# STRUCTURAL DETERMINANTS FOR NF-Y/DNA INTERACTION AT THE CCAAT BOX

Valentina Nardone[1¶], Antonio Chaves-Sanjuan[1¶], Marco Nardini[1*]

[1]Dipartimento di Bioscienze, Università degli Studi di Milano, Via Celoria 26, 20133, Milano, Italy.

[¶]These authors have contributed equally to the work

*Correspondence: Marco Nardini (marco.nardini@unimi.it)

**Abstract**

The recently determined crystal structures of the sequence-specific transcription factor NF-Y have illuminated the structural mechanism underlying transcription at the CCAAT box. NF-Y is a trimeric protein complex composed by the NF-YA, NF-YB, and NF-YC subunits. NF-YB and NF-YC contain a histone-like domain and assemble on a head-to-tail fashion to form a dimer, which provides the structural scaffold for the DNA sugar-phosphate backbone binding (mimicking the nucleosome H2A/H2B-DNA assembly) and for the interaction with NF-YA. The NF-YA subunit hosts two structurally extended α-helices; one is involved in NF-YB/NF-YC binding and the other inserts deeply into the DNA minor groove, providing exquisite sequence-specificity for recognition and binding of the CCAAT box. The analysis of these structural data is expected to serve as a powerful guide for future experiments aimed at understanding the role of post-translational modification at NF-Y regulation sites and to unravel the three-dimensional architecture of higher order complexes formed between NF-Y and other transcription factors that act synergistically for transcription activation. Moreover, these structures represent an excellent starting point to challenge the formation of a stable hybrid nucleosome between NF-Y and core histone proteins, and to rationalize the fine molecular details associated to the wide combinatorial association of plant NF-Y subunits.

**Keywords**

Transcription factor; X-ray Crystallography; NF-Y; Histones; DNA-binding.

## 1. Introduction

During the recent years considerable progress has been made towards understanding the structural basis for the fascinating biology of transcriptional regulation. Studies on eukaryotic and bacterial transcription factors (TFs) have shed light in the general architecture and in the DNA recognition mode of both general TFs (GTFs) and sequence-specific TFs, showing a considerable versatility in the formation of multiprotein complexes and in the interaction with DNA.

While GTFs bind to core promoters close to the transcription start site (TSS) and assemble in an ordered fashion with RNA polymerase to form a functional pre-initiation complex, TFs recognize and bind to short regulatory DNA sequences in promoters and enhancers [1,2]. Typically they contain a sequence-specific DNA-binding domain and a separate trans-activating region, which interacts with downstream factors and coactivators. Regulatory elements contain multiple binding sites for different TFs, which assemble in a unique combination and with a defined three-dimensional architecture. The DNA sequence specificity and the combinatorial cooperation of TFs delineate a complex transcriptional regulatory code, which is still very far from a complete elucidation. A lot of efforts are ongoing to build up comprehensive databases of TFs binding profiles [2-5]. In this perspective, the structural characterization of TFs in complex with their DNA cognate sequences at the atomic level represents a potent tool to feature new DNA-binding domains [6] and to identify protein interfaces likely involved in cooperative interactions with other TFs [7-9].

In this review we focus our attention on the recent progress made towards understanding the structural basis for the biology of the sequence-specific TF Nuclear Factor-Y (NF-Y). The knowledge of genomic binding of NF-Y is derived by ENCODE [10-12]: 25% of NF-Y sites are in promoters and a comparable number are located at tissue specific enhancers. NF-Y specifically recognizes the CCAAT box, a regulatory element typically located at a conserved distance of -60/-100 bp from the TSS and present in 30% of eukaryotic promoters [13-15]. This occurrence is similar to that of the TATA box [16], and the CCAAT box is typically found in TATA-less

promoters [15]. A multitude of genes has been described to be positively or negatively regulated by NF-Y, including prosurvival and cell-cycle-promoting genes as well as genes involved in metabolism [17-22]. Indeed, knockout of NF-Y is lethal during early embryonic development [23]. In general, NF-Y cooperates with neighboring TFs, including growth-controlling and oncogenic ones [11], consistent with the enrichment of CCAAT motifs in the promoters of genes overexpressed in cancer [24].

NF-Y is a protein complex minimally composed of three subunits: NF-YA, NF-YB, and NF-YC. All subunits are conserved throughout evolution and required for DNA-binding [25]. NF-YB and NF-YC contain a core region which belongs to the class of Histone Fold Domain (HFD) proteins (Figure 1). This module mediates the formation of a tight heterodimer between the NF-YB and NF-YC subunits, and is involved in non sequence-specific DNA-binding. Heterodimerization of NF-YB and NF-YC results in the formation of a surface for NF-YA association, allowing the resulting trimer to bind DNA with high affinity and specificity. The NF-YA core domain is less than 60 amino acids long and divided into two segments: an N-terminal region responsible for NF-YC/NF-YB binding, and a C-terminal region implicated in specific recognition of the CCAAT element (Figure 1) [26-30]. In addition to these highly conserved core domains involved in trimerization, both NF-YA and NF-YC subunits display much less conserved flanking regions, which include a large Glutamine-rich (Q-rich) domain with transcriptional activation potential (Figure 1) [31-32].

In yeast the three NF-Y subunits (called HAP2, HAP3 and HAP5, respectively) have no Q-rich regions and the activation function is encoded by a fourth subunit (HAP4), present also in fungi, with no apparent homologues in other species [33,34].

Here we review the structural determinants describing the NF-Y subunits, the three-dimensional architecture of their complex, and the molecular details of the DNA recognition and binding at the CCAAT box as derived from the available NF-Y crystallographic data from mammal,

*Aspergillus nidulans* (where NF-Y is called CBC and the three subunits HapB, HapC and HapE, respectively) and *Arabidopsis thaliana* [35-38].

## 2. The NF-YB/NF-YC HFD dimer and its trimerization with NF-YA

Both NF-YB and NF-YC host a HFD, as in core histones H2A, H2B, H3, and H4 [39], and in several other proteins involved in transcription, chromatin remodeling and DNA repair [35,40-43]. Typically, the HFD is formed by a minimum of 3 helices (α1, α2 and α3) separated by two loops (L1 and L2), with helices α1 and α3 flanking almost orthogonally to α2 (Figure 2A).

The HFDs of NF-YB and NF-YC are homologous in sequence and structure to H2B and H2A, respectively, and comparison between their HFD regions reveals relatively little differences (Figure 2B). This similarity extends also outside the HFD and includes the presence of conserved additional secondary structure elements at the HFD C-termini: in NF-YB an extra αC helix is found similar to that of H2B, while in NF-YC a loop-short helix-loop motif is present, reminiscent of the short αC helix found in H2A. The N-terminal regions of the HFD modules instead differ markedly, with core histone H2A and H2B hosting tails not present in NF-YC and NF-YB (Figure 2B). Interestingly, the *Aspergillus nidulans* HapE subunit (corresponding to NF-YC) comprises also an N-terminal extension, termed αN, but adopts a different orientation relative to that present in core-histones [37]. Other structural features of HFDs that have been recognized to be specific only for NF-Y subunits are: (i) the presence of an intra-chain Arg-Asp bidentate pair linking L2 to α3 and (ii) the presence in NF-YC of an absolutely conserved Trp, at the end of helix α2, sandwiched between loop L2 of NF-YC and loop L1 of NF-YB [35,36].

At the level of quaternary structure, the HFD modules of NF-YC and NF-YB associate in a head-to-tail fashion similar to H2A/H2B within the nucleosome (Figure 2C,D) [39]. This antiparallel "handshake" assembly juxtaposes the L1 loop of NF-YB and the L2 loop of NF-YC and *vice versa*, thus generating a twofold quasi-symmetry axis between the polypeptide chains (Figure 2C). Extensive interactions between these loop regions and the hydrophobic packing of residues

belonging to the long α2 helices of NF-YB and NF-YC contribute to the exceptional stability of this heterodimeric complex. The HFD modules are responsible not only for the establishment of contacts between the two NF-YB and NF-YC subunits, but also for building the molecular platform needed for binding and bending the DNA. Like H2A/H2B, the calculated electrostatic potential of the upper surface of the HFD portion of the NF-YB/NF-YC heterodimer is highly basic, and allows favorable polar and van der Waals interactions with the negatively charged phosphodiester backbone of the DNA (Figure 3). Such interactions are apparently devoid of DNA sequence specificity. About four turns of double-stranded DNA sit on the NF-YB/NF-YC platforms, with a bending angle of about 80° compared to the ideal B-DNA. This curvature is remarkably similar to the DNA bending in nucleosomes (Figure 3A,B) and highlights the evolutionary relationship of NF-YB/NF-YC to core histone proteins [44]. The α1-α1 region and the two L1-L2 loops build the central part and the two sides of the DNA contact regions (Figure 2E) [36,37]. Notably, the α1-α1/DNA contacts involve structurally equivalent residues in NF-Y and in the nucleosome H2A/H2B, and engineering experiments indicated that the integrity of α1 is essential for DNA-binding [45].

The structure and the DNA-binding mode of the NF-YB/NF-YC dimer are also highly reminiscent of that of other HFD-containing proteins, as in the case of negative cofactor 2 (NC2α/β) (Figure 3C), a protein that represses TATA box-dependent transcription, while increasing the activity of the distal promoter element [46,47]. Overall, despite limited insight available for the NC2α/β DNA interactions due to the short length of the bound DNA [40], DNA contacts involving NC2α at α1, L1, and L2, and NC2β at L2, are essentially equivalent in NF-YC and NF-YB, respectively (Figure 3A,C). Furthermore, other HFD-containing proteins, such as the Chrac14/Chrac16 HFD heterodimer [42] and the TAF12/TAF4 subunits of TFIID, resemble the geometry of NF-YB/NF-YC with some variations in the conservation of side-chain putatively involved in the DNA interaction [36,48,49].

What makes the NF-YB/NF-YC HFD dimer markedly different relative to other HFD-containing proteins is the presence of a wide negatively charged surface groove, built mostly by residues belonging to the NF-YC αC, NF-YC α1 and NF-YB α2, responsible for binding of the NF-YA subunit (Figure 3A). Heterotrimerization occurs mainly through interaction of the A1 helix of NF-YA with the NF-YB/NF-YC HFD dimer (Figures 2E and 3A). The NF-YA A1 helix contains several polar and mostly positively charged residues which interact via hydrogen bonds and salt bridges with residues belonging to the negatively charged surface groove of the NF-YB/NF-YC HFD dimer (Figures 3A). These contacts are extensive (contact interface of about 1,760 $Å^2$) and most of them involve residues proven to be crucial for heterotrimerization in mutational studies [27-30,35,50,51]. Furthermore, these interactions are protein selective since structural differences between the NF-YB/NF-YC and H2B/H2A are present in this region and are coupled to poor conservation in H2B/H2A of negatively charged residues important for NF-Y trimerization (Figure 3A,B). Similarly, sequence variations in the corresponding HFD region of NC2 (Figure 3A,C) was demonstrated to prevent NF-YA binding to the NC2α/β HFD module [35,45].

Thus, evolutionary adaptations of the HFD regions endowed NF-YB and NF-YC with novel functions compared to other histone proteins. Besides NF-YA recognition and binding, the αC helices of NF-YB and NF-YC have been shown to be important for association with the TATA binding protein (TBP) [52,53] and, therefore, to act as a scaffold for the assembly of the transcriptional preinitiation complex. Because αC helices are not directly engaged in DNA contacts (Figure 2C,E), protein-protein interactions with other TFs are possible in the presence and absence of DNA [14]. Moreover, the αC helix of NF-YC interacts with cell cycle-controlling proteins such as the proto-oncogene c-myc [54] and the tumor suppressor p53 [55]; such interactions might suggest that they contribute to the essentiality of NF-Y in eukaryotes.

All these structural and biochemical data on NF-Y support the general idea of HFDs as protein-protein interaction modules able to bind DNA in a core histone-like mode and to favor the binding of other TFs. However, this vision is probably reductive. Indeed, NF-YB has been shown to be

ubiquitinated at residue K138 of αC, which corresponds to K120 in H2B, and that this post-translational modification (PTM) is associated to transcriptional activation [36]. NF-YB ubiquitination is important for ubiquitination of H2B at residue K120 [25], which is genetically and biochemically upstream of important activating histone methylations, such as H3K4me3 and H3K79me2/3 [56]. Considering that core histones display PTMs that are not only localized in tails, but also within the HFD [57], and that several of these modified residues are conserved in their nature and locations in the HFD subunits of NF-Y and in other TFs, it is tempting to conclude that the PTMs of HFD-containing proteins might provide a further layer of potential epigenetic control [43].

**3. CCAAT recognition and binding**

The crystal structure of the NF-Y trimer bound to DNA reveal the strategy employed by the NF-YA subunit for specific DNA sequence recognition at the CCAAT box [36,37]. While the NF-YA helix A1 mediates trimerization with the HFD heterodimer, helix A2 and the following Gly-X-Gly-Gly-Arg-Phe loop motif (Gly-loop; X=any residue) provide sequence-specific contacts to the CCAAT box by inserting deeply into the DNA minor groove, resulting in a striking minor groove widening with a maximum of about 19 Å at the first CCAAT-box adenine (Figures 2E,F and 3A). The adjacent major groove regions are not affected by the NF-YA binding and, therefore, available for the potential binding of other TFs. This observation is in keeping with the ability of NF-Y to synergize with several TFs, most of which functionally recognize bases within the major groove, at a conserved distance from CCAAT [15,25].

The Gly-loop, located after the A2 helix (Figure 2F), displays a kinked backbone which allows close proximity of the two Gly-Gly carbonyl O atoms to the bases of the CCAAT complementary DNA strand (GG) (Figure 4A). Bases on the CCAAT DNA strand (AAT) are instead hydrogen-bonded to side-chains of Arg and His residues belonging to the NF-YA A2 helix (Figure 4B). In addition, the NF-YA/DNA interactions include several stabilizing contacts to the DNA phosphate

backbone and the minor groove insertion of the Phe residue belonging to the Gly-loop between the AT/CG base pairs at the CCAAT box (Figure 4A). Interestingly, this Phe positioning into the minor groove is reminiscent of that found in the TBP-DNA complex, where two Phe residues insert into the first and last base pairs of the targeted TATA sequence [58].

All NF-YA residues involved in sequence-specific recognition are strictly conserved in agreement with earlier mutagenesis data and with the matrix of DNA specificity [36]. Hence, a combination of proper main-chain conformation and side-chain distribution makes the NF-YA A2 helix and the Gly-loop able to recognize and bind the CCAAT box by selecting the correct sequence of pyrimidine or purine bases that favor hydrogen bond interactions and disfavor steric hindrance, thus providing the structural explanation for the reduced or abolished affinity of NF-Y to mutant CCAAT motifs [59]. It is likely that the NF-YA A2 helix slides along the DNA, thereby acting as a sequence sensor, until appropriate interactions with the DNA bases at the minor groove are provided, with the Gly-loop acting as an anchor to allow for high-affinity binding of NF-Y to the CCAAT box [36,37]. This is also consistent with the fact that in the absence of bound DNA the NF-YA A2 helix is structurally disordered [37].

The conformational flexibility required to direct the NF-YA A2 helix toward the DNA, while keeping the A1 helix stably linked to the NF-YB/NF-YC interface, is provided by the A1A2-linker (Figure 2F). Interestingly, the comparison of the crystal structures of mammalian and *Aspergillus* NF-Y/DNA complex suggests some indications of the sequence-structure relationship that drives the conformation of the A1A2-linker. Indeed, when the two structures are superimposed the trimerization and the DNA binding modes are perfectly conserved, with the only significant structural mismatch being localized at the C-terminus of the A1 helix and at the A1A2-linker on the NF-YA subunit (Figure 5). The A1A2-linker is a region of relative divergence across kingdoms and paralogs, and in the two structures it follows different pathways due to specific amino acid composition. In mammalian NF-YA two residues, Gly260 and Pro263, are localized at the beginning of the linker and guide its structure by providing flexibility and directionality, respectively (Figure 5A,C). These two

residues are absent in the *Aspergillus* HapB subunit (corresponding to NF-YA), where instead Pro267, located at the end of the linker, contributes to orient the A2 helix towards the DNA minor groove (Figure 5B,C). These sequence differences at the A1A2-linker are coupled with variations at the N-terminal region of the NF-YB subunits (i.e. Ile55 and Arg46 in mammalian NF-YB and *Aspergillus* HapC, respectively), which faces the A1A2-linker. As a result, in mammalian NF-YA and *Aspergillus* HapB the two A1A2-linkers have a certain degree of structure variability but, nevertheless, the following A2 helices are in register to each other and correctly inserted in the DNA minor groove at the CCAAT box (Figure5).

**4. The "hybrid nucleosome" hypothesis**

The structure of the NF-Y/DNA complex bears some intriguing implications for the description of the possible mechanistic bases of NF-Y action *in vivo*. The CCAAT motifs are at the top of the list of the regulatory sequences that are spared by nucleosomes [60], and NF-Y has been reported to interact directly with histone H3/H4, but not with H2A/H2B, during nucleosome reconstitution *in vitro* [61]. The NF-Y/DNA structural data offer an explanation for these observations.

In nucleosomes, two left-handed superhelical turns of bent DNA are wrapped around an octameric-histone core consisting of two copies each of the core histone proteins H2A, H2B, H3, and H4. At the heart of the octamer there are two H3/H4 heterodimers related by a twofold axis of symmetry, forming a stable $(H3/H4)_2$ heterotetramer, with two H2A/H2B heterodimers bound to opposite faces of $(H3/H4)_2$ [39]. When the NF-Y heterotrimer structure is superimposed on H2B/H2A within the nucleosome, the modeled NF-YB/NF-YC substituted almost ideally H2B/H2A, with the core interaction with H3/H4 relying mostly on NF-YB (Figure 6A). Considering that NF-YB Asp115 residue is structurally homologous to H2B Glu90 (salt-bridged to H4 His75 in the nucleosome) [39], it is tempting to propose that similar interactions can stabilize a hybrid structure. Binding of one NF-Y trimer to the $(H3/H4)_2$ heterotetramer for hybrid nucleosome formation would require only minor reorientations of the NF-YA A1 helix (and the following

A1A2-linker) relative to the H4 α2 and α3 helices (Figure 6B). Note that acetylation of H4 Lys91 in the H4 α3 helix is believed to affect nucleosome stability [62] suggesting an intrinsic structural adaptability of the H4 α3 region. On the other hand, two NF-Y trimers cannot coexist in the hybrid nucleosome, since the A1 helix of one NF-YA subunit would overlap with the corresponding helix of the other NF-YA subunit and partially with NF-YB. This is also in keeping with the fact that in this model the CCAAT box is located after about 20 bp from the DNA start, and the presence of two CCAAT boxes in the hybrid nucleosome would position them at 100bp distance, which is not the distance usually found in multiple-CCAAT-containing DNA [63,64].

Thus, the available structural data suggest a model whereby NF-Y would prevent nucleosome formation through association of the NF-YB/NF-YC heterodimer to a DNA-bound $(H3/H4)_2$ heterotetramer, the NF-YA A1A2-linker providing enough flexibility for the A2 helix to search for and bind to a CCAAT-box. Interestingly the putative formation of a hybrid nucleosome was also suggested by the analysis of the *Aspergillus* CBC structure, where it was shown that interaction of the CBC with the H3/H4 pair that is part of the first nucleosome turn (with the consequent formation of a tetrasome) is possible [37]. Furthermore, the NF-Y trimer in complex with a CCAAT-containing 25bp DNA [36] has been reported to fit within the nucleosome core particle bound to the DNA sequence of nucleosome A of the 3'-LTR of the mouse mammary tumor virus (MMTV-A), with a good match not only between the NF-YB/NF-YC and H2B/H2A dimers, but also in the DNA conformation in the region corresponding to the CCAAT box [65].

In this scenario, hybrid NF-Y/H3/H4 assemblies would locally halt nucleosome formation, while providing opportunities for other TFs to associate with the neighboring target sequences. Indeed, genome-wide studies have shown that the presence of Nucleosome Free Regions in active or poised core promoters is a widespread phenomenon [66]. The location of CCAAT at about -80 bp, relative to the TSS, matches the predicted positioning of H2A/H2B in the "average" core promoter nucleosome [60].

## 5. Structural features of plant NF-Y

In the plant lineage NF-Y also consists of three subunits. However, each subunit can be encoded not by one but by a family of genes (typically about 10), both in dicots and monocots, differentially expressed in various tissues. As a consequence, different subunit combination can lead to a wide variety of NF-Y trimeric complexes, suited to face the many environmental conditions that a plant can experience [67,68].

Recently the first crystal structure of a plant NF-Y, the *Arabidopsis thaliana* NF-YB6/NF-YC3 dimer, has been solved [38]. This structure is of particular interest because within the NF-YB genes, the *Arabidopsis* NF-YB6 and NF-YB9 form a conserved subfamily originally identified in genetic experiments for their key roles in embryo maturation [69,70]. NF-YB9, also known as LEAFY COTYLEDON 1 (LEC1), and NF-YB6 or LEC1-LIKE (L1L) are embryogenesis regulators acting during transition from embryo to adult status: *lec1* mutants display pleiotropic phenotypes [71-76] and L1L (NF-YB6) was shown to be able to partially complement the *lec1* defect [70]. Importantly, chimeric constructs have demonstrated that the LEC1 function in embryos can be associated specifically to the HFD region of the NF-YB subunit [69]. Besides a general interest related to their functional action, much interest on LEC1 and L1L is also associated to their agronomic potential, since overexpression of LEC1 or L1L in various species have been reported to result in significant changes in seed lipids/oils production [75,77,78].

The crystal structure of the *Arabidopsis thaliana* L1L/NF-YC3 HFD dimer reveals the typical features of a "classical" NF-YB/NF-YC dimer with some structural specificities of the LEC1 family. In particular, the *Arabidopsis* L1L and NF-YC3 subunits interact in a head-to-tail fashion, forming a classical histone-like pair (Figure 7A), with conserved positive and negative electrostatic distributions on the surfaces predicted to be involved in DNA contacts and NF-YA interaction, respectively (Figure 7B). Indeed, the L1L/NF-YC3 dimer was demonstrated to trimerize with the *Arabidopsis* NF-YA6 subunit and the trimer to bind DNA in EMSA experiments on a high affinity CCAAT box probe derived from the human HSP70 promoter [38]. This result is important since it

shows that a plant NF-Y trimer containing a L1L subunit is able to bind DNA directly. In fact, the LEC1/L1L capacity to bind DNA was previously questioned because of the LEC1-specific presence of an Asp residue in helix α2 (LEC1 Asp55 and L1L Asp84) where in "normal" NF-YB a Lys/Arg is present and contacts the DNA phosphate backbone, 2 bps upstream of CCAAT. Genetic experiments pinpointed Asp55 as crucial for LEC1 function *in vivo*: its mutation to Lys led to loss of LEC1 activity, while an Asp substitution in a canonical *Arabidopsis* NF-YB was sufficient to confer partial LEC1 behavior [69]. Superimposition of the mammalian NF-Y trimer in complex with DNA on the *Arabidopsis* L1L/NF-YC3 dimer suggests that the electrostatic repulsion between the negatively charged Asp84 and the DNA phosphate backbone would favor a slightly shifted DNA trajectory, stabilized on the opposite side of the double helix by His79, which has been proposed to be diagnostic for LEC1 family members (Figure 7C) [38].

Another interesting structural issue associated to plant NF-Y is related to the discovery of the involvement of plant NF-Y genes in the control of photoperiod-dependent flowering time [79,80]. In particular, members of the large NF-YB/NF-YC plant families have been shown to trimerize with COSTANS (CO), a key regulator of photoperiod-induced flowering time, *via* its conserved CCT (for CO, CO-like, and TOC1) domain [79,80]. This interaction is physiologically relevant since mutants in *Arabidopsis NF-YB/NF-YC* phenocopy the *co* mutants in terms of delayed flowering-time phenotype [79,81-83]. Furthermore, the overexpression of the *Arabidopsis* NF-YA1 subunit causes late flowering, thus suggesting the intriguing possibility that CO and NF-YA compete for the interaction with plant NF-YB/NF-YC dimers [84].

Sequence comparison of the DNA-binding region of NF-YA and CO indicate a marked similarity at the A2 helix, but a shorter A1A2-linker and some sequence variation in the following Gly-loop, with the absence of the first two Gly residues (Figure 8). Furthermore, the CO region corresponding to NF-YA A1 helix is rich in basic residues, as required for the heterotrimeric assembly formation with the NF-YB/NF-YC HFD dimer. Indeed, mutations reported to affect CO activity *in vivo* [80] correspond to amino acids essential for DNA-binding by NF-YA (Figure 8).

Based on this sequence/structure analysis it has been suggested that CO might have evolved to recognize slight variations of CCAAT, with a different 5' end, while using NF-YB/NF-YC as a DNA binding platform [36].

## 6. Conclusions

X-ray structures of NF-Y have provided crucial insights into the molecular mechanism responsible for the recognition and binding to DNA and into the architecture of the NF-YB/NF-YC HFD dimer and of the NF-Y trimer. These structures serve as a powerful guide for present and future biochemical experiments aimed at understanding the contributions of individual amino acid residues to the stability of the macromolecular assembly, to the sequence-specific DNA recognition, and to the, so far poorly understood, role of post-translational modification at regulation sites. Important objectives for the future will include the use of the whole arsenal of structural biology techniques, including cryo-electron microscopy, X-ray crystallography and small angle X-ray scattering (SAXS), in combination with biochemical and genetic analysis, to unravel the three-dimensional architecture of higher order complexes formed between NF-Ys bound at multiple CCAAT boxes, between NF-Y and core histone proteins to challenge the formation of a stable hybrid nucleosome, and between NF-Y and other TFs demonstrated to act synergistically for transcription activation.

As basic studies of transcription continue to provide insight into the molecular basis of human disease, one of the challenges for the future will be to exploit structural insights of TFs for the development of novel therapeutics. In this context, NF-Y may be seen as a target for cancer progression drugs. The available genomic data of NF-Y locations are consistent with the recurrence of CCAAT motifs in promoters of genes overexpressed in cancer, and indicate that NF-Y is a pioneer TF for oncogenic activators, which acts either synergistically with adjacent DNA-binding complexes or by "tethering" of individual oncogenic partners. Interfering with such pioneer action of NF-Y could bear key implications for cancer control. The chase for an anti-proliferative drug that

could act by displacing the NF-Y/CCAAT complex has already started, mostly focusing on minor-groove binding drugs able to block interactions of NF-Y with the promoter of topoisomerase IIα, thus blocking cell cycle progression without involving activation of p53 [85,86]. The availability of the crystal structure of NF-Y in complex with its CCAAT-containing DNA target completely changes the perspective in the field, providing the unique possibility to target NF-Y, and not the DNA, for drug binding.

We can also anticipate that the research field on plant NF-Y will have a dramatic development not only in the genetics area but also in structural biology. We expect that several new structures of NF-Y and, possibly, CCT-containing proteins in complex with HFD dimers and DNA will be solved in the near future. These data will provide the means for the rationalization and the full understanding of the fine molecular details associated to the wide combinatorial association of plant NF-Y subunits. For instance, they are expected to clarify whether sequence changes associated to specific phenotypes are mirrored by structural rearrangements that can affect the trimer formation and the DNA-binding affinity/specificity in plant NF-Y.

**Acknowledgements**

This work was supported by the AIRC Investigator Grant - IG 2014 to M.N. [grant number IG-15267].

**References**

[1] M. C. Thomas, C.-M. Chiang, The general transcription machinery and general cofactors, Crit. Rev. Biochem. Mol. Biol. 41 (2006) 105–178.

[2] A. Jolma, J. Yan, T. Whitington, J. Toivonen, K.R. Nitta, P. Rastas, E. Morgunova, M. Enge, M. Taipale, G. Wei, K. Palin, J.M. Vaquerizas, R. Vincentelli, N.M. Luscombe, T.R. Hughes, P. Lemaire, E. Ukkonen, T. Kivioja, J. Taipale, DNA-binding specificities of human transcription factors, Cell 152 (2013) 327–339.

[3] A. Mathelier, O. Fornes, D.J. Arenillas, C.Y. Chen, G. Denay, J. Lee, W. Shi, C. Shyr, G. Tan, R. Worsley-Hunt, A.W. Zhang, F. Parcy, B. Lenhard, A. Sandelin, W.W. Wasserman, JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles, Nucleic Acids Res. 44 (2016) D110–115.

[4] I.V. Kulakovskiy, et al, HOCOMOCO: a comprehensive collection of human transcription factor binding sites models, Nucleic Acids Res. 41 (2013) 195–202.

[5] M.T. Weirauch, A. Cote, R. Norel, M. Annala, Y. Zhao, T.R. Riley, J. Saez-Rodriguez, T. Cokelaer, A. Vedenko, S. Talukder; DREAM5 Consortium, H.J. Bussemaker, Q.D. Morris, M.L. Bulyk, G. Stolovitzky, T.R. Hughes, Evaluation of methods for modeling transcription factor sequence specificity, Nat. Biotechnol. 31 (2013) 126–134.

[6] B. Xu, D.E. Schones, Y. Wang, H. Liang, G. Li, A structural-based strategy for recognition of transcription factor binding sites, PLoS One 8 (2013) e52460.

[7] A.M. Jaeger, C.W. Pemble, L. Sistonen, D.J. Thiele, Structures of HSF2 reveal mechanisms for differential regulation of human heat-shock factors, Nat. Struct. Mol. Biol. 23 (2016) 147–154.

[8] D. Wu, N. Potluri, J. Lu, Y. Kim, F. Rastinejad, Structural integration in hypoxia-inducible factors, Nature 524 (2015) 303–308.

[9] A. Jolma, Y. Yin, K.R. Nitta, K. Dave, A. Popov, M. Taipale, M. Enge, T. Kivioja, E. Morgunova, J. Taipale, DNA-dependent formation of transcription factor pairs alters their binding specificity, Nature 527 (2015) 384–388.

[10] ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome, Nature 489 (2012) 57–74.

[11] J. Wang, J. Zhuang, S. Iyer, X. Lin, T.W. Whitfield, M.C. Greven, B.G. Pierce, X. Dong, A. Kundaje, Y. Cheng, O.J. Rando, E. Birney, R.M. Myers, W.S. Noble, M. Snyder, Z. Weng, Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors, Genome Res. 22 (2012) 1798–1812.

[12] J.D. Fleming, G. Pavesi, P. Benatti, C. Imbriano, R. Mantovani, K. Struhl, NF-Y coassociates with FOS at promoters, enhancers, repetitive elements, and inactive chromatin regions, and is stereo-positioned with growth-controlling transcription factors, Genome Res. 23 (2013) 1195–1209.

[13] S.N. Maity, B. de Crombrugghe, Role of the CCAAT-binding protein CBF/NF-Y in transcription, Trends Biochem. Sci. 23 (1998) 174–178.

[14] R. Mantovani, The molecular biology of the CCAAT-binding factor NF-Y, Gene 239 (1999) 15–27.

[15] D. Dolfini, F. Zambelli, G. Pavesi, R. Mantovani, A perspective of promoter architecture from the CCAAT box, Cell Cycle 8 (2009) 4127–4137.

[16] C. Yang, E. Bolotin, T. Jiang, F.M. Sladek, E. Martinez, Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters, Gene 389 (2007) 52–65.

[17] K.T. Bergh, O. Litzka, A.A. Brakhage, Identification of a major cis-acting DNA element controlling the bidirectionally transcribed penicillin biosynthesis genes acvA (pcbAB) and ipnA (pcbC) of *Aspergillus nidulans*, J. Bacteriol. 178 (1996) 3908–3916.

[18] M. Ceribelli, D. Dolfini, D. Merico, R. Gatta, A.M. Viganò, G. Pavesi, R. Mantovani, The histone-like NF-Y is a bifunctional transcription factor, Mol. Cell. Biol. 28 (2008) 2047–2058.

[19] H. Deng, Y. Sun, Y. Zhang, X. Luo, W. Hou, L. Yan, Y. Chen, E. Tian, J. Han, H. Zhang, Transcription factor NFY globally represses the expression of the *C. elegans* Hox gene Abdominal-B homolog egl-5, Dev. Biol. 308 (2007) 583–592.

[20] T.G. Littlejohn, M.J. Hynes, Analysis of the site of action of the amdR product for regulation of the amdS gene of *Aspergillus nidulans*, Mol. Gen. Genet. 235 (1992) 81–88.

[21] O. Litzka, K. Then Bergh, A.A. Brakhage, The *Aspergillus nidulans* penicillin-biosynthesis gene aat (penDE) is controlled by a CCAAT-containing DNA element, Eur. J. Biochem. 238 (1996) 675–682.

[22] S. Steidl, M.J. Hynes, A.A. Brakhage, The *Aspergillus nidulans* multimeric CCAAT binding complex AnCF is negatively autoregulated via its hapB subunit gene, J. Mol. Biol. 306 (2001) 643–653.

[23] A. Bhattacharya, J.M. Deng, Z. Zhang, R. Behringer, B. de Crombrugghe, S.N. Maity, The B subunit of the CCAAT box binding transcription factor complex (CBF/NF-Y) is essential for early mouse development and cell proliferation, Cancer Res. 63 (2003) 8167–8172.

[24] H. Goodarzi, O. Elemento, S. Tavazoie, Revealing global regulatory perturbations across human cancers, Mol. Cell 36 (2009) 900–911.

[25] D. Dolfini, R. Gatta, R. Mantovani, NF-Y and the transcriptional activation of CCAAT promoters, Crit. Rev. Biochem. Mol. Biol. 47 (2012) 29–49.

[26] J.T. Olesen, L. Guarente, The HAP2 subunit of yeast CCAAT transcriptional activator contains adjacent domains for subunit association and DNA recognition: model for the HAP2/3/4 complex, Genes Dev. 4 (1990) 1714–1729.

[27] S.N. Maity, B. de Crombrugghe, Biochemical analysis of the B subunit of the heteromeric CCAAT-binding factor. A DNA-binding domain and a subunit interaction domain are specified by two separate segments, J. Biol. Chem. 267 (1992) 8286–8292.

[28] Y. Xing, J.D. Fikes, L. Guarente, Mutations in yeast HAP2/HAP3 define a hybrid CCAAT box binding domain, EMBO J. 12 (1993) 4647–4655.

[29] Y. Xing, S. Zhang, J.T. Olesen, A. Rich, L. Guarente, Subunit interaction in the CCAAT-binding heteromeric complex is mediated by a very short alpha-helix in HAP2, Proc. Natl. Acad. Sci. U.S.A. 91 (1994) 3009–3013.

[30] R. Mantovani, X.Y. Li, U. Pessara, V.H. Hooft, C. Benoist, D. Mathis, Dominant negative analogs of NF-YA, J. Biol. Chem. 269 (1994) 20340–20346.

[31] F. Coustry, S.N. Maity, S. Sinha, B. de Crombrugghe, 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 (1996) 14485–14491.

[32] A. de Silvio, C. Imbriano, R. Mantovani, Dissection of the NF-Y transcriptional activation potential, Nucleic Acids Res. 27 (1999) 2578–2584.

[33] S.L. Forsburg, L. Guarente, Identification and characterization of HAP4: a third component of the CCAAT-bound HAP2/HAP3 heteromer, Genes Dev. 3 (1989) 1166–1178.

[34] D.S. McNabb, K.A. Tseng, L. Guarente, 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 (1997) 7008– 7018.

[35] C. Romier, F. Cocchiarella, R. Mantovani, D. Moras, 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 (2003) 1336–1345.

[36] M. Nardini, N. Gnesutta, G. Donati, R. Gatta, C. Forni, A. Fossati, C. Vonrhein, D. Moras, C. Romier, M. Bolognesi, R. Mantovani, Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination, Cell 152 (2013) 132-143.

[37] E.M. Huber, D.H. Scharf, P. Hortschansky, M. Groll, A.A. Brakhage, DNA minor groove sensing and widening by the CCAAT-binding complex, Structure 20 (2012) 1757–1768.

[38] N. Gnesutta, D. Saad, A. Chaves-Sanjuan, R. Mantovani, M. Nardini, Crystal structure of the *Arabidopsis thaliana* L1L/NF-YC3 histone-fold dimer reveals specificities of the LEC1 family of NF-Y subunits in plants, Mol. Plant (2016) Under revision.

[39] K. Luger, A.W. Mäder, R.K. Richmond, D.F. Sargent, T.J. Richmond, Crystal structure of the nucleosome core particle at 2.8 Å resolution, Nature 389 (1997) 251–260.

[40] K. Kamada, F. Shu, H. Chen, S. Malik, G. Stelzer, R.G. Roeder, M. Meisterernst, S.K. Burley, Crystal structure of negative cofactor 2 recognizing the TBP-DNA transcription complex, Cell 106 (2001) 71–81.

[41] E. Cler, G. Papai, P. Schultz, I. Davidson, Recent advances in understanding the structure and function of general transcription factor TFIID, Cell Mol. Life Sci. 66 (2009) 2123–2134.

[42] K.F. Hartlepp, C. Fernández-Tornero, A. Eberharter, T. Grüne, C.W. Müller, P.B. Becker, The histone fold subunits of Drosophila CHRAC facilitate nucleosome sliding through dynamic DNA interactions, Mol. Cell. Biol. 25 (2005) 9886–9896.

[43] N. Gnesutta, M. Nardini, R. Mantovani, The H2A/H2B-like histone-fold domain proteins at the crossroad between chromatin and different DNA metabolisms, Transcription 4 (2013) 114–119.

[44] C. Liberati, A. Ronchi, P. Lievens, S. Ottolenghi, R. Mantovani, NF-Y organizes the gamma-globin CCAAT boxes region, J. Biol. Chem. 273 (1998) 16880–16889.

[45] K. Zemzoumi, M. Frontini, M. Bellorini, R. Mantovani, NF-Y histone fold alpha1 helices help impart CCAAT specificity. J. Mol. Biol. 286 (1999) 327–337.

[46] E. Maldonado, M. Hampsey, D. Reinberg, Repression: targeting the heart of the matter, Cell 99 (1999) 455–458.

[47] P.J. Willy, R. Kobayashi, J.T. Kadonaga, A basal transcription factor that activates or represses transcription, Science 290 (2000) 982–985.

[48] K.J. Wright, M.T. II Marr, R. Tjian, TAF4 nucleates a core subcomplex of TFIID and mediates activated transcription from a TATA-less promoter, Proc. Natl. Acad. Sci. U.S.A. 103 (2006) 12347–12352.

[49] K. Gazit, S. Moshonov, R. Elfakess, M. Sharon, G. Mengus, I. Davidson, R. Dikstein, TAF4/4b x TAF12 displays a unique mode of DNA binding and is required for core promoter function of a subset of genes, J. Biol. Chem. 284 (2009) 26286–26296.

[50] S. Sinha, I.S. Kim, K.Y. Sohn, B. de Crombrugghe, S.N. Maity, 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 (1996) 328–337.

[51] I.S. Kim, S. Sinha, B. de Crombrugghe, S.N. Maity, Determination of functional domains in the C subunit of the CCAAT-binding factor (CBF) necessary for formation of a CBF-DNA complex: CBF-B interacts simultaneously with both the CBF-A and CBF-C subunits to form a heterotrimeric CBF molecule, Mol. Cell. Biol. 16 (1996) 4003–4013.

[52] M. Bellorini, D.K. Lee, J.C. Dantonel, K. Zemzoumi, R.G. Roeder, L. Tora, R. Mantovani, CCAAT binding NF-Y-TBP interactions: NF-YB and NF-YC require short domains adjacent to their histone fold motifs for association with TBP basic residues, Nucleic Acids Res. 25 (1997) 2174–2181.

[53] K. Matuoka, K. Yu Chen, Nuclear factor Y (NF-Y) and cellular senescence, Exp. Cell Res. 253 (1999) 365–371.

[54] H. Izumi, C. Molander, L.Z. Penn, A. Ishisaki, K. Kohno, K. Funa, Mechanism for the transcriptional repression by c-Myc on PDGF betareceptor, J. Cell Sci. 114 (2001) 1533–1544.

[55] C. Imbriano, N. Gnesutta, R. Mantovani, The NF-Y/p53 liaison: well beyond repression, Biochim. Biophys. Acta 1825 (2012) 131-139.

[56] T. Suganuma, J.L. Workman, Signals and combinatorial functions of histone modifications, Annu. Rev. Biochem. 80 (2011) 473–499.

[57] M.S. Cosgrove, J.D. Boeke, C. Wolberger, Regulated nucleosome mobility and the histone code, Nat. Struct. Mol. Biol. 11 (2004) 1037–1043.

[58] J.L. Kim, D.B. Nikolov, S.K. Burley, Co-crystal structure of TBP recognizing the minor groove of a TATA element, Nature 365 (1993) 520–527.

[59] W. Bi, L. Wu, F. Coustry, B. de Crombrugghe, S.N. Maity, DNA binding specificity of the CCAAT-binding factor CBF/NF-Y, J. Biol. Chem. 272 (1997) 26562–26572.

[60] E. Segal, Y. Fondufe-Mittendorf, L. Chen, A. Thåström, Y. Field, I.K. Moore, J.P. Wang, J. Widom, A genomic code for nucleosome positioning, Nature 442 (2006) 772–778.

[61] G. Caretti, M.C. Motta, R. Mantovani, NF-Y associates with H3-H4 tetramers and octamers by multiple mechanisms, Mol. Cell. Biol. 19 (1999) 8591–8603.

[62] J. Ye, X. Ai, E.E. Eugeni, L. Zhang, L.R. Carpenter, M.A. Jelinek, M.A. Freitas, M.R. Parthun, Histone H4 lysine 91 acetylation a core domain modification associated with chromatin assembly, Mol. Cell 18 (2005) 123–130.

[63] C. Liberati, A. di Savio, S. Ottolenghi, R. Mantovani R, NF-Y binding to twin CCAAT boxes: role of Q-rich domains and histone fold helices, J. Mol. Biol. 285 (1999) 1441–1455.

[64] D. Dolfini, F. Zambelli, M. Pedrazzoli, R. Mantovani, G. Pavesi, A high definition look at the NF-Y regulome reveals genome-wide associations with selected transcription factors, Nucleic Acids Res. 44 (2016) 4684–4702.

[65] T.D. Frouws, S.C. Duda, T.J. Richmond, X-ray structure of the MMTV-A nucleosome core, Proc. Natl. Acad. Sci. U.S.A. 113 (2016) 1214–1219.

[66] C. Jiang, B.F. Pugh, Nucleosome positioning and gene regulation: advances through genomics, Nat. Rev. Genet. 10 (2009) 161–172.

[67] T. Laloum, S. De Mita, P. Gamas, M. Baudin, A. Niebel, CCAAT-box binding transcription factors in plants: Y so many? Trends Plant Sci. 18 (2013) 157–166.

[68] K. Petroni, R.W. Kumimoto, N. Gnesutta, V. Calvenzani, M. Fornari, C. Tonelli, B.F. 3rd Holt, R. Mantovani, The promiscuous life of plant NUCLEAR FACTOR Y transcription factors, Plant Cell. 24 (2012) 4777–4792.

[69] H. Lee, R.L. Fischer, R.B. Goldberg, J.J. Harada, *Arabidopsis* LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor, Proc. Natl. Acad. Sci. U.S.A. 100 (2003) 2152–2156.

[70] R.W. Kwong, A.Q. Bui, H. Lee, L.W. Kwong, R.L. Fischer, R.B. Goldberg, J.J. Harada, LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development, Plant Cell 15 (2003) 5–18.

[71] S. Zhang, L. Wong, L. Meng, P.G. Lemaux, Similarity of expression patterns of knotted1 and ZmLEC1 during somatic and zygotic embryogenesis in maize (Zea mays L.), Planta 215 (2002) 191–194.

[72] K. Yazawa, K. Takahata, H. Kamada, Isolation of the gene encoding Carrot leafy cotyledon1 and expression analysis during somatic and zygotic embryogenesis, Plant Physiol. Biochem. 42 (2004) 215–223.

[73] M. Fambrini, C. Durante, G. Cionini, C. Geri, L. Giorgetti, V. Michelotti, M. Salvini, C. Pugliesi, Characterization of LEAFY COTYLEDON1-LIKE gene in Helianthus annuus and its relationship with zygotic and somatic embryogenesis, Dev. Genes Evol. 216 (2006) 253–264.

[74] L. Alemanno, M. Devic, N. Niemenak, C. Sanier, J. Guilleminot, M. Rio, J.L. Verdeil, P. Montoro, Characterization of leafy cotyledon1-like during embryogenesis in *Theobroma cacao* L, Planta 227 (2008) 853–866.

[75] J. Mu, H. Tan, Q. Zheng, F. Fu, Y. Liang, J. Zhang, X. Yang, T. Wang, K. Chong, X.J. Wang, J. Zuo, LEAFY COTYLEDON1 is a key regulator of fatty acid biosynthesis in *Arabidopsis*, Plant Physiol. 148 (2008) 1042–1054.

[76] P. Schellenbaum, A. Jacques, P. Maillot, C. Bertsch, F. Mazet, S. Farine, B. Walter, Characterization of VvSERK1, VvSERK2, VvSERK3 and VvL1L genes and their expression during somatic embryogenesis of grapevine (*Vitis vinifera* L.), Plant Cell Rep. 27 (2008) 1799–1809.

[77] B. Shen, W.B. Allen, P. Zheng, C. Li, K. Glassman, J. Ranch, D. Nubel, M.C. Tarczynski, Expression of ZmLEC1 and ZmWRI1 increases seed oil production in maize, Plant Physiol. 153 (2010) 980–987.

[78] H. Tan, X. Yang, F. Zhang, X. Zheng, C. Qu, J. Mu, F. Fu, J. Li, R. Guan, H. Zhang, G. Wang, J. Zuo, Enhanced seed oil production in canola by conditional expression of Brassica napus LEAFY COTYLEDON1 and LEC1-LIKE in developing seeds, Plant Physiol. 156 (2011) 1577–1588.

[79] O. Ben-Naim, R. Eshed, A. Parnis, P. Teper-Bamnolker, A. Shalit, G. Coupland, A. Samach, E. Lifschitz, The CCAAT binding factor can mediate interactions between CONSTANS-like proteins and DNA, Plant J. 46 (2006) 462–476.

[80] S. Wenkel, F. Turck, K. Singer, L. Gissot, J. Le Gourrierec, A. Samach, G. Coupland, CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of *Arabidopsis*, Plant Cell 18 (2006) 2971–2984.

[81] X. Cai, J. Ballif, S. Endo, E. Davis, M. Liang, D. Chen, D. DeWald, J. Kreps, T. Zhu, Y. Wu, A putative CCAAT binding transcription factor is a regulator of flowering timing in *Arabidopsis*, Plant Physiol. 145 (2007) 98–105.

[82] N.Z. Chen, X.Q. Zhang, P.C. Wei, Q.J. Chen, F. Ren, J. Chen, X.C. Wang, AtHAP3b plays a crucial role in the regulation of flowering time in *Arabidopsis* during osmotic stress, J. Biochem. Mol. Biol. 40 (2007) 1083–1089.

[83] R.W. Kumimoto, Y. Zhang, N. Siefers, B.F. III Holt, NF-YC3, NF-YC4 and NF-YC9 are required for CONSTANS-mediated, photoperiod-dependent flowering in *Arabidopsis thaliana*, Plant J. 63 (2010) 379–391.

[84] C. Li, A. Distelfeld, A. Comis, J. Dubcovsky, Wheat flowering repressor VRN2 and promoter CO2 compete for interactions with NUCLEAR FACTOR-Y complexes, Plant J. 67 (2011) 763–773.

[85] M. Kotecha, J. Kluza, G. Wells, C.C. O'Hare, C. Forni, R. Mantovani, P.W. Howard, P. Morris, D.E. Thurston, J.A. Hartley, D. Hochhauser, Inhibition of DNA binding of the NF-Y transcription factor by the pyrrolobenzodiazepine-polyamide conjugate GWL-78, Mol. Cancer Ther. 7 (2008) 1319–1328.

[86] H. Mackay, T. Brown, J.S. Sexton, M. Kotecha, B. Nguyen, W.D. Wilson, J. Kluza, B. Savic, C. O'Hare, D. Hochhauser, M. Lee, J.A. Hartley, Targeting the inverted CCAAT Box-2 of the topoisomerase IIalpha gene: DNA sequence selective recognition by a polyamide-intercalator as a staggered dimer, Bioorg. Med. Chem. 16 (2008) 2093–2102.

**Figure legends**

**Figure 1. Schematic representation of the NF-Y subunits.** Color shaded boxes highlight the evolutionarily conserved NF-YB/NF-YC and DNA binding domain (BC-DNA; green) in NF-YA, H2B-like or H2A-like HFD domains in NF-YB and NF-YC (orange and cyan, respectively), and the Gln-rich domains (Q-rich) in NF-YA and NF-YC (grey). The secondary structure composition of the NF-Y subunits is shown in the enlargement panels.

**Figure 2. Three dimensional structure of NF-Y. (A)** The tertiary structure of mammalian NF-YB (orange) and NF-YC (cyan) are shown as ribbon [36]. The secondary structure elements are indicated. **(B)** Structural superimposition of NF-YB and NF-YC (PDB-code 4CSR), illustrated as coils, to core histone H2B (magenta) and H2A (yellow) (PDB-code 1AOI [39], chains C/D, respectively). The N-terminal histone tails (grey) are indicated. **(C)** Ribbon diagram showing the NF-YB/NF-YC HFD dimer, and **(D)** structural superimposition with H2B/H2A. **(E)** Ribbon representations of the NF-YA/NF-YB/NF-YC heterotrimer in complex with the HSP70 25 bp-CCAAT-box oligonucleotide (PDB-code 4AWL [36]); secondary structure elements are labeled and those at the protein-DNA interface are highlighted in red. NF-YA and the DNA are colored in green and grey, respectively, with CCAAT nucleotides highlighted in violet. **(F)** The tertiary structure of NF-YA. The secondary structure elements, the A1A2-linker and the Gly-loop are indicated.

**Figure 3. Electrostatic surface and DNA binding-mode of HFD dimers.** Electrostatic surface of **(A)** NF-YB/NF-YC (NF-YA and DNA represented in ribbon and stick models color-coded as in Figure 2) [36], (B) H2B/H2A (DNA in grey) [39], and **(C)** NC2α/β (PDB-code 1JFI [40]; DNA in grey and TBP in brown). Blue and red colors indicate positively and negatively charged regions, respectively. In panel (A), the widening of the DNA minor groove due to the NF-YA A2 insertion in highlighted.

**Figure 4. Sequence-specific interactions of NF-Y/DNA at the CCAAT-box. (A)** Hydrogen bond interactions of NF-YA residues belonging to the Gly-loop with the CCAAT DNA complementary strand (grey), and **(B)** interactions between residues from the NF-YA A2 helix and the CCAAT DNA strand (CCAAT in violet) [36]. Key residues are labeled, and hydrogen bonds are shown by dashed lines.

**Figure 5. Structure of the A1A2-linker.** Structural comparison of the A1A2-linker in **(A)** the mammalian NF-Y (PDB-code 4AWL [36]) and **(B)** the *A. nidulans* CBC (PDB-code 4G92 [37]). The HFD subunits are shown in surface representation with NF-YB and the HapC in orange and NF-YC and HapE in cyan. The NF-YA and HapB subunits are shown in ribbon representation (green and blue, respectively). Relevant residues are represented in stick and indicated. **(C)** Structural-based sequence alignment of the A1A2-linker in NF-YA and HapB. Identical and similar residues are highlighted by yellow and grey shadings, respectively. Relevant Pro and Gly residues are shaded in black.

**Figure 6. Overlay of NF-Y on nucleosome $(H3/H4)_2$ tetramer. (A)** Side and top views of the hybrid nucleosome model. The NF-Y subunits (colour code as in Figure 1) of one NF-Y trimer [36] are shown after superimposition of the NF-YB/NF-YC on one nucleosome H2B/H2A (PDB-code 1AOI, chains D/C, respectively [39]). For clarity, the H2B/H2A dimers are not shown. The corresponding H3 and H4 subunits (PDB-code 1AOI: chains E/F, respectively [39]) are shown in red and yellow colours. The H3'and H4' subunits (PDB-code 1AOI: chains A/B, respectively [39]), forming with H3 and H4 the $(H3/H4)_2$ tetramer, are shown in blue and pink colours. **(B)** Close up of the NF-YA A1 helix. A small rearrangement of the NF-YA, H4 and H4' helices (indicated) is needed to avoid close contacts (highlighted by white arrows).

**Figure 7. The *Arabidopsis* L1L/NF-YC3 dimer. (A)** Ribbon diagram showing the L1L/NF-YC3 HFD dimer (L1L in light green and NF-Y3 in light pink), and **(B)** its electrostatic surface (PDB-code 5G49) [38]. The blue and red colors indicate positively and negatively charged regions, respectively. The orientation of the L1L/NF-YC3 dimer is similar to that of mammalian NF-YB/NF-YC in Figure 2C. **(C)** Superposition of the *Arabidopsis* L1L/NF-YC3 dimer on the mammalian NF-Y/DNA complex. The figure shows the position of L1L Asp84 and His79 (light green sticks) relative to DNA (grey, CCAAT box in violet) and to the corresponding NF-YB residues Lys78 and Thr73 (orange sticks). For clarity, the mammalian NF-YA and NF-YC and the *Arabidopsis* NF-YC3 subunits are omitted.

**Figure 8. Sequence alignment of the CCT domain with NF-YA**. The consensus sequence for plant CCT domains is aligned with the consensus sequence for *Arabidopsis thaliana* NF-YA proteins (AtNF-YA) and human NF-YA (hNF-YA). Residues identical and similar are highlighted in yellow and grey, respectively. Asterisks indicate the position of CCT domain mutations identified in plants [80]. The consensus sequence for plant CCT and AtNF-YA is derived from the sequence alignment of 24 CCT proteins and 10 AtNF-YA subunits, as reported in Petroni et al., 2012 [68], with residues belonging to the consensus if conserved in ≥ 50% of the aligned sequences. Variable residues are indicated by x.

# STRUCTURAL DETERMINANTS FOR NF-Y/DNA INTERACTION AT THE CCAAT BOX

Valentina Nardone[1¶], Antonio Chaves-Sanjuan[1¶], Marco Nardini[1*]

[1]Dipartimento di Bioscienze, Università degli Studi di Milano, Via Celoria 26, 20133, Milano, Italy.

[¶]These authors have contributed equally to the work

*Correspondence: Marco Nardini (marco.nardini@unimi.it)

**Abstract**

The recently determined crystal structures of the sequence-specific transcription factor NF-Y have illuminated the structural mechanism underlying transcription at the CCAAT box. NF-Y is a trimeric protein complex composed by the NF-YA, NF-YB, and NF-YC subunits. NF-YB and NF-YC contain a histone-like domain and assemble on a head-to-tail fashion to form a dimer, which provides the structural scaffold for the DNA sugar-phosphate backbone binding (mimicking the nucleosome H2A/H2B-DNA assembly) and for the interaction with NF-YA. The NF-YA subunit hosts two structurally extended α-helices; one is involved in NF-YB/NF-YC binding and the other inserts deeply into the DNA minor groove, providing exquisite sequence-specificity for recognition and binding of the CCAAT box. The analysis of these structural data is expected to serve as a powerful guide for future experiments aimed at understanding the role of post-translational modification at NF-Y regulation sites and to unravel the three-dimensional architecture of higher order complexes formed between NF-Y and other transcription factors that act synergistically for transcription activation. Moreover, these structures represent an excellent starting point to challenge the formation of a stable hybrid nucleosome between NF-Y and core histone proteins, and to rationalize the fine molecular details associated to the wide combinatorial association of plant NF-Y subunits.

**Keywords**

Transcription factor; X-ray Crystallography; NF-Y; Histones; DNA-binding.

## 1. Introduction

During the recent years considerable progress has been made towards understanding the structural basis for the fascinating biology of transcriptional regulation. Studies on eukaryotic and bacterial transcription factors (TFs) have shed light in the general architecture and in the DNA recognition mode of both general TFs (GTFs) and sequence-specific TFs, showing a considerable versatility in the formation of multiprotein complexes and in the interaction with DNA.

While GTFs bind to core promoters close to the transcription start site (TSS) and assemble in an ordered fashion with RNA polymerase to form a functional pre-initiation complex, TFs recognize and bind to short regulatory DNA sequences in promoters and enhancers [1,2]. Typically they contain a sequence-specific DNA-binding domain and a separate trans-activating region, which interacts with downstream factors and coactivators. Regulatory elements contain multiple binding sites for different TFs, which assemble in a unique combination and with a defined three-dimensional architecture. The DNA sequence specificity and the combinatorial cooperation of TFs delineate a complex transcriptional regulatory code, which is still very far from a complete elucidation. A lot of efforts are ongoing to build up comprehensive databases of TFs binding profiles [2-5]. In this perspective, the structural characterization of TFs in complex with their DNA cognate sequences at the atomic level represents a potent tool to feature new DNA-binding domains [6] and to identify protein interfaces likely involved in cooperative interactions with other TFs [7-9].

In this review we focus our attention on the recent progress made towards understanding the structural basis for the biology of the sequence-specific TF Nuclear Factor-Y (NF-Y). The knowledge of genomic binding of NF-Y is derived by ENCODE [10-12]: 25% of NF-Y sites are in promoters and a comparable number are located at tissue specific enhancers. NF-Y specifically recognizes the CCAAT box, a regulatory element typically located at a conserved distance of -60/-100 bp from the TSS and present in 30% of eukaryotic promoters [13-15]. This occurrence is similar to that of the TATA box [16], and the CCAAT box is typically found in TATA-less

promoters [15]. A multitude of genes has been described to be positively or negatively regulated by NF-Y, including prosurvival and cell-cycle-promoting genes as well as genes involved in metabolism [17-22]. Indeed, knockout of NF-Y is lethal during early embryonic development [23]. In general, NF-Y cooperates with neighboring TFs, including growth-controlling and oncogenic ones [11], consistent with the enrichment of CCAAT motifs in the promoters of genes overexpressed in cancer [24].

NF-Y is a protein complex minimally composed of three subunits: NF-YA, NF-YB, and NF-YC. All subunits are conserved throughout evolution and required for DNA-binding [25]. NF-YB and NF-YC contain a core region which belongs to the class of Histone Fold Domain (HFD) proteins (Figure 1). This module mediates the formation of a tight heterodimer between the NF-YB and NF-YC subunits, and is involved in non sequence-specific DNA-binding. Heterodimerization of NF-YB and NF-YC results in the formation of a surface for NF-YA association, allowing the resulting trimer to bind DNA with high affinity and specificity. The NF-YA core domain is less than 60 amino acids long and divided into two segments: an N-terminal region responsible for NF-YC/NF-YB binding, and a C-terminal region implicated in specific recognition of the CCAAT element (Figure 1) [26-30]. In addition to these highly conserved core domains involved in trimerization, both NF-YA and NF-YC subunits display much less conserved flanking regions, which include a large Glutamine-rich (Q-rich) domain with transcriptional activation potential (Figure 1) [31-32].

In yeast the three NF-Y subunits (called HAP2, HAP3 and HAP5, respectively) have no Q-rich regions and the activation function is encoded by a fourth subunit (HAP4), present also in fungi, with no apparent homologues in other species [33,34].

Here we review the structural determinants describing the NF-Y subunits, the three-dimensional architecture of their complex, and the molecular details of the DNA recognition and binding at the CCAAT box as derived from the available NF-Y crystallographic data from mammal,

*Aspergillus nidulans* (where NF-Y is called CBC and the three subunits HapB, HapC and HapE, respectively) and *Arabidopsis thaliana* [35-38].

## 2. The NF-YB/NF-YC HFD dimer and its trimerization with NF-YA

Both NF-YB and NF-YC host a HFD, as in core histones H2A, H2B, H3, and H4 [39], and in several other proteins involved in transcription, chromatin remodeling and DNA repair [35,40-43]. Typically, the HFD is formed by a minimum of 3 helices (α1, α2 and α3) separated by two loops (L1 and L2), with helices α1 and α3 flanking almost orthogonally to α2 (Figure 2A).

The HFDs of NF-YB and NF-YC are homologous in sequence and structure to H2B and H2A, respectively, and comparison between their HFD regions reveals relatively little differences (Figure 2B). This similarity extends also outside the HFD and includes the presence of conserved additional secondary structure elements at the HFD C-termini: in NF-YB an extra αC helix is found similar to that of H2B, while in NF-YC a loop-short helix-loop motif is present, reminiscent of the short αC helix found in H2A. The N-terminal regions of the HFD modules instead differ markedly, with core histone H2A and H2B hosting tails not present in NF-YC and NF-YB (Figure 2B). Interestingly, the *Aspergillus nidulans* HapE subunit (corresponding to NF-YC) comprises also an N-terminal extension, termed αN, but adopts a different orientation relative to that present in core-histones [37]. Other structural features of HFDs that have been recognized to be specific only for NF-Y subunits are: (i) the presence of an intra-chain Arg-Asp bidentate pair linking L2 to α3 and (ii) the presence in NF-YC of an absolutely conserved Trp, at the end of helix α2, sandwiched between loop L2 of NF-YC and loop L1 of NF-YB [35,36].

At the level of quaternary structure, the HFD modules of NF-YC and NF-YB associate in a head-to-tail fashion similar to H2A/H2B within the nucleosome (Figure 2C,D) [39]. This antiparallel "handshake" assembly juxtaposes the L1 loop of NF-YB and the L2 loop of NF-YC and *vice versa*, thus generating a twofold quasi-symmetry axis between the polypeptide chains (Figure 2C). Extensive interactions between these loop regions and the hydrophobic packing of residues

belonging to the long α2 helices of NF-YB and NF-YC contribute to the exceptional stability of this heterodimeric complex. The HFD modules are responsible not only for the establishment of contacts between the two NF-YB and NF-YC subunits, but also for building the molecular platform needed for binding and bending the DNA. Like H2A/H2B, the calculated electrostatic potential of the upper surface of the HFD portion of the NF-YB/NF-YC heterodimer is highly basic, and allows favorable polar and van der Waals interactions with the negatively charged phosphodiester backbone of the DNA (Figure 3). Such interactions are apparently devoid of DNA sequence specificity. About four turns of double-stranded DNA sit on the NF-YB/NF-YC platforms, with a bending angle of about 80° compared to the ideal B-DNA. This curvature is remarkably similar to the DNA bending in nucleosomes (Figure 3A,B) and highlights the evolutionary relationship of NF-YB/NF-YC to core histone proteins [44]. The α1-α1 region and the two L1-L2 loops build the central part and the two sides of the DNA contact regions (Figure 2E) [36,37]. Notably, the α1-α1/DNA contacts involve structurally equivalent residues in NF-Y and in the nucleosome H2A/H2B, and engineering experiments indicated that the integrity of α1 is essential for DNA-binding [45].

The structure and the DNA-binding mode of the NF-YB/NF-YC dimer are also highly reminiscent of that of other HFD-containing proteins, as in the case of negative cofactor 2 (NC2α/β) (Figure 3C), a protein that represses TATA box-dependent transcription, while increasing the activity of the distal promoter element [46,47]. Overall, despite limited insight available for the NC2α/β DNA interactions due to the short length of the bound DNA [40], DNA contacts involving NC2α at α1, L1, and L2, and NC2β at L2, are essentially equivalent in NF-YC and NF-YB, respectively (Figure 3A,C). Furthermore, other HFD-containing proteins, such as the Chrac14/Chrac16 HFD heterodimer [42] and the TAF12/TAF4 subunits of TFIID, resemble the geometry of NF-YB/NF-YC with some variations in the conservation of side-chain putatively involved in the DNA interaction [36,48,49].

What makes the NF-YB/NF-YC HFD dimer markedly different relative to other HFD-containing proteins is the presence of a wide negatively charged surface groove, built mostly by residues belonging to the NF-YC αC, NF-YC α1 and NF-YB α2, responsible for binding of the NF-YA subunit (Figure 3A). Heterotrimerization occurs mainly through interaction of the A1 helix of NF-YA with the NF-YB/NF-YC HFD dimer (Figures 2E and 3A). The NF-YA A1 helix contains several polar and mostly positively charged residues which interact via hydrogen bonds and salt bridges with residues belonging to the negatively charged surface groove of the NF-YB/NF-YC HFD dimer (Figures 3A). These contacts are extensive (contact interface of about 1,760 $Å^2$) and most of them involve residues proven to be crucial for heterotrimerization in mutational studies [27-30,35,50,51]. Furthermore, these interactions are protein selective since structural differences between the NF-YB/NF-YC and H2B/H2A are present in this region and are coupled to poor conservation in H2B/H2A of negatively charged residues important for NF-Y trimerization (Figure 3A,B). Similarly, sequence variations in the corresponding HFD region of NC2 (Figure 3A,C) was demonstrated to prevent NF-YA binding to the NC2α/β HFD module [35,45].

Thus, evolutionary adaptations of the HFD regions endowed NF-YB and NF-YC with novel functions compared to other histone proteins. Besides NF-YA recognition and binding, the αC helices of NF-YB and NF-YC have been shown to be important for association with the TATA binding protein (TBP) [52,53] and, therefore, to act as a scaffold for the assembly of the transcriptional preinitiation complex. Because αC helices are not directly engaged in DNA contacts (Figure 2C,E), protein-protein interactions with other TFs are possible in the presence and absence of DNA [14]. Moreover, the αC helix of NF-YC interacts with cell cycle-controlling proteins such as the proto-oncogene c-myc [54] and the tumor suppressor p53 [55]**; such interactions might suggest that they contribute to the essentiality of NF-Y in eukaryotes.**

All these structural and biochemical data on NF-Y support the general idea of HFDs as protein-protein interaction modules able to bind DNA in a core histone-like mode and to favor the binding of other TFs. However, this vision is probably reductive. Indeed, NF-YB has been shown to be

ubiquitinated at residue K138 of αC, which corresponds to K120 in H2B, and that this post-translational modification (PTM) is associated to transcriptional activation [36]. NF-YB ubiquitination is important for ubiquitination of H2B at residue K120 [25], which is genetically and biochemically upstream of important activating histone methylations, such as H3K4me3 and H3K79me2/3 [56]. Considering that core histones display PTMs that are not only localized in tails, but also within the HFD [57], and that several of these modified residues are conserved in their nature and locations in the HFD subunits of NF-Y and in other TFs, it is tempting to conclude that the PTMs of HFD-containing proteins might provide a further layer of potential epigenetic control [43].

**3. CCAAT recognition and binding**

The crystal structure of the NF-Y trimer bound to DNA reveal the strategy employed by the NF-YA subunit for specific DNA sequence recognition at the CCAAT box [36,37]. While the NF-YA helix A1 mediates trimerization with the HFD heterodimer, helix A2 and the following Gly-X-Gly-Gly-Arg-Phe loop motif (Gly-loop; X=any residue) provide sequence-specific contacts to the CCAAT box by inserting deeply into the DNA minor groove, resulting in a striking minor groove widening with a maximum of about 19 Å at the first CCAAT-box adenine (Figures 2E,F and 3A). The adjacent major groove regions are not affected by the NF-YA binding and, therefore, available for the potential binding of other TFs. This observation is in keeping with the ability of NF-Y to synergize with several TFs, most of which functionally recognize bases within the major groove, at a conserved distance from CCAAT [15,25].

The Gly-loop, located after the A2 helix (Figure 2F), displays a kinked backbone which allows close proximity of the two Gly-Gly carbonyl O atoms to the bases of the CCAAT complementary DNA strand (GG) (Figure 4A). Bases on the CCAAT DNA strand (AAT) are instead hydrogen-bonded to side-chains of Arg and His residues belonging to the NF-YA A2 helix (Figure 4B). In addition, the NF-YA/DNA interactions include several stabilizing contacts to the DNA phosphate

backbone and the minor groove insertion of the Phe residue belonging to the Gly-loop between the AT/CG base pairs at the CCAAT box (Figure 4A). Interestingly, this Phe positioning into the minor groove is reminiscent of that found in the TBP-DNA complex, where two Phe residues insert into the first and last base pairs of the targeted TATA sequence [58].

All NF-YA residues involved in sequence-specific recognition are strictly conserved in agreement with earlier mutagenesis data and with the matrix of DNA specificity [36]. Hence, a combination of proper main-chain conformation and side-chain distribution makes the NF-YA A2 helix and the Gly-loop able to recognize and bind the CCAAT box by selecting the correct sequence of pyrimidine or purine bases that favor hydrogen bond interactions and disfavor steric hindrance, thus providing the structural explanation for the reduced or abolished affinity of NF-Y to mutant CCAAT motifs [59]. It is likely that the NF-YA A2 helix slides along the DNA, thereby acting as a sequence sensor, until appropriate interactions with the DNA bases at the minor groove are provided, with the Gly-loop acting as an anchor to allow for high-affinity binding of NF-Y to the CCAAT box [36,37]. This is also consistent with the fact that in the absence of bound DNA the NF-YA A2 helix is structurally disordered [37].

The conformational flexibility required to direct the NF-YA A2 helix toward the DNA, while keeping the A1 helix stably linked to the NF-YB/NF-YC interface, is provided by the A1A2-linker (Figure 2F). Interestingly, the comparison of the crystal structures of mammalian and *Aspergillus* NF-Y/DNA complex suggests some indications of the sequence-structure relationship that drives the conformation of the A1A2-linker. Indeed, when the two structures are superimposed the trimerization and the DNA binding modes are perfectly conserved, with the only significant structural mismatch being localized at the C-terminus of the A1 helix and at the A1A2-linker on the NF-YA subunit (Figure 5). The A1A2-linker is a region of relative divergence across kingdoms and paralogs, and in the two structures it follows different pathways due to specific amino acid composition. In mammalian NF-YA two residues, Gly260 and Pro263, are localized at the beginning of the linker and guide its structure by providing flexibility and directionality, respectively (Figure 5A,C). These two

residues are absent in the *Aspergillus* HapB subunit (corresponding to NF-YA), where instead Pro267, located at the end of the linker, contributes to orient the A2 helix towards the DNA minor groove (Figure 5B,C). These sequence differences at the A1A2-linker are coupled with variations at the N-terminal region of the NF-YB subunits (i.e. Ile55 and Arg46 in mammalian NF-YB and *Aspergillus* HapC, respectively), which faces the A1A2-linker. As a result, in mammalian NF-YA and *Aspergillus* HapB the two A1A2-linkers have a certain degree of structure variability but, nevertheless, the following A2 helices are in register to each other and correctly inserted in the DNA minor groove at the CCAAT box (Figure5).

**4. The "hybrid nucleosome" hypothesis**

The structure of the NF-Y/DNA complex bears some intriguing implications for the description of the possible mechanistic bases of NF-Y action *in vivo*. The CCAAT motifs are at the top of the list of the regulatory sequences that are spared by nucleosomes [60], and NF-Y has been reported to interact directly with histone H3/H4, but not with H2A/H2B, during nucleosome reconstitution *in vitro* [61]. The NF-Y/DNA structural data offer an explanation for these observations.

In nucleosomes, two left-handed superhelical turns of bent DNA are wrapped around an octameric-histone core consisting of two copies each of the core histone proteins H2A, H2B, H3, and H4. At the heart of the octamer there are two H3/H4 heterodimers related by a twofold axis of symmetry, forming a stable $(H3/H4)_2$ heterotetramer, with two H2A/H2B heterodimers bound to opposite faces of $(H3/H4)_2$ [39]. When the NF-Y heterotrimer structure is superimposed on H2B/H2A within the nucleosome, the modeled NF-YB/NF-YC substituted almost ideally H2B/H2A, with the core interaction with H3/H4 relying mostly on NF-YB (Figure 6A). Considering that NF-YB Asp115 residue is structurally homologous to H2B Glu90 (salt-bridged to H4 His75 in the nucleosome) [39], it is tempting to propose that similar interactions can stabilize a hybrid structure. Binding of one NF-Y trimer to the $(H3/H4)_2$ heterotetramer for hybrid nucleosome formation would require only minor reorientations of the NF-YA A1 helix (and the following

A1A2-linker) relative to the H4 α2 and α3 helices (Figure 6B). Note that acetylation of H4 Lys91 in the H4 α3 helix is believed to affect nucleosome stability [62] suggesting an intrinsic structural adaptability of the H4 α3 region. On the other hand, two NF-Y trimers cannot coexist in the hybrid nucleosome, since the A1 helix of one NF-YA subunit would overlap with the corresponding helix of the other NF-YA subunit and partially with NF-YB. This is also in keeping with the fact that in this model the CCAAT box is located after about 20 bp from the DNA start, and the presence of two CCAAT boxes in the hybrid nucleosome would position them at 100bp distance, which is not the distance usually found in multiple-CCAAT-containing DNA [63,64].

Thus, the available structural data suggest a model whereby NF-Y would prevent nucleosome formation through association of the NF-YB/NF-YC heterodimer to a DNA-bound $(H3/H4)_2$ heterotetramer, the NF-YA A1A2-linker providing enough flexibility for the A2 helix to search for and bind to a CCAAT-box. Interestingly the putative formation of a hybrid nucleosome was also suggested by the analysis of the *Aspergillus* CBC structure, where it was shown that interaction of the CBC with the H3/H4 pair that is part of the first nucleosome turn (with the consequent formation of a tetrasome) is possible [37]. Furthermore, the NF-Y trimer in complex with a CCAAT-containing 25bp DNA [36] has been reported to fit within the nucleosome core particle bound to the DNA sequence of nucleosome A of the 3'-LTR of the mouse mammary tumor virus (MMTV-A), with a good match not only between the NF-YB/NF-YC and H2B/H2A dimers, but also in the DNA conformation in the region corresponding to the CCAAT box [65].

In this scenario, hybrid NF-Y/H3/H4 assemblies would locally halt nucleosome formation, while providing opportunities for other TFs to associate with the neighboring target sequences. Indeed, genome-wide studies have shown that the presence of Nucleosome Free Regions in active or poised core promoters is a widespread phenomenon [66]. The location of CCAAT at about -80 bp, relative to the TSS, matches the predicted positioning of H2A/H2B in the "average" core promoter nucleosome [60].

## 5. Structural features of plant NF-Y

In the plant lineage NF-Y also consists of three subunits. However, each subunit can be encoded not by one but by a family of genes (typically about 10), both in dicots and monocots, differentially expressed in various tissues. As a consequence, different subunit combination can lead to a wide variety of NF-Y trimeric complexes, suited to face the many environmental conditions that a plant can experience [67,68].

Recently the first crystal structure of a plant NF-Y, the *Arabidopsis thaliana* NF-YB6/NF-YC3 dimer, has been **solved** [38]. This structure is of particular interest because within the NF-YB genes, the *Arabidopsis* NF-YB6 and NF-YB9 form a conserved subfamily originally identified in genetic experiments for their key roles in embryo maturation [69,70]. NF-YB9, also known as LEAFY COTYLEDON 1 (LEC1), and NF-YB6 or LEC1-LIKE (L1L) are embryogenesis regulators acting during transition from embryo to adult status: *lec1* mutants display pleiotropic phenotypes [71-76] and L1L (NF-YB6) was shown to be able to partially complement the *lec1* defect [70]. Importantly, chimeric constructs have demonstrated that the LEC1 function in embryos can be associated specifically to the HFD region of the NF-YB subunit [69]. Besides a general interest related to their functional action, much interest on LEC1 and L1L is also associated to their agronomic potential, since overexpression of LEC1 or L1L in various species have been reported to result in significant changes in seed lipids/oils production [75,77,78].

The crystal structure of the *Arabidopsis thaliana* L1L/NF-YC3 HFD dimer reveals the typical features of a "classical" NF-YB/NF-YC dimer with some structural specificities of the LEC1 family. In particular, the *Arabidopsis* L1L and NF-YC3 subunits interact in a head-to-tail fashion, forming a classical histone-like pair (Figure 7A), with conserved positive and negative electrostatic distributions on the surfaces predicted to be involved in DNA contacts and NF-YA interaction, respectively (Figure 7B). Indeed, the L1L/NF-YC3 dimer was demonstrated to trimerize with the *Arabidopsis* NF-YA6 subunit and the trimer to bind DNA in EMSA experiments on a high affinity CCAAT box probe derived from the human HSP70 promoter [38]. This result is important since it

shows that a plant NF-Y trimer containing a L1L subunit is able to bind DNA directly. In fact, the LEC1/L1L capacity to bind DNA was previously questioned because of the LEC1-specific presence of an Asp residue in helix α2 (LEC1 Asp55 and L1L Asp84) where in "normal" NF-YB a Lys/Arg is present and contacts the DNA phosphate backbone, 2 bps upstream of CCAAT. Genetic experiments pinpointed Asp55 as crucial for LEC1 function *in vivo*: its mutation to Lys led to loss of LEC1 activity, while an Asp substitution in a canonical *Arabidopsis* NF-YB was sufficient to confer partial LEC1 behavior [69]. Superimposition of the mammalian NF-Y trimer in complex with DNA on the *Arabidopsis* L1L/NF-YC3 dimer suggests that the electrostatic repulsion between the negatively charged Asp84 and the DNA phosphate backbone would favor a slightly shifted DNA trajectory, stabilized on the opposite side of the double helix by His79, which has been proposed to be diagnostic for LEC1 family members (Figure 7C) [38].

Another interesting structural issue associated to plant NF-Y is related to the discovery of the involvement of plant NF-Y genes in the control of photoperiod-dependent flowering time [79,80]. In particular, members of the large NF-YB/NF-YC plant families have been shown to trimerize with COSTANS (CO), a key regulator of photoperiod-induced flowering time, *via* its conserved CCT (for CO, CO-like, and TOC1) domain [79,80]. This interaction is physiologically relevant since mutants in *Arabidopsis NF-YB/NF-YC* phenocopy the *co* mutants in terms of delayed flowering-time phenotype [79,81-83]. Furthermore, the overexpression of the *Arabidopsis* NF-YA1 subunit causes late flowering, thus suggesting the intriguing possibility that CO and NF-YA compete for the interaction with plant NF-YB/NF-YC dimers [84].

Sequence comparison of the DNA-binding region of NF-YA and CO indicate a marked similarity at the A2 helix, but a shorter A1A2-linker and some sequence variation in the following Gly-loop, with the absence of the first two Gly residues (Figure 8). Furthermore, the CO region corresponding to NF-YA A1 helix is rich in basic residues, as required for the heterotrimeric assembly formation with the NF-YB/NF-YC HFD dimer. Indeed, mutations reported to affect CO activity *in vivo* [80] correspond to amino acids essential for DNA-binding by NF-YA (Figure 8).

Based on this sequence/structure analysis it has been suggested that CO might have evolved to recognize slight variations of CCAAT, with a different 5' end, while using NF-YB/NF-YC as a DNA binding platform [36].

## 6. Conclusions

X-ray structures of NF-Y have provided crucial insights into the molecular mechanism responsible for the recognition and binding to DNA and into the architecture of the NF-YB/NF-YC HFD dimer and of the NF-Y trimer. These structures serve as a powerful guide for present and future biochemical experiments aimed at understanding the contributions of individual amino acid residues to the stability of the macromolecular assembly, to the sequence-specific DNA recognition, and to the, so far poorly understood, role of post-translational modification at regulation sites. Important objectives for the future will include the use of the whole arsenal of structural biology techniques, including cryo-electron microscopy, X-ray crystallography and small angle X-ray scattering (SAXS), in combination with biochemical and genetic analysis, to unravel the three-dimensional architecture of higher order complexes formed between NF-Ys bound at multiple CCAAT boxes, between NF-Y and core histone proteins to challenge the formation of a stable hybrid nucleosome, and between NF-Y and other TFs demonstrated to act synergistically for transcription activation.

As basic studies of transcription continue to provide insight into the molecular basis of human disease, one of the challenges for the future will be to exploit structural insights of TFs for the development of novel therapeutics. In this context, NF-Y may be seen as a target for cancer progression drugs. The available genomic data of NF-Y locations are consistent with the recurrence of CCAAT motifs in promoters of genes overexpressed in cancer, and indicate that NF-Y is a pioneer TF for oncogenic activators, which acts either synergistically with adjacent DNA-binding complexes or by "tethering" of individual oncogenic partners. Interfering with such pioneer action of NF-Y could bear key implications for cancer control. The chase for an anti-proliferative drug that

could act by displacing the NF-Y/CCAAT complex has already started, mostly focusing on minor-groove binding drugs able to block interactions of NF-Y with the promoter of topoisomerase IIα, thus blocking cell cycle progression without involving activation of p53 [85,86]. The availability of the crystal structure of NF-Y in complex with its CCAAT-containing DNA target completely changes the perspective in the field, providing the unique possibility to target NF-Y, and not the DNA, for drug binding.

We can also anticipate that the research field on plant NF-Y will have a dramatic development not only in the genetics area but also in structural biology. We expect that several new structures of NF-Y and, possibly, CCT-containing proteins in complex with HFD dimers and DNA will be solved in the near future. These data will provide the means for the rationalization and the full understanding of the fine molecular details associated to the wide combinatorial association of plant NF-Y subunits. For instance, they are expected to clarify whether sequence changes associated to specific phenotypes are mirrored by structural rearrangements that can affect the trimer formation and the DNA-binding affinity/specificity in plant NF-Y.

**Acknowledgements**

This work was supported by the AIRC Investigator Grant - IG 2014 to M.N. [grant number IG-15267].

**References**

[1] M. C. Thomas, C.-M. Chiang, The general transcription machinery and general cofactors, Crit. Rev. Biochem. Mol. Biol. 41 (2006) 105–178.

[2] A. Jolma, J. Yan, T. Whitington, J. Toivonen, K.R. Nitta, P. Rastas, E. Morgunova, M. Enge, M. Taipale, G. Wei, K. Palin, J.M. Vaquerizas, R. Vincentelli, N.M. Luscombe, T.R. Hughes, P. Lemaire, E. Ukkonen, T. Kivioja, J. Taipale, DNA-binding specificities of human transcription factors, Cell 152 (2013) 327–339.

[3] A. Mathelier, O. Fornes, D.J. Arenillas, C.Y. Chen, G. Denay, J. Lee, W. Shi, C. Shyr, G. Tan, R. Worsley-Hunt, A.W. Zhang, F. Parcy, B. Lenhard, A. Sandelin, W.W. Wasserman, JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles, Nucleic Acids Res. 44 (2016) D110–115.

[4] I.V. Kulakovskiy, et al, HOCOMOCO: a comprehensive collection of human transcription factor binding sites models, Nucleic Acids Res. 41 (2013) 195–202.

[5] M.T. Weirauch, A. Cote, R. Norel, M. Annala, Y. Zhao, T.R. Riley, J. Saez-Rodriguez, T. Cokelaer, A. Vedenko, S. Talukder; DREAM5 Consortium, H.J. Bussemaker, Q.D. Morris, M.L. Bulyk, G. Stolovitzky, T.R. Hughes, Evaluation of methods for modeling transcription factor sequence specificity, Nat. Biotechnol. 31 (2013) 126–134.

[6] B. Xu, D.E. Schones, Y. Wang, H. Liang, G. Li, A structural-based strategy for recognition of transcription factor binding sites, PLoS One 8 (2013) e52460.

[7] A.M. Jaeger, C.W. Pemble, L. Sistonen, D.J. Thiele, Structures of HSF2 reveal mechanisms for differential regulation of human heat-shock factors, Nat. Struct. Mol. Biol. 23 (2016) 147–154.

[8] D. Wu, N. Potluri, J. Lu, Y. Kim, F. Rastinejad, Structural integration in hypoxia-inducible factors, Nature 524 (2015) 303–308.

[9] A. Jolma, Y. Yin, K.R. Nitta, K. Dave, A. Popov, M. Taipale, M. Enge, T. Kivioja, E. Morgunova, J. Taipale, DNA-dependent formation of transcription factor pairs alters their binding specificity, Nature 527 (2015) 384–388.

[10] ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome, Nature 489 (2012) 57–74.

[11] J. Wang, J. Zhuang, S. Iyer, X. Lin, T.W. Whitfield, M.C. Greven, B.G. Pierce, X. Dong, A. Kundaje, Y. Cheng, O.J. Rando, E. Birney, R.M. Myers, W.S. Noble, M. Snyder, Z. Weng, Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors, Genome Res. 22 (2012) 1798–1812.

[12] J.D. Fleming, G. Pavesi, P. Benatti, C. Imbriano, R. Mantovani, K. Struhl, NF-Y coassociates with FOS at promoters, enhancers, repetitive elements, and inactive chromatin regions, and is stereo-positioned with growth-controlling transcription factors, Genome Res. 23 (2013) 1195–1209.

[13] S.N. Maity, B. de Crombrugghe, Role of the CCAAT-binding protein CBF/NF-Y in transcription, Trends Biochem. Sci. 23 (1998) 174–178.

[14] R. Mantovani, The molecular biology of the CCAAT-binding factor NF-Y, Gene 239 (1999) 15–27.

[15] D. Dolfini, F. Zambelli, G. Pavesi, R. Mantovani, A perspective of promoter architecture from the CCAAT box, Cell Cycle 8 (2009) 4127–4137.

[16] C. Yang, E. Bolotin, T. Jiang, F.M. Sladek, E. Martinez, Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters, Gene 389 (2007) 52–65.

[17] K.T. Bergh, O. Litzka, A.A. Brakhage, Identification of a major cis-acting DNA element controlling the bidirectionally transcribed penicillin biosynthesis genes acvA (pcbAB) and ipnA (pcbC) of *Aspergillus nidulans*, J. Bacteriol. 178 (1996) 3908–3916.

[18] M. Ceribelli, D. Dolfini, D. Merico, R. Gatta, A.M. Viganò, G. Pavesi, R. Mantovani, The histone-like NF-Y is a bifunctional transcription factor, Mol. Cell. Biol. 28 (2008) 2047–2058.

[19] H. Deng, Y. Sun, Y. Zhang, X. Luo, W. Hou, L. Yan, Y. Chen, E. Tian, J. Han, H. Zhang, Transcription factor NFY globally represses the expression of the *C. elegans* Hox gene Abdominal-B homolog egl-5, Dev. Biol. 308 (2007) 583–592.

[20] T.G. Littlejohn, M.J. Hynes, Analysis of the site of action of the amdR product for regulation of the amdS gene of *Aspergillus nidulans*, Mol. Gen. Genet. 235 (1992) 81–88.

[21] O. Litzka, K. Then Bergh, A.A. Brakhage, The *Aspergillus nidulans* penicillin-biosynthesis gene aat (penDE) is controlled by a CCAAT-containing DNA element, Eur. J. Biochem. 238 (1996) 675–682.

[22] S. Steidl, M.J. Hynes, A.A. Brakhage, The *Aspergillus nidulans* multimeric CCAAT binding complex AnCF is negatively autoregulated via its hapB subunit gene, J. Mol. Biol. 306 (2001) 643–653.

[23] A. Bhattacharya, J.M. Deng, Z. Zhang, R. Behringer, B. de Crombrugghe, S.N. Maity, The B subunit of the CCAAT box binding transcription factor complex (CBF/NF-Y) is essential for early mouse development and cell proliferation, Cancer Res. 63 (2003) 8167–8172.

[24] H. Goodarzi, O. Elemento, S. Tavazoie, Revealing global regulatory perturbations across human cancers, Mol. Cell 36 (2009) 900–911.

[25] D. Dolfini, R. Gatta, R. Mantovani, NF-Y and the transcriptional activation of CCAAT promoters, Crit. Rev. Biochem. Mol. Biol. 47 (2012) 29–49.

[26] J.T. Olesen, L. Guarente, The HAP2 subunit of yeast CCAAT transcriptional activator contains adjacent domains for subunit association and DNA recognition: model for the HAP2/3/4 complex, Genes Dev. 4 (1990) 1714–1729.

[27] S.N. Maity, B. de Crombrugghe, Biochemical analysis of the B subunit of the heteromeric CCAAT-binding factor. A DNA-binding domain and a subunit interaction domain are specified by two separate segments, J. Biol. Chem. 267 (1992) 8286–8292.

[28] Y. Xing, J.D. Fikes, L. Guarente, Mutations in yeast HAP2/HAP3 define a hybrid CCAAT box binding domain, EMBO J. 12 (1993) 4647–4655.

[29] Y. Xing, S. Zhang, J.T. Olesen, A. Rich, L. Guarente, Subunit interaction in the CCAAT-binding heteromeric complex is mediated by a very short alpha-helix in HAP2, Proc. Natl. Acad. Sci. U.S.A. 91 (1994) 3009–3013.

[30] R. Mantovani, X.Y. Li, U. Pessara, V.H. Hooft, C. Benoist, D. Mathis, Dominant negative analogs of NF-YA, J. Biol. Chem. 269 (1994) 20340–20346.

[31] F. Coustry, S.N. Maity, S. Sinha, B. de Crombrugghe, 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 (1996) 14485–14491.

[32] A. de Silvio, C. Imbriano, R. Mantovani, Dissection of the NF-Y transcriptional activation potential, Nucleic Acids Res. 27 (1999) 2578–2584.

[33] S.L. Forsburg, L. Guarente, Identification and characterization of HAP4: a third component of the CCAAT-bound HAP2/HAP3 heteromer, Genes Dev. 3 (1989) 1166–1178.

[34] D.S. McNabb, K.A. Tseng, L. Guarente, 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 (1997) 7008– 7018.

[35] C. Romier, F. Cocchiarella, R. Mantovani, D. Moras, 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 (2003) 1336–1345.

[36] M. Nardini, N. Gnesutta, G. Donati, R. Gatta, C. Forni, A. Fossati, C. Vonrhein, D. Moras, C. Romier, M. Bolognesi, R. Mantovani, Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination, Cell 152 (2013) 132-143.

[37] E.M. Huber, D.H. Scharf, P. Hortschansky, M. Groll, A.A. Brakhage, DNA minor groove sensing and widening by the CCAAT-binding complex, Structure 20 (2012) 1757–1768.

[38] N. Gnesutta, D. Saad, A. Chaves-Sanjuan, R. Mantovani, M. Nardini, Crystal structure of the *Arabidopsis thaliana* L1L/NF-YC3 histone-fold dimer reveals specificities of the LEC1 family of NF-Y subunits in plants, Mol. Plant (2016) Under revision.

[39] K. Luger, A.W. Mäder, R.K. Richmond, D.F. Sargent, T.J. Richmond, Crystal structure of the nucleosome core particle at 2.8 Å resolution, Nature 389 (1997) 251–260.

[40] K. Kamada, F. Shu, H. Chen, S. Malik, G. Stelzer, R.G. Roeder, M. Meisterernst, S.K. Burley, Crystal structure of negative cofactor 2 recognizing the TBP-DNA transcription complex, Cell 106 (2001) 71–81.

[41] E. Cler, G. Papai, P. Schultz, I. Davidson, Recent advances in understanding the structure and function of general transcription factor TFIID, Cell Mol. Life Sci. 66 (2009) 2123–2134.

[42] K.F. Hartlepp, C. Fernández-Tornero, A. Eberharter, T. Grüne, C.W. Müller, P.B. Becker, The histone fold subunits of Drosophila CHRAC facilitate nucleosome sliding through dynamic DNA interactions, Mol. Cell. Biol. 25 (2005) 9886–9896.

[43] N. Gnesutta, M. Nardini, R. Mantovani, The H2A/H2B-like histone-fold domain proteins at the crossroad between chromatin and different DNA metabolisms, Transcription 4 (2013) 114–119.

[44] C. Liberati, A. Ronchi, P. Lievens, S. Ottolenghi, R. Mantovani, NF-Y organizes the gamma-globin CCAAT boxes region, J. Biol. Chem. 273 (1998) 16880–16889.

[45] K. Zemzoumi, M. Frontini, M. Bellorini, R. Mantovani, NF-Y histone fold alpha1 helices help impart CCAAT specificity. J. Mol. Biol. 286 (1999) 327–337.

[46] E. Maldonado, M. Hampsey, D. Reinberg, Repression: targeting the heart of the matter, Cell 99 (1999) 455–458.

[47] P.J. Willy, R. Kobayashi, J.T. Kadonaga, A basal transcription factor that activates or represses transcription, Science 290 (2000) 982–985.

[48] K.J. Wright, M.T. II Marr, R. Tjian, TAF4 nucleates a core subcomplex of TFIID and mediates activated transcription from a TATA-less promoter, Proc. Natl. Acad. Sci. U.S.A. 103 (2006) 12347–12352.

[49] K. Gazit, S. Moshonov, R. Elfakess, M. Sharon, G. Mengus, I. Davidson, R. Dikstein, TAF4/4b x TAF12 displays a unique mode of DNA binding and is required for core promoter function of a subset of genes, J. Biol. Chem. 284 (2009) 26286–26296.

[50] S. Sinha, I.S. Kim, K.Y. Sohn, B. de Crombrugghe, S.N. Maity, 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 (1996) 328–337.

[51] I.S. Kim, S. Sinha, B. de Crombrugghe, S.N. Maity, Determination of functional domains in the C subunit of the CCAAT-binding factor (CBF) necessary for formation of a CBF-DNA complex: CBF-B interacts simultaneously with both the CBF-A and CBF-C subunits to form a heterotrimeric CBF molecule, Mol. Cell. Biol. 16 (1996) 4003–4013.

[52] M. Bellorini, D.K. Lee, J.C. Dantonel, K. Zemzoumi, R.G. Roeder, L. Tora, R. Mantovani, CCAAT binding NF-Y-TBP interactions: NF-YB and NF-YC require short domains adjacent to their histone fold motifs for association with TBP basic residues, Nucleic Acids Res. 25 (1997) 2174–2181.

[53] K. Matuoka, K. Yu Chen, Nuclear factor Y (NF-Y) and cellular senescence, Exp. Cell Res. 253 (1999) 365–371.

[54] H. Izumi, C. Molander, L.Z. Penn, A. Ishisaki, K. Kohno, K. Funa, Mechanism for the transcriptional repression by c-Myc on PDGF betareceptor, J. Cell Sci. 114 (2001) 1533–1544.

[55] C. Imbriano, N. Gnesutta, R. Mantovani, The NF-Y/p53 liaison: well beyond repression, Biochim. Biophys. Acta 1825 (2012) 131-139.

[56] T. Suganuma, J.L. Workman, Signals and combinatorial functions of histone modifications, Annu. Rev. Biochem. 80 (2011) 473–499.

[57] M.S. Cosgrove, J.D. Boeke, C. Wolberger, Regulated nucleosome mobility and the histone code, Nat. Struct. Mol. Biol. 11 (2004) 1037–1043.

[58] J.L. Kim, D.B. Nikolov, S.K. Burley, Co-crystal structure of TBP recognizing the minor groove of a TATA element, Nature 365 (1993) 520–527.

[59] W. Bi, L. Wu, F. Coustry, B. de Crombrugghe, S.N. Maity, DNA binding specificity of the CCAAT-binding factor CBF/NF-Y, J. Biol. Chem. 272 (1997) 26562–26572.

[60] E. Segal, Y. Fondufe-Mittendorf, L. Chen, A. Thåström, Y. Field, I.K. Moore, J.P. Wang, J. Widom, A genomic code for nucleosome positioning, Nature 442 (2006) 772–778.

[61] G. Caretti, M.C. Motta, R. Mantovani, NF-Y associates with H3-H4 tetramers and octamers by multiple mechanisms, Mol. Cell. Biol. 19 (1999) 8591–8603.

[62] J. Ye, X. Ai, E.E. Eugeni, L. Zhang, L.R. Carpenter, M.A. Jelinek, M.A. Freitas, M.R. Parthun, Histone H4 lysine 91 acetylation a core domain modification associated with chromatin assembly, Mol. Cell 18 (2005) 123–130.

[63] C. Liberati, A. di Savio, S. Ottolenghi, R. Mantovani R, NF-Y binding to twin CCAAT boxes: role of Q-rich domains and histone fold helices, J. Mol. Biol. 285 (1999) 1441–1455.

[64] D. Dolfini, F. Zambelli, M. Pedrazzoli, R. Mantovani, G. Pavesi, A high definition look at the NF-Y regulome reveals genome-wide associations with selected transcription factors, Nucleic Acids Res. 44 (2016) 4684–4702.

[65] T.D. Frouws, S.C. Duda, T.J. Richmond, X-ray structure of the MMTV-A nucleosome core, Proc. Natl. Acad. Sci. U.S.A. 113 (2016) 1214–1219.

[66] C. Jiang, B.F. Pugh, Nucleosome positioning and gene regulation: advances through genomics, Nat. Rev. Genet. 10 (2009) 161–172.

[67] T. Laloum, S. De Mita, P. Gamas, M. Baudin, A. Niebel, CCAAT-box binding transcription factors in plants: Y so many? Trends Plant Sci. 18 (2013) 157–166.

[68] K. Petroni, R.W. Kumimoto, N. Gnesutta, V. Calvenzani, M. Fornari, C. Tonelli, B.F. 3rd Holt, R. Mantovani, The promiscuous life of plant NUCLEAR FACTOR Y transcription factors, Plant Cell. 24 (2012) 4777–4792.

[69] H. Lee, R.L. Fischer, R.B. Goldberg, J.J. Harada, *Arabidopsis* LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor, Proc. Natl. Acad. Sci. U.S.A. 100 (2003) 2152–2156.

[70] R.W. Kwong, A.Q. Bui, H. Lee, L.W. Kwong, R.L. Fischer, R.B. Goldberg, J.J. Harada, LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development, Plant Cell 15 (2003) 5–18.

[71] S. Zhang, L. Wong, L. Meng, P.G. Lemaux, Similarity of expression patterns of knotted1 and ZmLEC1 during somatic and zygotic embryogenesis in maize (Zea mays L.), Planta 215 (2002) 191–194.

[72] K. Yazawa, K. Takahata, H. Kamada, Isolation of the gene encoding Carrot leafy cotyledon1 and expression analysis during somatic and zygotic embryogenesis, Plant Physiol. Biochem. 42 (2004) 215–223.

[73] M. Fambrini, C. Durante, G. Cionini, C. Geri, L. Giorgetti, V. Michelotti, M. Salvini, C. Pugliesi, Characterization of LEAFY COTYLEDON1-LIKE gene in Helianthus annuus and its relationship with zygotic and somatic embryogenesis, Dev. Genes Evol. 216 (2006) 253–264.

[74] L. Alemanno, M. Devic, N. Niemenak, C. Sanier, J. Guilleminot, M. Rio, J.L. Verdeil, P. Montoro, Characterization of leafy cotyledon1-like during embryogenesis in *Theobroma cacao* L, Planta 227 (2008) 853–866.

[75] J. Mu, H. Tan, Q. Zheng, F. Fu, Y. Liang, J. Zhang, X. Yang, T. Wang, K. Chong, X.J. Wang, J. Zuo, LEAFY COTYLEDON1 is a key regulator of fatty acid biosynthesis in *Arabidopsis*, Plant Physiol. 148 (2008) 1042–1054.

[76] P. Schellenbaum, A. Jacques, P. Maillot, C. Bertsch, F. Mazet, S. Farine, B. Walter, Characterization of VvSERK1, VvSERK2, VvSERK3 and VvL1L genes and their expression during somatic embryogenesis of grapevine (*Vitis vinifera* L.), Plant Cell Rep. 27 (2008) 1799–1809.

[77] B. Shen, W.B. Allen, P. Zheng, C. Li, K. Glassman, J. Ranch, D. Nubel, M.C. Tarczynski, Expression of ZmLEC1 and ZmWRI1 increases seed oil production in maize, Plant Physiol. 153 (2010) 980–987.

[78] H. Tan, X. Yang, F. Zhang, X. Zheng, C. Qu, J. Mu, F. Fu, J. Li, R. Guan, H. Zhang, G. Wang, J. Zuo, Enhanced seed oil production in canola by conditional expression of Brassica napus LEAFY COTYLEDON1 and LEC1-LIKE in developing seeds, Plant Physiol. 156 (2011) 1577–1588.

[79] O. Ben-Naim, R. Eshed, A. Parnis, P. Teper-Bamnolker, A. Shalit, G. Coupland, A. Samach, E. Lifschitz, The CCAAT binding factor can mediate interactions between CONSTANS-like proteins and DNA, Plant J. 46 (2006) 462–476.

[80] S. Wenkel, F. Turck, K. Singer, L. Gissot, J. Le Gourrierec, A. Samach, G. Coupland, CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of *Arabidopsis*, Plant Cell 18 (2006) 2971–2984.

[81] X. Cai, J. Ballif, S. Endo, E. Davis, M. Liang, D. Chen, D. DeWald, J. Kreps, T. Zhu, Y. Wu, A putative CCAAT binding transcription factor is a regulator of flowering timing in *Arabidopsis*, Plant Physiol. 145 (2007) 98–105.

[82] N.Z. Chen, X.Q. Zhang, P.C. Wei, Q.J. Chen, F. Ren, J. Chen, X.C. Wang, AtHAP3b plays a crucial role in the regulation of flowering time in *Arabidopsis* during osmotic stress, J. Biochem. Mol. Biol. 40 (2007) 1083–1089.

[83] R.W. Kumimoto, Y. Zhang, N. Siefers, B.F. III Holt, NF-YC3, NF-YC4 and NF-YC9 are required for CONSTANS-mediated, photoperiod-dependent flowering in *Arabidopsis thaliana*, Plant J. 63 (2010) 379–391.

[84] C. Li, A. Distelfeld, A. Comis, J. Dubcovsky, Wheat flowering repressor VRN2 and promoter CO2 compete for interactions with NUCLEAR FACTOR-Y complexes, Plant J. 67 (2011) 763–773.

[85] M. Kotecha, J. Kluza, G. Wells, C.C. O'Hare, C. Forni, R. Mantovani, P.W. Howard, P. Morris, D.E. Thurston, J.A. Hartley, D. Hochhauser, Inhibition of DNA binding of the NF-Y transcription factor by the pyrrolobenzodiazepine-polyamide conjugate GWL-78, Mol. Cancer Ther. 7 (2008) 1319–1328.

[86] H. Mackay, T. Brown, J.S. Sexton, M. Kotecha, B. Nguyen, W.D. Wilson, J. Kluza, B. Savic, C. O'Hare, D. Hochhauser, M. Lee, J.A. Hartley, Targeting the inverted CCAAT Box-2 of the topoisomerase IIalpha gene: DNA sequence selective recognition by a polyamide-intercalator as a staggered dimer, Bioorg. Med. Chem. 16 (2008) 2093–2102.

**Figure legends**

**Figure 1. Schematic representation of the NF-Y subunits.** Color shaded boxes highlight the evolutionarily conserved NF-YB/NF-YC and DNA binding domain (BC-DNA; green) in NF-YA, H2B-like or H2A-like HFD domains in NF-YB and NF-YC (orange and cyan, respectively), and the Gln-rich domains (Q-rich) in NF-YA and NF-YC (grey). The secondary structure composition of the NF-Y subunits is shown in the enlargement panels.

**Figure 2. Three dimensional structure of NF-Y. (A)** The tertiary structure of mammalian NF-YB (orange) and NF-YC (cyan) are shown as ribbon **[36]**. The secondary structure elements are indicated. **(B)** Structural superimposition of NF-YB and NF-YC (PDB-code 4CSR), illustrated as coils, to core histone H2B (magenta) and H2A (yellow) (PDB-code 1AOI **[39]**, chains C/D, respectively). The N-terminal histone tails (grey) are indicated. **(C)** Ribbon diagram showing the NF-YB/NF-YC HFD dimer, and **(D)** structural superimposition with H2B/H2A. **(E)** Ribbon representations of the NF-YA/NF-YB/NF-YC heterotrimer in complex with the HSP70 25 bp-CCAAT-box oligonucleotide (PDB-code 4AWL **[36]**); **secondary structure elements are labeled and those at the protein-DNA interface are highlighted in red**. NF-YA and the DNA are colored in green and grey, respectively, with CCAAT nucleotides highlighted in violet. **(F)** The tertiary structure of NF-YA. The secondary structure elements, the A1A2-linker and the Gly-loop are indicated.

**Figure 3. Electrostatic surface and DNA binding-mode of HFD dimers.** Electrostatic surface of **(A)** NF-YB/NF-YC (NF-YA and DNA represented in ribbon and stick models color-coded as in Figure 2) **[36]**, **(B)** H2B/H2A (DNA in grey) **[39]**, and **(C)** NC2α/β (PDB-code 1JFI **[40]**; DNA in grey and TBP in brown). Blue and red colors indicate positively and negatively charged regions,

respectively. In panel (A), the widening of the DNA minor groove due to the NF-YA A2 insertion in highlighted.

**Figure 4. Sequence-specific interactions of NF-Y/DNA at the CCAAT-box. (A)** Hydrogen bond interactions of NF-YA residues belonging to the Gly-loop with the CCAAT DNA complementary strand (grey), and **(B)** interactions between residues from the NF-YA A2 helix and the CCAAT DNA strand (CCAAT in violet) **[36]**. Key residues are labeled, and hydrogen bonds are shown by dashed lines.

**Figure 5. Structure of the A1A2-linker.** Structural comparison of the A1A2-linker in **(A)** the mammalian NF-Y (PDB-code 4AWL **[36]**) and **(B)** the *A. nidulans* CBC (PDB-code 4G92 **[37]**). The HFD subunits are shown in surface representation with NF-YB and the HapC in orange and NF-YC and HapE in cyan. The NF-YA and HapB subunits are shown in ribbon representation (green and blue, respectively). Relevant residues are represented in stick and indicated. **(C)** Structural-based sequence alignment of the A1A2-linker in NF-YA and HapB. Identical and similar residues are highlighted by yellow and grey shadings, respectively. Relevant Pro and Gly residues are shaded in black.

**Figure 6. Overlay of NF-Y on nucleosome $(H3/H4)_2$ tetramer. (A)** Side and top views of the hybrid nucleosome model. The NF-Y subunits (colour code as in Figure 1) of one NF-Y trimer **[36]** are shown after superimposition of the NF-YB/NF-YC on one nucleosome H2B/H2A (PDB-code 1AOI, chains D/C, respectively **[39]**). For clarity, the H2B/H2A dimers are not shown. The corresponding H3 and H4 subunits (PDB-code 1AOI: chains E/F, respectively **[39]**) are shown in red and yellow colours. The H3'and H4' subunits (PDB-code 1AOI: chains A/B, respectively **[39]**), forming with H3 and H4 the $(H3/H4)_2$ tetramer, are shown in blue and pink colours. **(B)** Close up of

the NF-YA A1 helix. A small rearrangement of the NF-YA, H4 and H4' helices (indicated) is needed to avoid close contacts (highlighted by white arrows).

**Figure 7. The *Arabidopsis* L1L/NF-YC3 dimer. (A)** Ribbon diagram showing the L1L/NF-YC3 HFD dimer (L1L in light green and NF-Y3 in light pink), and **(B)** its electrostatic surface (PDB-code 5G49) **[38]**. The blue and red colors indicate positively and negatively charged regions, respectively. The orientation of the L1L/NF-YC3 dimer is similar to that of mammalian NF-YB/NF-YC in Figure 2C. **(C)** Superposition of the *Arabidopsis* L1L/NF-YC3 dimer on the mammalian NF-Y/DNA complex. The figure shows the position of L1L Asp84 and His79 (light green sticks) relative to DNA (grey, CCAAT box in violet) and to the corresponding NF-YB residues Lys78 and Thr73 (orange sticks). For clarity, the mammalian NF-YA and NF-YC and the *Arabidopsis* NF-YC3 subunits are omitted.

**Figure 8. Sequence alignment of the CCT domain with NF-YA**. The consensus sequence for plant CCT domains is aligned with the consensus sequence for *Arabidopsis thaliana* NF-YA proteins (AtNF-YA) and human NF-YA (hNF-YA). Residues identical and similar are highlighted in yellow and grey, respectively. Asterisks indicate the position of CCT domain mutations identified in plants [80]. The consensus sequence for plant CCT and AtNF-YA is derived from the sequence alignment of 24 CCT proteins and 10 AtNF-YA subunits, as reported in Petroni et al., 2012 [68], with residues belonging to the consensus if conserved in ≥ 50% of the aligned sequences. Variable residues are indicated by x.

Figure 1
Click here to download high resolution image

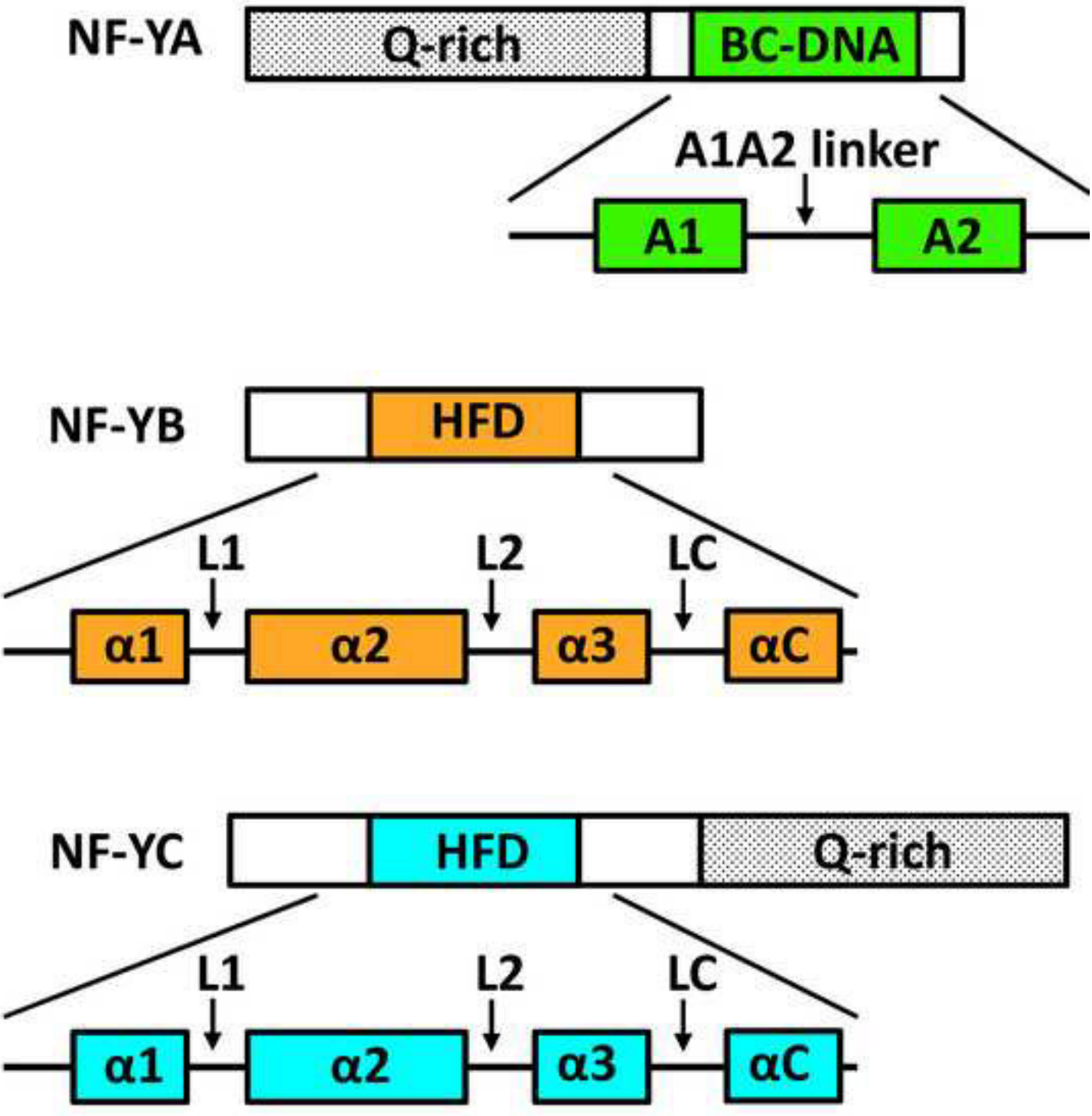


**Figure 2 (revised)**
**Click here to download high resolution image**

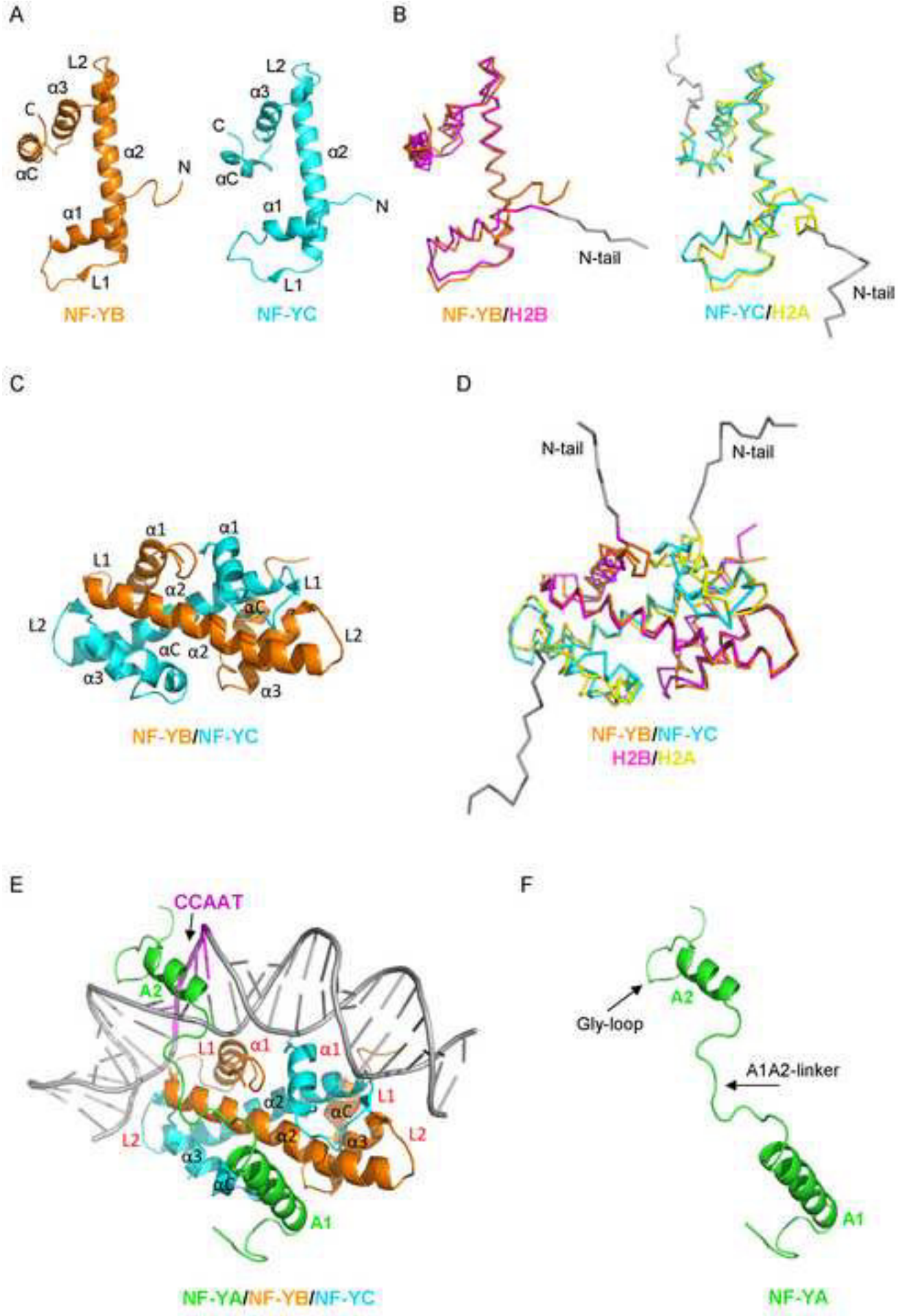


Figure 3
Click here to download high resolution image

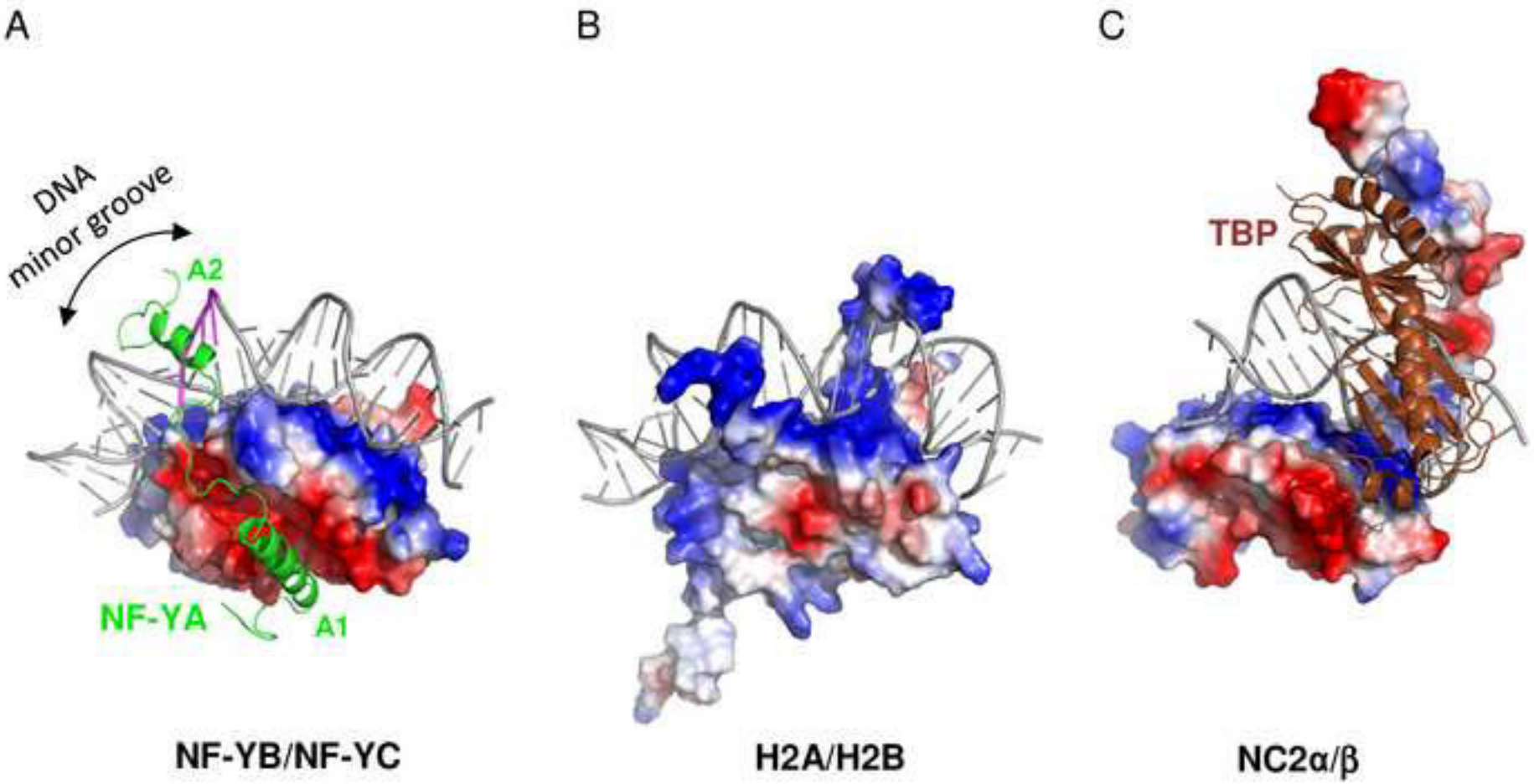


Figure 4

Click here to download high resolution image

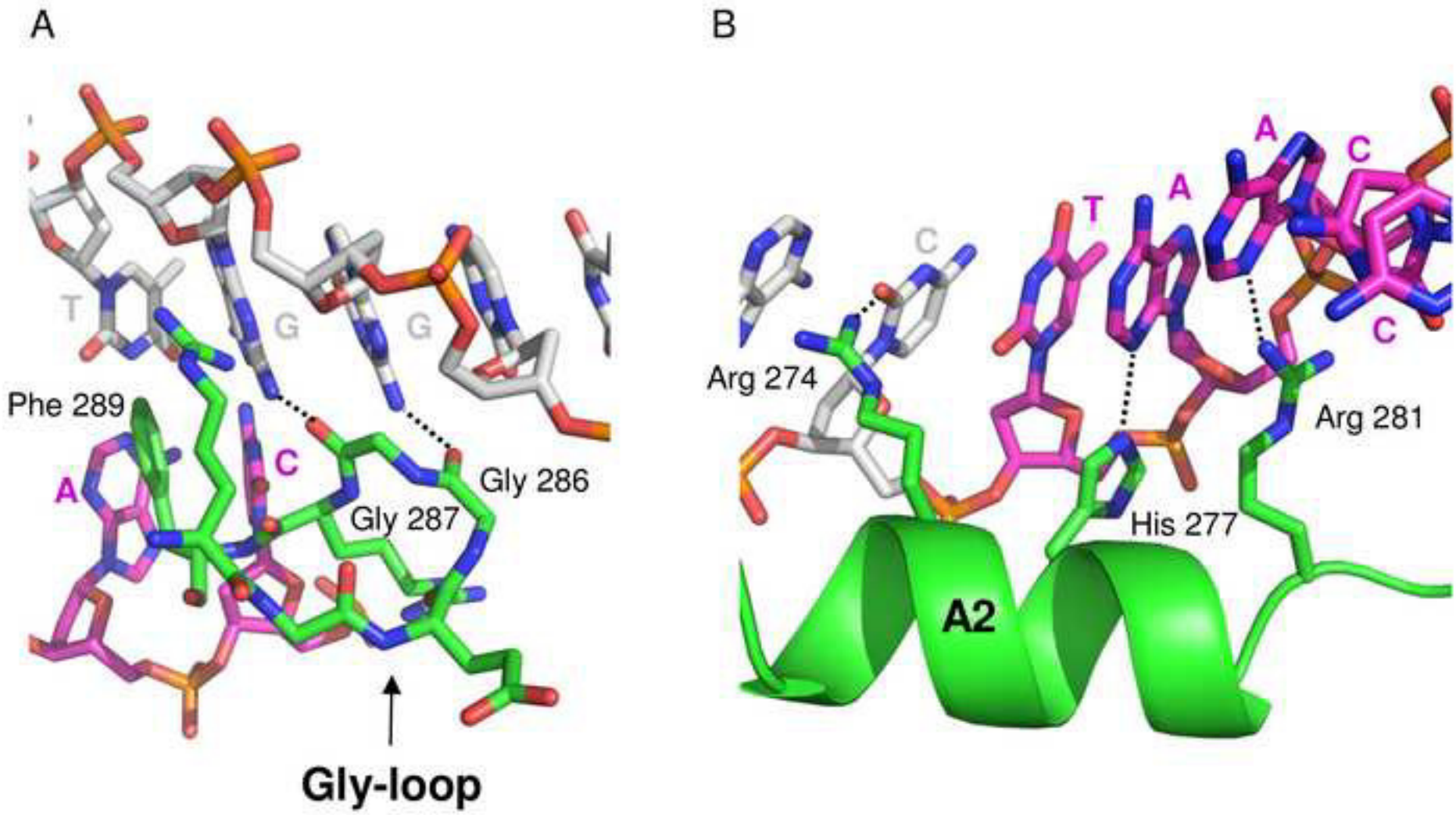


**Figure 5 (revised)**
Click here to download high resolution image

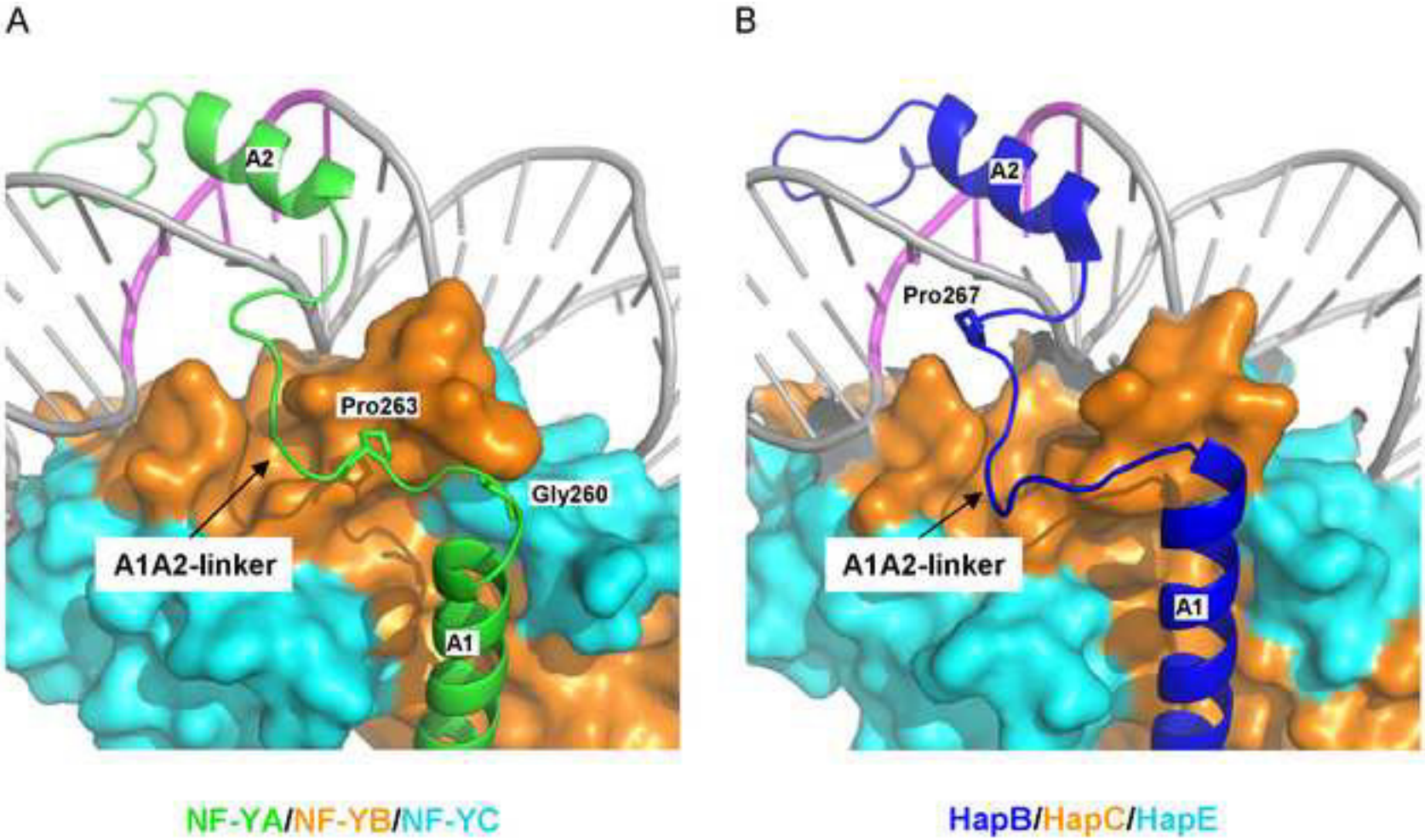


Figure 6

Click here to download high resolution image

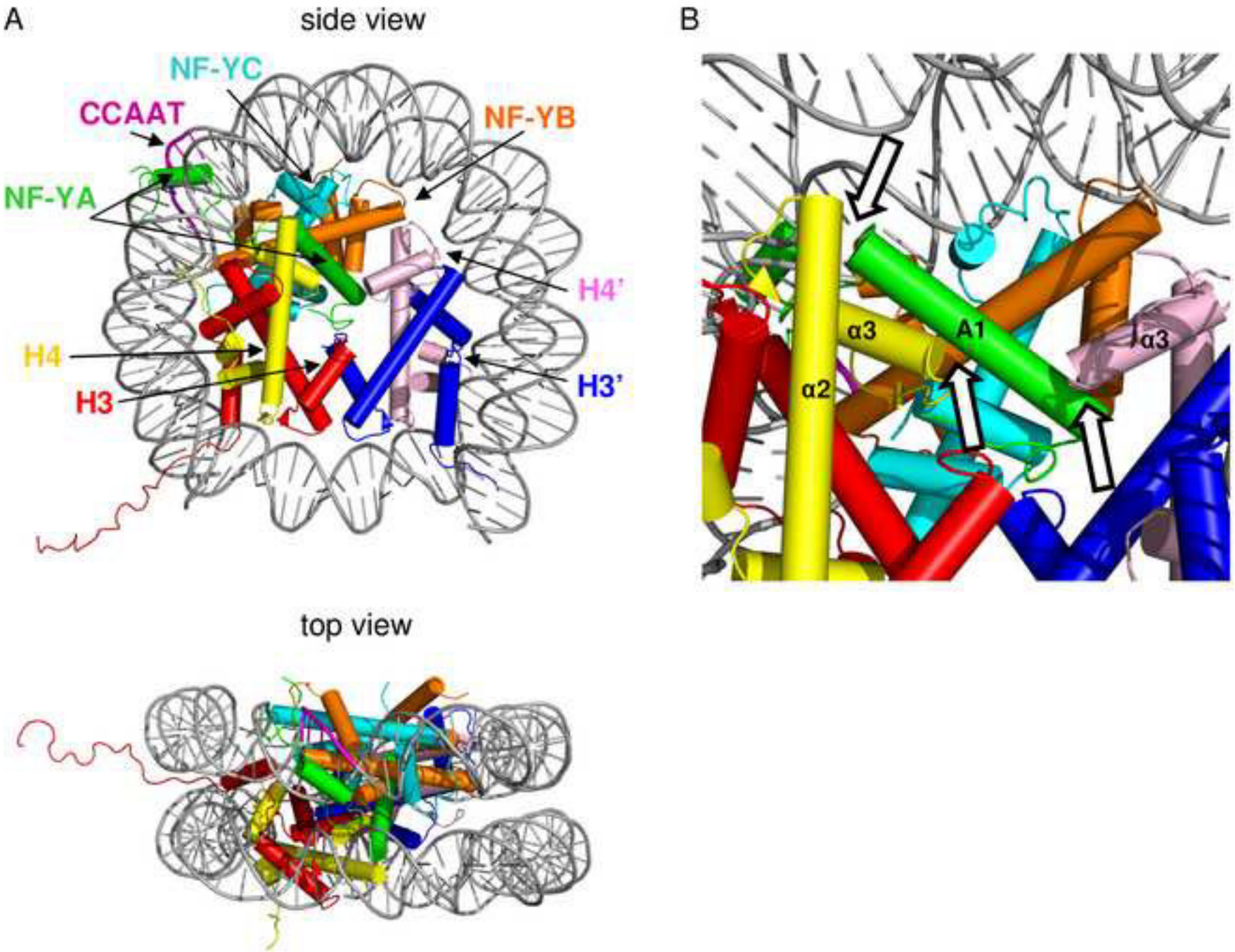


Figure 7
Click here to download high resolution image

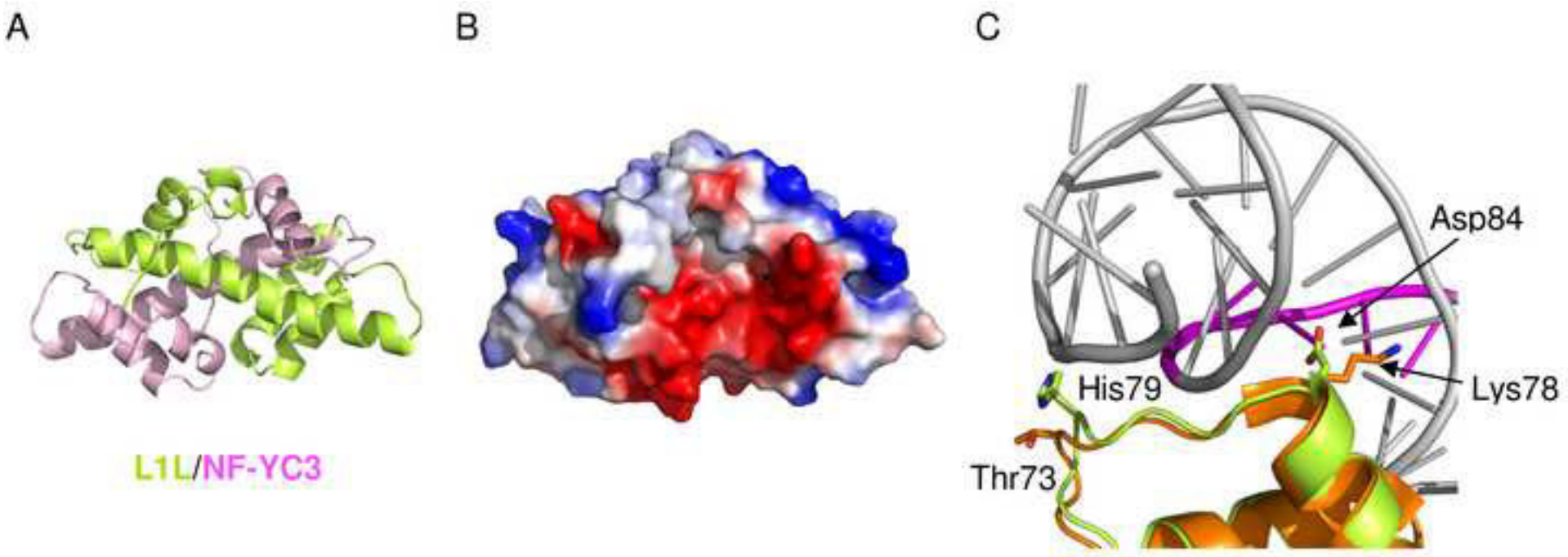
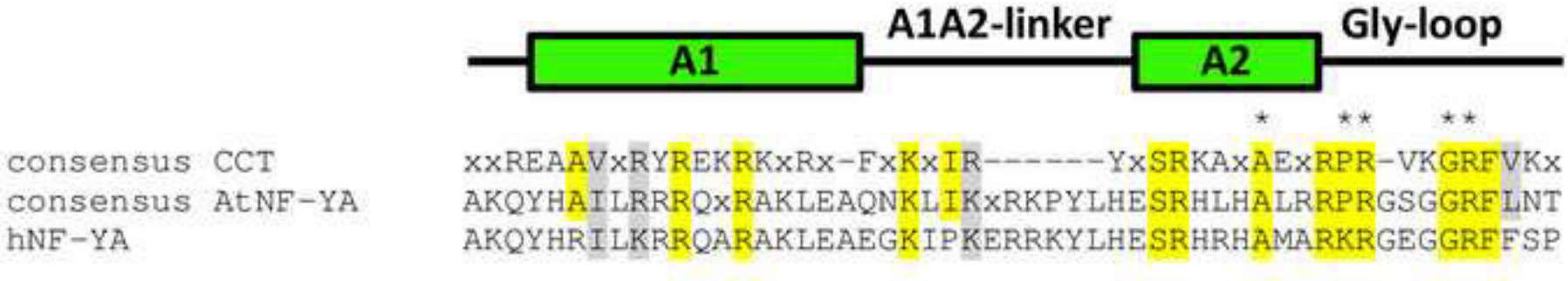


**Figure 8**

Click here to download high resolution image



***Conflict of Interest**

**Click here to download Conflict of Interest: Conflict of interest.docx**