UNIVERSITÀ DEGLI STUDI DI MILANO

Scuola di Dottorato in Scienze Biologiche e Molecolari

XXVII Ciclo

The interplay among NF-Y and E-BOX Transcription Factors: Myc, MAX and USF1

M. Lorenzo

PhD Thesis

Scientific tutor: Prof. Roberto Mantovani

Academic year: 2013-2014

SSD: [BIO/10; BIO/11]

Thesis performed at Department of Biosciences, University of Milan.

CONTENTS

PART I

ABSTRACT

Transcription is a process finely regulated by different transcription factors (TFs) which bind regulatory sequences present in gene promoters and allow the precise execution of gene expression programs. Misregulation of such process can lead to different pathologies, including development/differentiation defects, uncontrolled cell growth and cancer. For these reasons it is important to understand the molecular details of the interplay that occurs between different TFs to modulate gene expression.

NF-Y, the heterotrimeric complex composed by NF-YA, NF-YB and NF-YC subunits, all required for DNA binding, recognizes the consensus sequence CCAAT, present in about 30% of eukaryotic promoters, at -60/-100 bp from the Transcription Start Site (TSS).

One of the most important roles of NF-Y in transcription is to interact synergistically with other TFs to activate, or to repress, gene expression.

In this study we focused on the relationship occurring between NF-Y and the TFs MAX, Myc and USF1, which recognize the E-BOX sequence CACGTG. We made this choice relying on genomic information and ChIP-seq data derived from the ENCODE project. These studies revealed that in CCAAT promoters, several Transcription Factor Binding Sites (TFBS) are consistently present. Among these, the E-BOX consensus is one of the most frequently found, generally at a distance from CCAAT of 10/12 bp, in a precise orientation; moreover CCAAT and E-BOXes, at these distances, are bound *in vivo* by NF-Y an E-BOX TFs, respectively.

In addition to genomic information, the knowledge of NF-Y and E-BOX TFs domains structures in complex with DNA allows us to speculate on the possible interactions that might occur when DNA, with this configuration, is bound, and bent, by the two proteins.

Therefore, we analyzed and characterized the relationships among NF-Y and MAX, Myc or USF1 using Electrophoretic Mobility Shift Assay (EMSA) and ChIP analyses:

in vitro EMSAs were performed using E-BOX TFs (MAX or USF1 as homodimers, and Myc/MAX as heterodimers) and NF-Y proteins purified from *E. coli*; as probes, DNA sequences derived from genomic loci that reflect the observed distances and orientation of CCAAT and E-BOX. In particular, we chose DNA sequences derived from the HOXB4 and HOXB7 gene promoters, and the ERV9 Long Terminal Repeat (LTR) region.

With these experiments, we found that NF-Y can form stable ternary complexes with all recombinant E-BOX TFs and DNA configurations analyzed, with no negative interactions. We observed that USF1 and NF-Y display a facilitated binding on HOXB4, that becomes even more

stable on HOXB7, and cooperative on the LTR sequence. Moreover, we observed that the minimal domains of NF-Y and USF1 maintain these DNA binding features.

ChIP experiments in HeLa cells, previously silenced for NF-Y, revealed that the interplay between NF-Y and E-BOX TFs exists also *in vivo*. In fact, after NF-Y silencing, Myc, MAX and USF1 binding is affected, directly or indirectly, further supporting the notion that NF-Y and E-BOX TFs cooperate in the regulation of target genes that contain CCAAT and E-BOXes in their promoters.

1. STATE OF ART

1.1 TRANSCRIPTION AND TRANSCRIPTION FACTORS

Transcription is the process through which the genetic information carried by DNA is transferred to a complementary molecule of RNA. This process is finely regulated through different groups of regulatory proteins called Transcription Factors (TFs). Usually, TFs recognize regulatory sequences present in gene promoters to induce the expression and/or repression of target genes in different cellular contexts.

There are two groups of eukaryotic TFs: the first participates to the assembly of transcription-initiation complexes (composed by co-activators, co-repressors, general TFs, histone acetylases, deacetylases, kinases, and methylases), the second is composed by TFs that recognize in a specific manner regulatory sequences present in gene promoters (Lemon and Tjian, 2000). This latter class could be divided in subclasses following the order: constitutively expressed and regulatory, that are induced by specific cellular and extracellular signals (Brivanlou and Darnell, 2002) (Fig. 1).

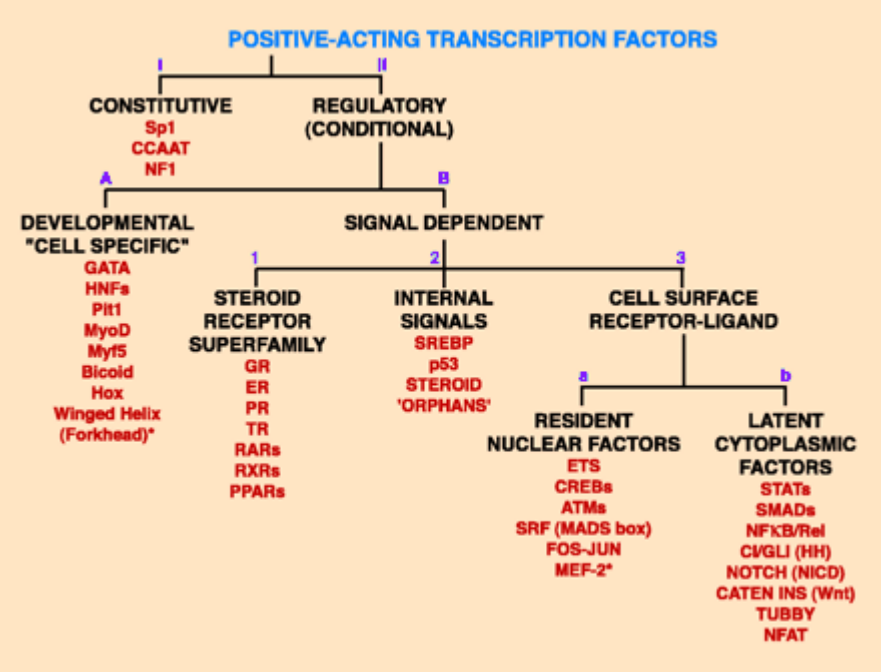


Figure 1: Scheme of positive-acting transcription factors. These transcription factors usually have a DNA-binding domain and a transactivation domain that exhibits transcriptional activation potential. In black principal groups, in red some examples of different classes (Brivanlou and Darnell, 2002).

The regulation of the expression through binding of TFs at specific consensus sequences is the first step by which the complex mechanisms of differentiation and development are regulated.

The knowledge of the possible interaction between 2 or more TFs could be the tool to understand how specific diseases develop and, at the same time, the starting point for a possible therapy. For example, Darnell (2002) suggests the use of Transcription Factors as targets for cancer therapy, because it is clear that specific TFs are over expressed and/or down regulated in specific types of cancer (Table 1).

Table 1: Examples of tumors with high levels of NF-kB (Darnell, 2002)

	Tumours with persistently high levels of NF-κB	
Gene	**Haematopoietic tumours**	**Solid tumours**
c-REL	Diffuse large-cell lymphoma; primary mediastinal B-cell lymphoma; follicular large-cell lymphoma; follicular lymphoma; diffuse large-cell lymphoma	Non-small-cell lung carcinoma
RELA	B-cell non-Hodgkin's lymphoma; multiple myeloma; diffuse large-cell lymphoma	Squamous head and neck carcinoma; breast adenocarcinoma; stomach adenocarcinoma; thyroid carcinoma cell lines; non-small-cell lung carcinoma
NFKB1	Acute lymphoblastic leukaemia	Non-small-cell lung carcinoma; colon cancer cell lines; prostate cancer cell lines; breast cancer cell lines; bone cancer cell lines; brain cancer cell lines
NFKB2	Cutaneous T-cell lymphoma; B-cell non-Hodgkin's lymphoma; B-cell chronic lymphocytic leukaemia; multiple myeloma	Breast carcinoma; colon carcinoma
BCL2	B-cell non-Hodgkin's lymphoma; B-cell chronic lymphocytic leukaemia	

An example of ubiquitous Transcription Factors that recognize in a specific manner the consensus sequence CCAAT is Nuclear Factor Y-box (NF-Y).

1.2 NF-Y

NF-Y is an heterotrimeric complex composed of three independent subunits, NF-YA, NF-YB and NF-YC , their assembly is required for CCAAT binding. There are different proteins that can recognize the CCAAT-box (for example YY1 or C/EBPs), but NF-Y is the only one that recognizes in a specific manner all 5 nucleotides and also the flanking regions. The three subunits of NF-Y are coded by 3 independent genes. Their transcripts are subjected to alternative splicing generating different isoforms for NF-YA and NF-YC. These isoforms combine with each other and with NF-YB to form different trimers differentially expressed in cell lines and in different stages of cell development (Ceribelli *et al*, 2009; Dolfini *et al*, 2002) (Fig 2).

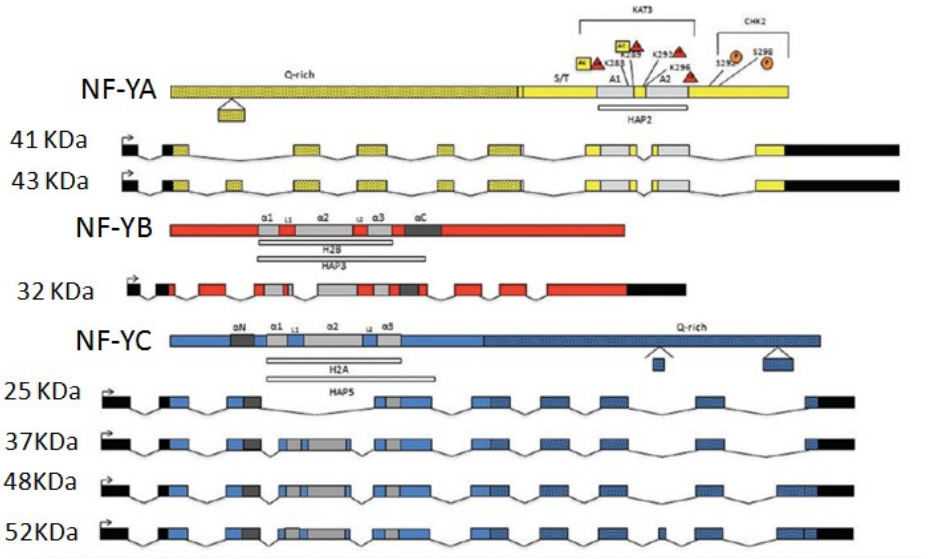


Figure 2:NF-Y is an heterotrimer composed by 3 independent subunits: NF-YA, NF-YB and NF-YC. For NF-YA and NF-YC different splicing forms exist and they are differentially expressed basing on cellular types and development stages of the cells (Dolfini *et al,* 2013).

The 3 subunits present regions that are very conserved throughout evolution: NF-YB and NF-YC present the histone fold domain that are homologues in primary sequence to histone H2B and H2A, respectively. Instead, the conserved region of NF-YA is localized at the C-terminus where we can find the domain necessary for the recognition of NF-YB/NF-YC dimer and CCAAT sequence. NF-YA and NF-YC contain transactivation domains localized at the N-terminus and C-terminus, respectively; these domains are enriched in hydrophobic residues and glutamines, in mammalians, and they are the regions least conserved in evolution (Mantovani R, 1999).

All 3 subunits are essential for DNA binding, in fact, the NF-Y assembly process takes place in different steps: first of all, NF-YB and NF-YC associate, then the trimer regulatory subunit, NF-YA, binds to the heterodimer; after the reconstruction of the functional heterotrimer, the TF recognizes the CCAAT sequence and regulates the expression of target genes (Ly LL *et al*, 1995) (Fig 3).

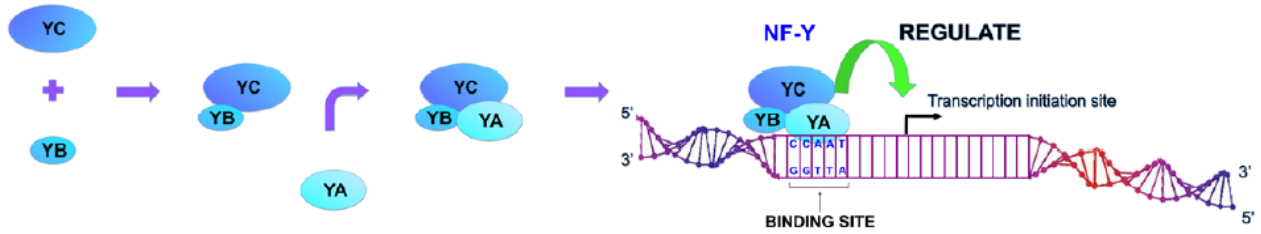


Figure 3: NF-Y complex formation. NF-Y consists of three independent subunits, NF-YA, NF-YB and NF-YC, which are all necessary for the formation of a functional trimer and binding to the CCAAT boxes (Ly LL *et al,* 1995).

1.3 NF-Y STRUCTURE

One of the goals in the study of TFs is the possibility to have 3D structures of TFs bound to DNA with a high resolution. Exploiting this technique, we can have the information about the bending of DNA bound by TFs, the residues of TFs that contact DNA and more in general the molecular information about the interaction between TFs and DNA. In 2013, the 3D structure of the conserved domains of the 3 subunits of NF-Y (Fig 4C) was published in complex with an oligonucleotide of 25 bp (Fig 4D), derived from the promoter of the HSP70 gene with a resolution of 3,1 Å (Fig 4A).

The NF-YB and NF-YC head-to-tail assembly provides a stable scaffold ready for the association of NF-YA. The Histone Fold Domains (HFDs) of NF-YB and NF-YC are formed by 3 helices (α1, α2 and α3) separated by 2 loops (L1 and L2). The interaction between the HFD dimer of NF-Y and DNA occurs through van der Waals and electrostatic interactions established between the highly basic upper surface of NF-YB/NF-YC dimer and negatively charged DNA sugar phosphate backbone (Fig 4B). Interactions between α1-α1 contribute to stabilize the two L1-L2 loops and the binding of the dimer to DNA. One of the major differences from the H2A/H2B dimer is the presence of a negatively charged groove assembled by αC and α1 helices of NF-YC and α2 helix of NF-YB, that represents the docking site for NF-YA (Fig 4A).

The NF-YA A1 helix contacts the HFD dimer and is connected to A2 helix, that directly contacts the CCAAT sequence, through the A1A2 linker. In particular, Phe289 of YA (in A2 helix) is positioned in the minor groove between C and A in the CCAAT-box. This causes, together with the motif GxGGRF, following the A2 helix of YA, a global DNA bending of about 80°. As a result, NF-Y conjugates the HFD capacity to form a stable binding to DNA with sequence specificity given by NF-YA (Nardini *et al*, 2013).

The binding of NF-Y to DNA induces a significant bending of DNA, that could promote the binding of other TFs in the adjacent mayor/minor grooves, because it could make the close DNA regions more accessible.

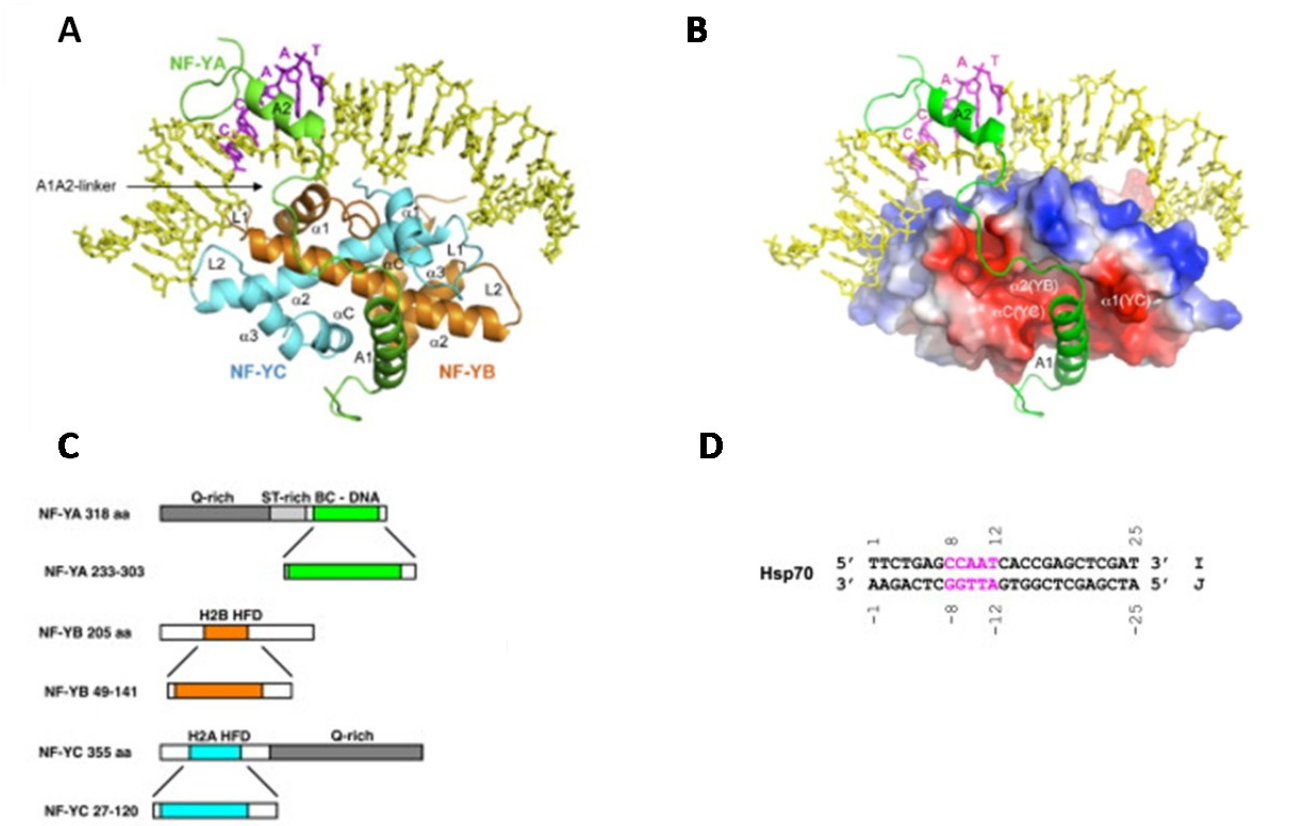


Figure 4: 3D structure of conserved domains of NF-Y bound to DNA. A) Structure of the binary complex NF-Y/DNA. NF-YA in green, NF-YB in orange and NF-YC in cyan, CCAAT nucleotide in magenta. B) Electrostatic surface of NF-YB/NF-YC dimer, the blue and red colours indicate positively and negatively charged regions respectively. C) NF-Y domains used for structure determination. D) Sequence of DNA derived from HSP70 promoter containing CCAAT sequence (Nardini *et al*, 2013).

1.4 NF-Y FUNCTIONS

The function of NF-Y, as for other TFs, is to regulate the expression of genes that present in their promoter the consensus sequence CCAAT-box. In literature, many examples are reported to elucidate the function of NF-Y and to characterize cluster gene targets. The genes regulated by NF-Y belong to different classes: cell cycle-related genes, development genes and human disease-related genes, that include examples important for hematopoietic disorders and cancer.

The role of NF-Y in the modulation of cell cycle genes, and consequently in cancer, was analyzed in different ways: NF-Y regulates the expression of topoisomerase II alpha, cdc2, cyclins and cdc25C genes and first of all its *liaison* with p53 (usually mutated in 50% of human cancers; Beroud and Soussi, 1998 and Hainaut and Hollstein, 2000) is known.

The topoisomerase II alpha promoter is usually activated during the late S and G2/M phases of the cell cycle and it is involved in different cellular functions, like replication or chromatin condensation. Qianghua Hu et al (2002) demonstrated that NF-Y activates the promoter of topoisomerase II alpha. The activity of topoisomerase II alpha promoter decreases considerably after the destruction of 3 out of 7 CCAAT-boxes present in its promoter, resulting in a block of cellular cycle.

Other examples of cell cycle genes directly regulated by NF-Y are Cyclin B1, Aurora A and CDK1 that are genes increased in various human tumors. The promoters of these genes are characterized by the presence of the CCAAT-box, a cell-cycle dependent element (CDE) and a cell cycle homology region (CHR). Hu *et al* (2006) demonstrated that the transactivation domain of NF-YA is specifically needed for Cyclin B1, Aurora A and CDK1 transcription activation and consequently for cell cycle progression at G2/M phase.

The importance of NF-Y role in cell cycle progression could also be noted relying on its interplay with p53 (Fig. 5).

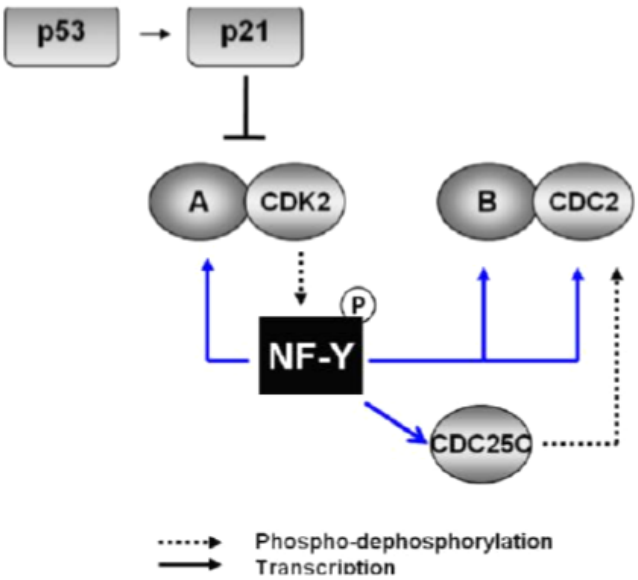


Figure 5. *Liaison* between NF-Y and p53 (Hee-Don Chae, *et al*, 2011)

After DNA damage, the p53 cascade is activated: this causes the arrest of cell cycle at G2 checkpoint through p21-mediated inactivation of CDK2. Consequently, the CDK2-dependent NF-Y

phosphorylation is inhibited and therefore the NF-Y dependent transcription of cell cycle G2-regulatory genes (Yun J *et al,* 2003).

In general, NF-Y is indicated as a TF that checks cell cycle controlled genes (Elkon *et al*, 2003) (Fig 6).

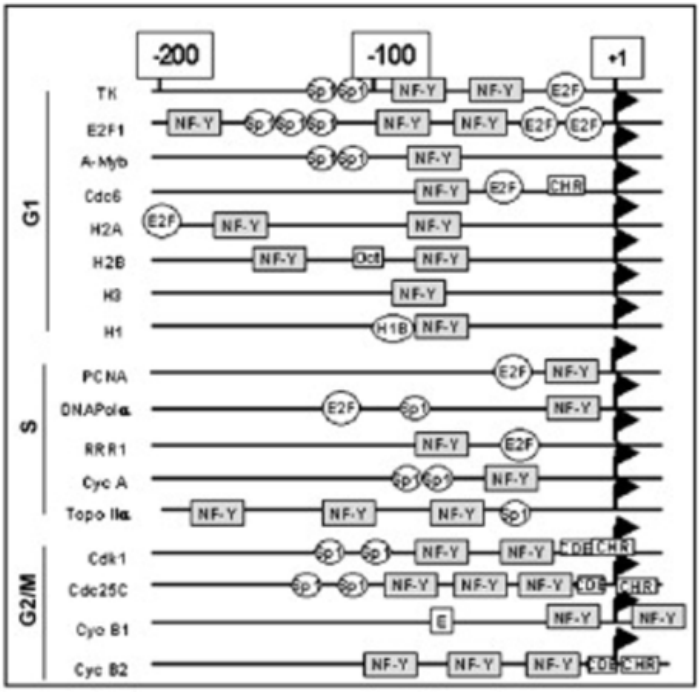


Figure 6. Scheme of some cell cycle genes regulated by NF-Y. (Di Agostino *et al*, 2006)

Emerson reported an example of the role of NF-Y in the development and differentiation of hematopoietic stem cells (HSC). Using an animal model, he demonstrated that deletions in NF-YA domains cause the stop of HSC in G2/M phase and the subsequent apoptosis. These effects are coupled with the deregulation of multiple genes that include: cell cycle genes like cyclin b1 an p21, apoptosis genes like Bcl-2, self-renewal genes like HOXB4 and Notch. These effects are independent by p53 cascade activation (Emerson, 2011).

From another point of view, NF-Y can be classified as a "pioneer". In fact, it can bind DNA in an inactive and condensed chromatin context to allow the binding of other TFs (Fleming, 2013). The pioneer feature of NF-Y is also supported by its interaction with co-activators and co-repressors, including chromatin modifiers enzymes and chromatin remodelers (Donati , 2008). Indeed, NF-Y is able to promote the establishment of specific positive or negative histones post-translational modifications (PTMs) along the genomic regions it controls. In addition, we must not forget that the similarity of NF-Y to histones is not limited to its structural properties, but reflects also the combination of PTMs that occurs for histone proteins (within the HFD) and that is crucial for their

behaviour. These modifications include acetylations, ubiquitynations (Nardini, 2013) (Fig 8) and phosphorylations (Yun, 2003).

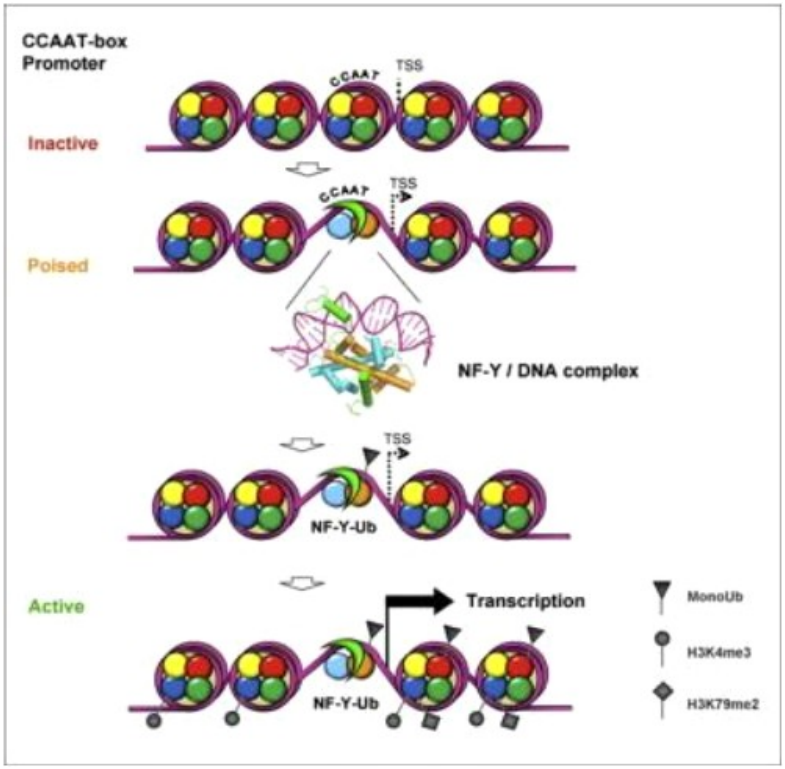


Figure 7: Cartoon representation of NF-Y induced establishment of active chromatin environment. (Nardini *et al*, 2013)

1.5 THE CCAAT-BOX AND OTHER CONSENSUS SEQUENCES: THEIR RELATIVE PRESENCE IN THE HUMAN GENOME

The combination of bioinformatic tools and ChIP-seq data gives us more information about the position of the CCAAT-box in human promoters and the CCAAT-box, bound by NF-Y, in a genome-wide vision. This allows to analyze genes regulated by NF-Y more in details, and to better characterize one of the most important roles of NF-Y in transcription that is to interact synergistically with other transcription factors.

The eukaryotic promoter is a DNA sequence characterized by several regulatory sequences recognized by different TFs and RNA polymerases. The CCAAT-box is one of the most typical elements present in about 30 % of eukaryotic promoters. The genomic study done by Dolfini *et al* highlights some features of CCAAT: position, orientation, correlation with CpG islands, the co-occurrence of NF-Y sites and other TFs and finally the identification of positional correlation between CCAAT and selected TFs sites. They started collecting experimentally validated NF-Y binding sites from Medline and obtained the p-CCAAT matrix (p as promoters) that was used to

optimize the analysis of ChIP on ChIP data. RefSeq promoters were excluded from these analyses, to have a matrix reflecting the regions far from promoters. After alignment of different oligos, they obtained a new matrix called g-CCAAT (g as genomic). Crossing p-CCAAT and g-CCAAT data analyses, they found that the flanking regions of the pentanucleotide are very similar, and now we can extend the consensus sequence: S-R-R-C-C-A-A-T-S-R-S-N-V-N-S-S.(S: C or G, R: Purine; N: any nucleotide; V: A, C or G) (Fig 8A).

Another important aspect that was investigated is the position of the CCAAT-box versus TSS, including the analysis of the position of TATA-BOX. They found that the CCAAT-boxes are preferentially found in -80 region, instead TATA box is localized at -25/-30 (Fig 8B).

To analyze the presence of other consensus sequences near the CCAAT-box, they selected the 20% of the top CCAAT promoters and analyzed them with Pscan algorithm using the matrices of the JASPAR database. At the top of the list there are consensus sequences recognized by SP1, E2F and CREB. In addition, another class of TFBS emerged: E-BOX (CACGTG). Last but not least, they determined the distance between CCAAT and E-BOX that it is around 20 bp (Fig 8C and D).

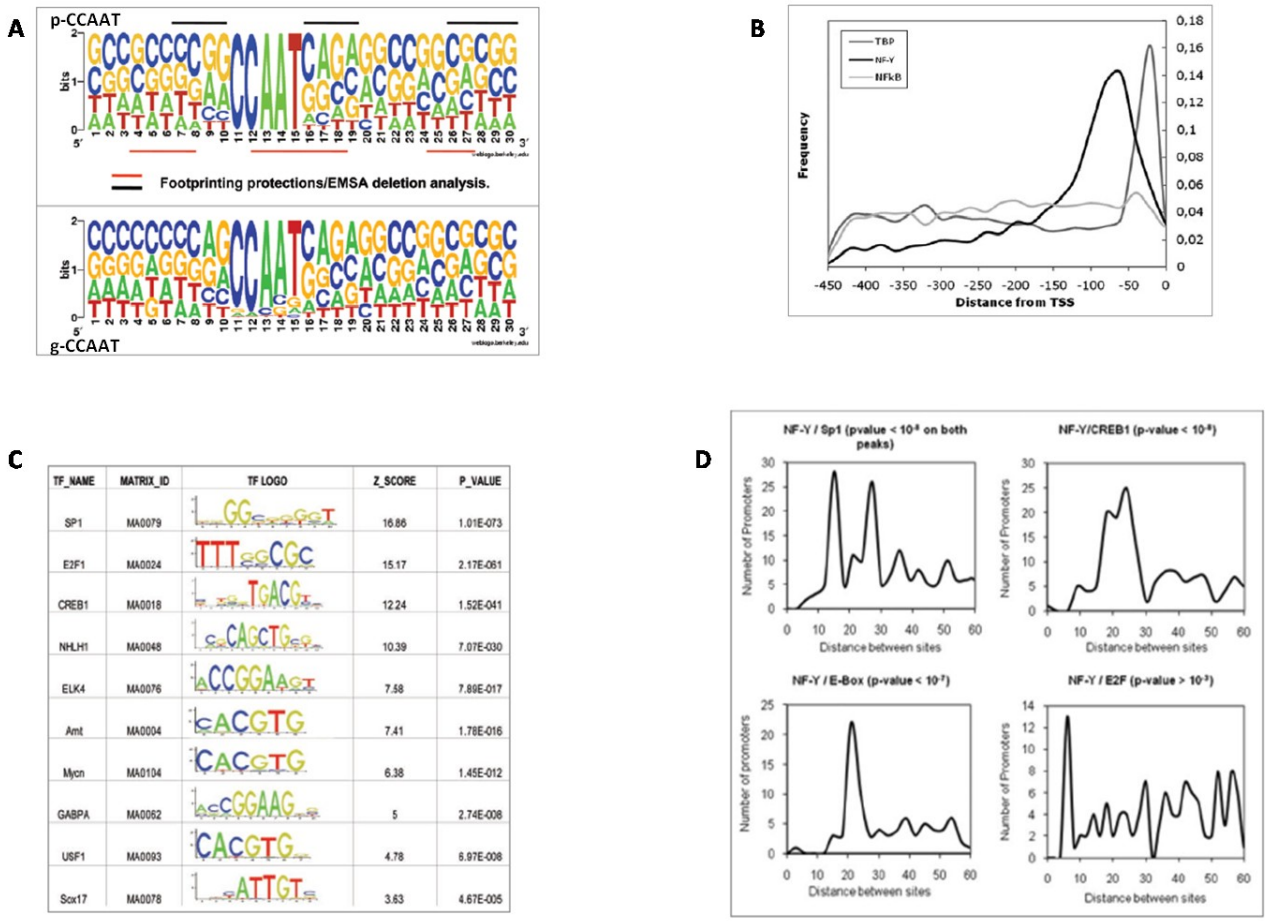


Figure 8: The CCAAT promoters features. A) Flanking region derived from p-CCAAT and g-CCAAT. B) Position of CCAAT-box and TATA box versus TSS. C) The consensus sequences for other TFs found in the top of

20% CCAAT promoters. D) Distance between CCAAT-box and other consensus sequences in the promoters analyzed (Dolfini *et al,* 2009).

The subsequent analysis, performed *in vivo* by Fleming *et al*, gives us more information about the relationship between NF-Y and other TFs that present their consensus sequences near the CCAAT-box. The data analyzed derived from ChIP-seq performed on 3 different cell lines: GM12878, HeLa S3 and K562. This data revealed that NF-Y targets include genes of cellular signalling (like MAPK12, MAPK13), DNA repair (base excision and homologues recombination), cell cycle (G2/M checkpoint, regulation of DNA replication), metabolic (cholesterol biosynthesis, polyamines) and expression.

NF-Y can bind the CCAAT-box also far from promoter regions like in tissue specific Enhancers and LTR/nonmodified-chromatin. The role of NF-Y in positions that are far from promoters was verified through silencing of NF-YA by lentiviral small hairpin RNA. These analysis revealed that the effect of NF-YA silencing regards also those genes (89% of down expressed and 61% of up expressed) that have NF-Y binding sites distal from promoters. The presence of NF-Y binding in LTR regions, in an heterochromatin context, is probably due to the fact that LTR proviral sites are under strong control of the host organism. Probably, NF-Y participates in inhibiting their moving in the genome in somatic tissues and/or in their activation during the embryonic development where many repetitive elements are expressed.

The study also clarified the relationship between NF-Y and other TFs at promoter and enhancer level finding 44 TFs at the promoters and 50 at the enhancers (Fig 9A). At the promoter level, the group that mostly interacts with NF-Y is composed by Myc, Pol2, FOS and others. At the enhancer level, NF-Y interacts with FOS, USF1/2, MAX and others. All E-BOX TFs present in ENCODE are statistically enriched at NF-Y locations, and interestingly MAX/NF-Y bound promoters exceed than Myc/NF-Y bound promoters, probably because Myc uses another system to bind DNA when there is NF-Y. In LTR regions they found colocalization of NF-Y sites and USF sites, suggesting that this two TFs can bind nonmodified chromatin. To investigate if also a specific distance exists, they considered TATA-box, E-BOX, E2F and AP-1 motifs and plotted them with NF-Y picks (indicated as predicted in Fig 9), after this they considered the real sites bound by TATA, E-BOX TFs and others using the ChIP seq data (indicated as verified in Fig 9). It was clear that the E-BOX consensus sequence is present at -12/-11 bp from the CCAAT-box in a specific configuration and that the positioning is maintained if we consider MAX and USF. For Myc it is different because the predicted sites do not correspond to the verified ones, probably because Myc, in some cases, binds DNA indirectly, maybe through NF-Y (Fig 9B).

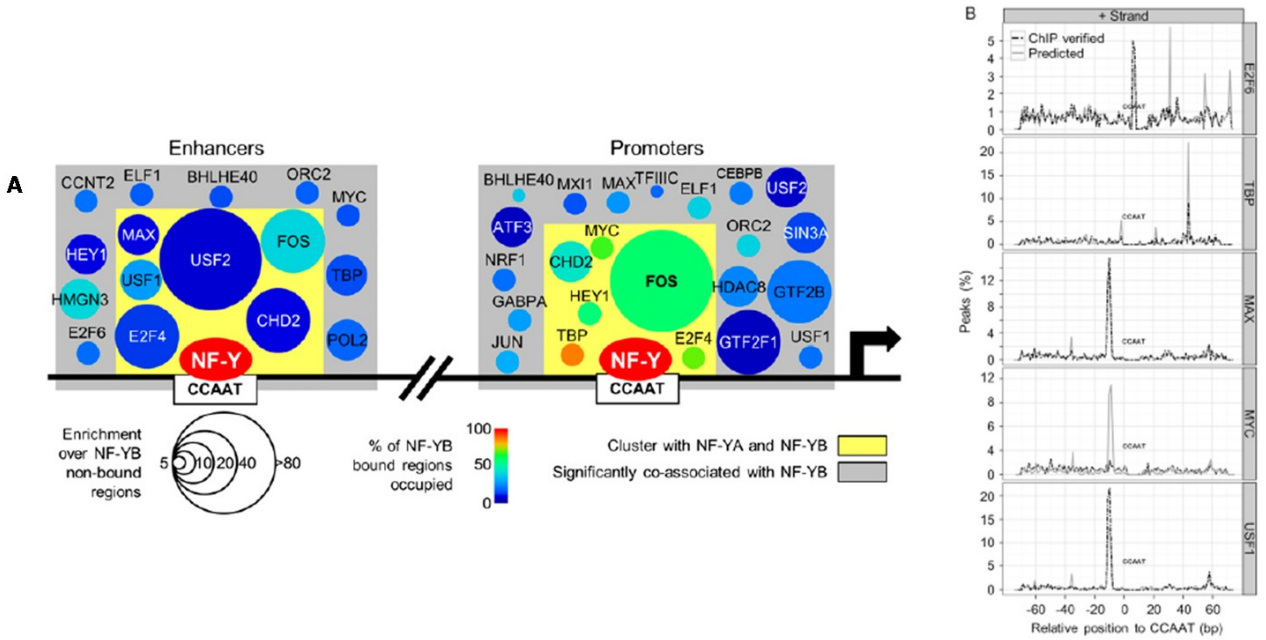


Figure 9: Interaction between NF-Y and other TFs. A) Interaction between NF-Y and other transcription factors at promoters and enhancer levels. B) The E-BOX TFs binding in all genomic regions near the CCAAT box bound by NF-Y; the predicted peaks are reported together with the verified ones (Fleming, 2013).

These two studies reveal three important new aspects: (i) in CCAAT promoters, with high relevance, there are other TFBSs, with one of the most frequent being the E-BOX, recognized by E-BOX TFs like Myc, MAX and USF1; (ii) the distance and the orientation of the 2 consensus sequences is constant and it is of 11/12 bp; (iii) *in vivo,* these 2 sequences are bound by the respective TFs. Nevertheless, USF1 and NF-Y can also bind together the nonmodified chromatin of the LTR12 sequence, probably modelling LTR moving. The subsequent question is: could an interplay between NF-Y and E-BOX TFs (Myc, MAX and USF1) exist in those genomic regions in which the 2 consensus sequences are separated by 11/12 bp?

1.6 E-BOX BINDING FACTORS: Myc, MAX AND USF

The E-BOX binding factors are a class of transcription factors that recognize a specific consensus sequence called as E-BOX: CANNTG (N: A, T, G or C). This class includes USF (Upstream Stimulatory Factor) (Ferre-D'Amare AR *et al*, 1994 - Ayer DE, 1993), c-Myc (Amati B, 1993 - Littlewood TD, 1992), MAX, MAD, TFEB (Carr CS, 1990 - Muhle-Goll C, 1995) among others.
These proteins contain specific domains: a basic-helix-loop-helix domain, necessary for DNA binding and a leucine zipper domain, necessary for homo/hetero-dimerization. These transcription factors usually bind DNA as dimers (Fig 10).

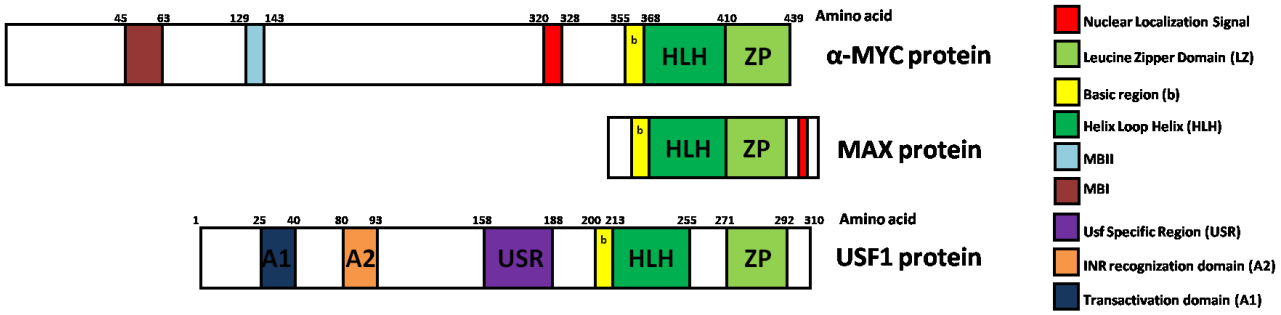


Figure 10: The E-BOX TFs MAX, MYC and USF1. The different domains are indicated.

E-BOX TFs can regulate the expression of different genes, for example the network MAX/Myc/MAD is essential for the regulation of cellular cycle progression (Fig 11).

The heterodimerization of Myc/MAX forces the expression of S phase progression genes (Littlewood TD, 1992 - Carr CS, 1990) and inhibits differentiation and entry into a quiescent state (Carr CS, 1990); instead the heterodimerization of MAX/MAD negatively regulates cell growth (Austen M, 1997). While Myc or MAD can't bind DNA as homodimers, MAX can, but its role is still not clear (Fig 11).

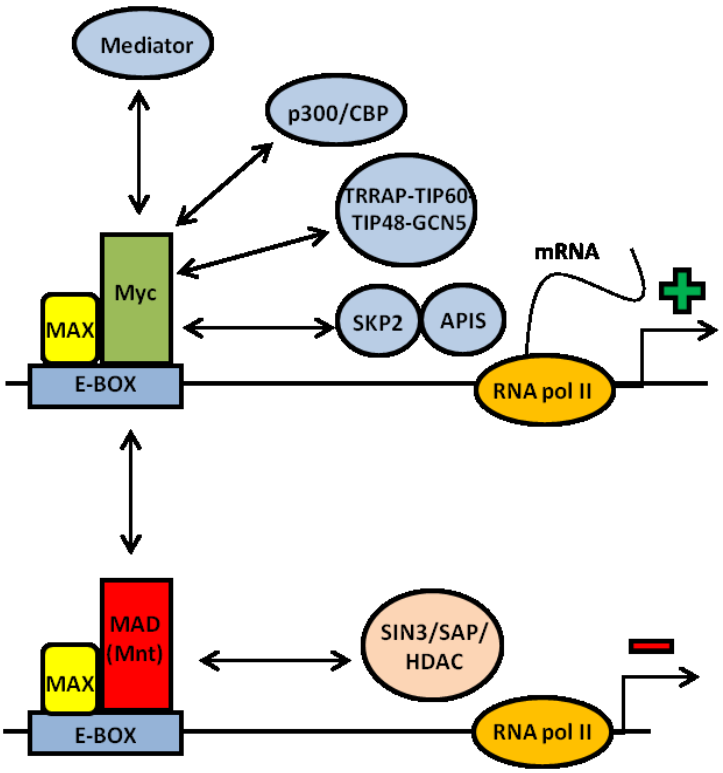


Figure 11. The network MAX/Myc/MAD regulates the expression of target genes. The heterodimerization of Myc/MAX forces expression, instead the heterodimerization of MAX/MAD negatively regulates cell growth. Myc and MAD can't bind DNA as homodimers, MAX can, but its role is unknown.

The genes that code for USF1 and USF2 are localized on chromosome 1 and chromosome 19, respectively. These proteins usually interact with DNA as heterodimers but also form homodimers. The presence and abundance of heterodimers (USF1/USF2) or USF1 homodimers vary in different cell types and at the various stages of cellular differentiation (Sirito M *et al,* 1994). USF1/2 can regulate different genes, like cell cycle components including cyclin-dependent kinases and cyclins. USF can also have antagonizing function for oncogenes like Myc, E1A and Ras and can activate the expression of tumor suppressors like p53, BRCA2, and APC. USF can also recruit different chromatin-modifying enzymes which are thought to establish active chromatin marks and to prevent the spread of heterochromatin into the active globin gene *locus* (Anantharaman *et al,* 2011).

The 3D structures of MAX/MAX, USF1 and Myc/MAX minimal domains bound to DNA are known.

- The 3D structure of MAX/MAX bound to DNA tells us that the dimer directly contacts the major groove of DNA with its alpha-helical basic region. This symmetric homodimer is composed by four-helix bundles coupled two by two through a loop. The two alpha-helical segments are composed of the basic region plus helix 1 and helix 2 plus the leucine repeat, respectively. The leucine repeat forms a parallel coiled coil (Ferre-D'Amare *et al,* 1993) (Fig 12A).
- The USF b/HLH/LZ structure is similar to the b/HLH/LZ domain of MAX and reveals that the b/HLH DNA binding domain homodimerizes, forming a parallel, left-handed four-helix bundle, and that the basic region becomes alpha-helical on binding to the major groove of the DNA. The b/HLH/LZ DNA binding domain of USF exists as a bivalent homotetramer that could be present in cell at the USF physiological intranuclear concentration and depends on the integrity of the leucine zipper motif (Ferre-D'Amare *et al*, 1994) (Fig 12B).
- The b/HLH/LZ domains of Myc/MAX and MAD/MAX heterodimers bound to the E-BOX have been determined at 1.9 Å and 2.0 Å resolution, respectively. The Myc/MAX heterodimer, but not its MAD/MAX counterpart, dimerizes to form a bivalent heterotetramer, which explains how Myc can upregulate the expression of genes with promoters bearing widely separated E-BOXes (Nair, S.K *et al,* 2003) (Fig 12C).

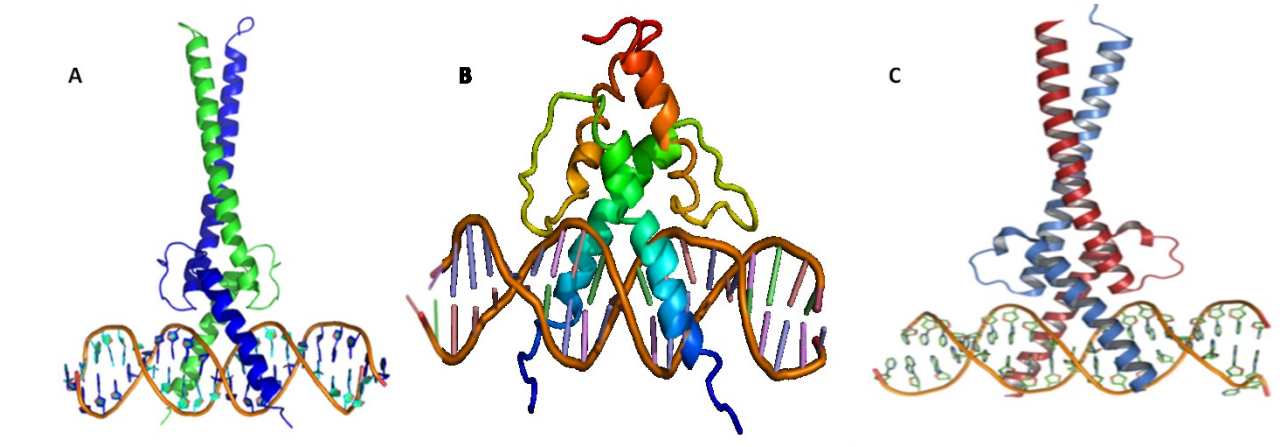


Figure 12: 3D structures of E-BOX TFs on DNA. A) MAX/MAX and DNA B) b/HLH/LZ domains of USF1 and DNA C) b/HLH/LZ domains of Myc-MAX (Ferre-D'Amare *et al,* 1993, 1994; Nair SK *et al,* 2003)

1.7 RELATIONSHIP BETWEEN NF-Y AND E-BOX TRANSCRIPTION FACTORS

Many examples of interplays between NF-Y and E-BOX TFs are reported in literature.

For example, Taira *et al* suggested that the complex composed by Myc and NF-Y is essential for the activation of HSP70 promoter. In particular, when the level of Myc is high, during G1 phase, the complex composed by Myc-NF-Y activates the promoter; instead, when the level of Myc is low another molecule that interacts with Myc turns off the HSP70 promoter and the expression of that gene is repressed. They also demonstrated the direct interaction between c-Myc and the transactivation domain of NF-YC.

On the other hand, Izumi *et al* demonstrated that Myc-NF-Y binding on PDGF β receptor promoter is necessary for its repression. A direct interaction between the transactivation domain of NF-YC and the transactivation domain of Myc exists.

Ito *et al* demonstrated that NF-Y and USF1 cooperate in the regulation of the expression of CP27 gene. This gene codes for a chromatin factor required for Embrionic Stem (ES) cells pluripotency.

Another example of the cooperation between NF-Y and USF1 is on the SVCT2 exon1a promoter. This gene codes for the human ascorbate transporter. In this case, the two consensus sequences are separated by 12 bp. Qiao *et al* demonstrated that the integrity of the CCAAT-box is essential for promoter activity and that NF-Y and USF1 cooperate in DNA binding to activate the expression of the target gene. They also demonstrated that the spacing between the 2 consensus sequences is essential for the activation of the promoter, because 4 additional bp are sufficient to lose the activity of the reporter gene in luciferace assays.

Moreover, USF1 and NF-Y cooperate in the regulation of the expression of HOXB4 gene, as demonstrated by Zhu *et al*. HOXB4 is essential for the maintenance of staminality of HSC. Also in

this case the integrity of the CCAAT-box is crucial for the cooperative binding with USF1. The interesting thing is that the HOX4 genes all present this configuration: E-BOX—10bp—CCAAT-BOX ; instead other HOX genes, as HOX7 gene, present the following configuration: E-BOX—12bp—CCAAT. HOXB4 is expressed only in hematopoietic cells, while other HOX genes, like HOXB7 , are normally expressed and contribute to the correct development of different tissues and organogenesis. Meccia *et al* demonstrated that NF-Y and USF1 are essential for the regulation of the expression of HOXB7 gene.

1.8 HOX GENES

Homeobox genes comprise a super-family of evolutionarily conserved genes, which play essential roles in controlling body plan specification and cell fate determination (Dhwani Haria and Honami Naora 2013). These genes are characterized by the presence of a 180- nucleotide sequence element (the homeobox) which encodes a 60 amino-acid structural motif, the homodomain.

This homodomain recognizes specific DNA elements containing a TAAT motif. Through this recognition, proteins coded by Homeobox genes can regulate the expression of different genes that usually specify structures along the anteroposterioir axis in bilaterians. This function was discovered for the first time in *Drosophila melanogaster*, in which Homeobox gene mutations caused formation of body parts in inappropriate contexts (Gehring & Hiromi, 1986; McGinnis & Krumlauf, 1992).

In humans, mutations in these genes cause a wide range of complex developmental abnormalities, including limb malformations and sensory defects (Mortlock & Innis, 1997).

Different sets of Homeobox genes control skeletal patterning, limb and craniofacial morphogenesis and development of virtually all organ systems (McGinnis & Krumlauf, 1992; Panganiban & Rubenstein, 2002). In adults, Homeobox genes also regulate tissue regeneration and play critical roles in controlling self renewal and differentiation of hematopoietic progenitors (Argiropoulos & Humphries, 2007 12.McGinnis & Krumlauf, 1992).

Mammalian Homeobox genes are categorized into several families, that are named following their homologues in the fly. Examples include DLX (distal-less), PAX (paired), MSX (muscle segment), CDX (caudal), EN (engrailed) and OTX (orthodenticle).

Most Homeobox genes are dispersed throughout the genome, whereas the members of the mammalian HOX and DLX gene families are organized in clusters (Fig 13).

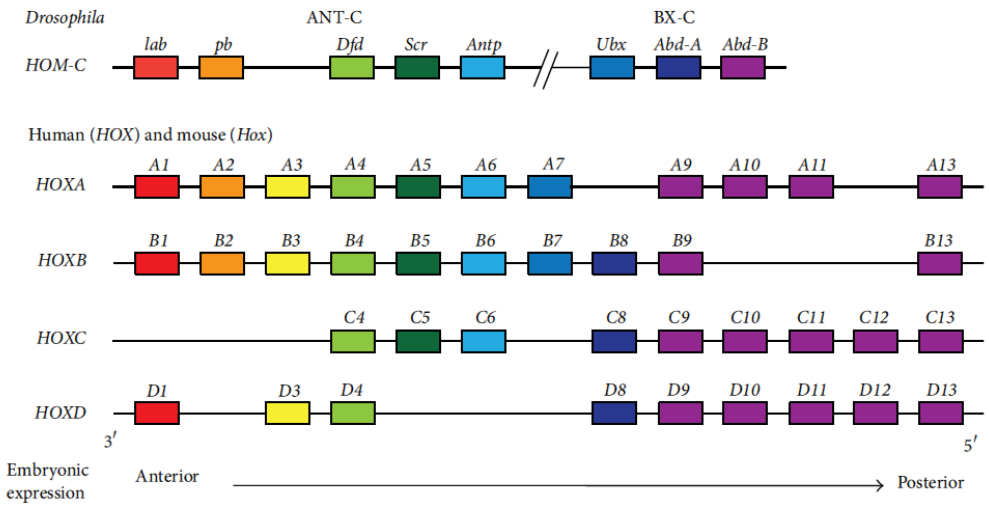


Figure 13: Arrangement of HOX genes in the *Drosophila* and mammalian genomes. 39 HOX genes are divided into four separate clusters (HOXA, HOXB, HOXC and HOXD) on four different chromosomes. In each cluster, HOX genes are tandem arranged in sequence from 3' to 5'. HOX genes with the same number are referred to as paralogs. (Taniguchi, 2014)

AIM OF THE THESIS

The aim of this work is to discover the possible interplay between NF-Y and selected E-BOX TFs and to characterize its molecular features.

The genomic data, ChIP-seq data and 3D structural analyses, give us many information about the existence of the interplay between NF-Y and Myc, MAX and USF1. Therefore, we decided to study this relationship with different approaches: EMSA, SAXS and ChIP assays.

EMSAs allow us to study the formation of ternary complexes composed by NF-Y, E-BOX TF and DNA *in vitro*. Moreover, through this technique we can characterize the features of the binding of ternary complexes, using as DNA probes natural configurations emerged from genomic data (10/12 bp spacing between CCAAT and E-BOX). We can also study the protein domains involved in the interaction using different constructs that include different domains of NF-Y and E-BOX TFs. We can also study the importance of DNA configurations using, as probes, artificial DNA sequences in which the two sites are separated by 15 or 17 bp.

Instead, SAXS allow us to obtain an approximate 3D structure of different ternary complexes to better understand the molecular details of the relationship, when it exist.

Finally, ChIP assays on HeLa cells silenced for NF-Y, give us information about the *in vivo* relationship between NF-Y and E-BOX TFs at the genomic level, to verify and validate *in vitro* data.

MAIN RESULTS

1.1 Studies on the relationship between NF-Y and USF1, MAX and Myc *in vitro*.

1.1.1 Expression and purification of E-BOX Binding factors as recombinant soluble proteins in *E. coli.*

In order to study NF-Y/ TFs interactions *in vitro*, we first set up the conditions for the expression and purification of selected E-BOX TFs: USF1 and MAX as homodimers and Myc/MAX as heterodimers. As already discussed in the introduction, these proteins were chosen based on known functional interactions with NF-Y and availability of their 3D structures in complex with DNA.

I set up and optimized specific protocols for the expression and purification of all E-BOX recombinant proteins present in this work. E-BOX TFs coding sequences were cloned in pET expression vectors that exploit the T7 RNA polymerase system and allow the induction of protein expression with IPTG (0.1-0.4 mM). In all cases, purification was obtained through in-batch Ni^{2+}- affinity chromatography, given the presence of a C-terminal or N-terminal 6His-tag. The optimization of the different purification protocols allowed me to obtain a good protein yield for each E-BOX TF and a high purity of the final protein preparations.

The recombinant E-BOX TFs produced were the following:

1.Full length MAX homodimers (aa 1-151). The recombinant cDNA fragment encodes for a protein with a molecular weight of 18.8 KDa , with the 6His-tag fused at the C-Terminus. Yield : 8 mg of protein / 1 Lt cell culture

2. Full length USF1 homodimers (aa 1-310). The recombinant cDNA fragment encodes for a protein with a molecular weight of 36.3 KDa, with the 6His-tag fused at the N-Terminus. Yield: 0.7 mg of protein / 1 Lt cell culture

3. Full length Myc/MAX heterodimer was also produced, however the purified protein lacked DNA binding activity and was not further used in the following analyses.

4. Full Length NF-Y heterotrimer (NF-YA: aa 1-318; NF-YB: aa 1-205; NF-YC: aa 1-355): The protocol for the expression and purification of NF-Y was already available in the laboratory. The 3 subunits are expressed individually in *E.coli* and are purified from inclusion bodies. The subsequent refolding occurs *in vitro* through dialysis steps using an appropriate ratio among protein subunits.

In order to better define the proteins domains involved in the interactions, we also produced the following minimal domains constructs:

5. USF short (USF1s) homodimers (aa 106-310). The recombinant cDNA encodes for a protein of 24,6 KDa that includes the b/HLH/LZ domains, with the 6His-tag fused at the N-Terminus. Yield:

6mg of protein / 1 Lt cell culture. This construct lacks a portion of USF1 N-terminal domain, in particular the regions A1 and A2, that correspond to transactivation domain and initiator element recognition domain, respectively.

6. Myc minimal domain/ MAX (Myc md/MAX) heterodimer (Myc: aa 353-433; MAX: aa 1-151). The construct allows the coexpression of the DNA binding domain of Myc and full length MAX, encoding 2 proteins with a molecular weight of 11,2 KDa and 16,6 KDa, respectively. The 6His-tag is located at the C-Terminus of Myc md. Yield: 2.4 mg of protein / 1 Lt cell culture.

7. Myc342/MAX heterodimer (Myc: aa 335-433; MAX: aa 1-151). This construct, similar to construct 6, contains a larger portion at the N-Terminus of the DNA binding domain of Myc, and allows the coexpression of the 2 proteins with a molecular weight of 12,9 KDa and 16,6 KDa for Myc and full length MAX, respectively. The 6His-tag is located at the C-Terminus of Myc 342. Yield: 8 mg of protein / 1 Lt cell culture.

6. NF-Ymd heterotrimer (NF-YA: aa 233-303; NF-YB: aa 49-141; NF-YC aa 27-120). The protocol for the expression and purification of NF-Ymd heterotrimer was already available in the laboratory. In this case, the 3 subunits are co-expressed as soluble heterotrimer in *E. coli* and purified by affinity chromatography, followed by gel filtration, after cleavage of the 6His-tag (Nardini et al, 2013).

1.1.2 Study of molecular interplays between NF-Y and E-BOX TFs *in vitro*, using Electro Mobility Shift Assay (EMSA)

The EMSA is based on the slower mobility, during native gel electrophoresis, of a DNA probe bound by one or more proteins respect to the free DNA. We used this technique to evaluate *in vitro* the formation of ternary complexes composed by DNA/NF-Y/E-BOX TF, and to characterize the nature of their binding properties. As DNA probes we selected natural genomic sequences that display the most representative configurations of the E-BOX and CCAAT elements, identified by genomic and ChIP-seq data analyses, in which the two consensus sequences are separated by 10/12 bp. We choose 2 different types of DNA sequences: two derived from HOX genes promoters, and one derived from ERV9 LTR region present in human chromosome 5. In particular the DNA sequences used as probes were the following:

- A 48 bp portion of the HOXB4 promoter containing E-BOX and CCAAT box separated by 10 bp.

- A 50 bp portion of the HOXB7 promoter containing E-BOX and CCAAT box separated by 12 bp.
- A 50 bp portion of the ERV9 LTR, present in human chromosome 5, containing E-BOX and CCAAT box separated by 12 bp.

All probes are labelled on one strand with a Cy5 fluorophore at the 5' end, necessary for the detection. The fluorescent gels images were acquired with the MP Chemidoc Imager.

First of all we set up the conditions to verify the specific formation of a single binary complex of each TF on their target sequences. As expected, all TFs bound their consensus sequences.

Secondly we set out to determine the eventual formation of the ternary complexes by dose response assays: with the DNA and one of the two TFs concentration kept constant, the second TF is added at increasing concentration to evaluate the formation of band migrating slower than the single binary complexes, possibly representing the complex composed by DNA, NF-Y and one E-BOX TF.

Once we demonstrated the complexes we observed containing both NF-Y and the consider E-BOX TF, we set up the conditions to characterize them. The ideal condition when we incubate two proteins together with DNA is that we can detect 4 complexes on gel: one of free probe, 2 belonging to binary complexes and the forth is the ternary complex. In this way, during completion assays, each complexes can be follow in each condition and in timing.

The dose response assay give us following information:

- The E-BOX TFs purified can bind their consensus sequence CACGTG on the different probes used.
- NF-Y can form ternary complexes with E-BOX TFs and probes
- The ternary complexes are purposely formed by NF-Y an E-BOX TFs, because using specific unlabelled oligos carrying the CCAAT or E-BOX elements, the slower migrating band is competed together with binary complexes. The specificity is confirmed by the fact that mutated CCAAT box or E-BOX can't compete the respective binary complexes binding and the ternary one.
- Considering that the ternary complexes can be efficiently formed *in vitro* we exploited their stability to develop large scale purification protocols to form and isolate ternary complexes suitable for structural analyses that were performed through Small-Angle X-ray Scattering (SAXS), in collaboration with Prof Nardini. This technique allows to determine the molecular envelop of a concentrated solute when it is present in non aggregate state. We have purified and isolated the complexes: MAX/NF-Ymd/HOXB4

and Mycmd/MAX/NF-Ymd/HOXB4 which were measured at EMBL Petra synchrotron in Hamburg; the data were collected and are currently under analysis.

In order to better characterize the features of the interplay among NF-Y and E-BOX TFs, we performed OFF RATE assays.

This type of EMSA allows us to have information about the stability of ternary complexes over time and to compare it with that of binary complexes. After the formation of protein/DNA complexes, we add an excess of unlabelled CCAAT or E-BOX competitor in reaction, and load the samples on the migrating gel at different time points. As a result, we will obtain a stairway migration of the samples and we can analyze the behavior of each binary or ternary complex, which will disappear with different timings depending on the stability of the complex itself.

Using these assays, we decided to classify the binding of NF-Y with USF1, MAX or Myc/MAX on different probes, in three classes:

- Not cooperative: if there is no difference in stability over time between binary and ternary complexes (on the gel, the bands of binary and ternary complexes disappear at the same time)
- Facilitated: if the ternary complex is more stable in time than the binary complex (on the gel, the band of ternary complex is still detected also when the band of binary complex has already disappeared)
- Cooperative: if the ternary complex is firmly present for a time longer than what observed for the facilitated binding (on the gel, the band of ternary complex is detectable to the end of analysis, while the band of binary complex is not)

  With our OFF RATE analyses, we can conclude that:

  - On HOXB4 probe, when together, USF1 and NF-Y display a facilitated binding, instead NF-Y with MAX homodimers or Myc/MAX heterodimer do not.
  - The facilitated binding observed on HOXB4 between NF-Y and USF1 becomes more stable on HOXB7 probe. In addition, no facilitated or cooperative binding was observed with NF-Y/MAX or NF-Y/Myc/MAX.
  - Surprisingly, the only cooperative binding observed was that displayed by NF-Y and USF1 on LTR sequence. Also in this case NF-Y/MAX or NF-Y/Myc/MAX didn’t show any cooperative or facilitated behaviors.

To verify which the domains of NF-Y and USF1 were involved in the interaction, we used NF-Ymd and USFs to perform the same OFF RATE analyses on HOXB4, HOXB7 and LTR probes. We observed that the minimal domains maintain the behavior already observed with full length proteins: facilitated binding on HOXB4, higher facilitated binding on HOXB7 and cooperative binding on LTR.

To study how the spacing of the two consensus sequences may influence the stability of NF-Y/USF1 ternary complexes, we used two artificial DNA sequences in which the binding sites are separated by 5 extra bp, one derived from HOXB4, resulting in a 15 bp spacer length, and the other derived from LTR sequence, in which the 2 elements are separated by 17 bp. These 5 bp entailed an additional half helix turn of DNA. Using OFF RATE assays, we observed that this is sufficient to lose the facilitated binding observed on HOXB4 and the cooperative binding on LTR sequence.

1.2 Studies on the relationship between NF-Y and USF1, MAX and Myc *in vivo*.

1.2.1 NF-Y contribution on *in vivo* DNA binding of E-BOX TFs

To understand the contribution of NF-Y on the E-BOX TFs DNA binding *in vivo*, we analyzed through ChIP assays the binding of Myc, MAX and USF1 on their target promoters, after NF-Y silencing, in HeLa cells.

To cause the dissociation of NF-Y from its target promoters, we silenced NF-YB subunits through lentiviral delivery of shRNAs, in fact it is sufficient to silence one of the three subunits to avoid a functional trimer formation. A non-target shRNA was used as control (shCTR). Viruses were collected with the culture medium after 36 hours post transfection of HEK 293T cells with lentiviral packaging plasmid. HeLa cells were infected and collected at 72 hours post infection. NF-YB silencing was assessed by Western blotting and we could also confirm that the expression of the E-BOX binding factors of interest was not influenced by NF-YB silencing.

For each antibody, we performed 2 ChIP assays with chromatin derived from 2 independent NF-YB silencing experiments. The subsequent analysis on TFs DNA binding were performed using quantitative Real Time PCR (qPCR). In particular, we focused on target gene promoters which display a distance between the two elements of 11 bp. In general, we can observe that NF-Y silencing causes an effect on E-BOX TFs DNA binding, and such effects depend on the efficiency of the NF-YB silencing.

In particular, in NF-YB silenced HeLa samples, we can observe, as expected, that the binding of NF-Y is decreased, on all its target promoters analyzed.

MAX was found on LMNB1, HOXB7, MLX; USF1 on LMNB1, HOXB7, PTPDC1, MLX and SERF2, Myc on MLX, SERF2 e CCNB1. After NF-YB silencing, the binding of USF1 and MAX on LMNB1 and PTPDC1, that are negative for NF-Y, doesn't change. Instead on HOXB7 and MLX, the binding of USF1 and MAX decreases together with NF-YB binding. On SERF2 the binding of USF1 increases after NF-YB silencing. The binding of Myc on MLX, SERF2 and CCNB1 decreases with the binding of NF-Y, instead on ACTL6a, the binding of Myc increases after NF-YB silencing.

In conclusion, we observed that NF-YB silencing causes an effect on USF1, Myc and MAX binding and this effect is different if we consider the different promoters.

DISCUSSION AND FUTURE PERSPECTIVES

The regulation of transcription is the result of the interaction among different TFs that bind the regulatory sequences present in the promoters of target genes allowing their activation or repression in the different cellular contexts. NF-Y is one of the most important TFs that regulates the expression of different genes. NF-Y is a "pioneer" TF, because its binding to DNA recruits the chromatin modification complexes and allows the execution of genetic programs. The most important role of NF-Y in transcription is to interact synergistically with other TFs. The new genome-wide maps, obtained by ENCODE project, of TFs loci, highlight the architecture of different human promoters. These data pointed out that near the CCAAT box, bound by NF-Y, there are several specific consensus sequences recognized and bound by a specific set of TFs. The E-BOX is one of the most represented elements found in CCAAT promoters. Moreover, it was observed that the distance between CCAAT box bound by NF-Y and E-BOX bound by some of the E-BOX TFs is constant and it is 10-12 bp.

The aim of this study was to discover the possible relationship between NF-Y and selected E-BOX binding factors: Myc, MAX and USF1; to characterize the features of DNA binding Myc, MAX or USF1 in presence of NF-Y, and to understand the molecular basis of these interplays on specific genomic regions that present a 10-12 bp distance between the two relative elements. To do this, we decided to use both *in vitro* and *in vivo* approaches.

For *in vitro* studies, we used purified TFs proteins and specific DNA configurations as probes in EMSAs. The DNA configuration used for these assays derived from HOX genes and LTR region. The HOX genes are involved in different fundamental processes for organism development, like HOXB7, and for the differentiation of specific stem cells lineages, like HOXB4 gene. In fact, the up-regulation of the first is usually associated to tumorigenesis, and misregulation of HOXB4 expression in hematopoietic stem cells, causes problems in their differentiation. The LTR sequences are portions of DNA although found in heterochromatic state contain sequences recognized by different TFs like SP1, AP-1, NF-kB. Recent data explained the role of these repetitive sequences, that, in total, represent 40% of the human genome: in some cases they allow accessory regulatory sequences for gene expression, on the other hand, it was also demonstrated that their hypometilation is associated with specific types of cancer . In all these cases NF-Y plays a fundamental role, because it was reported that it modulates the expression of HOXB4 and HOXB7 genes and that CCAAT bound by NF-Y is one of the most represented elements in LTR sequences.

From our *in vitro* study, it emerges that NF-Y can form stable and specific ternary complexes with all E-BOX TFs, and on all natural DNA configurations considered. More importantly, regarding

DNA binding, no type of negative interplays were detected. Our EMSA analyses allow us to classify the binding of the observed ternary complexes in 3 classes: independent, facilitated and cooperative. We found that the binding of MAX/NF-Y/DNA is independent in all configuration considered, as is the binding of Myc/MAX/NF-Y/DNA. Instead, the binding of USF1/NF-Y/DNA complexes is facilitated on the analyzed gene promoters, and cooperative on the ERV9 LTR. Moreover, the interaction between NF-Y and USF1 involves their minimal domains, because they are sufficient to observe these facilitated/cooperative behaviours. Another important aspect is that the interactions observed are dependent on the precise spacing between the two consensus sequences. In fact, when the binding sites are artificially positioned on opposite configuration, by the addition of 5 bp (half helix turn) the facilitated or cooperative binding observed on 10 or 12 bp, disappears.

The interactions observed *in vitro* for USF1, were also confirmed *in vivo*. In fact, after NF-Y silencing, the genomic binding of USF1 is affected, usually decreased, on target promoters analyzed. In addition, for Myc and MAX, the data obtained *in vivo* tell us that an interplay with NF-Y also exists.

Therefore, we conclude that a physical interaction involving the minimal domains of USF1 and NF-Y can occur on specific DNA configurations. Considering the 3D structures available for all E-BOX TFs, we can note that the loops of USF1 b/HLH/ZP domains are longer and tend to expand more on DNA than those of MAX and Myc/MAX b/HLH/LZ, it could be possible that the interaction with NF-Y might occur through a specific conformation of USF1 loops. However we still need to define the USF1 protein regions involved.

The interplay between NF-Y and MAX observed *in vivo*, but not *in vitro*, can be possibly explained because another molecule or chromatin modifications could take part in this interplay; for Myc we conclude that the interplay with NF-Y didn't involve its minimal domain, in fact, in literature it was already reported that NF-Y and Myc interact through their transactivation domains, but also in this case an interplay exists.

The future work will include the isolation and purification of the ternary complex composed by DNA/NF-Y md/USF1s to obtain its 3D structure. In parallel, to better characterize the regions involved in the interaction with NF-Y, we could reduce the protein domains of USF1.

Another aspect that we could investigate is the analyses of histone modifications, given the fact that NF-Y is a pioneer transcription factor. We could verify if the silencing of NF-Y is associated to a specific chromatin state modifications, and if it could be related to a local recruitments or dissociation of other TFs.

We can conclude that a relationship among NF-Y and E-BOX binding factors is actually present not only at the promoter level, but also in LTR regions. Considering HOX gene promoters, we must remember that HOXC4 and HOXD4 promoters present the same configuration of HOXB4 promoter, instead HOXA4 presents the sequences configuration seen in HOXB7 and HOXC7, so our considerations on the interplay between NF-Y and USF1 could be extended also to these other targets. The different processes in which the expression of these genes are involved, are different and include differentiation, cell cycle regulation and genomic remodelling.

Through this study we can better understand the molecular bases of TFs interactions, and how they could impact on the gene expression programs and, more in general, on the complexity of genome organization.

5.REFERENCES

Amati B., Littlewood TD., Evan GI., Land H. The c-Myc protein induces cell cycle progression and apoptosis through dimerization with Max. 1993 *EMBO J.* **15;12(13):5083-7.**

Anantharaman A., Lin IJ., Barrow J., Liang SY., Masannat J., Strouboulis J., Huang S., Bungert J. Role of helix-loop-helix proteins during differentiation of erythroid cells. 2011 *Mol Cell Biol*. **31(7):1332-43**.

Argiropoulos & Humphries. Hox genes in hematopoiesis and leukemogenesis. 2007 *Oncogene* **26(47):6766-762007.**

Ayer DE., Kretzner L., Eisenman RN. Mad: a heterodimeric partner for Max that antagonizes Myc transcriptional activity. 1993 *Cell*. **29;72(2):211-22**.

Béroud C., Soussi T. p53 gene mutation: software and database.1998 *Nucleic Acids Res*. **26(1):200-4.**

Brivanlou and Darnell, Signal transduction and the control of gene expression. 2002 *Science*. **295(5556):813-8.**

Bungartz G, Land H., Scadden DT., Emerson SG. NF-Y is necessary for hematopoietic stem cell proliferation and survival. 2012 *Blood*. **119(6):1380-9.**

Carr CS., Sharp PA. A helix-loop-helix protein related to the immunoglobulin E box-binding proteins. 1990 *Mol Cell Biol*. **10(8):4384-8.**

Ceribelli M., Benatti P., Imbriano C., Mantovani R., NF-YC complexity is generated by dual promoters and alternative splicing. 2009 *J Biol Chem* **284(49):34189-200.**

Dhwani Haria and Honami Naora Homeobox Gene Deregulation: Impact on the Hallmarks of Cancer**.** 2013 *Cancer Hallm* **1(2-3):67-76.**

Di Agostino S., Strano S., Emiliozzi V., Zerbini V., Mottolese M., Sacchi A., Blandino G., Piaggio G. Gain of function of mutant p53: the mutant p53/NF-Y protein complex reveals an aberrant transcriptional mechanism of cell cycle regulation. 2006 *Cancer Cell*. **10(3):191-202.**

Dolfini D., Zambelli F., Pavesi G., Mantovani R. A perspective of promoter architecture from the CCAAT box. 2009 *Cell Cycle*, **8:24, 4127-4137.**

Dolfini D, Gatta R, Mantovani R., NF-Y and the transcriptional activation of CCAAT promoters. 2012 . *Crit Rev Biochem Mol Biol*;**47(1):29-49**.

Dolfini D, Mantovani R., Targeting the Y/CCAAT box in cancer: YB-1 (YBX1) or NF-Y? 2013 *Cell Death Differ*; **20(5):676-85**.

Donati G., Gatta R., Dolfini D., Fossati A., Ceribelli M., Mantovani R. An NF-Y-dependent switch of positive and negative histone methyl marks on CCAAT promoters. 2008 *PLoS One*. **30;3(4):e2066**.

Elkon R., Linhart C., Sharan R., Shamir R., Shiloh Y. Genome-wide in silico identification of transcriptional regulators controlling the cell cycle in human cells. 2003 *Genome Res* **13:773–780.**

Ferré-D'Amaré AR., Pognonec P., Roeder RG., Burley SK. Structure and function of the b/HLH/Z domain of USF 1994 *EMBO J.* **13(1):180-9**.

Ferré-D'Amaré AR., Prendergast GC., Ziff EB., Burley SK. Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. 1993. *Nature* **363(6424):38-45**.

Fleming JD., Pavesi G., Benatti P., Imbriano C., Mantovani R., Struhl K. NF-Y coassociates with FOS at promoters, enhancers, repetitive elements, and inactive chromatin regions, and is stereo-positioned with growth-controlling transcription factors. 2013 *Genome Res* **23(8):1195-209.**

Gehring & Hiromi, Homeotic genes and the homeobox, 1986 *Annu Rev Genet*. **20:147-73.**

Hackzell A., Uramoto H., Izumi H., Kohno K., Funa K. p73 independent of c-Myc represses transcription of platelet-derived growth factor beta-receptor through interaction with NF-Y. 2002 *J Biol Chem*. **277(42):39769-76**.

Hainaut and Hollstein, p53 and human cancer: the first ten thousand mutations 2000 *Adv Cancer Res*.;**77:81-137**.

Hee-Don Chae, Jeanho Yun, Yung-Jue Bang and Deug Y Shin. Cdk2-dependent phosphorylation of the NF-Y transcription factor is essential for the expression of the cell cycle-regulatory genes and cell cycle G1/S and G2/M transitions. 2004. *Oncogene*. **23:4084–4088.**

Hee-Don Chae, Jungbin Kim & Deug Y. Shin. NF-Y binds to both G1- and G2-specific cyclinpromoters; a possible role in linking CDK2/Cyclin A to CDK1/Cyclin B. 2011. *BMB Rep*. **44(8):553-7**.

Hu Q, Lu JF, Luo R, Sen S, Maity SN Inhibition of CBF/NF-Y mediated transcription activation arrests cells at G2/M phase and suppresses expression of genes activated at G2/M phase of the cell cycle. 2006 *Nucleic Acids Res*.;**34(21):6272-85**.

Hu Q1, Bhattacharya C, Maity SN. CCAAT binding factor (CBF) binding mediates cell cycle activation of topoisomerase IIalpha. Conventional CBF activation domains are not required. 2002 *J Biol Chem*. **277(40):37191-200.**

Ito Y1, Zhang Y, Dangaria S, Luan X, Diekwisch TG. NF-Y and USF1 transcription factor binding to CCAAT-box and E-box elements activates the CP27 promoter Gene. 2011 *Gene* ;**473(2):92-9**.

Lemon and Tjian, Orchestrated response: a symphony of transcription factors for gene control, 2000 *Genes Dev*; **14(20):2551-69**.

Littlewood TD., Amati B., Land H., Evan GI. Max and c-Myc/Max DNA-binding activities in cell extracts. 1992 *Oncogene*; **7(9):1783-92**.

Ly LL, Yoshida H, Yamaguchi M., Nuclear transcription factor Y and its roles in cellular processes related to human disease 2013 *Am J Cancer Res*.;**3(4):339-46.**

Mantovani R, The molecular biology of the CCAAT-binding factor NF-Y. 1999 *Gene*; **239(1):15-27.**

McGinnis & Krumlauf, Homeobox genes and axial patterning, 1992 *Cell*; **68(2):283-302.**

Meccia E., Bottero L., Felicetti F., Peschle C., Colombo MP., Carè A. HOXB7 expression is regulated by the transcription factors NF-Y, YY1, Sp1 and USF-1. 2003 *Biochim Biophys Acta*.; **1626(1-3):1-9**.

Mortlock & Innis, Mutation of HOXA13 in hand-foot-genital syndrome. 1997 *Nat Genet*.; **15(2):179-80.**

Muhle-Goll C1, Nilges M, Pastore A. The leucine zippers of the HLH-LZ proteins Max and c-Myc preferentially form heterodimers. 1995 *Biochemistry*; **34(41):13554-6.**

Nair SK1, Burley SK. X-ray structures of Myc-Max and Mad-Max recognizing DNA. Molecular bases of regulation by proto-oncogenic transcription factors. 2003 *Cell*; **112(2):193-205**.

Nardini M., Gnesutta N., Donati G., Gatta R., Forni C., Fossati A., Vonrhein C., Moras D., Romier C., Bolognesi M., Mantovani R. Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination 2013, *Cell*; **152(1-2):132-43**.

Panganiban & Rubenstein, Developmental functions of the Distal-less/Dlx homeobox genes 2002 *Development*; **129(19):4371-86.**

Qiao H1, May JM. CpG methylation at the USF-binding site mediates cell-specific transcription of human ascorbate transporter SVCT2 exon 1a. 2011 *Biochem J*.; **440(1):73-84**.

Sirito M., Lin Q., Maity T., Sawadogo M. Ubiquitous expression of the 43- and 44-kDa forms of transcription factor USF in mammalian cells. 1994. *Nucleic Acids Res*.;**22(3):427-33**.

Taira T., Sawai M., Ikeda M., Tamai K., Iguchi-Ariga SM., Ariga H. Cell cycle-dependent switch of up- and down-regulation of human hsp70 gene expression by interaction between c-Myc and CBF/NF-Y. 1999 *J Biol Chem*.; **274(34):24270-9**.

Taniguchi, Hox transcription factors: modulators of cell-cell and cell-extracellular matrix adhesion. 2014 *Biomed Res Int*.; **2014:591374**.

Yun J., Chae HD., Choi TS., Kim EH., Bang YJ., Chung J., Choi KS., Mantovani R., Shin DY. Cdk2-dependent phosphorylation of the NF-Y transcription factor and its involvement in the p53-p21 signaling pathway. 2003 *J Biol Chem*.;**278(38):36966-72**.

Zhu J., Giannola DM., Zhang Y., Rivera AJ., Emerson SG. NF-Y cooperates with USF1/2 to induce the hematopoietic expression of HOXB4. 2003 *Blood*.;**102(7):2420-7**

PART II

List of published papers:

Calvenzani V., Testoni B., Gusmaroli G., Lorenzo M.,Gnesutta N., Petroni K., Mantovani R., Tonelli C., (2012)
Interactions and CCAAT-binding of Arabidopsis thaliana NF-Y subunits. *Plos one* **3,73** 7 (8): e 42902

Interactions and CCAAT-Binding of *Arabidopsis thaliana* NF-Y Subunits

Valentina Calvenzani, Barbara Testoni[¤a], Giuliana Gusmaroli[¤b], Mariangela Lorenzo, Nerina Gnesutta, Katia Petroni, Roberto Mantovani*, Chiara Tonelli*

Dipartimento di Bioscienze, Universita` degli Studi di Milano, Milano, Italy

Abstract

Background: NF-Y is a transcription factor that recognizes with high specificity and affinity the widespread CCAAT box promoter element. It is formed by three subunits: NF-YA and the NF-YB/NF-YC- heterodimer containing histone fold domains (HFDs). We previously identified a large *NF-Y* gene family in *Arabidopsis thaliana*, composed of 29 members, and characterized their expression patterns in various plant tissues.

Methods: We used yeast Two-hybrids assays (Y2H), pull-down and Electrophoretic Mobility Shift Assay (EMSA) *in vitro* experiments with recombinant proteins to dissect AtNF-YB/AtNF-YC interactions and DNA-binding with different AtNF-YAs.

Results: Consistent with robust conservation within HFDs, we show that heterodimerization is possible among all histone-like subunits, including the divergent and related LEC1/AtNF-YB9 and L1L/AtNF-YB6 required for *embryo* development. DNA-binding to a consensus CCAAT box was investigated with specific AtNF-YB/AtNF-YC combinations and observed with some, but not all AtNF-YA subunits.

Conclusions: Our results highlight (i) the conserved heterodimerization capacity of AtNF-Y histone-like subunits, and (ii) the different affinities of AtNF-YAs for the CCAAT sequence. Because of the general expansion of NF-Y genes in plants, these results most likely apply to other species.

Citation: Calvenzani V, Testoni B, Gusmaroli G, Lorenzo M, Gnesutta N, et al. (2012) Interactions and CCAAT-Binding of *Arabidopsis thaliana* NF-Y Subunits. PLoS ONE 7(8): e42902. doi:10.1371/journal.pone.0042902

Editor: Pierre-Antoine Defossez, Universite´ Paris-Diderot, France

Received May 24, 2012; Accepted July 12, 2012; Published August 17, 2012

Copyright: ®2012 Calvenzani et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by a Progetto di Rilevante Interesse Nazionale (PRIN) grant to RM and by European Research Area-Plant Genomics, Fondo Italiano Ricerca di Base (ERA-PG FIRB) to CT. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist. * E-mail: mantor@unimi.it (RM); chiara.tonelli@unimi.it (CT)

¤a Current address: Centre National de la Recherche Scientifique (CNRS), Lyon, France

¤b Current address: Department of Science and Mathematics, University of South Carolina Beaufort, Bluffton, South Carolina, United States of America

Introduction

The CCAAT box is one of the most ubiquitous promoter elements, being present in many, if not most of eukaryotic promoters [1]. Typically, it is found between −60 and −100 base-pairs from the transcriptional start site. The functional importance of the evolutionarily conserved consensus pentanucleotide has been widely established in several experimental systems. Twenty years of biochemical and genetic analyses have clarified that NF-Y [HAP2/3/5 in yeast] is a trimeric protein complex composed of NF-YA [HAP2], NF-YB [HAP3] and NF-YC [HAP5]. All subunits are required for DNA-binding and conserved throughout evolution [2]. NF-YB/NF-YC belong to the class of Histone Fold Domain [HFD] proteins, forming a tight dimer, structurally similar to H2A/H2B, with DNA-binding interaction modules [3]. Heterodimerization results in the formation of a surface for NF-YA association, allowing the resulting trimer to bind DNA with high specificity and affinity. The *fungi* HAP complex activates transcription through an additional subunit, HAP4, containing an acidic activation domain [4,5], unlike the mammalian NF-YA and NF-YC subunits which display large domains rich in Glutamines with transcriptional activation potential [6,7]. In plants, NF-Y also consists of three subunits and we and others have identified and classified them in *Arabidopsis* [8–10], and other species [11–15]. In general, plants have large families of genes, differentially expressed in various tissues: typically, 4–6 members are abundant and ubiquitous, while the others are restricted to certain tissues or developmental stages.

Genetic experiments were initially described for LEAFY COTYLEDON 1 (LEC1, AtNF-YB9) which has a role in *embryo* maturation and specification of cotyledon identity, with a unique pattern of expression confined to *embryos* ([16–18], reviewed in [19]). A LEC1 related member, L1L/AtNF-YB6, was shown to be able to partially complement the *lec1* defect [20], and chimeric constructs demonstrated that the HFD domain is necessary and sufficient for LEC1 function in *embryos* [17]. The LEC1 homologues have similar roles in carrot [21,22] and *Theobroma cacao* [23]. Genetic analysis of AtNF-YA5 mutants indicate that it is involved in both ABA and blue light responses, together with LEC1 [24], and in drought resistance [25], similarly to AtNF-YB1 and YB2 and maize ZmNF-YB2 [26,27]. AtNF-YB2 and AtNF-

YB3 are important for flowering [28–31], and MtHAP2-1 regulates symbiotic nodules in *Medicago truncatula* [32].

The growing wealth of genetic data is poorly matched by biochemical advancements. The presence of 29 *bona fide* NF-Y genes in the *Arabidopsis* genome could potentially result in the formation of .900 alternative heterotrimeric combinations with different DNA-binding capabilities: the most obvious questions are whether there is specificity in interactions and whether all combinations are capable to bind to the CCAAT box. DNA-binding has been scored with carrot LEC1, one cNF-YB and two cNF-YCs [33], with OsHAP3A (NF-YB), six OsHAP5s (NF-YC) and one OsHAP2 [13], and AtNF-YB2 and AtNF-YB3 coupled to yeast HAP2 and HAP3 subunits [30]. A recent systematic study conducted on *Arabidopsis* NF-Y subunits using Y2H assays reached the following conclusions [34]: (i) the HFD subunits do not homodimerize, (ii) they heterodimerize among them, with a notable degree of specificity, and (iii) AtNF-YAs can only bind to HFD dimers, and not to single subunits. The last point was expected, given the wealth of previous biochemical and genetic work. To clarify the stunning complexity of this system, we undertook Y2H assays, *in vitro* pull-down and Electrophoresis Mobility Shift Assay (EMSAs), reporting the interaction map and DNA-binding activity of 24 members of the *Arabidopsis* NF-Y gene family.

Results

Yeast Two-Hybrids assays

Since NF-YB and NF-YC are known to form a tight heterodimer, whose interaction generates an optimal surface for NF-YA association, we used Y2H assays to systematically dissect the ability of each member of the AtNF-YB and AtNF-YC family to interact with each other. The bait and prey vectors contained the GAL4 DNA-binding domain (DBD) and GAL4 activation domain (AD), respectively. For each pair of AtNF-YB/AtNF-YC constructs, the Yeast Two-Hybrid interactions were tested in both configurations, to minimize the possibility of false positive and negative results. For both *NF-Y* gene families, we used the full length cDNAs corresponding to all *AtNF-YB* and *AtNF-YC* genes previously classified [9]. Three readouts were considered: His, Ade and LacZ, each driven by a different promoter under the control of the GAL4 responsive elements. Figure 1A shows the results of the different combinations with AtNF-YCs fused to the GAL4 DBD, and AtNF-YBs to GAL4 AD. On the other hand, Figure 1B shows the result obtained with AtNF-YCs fused to the GAL4 activation domain and AtNF-YBs to GAL4 DBD. Note that, in both cases, 3-AT was added to the yeast medium to minimize the growth due to self-activation. A first result is that the vast majority of the NF-YB and NF-YC family members can interact with each other in this *in vivo* assay. The only exception to this general observation is LEC1/AtNF-YB9, which does not interact significantly with any of the AtNF-YCs, in both configurations (Fig. 1), except for a suboptimal interaction with AtNF-YC3 and only with the AD configuration (Fig. 1A). A weaker interaction can be observed between specific pairs, like AtNF-YB2/AtNF-YC6 and AtNF-YB3/AtNF-YC7, in both configurations. Other pairs with suboptimal affinity are AtNF-YB2/AtNF-YC2, AtNF-YB3/AtNF-YC2, AtNF-YB3/AtNF-YC6, AtNF-YB4/AtNF-YC7 and AtNF-YC3/AtNF-YB10 (Fig. 1B). To further confirm these interactions and better quantify their strength, liquid Y2H Assays were performed by measuring b-GAL activity under conditions of exponential growth. For the liquid assay, we used the AtNF-YB (DBD) and AtNF-YC (AD) configurations shown in Figure 1B. The results of these experiments are shown in Figure 2. As previously determined by in plate assays, the liquid assay confirmed that LEC1/AtNF-YB9 (DBD) does not significantly interact above background levels with any AtNF-YC subunits. The liquid assay confirmed the weak interactions detected by the in plate assay. Furthermore, it was possible to detect a couple of additional weak interactions between AtNF-YB7/AtNF-YC6 and AtNF-YB10/AtNF-YC6. On the other hand, AtNF-YB1, AtNF-YB5 and AtNF-YB6, and to a lesser degree AtNF-YB2, showed robust interactions with all AtNF-YC family members. Overall, this set of experiments indicate that the vast majority of the HFD combinations heterodimerize, with few very specific exceptions.

In vitro analysis

The negativity of LEC1/AtNF-YB9, unable to interact with any AtNF-YC, and the positivity of L1L/NF-YB6, which binds to all partners, are not expected. To substantiate the Y2H assays, we produced and purified recombinant proteins, as well as *in vitro* produced proteins by transcription and translation [TnT] of different subunits (Fig. S1). We chose AtNF-YB2 and AtNF-YA6 because they are rather "conventional"structure-wise when compared to the mammalian homologues. AtNF-YC were mixed with an excess of His-tagged recombinant AtNF-YB2 and loaded on NTA-Nickel columns. Figure 3 shows the results of such experiments. As expected, control columns did not retain any AtNF-YC subunit in the bound fractions in the absence of AtNF-YB2 (Fig. 3A and 3B, lanes 5). On the other hand, all AtNF-YCs were bound, with varying degrees of efficiency, in the presence of AtNF-YB2 (Fig. 3A and 3B, lanes 3), or L1L/AtNF-YB6 (Data not shown). While this assay is not quantitative, it does confirm that the two AtNF-YBs are able to retain on the column all AtNF-YCs, consistent with the results obtained by the Y2H assay. In the same assay, AtNF-YA6 was also retained with different AtNF-YC combinations when His-tagged AtNF-YB2 was added (Fig. 3B), indicating that interactions are observed in the presence of the three subunits.

Having shown that most HFD subunits are able to interact both *in vivo* and *in vitro*, the next relevant question concerns the affinity of combinations for the CCAAT box. To answer this question, recombinant proteins were produced by TnT and used in EMSAs with a consensus, high affinity NF-Y oligonucleotide [1]. In Figure 4, several members of the *Arabidopsis* subunits were first assayed in the presence of the mouse NF-YA/NF-YC heterodi-mer. As negative controls we used the mouse dimeric combinations alone (Fig. 4A lane 2, Fig. 4B lane 1 and Fig. 4C lane 4). *In vitro* transcribed and translated Luciferase was added to the same mouse dimers as an additional negative control (Fig. 4A, lane 1). Recombinant mouse NF-YA alone was also used as negative control (Fig. 4C, lane 2). Positive controls were the mouse recombinant NF-Y trimer (Fig. 4C, lane 1), and single mouse NF-Y subunits added to the corresponding mouse dimeric combinations: NF-YB to NF-YA/NF-YC (Fig. 4A, lane 9), NF-YC to NF-YA/NF-YB (Fig. 4B, lane 11) and NF-YA to NF-YB/NF-YC (Fig. 4C, lane 11). Surprisingly, none of the AtNF-YBs added to the mouse NF-YA/NF-YC led to the formation of a complex with an electrophoretic activity different from the negative controls (Fig. 4A). In the case of the AtNF-YCs, instead, all subunits generated a discrete band with mouse NF-YA/NF-YB, with mobility somewhat similar to that of mouse NF-Y: the bands were weak for AtNF-YC3, AtNF-YC7 and AtNF-YC8, but quite robust for the other six AtNF-YCs tested. For AtNF-YAs, AtNF-YA2 and AtNF-YA4 were negative, whereas AtNF-YA3, AtNF-YA6, AtNF-YA8 and AtNF-YA9 were all capable of generating bands with mobilities similar to mouse NF-Y. These results indicate that the majority of the AtNF-YA and AtNF-YC members behave as

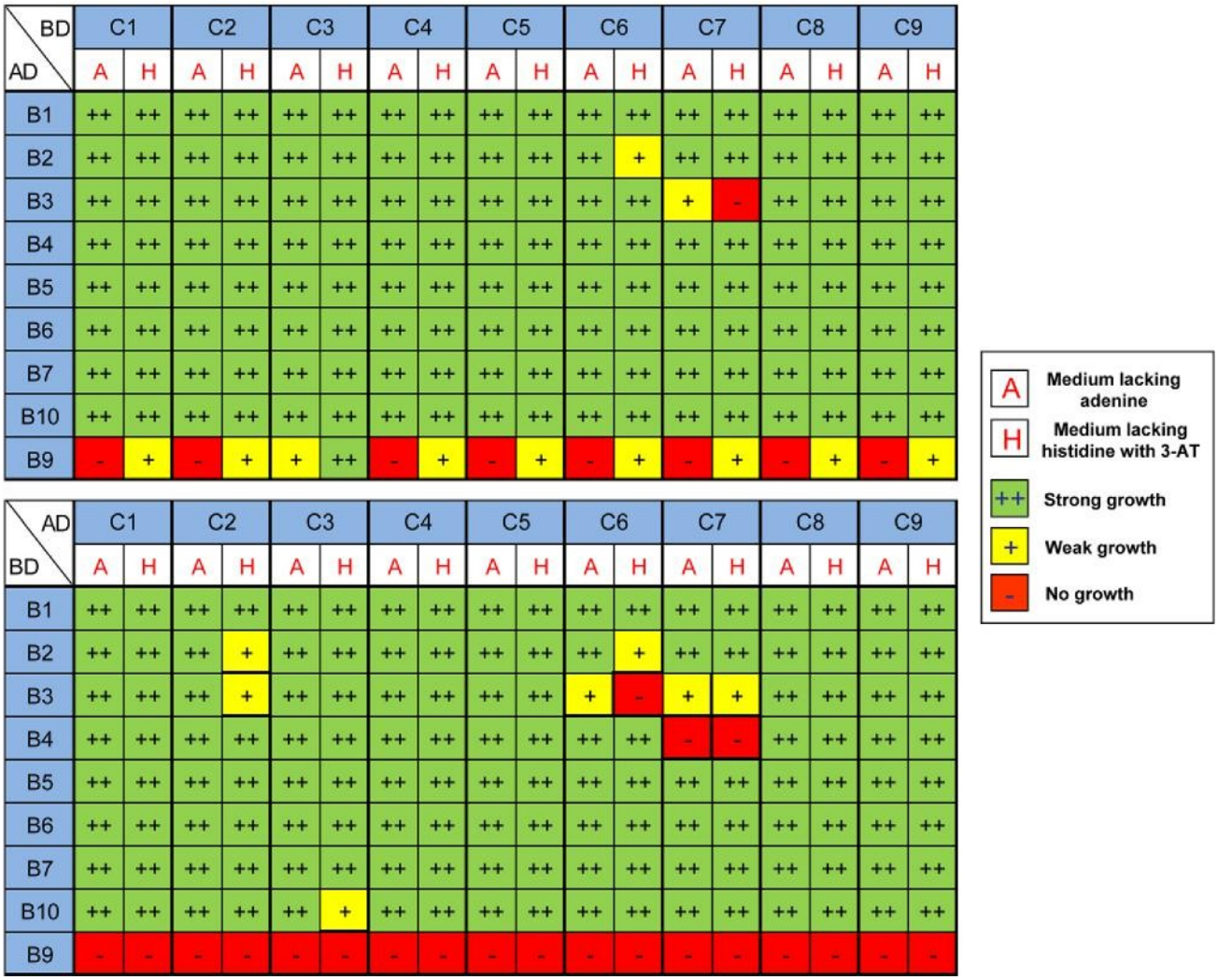


Figure 1. AtNF-YB-AtNF-YC interactions by colony yeast two hybrids assays. A.The indicated AtNF-YCs were fused to the Activation Domain (AD) and tested with AtNF-YB fused to the DNA-binding domain of GAL4. B. Same as A, except that the reverse experiment was tested, namely the AtNF-YBs fused to the Activation Domain were matched to the AtNF-YCs fused to the DNA Binding Domain. ++ refers to robust growth on the selective medium, + weak growth, and 2 no growth.
doi:10.1371/journal.pone.0042902.g001

canonical NF-Ys, as they associate with mouse subunits and bind to the CCAAT box.

The negativity of the AtNF-YBs in the TnT-EMSA assays (Fig. 4A) was troubling: therefore, we decided to investigate whether this was an artefact due to the use of mouse recombinant NF-YA and NF-YC subunits and/or to the TnT system used. First, we selected two AtNF-YBs -AtNF-YB2 and L1L/AtNF-YB6- which are proficient in interactions with all AtNF-YCs according to the Y2H assay. We produced and purified single His-tagged AtNF-YB2 and AtNF-YB6 recombinant proteins in *E. coli*, together with two AtNF-YC subunits, namely AtNF-YC3 and AtNF-YC7 (Fig. S1). The choice of these members were driven by two types of considerations, the first being expression patterns, the second relatedness to mouse subunits: AtNF-YB2 and AtNF-YC3 are the most ubiquitously expressed and less "variant", whereas L1L/AtNF-YB6 and AtNF-YC7 are strictly tissue-specific and the most deviant. The HFD proteins were found in inclusion bodies, as expected, denatured and efficiently renatured when mixed together [35,36]. In one set of experiments, to the *Arabidopsis* NF-YB/NF-YC dimers we added recombinant AtNF-YA6, one of the AtNF-YAs positive in the EMSAs with mouse subunits (Fig. 5A). The *Arabidopsis* NF-YB/NF-YC dimers were also added to recombinant mouse NF-YA (Fig. 5A). AtNF-YA6 is able to generate NF-Y-like bands when AtNF-YC3 dimerized with either AtNF-YB2 or AtNF-YB6; the AtNF-YC7 combinations, on the other hand, yielded either no band or a smeary pattern. The same was essentially observed with mouse NF-YA (Fig. 5A), except that the AtNF-YB6/AtNF-YC3 combination was more efficient in binding, paralleling the efficiency of the mouse NF-Y trimer. The difference in mobilities of At-NF-YA6 and mouse NF-YA complexes are visible and most likely due to the different molecular mass of these two NF-YA proteins (308 and 347 residues, respectively). Again, the AtNF-YC7 combinations gave no band or a smear, indicating that heterotrimers with this subunit are very inefficient in CCAAT-binding. We decided to further dissect the DNA-binding activity of this heterodimer in the presence of other AtNF-YA family members: Figure 5B shows that an NF-Y complex was obtained with AtNF-YA3, AtNF-YA6, AtNF-YA8 and AtNF-YA9. Taken together, the results of Figure 5 are consistent with the set of experiments previously performed by

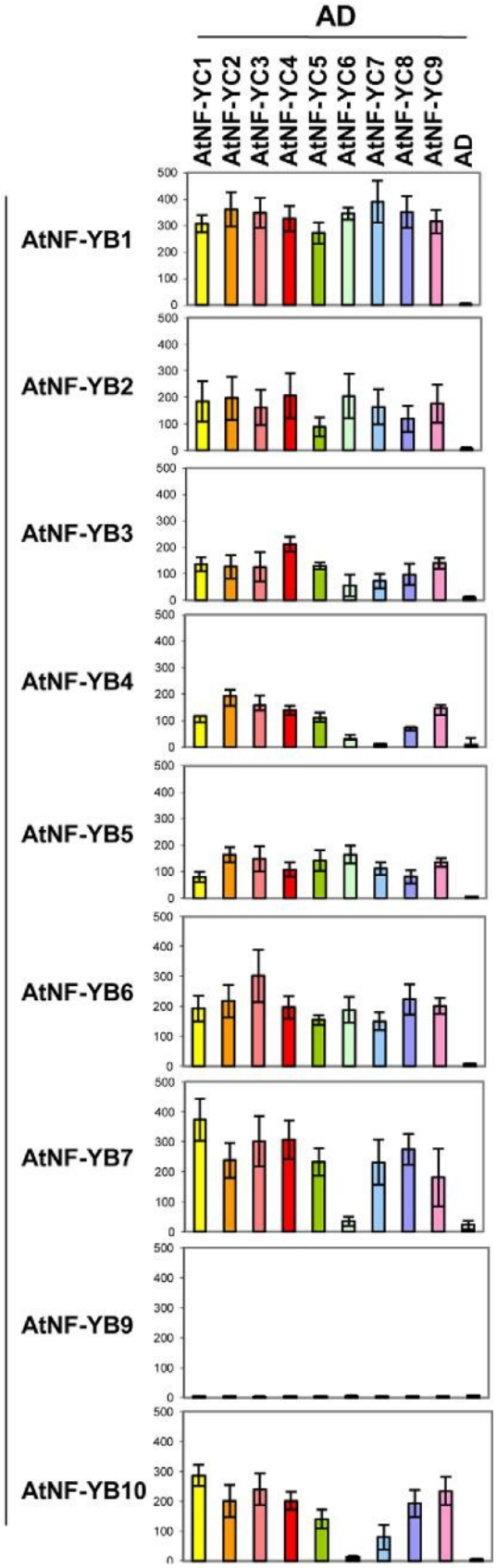


Figure 2. AtNF-YB-AtNF-YC interactions by liquid yeast two hybrids assays. Yeast two hybrids assays in liquid cultures using the AtNF-YB (DBD) and AtNF-YC (AD) configuration are depicted. b-Galactosidase Units were measured as detailed in Materials and Methods. The experiments were repeated three times and the standard deviations are indicated.
doi:10.1371/journal.pone.0042902.g002

using mouse recombinant subunits (Fig. 4). These data indicate that AtNF-YA2 and AtNF-YA4 are either incapable to associate to AtNF-YB6/AtNF-YC3 -and mouse HFD dimers- or to bind to DNA.

L1L/AtNF-YB6 and LEC1/AtNF-YB9 belong to the same clade and are genetically linked. Having shown that L1L/AtNF-YB6 is capable to heterotrimerize and bind to CCAAT, we wondered whether the lack of heterodimerization of LEC1/AtNF-YB9 was due to some artefacts of the Y2H system. We decided to use an *E. coli* coexpression system in which the HFDs of either protein was coexpressed with the HFD of AtNF-YC3: Figure 6A shows that both heterodimers are produced and purified from soluble bacterial extracts. The copurification of (untagged) AtNF-YC3 with the His-tagged AtNF-YBs is a clear sign of hetero-dimerization. Surprisingly, when we expressed LEC1/AtNF-YB9 alone, rather than being confined to inclusion bodies, the protein was very efficiently produced in a soluble form, which is very unusual for HFD proteins. Next, we performed EMSAs with a CCAAT oligonucleotide, using the two heterodimers and mouse NF-YA: Figure 6B shows that both gave shifted bands, with mobilities similar to NF-Y. The affinities were lower with respect to the mouse NF-YB/NF-YC used as positive control, but similar among them. Note that in this particular experiments, we used the minimal heterotrimerization/DNA binding domain constructs consisting of the evolutionarily conserved regions of each subunit [3], with a 31 bp Cy5-labelled probe in Agarose-EMSA, which resulted in faster DNA-protein complexes. AtNF-YB9 alone did not show any DNA binding. Taken together, these data prove that LEC1/AtNF-YB9 can heterodimerize, trimerize with NF-YA and bind to CCAAT as efficiently as its closest relative, L1L/AtNF-YB6.

Discussion

One of the most pressing questions in biology is to understand the interactions of transcription factors among each other and with their natural DNA targets. As they have often evolved in complex families, whose members share some common features and diverge in others, the intricacies of the role of each member needs to be clarified. This is particularly challenging in plants, where genes encoding TFs have expanded to amazing numbers. One such example is NF-Y, the heteromeric CCAAT-binding protein, whose subunits are encoded by single copy genes in most eukaryotes, including mammals [2], while in *Arabidopsis* and other plants each is represented by large families. We began a systematic investigation of the interactions among 24 AtNF-Y subunits, by using Y2H *in vivo* and *in vitro* assays. Some of our data, notably those on LEC1/AtNF-YB9, indicate that negative results of Y2H assays should be confirmed by independent biochemical means, before interactions can be ruled out.

Dimerization

By and large, the *Arabidopsis* NF-Y HFD subunits -AtNF-YBs and AtNF-YCs- are able to heterodimerize, whether by Y2H assays or by *in vitro* interactions. These results are not surprising, based on considerations made on the crystallographic structure of the mouse NF-YB/NF-YC dimer [3], and of similar HFD dimers, including H2A/H2B [35]. These analyses have revealed that the a2 helix of the HFD is the core of the dimerization surface, thanks

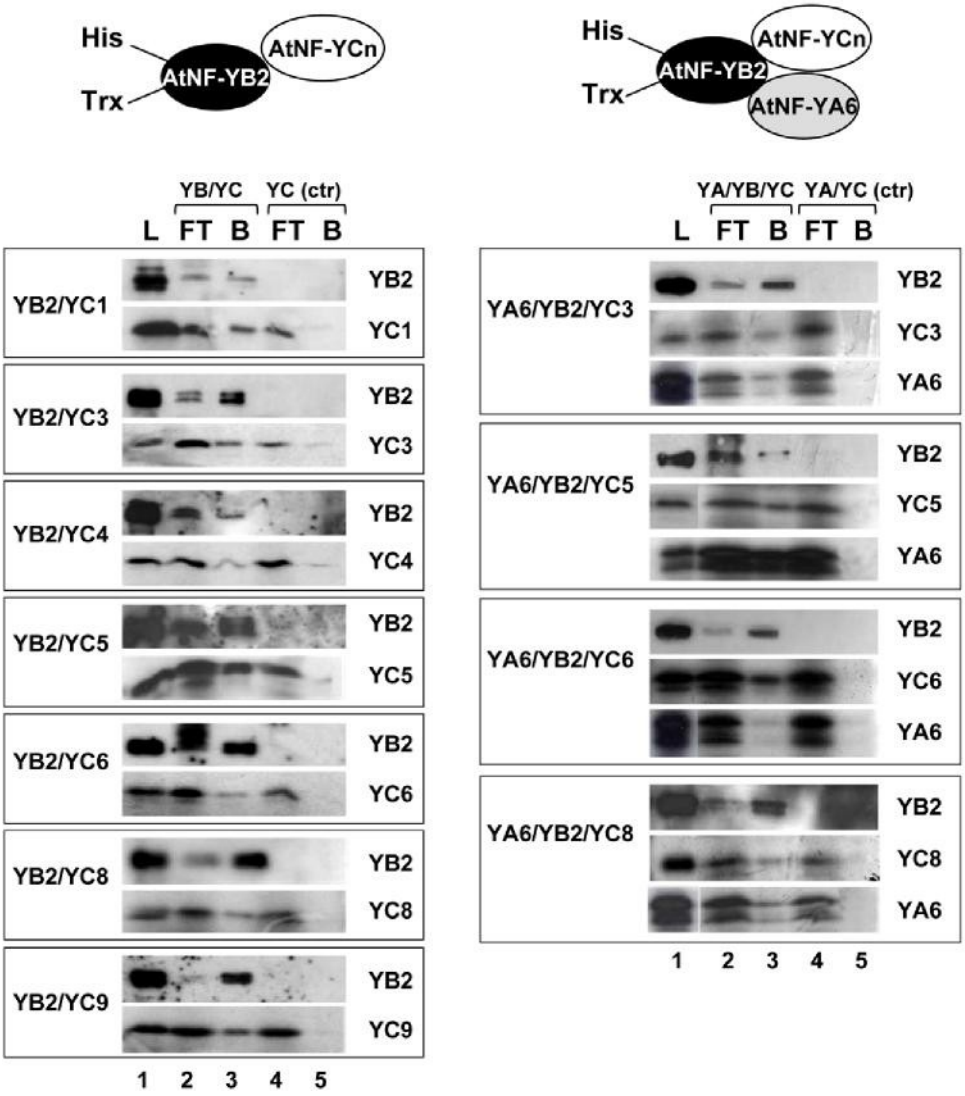


Figure 3. AtNF-Y Subunits interactions *in vitro*. A. The indicated labelled, TnT produced NF-YCs were assayed in affinity assays with recombinant AtNF-YB2 containing an His-tag. Load (L), flow-.through (FT) and bound (B) fractions of NTA Nickel columns, with (Lanes 2 and 3) and without (Lanes 4 and 5) His-AtNF-YB2 were run on SDS-PAGE gels and labelled proteins were revealed by autoradiography. B. Same as A, except that labelled, TnT produced AtNF-YA6 was added to the load fraction.
doi:10.1371/journal.pone.0042902.g003

primarily to hydrophobic contacts. Another region of importance is the a1 helix, stabilized by hydrophobic interactions that are stacked against a conserved tryptophan at position 85 of a2 helix of NF-YC. Essentially all AtNF-YBs and AtNF-YCs have hydrophobic residues at appropriate positions, thus the widespread heterodimerizations we observed came to modest surprise. With respect to the Y2H experiments reported by Hackenberg et al. [34], as well as previous data [13,31], we note the following differences.

(i) Some clear preferences for heterodimer formation between specific AtNF-YBs and AtNF-YCs were observed in the present and in the Hackerberg studies: only AtNF-YC6 and AtNF-YC7 show a marked preference for selected AtNF-YB subunits, in our study; all AtNF-YCs, except AtNF-YC3 and AtNF-YC9, have clear preferences in the Hackenberg data. On the AtNF-YBs side, AtNF-YB2, AtNF-YB3, AtNF-YB4, AtNF-YB7 and AtNF-YB10 show reduced affinity for one, or sometimes two AtNF-YCs: the same applies in the reported study. Even with our knowledge of the structure, it is quite difficult to rationalize these preferences, which seems quantitative more than qualitative.

(ii) In the Hackenberg study, AtNF-YB11, B12, B13 and AtNF-YC10, 11, 13 are very selective, with a tendence to heterodimerize among them: the likeliest explanation is that these are not true NF-Y subunits, but rather resemble to other H2A/H2B-likes [KT, CT, RM, in preparation]: it should be remembered, in fact, that the TBP/TATA-binding NC2a/NC2b and the chromatin remodeling and DNA-Polymerase e subunits Chrac15/Dpb3/Dpb4 share extended conservation and have highly similar structures

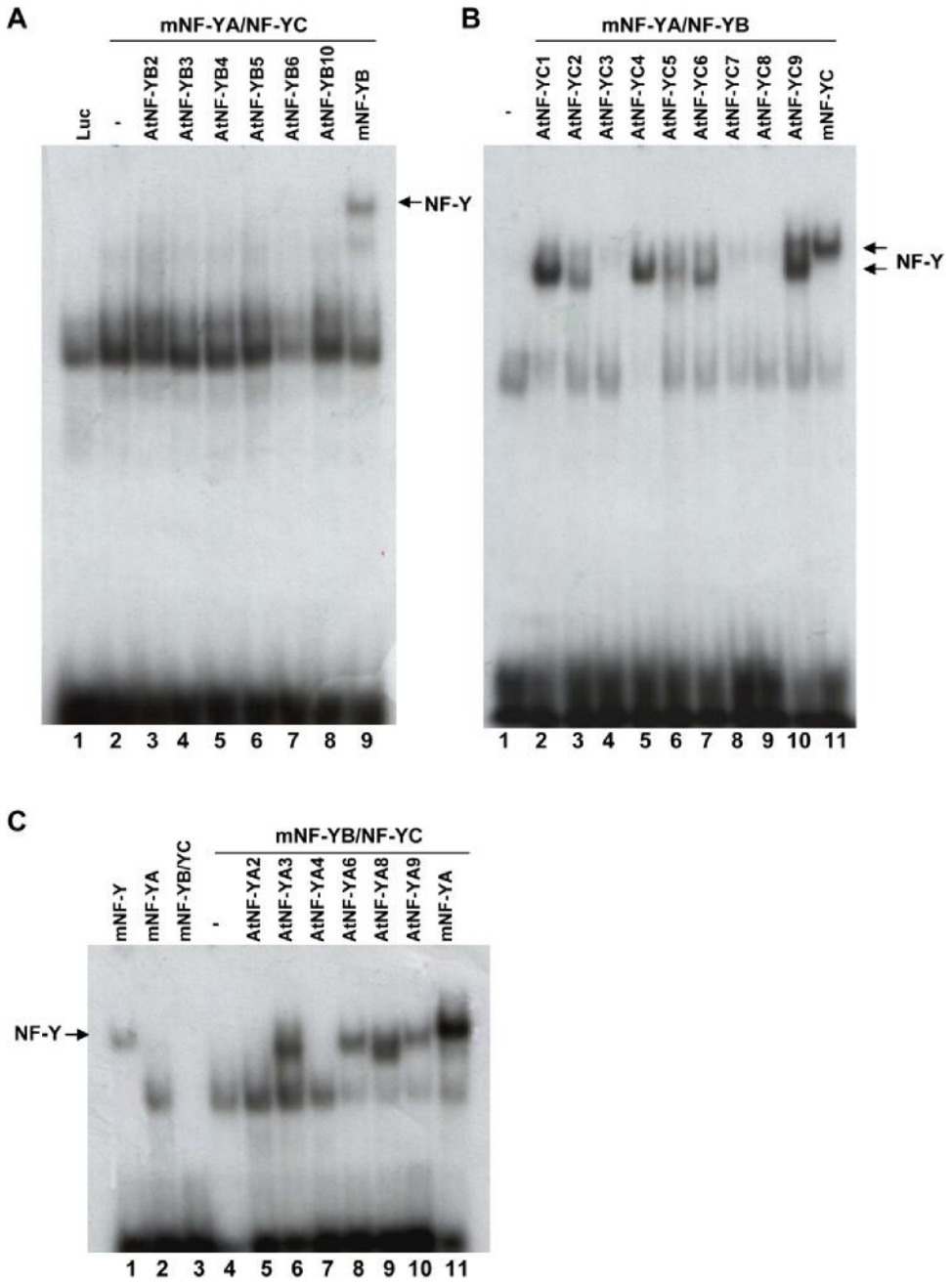


Figure 4. EMSAs of AtNF-Y subunits with mouse NF-Y. A.Electrophoretic Mobility Shift Assay of the indicated AtNF-YB with recombinant mouse NF-YA and NF-YC using a labeled CCAAT-containing oligonucleotide. B. Same as A, except that At NF-YCs were used with recombinant mouse NF-YA and NF-YB. C. Same as A, except that AtNF-YA were used with recombinant mouse NF-YB and NF-YC. The migration of the mouse NF-Y complex is indicated.
doi:10.1371/journal.pone.0042902.g004

[2,37,39]. We have previously reported that mammalian NF-YB/NF-YC do not cross heterodimerize with NC2 subunits [38], and there are structural reasons for this [3].

(iii) Kumimoto et al. have shown that AtNF-YB2 and B3 interact strongly with AtNF-YC3, C4 and C9 in Y2H and in genetic terms [31], which is in line with our data, but not with the Hackerberg study, in which they lack AtNF-YC4 binding. In rice, the homologue of AtNF-YB2 (OsHAP3A) interacts in Y2H with all OsHAP5s, including the homologue of AtNF-YC4, except for homologues of AtNF-YC2 and C3 [13].

(iv) AtNF-YC5, AtNF-YC7 and AtNF-YC8, which belongs to a common clade and are the most tissue-restricted members of the AtNF-YC [8], were negative for AtNF-YB heterodimerization in the Hackenberg et al., study, the former in both AD and DB combinations, the latters in one. In our data, these AtNF-YCs were generally positive for all AtNF-YBs, except AtNF-YC7, negative with AtNF-YB3 and AtNF-YB4.

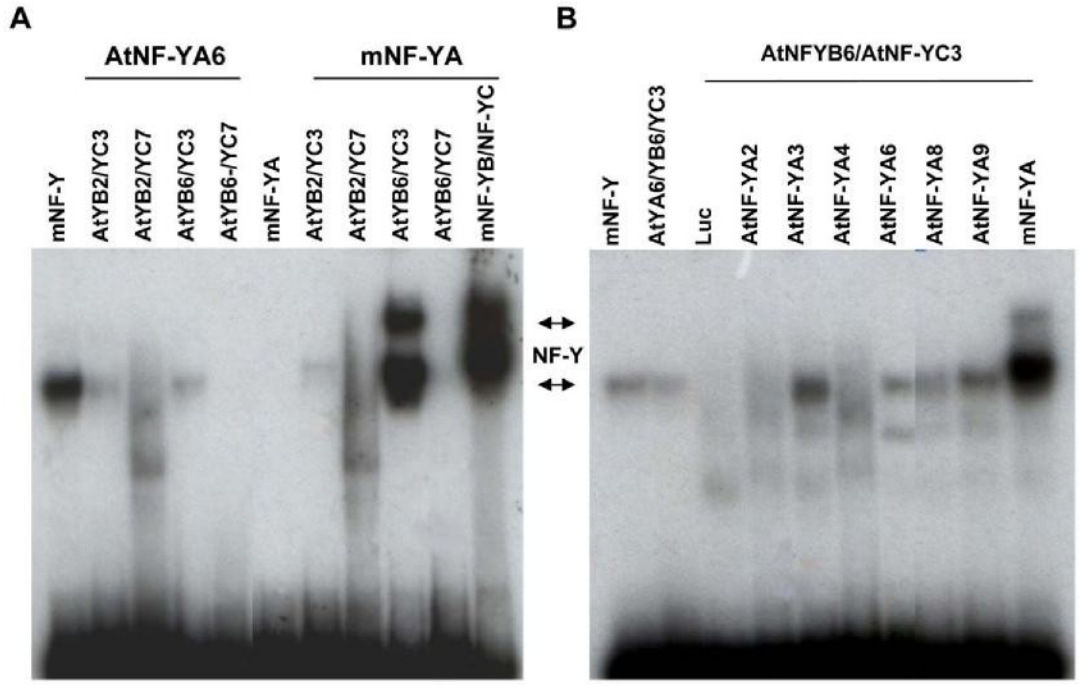


Figure 5. EMSAs of At NF-Y subunits. Electrophoretic Mobility Shift Assay of the indicated AtNF-Y subunits with a labelled CCAAT-containing oligonucleotide.
doi:10.1371/journal.pone.0042902.g005

(v) LEC1/AtNF-YB9 showed a dual behaviour in our hands: no interactions with AtNF-YCs in the Y2H assays, yet efficient heterodimerization with AtNF-YC3 by co-expression of the two proteins in *E. coli*. In the Hackenberg study, LEC1/AtNF-YB9 was positive with most AtNF-YCs; moreover, the carrot homologue of LEC1 was able to bind DNA *in vitro* with two NF-YC homologues [33]. Thus, in this specific case, our Y2H was clearly misleading.

All in all, different Y2H data show some discrepancies, most likely due to technicalities in the expression vectors, yeast productions or activation assays. We also have to bear in mind that yeast possesses endogenous HAPs (as well as NC2 and Dpb3/ 4), indeed shown to interact with some of the plant members [30], thus possibly influencing the results of such assays. Our experiments with LEC1/AtNF-YB9 illustrates the dangers of relying only on this assay in the case of negative results.

We are intrigued by the unusual capacity of LEC1/AtNF-YB9 to form homodimers and remain soluble in bacteria: to the best of our knowledge, this is unique among HFDs, which are normally found as inactive, precipitated proteins in inclusion bodies, when not overexpressed with the appropriate partner [36]. This brings up the question of whether LEC1 homodimers are formed in plants. We found that they do not bind DNA, most likely because of lack of interactions with NF-YA, which absolutely requires NF-YC. It is possible that there is regulation of homo- to heterodimer formation: for example, post-translational modifications (PTMs), not performed in bacteria, could be required to render the HFD prone to heterodimerization: these are histone-like proteins, and histones are crucially controlled by a wealth of PTMs, and we have recently obtained evidence that mouse NF-YB is modified *a-la* H2B (RM, in preparation).

LEC1/AtNF-YB9 and L1L/AtNF-YB6 are capable to efficiently heterodimerize with AtNF-YC3, trimerize and bind to DNA, and the latter also with all AtNF-YC partners in Y2H assays. These data fits with genetic experiments, which established that L1L complements the LEC1 mutants, and in domain swapping experiments with other AtNF-YBs, the B domain -corresponding to the HFD [40]- is required for complementation. In addition to the AtNF-YC and AtNF-YA partners, LEC1 and L1L could exert their roles through interacting proteins, such as MADS box OsMADS6 and OsMADS18 [41], Pirin1, an iron-containing member of the cupin superfamily involved in a pathway leading to an ABA-mediated delay in seed germination [24]. Additional proteins interacting with AtNF-Ys are bZIP67, interacting with AtNF-YC2 in the regulation of *CRUCIFERIN C* [*CRC*] and *SUCROSE SYNTHASE2* [*SUS2*] in *Arabidopsis* proto-plasts [21], and, most importantly, CONSTANS and CON-STANS-like proteins in *Arabidopsis* and tomato [42,43] involved in determining the proper flowering timing with specific members of AtNF-YBs and AtNF-YCs [31,44].

DNA-binding

The formation of NF-Y heterotrimers was tested with selected AtNF-YB/AtNF-YC HFD dimers. While the HFD dimer contributes substantially to DNA-binding, mostly through a1 helices, L1 and L2 loops, the subunit that confers the sequence-specificity is NF-YA. On the HFD side, the heterotrimerization surface relies in selected residues in the a2 helix of NF-YB and in the aC helix of NF-YC. The E90 and E98 of mouse NF-YB, important for NF-YA binding [45], are conserved in all AtNF-YBs [8–10,34]. The aC helix of AtNF-YC, on the other hand, shows differences in at least three members: AtNF-YC5 possesses an R at position 109 of mouse NF-YC, instead of an hydrophobic residue; AtNF-YC8 has two Aspartates at position 111 and 112, instead of hydrophobics, together with Isoleucine at position 113, instead of the helix capping Proline [3]; finally, AtNF-YC7 has a four aminoacids addition in the a3 helix, which extends it for an additional turn, hence displacing the LC domain and aC helix

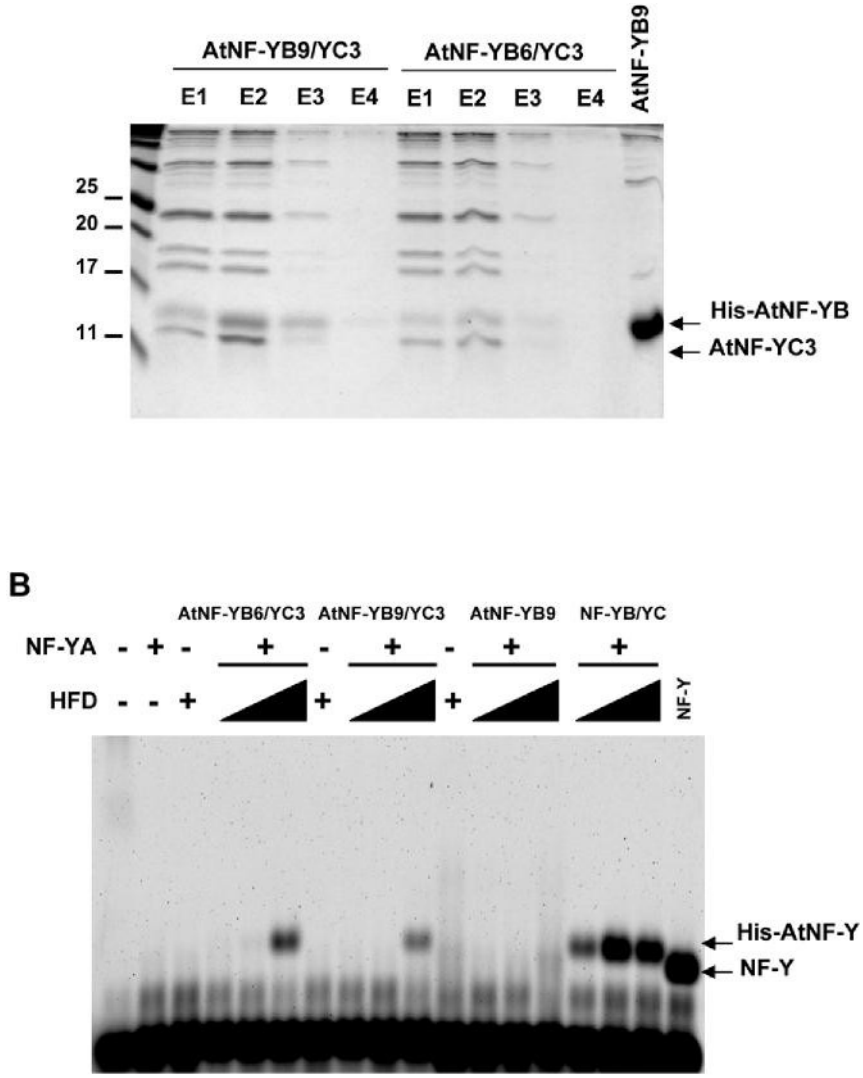


Figure 6. *E. coli* co-expression of LEC1/AtNF-YB9 with AtNF-YC3 allows functional heterodimerization, heterotrimerization and CCAAT-binding. A. Purification of soluble LEC1/AtNF-YB9 or L1L/AtNF-YB6 HFD heterodimers by co-expression with AtNF-YC3. Nickel-affinity purification elution profiles obtained from soluble fractions of 6His-LEC1/AtNF-YB9 or 6His-L1L/AtNF-YB6 with AtNF-YC3. Equal volumes of indicated elution fractions (E) in 100 mM Imidazole of LEC1/AtNF-YB9 or L1L/AtNF-YB6 with AtNF-YC3 were analysed by SDS-PAGE and Coomassie staining. E2, were dialysed and used in Agarose gel non-radioactive EMSAs shown in (B). B. Fluorescence agarose gel EMSAs of trimer reconstitution with mouse NF-YA. 5^{9}-Cy5 labeled CCAAT oligonucleotide probe was incubated with increasing amounts of the indicated 6His-tagged HFD dimers isolated by Ni-affinity purification, or mouse 6His-NF-YB/NF-YC as positive control, in the presence, or absence, of purified mouse NF-YA. Purified (untagged) mouse NF-Y trimer was used as a reference for NF-Y complex migration.
doi:10.1371/journal.pone.0042902.g006

from their natural positions. Not coincidentally, these three members were not proficient in DNA-binding in our assays. Although the interaction with AtNF-YA6 appears to be visible with recombinant proteins, it remains to be seen whether other residues directly contacting DNA in L1 and L2 loops (N86 in AtNF-YC5 and G113 in AtNF-YC7, instead of a conserved Lysine) might explain the decrease in DNA affinity of this group of AtNF-YCs.

It was initially troubling to obtain negative results in EMSAs with the TnT-produced AtNF-YBs, but this was most likely due to technical problems of the translation extract, possibly inhibiting trimerization, or production of inactive AtNF-YBs in the absence of coexpression of AtNF-YCs: in fact, recombinant AtNF-YBs produced from *E. coli*, including the divergent LEC1/AtNF-YB9 and L1L/AtNF-YB6, were positive in DNA-binding. Interestingly, mutation of an Aspartate at position 55 of LEC1 is sufficient to abrogate LEC1 function *in vivo* [17]. D55 is located at the beginning of the a2 Helix, in a region that lies on the surface of the dimer: most other *Arabidopsis* and mammalian NF-YBs have a Lysine, conserved in H2B, and predicted to be involved in protein-DNA interactions [3]. L1L/AtNF-YB6 also has an Aspartate at this position, which might be considered as a "signature"for this subfamily: the change might decrease affinity for DNA, but an important result in our study is that it certainly does not abolish it: in essence, no AtNF-YB is "deviant"enough to have lost the DNA-binding capacity.

On the NF-YA side, the evolutionarily conserved domain is responsible for trimerization and CCAAT-binding. Protein-

protein interaction assays and EMSAs indicate that the majority of AtNF-YAs are able to interact with AtNF-YB2/AtNF-YC3 and L1L/AtNF-YB6/AtNF-YC3. Indeed, they are quite proficient in association to the mouse NF-YB/NF-YC dimer. In particular, the AtNF-YA6 shows robust CCAAT binding, which strongly suggests that AtNF-YA5, not tested here, behaves similarly: the two belong to a common clade, and the DNA-binding subdomain is absolutely conserved. AtNF-YA5 is so far the only AtNF-YA for which genetic experiments were reported: mutation causes drought stress and overexpression drought resistance [25]; our data suggest that the mechanisms are related to prototypical CCAAT-binding.

Only AtNF-YA2 and AtNF-YA4 were negative, suggesting that they are either incapable to trimerize or bind DNA. Note that AtNF-YA7 and AtNF-YA10 not tested here might behave similarly, since the residues required for subunits interactions and DNA-binding are identical to AtNF-YA4 and AtNF-YA2, respectively. Several papers described two separate 20 aminoacid stretches as required for subunits interactions and DNA-binding [46–48]. Detailed mutagenesis of the mouse and yeast subunits pinpointed several aminoacids necessary for the two functions. In the subunits interaction domain, no dramatic changes are observed, and indeed important residues are conserved in AtNF-YA2 and AtNF-YA4, with the notable exception of R273 (mouse), which is G147 in AtNF-YA2 and G137 in AtNF-YA4: potentially, this could affect trimerization, since an R to G mutation in yeast HAP2 does decrease the efficiency of HFD association significantly [46]. We note, however, that in none of the other AtNF-YAs, nor in most other plant NF-YA genes, there is an Arginine at this position: in proficient members of the family tested here, an Alanine is present. Most importantly, AtNF-YA2 and AtNF-YA4 were previously tested for heterotrimerization, and indeed showed efficient association with HFDs [34]: in all likelyhood, therefore, they have decreased DNA-binding affinity, despite an overall conservation of key DNA-binding residues. Can we take these data as an indication that some of the AtNF-YAs have lost the capacity to bind DNA? If it is indeed so, what might be their function? The most obvious answer is that if they do bind NF-YB/NF-YC dimmer, they might act as Dominant Negative in terms of CCAAT binding: indeed, introduction of mutations in the DNA-binding subdomain of mouse NF-YA transforms it into a DN protein ([1] and References therein).

The alternative, more appealing possibility to explain these results is that trimers with these subunits have subtly changed sequence-specificity. Residues that are variant in these genes, such as C176 in AtNF-YA4 -a Serine in the other AtNF-YAs- and H178 in AtNF-YA2 -a Glutamate in the other AtNF-YAs- or the longer linker of AtNF-YA2 might account for this. Bioinformatic analysis performed in our lab on human genome-wide data has established that the NF-Y consensus, even in mammals, can, moderately, deviate from a perfect pentanucleotide CCAAT, provided that additional flanking nucleotides are present [2]: indeed, some 30% of NF-Y bound *in vivo* in human cells show a deviation of one nucleotide of the core CCAAT sequence. It seems reasonable therefore to postulate that subclasses of AtNF-YAs might bind variant versions of the CCAAT box: this hypothesis can be tested more thoroughly by the biochemical assays we set up with recombinant proteins, as we have started to do here. Even so, rationalization and full understanding of the molecular details of the enormous combinatorial possibilities of plant NF-Ys will have to ultimately await crystallization of NF-Y/CCAAT complexes.

Materials and Methods

Yeast strains and plasmid construction

The cDNAs corresponding to each AtNF-Y subunit used in the Yeast-Two Hybrid assay, were amplified from *Arabidopsis* cDNA libraries using gene specific primers containing the attB1 and attB2 sequences for homologous recombination and subsequently cloned into pDONOR201 vector (Life Technology). *AtNF-YB* and *AtNF-C* coding sequences in pDONOR201 were subsequently cloned in the GAL4 Gateway vector system: pDEST32 for DNA binding domain fusions (pDBD) and pDEST22 for activation domain fusions (pAD). The pDEST32 and pDEST22 vectors were transformed into *Saccharomyces cerevisiae* strain PJ69-4A (trp1-901 leu2-3, 112 ura3-52 his3-200 gal4Δ gal80Δ LYS2::GAL1-HIS3 GAL2-ADE2 met2::GAL7-lacZ) [49]. Yeast Two-Hybrid assay was performed as described below.

Yeast Two-Hybrid (Y2H) analysis

Haploid Yeast a and A were transformed respectively with pBD and pAD vector constructs using the lithium acetate method [47] and selected on Yeast Synthetic Dropout [YSD] medium lacking Leu and Trp, respectively. Yeast carrying pBD vectors were tested for autoactivation on selective medium with 5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside (X-Gal), on medium lacking histi-dine and supplemented with different concentrations of 3-aminotriazole (0, 3, 5, 10, 25 and 50 mM) and on medium lacking adenine. Mating type a and A were mated and diploids selected on YSD medium lacking Leu and Trp.

Two-hybrid interactions were assayed on selective YSD medium lacking Leu, Trp, and Ade or His supplemented with 50 mM 3-aminotriazole. Selection was performed at 28°C for 4 days.

Liquid Two-Hybrid Assay

Semi-quantitative assay for comparing the strength of AtNF-YB and AtNF-YC subunits interactions was performed by liquid LacZ assay. For the liquid assay, we used the AtNF-YB (DBD) and AtNF-YC (AD) configuration.

Yeast was inoculated in selective medium and grown for 8–9 h, then centrifuged at 4500 rpm for 5 min. Pelletted cells were resuspended in 5 ml of selective medium and grown O/N at 28°C. Cells were centrifuged at 4500 rpm for 5 min and the pellet was resuspended in 0.5 ml of cold water, centrifuged again for 30 sec at 14000 rpm. Pellet was resuspended in 250 ml of pre-cooled Breaking Buffer (100 mM Tris-HCl pH 8.0, 10% Glycerol, 1 mM Dithiothreitol, protease inhibitors) and frozen in liquid nitrogen. Unfrozen samples were subjected to 10 cycles of vortex/ice with glass beads, centrifuged at 14000 rpm for 10 min and supernatant has been recovered. Then 20 ml of protein extract were transferred to a 1.5 ml centrifuge tube and added with 800 ml of Z-Buffer 1Δ (60 mM NaH_2PO_4, 40 mM Na_2HPO_4 anhydrous, 10 mM KCl, 1 mM Mg_2SO_4, 50 mM (3-Mercaptoethanol) and 200 ml ortho-Nitrophenyl-(3-galactoside (ONPG) 4 mg/ml. The tube was incubated at 37°C until the solution became yellow, for a maximum of 45 min and the reaction was stopped adding 400 ml of 1.5 M Na_2CO_3. The samples were centrifuged for 30 sec at 13000 rpm and the optical density at 420 nm (OD_{420}) was determined. Activity in Miller Units was calculated according to the formula $(OD_{420} *1.4)/(0.0045*C*V*t)$ where C = concentration of protein extract (mg/ml); V =volume of protein extract (ml); t =time (min). Activity of AtNF-YB GAL4-DBD with GAL4-AD fused with no AtNF-YC subunit has been used as control.

Production of recombinant AtNF-YB and generation of ^{35}S-Labeled AtNF-YC

To examine the *in vitro* interaction between AtNF-YB and AtNF-YC subunits, His-tagged AtNF-YBs and ^{35}S labelled AtNF-YC were produced and used for pull-down experiments.

Chemically competent *E. coli* BL21 cells were transformed by thermal shock with 100 ng of pET32A or pET32B, in which AtNF-YB coding sequences were cloned. Transformed cells were inoculated in LB broth (5 ml) with ampicillin (100 ng/ml) at 37°C for 16 h. An aliquot (3 ml) of this culture was inoculated in 200 ml of the same medium and let grow until an $_{OD600}$ of 0.6 was reached. The expression of each protein was induced with IPTG (1 mM) for 3 h.

Cells were harvested by centrifugation at 6000 rpm for 10 min at 4°C and suspended in Sonication Buffer (300 mM KCl, 20 m M Tris-HCl pH 7.8, 0.05% NP40, 1 mM EDTA, 1 mM PMSF, 5 mM (3-Mercaptoethanol) containing a cocktail of Protease inhibitors (12.5 mg/ml leupeptin, 5 mg/ml trypsin inhibitor, 5 mg/ml pepstatin, 10 mg/ml chymostatin). The cells were then thoroughly disrupted with a sonicator (10 cycles, 20 sec each). The samples were centrifuged at 23000 rpm at 4°C for 90 min to separate supernatant (SN) from inclusion bodies (IB). The SN and IB (Resuspended in 100 mM KCl, 20 mM Tris-HCl pH 7.8, 5 mM (3-Mercaptoethanol, 1 mM PMSF, 6 M GnCl) were loaded onto Nichel-Agarose columns (Sigma). After thoroughly washing with Washing Buffer (20 mM Tris-HCl pH 7.8 10% glycerol, 300/1000/100 mM KCl), the proteins bound to the columns were eluted in Elution Buffer (2 mM Tris-HCl pH 7.8, 10% glycerol, 100 mM KCl, 1 mM PMSF, 5 mM (3-Mercaptoethanol, 300 mM Imidazol). Finally, eluted fractions from SN and IB were subjected to dialysis to remove Imidazol.

AtNF-YC subunits, cloned in pCR4TOPO (Invitrogen), were synthesized and ^{35}S-labeled by coupled transcription and translation in 25 µl of nuclease-treated rabbit reticulocyte lysate (TnT, Promega).

His pull-down assay

His-tagged AtNF-YB recombinant proteins (500 ng) and 10 µl of AtNF-YCs produced by TnT were incubated together at 37°C for 30 min in 100 µl of NDB100 (20% glycerol, 100 mM KCl, 20 mM Tris-HCl pH 7.8, 0.5 mM EDTA, 5 mM (3-Mercaptoethanol). After incubation, recombinant proteins were loaded onto a Nichel-Agarose column (Sigma), incubated for 3 h at 4°C, and then centrifuged at 4000 rpm for 1 min at 4°C to recover the "flow through"(FT). After washing 3 times, they were eluted ("bound", B) with 30 µl of Elution Buffer (NDB100 containing 5 mM (3-Mercaptoethanol, 0.25 M imidazole, PIC 1%). As negative controls, aliquots (10 µl) of the same AtNF-YC subunits producted by TnT were incubated with the Nichel-Agarose column. We did not observe any aspecifically bound AtNF-YC subunits in the negative controls performed in the absence of His-tagged AtNF-YBs. One third of FT and B samples were subjected to SDS-PAGE, transferred to a nitrocellulose membrane (150 mA/gel, 1.5 h), and analyzed by Western blotting using anti-His antibodies; the remaining two thirds of each sample were analysed by autoradiography to detect AtNF-YC subunits.

HFD heterodimer Protein expression and purification

The 6His-AtNF-YB/AtNF-YC soluble HFD dimers were purified exploiting the T7-driven co-expression system described in [3,50]. AtNF-YC3 (AA 55–148) (corresponding to the HFD region of mouse NF-YC AA 27–120) was subcloned in the pmncYC vector; LEC1/AtNF-YB9 (AA 56–148) or L1L/AtNF-YB6 (AA 26–118) subunits (corresponding to mouse NF-YB HDF AA 49–141) were subcloned in pET15b, resulting in 6His-N-terminal fusions. 6His-LEC1/AtNF-YB9, or 6His-L1L/AtNF-YB6, was expressed in *E. coli* BL21(DE3) together with, or not, AtNF-YC3, and purified by Ni-chelate affinity chromatography (HisSelect, SIGMA-Aldrich), as described in [3], in buffer A (10 mM Tris pH 8.0, 400 mM NaCl, 2 mM $MgCl_2$, 5 mM imidazole), and eluted by subsequent additions of 1 bed volume of buffer B containing 100 mM Imidazole. Indicated 6His-HFD protein purification eluates were dialysed against buffer B (10 mM Tris pH 8.0, 400 mM NaCl, 2 mM DTT) containing 10% glycerol, and used in Fluorescensce Agarose gel EMSAs. The soluble NF-Y heterotrimeric subunit complex and 6His-NF-YA were produced as described in [50], and purified by Ni-chelate affinity chromatography (HisSelect, SIGMA-Aldrich) in buffer A, followed by thrombin cleavage of the NF-YA C-terminal His-tag, and gel filtration (GF) chromatography (HiLoad Superdex75, Amersham Pharmacia) in buffer B. GF fractions corresponding to the NF-Y heterotrimer, or the NF-YA isolated subunit, were collected, and used in Fluorescensce Agarose gel EMSAs, after addition of 10% glycerol for storage.

Electrophoretic Mobility Shift Assays

For electrophoretic mobility shift assays ^{32}P labelled fragments ~10000 CPMs- are incubated in NF-Y Buffer (20 mM Hepes pH 7.9, 50 mM NaCl, 5% Glycerol, 5 mM $MgCl_2$, 5 mM (3-ME) with the recombinant proteins (1–5 ng), in a total volume of 10 µl; after incubation for 15' at 20°C, we added 2 µl of 1% NF-Y buffer containing Bromophenol Blue and samples loaded on a 4.5% Polyacrylamide in 0.5% TBE. Gels were dried and exposed. For Fluorescence Agarose Gel EMSAs of Figure 6, heterotrimer formation and CCAAT-box DNA-binding of the 6His-AtNF-YB/AtNF-YC soluble dimers was assessed with Cy5-labeled oligos, by addition of GF purified mouse NF-YA (AA 233–303). Equal protein amounts of Ni-purified 6His-AtNF-YB/NF-YC HFD dimers (3, 6, or 9 ng/ul) were mixed in 15 µl reactions with the 5'-labeled 31 bp oligo probe derived from human HSP70 promoter CCAAT box sequence (Cy5-CTTCTGAGCCAAT-CACCGAGCTCGATGAGGC) in DNA binding mix (20 nM ds oligo, 20 mM Tris pH 7.5, 50 mM NaCl, 0.5 mM EDTA, 5 mM $MgCl_2$, 2.5 mM DTT, 0.1 mg/ml BSA, 5% glycerol), in the presence of 40 nM NF-YA, where indicated. Ni-purified mouse 6His-NF-YB/NF-YC (1, 3, 6 ng), or GF purified NF-Y trimer (60 nM) were used a positive controls. After 30 min incubation at 23°C, binding reactions were loaded on a 2.5% agarose gel and separated by electrophoresis in 0.5% TBE. Fluorescence gel images were obtained with a Typhoon 8610 Variable Mode Imager (Molecular Dynamics).

Supporting Information

Figure S1 TnT and recombinant proteins production. A. AtNF-YA, AtNF-YB and AtNF-YC subunits were synthesized and ^{35}S-labeled by coupled transcription and translation in nuclease-treated rabbit reticulocyte lysate (TnT, Promega). B. His-tagged AtNF-YA6, AtNF-YB2 and AtNF-YB6, AtNF-YC3 and AtNF-YC7 have been produced in *E. Coli* and purified by Nichel-Agarose columns (Sigma). Load (L), flow-through (FT), wash (W) and eluted (E) fractions of NTA Nickel columns are shown.
(TIF)

Author Contributions

Conceived and designed the experiments: RM CT. Performed the experiments: VC BT GG ML NG KP. Analyzed the data: RM CT NG. Contributed reagents/materials/analysis tools: ML NG GG. Wrote the paper: RM CT.

References

1. Dolfini D, Zambelli F, Pavesi G, Mantovani R (2009) A perspective of promoter architecture from the CCAAT box. Cell Cycle 9: 4127–4237.
2. Dolfini D, Gatta R, Mantovani R (2012) NF-Y and the transcriptional activation of CCAAT promoters. Crit Rev Biochem Mol Biol 47: 29–49.
3. Romier C, Cocchiarella F, Mantovani R, Moras D (2003) The crystal structure of the NF-YB/NF-YC heterodimer gives insight into transcription regulation and DNA binding and bending by transcription factor NF-Y. J Biol Chem 278: 1336–1345.
4. Forsburg SL, Guarente L (1989) Identification and characterization of HAP4: a third component of the CCAAT-bound HAP2/HAP3 heteromer. Genes Dev 3: 1166–1178.
5. McNabb DS, Tseng KA, Guarente L (1997) The *Saccharomyces cerevisiae* Hap5p homolog from fission yeast reveals two conserved domains that are essential for assembly of heterotetrameric CCAAT-binding factor. Mol Cell Biol 17: 7008–7018.
6. Coustry F, Maity SN, Sinha S, de Crombrugghe B (1996) The transcriptional activity of the CCAAT-binding factor CBF is mediated by two distinct activation domains, one in the CBF-B subunit and the other in the CBF-C subunit. J Biol Chem 271: 14485–14491.
7. de Silvio A, Imbriano C, Mantovani R (1999) Dissection of the NF-Y transcriptional activation potential. Nucleic Acids Res 27: 2578–2584.
8. Gusmaroli G, Tonelli C, Mantovani R (2001) Regulation of the CCAAT-binding NF-Y subunits in *Arabidopsis thaliana*. Gene 264: 173–185.
9. Gusmaroli G, Tonelli C, Mantovani R (2002) Regulation of novel members of the *Arabidopsis thaliana* CCAAT-binding nuclear factor Y subunits. Gene 283: 41– 48.
10. Siefers N, Dang KK, Kumimoto RW, Bynum WE IV, Tayrose G, et al. (2009) Tissue specific expression patterns of *Arabidopsis thaliana* NF-Y transcription factors suggest potential for extensive combinatorial complexity. Plant Physiol 149: 625–641.
11. Yang J, Xie Z, Glover BJ (2005) Asymmetric evolution of duplicate genes encoding the CCAAT-binding factor NF-Y in plant genomes. New Phytol 165: 623–631.
12. Stephenson TJ, McIntyre CL, Collet C, Xue GP (2007) Genome-wide identification and expression analysis of the NF-Y family of transcription factors in *Triticum aestivum*. Plant Mol Biol 65: 77–92.
13. Thirumurugan T, Ito Y, Kubo T, Serizawa A, Kurata N (2008) Identification, characterization and interaction of HAP family genes in rice. Mol Genet Genomics 279: 279–289.
14. Xie Z, Li X, Glover BJ, Bai S, Rao GY, et al. (2008) Duplication and functional diversification of HAP3 genes leading to the origin of the seed-developmental regulatory gene, LEAFY COTYLEDON1 (LEC1), in nonseed plant genomes. Mol Biol Evol 25: 1581–1592.
15. Cao S, Kumimoto RW, Siriwardana CL, Risinger JR, Holt BF III (2011) Identification and characterization of NF-Y transcription factor families in the monocot model plant *Brachypodium distachyon*. PLoS One 6:e21805.
16. Lotan T, Ohto M, Yee MK, West MA, Lo R, et al. (1998) *Arabidopsis* LEAFY COTILEDON1 is sufficient to induce embryo development in vegetative cells. Cell 93: 1195–1205.
17. Lee H, Fischer RL, Goldberg RB, Harada JJ (2003) *Arabidopsis* LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor. Proc Natl Acad Sci USA 100: 2152–2156.
18. Gaj MD, Zhang S, Harada JJ, Lemaux PG (2005) Leafy cotyledon genes are essential for induction of somatic embryogenesis of *Arabidopsis*. Planta 222: 977– 988.
19. Braybrook SA, Harada JJ (2008) LECs go crazy in *embryo* development. Trends Plant Sci 13: 624–30.
20. Kwong RW, Bui AQ, Lee H, Kwong LW, Fischer RL, et al. (2003) LEAFY COTYLEDON1-LIKE defines a class of regulators essential for *embryo* development. Plant Cell 15: 5–18.
21. Yamamoto A, Kagaya Y, Toyoshima R, Kagaya M, Takeda S, et al. (2009) *Arabidopsis* NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors. Plant J 58:843–56.
22. Yazawa K, Takahata K, Kamada H (2004) Isolation of the gene encoding Carrot leafy cotyledon1 and expression analysis during somatic and zygotic embryogenesis. Plant Physiol Biochem 42: 215–223.
23. Alemanno L, Devic M, Niemenak N, Sanier C, Guilleminot J, et al. (2008) Characterization of leafy cotyledon1-like during embryogenesis in *Theobroma cacao* L. Planta 227: 853–866.
24. Warpeha KM, Upadhyay S, Yeh J, Adamiak J, Hawkins SI, et al. (2007) The GCR1, GPA1, PRN1, NF-Y signal chain mediates both blue light and abscisic acid responses in *Arabidopsis*. Plant Physiol 143: 1590–1600.
25. Li WX, Oono Y, Zhu J, He XJ, Wu JM, et al. (2008) The *Arabidopsis* NFYA5 transcription factor is regulated transcriptionally and posttranscriptionally to promote drought resistance. Plant Cell 20: 2238–2251.
26. Nelson DE, Repetti PP, Adams TR, Creelman RA, Wu J, et al. (2007) Plant nuclear factor Y (NF-Y) B subunits confer drought tolerance and lead to improved corn yields on water-limited acres. Proc Natl Acad Sci USA 104: 16450–16455.
27. Ito Y, Thirumurugan T, Serizawa A, Hiratsu K, Ohme-Takagi M, et al. (2011) Aberrant vegetative and reproductive development by overexpression and lethality by silencing of OsHAP3E in rice. Plant Sci 181: 105–110.
28. Cai X, Ballif J, Endo S, Davis E, Liang M, et al. (2007) A putative CCAAT-binding transcription factor is a regulator of flowering timing in *Arabidopsis*. Plant Physiol 145: 98–105.
29. Chen NZ, Zhang XQ, Wei PC, Chen QJ, Ren F, et al. (2007) AtHAP3b plays a crucial role in the regulation of flowering time in *Arabidopsis* during osmotic stress. J Biochem Mol Biol 40: 1083–1089.
30. Kumimoto RW, Adam L, Hymus GJ, Repetti PP, Reuber TL, et al. (2008) The Nuclear Factor Y subunits NF-YB2 and NF-YB3 play additive roles in the promotion of flowering by inductive long-day photoperiods in *Arabidopsis*. Planta 228: 709–723.
31. Kumimoto RW, Zhang Y, Siefers N, Holt BF III (2010) NF-YC3, NF-YC4 and NF-YC9 are required for CONSTANS-mediated, photoperiod-dependent flowering in *Arabidopsis thaliana*. Plant J 63: 379–391.
32. Combier JP, Frugier F, de Billy F, Boualem A, El-Yahyaoui F, et al. (2006) MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in *Medicago truncatula*. Genes Dev 20: 3084–3088.
33. Yazawa K, Kamada H (2007) Identification and characterization of carrot HAP factors that form a complex with the *embryo*-specific transcription factor C-LEC1. J Exp Bot 13: 3819–3828.
34. Hackenberg D, Wu Y, Voigt A, Adams R, Schramm P, et al. (2012) Studies on Differential Nuclear Translocation Mechanism and Assembly of the Three Subunits of the *Arabidopsis thaliana* Transcription Factor NF-Y. Mol Plant In press.
35. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of nucleosome core particle at 2.8A resolution. Nature 389: 251–260.
36. Li XY, Mantovani R, Hooft van Huijsduijnen R, Andre I, Benoist C, et al. (1992) Evolutionary variation of the CCAAT-binding transcription factor NF-Y. Nucleic Acids Res 20: 1087–1091.
37. Hartlepp KF, Ferna´ndez-Tornero C, Eberharter A, Gru¨ne T, Mu¨ller CW, et al. (2005) The histone fold subunits of Drosophila CHRAC facilitate nucleosome sliding through dynamic DNA interactions. Mol Cell Biol 25: 9886–9896.
38. Zemzoumi K, Frontini M, Bellorini M, Mantovani R (1999) NF-Y histone fold a1 helices help impart CCAAT specificity. J Mol Biol 286: 327–337.
39. Kamada K, Shu F, Chen H, Malik S, Stelzer G, et al. (2001) Crystal structure of negative cofactor 2 recognizing the TBP-DNA transcription complex. Cell 106: 71–81.
40. Masiero S, Imbriano C, Ravasio F, Favaro R, Pelucchi N, et al. (2002) Ternary complex formation between MADS-box transcription factors and the histone fold protein NF-YB. J Biol Chem 277: 26429–26435.
41. Wenkel S, Turck F, Singer K, Gissot L, Le Gourrierec J, et al. (2006) CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of *Arabidopsis*. Plant Cell 18: 2971–2984.
42. Ben-Naim O, Eshed R, Parnis A, Teper-Bamnolker P, Shalit A, et al. (2006) The CCAAT binding factor can mediate interactions between CONSTANS-like proteins and DNA. Plant J 46: 462–476.
43. Li C, Distelfeld A, Comis A, Dubcovsky J (2011) Wheat flowering repressor VRN2 and promoter CO2 compete for interactions with NUCLEAR FACTOR-Y complexes. Plant J 67: 763–773.
44. Sinha S, Kim IS, Sohn KY, deCrombrugghe B, Maity SN (1996) Three classes of mutations in the A subunit of the CCAAT-binding factor CBF delineate functional domains involved in the three-step assembly of the CBF-DNA complex. Mol Cell Biol 16: 328–337.
45. Xing Y, Fikes JD, Guarente L (1993) Mutations in yeast Hap/HAP3 define a hybrid CCAAT box binding domain. EMBO J 12: 4647–4655.
46. Xing Y, Zhang S, Olesen JT, Rich A, Guarente L (1994) Subunit interaction in the CCAAT-binding heteromeric complex is mediated by a very short alpha-helix in HAP2. Proc Natl Acad Sci USA 91: 3009–3013.
47. Mantovani R, Li XY, Pessara U, Hooft van Huijsduijnen R, Benoist C, et al. (1994) Dominant negative analogs of NF-YA. J Biol Chem 269: 20340–20346.
48. Gietz D, St Jean A, Woods RA, Schiestl RH (1992) Improved method for high efficiency transformation of intact yeast cells. Nucleic Acids Res 20: 1425.
49. James P, Halladay J, Craig EA (1996) Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. Genetics 144: 1425–1436.
50. Diebold ML, Fribourg S, Koch M, Metzger T, Romier C (2011) Deciphering correct strategies for multiprotein complex assembly by co-expression: application to complexes as large as the histone octamer. J Struct Biol 175: 178–188.

PART III

MANUSCRIPT IN PREPARATION:

Dissection of the interplay between NF-Y and MAX, MYC and USF1.

Mariangela Lorenzo, Matteo Pigni, Giulio Pavesi, Marco Nardini, Roberto Mantovani and Nerina Gnesutta*.

Dipartimento di Bioscienze, Via Celoria 26, 20143 Milano Italy.

* To whom correspondance should be addressed; nerina.gnesutta@unimi.it.

ABSTRACT

The combinatorial interplay among Transcription Factors –TFs- is at the heart of the regulation of gene expression. The CCAAT-box binding NF-Y is known to cooperate with many TFs and believed to be a "pioneer" TF. Analysis of genome-wide ENCODE data indicates that TFs binding to the E-box have precise alignment with NF-Y, 10/12 bps 5' of CCAAT. Here, we focus on NF-Y interplay with MAX, MYC and USF1. The 12 bps distance in LTR repetitive sequences and HOXB7 promoter is important for synergistic binding of USF1/NF-Y; facilitated binding is observed in the 10 bps HOXB4 configuration. The NF-Y and USF1 DNA-binding domains are sufficient for cooperativity. However, DNA-binding of MAX homodimers and MYC/MAX heterodimers is unaffected by NF-Y. NF-Y inactivation leads to a decrease *in vivo* binding of USF1, but also of MAX to HOXB7; negative, as well as positive effects on recruitment of MYC and USF1 on various promoters are observed. Finally, SAXS experiments allowed us to visualize the 3D structures of NF-Y/MAX and NF-Y/MYC/MAX associated to the respective sites. On the one hand, these data rationalize the function of a conserved distance between E and CCAAT boxes in promoters and LTR repetitive sequences, and a special partnership with USF1; on the other, they indicate that cooperative interactions with NF-Y at the level of DNA binding is but one of the possible mechanisms guiding recruitment of E-box TFs to CCAAT promoters.

Figure 1: Formation of binary and ternary complexes between NF-Y and E-BOX TFs with different probes.

The ratio used in these dose response assays is 1 (probes: 20nM):3 (NF-Y) :2 (E-BOX TFs).

A) Formation of binary complexes composed by HOXB4 and NF-Y (lane 2), HOXB4 and USF1 (lane 3), HOXB4 and MAX (lane 4). When NF-Y and different E-BOX TFs are incubated together with HOXB4 probe (lanes 5 and 10) slower migrating complexes appear. The specificity of ternary and binary complexes is demonstrated using an excess of unlabelled competitors (1 μM) that contain the CCAAT-box (lanes 6 and 11) or E-BOX (lanes 8 and 13); in these cases we can observe that the bands of ternary complexes and the bands of binary complexes competed disappear. We can' t observe the same behaviour when we use mutated CCAAT (lanes 7 and 12) or mutated E-BOX (lanes 9 and 14). It is indicated that the ternary and binary complexes are composed by DNA, NF-Y and E-BOX TFs. These considerations also apply to the subsequent panels.

B) Formation of binary complexes composed by NF-Y and HOXB4 (lane 2), Myc 342/MAX and HOXB4 (lane 3), Myc md/MAX and HOXB4 (lane 4). Formation of ternary complex composed by HOXB4, NF-Y and Myc 342/MAX (lane 5) and ternary complex composed by HOXB4, NF-Y and Myc md/MAX (lane 10). Specificity studies using wild type or mutated competitors: CCAAT (lanes 8 and 13), mutated CCAAT (lanes 9 and 14), E-BOX (lanes 6 and 11) and mutated E-BOX (lanes 7 and 12).

C) Formation of binary complexes composed by NF-Y and LTR (lane 2), USF1 and LTR (lane 3), MAX and LTR (lane 4). Formation of ternary complex composed by LTR, NF-Y and USF1 (lane 5) and ternary complex composed by LTR, NF-Y and MAX (lane 10). Specificity studies using wild type or mutated competitors: CCAAT (lanes 6 and 11), mutated CCAAT (lanes 7 and 12), E-BOX (lanes 8 and 13) and mutated E-BOX (lanes 9 and 14).

D) Formation of binary complexes composed by NF-Y and LTR (lane 2), Myc 342/MAX and LTR (lane 3), Myc md/MAX and LTR (lane 4). Formation of ternary complex composed by LTR, NF-Y and Myc 342/MAX (lane 5) and ternary complex composed by LTR, NF-Y and Myc md/MAX (lane 10). Specificity studies using wild type or mutated competitors: CCAAT (lanes 8 and 13), mutated CCAAT (lanes 9 and 14), E-BOX (lanes 6 and 11) and mutated E-BOX (lanes 7 and 12).

E) Formation of binary complexes composed by NF-Y and HOXB7 (lane 2), MAX and HOXB7 (lane 3), USF1 and HOXB7 (lane 4), Myc 342/MAX and HOXB7 (lane 5). Formation of ternary complex composed by HOXB7, NF-Y and MAX (lane 6), ternary complex composed by HOXB7, NF-Y and USF1 (lane 9) ternary complex composed by HOXB7, NF-Y and Myc342/MAX (lane 12). Studies of specificity using wild type competitors: CCAAT (lanes 7, 10 and 13) and E-BOX (lanes 8, 11 and 14).

F) Scheme of the interplays between NF-Y and E-BOX TFs on the DNA configuration present in HOXB4, LTR and HOXB7 probes.

Figure 2: OFF rate quantifications of different ternary complexes between NF-Y/E-BOX TFs/DNA

In the OFF rate experiments we measure the stability of ternary complexes in time, using an excess of unlabelled competitors and loading the samples at different time points. These assays allow us to classify the nature of ternary complex binding in 3 classes: independent, facilitated or cooperative.

Panels A), B) and C) represent the OFF rate EMSAs of ternary complexes composed by DNA/MAX/NF-Y. In D), E) and F) the respective quantifications of two independent experiments. These are the typical behaviours of an independent binding because the ternary complexes are as stable as the binary ones in time on the different probes

analyzed. Myc342/MAX/NF-Y/DNA complexes display the same behaviour of MAX/NF-Y/DNA complexes (panels G, H and I), showing an independent binding of the two proteins on considered probes.

Instead, the ternary complex USF1/NF-Y/DNA shows a facilitated binding on HOXB4 (panel L) , that becomes more stable on HOXB7 (panel N) and cooperative on LTR (panel M). These considerations derived from the observation that the complex USF1/NF-Y/DNA is more stable in time than the binding of the single TFs on the analyzed probes, suggesting that the binding of the one stabilizes the binding of the other.

The OFF rate quantifications regards two independent experiments performed. The gels, quantified with ImageLab program (BIORAD), are reported in Supplementary 2, 3 and 4. For the concentration of TFs and probes used see Figure 1.

Figure 3: Analyses of USF1 and NF-Y domains involve in the interaction (A-C) and the importance of consensus sequences distance (D-H)

To better analyze the protein domains of USF1 and NF-Y involved in the interplay observed on HOXB4, HOXB7 and LTR probes, we performed the OFF rate assays on three probes using NF-Y md and USFs.

As we can observe from the quantification analyses of OFF rate EMSAs, the behaviour of the minimal domains of the two proteins are the same of full length TFs: facilitated binding on HOXB4 (Panel A), more stable binding of the ternary complex on HOXB7 (Panel B) and cooperative binding on LTR sequence (Panel C).

The importance of the spacing between the two consensus sequences is showed in the OFF rate experiments reported in panel F and H. Panel F and the respective quantification reported in panel D and E show the behaviour of USF1/NF-Y/DNA complexes on two probes: lanes 2-7 HOXB4, lanes 9-14 HOXB4+5. This last probe is an artificial DNA configuration derived from HOXB4 in which the spacing between CCAAT and E-BOX is of 15 bp. As we can detect from the gels and from the respective quantification, these 5 additional bp are sufficient to lose the facilitated behaviour.

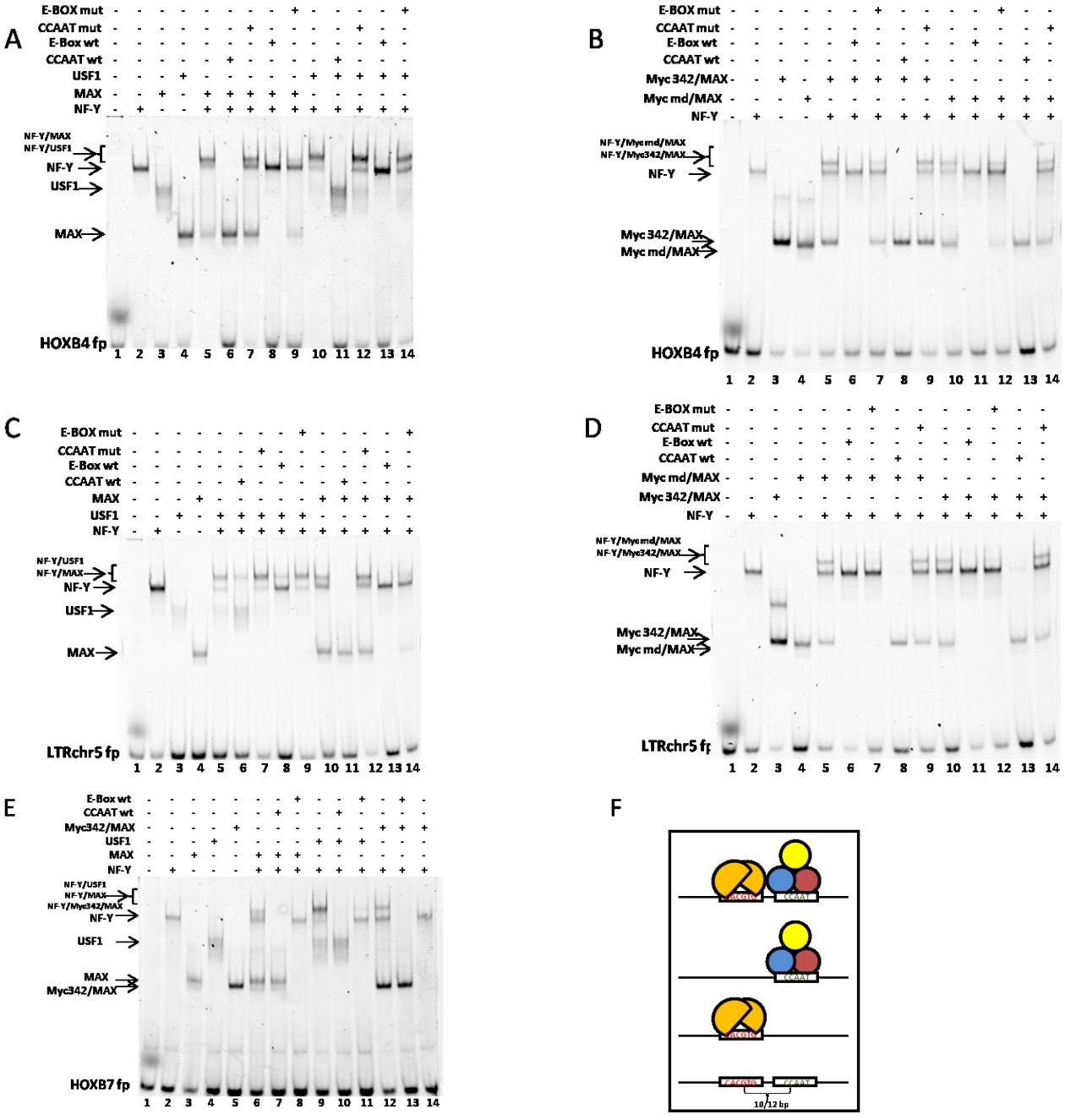
The same considerations will occur on another artificial configuration derived from LTR in which the two elements are separated by 17 bp. Panel G reports a dose response assay, in which it is clear that at the same protein concentration, there is no difference in binary complexes formation (lanes 2, 5, 9, 12, 15, 16). Instead, when we incubated together the two proteins and the DNA, the ternary complex USF1/NF-Y/LTR is more detectable than USF1/NF-Y/LTR+5 (lanes 3 and 4 versus 10 and 11; lanes 6 and 7 versus 13 and 14) .

The OFF rate experiment (panel H) shows that USF1/NF-Y/LTR+5 complex is characterized by an independent binding, that is instead cooperative in the case of USF1/NF-Y/LTR. In fact, at 60' (lanes7 and 15) the USF1/NF-Y/LTR+5 complex is no more visible (lane 15), while NF-Y still is; on the contrary the USF1/NF-Y/LTR complex is visible (lane7), while NF-Y is not.

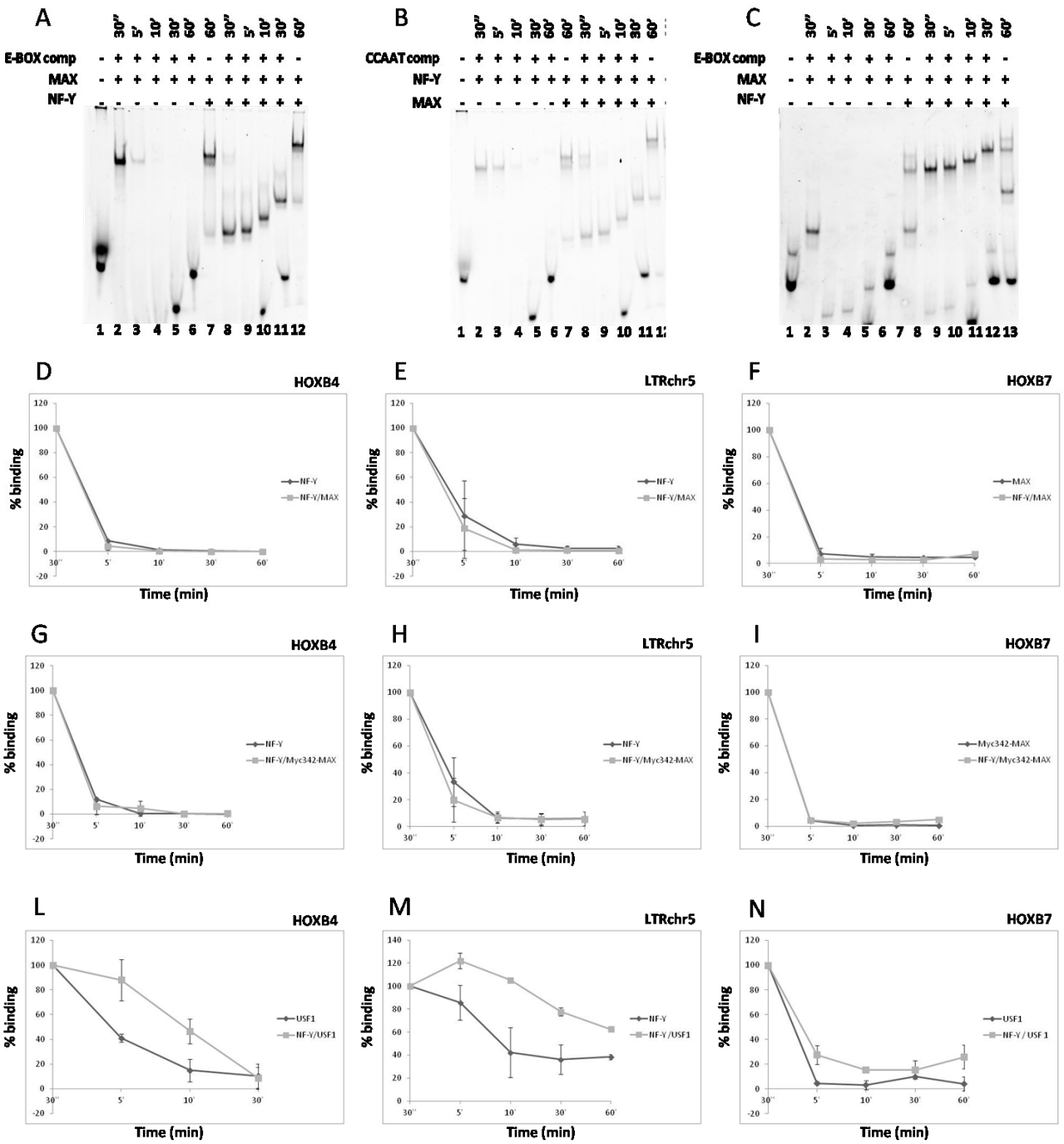
Figure 4: Western Blot analyses and Chip assays on HeLa cells silenced for NF-YB.

The western blot (WB) analyses reported in panel A and F, reveal that, after NF-YB silencing, the expression levels of Myc, MAX and USF1 doesn't change. The WB analysis were performed on nuclear extracts (20 µg) performed 72 hours post HeLa cells infection.

The 2 independent ChIP assays, performed on 2 independent NF-YB silencing, reveal that the interplays observed *in vitro* between NF-Y and USF1, Myc and MAX, occurs also *in vivo*. After NF-YB silencing, the binding of NF-YB, in Sh-NF-YB samples, decreases on its target promoters (CSK2, HOXB7, ACTL6a, LTN1, CCNB1, SERF2 and MLX). The binding of USF1 and MAX on negative targets for NF-Y, doesn't change. The binding of USF1, MAX and Myc decreases in Sh-NF-YB samples, on specific targets like MLX, CCNB1 and HOXB7. Instead after NF-YB silencing the binding of USF1 on SERF2 increases, as the binding of Myc on ACTL6a**.**



Lorenzo *et al* Figure 1



Lorenzo *et al*, Figure 2

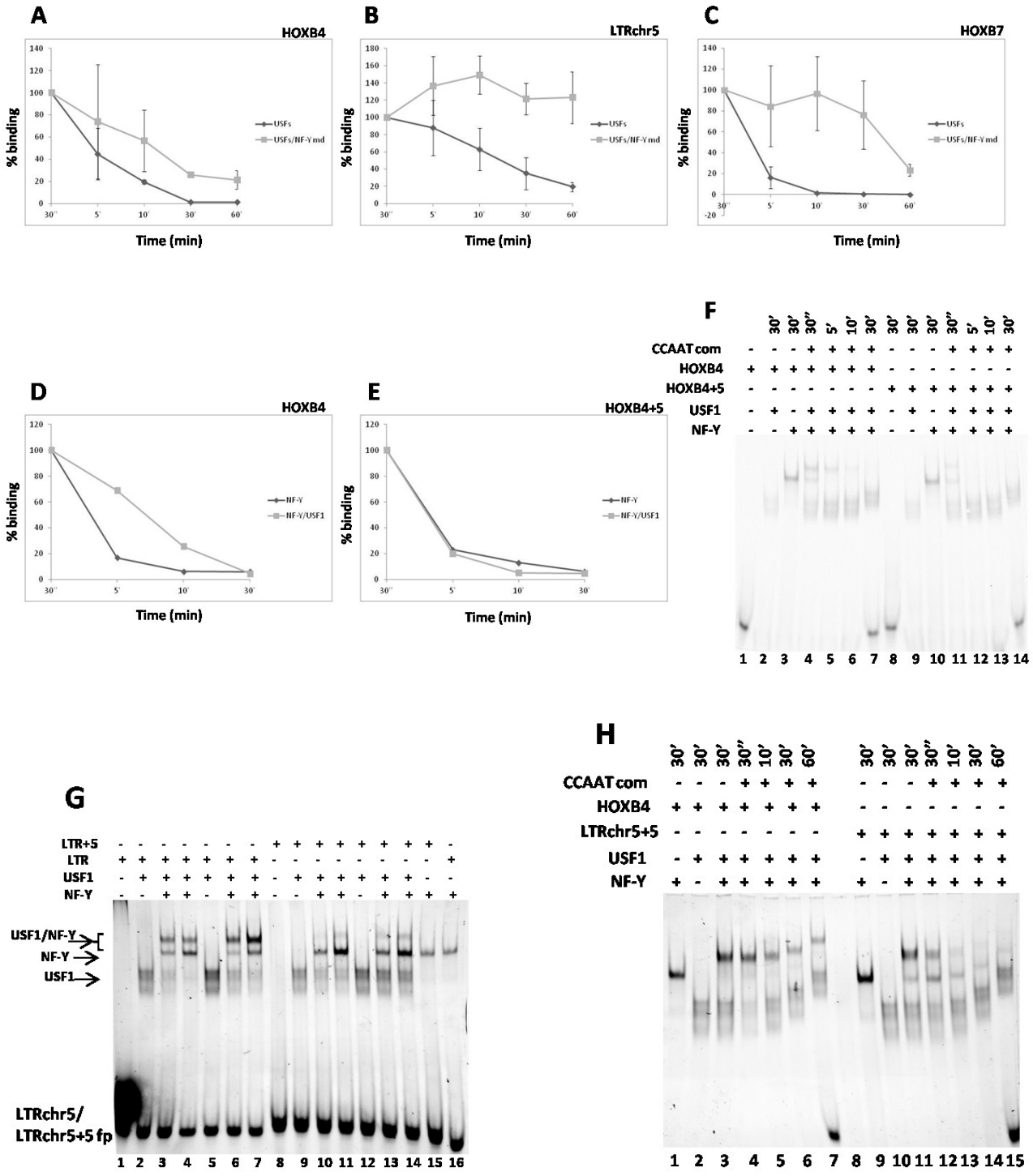


Figure 3 Lorenzo et al.

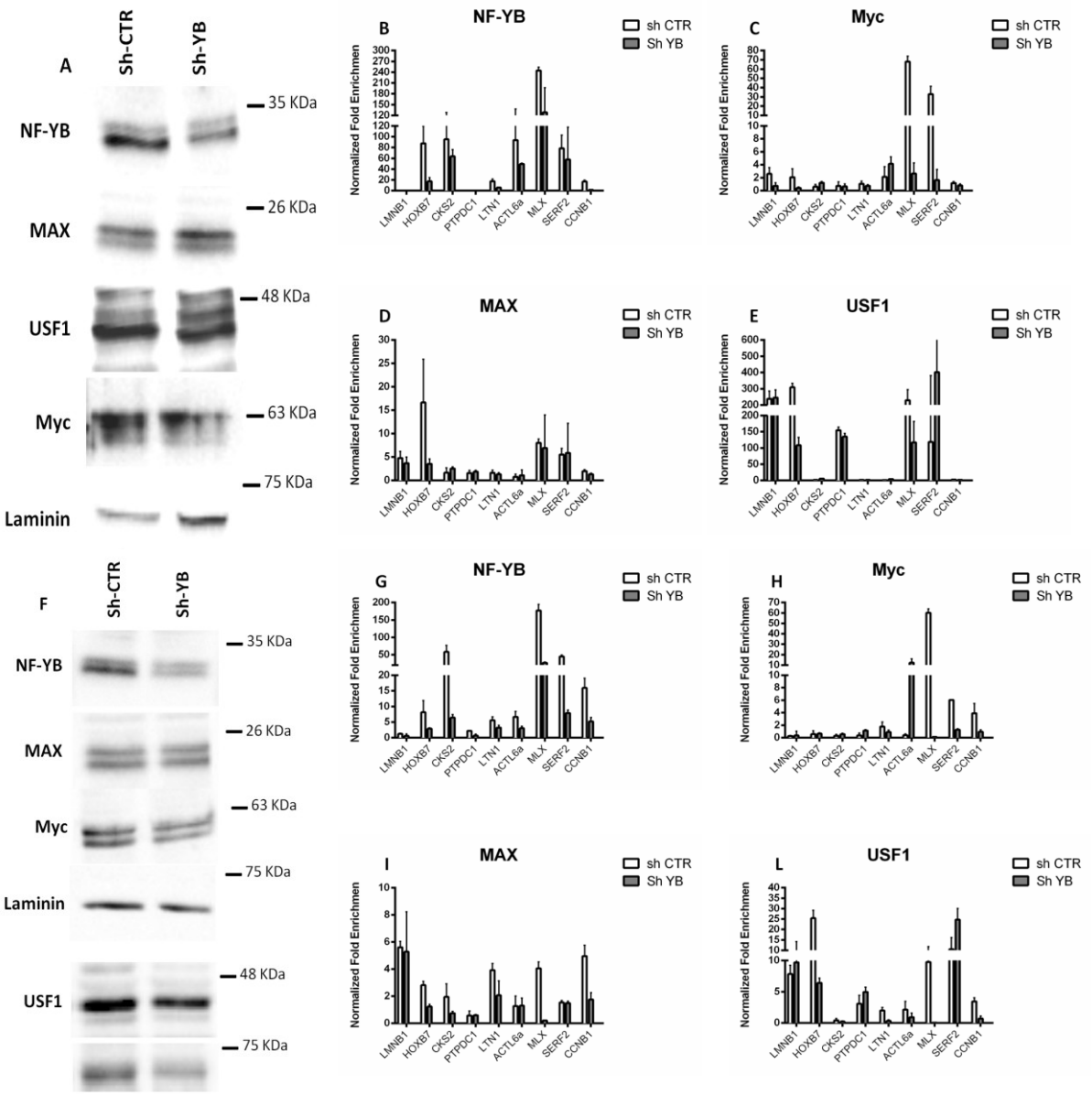
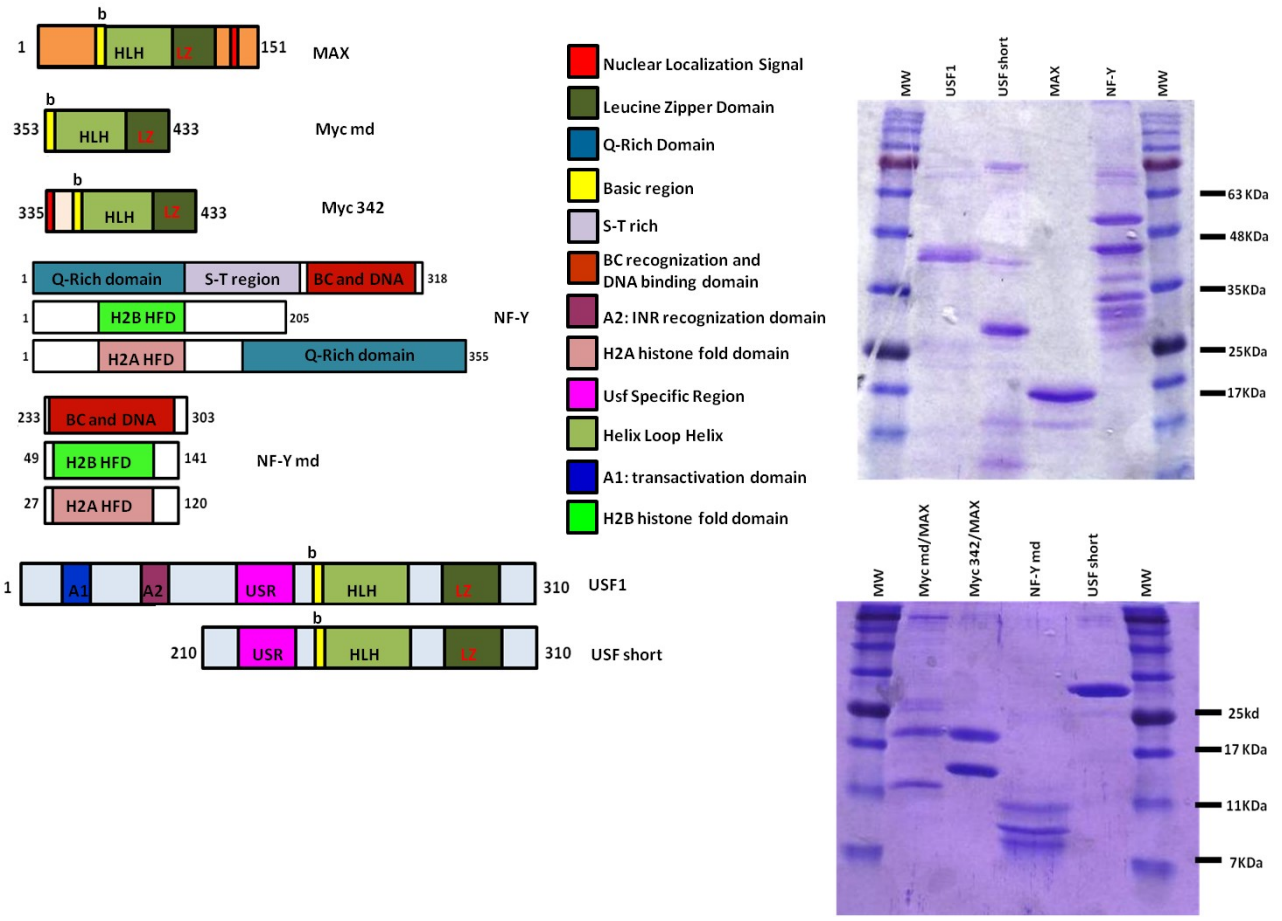
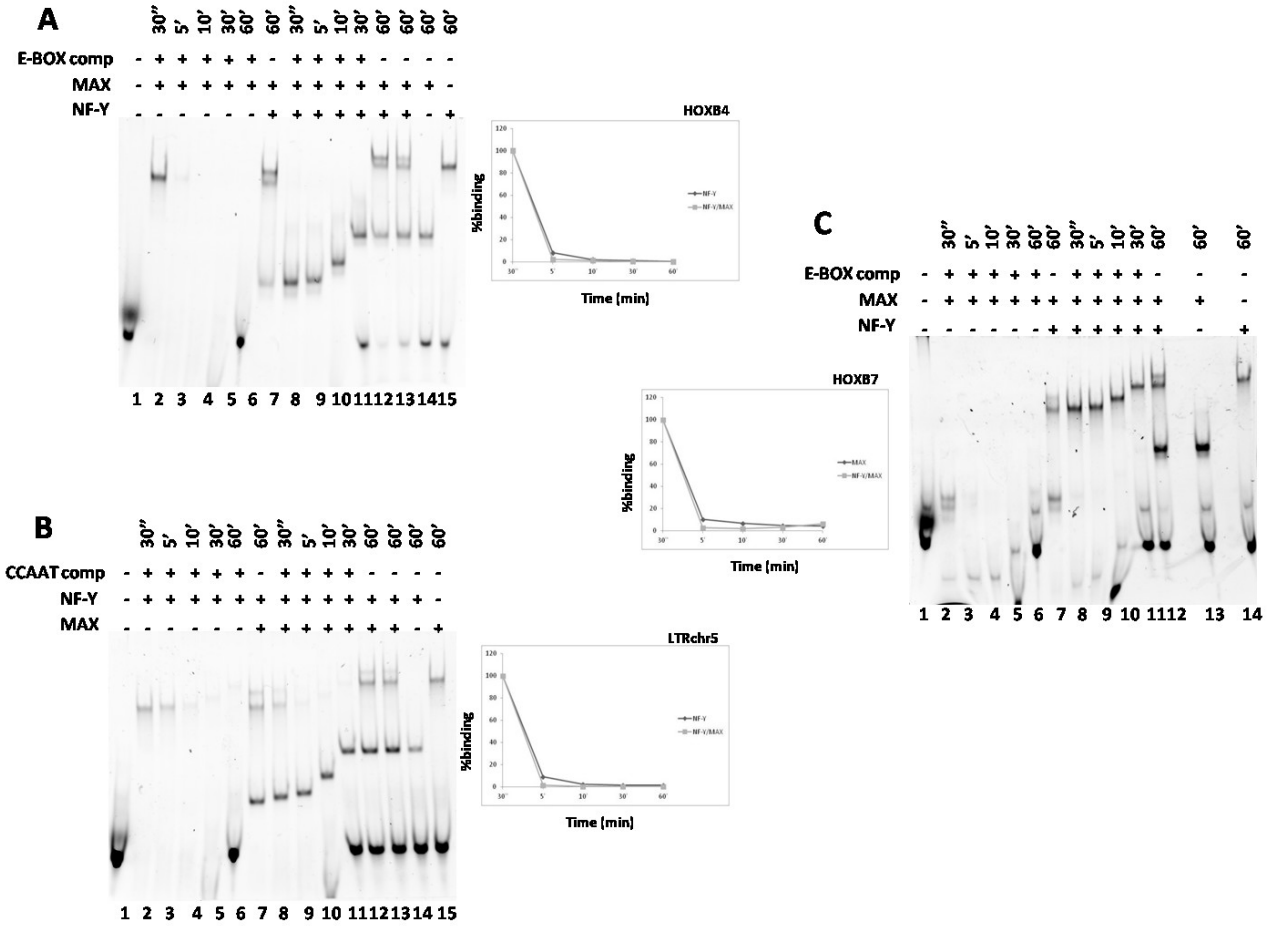


Figure 4 Lorenzo *et al*

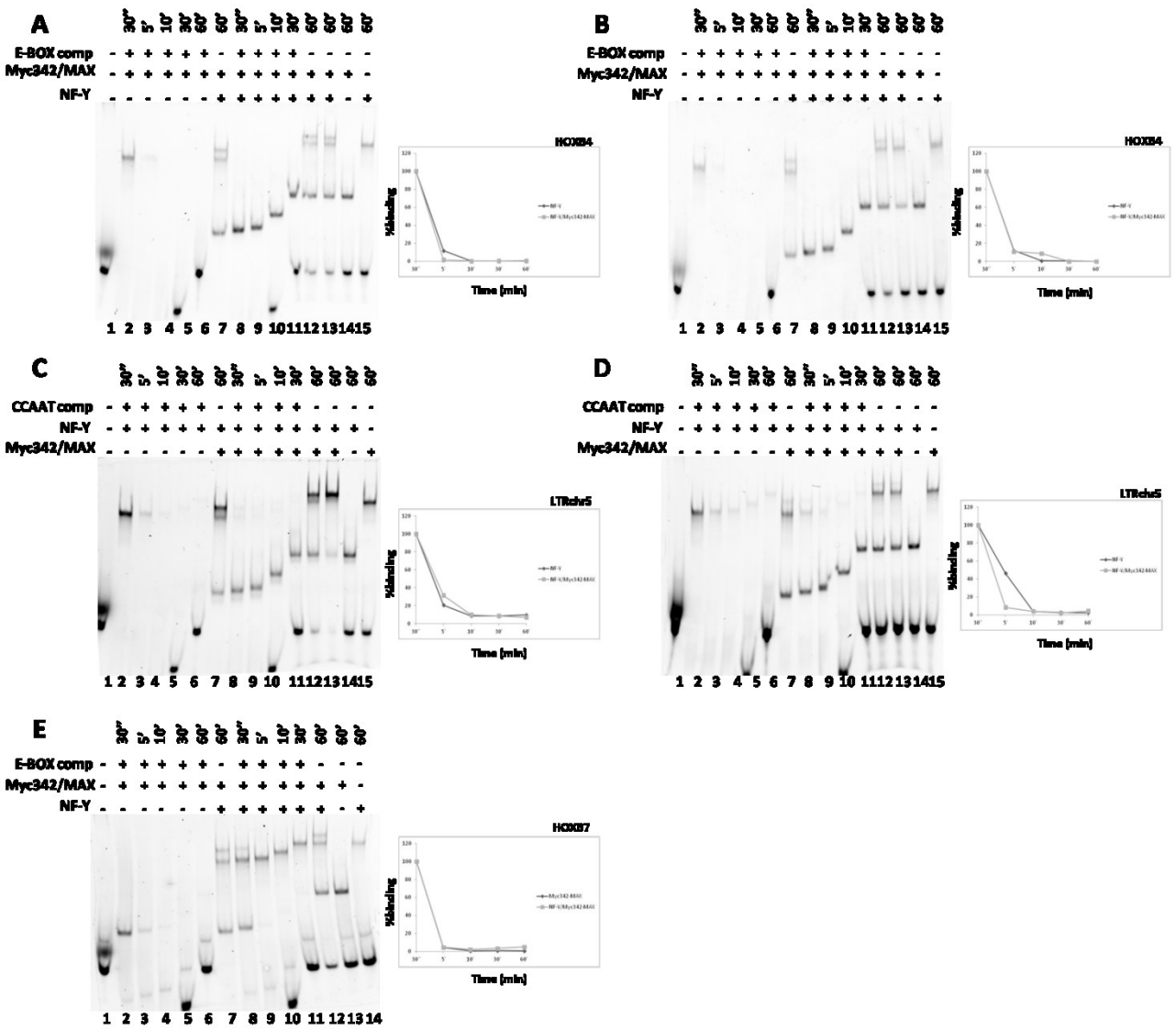


Supplementary 1: Scheme of TFs used in EMSAs and SDS-PAGE of purification steps.

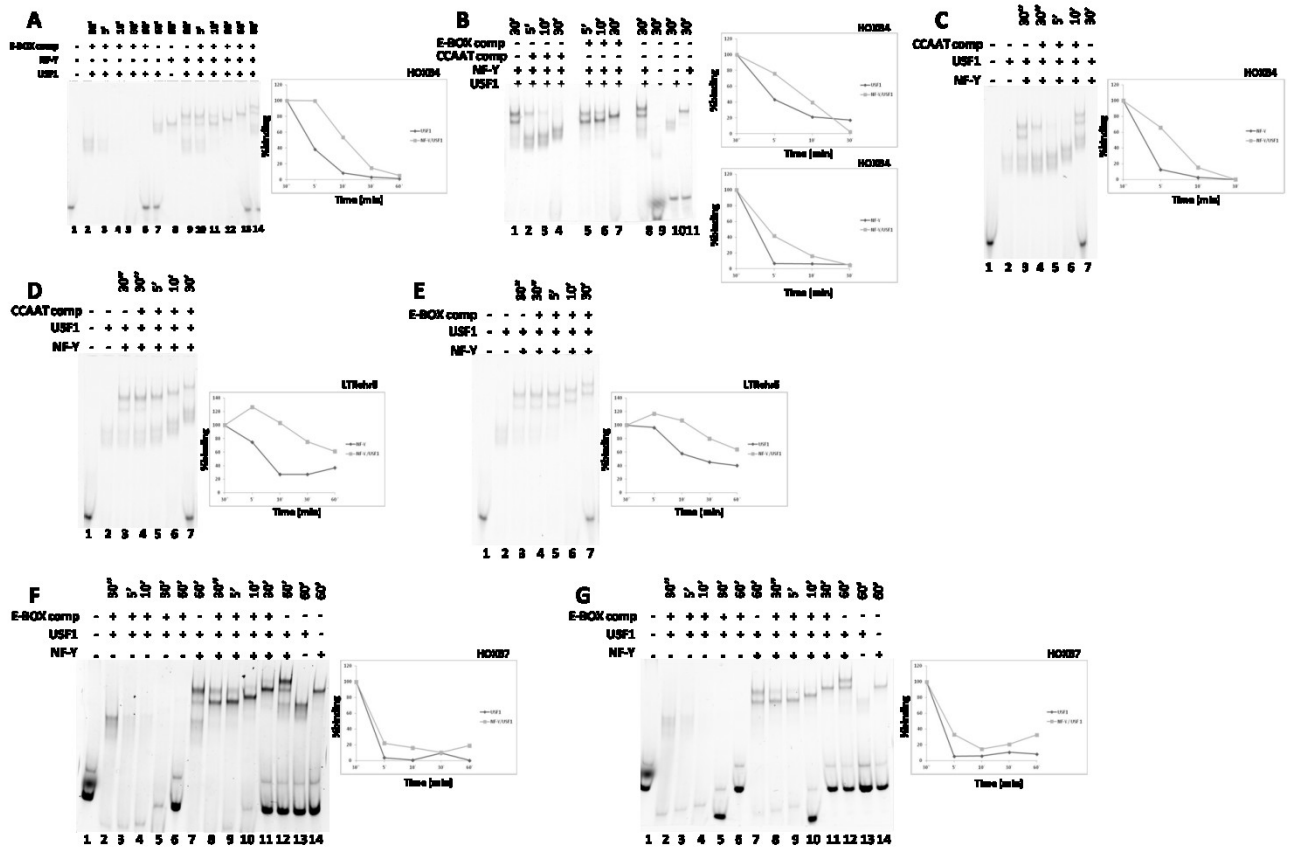
In figure are reported the scheme of full length (MAX and USF1) and minimal domains (Myc342, Myc md and USFs) E-BOX TFs used in EMSAs, which were expressed as recombinant proteins in *E. coli* and purified with Ni-affinity chromatography. The constructs of NF-Y and NF-Y md used are also indicated. The purity of different purifications can be observed from Coomassie stained gels included.



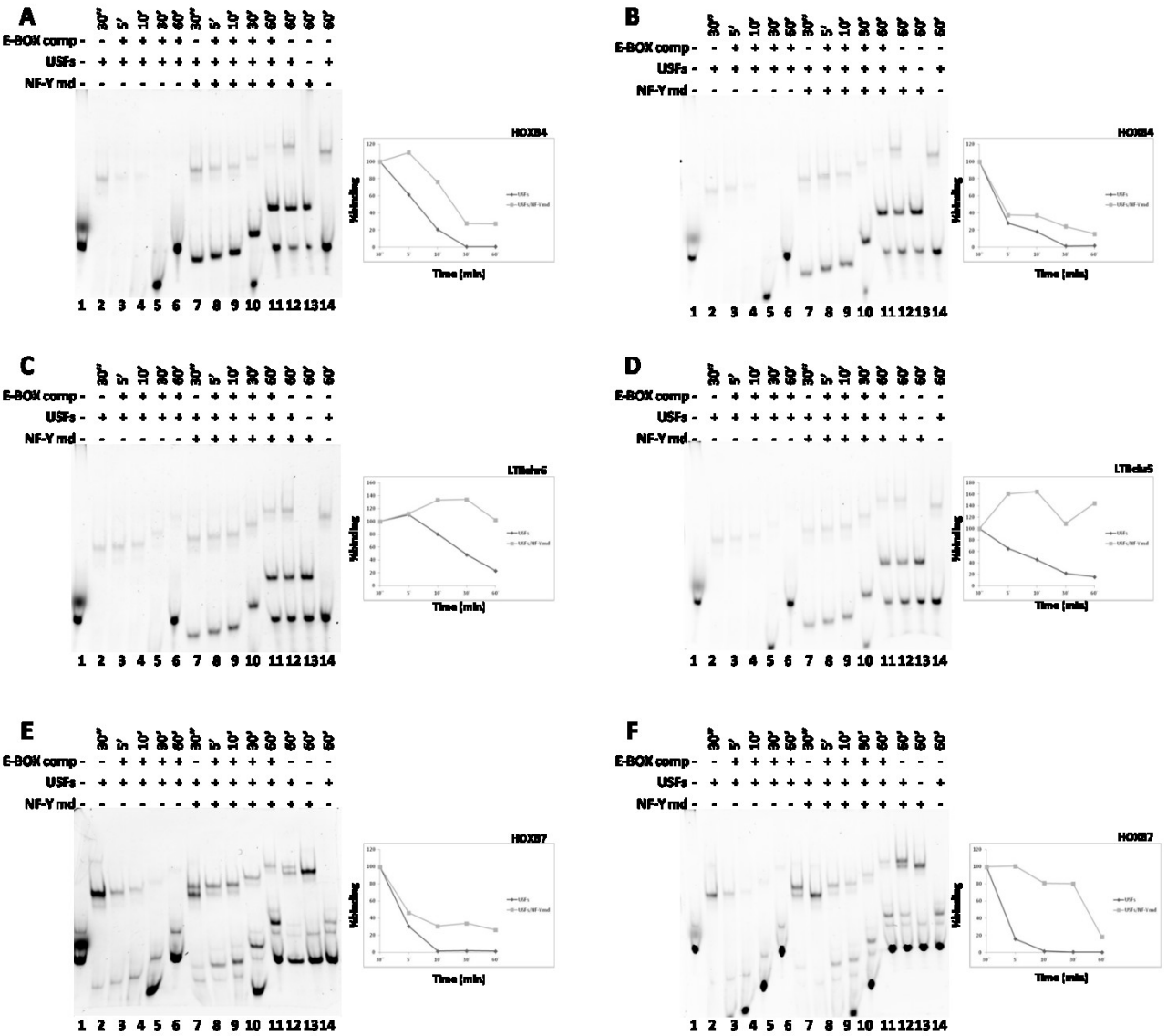
Supplementary 2: Off rate quantifications and gels of NF-Y/MAX/DNA complexes on HOXB4 (panel A), on LTR (panel B) and on HOXB7 (panel C)



Supplementary 3: Off rate quantifications and gels of Myc342/MAX/NF-Y/DNA complexes on HOXB4 (panel A and B), on LTR (panel c and D) and on HOXB7 (panel E)



Supplementary 4: Off rate quantifications and gels of NF-Y/USF1/ DNA complexes on HOXB4 (panels A, B and C), on LTR (panels D and E) and on HOXB7 (panels F and G).



Supplementary 5: Off rate quantifications and gels of USFs/NF-Ymd/DNA complexes on HOXB4 (panels A and B), on LTR (panels C and D) and on HOXB7 (panels E and F).

Supplementary 6: Probes and competitor sequences used in EMSAs.

Name	**Sequence**
HOXB4	[Cy5]-5'-TTAGGCGCC**CACGTG**ATCCTCCGAG**CCAAT**GGCCGCCCCGCCTGCGAT-3'
HOXB7	[Cy5]-5'-GACCCCGCC**CACGTG**ACGTCCCCTCCG**CCAAT**GGCCGGGCCGTCTCCCCA-3'
LTR	[Cy5]- 5'-GGACACCAT**CACGTG**ATCCAAGCCCGG**CCAAT**CAGGTTCTTTCTTTCTCA-3'
HOXB4+5	[Cy5]- 5'-TTAGGCGCC**CACGTG**AT**AAGAT**CCTCCGAG**CCAAT**GGCCGCCCCGCCTGCGAT-3'
LTR+5	[Cy5]- 5'-GGACACCAT**CACGTG**AT**AAGAT**CCAAGCCCGG**CCAAT**CAGGTTCTTTCTTTCTCA-3'
E-BOX competitor	5'-TTAGGCGCC**CACGTG**ATCCTCCGA-3'
CCAAT competitor	5'- TTCTGAG**CCAAT**CACCGAGCTCGAT -3'

Supplementary 7: Primer sequences used in qPCR

Name	Left Primer	Right Primer	Length of Fragment (bp)	Tm
LMNB1	TCCTTCCTTACAGCCCTGAGC	CGCCTCCACGTGACTACCAT	163	63°C
PTPDC1	CCTCACGCCGTCCCCTTACTT	GTGACGTGTGCTGGCCAATG	159	63°C
SERF2	CTACGTCTCACTCGGGAAGC	GCACAGTCGTTCCTTTCCTC	168	63°C
MLX	GCTGGAACCGTGGAGTAAAG	CGGCTCCGTCATCTTGTACC	176	63°C
HOXB7	AAAAGGACCCCCTTTTTCCT	CTCGGCTTTCCCATTCATTAT	160	63°C
CSK2	CCCCGTGACGTACCTATCTT	ACAACTCGCCGGAGACTAAC	217	60°C
CCNB1	GGCTTCCTCTTCACCAGGCA	CGCGATCGCCCTGGAAAC	241	66°C
ACTL6a	GGCTGCCTTCGATTGGCTAAA	AGGAGCGATAGCCCCTGACTC	207	66°C
LTN1	GACACAACTTCCGGCTTCTGG	GACAGGCGTGGCTTTAGCTGT	169	66°C
SAT	GGCGACCAATAGCCAAAAAGTGAG	CAATTATCCCTTCGGGGAATCGG	162	60°C