THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 1992 by The American Society for Biochemistry and Molecular Biology, Inc.

Vol. 267, No. 13, Issue of May 5, pp. 8984-8990, 1992
Printed in U.S.A.

Intron-Exon Organization of the NF-Y Genes

TISSUE-SPECIFIC SPLICING MODIFIES AN ACTIVATION DOMAIN*

(Received for publication, October 10, 1991)

Xiao-Yan Li‡, Rob Hooft van Huijsduijnen§, Roberto Mantovani¶, Christophe Benoist, and Diane Mathis

From the Laboratoire de Génétique Moléculaire des Eucaryotes du Centre National de la Recherche Scientifique, Unité 184 de Biologie Moléculaire et de Génie Génétique de l'Institut National de la Santé et de la Recherche Médicale Institut de Chimie Biologique, Faculté de Médecine, 67085 Strasbourg Cedex, France

NF-Y is a highly conserved transcription factor which recognizes CCAAT motifs in a variety of genes. We report here the genomic organization of the genes encoding both subunits, which gives interesting clues about the functional organization of the proteins. We also report the existence of isoforms of NF-YA which result from differential splicing. These alternative splicing events map within a glutamine-rich activation domain and show a marked cell- and tissue-specific bias.

NF-Y is a eukaryotic transcription factor that binds with high specificity to CCAAT motifs in the promoter regions of genes transcribed by RNA polymerase II(B). Its characteristic binding site (typically 5′$^{AA}_{GG}$CCAATCAG ... 3′) is found in direct or inverted orientation in the 5′ promoter region of a wide variety of genes: *e.g.* the albumin, α-globin, β-actin, α-collagen, or histone genes and the class II genes of the major histocompatibility complex (MHC)[1] (Barberis *et al.*, 1987; Dorn *et al.*, 1987; Dutta *et al.*, 1990; Hatamochi *et al.*, 1988; Knight *et al.*, 1987; Quitschke *et al.*, 1989; Raymondjean *et al.*, 1988; van Wijnen *et al.*, 1988; Wuarin *et al.*, 1990). It is certainly identical to CP1 (Chodosh *et al.*, 1988a) and CBF, the latter shown by de Combrugghe's group to be involved in transcriptional initiation of α-collagen genes (Maity *et al.*, 1988); NF-Y also probably corresponds to a number of less characterized CCAAT binding factors (for references, see Hooft van Huijsduijnen *et al.*, 1990). In keeping with its widespread influence, NF-Y is a ubiquitous factor, present in all cell types tested, including undifferentiated teratocarcinomas (Hooft van Huijsduijnen *et al.*, 1987).

NF-Y is composed of two subunits (referred to as NF-YA and NF-YB), both of which are required for DNA binding (Chodosh *et al.*, 1988a; Hatamochi *et al.*, 1988; Hooft van Huijsduijnen *et al.*, 1990). We have recently cloned the cDNAs coding for both subunits in the mouse (Hooft van Huijsduijnen *et al.*, 1990). Vuorio *et al.* (1990) and Maity *et al.* (1990) also reported the cloning of the rat NF-Y and Becker *et al.* (1991) cloned a portion of human NF-YA by complementation in yeast. Sequence analysis of the clones confirmed the suspected analogy, suggested by Chodosh *et al.* (1988b), between the NF-Y subunits and the transcription factors HAP2 and HAP3 from *Saccharomyces cerevisiae*: highly conserved domains (~70% homology) were shared between NF-YA and HAP2 and between NF-YB and HAP3.

In the course of analyzing our cDNA clones, we noticed the existence of several different types of NF-YA cDNAs. To understand how these isoforms arise and to explore the possibility of NF-YB isoforms, we have analyzed genomic clones for both subunits. We now demonstrate that the NF-YA isoforms result from differential splicing; interestingly, one of the differential splicing events shows a strong tissue-specific bias. We have followed several approaches to establish the functional characteristics of these various isoforms.

MATERIALS AND METHODS

PCR Amplification and Library Screenings—Poly(A$^+$) RNA was prepared from a variety of cell lines or mouse tissues, essentially as described (Koch *et al.*, 1989). The RNA was reverse-transcribed and amplified by PCR under standard conditions, using primers mapping just outside the NF-YA coding sequences (5′-CTAGGAATTCAGATCTGTAGAGTGAAGCTTCAGG-3′) and 5′-CTAGGAATTCAGATCTCCCCACTGAAGTCAGTCC-3′) and incorporating artificial cloning sites. PCR products were cut with *Eco*RI and cloned into pBluescript (Stratagene, La Jolla, CA). cDNA and genomic λ libraries were screened according to standard protocols, using cDNA fragments from clones YA-EM13 and YB-EM38 (Hooft van Huijsduijnen *et al.*, 1990).

S1 Nuclease Mapping—Relative amounts of short and long form NF-YA mRNA were determined by quantitative S1 nuclease analysis, performed as described (Koch *et al.*, 1989). The probe was a single-stranded complementary DNA fragment from long form NF-YA, 5′ end-labeled (at position 235) and extending across the variable element and into the plasmid vector sequences.

Western Blots—Crude nuclear extracts (50 μg) were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis) and the proteins electrophoretically transferred to nitrocellulose. The filters were probed with an affinity purified antiserum raised against the C-terminal peptide of NF-YA, as described.[2]

Retardation Assays—Standard or pore-gradient gel retardation assays were performed as previously described, with a 22-mer *ds*-oligonucleotide encompassing the Y box of the Eα promoter (Hooft van Huijsduijnen *et al.*, 1987; Dorn *et al.*, 1987).

* This work was supported by institutional funds from the Institut National de la Santé et de la Recherche Médicale and the Centre National de la Recherche Scientifique and by a grant from the Association pour la Recherche sur le Cancer. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) M86214 and M86215.

‡ Received a fellowship from the Université Louis Pasteur.

§ Received a fellowship from the Fondation pour la Recherche Médicale. Present address: Glaxo Institute for Molecular Biology S. A., 14, chemin des Aulx, 1228 Geneva, Switzerland.

¶ Received a fellowship from the European Molecular Biology Organization and the Commission des Communautés Europeénnes.

[1] The abbreviations used are: MHC, major histocompatibility complex; PCR, polymerase chain reaction; CAT, chloramphenicol acetyltransferase.

[2] U. Pessara, submitted for publication.

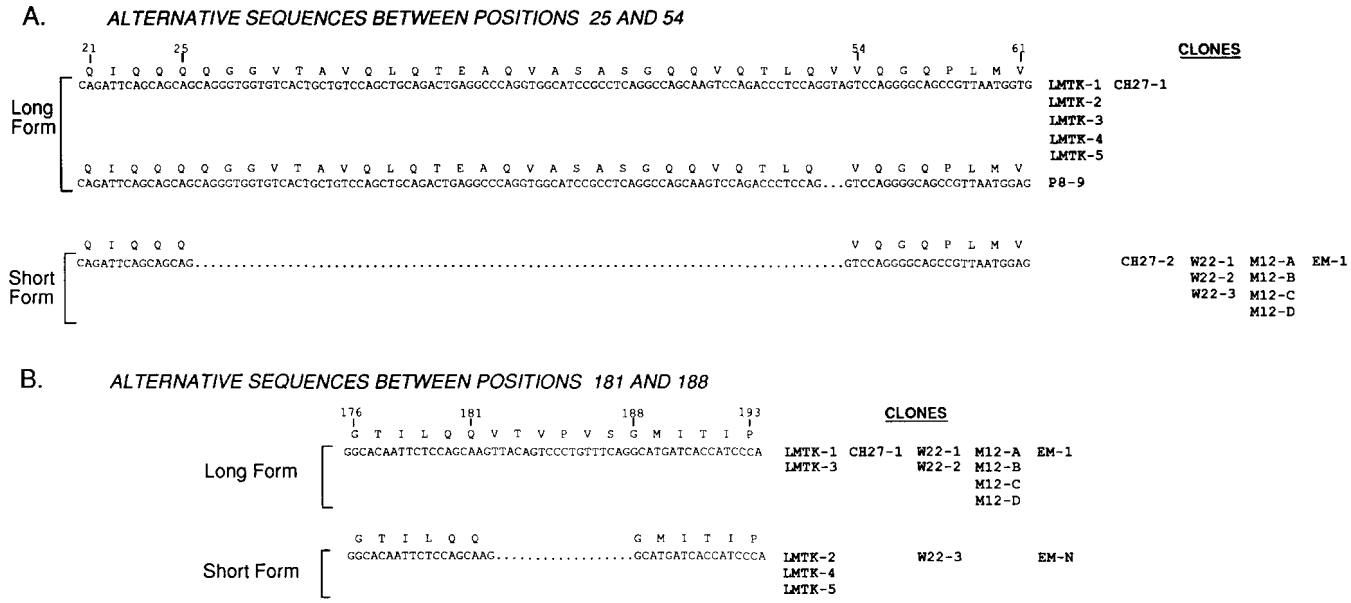


FIG. 1. **Sequence variations in murine NF-YA clones.** Sixteen cDNA clones were isolated by cDNA library screening or by PCR amplification from poly(A^+) RNA. RNAs for amplification or cDNA library construction were isolated from MHC class II-positive B lymphomas (*M12* and *CH27*) or from class II-negative cells: fibroblasts (*LMTK*), the WEHI-22 T lymphoma (*W22*), or the P8 pre-B cell line. Clones EM-1 and EM-N were isolated from a library made with A^+ RNA from whole mouse embryos at 10 days of gestation. *A*, sequence between amino acid positions 21 and 61. The numbering refers to the full long form sequence (see Fig. 3) and is thus different from our earlier numbering (Hooft van Huijsduijnen *et al.*, 1990). The identity of the cDNA clones carrying each form is shown on the *right*, grouped according to cell type. *B*, sequence between amino acid positions 177 and 193.

In Vitro Translation—Recombinant NF-Y subunits were synthesized in a wheat germ cell-free translation system and the products tested in a gel retardation assay, as described (Hooft van Huijsduijnen, 1990).

Transfections and CAT Assays—The construction of the vectors containing GAL4-NF-YA fusions will be described elsewhere.[3] Cells were transfected with 1 μg of reporter plasmid M_2TATA-CAT (a gift from L. Tora, see Webster *et al.*, 1988 and the legend to Fig. 6) and 2 μg of activator test plasmid; 2 μg of PMβGAL (a gift from A. Weisz of this Institute), a β-Gal expression vector, was also included as an internal standard. A standard calcium phosphate coprecipitation method was used (see Koch *et al.*, 1989). Extract preparations were performed according to a standard protocol (Webster *et al.*, 1988 and references therein). β-Gal activity was measured as described by Webster *et al.* (1988) and used to determine the amount of extract to be used in the CAT assays.

RESULTS AND DISCUSSION

Isoforms of NF-Y—The ubiquitous nature of NF-Y, as judged from gel retardation assays, contrasts with the marked tissue specificity of the genes whose transcription it regulates (*e.g.* albumin or class II genes of the MHC). To ascertain whether minor sequence variations of NF-Y could correlate with this specificity, we isolated by direct library screening or by PCR amplification a large number of cDNA clones for both subunits of NF-Y, clones derived from either MHC class II-positive or -negative cell types. For NF-YB, the coding sequences were identical in all clones sequenced. For NF-YA, we detected several different isoforms, with differences in two areas of the protein (Fig. 1). Between positions 25 and 54, three different forms exist. One type of clone, the "long form," carries the full complement of sequence. In several other clones, 84 nucleotides are missing; the reading frame is maintained, so that the protein encoded by "short form" mRNAs is shorter by 28 amino acids. In a third variant, found only once, a 3-base deletion removes a single amino acid from the long form.

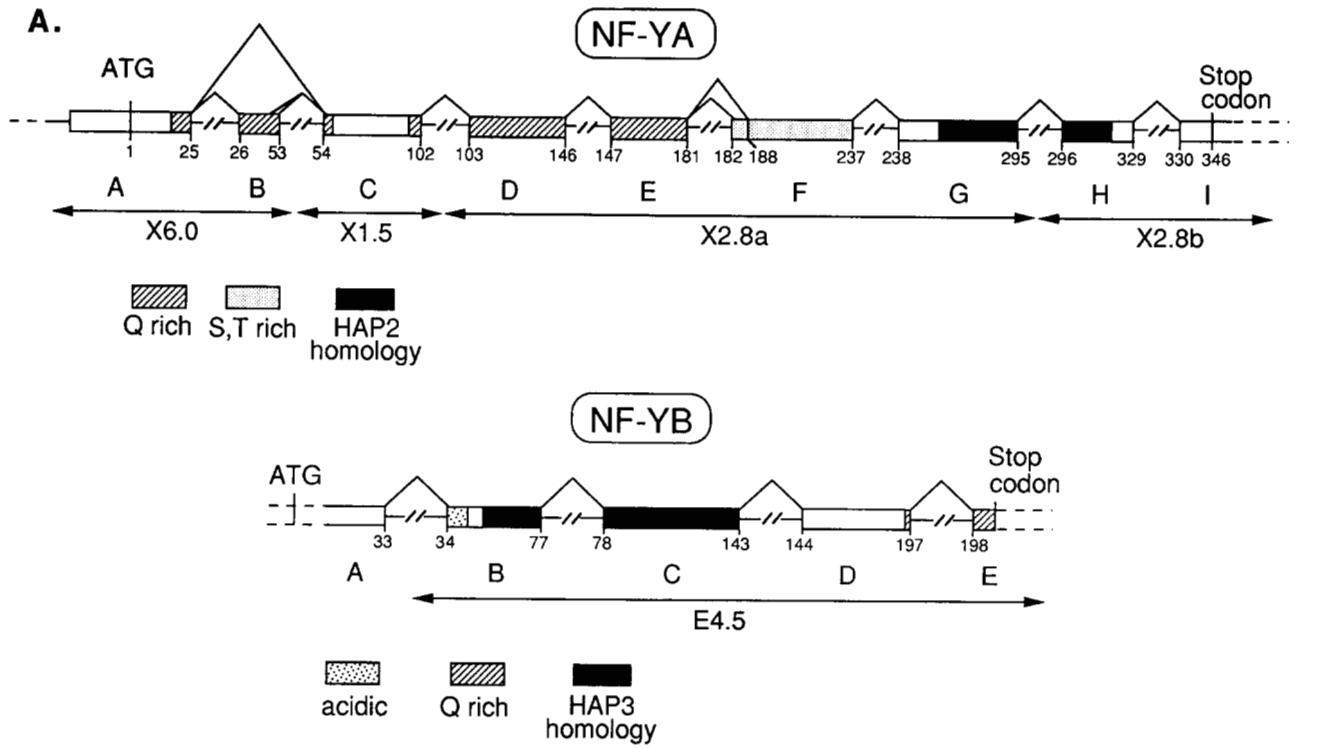
[3] R. Mantovani, manuscript in preparation.

We also found two sequence variants between positions 181 and 188, differing by the presence or absence of an 18-base pair stretch. For individual clones, there was no correlation between the sequence type at the two variable stretches: clones with the full sequence between positions 25 and 54 split evenly into the two sequence groups for positions 181–188, as did clones with the abbreviated 25–54 sequence.

NF-Y Genomic Structure—The positions of the variant sequences in NF-YA cDNA clones strongly implied that alternative splicing is involved in their generation. To establish this point, we decided to investigate the intron-exon structure of the gene encoding the NF-YA subunit; for completeness, we also studied the structure of the NF-YB gene. Several clones were obtained by screening genomic DNA libraries carried in λ vectors. For NF-YA, we isolated two overlapping clones from a mouse library in λEMBL12; for NF-YB, a fragment was isolated from a λ-607 library of mouse DNA. Several *Xba*I subfragments of the NF-YA clones hybridizing to the cDNA probe were subcloned into pBluescript and these subclones sequenced directly, using specific primers derived from the cloned cDNA sequences. We thus obtained the sequence of all the intron-exon junctions. The results are shown in Fig. 2, and the borders of the individual exons are shown on the cDNA sequences of Fig. 3.

The NF-YA protein-coding sequences derive from nine exons; these are all contained within four *Xba*I fragments, together encompassing 13 kilobase pairs. Yet another exon (at least) must encode the remainder of the 5′-untranslated sequence. For NF-YB, at least five exons could be identified; the genomic clones we obtained did not contain the sequences coding for the first 33 amino acids, which could theoretically be split further.

The data confirm that alternative splicing is operating to generate the isoforms of NF-YA. Variations between positions 25 and 54 are due to the facultative use of exon B. The short form is thus due to a splicing event that bypasses exon B.



B.

NF-YA

```
GGGCCTTTACTGATTGACTCATTTTAGGTTTTTTCCTTCCCTTTTTATTAG GAGTCCCTAC---EXON A---GCAGCAGCAG GTATGGGAGCAAGTGTGCTTTGACTATCACAATAGCACAGTGATGGAGACCTCTGTGGAAGG
AGTGATCTGCTGTCTTGGGCTGTATATTTCCATCCTGATACATTTCAAGCTCTTCCTCTTCCTGTTCTCAGCAG CAGGGTGGTG---EXON B---GACCCTCCAG GTA GTGGTGCCCTCTCTGATTCCCTGTGACGACTCGAGAACATGCCTGTCTCGTCGTGTTTGTGTTTGGATCTCAAATT
GGTGAAGCTGCTTGTGTCTGACACTATTTCATTGTTATTTTATTTCTTTCTAG GTCCAGGGGC---EXON C---TTTACAGCAG GTGAGCGGTGTTTTAAAAGTATTTAGGTTGTATGTCAGAGTAG
GAGTGGTGTCAGTGTGGTGCCAAGCTGTTTGTGTATTCTCTTCCACAG ATACAGTTGG---EXON D---CCAGACTCAG GTGAGTCCACCTGCTCCTCTGTCATCAGGGACACGAACCGACCTGGTTCACATGGG
ATATGTCATTGTTTTCTACAG ACACAACAAC---EXON E---CTCCAGCAAG GTAAGCTGCCGCTATTGCAGTGGCCCTTTTTAGGGCATTAGTTTTAGCCTTTGAGTACC
TTTTATTTTGTTTTGTTTTGTTTTTGTTTTATGTTTTGTTTTCTTAAAG TTACAGTCCCTGTTTCAG GCATGATCAC---EXON F---AATGGTCATG GTAAGGAAAGTGTCCCTGGCACTGGAGCTTCCTGTGGGGTTCTCTTGTC
ACTGCCAGTAATAAGGTCTGGTCTCCTGTTCATGCAG ATGGTACCAG---EXON G---GGAACGAAGG GTACGTGTGCCCTGGGGAGTGCTAAGCAGGAAGCAGTGCAGTTCTGCATCAA
GCTTGATATCGAATTCCTGCAGCCCGGGGGATCCACTAGTTCTAG AAATACCTCC---EXON H---TCACATGCAG GTAAGGACATTGCATCATCTTTCTCTGTTGCCGCCCTTCTCCATCAACAA
CCTAAGTGTTGATGTAAAGGATGTTGTGCTGTCTCTCTCTGTAG GATCCAAACC---EXON I----------
```

NF-YB

```
AAGTGCTCTTAACCGCTGAGCCATCTCTCCAGCCCCTAAACCTAATTCTTAAAAAACTTTCTTTAG ATACTGAGGA---EXON B---AACAGGAAAG GTAATGTAGAGGAGACTTTGACTGGCTTTTCTTTTGGCAGCGGGAGGGAAGGGGAATGGAAACTGTAAGAC
AATGGGCGCGCTCTGTGTGGCTCTTTGTTTTTGCAGCTGTTAAGAAAGTAATACTTACAGCTTCTTTTCTTCAG ATCGCAAAAG---EXON C---GTTCAGAGAG GTAGGAGACACTTACGGTTTACAATATGTCACTTACACATGCTTTGCGGCAGGTAACATTTCCGGTAGTAG
AATTTTGGACATTACATGAGCATTCTCATTTAGAAATAGTTTGATATGAATGCTTTTCTTTATTCCCTAG GCCATGAAAG---EXON D---TTACCAACAG GTATGGTGGGTGCCGCCTCTGCATTTCAAAAGACAGAGTCGATTCACTGACTATTTTAATCACCCATACTC
GTTCCCTGCTGTGGCCTGTTTGGTGTTGTCTGACATAAAAAAAAAAAAATCTTAATTTATATTTTCCAATCTTAG ATTTCTGGTG---EXON E----------
```

FIG. 2. **Genomic structure of NF-Y subunit genes.** *A*, individual exons are drawn to scale, shaded according to their amino acid composition or to their homology to yeast protein. Introns are not drawn to scale. The map does not extend to the 5′- and 3′-untranslated regions but focuses on the protein coding exons. *Horizontal arrows* deliminate the individual *Xba*I or *Eco*RI genomic fragments. Numbering refers to the amino acid positions on the longest protein isoform (see Fig. 3). *B*, nucleotide sequence at the intron-exon borders. The exon sequences are *underlined* (*double-underlined* to indicate alternative splicing pathways).

When exon B is used, the most frequently used splice donor at position 53 can also be replaced by an alternative donor at position 52, thus generating the sequence seen in clone p8-9 (Fig. 1*A*). The different isoforms between positions 181 and 188 result from alternative splice acceptor sites, at positions 182 or 188.

As usual, the positions of the introns in the NF-Y genes delineate clear protein structural domains, each exon encoding a protein segment of characteristic amino acid composition. In NF-YA, exon B constitutes the bulk of the first glutamine-rich domain, although this does continue into exon C. The latter does not have quite as high a frequency of glutamine residues, but is nevertheless marked by a high frequency of hybrophobics and a complete absence of charged residues. Exons D and E code for typical Q-rich domains, as found for example in the SP1 or Oct transcription factors (Kadonaga *et al.*, 1987; Wirth *et al.*, 1990 and references therein): high glutamine content (38 and 31%, respectively), virtually no charged amino acids. Glutamines are absent from the sequence of exon F, which has instead a high frequency of Ser and Thr residues, but still no charges.

We were initially surprised to find that the sequences coding for the HAP2 homology domain are split into two exons at position 295. But this observation made more sense in view of the results of Olesen and Guarente (1990), analyzing the effect of HAP2 point mutations and truncations on subunit interaction and DNA binding activity. These authors showed that the segment of *S. cerevisiae* HAP2 with high homology to the HAP2 protein of *Saccharomyces pombe* or to NF-YA was responsible for subunit interaction and DNA binding, but that these two functions were spatially segretated. As indicated in Fig. 3*C*, the intron in murine NF-YA falls within the

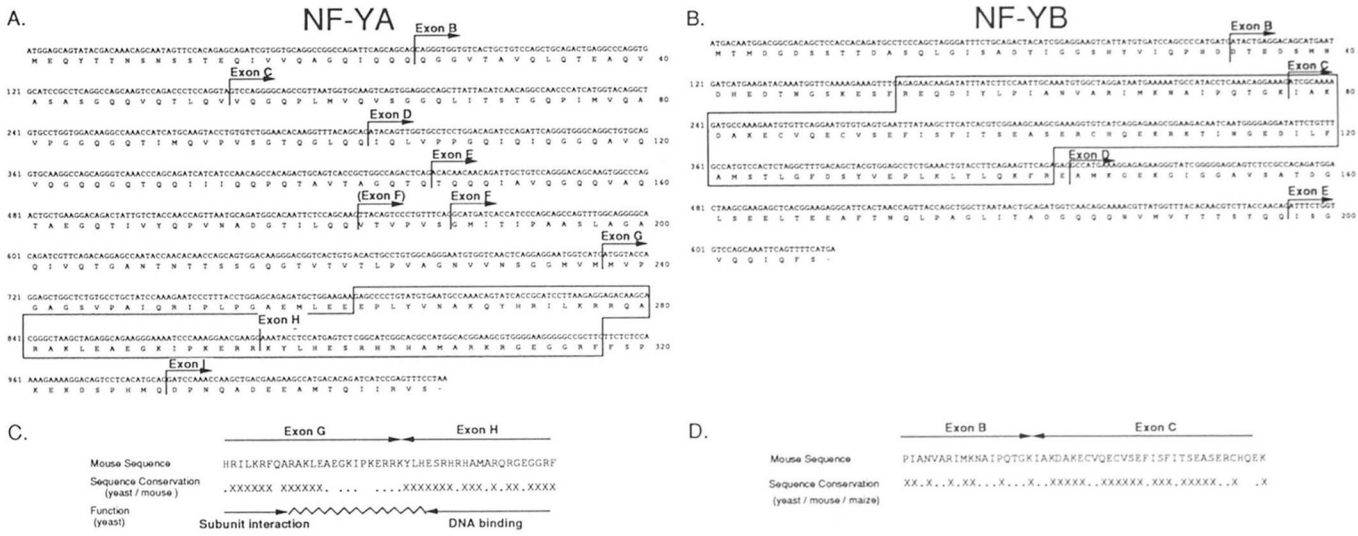


FIG. 3. **NF-YA and NF-YB sequences.** *A* and *B*, the borders between exons are shown on the cDNA sequences of the protein coding regions of NF-YA and NF-YB. The longest isoform is shown for NF-YA. The "yeast homology" domain is *boxed*. *C* and *D*, exon breakpoints within the "homology domains" of NF-YA and NF-YB. The sequence of the murine subunits is shown under *arrows* defining the exon boundaries. The degree of sequence conservation was evaluated by comparing sequences from mouse, *S. cerevisiae*, and *S. pombe* (for NF-YA) and mouse, *S. cerevisiae*, and maize (for NF-YB; sequences taken from the following references (Becker *et al.*, 1991; Hahn and Guarente, 1988; Hooft van Huijsduijnen *et al.*, 1990). An *X* denotes an amino acid present in all three sequences compared, whereas a *dot* indicates identity in two of three sequences. The boundaries of the yeast functional regions of HAP2, determined by Olesen and Guarante (1990), are indicated below. The *wavy line* between the *arrows* denotes the uncertainty of the border between the functional regions.

border between these regions. We have shown recently that the HAP2 homology domain of the mammalian protein is also sufficient for subunit interaction and DNA binding.[4] Thus, it seems reasonable to speculate that the intron which splits the homology domain in NF-YA might also separate the two functions. Interestingly, the intron falls within a 10-amino acid stretch that is rather poorly conserved when compared to the flanking sequences (Fig. 3*C*). These conserved flanks seem to correspond to two long α-helices according to secondary structure prediction programs.

An intron also bisects the homology domain in NF-YB. Again, the position of the intron falls in a short stretch between two putative α-helices, the helix-forming residues being more conserved than the intervening seven amino acids (Fig. 3*D*). From this disposition, one might speculate that NF-YB and NF-YA have the same basic structural organization: an intron separating two α-helical components, one interacting with the companion subunit, the other contacting DNA. This would imply that the NF-Y/HAP family of transcription factors follows the same organizational blueprint as the more extensive leucine-zipper or helix-turn-helix families (Busch and Sassone-Corsi, 1990; Harrison and Aggarwal, 1990; Kingston, 1989; Vinson *et al.*, 1989).

Cell-type Specificity of NF-Y Isoforms—While analyzing NF-YA cDNA clones from various cell types, we realized that the distribution of sequence variants was not homogeneous: all clones derived from LMTK (a class II-negative fibroblast line) were of the long form at positions 25–54, whereas all clones from M12 (a class II-positive B lymphoma line) were of the short form. To determine whether this distribution was due to true cell-type specificity of the alternative splicing, we performed S1 nuclease analysis of poly(A^+) RNA from a number of different cultured cell lines or from fresh mouse tissues (Fig. 4). These experiments confirmed the distribution seen in the cDNA clones: the long form accounted for most

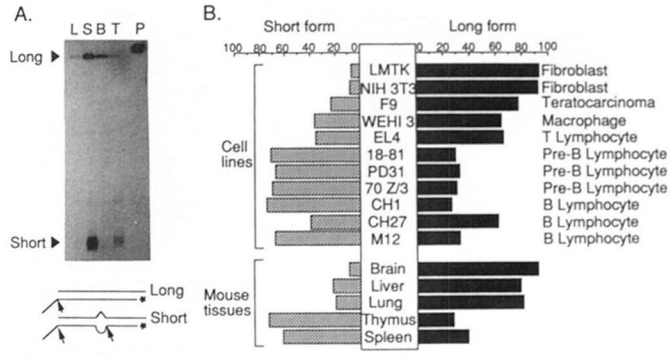


FIG. 4. **Tissue- and cell-type specificity of NF-YA gene alternative splicing.** *A*, estimation of the relative proportions of long and short form mRNA by S1 nuclease mapping. As illustrated in the diagrams on the bottom of the figure, a 5′ end-labeled probe derived from long form cDNA gives rise to different S1 digestion products after hybridization to long or short form mRNA. Long form mRNA protects the full length of the hybridizing segment of the probe; short form mRNA leaves a single-stranded loop in the probe, which is cleaved by S1, resulting in a shorter labeled fragment. RNAs in this experiment derive from liver (*L*), spleen (*S*), brain (*B*), thymus (*T*), and intact probe (*P*). *B*, from several such S1 nuclease mapping experiments we estimated by densitometry the proportion of long and short form NF-YA message in RNA from a number of cell lines or normal tissues. The values shown here are an average of two to four determinations. Although there was some variation between different experiments and RNA preparations, these usually did not exceed 15–20%.

of the NF-YA mRNA in LMTK, but the short form was the most frequent in RNA from M12. It was also apparent that the dichotomy does not correlate with MHC class II expression status, but rather breaks roughly along lymphoid/non-lymphoid lines. The long form is more prevalent in the brain, liver, lung, and in fibroblast and teratocarcinoma cells. Conversely, the short form predominates in thymus and spleen and in cells of the B lymphoid lineage: class II-negative pre-

[4] U. Pessara *et al.*, unpublished data.

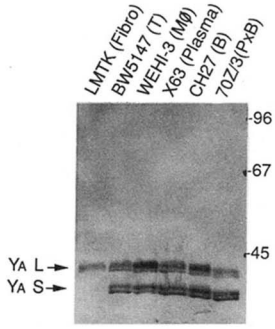


FIG. 5. **NF-YA isoforms at the protein level.** *A*, Western blot analysis of nuclear extracts from several cell lines, probed with an affinity purified antipeptide antiserum directed against the C terminus of NF-YA. The doublets which presumably correspond to long and short form NF-YA are indicated.

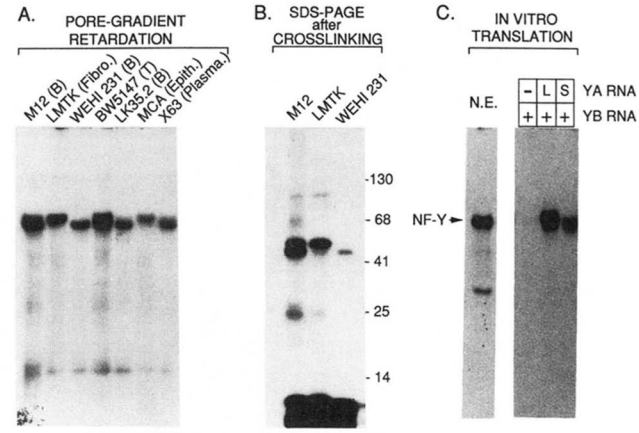


FIG. 6. **DNA binding potential of NF-YA isoforms.** *A*, NF-Y DNA binding activity in several cell lines was revealed by electrophoresis on pore-gradient gels, under very long electrophoresis times, so that each complex migrates until it reaches a pore size limit (see Hooft van Huijsduijnen, 1987). M12, WEHI231, and LK35.2 are B cell lines or hybrids, X63 is a plasmacytoma, BW-5147 a thymoma, LMTK a fibroblast, and MCA a skin epithelial cell line. The labeled oligonucleotide was a double-stranded 22-mer, encompassing the Y box of the Eα promoter. For details, see Hooft van Huijsduijnen *et al.*, 1987; Koch *et al.*, 1989. *B*, NF-Y complexes from nuclear extracts of LMTK (90% large), WEHI-231 (90% small), or M12 (75% small, 25% large) were cross-linked by incubating gel slices with 1% paraformaldehyde, as described (Dorn *et al.*, 1989). This material was then eluted from the gel and electrophoresed on a denaturing SDS-polyacrylamide gel. *C*, *in vitro* translation. Cloned long (*L*) and short (*S*) form NF-YA cDNAs were placed under the control of the T3 RNA polymerase promoter in pBluescript and then transcribed and translated *in vitro* as described (Hooft van Huijsduijnen, 1990). The products were mixed with a similar transcription/translation reaction for NF-YB and tested in a gel retardation assay with the Y box 22-mer. *N.E.*, a control retardation assay, run in parallel with standard nuclear extracts.

B cells (18–81, PD31, 70Z/3) or class II-positive mature B cells (M12, CH1); one exception to this trend is CH27, also a class II-positive B lymphoma, in which the long form appears slightly more frequently. Intermediate distributions, with only a moderate prevalence of the long form, were found for EL4 (a T lymphoma) and WEHI-3 (a myelomonocytic line); in the latter cells, γ-interferon induces the expression of MHC class II genes, but did not affect the proportion of long versus short form NF-YA (not shown).

Similar S1 nuclease analyses were performed with a probe spanning the variable sequences at positions 181–188. The data confirmed the observation with cDNA clones, namely that the isoforms at that position were equally represented in all cell and tissue types (not shown).

NF-Y mRNA Isoforms Are Translated—To determine whether the long and short forms of NF-YA mRNA do give rise to protein, we exploited highly specific anti-NF-YA antibodies, raised against a peptide derived from the C-terminal sequence of NF-YA.[5] SDS-polyacrylamide gels were run under conditions enhancing the resolution of proteins in the 30–50-kDa range; the proteins were transferred onto a filter and probed with the anti NF-YA antibody. As shown in Fig. 5, several bands can be detected in this assay. The two doublets most likely correspond to the long and short forms at positions 25–54, based on their relative sizes and on their distribution in different cell lines: very predominant long form in LMTK, somewhat predominant long form in WEHI3, predominant short form in B cells. Within each doublet, the two bands very likely correspond to the alternative forms at positions 181–188.

Different Function for NF-YA Isoforms—Having established the existence of NF-YA isoforms, we investigated the functional consequences of this sequence variability. To evaluate their DNA binding potential, we assayed a panel of cell lines with different proportions of long and short NF-YA forms, making use of the resolving power of pore-gradient gel retardation assays. More than standard gel retardation, this technique is sensitive to differences in the molecular weight of protein-DNA complexes (for references, see Hooft van Huijsduijnen *et al.*, 1987). As shown in Fig. 6*A*, we detected two NF-Y/DNA complexes by this assay, differing slightly in their migration endpoint on the pore gradient gel (it is actually possible to detect the difference between these complexes in ordinary gel retardations, but far less convincingly; data not shown). The relative proportion of the two forms in different cells correlates well with the tissue distribution of the long and short mRNA species (Fig. 6*A*). The larger protein form predominated in non-lymphoid cells (LMTK, MCA), whereas the smaller form was prevalent in lymphoid cells (M12, WEHI 231, LK-35.2, X63 or BW-5147). Sodium dodecyl sulfate-polyacrylamide gel electrophoresis after cross-linking of the two complexes showed that their differential mobility is due to a difference in molecular weight of the NF-YA subunit, of approximately 2–4 kDa (Fig. 6*B*). This value is in good agreement with the size of the protein segment encoded by the facultative exon B. To confirm by another means that both isoforms of NF-YA could bind DNA, we used NF-YA cDNAs of either isoform as a template for RNA synthesis *in vitro*, using a T7 RNA polymerase system. This RNA was then used to program a wheat germ *in vitro* translation system. The translation products were tested, together with NF-YB translation products, in a gel retardation assay. As shown in Fig. 6*C*, both long and short form NF-YA could complement NF-YB to yield specific complexes with a Y box oligonucleotide.

We thus feel confident that both mRNA isoforms differing at positions 25–54 do actually give rise to protein, capable of binding DNA. To further assess the binding specificity of NF-YA isoforms, we tested nuclear extracts from LMTK or WEHI-231 against a panel of mutant *ds*-oligonucleotides, each of which bears a single-site substitution from the wild-type Y box oligonucleotide (Dorn *et al.*, 1987). As demonstrated in Fig. 7 the patterns are superimposable, whether the NF-Y be from WEHI 231 (predominantly short form NF-YA) or from LMTK (predominantly long form), indicating the same specificity for DNA.

Although the DNA-binding function seemed identical for both NF-Y isoforms, it was conceivable that their transcrip-

[5] U. Pessara *et al.*, manuscript in preparation.

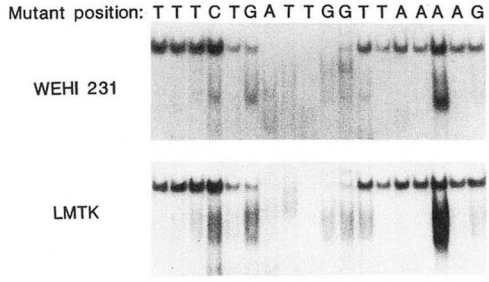


FIG. 7. **Fine specificity of DNA contacts of NF-YA isoforms.** Nuclear extracts from WEHI-231 or LMTK (predominantly short and long form NF-YA, respectively, see Fig. 6) were tested in gel retardation assays, with a panel of 18 single-base mutants of the standard Y box *ds*-oligonucleotide. Each mutant bears a transition/transversion at the indicated position of the wild-type sequence.

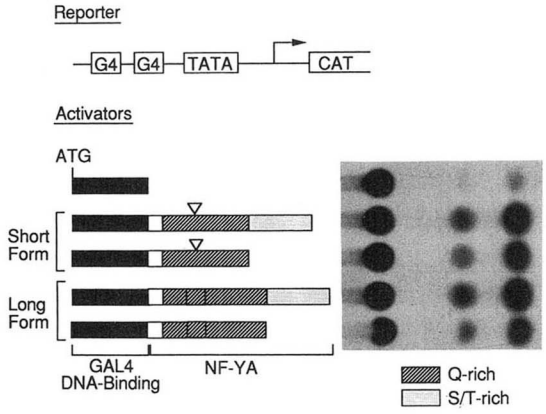


FIG. 8. **Activation potential of the NF-YA isoforms.** The reporter gene is composed of two GAL4 binding sites linked to the TATA box and CAP site from the adenovirus major late promoter (positions −34/+14; see Webster *et al.*, 1988). The activator was obtained fusing the GAL4 DNA binding domain (amino acids 1–147; black box) to the N terminus of NF-YA through a linker sequence (Ile-Gly-Arg-Ala). In some constructs the NF-YA moiety only extended to the end of the Q-rich region (to position 132) whereas in others it also included the S/T-rich region (to position 233). The expression of the activator proteins was under the control of the SV40 early promoter, in the plasmid pGAL(1–147)poly (Webster *et al.*, 1989).

tional activation potential might differ. We have recently found that the N-terminal region of NF-YA possesses significant transcriptional activation potential.[6] This activity maps to the glutamine-rich stretch and thus could be affected by the differential splicing at position 25–54. As diagrammed in Fig. 8, and following a now standard strategy, we cloned the long and short forms of NF-YA cDNA or fragments thereof into a vector driving the expression of a hybrid protein that consists of the NF-YA activation region fused to the DNA binding domain of GAL-4 (Webster *et al.*, 1988). We tested the ability of these constructs to trans-activate the transcription of a cotransfected test plasmid; the latter consisted of two tandem GAL-4 binding sites linked to the TATA box and capsite of the adenovirus major late promoter, driving the transcription of a reporter chloramphenicol acetyltransferase (CAT) gene. A representative experiment is shown in Fig. 8. Whether alone or in combination with the flanking serine/theonine-rich region, the glutamine-rich activation regions of both NF-YA isoforms were similarly effective in activating transcription. This observation applied whether the cells used for transfection were lymphoid or non-lymphoid (M12, LMTK, Hela, Raji, and HepG2 cells were tested in this manner).

In summary, we find that tissue-specific splicing leads to two forms of NF-YA, differing by 28 amino acids in the glutamine-rich transcriptional activation region. Alternatively splicing has been reported for a variety of transcription factors (see, for example, (Busturia *et al.*, 1990; Foulkes *et al.*, 1991; Gaire *et al.*, 1990; Izumo and Mahdavi, 1988; Koenig *et al.*, 1989; Nakabeppu *et al.*, 1991; Wirth *et al.*, 1990), although the tissue specificity we observe is unusual. In some instances, a functional correlate has been demonstrated for alternative forms of the proteins, often with a conservation of DNA binding but a loss of the activation function, generating, as a consequence, negative transcriptional regulators (*e.g.* Foulkes *et al.*, 1991; Izumo *et al.*, 1988; Koenig *et al.*, 1989; Nakabeppu *et al.*, 1991). In several other instances, functional differences between isoforms have not been demonstrable (Busturia *et al.*, 1990; Gaire *et al.*, 1990; Wirth *et al.*, 1990). In the present case, DNA binding is not affected, either in specificity or apparent affinity. Transcriptional activation, which maps to the domain subject to sequence variation, was equally active for the long and short forms. Yet, we consider the position of this alternative splice site to be too conspicuous to be happenstance, particularly since human NF-YA also appears as two forms differing at positions 25–54.[7] We anticipate, then, that the tissue-specific modification of the activation domain will prove, in situations less artificial than the assays used here, to be important in controlling the protein-protein contacts that NF-YA must make with other components of the transcriptional initiation complex.

Acknowledgments—We thank L. Tora and J. M. Garnier for the gift of plasmids, libraries, and advice and P. Gerber, C. Ebel, and the staff of the central culture facility for assistance.

REFERENCES

Barberis, A., Superti-Furga, G., and Busslinger, M. (1987) *Cell* **50,** 347–359
Becker, D. M., Fikes, J. D., Guarente, L. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 1968–1972
Busch, S. J., and Sassone-Corsi, P. (1990) *Trends Genet.* **6,** 36–41
Busturia, A., Vernos, I., Macias, A., Casanova, J., and Morata, G. (1990) *EMBO J.* **9,** 3551–3555
Chodosh, L. A., Baldwin, A. S., Carthew, R. W., and Sharp, P. A. (1988a) *Cell* **53,** 11–24
Chodosh, L. A., Olesen, J., Hahn, S., Baldwin, A. S., Guarente, L., and Sharp, P. A. (1988b) *Cell* **53,** 25–35
Dorn, A., Bollekens, J., Staub, A., Benoist, C., and Mathis, D. (1987) *Cell* **50,** 863–872
Dorn, A., Benoist, C., and Mathis, D. (1989) *Mol. Cell. Biol.* **9,** 312–320
Dutta, A., Stoeckle, M. Y., and Hanafusa, H. (1990) *Genes & Dev.* **4,** 243–254
Foulkes, N. S., Borrelli, E., and Sassone-Corsi, P. (1991) *Cell* **64,** 739–749
Gaire, M., Chatton, B., and Kedinger, C. (1990) *Nucleic Acids Res.* **18,** 3467–3473
Hahn, S., and Guarente, L. (1988) *Science* **240,** 317–321
Harrison, S. C., and Aggarwal, A. K. (1990) *Annu. Rev. Biochem.* **59,** 933–969
Hatamochi, A., Golumbek, P. T., Van Schaftingen, E., and de Crombrugghe, B. (1988) *J. Biol. Chem.* **263,** 5940–5947
Hooft van Huijsduijnen, R., Bollekens, J., Dorn, A., Benoist, C., and Mathis, D. (1987) *Nucleic Acids Res.* **15,** 7265–7272
Hooft van Huijsduijnen, R., Li, X.-Y., Black, D., Matthes, H., Benoist, C., and Mathis, D. (1990) *EMBO J.* **9,** 3119–3127
Izumo, S., and Mahdavi, V. (1988) *Nature* **334,** 539–542
Kadonaga, J. T., Carner, K. R., Masiarz, F. R., and Tjian, R. (1987) *Cell* **51,** 1079–1090

[6] R. Mantovani, unpublished data.

[7] Li, X.-Y., Hooft van Huijsduijen, R., Mantovani, R., Benoist, C., and Mathis, D. (1992) *Nucleic Acid Res.*, in press.

Kingston, R. E. (1989) *Cell Biol.* **1,** 1081–1087
Knight, G. B., Gudas, J. M., and Pardee, A. B. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 8350–8354
Koch, W., Benoist, C., and Mathis, D. (1989) *Mol. Cell. Biol.* **9,** 303–311
Koenig, R. J., Lazar, M. A., Hodin, R. A., Brent, G. A., Larsen, P. R., Chin, W. W., and Moore, D. D. (1989) *Nature* **337,** 659–661
Maity, S. N., Golumbek, P. T., Karsenty, G., and de Crombrugghe, B. (1988) *Science* **241,** 582–585
Maity, S. N., Vuorio, T., and de Crombrugghe, B. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 5378–5382
Nakabeppu, Y., and Nathans, D. (1991) *Cell* **64,** 751–759
Olesen, J. T., and Guarente, L. (1990) *Genes & Dev.* **4,** 1714–1729
Quitschke, W. W., Lin, Z-Y., DePonti-Zilli, L., and Paterson, B. M. (1989) *J. Biol. Chem.* **264,** 9539–9546
Raymondjean, M., Cereghini, S., and Yaniv, M. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85,** 757–761
van Wijnen, A. J., Massung, R. F., Stein, J. L., and Stein, G. S. (1988) *Biochemistry* **27,** 6534–6541
Vinson, C. R., Sigler, P. B., and McKnight, S. L. (1989) *Science* **246,** 911–916
Vuorio, T., Maity, S. N., and de Crombrugghe, B. (1990) *J. Biol. Chem.* **265,** 22480–22486
Webster, N., Jin, J. R., Green, S., Hollis, M., and Chambon, P. (1988) *Cell* **52,** 169–178
Webster, N. J. G., Green, S., Tasset, D., Ponglikitmongkol, M., and Chambon, P. (1989) *EMBO J.* **8,** 1441–1446
Wirth, T., Priess, A., Annweiler, A., Zwilling, S., and Oeler, B. (1991) *Nucl. Acids Res.* **19,** 43–51
Wuarin, J., Mueller, C., and Schibler, U. (1990) *J. Mol. Biol.* **214,** 865–874