate vectors for *in vitro* transcription/translation and the products so generated had substantial NF-Y activity (Figure 7). We believe that our clones also correspond to the independently described transcription factors CP1 and CBF: CP1 subunits are known to complement HAP2/3 subunits for DNA binding, and we find extensive homology between NF-Y and HAP2/3; CBF and NF-Y are indistinguishable in retardation and competition assays.

NF-Y and HAP2/3

The most striking finding that emerges from the sequence analysis is the presence of domains in the NF-Y subunits that are highly homologous to regions of the HAP2 and HAP3 transcription factors from *Saccharomyces cerevisiae*. Homology is restricted to a single domain for each subunit (57 or 90 amino acids) and within the domains, amino acid identities between the yeast and mouse proteins are extremely frequent (66 and 75%; Figure 7). This high degree of homology far exceeds the similarities previously noted between domains of the 'homeo' family (for review see Scott *et al.*, 1989) or between GCN4 and members of the fos/jun family of transcription factors (Vogt *et al.*, 1987; for review and refs, see Busch and Sassone-Corsi, 1990). Barring convergent evolution, which seems intuitively unlikely, the superimposable domains of NF-Y and HAP must trace back to a common ancestor that existed before the evolution of

Fig. 8. NF-Y gene number and expression. (**A**) Southern blots: mouse, hamster or human DNA was digested with *Bam*HI (B), *Eco*RI (E) or *Hind*III (H) and probed with labelled NF-YA (left) and NF-YB (right) cDNA (YA-EM13 and YB-EM38, respectively; Figure 4). (**B**) Northern blots with 5 μg poly(A)⁺ RNA from CH1, a B cell line (B); W22, a T cell line (T); liver (Li); mammary gland (Mg); salivary glands (Sg); intestine (In); and colon (Co). Probes as in A. Size markers are in kb.

multicellular eukaryotes. The evolutionary process seems to have taken some interesting turns. First of all, intron elimination must have taken place in yeast (or intron insertion in the mouse) because preliminary analysis shows that the homology domain in NF-YB is split into two by an intron (R.Hooft van Huijsduijnen, unpublished). Another difference between yeast and mouse is that the *HAP3* gene generates antisense transcripts with an ORF although they are of unknown function (Hahn *et al.*, 1988). There are no antisense transcripts from the NF-YB gene, and the corresponding ORF is closed (R.Hooft van Huijsduijnen, unpublished). The acquisition of antisense transcripts in yeast is probably a late event, as intron elimination would hardly be compatible with the conservation of ORFs of different polarities.

Another interesting evolutionary turn concerns HAP4. Forsburg and Guarente (1989) have convincingly shown that that the HAP2/3 complex in yeast actually contains a third component, HAP4. This additional protein is necessary for the formation of and is included in the complexes detected in gel retardation assays *in vitro*; moreover, it is indispensable for transcriptional activation of the cytochrome genes controlled by HAP2/3. However, several arguments lead us to believe that mammalian cells do not contain an individualized equivalent of HAP4: two fractions are necessary and sufficient to restore NF-Y activity after high resolution preparative electrophoresis of purified NF-Y (Figure 2; see also Hatamochi *et al.*, 1988); two fractions were also all that was needed to generate CP1 activity after column purification (Chodosh *et al.*, 1988a); and RNAs transcribed from NF-YA and NF-YB cDNAs are necessary and sufficient for the generation of NF-Y DNA binding activity after *in vitro* translation. Each of these points would not be definitive in and of itself—one could always argue comigration during electrophoresis, copurification on columns, or that the wheat germ extract provides a HAP4 equivalent—but we feel that, together, these points make a convincing case for the absence of an individualized HAP4-like entity in mammalian cells. HAP4 function may have been somehow incorporated into NF-YA and NF-YB, perhaps accounting for the larger size of the mouse proteins relative to their yeast counterparts, but we have not detected significant sequence similarities between HAP4 and NF-YA or B. Alternatively, the requirement for HAP-4 might be specific to yeast, and its activity simply dispensable for NF-Y function.

The boundaries of the mouse/yeast homology domains are very sharp. We have detected no obvious sequence resemblance in the flanking regions of NF-Y and HAP2/3, aside from a similar disposition of glutamine-rich and acidic stretches. This implies that the homology domains concentrate the residues required for protein−protein interactions as well as for specific DNA binding, and that both functions have been stringently conserved. Interestingly, the homology domains contain the CX_3C and HX_3H motifs, very similar to metal binding blocks in a variety of proteins. As NF-Y is a metallo-protein (Hooft van Huijsduijnen *et al.*, 1987), it seems reasonable to suggest that the 'half-fingers' we observe are involved in metal binding. One might actually hypothesize a tetrahedral metal coordination complex made of the CX_3C of NF-YB and the HX_3H found in NF-YA. The metal ion coordinated by the cysteines and histidines from both chains would then be the core of the protein/protein interface. Such metal-mediated interactions have
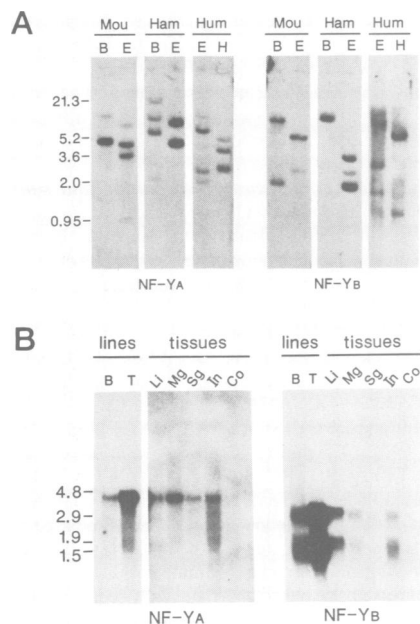
previously been proposed for dimerization of the tat protein of HIV-1 and for the interaction of CD4 and CD8 with p56lck (Frankel *et al.*, 1988; Turner *et al.*, 1990). Interestingly, the CX_3C motif in NF-YB is embedded within a stretch of internally repetitive amino acid sequence AKDAKECVQECVSEFISFITSEASER which, based on the 3−4 amino acid periodicity of its repeats, looks unmistakably like an amphipathic α-helical stretch. Although leucines are conspicuously absent, this brings to mind the dimerization of 'leucine-zipper' proteins through parallel α-helices (see Vinson *et al.*, 1989; Busch and Sassone-Corsi, 1990; and refs therein).

NF-Y and other transcription factors
Aside from the 'half-fingers' discussed above, the homology domains of NF-YA and NF-YB do not include any of the motifs common to other DNA binding proteins. There are no zinc fingers, leucine-zippers or helix-loop-helix motifs. Neither is there a clear example of the helix-turn-helix motif frequently found in prokaryotic and eukaryotic regulatory factors (following the loose consensus defined by Johnson and McKnight, 1989).

We did not observe any resemblance between NF-Y and NF-1 or C/EBP, as was previewed by the lack of similarity between HAP2/3 and these factors (Landschulz *et al.*, 1988; Santoro *et al.*, 1988). This is not really so surprising given that the three mammalian factors actually have rather different binding specificities. NF-Y has an absolute requirement for the CCAAT motif (Dorn *et al.*, 1987b). C/EBP has a rather loose specificity, that includes GCAAT but also the enhancer core motif $GTGG^{AAA}_{TTT}G$ (Graves *et al.*, 1986; Johnson *et al.*, 1987). NF-I has the palindromic recognition sequence $TTGGCTN_3AGCCAA$ that sometimes overlaps a CCAAT box (Gronostajski, 1986; Jones *et al.*, 1987). There is thus no reason to expect that these three proteins would be related, nor is there any reason to invoke families of CCAAT binding transcription factors.

Perhaps more relevant are NF-Y* and CP2, CCAAT box binding proteins with specificities very similar to yet demonstrably different from NF-Y and CP1 (Dorn *et al.*, 1987b; Chodosh *et al.*, 1988a). One might have expected that these are members of a gene family that includes NF-Y. Yet hybridizing Southern blots of mouse DNA with NF-YA and NF-YB probes revealed only very few bands, even at hybridization stringencies which easily allow detection of the human homologues (Figure 8A; preliminary data indicates that human and murine NF-Y share ~85% identity at the nucleotide level—R.Mantovani, unpublished). Although further analyses, including cDNA library screenings at lower stringencies, are still necessary, one can already conclude that there is no large family of genes closely related to NF-Y.

Materials and methods

Protein purification
Nuclear extracts were prepared from 30 l batches of spinner cultured CH27 cells as described (Dorn *et al.*, 1987b), except that the nuclear pellets were disrupted, upon resuspension in 1 M NaCl buffer, with a Polytron homogenizer. The protein in the supernatant was precipitated overnight in 60% saturated ammonium sulphate, pelleted by centrifugation (30 min 46 000 g), resuspended in 100 ml of 25 mM HEPES pH 7.9, 2 mM $MgCl_2$, 0.1 mM EDTA, 25% glycerol, 0.5 mM DTT and 40 mM KCl (buffer Z), and dialysed against the same buffer. The solution was clarified by centrifugation at 46 000 g and passed twice over a heparin A4R (IBF)

column. The column was washed and eluted with 100 ml portions of buffer Z supplemented with 0.1, 0.2, 0.3, 0.4, 0.5 and 0.7 M KCl. The NF-Y containing fractions (usually eluting at 0.3 and 0.4 M KCl as assayed by the gel retardation assay) were collected, dialysed against buffer Z (Kadonaga and Tjian, 1986) supplemented with 0.5 mM phenylmethylsulphonylfluoride (PMSF) and passed over a Y-box affinity column. The oligonucleotides used for this column were the E_α 22mer oligos (Dorn *et al.*, 1987b) plus a 5' TGAAAC addition for the coding and a 5' GTTTCA extension for the non-coding strand. The oligos were concatenated and linked to CNBr-activated Sepharose 4B (Pharmacia) as described [Kadonaga and Tjian (1986) ~50 µg DNA per 1 ml resin]. The column was washed at 0.3 M KCl and eluted at 1 M KCl. The sample was dialysed back to 0.1 M KCl in buffer Z/PMSF and passed over an albumin CCAAT box affinity column (GGGTAGGAACCAATGAAATGAAAGCAGGT and its complement, with the same extensions as the E_α-CCAAT box oligos). This second affinity column was washed at 0.1 M KCl and eluted stepwise with one column-volume of buffer Z/PMSF supplemented with 0.1−1.2 M KCl (0.1 M increments). NF-Y elution was monitored by gel retardation assays, and the pooled fractions were dialysed to 40 mM KCl. The mutated Y-box affinity column was constructed using the oligonucleotides ATTTTTCTG-AGGTGTTAAAAGT and its complement, carrying the same 5' extensions described above.

Peptide generation and microsequencing
The active fractions from five to eight purification batches were acetone precipitated (4 vols), run into a denaturing protein gel, and transferred in 20 min at 1.25 mA/cm^2 to a polyvinylidenedifluoride membrane (Millipore) in 10 mM CAPS, 10% methanol pH 11. The protein was stained for 5 min in 0.2% Ponceau S in 3% TCA, rinsed with water, destained 5−10 min in 200 µM NaOH, and rinsed in water. Strips with protein were digested overnight at 37°C with endoproteinase Lys-C (WAKO; 1 µg enzyme per 10 µg estimated substrate) in 10% acetonitrile, 0.2% hydrogenated Triton X-100 in 50 mM Tris−HCl pH 9.5 (M.Moos, personal communication). The resulting peptides were injected on a C8 Brownlee RP-300 2.1 × 30 mm column and eluted with a gradient of 0−50% acetonitrile in 0.085% TFA. Selected pure peptides were sequenced on an Applied Biosystems 4770 protein sequencer, using standard cycles supplied by the manufacturer.

Antiserum preparation and gel retardation interference
The synthetic oligopeptide TRYQQISGVQQC (P3) was conjugated to BSA via its C-terminal sulphide group (Rothbard *et al.*, 1984) and injected i.p. in female SJLxA/J mice at 3 week intervals (200 µg with complete Freund's adjuvant for the first, 100 µg with incomplete adjuvant for the following injections). Mice were bled 1 week after the boosts. For interference with the gel retardation assay, 4 µl serum was preincubated 5 min on ice with 2 µl PBS or peptide [the antigen itself or ESFREQDIYLC (P2) as a control, both at 0.1 mg/ml], then 2 µl M12 nuclear extract was added for another 25 min. Finally, the other components (buffer, poly d[IC] and labelled oligonucleotide) were added and loaded onto the gel after a further 20 min.

Polymerase chain reactions, library screenings and sequencing
One hundred pmoles oligonucleotide were end-labelled with 200 µCi (7.4 MBq) [γ-^{32}P]ATP in 40 µl final volume, and after heat-inactivation of the kinase the other oligonucleotide primer and all other components of the PCR were added (Gil *et al.*, 1988). The substrate for the reaction was 0.5 µg of a CsCl purified CDM8 plasmid cDNA library derived from the B/lymphoma M12.4.1 (>500 bp inserts; a kind gift from P.Dellabona). The NF-YA oligonucleotide GGGGAGGAAGAACC^T_CGCCCCCCTCNCC deduced from peptide P1 was used for amplification with the HAP2 derived oligonucleotides GTNAAYGCNAARCANTA, AARCCNTAYYTNCAYGA (failed to amplify) and (A/C)GNCAYAARCAYGCNATG, in 35 cycles of PCR (1 min 94°C, 1.5 min 55°C and 2.5 min 72°C). The NF-YB oligonucleotides TTYAGRGARCARGA and TTCAT(A/G/T)ATYTIGC were similarly used under the conditions described by Gil *et al.* (1988). Three other combinations of oligonucleotides that were based on different arginine codons were empirically found to be much less efficient. The PCR products were precipitated, run on acrylamide gels and major products or products of the expected size eluted and chemically sequenced (Maxam and Gilbert, 1980).

Libraries were plated and screened by standard methods; M13 subclones were sequenced by the dideoxynucleotide technique (Sanger *et al.*, 1980).

Northern and Southern hybridization
Poly(A) RNA samples were denatured in formamide/formaldehyde and run on agarose gels containing 0.37 M formaldehyde, transferred to Hybond N (Amersham), stained with methylene blue and hybridized with randomly primed (Feinberg and Vogelstein, 1984) cDNA insert in 50% formamide at 42°C. Final washing was at 65°C in 0.1 × SSC/0.1% SDS.

DNA samples (10 μg) were digested overnight, run into a 1% agarose gel, stained, depurinated, denatured and transferred by Hybond N plus. UV crosslinking (Amersham protocol) hybridization and washing were as above, except that final washing (1 h) was at 55°C.

In vitro transcription/translation

The *Eco*RI digested cDNA fragments for NF-YA and B (fragments YA-EM13 and YB-EM38) were subcloned into pBluescriptII or pGEM-1, linearized with *Xho*I and *Ava*I, and 2 μg transcribed by RNA polymerase T3 or T7 (Promega), respectively, in the presence of 0.5 mM cap analogue in 20 μl final volume. The transcripts were purified from 'low-melt' agarose gels, and translated in a wheat-germ extract (Promega). Of this 4 μl was chilled on ice and further incubated for 20 min with 2 μl binding buffer, polyd[IC] and radiolabelled Y-box oligonucleotide and run on a 4% non-denaturing polyacrylamide gel (Dorn et al., 1987b). Radiolabelled protein was obtained by performing translations with [^{35}S]methionine, 2 μl of which was run in SDS–polyacrylamide gels.

References

Barberis,A., Superti-Furga,G. and Busslinger,M. (1987) *Cell*, **50**, 347–352.
Benoist,C. and Mathis,D. (1990) *Annu. Rev. Immunol.*, **8**, 681–715.
Berg,J.M. (1986) *Science*, **232**, 485–487.
Bucher,P. and Trifonov,E.N. (1988) *J. Biomol. Struct. Dyn.*, **5**, 1231–1242.
Busch,S.J. and Sassone-Corsi,P. (1990) *Trends Genet.*, **6**, 36–38.
Chodosh,L.A., Baldwin,A.S., Carthew,R.W. and Sharp,P.A. (1988a) *Cell*, **53**, 11–17.
Chodosh,L.A., Olesen,J., Hahn,S., Baldwin,A.S., Guarente,L. and Sharp,P.A. (1988b) *Cell*, **53**, 25–31.
Collins,F.S., Metherall,J.E., Yamakawa,M., Pan,J., Weissman,S.M. and Forget,B.G. (1985) *Nature*, **313**, 325–326.
Connelly,S. and Manley,J.L. (1989) *Cell*, **57**, 561–569.
Dorn,A., Durand,B., Marfing,C., Lemeur,M., Benoist,C. and Mathis,D. (1987a) *Proc. Natl. Acad. Sci. USA*, **84**, 6249–6253.
Dorn,A., Bollekens,J., Staub,A., Benoist,C. and Mathis,D. (1987b) *Cell*, **50**, 863–870.
Feinberg,A. and Vogelstein,B. (1983) *Anal. Biochem.*, **132**, 6–11.
Forsburg,S.L. and Guarenet,L. (1989) *Genes Dev.*, **3**, 1166–1172.
Frankel,A.D. and Pabo,C.O. (1988) *Cell*, **53**, 675–677.
Frankel,A.D., Bredt,D.S. and Pabo,C.O. (1988) *Science*, **240**, 70–72.
Gelinas,R., Endlich,B., Pfeiffer,C., Yagi,M. and Stamatoyannopoulos,G. (1985) *Nature*, **313**, 323–325.
Gibson,T.J., Postma,J.P.M., Brown,R.S. and Argos,P. (1988) *Prot. Eng.*, **2**, 209–214.
Gil,G., Smith,J.R., Goldstein,J.L., Slaughter,C.A., Orth,K., Brown,M.S. and Osborne,T.F. (1988) *Proc. Natl. Acad. Sci. USA*, **35**, 8963–8967.
Goding,C.R., Temperley,S.M. and Fisher,F. (1987) *Nucleic Acids Res.*, **15**, 7761–7773.
Gonzalez,G.A., Yamamoto,K.K., Fischer,W.H., Karr,D., Menzel,P., Biggs III,W., Vale,W.W. and Montminy,M.R. (1989) *Nature*, **337**, 749–751.
Graves,B.J., Johnson,P.F. and McKnight,S.L. (1986) *Cell*, **44**, 565–571.
Gronostajski,R.M. (1986) *Nucleic Acids Res.*, **14**, 9117–9129.
Hahn,S., Pinkham,J., Wei,R., Miller,R. and Guarente,L. (1988) *Mol. Cell. Biol.*, **8**, 655–662.
Hatamochi,A., Golumbek,P.T., van Schaftingen,E. and de Crombrugghe,B. (1988) *J. Biol. Chem.*, **263**, 5940–5945.
Hooft van Huijsduijnen,R.A.M., Bollekens,J., Dorn,A., Benoist,C. and Mathis,D. (1987) *Nucleic Acids Res.*, **15**, 7265–7273.
Johnson,P.F. and McKnight,S.L. (1989) *Annu. Rev. Biochem.*, **58**, 799–848.
Johnson,P.F., Landschulz,W.H., Graves,B.J. and McKnight,S.L. (1987) *Genes Dev.*, **1**, 133–140.

Jones,K.A., Kadonaga,J.T., Rosenfeld,P.J., Kelly,T.J. and Tjian,R. (1987) *Cell*, **48**, 79–87.
Kadonaga,J.T. and Tjian,R. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 5889–5893.
Kim,C.G., Barnhart,K. and Sheffery,M. (1988) *Mol. Cell. Biol.*, **8**, 4270–4277.
Knight,G.B., Gudas,J.M. and Pardee,A.B. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 8350–8353.
Landschulz,W.H., Johnson,P.F., Adashi,E.Y., Graves,B.J. and McKnight, S.L. (1988) *Genes Dev.*, **2**, 786–792.
La Thangue,N.B. and Rigby,P.W.J. (1988) In Hames,B.D. and Glover,D.M. (eds), *Frontiers in Molecular Biology*. IRL Press, Oxford, pp. 1–42.
Lathe,R. (1985) *J. Mol. Biol.*, **183**, 1–8.
Lichtsteiner,S. Wuarin,J. and Schibler,U. (1987) *Cell*, **51**, 963–973.
Maity,S.N., Golumbek,P.T., Karsenty,G. and de Crombrugghe,B. (1988) *Science*, **241**, 582–584.
Maniatis,T., Goodbourn,S. and Fischer,J.A. (1987) *Science*, **236**, 1237–1241.
Mantovani,R., Superti-Furga,G., Gilman,J. and Ottolenghi,S. (1989) *Nucleic Acids Res.*, **17**, 6681–6686.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–548.
Mitchell,P.J. and Tjian,R. (1989) *Science*, **245**, 371–375.
Miwa,K., Doyle,C. and Strominger,J.L. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 4939–4943.
Oikarinen,J., Hatamochi,A. and de Crombrugghe,B. (1987) *J. Biol. Chem.*, **262**, 11064–11070.
Olesen,J., Hahn,S. and Guarente,L. (1987) *Cell*, **51**, 953–960.
Pinkham,J.L., Olesen,J.T. and Guarente,L.P. (1987) *Mol. Cell. Biol.*, **7**, 578–587.
Raymondjean,M., Cereghini,S. and Yaniv,M. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 757–761.
Rothbard,J.B., Fernandez,R. and Schoolnik,G.K. (1984) *J. Exp. Med.*, **160**, 208–216.
Sanger,F., Coulson,A.R., Barrell,B.G., Smith,A.J.H. and Roe,B.A. (1980) *J. Mol. Biol.*, **143**, 161–171.
Santoro,C., Mermod,N., Andrews,P.C. and Tjian,R. (1988) *Nature*, **334**, 218–222.
Scott,M.P., Tamkun,J.W. and Hartzell III,G.W. (1989) *Biochim. Biophys. Acta*, **989**, 25–58.
Turner,J.M., Brodsky,M.H., Irving,B.A., Levin,S.D., Perlmutter,R.M. and Littman,D.R. (1990) *Cell*, **60**, 755–762.
van Wijnen,A.J., Massung,R.F., Stein,J.L. and Stein,G.S. (1988) *Biochemistry*, **27**, 6534–6539.
Vinson,C.R., Sigler,P.B. and McKnight,S.L. (1989) *Science*, **246**, 911–914.
Vogt,P.K., Bos,T.J. and Doolittle,R.F. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 3316–3320.