Conservation and divergence of NF-Y transcriptional activation function

Esteban Serra*, Khalid Zemzoumi§, Alberto di Silvio¹, Roberto Mantovani¹, Vinca Lardans and Colette Dissous*

Unité INSERM 167, Institut Pasteur Lille, 1, rue du Professeur Calmette, 59019 Lille, France and
¹Dipartimento di Genetica e Biologia dei Microrganismi, Università di Milano, Via Celoria 26, 20133 Milano, Italy

Received April 1, 1998; Revised and Accepted July 2, 1998 DDBJ/EMBL/GenBank accession no. AF037602

ABSTRACT

The CCAAT-binding protein NF-Y is involved in the regulation of a variety of eukaryotic genes and is formed in higher eukaryotes by three subunits NF-YA/B/C. We have characterized NF-Y of the trematode parasite Schistosoma mansoni and studied the structure and the function of the SMNF-YA subunit. In this work, we present the cloning and sequence analysis of the B subunit of the parasite factor. SMNF-YB contains the conserved HAP-3 homology domain but the remaining part of the protein was found to be highly divergent from all other species. We demonstrated by transfections of GAL4 fusion constructs, that mouse NF-YB does not contain activation domains while the C-terminal part of SMNF-YB has transcriptional activation potential. On the other hand, the N-terminal parts of SMNF-YA and mouse NF-YA were shown to mediate transactivation; the integrity of a large 160 amino acid glutamine-rich domain of NF-YA was required for this function and an adjacent serine- and threonine-rich domain was necessary for full activity in HepG2, but redundant in other cell types. Transactivation domains identified in SMNF-YB are also rich in serine and threonine residues. Our results indicate that serine/threonine-rich sequences from helminth parasites potentiates transcription and that such structures have diverged during evolution within the same transcription factor.

INTRODUCTION

NF-Y (also called CBF, CP1) is a ubiquitous heteromeric complex that binds to the widespread CCAAT sequence present in the promoters and enhancers of a variety of eukaryotic genes (1,2). NF-Y has an absolute requirement for the CCAAT box as well as a strong preference for specific flanking sequences (3,4).

NF-Y from higher eukaryotes is formed by three subunits, NF-YA, NF-YB and NF-YC (5,6). All subunits are required for DNA binding (6). Saccharomyces cerevisiae also has a CCAAT-binding activity displayed by the HAP-2/3/4/5 complex that is specifically required to activate transcription when cells are grown on a non fermentable carbon source (7–9). NF-YA/B/C subunits present a significant degree of identity with yeast HAP-2, HAP-3 and HAP-5 factors, respectively, in conserved regions called the HAP-2/3/5 homology domains (6,7,10). The yeast HAP-4 protein is not needed for DNA-binding but contains an acidic domain that is essential to promote transactivation when associated with the HAP-2/3/5 complex (11).

NF-YB and NF-YC form a dimer which in turn interacts with NF-YA allowing the whole complex to bind DNA (6). Interestingly, the conserved HAP-3/5 homology domains within NF-YB–NF-YC have sequence similarities to the histone fold motifs (HFM) of the H2B and H2A histones respectively, and to the archaeabacterial histone-like protein Hmf-2 (12). A previous mutational analysis of NF-YA indicated that the HAP-2 homology domain contains a segment for interaction with other NF-Y subunits and a domain necessary for DNA binding (13,14). NF-YA does not present similarity to known interfaces and motifs and therefore constitutes an unusual heteromeric DNA-binding factor.

Transcriptional activation functions of transacting factors generally lie on domains localized outside of the DNA-binding regions. Several transcription factors are classified according to their motifs capable of mediating transcriptional activation such as acidic, glutamine-rich and proline-rich (15,16). In higher eukaryotes, NF-Y transcriptional activation has been shown to be essentially displayed by the NF-YA and NF-YC subunits. NF-YA transactivation domains were predominantly localized in the N-terminal region of the subunit which is highly rich in glutamine (35%) and flanked on its C-terminal side by a stretch of serine/threonine residues. On the other hand, the NF-YC transactivation activity is contained in the C-terminal part of the protein, which is very hydrophobic and also highly rich in glutamine residues (17,18).

We have already characterized the NF-YA subunit of the human blood fluke Schistosoma mansoni (SMNF-YA). We have shown that the HAP-2 homology domain is highly conserved in SMNF-YA, that is able to associate with the mammalian NF-YB/C subunits to form a complex with affinity to Y boxes derived from the 5’ flanking region of the Sm28GST parasite gene (19,20). These results are in line with the functionality of the schistosome HAP-2 homology domain and the existence of NF-YB/C subunits in this parasite.
In this paper, we present and analyse the complete sequence of the B subunit from the SMNF-Y complex (SMNF-YB). As expected, SMNF-YB shows a high degree of conservation in the HAP-3 homology domain, containing the HFM and the TATA binding protein (TBP)-binding domain (21). By contrast, the remaining part of the protein is completely divergent from the mammalian and yeast counterparts. No glutamine stretches are found in the SMNF-YB N-terminus, while a large C-terminal region is rich in serine and threonine. Interestingly, the parasite NF-YB, unlike its eukaryotic counterpart, is herein shown to mediate transactivation via its serine/threonine-rich domain, suggesting the ability of such a structure to potentiate transactivation in helminth parasites.

MATERIALS AND METHODS

Cloning of the S.mansoni NF-YB subunit

The following degenerated oligonucleotides were designed following comparison of the HAP-3 homology region from yeast and mammalian proteins: 5′-AA(GA/TG)CAAA(GA/G)-3′ and 5′-AT(GA/TC/CT)TICCC(GA/TT)TA-3′. These oligonucleotides were used in a touchdown PCR (22) on first strand adult S.mansoni cDNA. The touchdown PCR was performed using decreasing annealing temperatures as follows: 45°C for 5 cycles, 40°C for 5 cycles and 37°C for the last 25 cycles. Amplification products were cloned in the pCRII plasmid (InVitrogen), sequenced and used as probes for screening of a λgt10 adult S.mansoni cDNA library.

NF-YA and NF-YB transactivation regions—DBD Gal constructs

Different oligonucleotide pairs were synthesized in order to generate an SMNF-YA segment coding for yeast and mammalian proteins: 5′-AA(A/G/A/T)(G/A/C/A)AA-3′ and 5′-AT(A/G)(A/T)(C/T)(T/C)TICCC(G/A/TT)TA-3′. These oligonucleotides were used as a touchdown PCR (22) on first strand adult S.mansoni cDNA. The touchdown PCR was performed using decreasing annealing temperatures as follows: 45°C for 5 cycles, 40°C for 5 cycles and 37°C for the last 25 cycles. Amplification products were cloned in the pCRII plasmid (InVitrogen), sequenced and used as probes for screening of a λgt10 adult S.mansoni cDNA library.

RESULTS AND DISCUSSION

The DNA-binding and subunit interaction (HAP-3) domain of NF-YB appeared to be highly conserved at the primary sequence level. Based on this observation, we considered the possibility of isolating the S.mansoni NF-YB counterpart by a PCR-based strategy using degenerate oligonucleotides similar to the flanking regions of the conserved HAP-3 homology domain. Amplification was performed on adult S.mansoni cDNA that generated a major product of 130 bp. The sequence of the PCR product was determined, showing a high identity with the HAP-3 homology domain from yeast and mammalian factors. Further screening of an adult S.mansoni cDNA library, using the 130 bp DNA fragment as a probe, led us to the isolation of several clones containing a unique inserted sequence of 1547 bp with an open reading frame encoding a polypeptide of 242 amino acids homologous to the yeast HAP-3 and human NF-YB (CBF-A) proteins, named SMNF-YB.
As shown in Figure 1, a region highly similar to human and yeast proteins (80 and 61.5% identity, respectively) was observed between the P30 and the L104 residues of the schistosome protein. The HAP-3 homology domain is boxed. The HFM is underlined and necessary for TBP-binding (21). A similar TBP-binding region was larger than that of yeast and mammalian proteins and may adopt a structure formed by six β-sheets, each composed of 13–23 residues and homogeneously distributed from the end of the C-terminal activation domain of Oct-2 (31,32), which also possesses a high proline content, and the enhancer binding proteins ITF-1, ITF-2 and TFE3 (for review see 33). Moreover, the concept that activation domains often harbour multiple small subdomains is now well established. This idea first proposed for the yeast activators GAL4 and GCN4 has since been demonstrated for a large number of activators from different origins (for review see 34).

In addition, the two fusion proteins containing each half of the C-terminal region of SMNF-YB and the N-terminal part of SMNF-YA, as well as the modular structure predicted for the SMNF-YB region, we decided to study the transactivation potential of these two factors in order to provide new information about transactivation processes in helminths. Proteins were fused to the DNA-binding domain of the GAL4 yeast factor, which conserves the ability to bind the GAL4-DNA target sequences, but lacks activation activity. Four fusion proteins were prepared, as shown in Figure 2A and B. Two of these proteins contained the complete N-terminal sequence of SMNF-YA and the complete C-terminal sequence of SMNF-YB, respectively. The other fusion proteins contained halves of the C-terminal region of SMNF-YB. These plasmids were co-transfected with a construct containing the luciferase reporter gene under the control of five tandem GAL4 specific target sequences. The results presented in Figure 2C demonstrate that SMNF-YA and SMNF-YB chimeric proteins exhibited transactivation activity. In the three mammalian cell lines tested, the luciferase activity was higher than that obtained with the control DBD Gal plasmid. Interestingly, for two cell lines, the maximal luciferase activity was detected with the construct containing the whole C-terminal region of SMNF-YB. In addition, the two fusion proteins containing each half of the SMNF-YB activation region mediate nearly half the activity obtained with the whole C-terminal segment of SMNF-YB. These observations demonstrate the potential of SMNF-YB for transactivation and support the concept of a modular structure of the transactivation domain.

In parallel, we subjected mouse NF-YA and NF-YB to the same type of analysis. Figure 3A shows that, of the several mutants used, only YA5, YA6, YA11 and YA12 were active on the M2TA TA-CA T reporter plasmid, both in HepG2 and in M12 B lymphocytes. Mutants lacking the N-terminal glutamine-rich activators were all inactive, as were constructs in which the glutamine-rich domain was split (YA14, YA15 and YA16). Moreover, mutants containing the serine- and threonine-rich domain showed distinctively higher activities in HepG2, while mutants YA6 and YA12, lacking this domain, were all inactive. The modular structure analysis was not able to predict any similar structure in yeast or mammalian protein sequences. In mammals, NF-Y-dependent activation processes are mainly mediated by NF-YA and NF-YC subunits (17,30) while in yeast, CCAAT-dependent transactivation activity has been attributed to the HAP-4 component (11). The SMNF-YA subunit does show the typical glutamine-rich transactivation domain found in the N-terminal part of NF-YA proteins from higher eukaryotes but contains also an elevated proportion of serine/threonine residue in its N-terminal region (19). A small group of transcription activators contain activation domains that have been already characterized as serine- and threonine-rich. These include the C-terminal activation domain of Oct-2 (31,32), which also possesses a high proline content, and the enhancer binding proteins ITF-1, ITF-2 and TFE3 (for review see 33). Moreover, the concept that activation domains often harbour multiple small subdomains is now well established. This idea first proposed for the yeast activators GAL4 and GCN4 has since been demonstrated for a large number of activators from different origins (for review see 34).

Taking into account the serine/threonine contents of both the C-terminal region of SMNF-YB and the N-terminal part of SMNF-YA, as well as the modular structure predicted for the SMNF-YB region, we decided to study the transactivation potential of these two factors in order to provide new information about transactivation processes in helminths. Proteins were fused to the DNA-binding domain of the GAL4 yeast factor, which conserves the ability to bind the GAL4-DNA target sequences, but lacks activation activity. Four fusion proteins were prepared, as shown in Figure 2A and B. Two of these proteins contained the complete N-terminal sequence of SMNF-YA and the complete C-terminal sequence of SMNF-YB, respectively. The other fusion proteins contained halves of the C-terminal region of SMNF-YB. These plasmids were co-transfected with a construct containing the luciferase reporter gene under the control of five tandem GAL4 specific target sequences. The results presented in Figure 2C demonstrate that SMNF-YA and SMNF-YB chimeric proteins exhibited transactivation activity. In the three mammalian cell lines tested, the luciferase activity was higher than that obtained with the control DBD Gal plasmid. Interestingly, for two cell lines, the maximal luciferase activity was detected with the construct containing the whole C-terminal region of SMNF-YB. In addition, the two fusion proteins containing each half of the SMNF-YB activation region mediate nearly half the activity obtained with the whole C-terminal segment of SMNF-YB. These observations demonstrate the potential of SMNF-YB for transactivation and support the concept of a modular structure of the transactivation domain.

In parallel, we subjected mouse NF-YA and NF-YB to the same type of analysis. Figure 3A shows that, of the several mutants used, only YA5, YA6, YA11 and YA12 were active on the M2TA TA-CA T reporter plasmid, both in HepG2 and in M12 B lymphocytes. Mutants lacking the N-terminal glutamine-rich activators were all inactive, as were constructs in which the glutamine-rich domain was split (YA14, YA15 and YA16). Surprisingly, constructs containing both the activation domain and the HAP2 conserved domain were also devoid of any activating potential (YA1, YA4, YA10 and YA13). Moreover, mutants containing the serine- and threonine-rich domain showed distinctively higher activities in HepG2, while mutants YA6 and YA12, lacking this part of NF-YA, were higher in M12 (Fig. 3A) as well as in NIH3T3 and HeLa cells (data not shown), suggesting that cell-type preference plays a role in activation potential by this domain. To rule out the possibility that these results were due to poor expression of some of the mutants, we checked expression of all mutants in transfections by performing western blot analysis.
with an anti-GAL4 antibody: Figure 3B shows that mutants negative in transcription were present at equal levels compared to the functioning fusion proteins.

The same type of experiment was tried with mouse NF-YB–GAL4DBD fusions (Fig. 4): in this case, neither the wt NF-YB, nor any of the mutants showed any detectable degree of activation on the M2TA-CA-T vector. To further prove that mouse NF-YB subunit is truly devoid of any activating potential in this assay and rule out that more than two GAL4 sites are required for NF-YB proteins, we tried the more sensitive cis-vector containing the TK promoter in front of two GAL4 sites. As shown in Figure 4, this vector was efficiently activated by the positive GAL4 control, but not by any of the NF-YB constructs: if anything, a small but reproducible degree of repression was observed. We also transfected NIH3T3 cells in parallel with SMNF-YB, mouse NF-YB and YA6, with the 5xGAL4 sites-containing vectors used before: the results shown in Table 1 indicate that unlike SMNF-YB, mouse NF-YB does not activate. This experiment also allowed us to quantify the YA6 potential as being of similar magnitude with respect to the SMNF-YA and SMNF-YB. In these experiments, the expression efficiency of all mutants was controlled by western blot analysis, as described for NF-YA constructs (data not shown). Moreover, the mouse GAL4-NF-YB constructs do activate, provided that the two other subunits are cotransfected (A.di Silvio and R.Mantovani, in preparation). The conclusion of this set of experiments is that mouse and schistosome NF-YA are both able to activate transcription, while only the helminth NF-YB harbours such potential.

Table 1. Comparative analysis of schistosome and mouse NF-YB transcriptional potential

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Luciferase activity (fold activation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAL4Poly</td>
<td>1</td>
</tr>
<tr>
<td>GAPYB (S.mansoni)</td>
<td>10.4</td>
</tr>
<tr>
<td>GALYB (mouse)</td>
<td>1.6</td>
</tr>
<tr>
<td>GALYA (mouse)</td>
<td>11.1</td>
</tr>
</tbody>
</table>

Luciferase activity was measured in NIH3T3 cells co-transfected with NF-Y constructs and pGSTATALuc plasmid.

NF-Y represents an ideal system to follow conservation/divergence of transcriptional activation domains. Unlike many general transcription factors (such as TBP, TFIIIB, TFIIA, etc.) which are highly conserved, but do not possess a distinct transcription activating domain, and unlike many DNA-binding factors, whose conservation in different phyla is much lower and restricted to the DNA-binding domain, NF-Y shows among the highest degree of conservation (67–73% identity, 85–90% similarity) in the subunit association and DNA-contacting parts, but also in recognizable activating domains. At this time, little is known about transcription factors in helminth parasites. Besides the A subunit of the SMNF-Y complex, a limited number of Schistosoma transcription factors have been described, that include the Heat Shock Factor (HSF) (35) and several nuclear receptors considered as transcriptional regulators (36). In all cases, a relatively high conservation of sequences involved in the binding
Figure 3. Transcriptional activation of mouse NF-YA. (A) The scheme depicts the wt short (YA1) and long (YA13) forms of NF-YA proteins fused to the DNA-binding domain of GAL4. YA2 (amino acids 161–346), YA3 (amino acids 262–346), YA4 (amino acids 1–317), YA5 (amino acids 1–261), YA6 (amino acids 1–160), YA8 (amino acids 161–317), YA9 (amino acids 262–317). YA10, YA11, YA12 are equivalent to YA4, YA5, YA6, except that they contain the 28 amino acids of the optional exon B (46). YA14 (amino acids 1–70, short form), YA15 (amino acids 1–70 long form), YA16 (amino acids 71–160). The reporter CAT vector M2TA TA-CA T contains two GAL4-binding 17mers in front of the AdML TATA box (24). (B) Western blot analysis of extracts derived from HeLa cells transfected with the indicated YA mutants using an anti-GAL4 monoclonal antibody.

Figure 4. Transcriptional potential of mouse NF-YB. GAL4–NF-YB constructs were made essentially as described for NF-YA in Figure 3. YB1 (amino acids 1–161), YB2 (amino acids 1–50), YB4 (amino acids 51–140), YB6 (amino acids 141–207). The N- and C-terminal domains of NF-YB are indicated by dotted boxes, while a grey box indicates the central conserved domain of NF-YB. Results of transfection of HeLa cells with M2TA TA-CA T (left panel) or MX2-CA T (containing a tandem of GAL4-binding 17mers fused to the TK promoter) are presented.

to specific DNA targets was observed between schistosomes and higher eukaryotes. However, no significant identity was found when other regions, particularly activation domains of mammalian factors, were compared to their Schistosoma counterparts.

The NF-YA N-terminal domains of mammals, sea urchin, S.mansoni and Aspergillus nidulans contain clearly recognizable glutamine-rich domains (19,37, R.Mantovani, unpublished). Similar in overall amino acid composition, hydrophobics predominate as well as glutamines, they are largely different in the primary structure. Many factors of different phylogenetic origins have been shown to activate transcription by their acidic- and/or proline-rich transactivation domains in the budding yeast (38–40). The S.cerevisiae acidic transcription factor GAL4 also activates transcription in mammals as well as in plants, suggesting a conserved and interchangeable mechanism of acidic transactivation factors throughout evolution (41,42). In contrast, glutamine-rich domains, such as those found in the human SP1 factor, failed to potentiate activation of transcription in S.cerevisiae (43), while they were functional in Schizosaccharomyces pombe (44). Results indicating that glutamine-rich sequences have different functions in human, S.pombe and S.cerevisiae could mean that such domains need additional transcription cofactors which are present in mammals and absent, or playing distinct functions, in S.cerevisiae. In keeping with this observation, a recent paper reported binding of NF-YA–NF-YC Q-rich activation domains to dTAFII110 (45), one of the few TBP-associated factors for which no S.cerevisiae counterpart has been found so far. In S.cerevisiae, a fourth subunit, HAP-4, containing
a powerful highly acidic domain is required to activate the HAP-2/3/5 complex: this subunit is apparently absent in other phyla, where the activating information has been incorporated within the core subunits within complex, perhaps modulable domains. In NF-YA, a serine- and threonine-rich segment was described as being essential for full transactivation in concert with the glutamine-rich region in an *in vitro* system (17). The data of transfections in different cell lines presented here support the idea that this domain is necessary for full activation in some but not all cell lines, thus suggesting that activation is influenced by the cell-type context, probably reflecting the existence, or abundance, of intermediate cell-specific cofactors capable of mediating activation in a given cell type. It is worth noting that mouse NF-YA activation emerges only in GAL4 constructs devoid of the HAP-2 homology domain: this could be explained by supposing that such fusion proteins are coupled by the endogenous NF-YB–NF-YC dimer that can then bind the plethora of CCAAT-box targets in the genome, or that additional nuclear factors exist that mask the homology domain. Further experiments are needed to discriminate between these possibilities. Our observations suggest that NF-Y protein identity is restricted to the HAP-2/3/5 homology domains, the latter two containing putative histone fold domains and function of a complex evolution of the structure/function of NF-Y factors as well as of other transcription factors.

Our study, performed in mammalian cells, represents the first indication that activation domains of a *Schistosoma* transcription factor, as divergent from human as it is from yeast, efficiently operates in a mammalian context. The results presented here also show, for the first time, the existence of a functional activation domain in the B subunit of an NF-Y complex. We found that, unlike NF-YA and NF-YC that mediate transactivation via their glutamine-rich domains, the mouse NF-YB subunit is completely unable to potentiate transcription in higher eukaryotes. Our experiments suggest a model whereby transactivation by NF-Y in helminth parasites could be mediated by the A and B subunits via glutamine- and serine/threonine-rich domains, with the latter being slightly more efficient. This observation could also signify the preference for serine/threonine-rich configurations as activators for the schistosome RNA PolII, depending on the context of the parasite promoters and the position of DNA-binding sites that govern responsiveness to different regulatory signals. The fact that the C-terminal region of SMNF-YA also contains serine and threonine residues in addition to glutamines, when compared with the mammalian NF-YA (15), is in line with this hypothesis. Further cloning and analysis of the third SMNF-YC subunit will certainly shed more light on the evolution of transcription activating function of this conserved transcription factor.

**ACKNOWLEDGEMENTS**

We thank J. Trolet and J.M. Merczez for technical assistance. This work was supported by the Institut National de la Santé et de la Recherche Medicale (INSERM), the Institut Pasteur de Lille and grant PRIN ‘Protein–Nucleic Acids Interactions’ to R.M. E.S. was a fellow of the Fondation pour la Recherche Medicale (France).

**REFERENCES**


