DR MARCO NARDINI (Orcid ID : 0000-0002-3718-2165)

Article type : Original Article

Structural determinants for NF-Y subunit organization and NF-Y/DNA association in plants

Antonio Chaves-Sanjuan, Nerina Gnesutta, Andrea Gobbini, Damiano Martignago, Andrea Bernardini, Fabio Fornara, Roberto Mantovani* and Marco Nardini*

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

* These Authors are co-corresponding and co-last Authors

* To whom correspondence should be addressed:
R.M. Tel: +39 02 5031 5005; Fax: +39 02 5031 4925; Email: mantor@unimi.it;
M.N. Tel: +39 02 503 14893; Fax: +39 02 5031 4895; Email: marco.nardini@unimi.it

RUNNING TITLE
Structure of the NF-Y/DNA complex in plants

ABSTRACT

NF-Y transcription factor comprises three subunits; NF-YA, NF-YB, and NF-YC. NF-YB and NF-YC dimerize through their histone fold domain (HFD), which can bind DNA in a non-sequence-specific fashion while serving as a scaffold for NF-YA trimerization. Upon trimerization, NF-YA specifically recognizes the *CCAAT* box sequence on promoters and enhancers. In plants, each NF-Y subunit is encoded by several genes giving rise to hundreds of potential heterotrimeric combinations. In addition, plant NF-YBs and NF-YCs interact with other protein partners to recognize a plethora of genomic motifs, as the CCT protein family that binds *CORE* sites. The NF-Y subunit organization and its DNA-binding properties, together with the NF-Y HFD capacity to adapt different protein modules, represent plant-specific features that play a key role in development, growth and reproduction. Despite their relevance, these features are still poorly understood at the molecular level. Here, we present the structures of *Arabidopsis* and rice NF-YB/NF-YC dimers, and of an *Arabidopsis* NF-Y trimer in complex with the *FT CCAAT* box, together with biochemical data on NF-Y mutants. The dimeric structures identify the key residues for NF-Y HFD stabilization. The NF-Y/DNA structure and the mutation experiments shed light on HFD trimerization interface properties and the NF-YA sequence appetite for the bases flanking the *CCAAT* motif. These data explain the logic of plant NF-Y gene expansion: the trimerization adaptability and the flexible DNA-binding rules serve the scopes of accommodating the large number of NF-YAs, CCTs and possibly other NF-Y HFD binding partners and a diverse audience of genomic motifs.

KEYWORDS

Transcription factor; histone fold; *CCAAT* box; NF-Y; CCT; CONSTANS; Ghd8, flowering; *Arabidopsis thaliana*; *Oryza sativa.*

INTRODUCTION

Development, differentiation, and tissue homeostasis in eukaryotes depend upon gene expression fine-tuning, whose primordial step is the activation of transcription initiation and elongation (Levine *et al.*, 2014). This process is influenced by chromatin, wherein protein complexes with enzymatic activities modify histones through post-translational modifications (Hyun *et al.*, 2017). Transcription factors (TFs), bound to regulatory sequences in promoters and enhancers, recruit these complexes to specific genomic areas. Selective, precise, timely and synergistic binding of TFs to their DNA *cis*-elements govern transcription initiation. Mechanistically, some TFs serve a pioneering role, by penetrating highly packed chromatin regions, preparing the DNA for binding of other TFs and chromatin-modifying machines, required for transcription activation (Zaret and Carroll, 2011).

Nuclear Factor Y (NF-Y) is a heterotrimeric pioneer TF in plants (Tao *et al.*, 2017), mammals (Fleming *et al.*, 2013; Oldfield *et al.*, 2014; Lu *et al.*, 2016), and other eukaryotes kingdoms, formed by the NF-YA, NF-YB, and NF-YC subunits. NF-YB and NF-YC possess a histone fold domain (HFD) which mediates their heterodimerization and creates a molecular scaffold for NF-YA interaction (Nardone *et al*., 2017). The NF-Y DNA-target is the *CCAAT* box, found with high frequency in eukaryotic promoters (Dolfini *et al.*, 2009). Each NF-Y subunit has different and complementary DNA-binding roles: non-sequence-specific DNA binding and bending for the NF-YB/NF-YC dimer and sequence-specific *CCAAT* box recognition for NF-YA (Huber *et al.*, 2012; Nardini *et al.*, 2013).

Two aspects of the NF-Y system are peculiar to plants. First, the gene families encoding each of the three NF-Y subunits are expanded. Genetic data indicate that the different NF-Y subunit combinations participate in most, if not all, plant developmental, growth and reproduction decisions, under normal or adverse environmental conditions (Wenkel *et al.*, 2006; Gnesutta *et al.*, 2017a; Swain *et al.*, 2017). Second, the *CCAAT*-recognition domain of the NF-YA subunit shares sequence homology with other proteins, like those containing the CCT (Constans (CO), Constans-like, Timing of Cab Expression 1 (TOC1)) domain, which is found in a vast family of plant-specific TFs (Wenkel *et al.*, 2006). It has been demonstrated that two CCT-containing TFs, the *Arabidopsis* CO and the rice ortholog Heading date 1 (Hd1), bind NF-YB/NF-YC dimers to form functional heterotrimeric complexes (NF-CO) with a DNA-binding specificity different from *CCAAT* (Gnesutta et al., 2017a; Goretti et al., 2017). The regulation of the *Arabidopsis FT* florigen gene (*Hd3a* in rice) illustrates the complex and multi-task NF-Y molecular organization; the NF-Y

trimer binds to a functionally important enhancer *CCAAT* box at -5 kb from the transcription start site, and the NF-CO trimer binds to proximal promoter *CORE*s (*CO-Responsive Elements*) (Adrian *et al.*, 2010; Tiwari *et al.*, 2010; Cao *et al.*, 2014; Hackenberg *et al.*, 2012; Siriwardana *et al.*, 2016; Luo *et al.*, 2018).

The genetic expansion of multi-subunit TFs in plants, associated with neo-functionalization, could be explained by the different specialization layers provided by the diverse oligomerization possibilities. In the case of NF-Y and NF-CO, this involves their heterodimerization and trimerization potential. In this respect, systematic Y2H experiments suggested that most, and possibly all, plant NF-YB and NF-YC subunits can heterodimerize (Trigg *et al.*, 2017; Hackenberg et al., 2012; Calvenzani *et al.*, 2012; Gnesutta *et al.*, 2017b). Similarly, Y2H and Y3H experiments reported many interactions between single NF-YB and NF-YC subunits, or NF-YB/NF-YC heterodimers with NF-YAs, CCTs and other proteins (Wenkel *et al.*, 2006; Trigg *et al.*, 2017). This was further supported by the only atomic structure of plant NF-YB/NF-YC dimer reported so far (Gnesutta *et al.*, 2017b), the *Arabidopsis* AtNF-YC3 in complex with AtNF-YB6 or L1L (LEC1-Like), a divergent AtNF-YB of the LEC1 family, essential for embryo development.

The presence of evolutionarily non-conserved regions within each subunit, ultimately driving specific functions to the different heterotrimers, would represent a second regulation layer. However, elegant genetic experiments swapping the N-terminus, the C-terminus and the central HFD of the NF-YB embryo-specific LEC1 and AtNF-YB3, involved in flowering time, conclusively established that the embryo functions reside in the LEC1 HFD, and not in any external, non-conserved, parts (Lee *et al.*, 2003). The authors pinpointed a single amino acid substitution, the Asp55 within the HFD, can confer embryo functions to AtNF-YB3. Besides, our structural data on AtNF-YB6 (L1L) suggested ~~new~~ atypical DNA-binding modalities based on an Asp-His dyad diagnostic for the LEC1 protein family (Gnesutta *et al.*, 2017b). The latter results illustrate the importance to perform structural and biochemical studies on plant NF-Ys, despite the similarity of their amino acid sequences and the availability of homologous atomic structures from mammals and fungi.

Finally, trimerization yielding selectivity for different DNA targets constitutes a third regulation layer. Recently, we have reported the *CCAAT* matrix for a specific *Arabidopsis* NF-Y trimer (AtNF-Y), assayed through saturation mutagenesis and the analysis of genomic DNA sites bound *in vivo* (Gnesutta *et al.*, 2019). The plant NF-Y *CCAAT* matrix showed an absolute

requirement for the *CCAAT*-pentanucleotide and, even if the AtNF-Y trimer exhibited higher flexibility with respect to the mammalian matrix, a certain preference for nucleotides flanking the *CCAAT* was also noticed (Gnesutta *et al.*, 2019). In addition, the capacity of NF-YB/NF-YC dimer to form diverse functional trimeric complexes with other sequence-specific DNA-binding modules incredibly expands its potentials in the recognition of target DNA elements. For example, the NF-CO complex, formed between NF-YB/NF-YC dimers and a CCT protein (CO), specifically recognizes the *CORE* elements (Gnesutta *et al.*, 2017a). Therefore, the biochemical and *in vivo* data point to flexible NF-Y DNA-binding rules specific for plants, serving the scope of adapting to diverse genomic motifs.

Many relevant biological questions are still open for the NF-Y system in plants, and structural biology on this field is only at its early stage. First, the widespread plant NF-YB/NF-YC heterodimerization promiscuity is still poorly understood at structural level, as just one structure of the HDF dimer is reported in the literature (Gnesutta *et al.*, 2017b). Second, no structures for plant NF-Y heterotrimers, nor NF-COs, are available to settle questions as to the molecular basis of heterotrimeric assemblies in plants, nor for NF-YA *vs* CO trimer selectivity (Gnesutta *et al.*, 2018). Third, structural information is needed to have a complete picture of the residues providing *CCAAT* box interactions, different preferential binding to nucleotides flanking the *CCAAT* box (Gnesutta *et al.*, 2019), and to extend the model to the many plant NF-Ys and NF-COs complexes.

To start addressing these questions, we crystallized and solved the ~~first~~ atomic structures of the *Arabidopsis* NF-YA6/NF-YB2/NF-YC3 trimer in complex the *FT* -5.3 kb *CCAAT* box (AtNF-Y/DNA), as well as two NF-YB/NF-YC heterodimers, from *Arabidopsis* and rice, involved in flowering. The joint analysis of all NF-YB/NF-YC dimeric structures available to date provided a detailed mapping of residues contributing to the NF-Y HFD assembly. This result supports the "all combinations are possible" paradigm for the plant NF-YB/NF-YC heterodimerization. The analysis of the AtNF-Y trimer, and its comparison with those from mammals and fungi, demonstrates that the NF-YB/NF-YC dimer adaptive trimerization lies on a responsive electrostatic-interaction network efficient on hosting a number of different interaction partners. Based on such knowledge, we analysed and generated a set of NF-YB/NF-YC trimerization mutants that discriminate between two AtNF-YAs and CO. Furthermore, the AtNF-Y/DNA structure identified contacts to *CCAAT* box flanking bases that contribute to certain specificity, in agreement with recent mutagenesis studies that suggest flexible DNA-binding rules to recognize a variety of genomic CCAAT motifs (Gnesutta et al., 2019).

RESULTS

Structural determinants of NF-YB/NF-YC dimerization in plants

The genetic expansion of NF-Y subunits in plants raises the question about the general structural requirement for NF-YB/NF-YC dimerization through their HFD regions and for the ability of the HFD dimer to interact with NF-YAs and CCTs. Regarding the first aspect, two systematic studies on *Arabidopsis* NF-Y HFD (Calvenzani *et al*., 2012; Hackenberg *et al*., 2012), and TFs interaction screenings through Y2H (Trigg et al., 2017) indicated that they can form various heterodimers. However, no comparative structural analysis identifying residues involved in HFD dimerization in plants has been presented to date. For this reason, we determined the crystallographic structure of the HFD of two orthologous NF-YB/NF-YC dimers, AtNF-YB2/NF-YC3 (in two crystal forms: PDB-codes 6R0M and 6R0N) and *O. sativa* Ghd8/NF-YC7 (PDB-code 6R0L), involved in the regulation of flowering timing in *Arabidopsis* and rice. (Table 1). Both heterodimers adopt the typical head-to-tail architecture of HFDs (Figure 1a) and the structural comparisons reveal their similarity, with a rmsd of 0.85 Å, calculated over 164 Cα pairs (sequence identity of 83.0% for the NF-YB subunit and 83.9% for the NF-YC subunit). The previously reported AtNF-YB6/NF-YC3 plant dimer (PDB-code 5G49), involved in embryo development (Gnesutta *et al.*, 2017b), superposes well with the structures here presented, with a maximum rmsd of 1.23 Å, calculated over 164 Cα pairs for the most deviant plant NF-YB/NF-YC dimer (Figure 1a). The core of the HFD is conserved in terms of secondary structure, positive charge distribution at the DNA-binding surface and negative charge distribution at the NF-YA binding groove (Figure 1b). Divergent regions are only localized in flexible loops at the N- and C-terminal regions of each subunit.

We next analyzed the intermolecular contacts in plant NF-YB/NF-YC structures and we compared them with those present in the mammalian and the *Aspergillus* structures to determine residues involved in dimerization and the magnitude of their contributions. We then extended these results to other plant NF-YB and NF-YC proteins. Residues buried at the dimeric interface (having a solvation energy above 0.5 kcal/mol) show higher sequence conservation, and likely account for dimer integrity (Figure S1). This is in agreement with mutation experiments reported so far (Nardini *et al.*, 2013). It is worth noting that 37.5% of the total HFD residues face the dimer

interface and that the vast majority of them participate in van der Waals contacts, with few polar interactions that mostly involve main-chain atoms. When only these residues are used for structure superposition, the rmsd between AtNF-YB2/NF-YC3 and Ghd8/OsNF-YC7 drops down to 0.4 Å, and with AtNF-YB6/NF-YC3 to 0.8 Å, indicating a strong and conserved shape complementarity at the subunit interface.

In summary, we expect that the majority of plant NF-YB and NF-YC subunits accomplish mutual heterodimerization, in agreement with the promiscuity experimentally highlighted by Y2H experiments (Calvenzani et al. 2012; Hackenberg et al., 2012; Trigg *et al.*, 2017). Shape complementarity and sequence similarity are the major driving force for HFD heterodimerization, more than the presence of specific residue-residue interactions. However, sequence variations at the interface of the HFDs might tune the binding affinity between subunits and/or the stabilization of the complex, which might, in turn, finely balance dimer proportions within the cell.

The *Arabidopsis* NF-Y trimer in complex with DNA

The atomic structure of the *A. thaliana* NF-YA6/NF-YB2/NF-YC3 trimer (AtNF-Y) in complex with a 25 bp dsDNA, containing the *FT* enhancer *CCAAT* box was solved at 2.5 Å resolution (PDB-code 6R2V) (Figure 2). The rationale behind employing this combination of subunits was that (i) genetic experiments demonstrated that AtNF-YB2/NF-YC3 regulate *FT* gene expression, (ii) they are involved in control of flowering timing by binding the -5kb enhancer *in vivo* (Kumimoto *et al.*, 2008; 2010; Cao et al., 2014), (iii) they share overlapping expression patterns with AtNF-YA6 (Gusmaroli *et al.*, 2001; Siefers *et al.*, 2009) and (iv) the *CCAAT* of the *FT* enhancer is the only *bona fide CCAAT* box bound by NF-Y whose function has been experimentally tested in plants so far (Gnesutta *et al.*, 2019). Data collection and refinement statistics are reported in Table 1.

The AtNF-YB2 and AtNF-YC3 subunits provide the scaffold for AtNF-YA6 binding and DNA recruitment through their HFD (Figure 2). The AtNF-YB2/NF-YC3 dimer superposes well when it is part of the NF-Y/DNA complex or isolated (rmsd of 1.1 Å, calculated over 169 Cα pairs). AtNF-YA6 displays an elongated structure hosting two helices, A1 and A2, the A1A2-linker and the Gly-rich loop (Gly-X-Gly-Gly-Arg-Phe) located just after the A2 helix (Figures 2, and 3). The positively charged A1 helix mediates the trimerization with the HFD module, while the A2 helix and the Gly-rich loop deeply insert into the DNA minor groove at the *CCAAT* box making all sequence-specific DNA contacts (Figures 4, and S2). The structure of AtNF-Y/DNA

provides the ~~first~~ detailed picture of the NF-Y trimer in plants, and its interaction with a functional DNA sequence. When compared with its mammalian (NF-Y/DNA) and *Aspergillus* (AnHapB/C/E/DNA) counterparts, it appears that the overall architecture of the complex is mostly conserved (rmsd of 1.40 Å and 2.50 Å were calculated over 228 Cα pairs, respectively), yet with significant plant-specific features that emerge in both the NF-YA/HFD trimer interface and the NF-Y/DNA interactions.

Plant-specific features at the NF-Y trimerization interface

AtNF-Y heterotrimerization occurs mainly through interaction of the AtNF-YA6 A1 helix with the HFD dimer. The polar and basic residues of the A1 helix interact *via* hydrogen bonds and salt bridges with the large negatively charged groove of the AtNF-YB2/NF-YC3 dimer (contact interface of 1,223 $Å^2$). While all HFD residues involved in AtNF-Y trimerization are highly conserved in plant, mammal, and *Aspergillus* HFDs (Figure S1), specific amino acid substitutions underlie striking differences in the AtNF-YA6 subunit (Figure 3). The structure shows that the AtNF-YA6 A1 helix is shifted ~1.5 Å along its axis and rotated 4° and 8° when compared to the mammalian and *Aspergillus* counterparts, respectively (Figure S3a,b). The variable positioning of the A1 helix relative to the HFD dimer can be explained by the presence of a different interaction pattern between subunits, which is specific for each NF-YA. In mammals, NF-YA Arg245 interacts with NF-YB Ser97, Glu98 and Glu101. Similarly, *Aspergillus* Arg243 interacts with Ser88, Glu89 and Glu92 of the NF-YB subunit. In AtNF-YA6, instead, the equivalent Arg residue is substituted by the apolar Ala182 (Ala or Gly in all *A. thaliana* NF-YAs) (Figure 3). As a consequence, AtNF-YA6 Arg185, positioned one helix turn after Ala182, replaces the missing contact with Glu73 and Asp76 of AtNF-YB2 (equivalent to Glu98/Glu101 and Glu90/Glu92 in mammalian NF-YB and AnHapC, respectively) (Figure 5). Interestingly, the AtNF-YA6 Arg185 residue corresponds to a lysine in mammals and *fungi*, where it is irrelevant for trimerization, as shown in previous structural and mutational studies (Huber *et al.*, 2012; Nardini *et al.*, 2013; Xing *et al.*, 1993). Based on these observations, we conclude that, despite the conservation of a positively charged residue corresponding to AtNF-YA6 Arg185, its structural role differs in different NF-YA subunits depending on the polarity of the amino acid located two residues ahead in the sequence (Ala/Gly or Arg).

The shifted position in A1 is further stabilized by a double salt bridge between the HFD residues, AtNF-YB2 Asp76 and AtNF-YC3 Asp83, and AtNF-YA6 Arg189, modelled in two alternate conformations (Figure 5). This interaction is completely absent in mammalian and *Aspergillus* NF-Ys, where an Ala substitutes the AtNF-YA6 Arg189 and some water molecules satisfy the H-bonding acceptance of the conserved orphans Asp/Glu (Figure 3). It should be noted that Arg189 is not generally conserved in AtNF-YAs. Only the phylogenetically related AtNF-YA5 has a positively charged His residue (Petroni *et al.*, 2012). Yet, in the plant homolog CCT family a Lys residue is conserved in the same position and possibly fulfils a similar function (Figure 3). Another important difference between plant, mammalian and *Aspergillus* NF-Ys arises at the conserved NF-YC L1 loop acidic stretch (Asp83 and Asp85 in AtNF-YC3): it does contact AtNF-YA6 within the plant trimer (Figure 5), whereas it does not participate in trimerization in mammalian and *Aspergillus* NF-Ys.

In summary, the trimerization interface between the AtNF-YA subunit and the AtNF-YB2/NF-YC3 dimer differs respect to the mammalian and *Aspergillus* counterparts due to a shift in Arg distribution, with consequent reorganization of the interaction network. The extensive polar and acidic NF-YB/NF-YC interface grants alternative options for H-bonding and salt bridges formation with the variable NF-YA A1 positively charged residues. The differences in length and sequence composition typical of the A1A2-linker in plant NF-YAs, and homologous proteins, allow correct positioning of the A2 helix at the *CCAAT* box, independently of the A1 helix orientation relative to the NF-YB/NF-YC dimer.

Analysis of HFD mutants impacting trimerization

The comparison of the AtNF-Y structure with those from mammal and *Aspergillus* shows that the NF-Y trimerization interface is sufficiently plastic to reshape the interaction pattern of H-bonds and salt bridges between the NF-YA subunit and the HFD dimer to provide a similar trimer architecture despite sequence variability. This aspect is relevant not only because it would account for the flexible combinatorial NF-Y assemblies typically found in plants, but also because it is conceptually linked to the ability of the plant HFD dimer to accommodate sequence-specific subunits other than NF-YAs, notably CCTs. To asses this scenario, we produced and analyzed HFD mutants that could discriminate interactions among different NF-YAs and between NF-YA and CCT. Guided by the AtNF-Y structure, we produced HFD dimers with AtNF-YB2 or AtNF-YC3 mutated in acidic residues (E65R and D83R-D85R, respectively), and performed dose-

response EMSAs with AtNF-YA6, AtNF-YA2 - both sharing overlapping expression patterns with the AtNF-YB2/NF-YC3 dimer subunits (Gusmaroli *et al.*, 2001; Siefers *et al.*, 2009), and with CO (Wenkel *et al.*, 2006). We chose three Cy5-labelled DNA probes reported in the literature (Gnesutta *et al.*, 2019): the *CCAAT* box of *FT* and of *CHLOROPHYL A/B BINDING PROTEIN 2* (*CAB2*; *lhcb1*1*), both targets of AtNF-Y, and the *FT* promoter *CORE2* (*CCACA*) element, to test NF-Y/DNA and NF-CO/DNA complex formation, respectively (Figure 6). For the interpretation of EMSA results, it is important to note that, in the presence of intact heterodimerization and DNA-binding capability, a decrease in the formation of the protein/DNA complex exclusively accounts for a reduction in trimerization. Heterodimer integrity of mutant HFD proteins, is in fact confirmed by the dimeric assembly of purified proteins obtained by co-expression of 6His-tagged NF-YB subunits together with untagged NF-YC, as verified by electrophoresis (Figure S4).

The AtNF-YB2-E65R mutant dimer retains substantial binding for AtNF-YA6, but it is extremely detrimental for AtNF-YA2 trimerization and, to lesser extent, for CO (Figure 6a; see also Figure S5a). Molecularly, AtNF-YA differences might be explained by the multiple contacts of the flexible AtNF-YA6 Arg189 with AtNF-YC3 Asp83 (L1 loop of the HFD) and AtNF-YB2 Glu73 and Asp76 (α2 of the HFD) (Figure 5), sufficient to compensate for the loss of the Glu65(YB2)-Arg186(YA6) interaction (Figure 3). This is not the case for AtNF-YA2 and CO, where the residue corresponding to AtNF-YA6 Arg189 is Ser and Lys, respectively (Figure 3). The AtNF-YC3-D83R-D85R double mutant displays, instead, the most dramatic effect on CO, whose binding is completely obliterated (Figure 6b; see also Figure S5b), while residual binding is scored for AtNF-YA2/A6, particularly on the *CAB2 CCAAT*. Although it is not yet possible to rationalize the contacts of CO with HFDs, due to the absence of the NF-CO/DNA structure, the AtNF-YC3-D83R-D85R mutant confirms the critical role of AtNF-YA6 Arg189 that when Lys, as in the case of CCTs, possibly only interacts with the acidic AtNF-YC L1 loop region and not with the AtNF-YB α2.

In conclusion, these data support the notion of the differential trimerization rules between NF-YAs and among NF-YAs and CCTs. In particular, we showed the impact of mutations of the acidic patch formed by AtNF-YC L1 (D83R-D85R in AtNF-YC3) and AtNF-YB α2 (E65R in AtNF-YB2), discriminating between CO and AtNF-YAs and between different AtNF-YAs, respectively.

Plant-specific NF-Y DNA-binding properties

In the AtNF-Y/DNA complex, the bound oligonucleotide bends about 80º, coating the positively charged surface of the HFD. The insertion of AtNF-YA6 A2 helix into the minor groove, containing the *CCAAT* sequence, with its consequent widening, reinforces dsDNA bending (Figure 2). AtNF-Y contacts the DNA through 32 residues from all three subunits, fewer than mammalian NF-Y and AnHapB/C/E complexes. Most of these residues belong to the L1, L2 and α1 regions of the HFDs and they all establish polar interactions with the phosphate backbone distributed along four helical turns of the bound DNA (Figure S2). These non-sequence-specific interactions mostly reproduce the interaction pattern detailed for the NF-Y/DNA and AnHapB/C/E/DNA complexes (Huber *et al.*, 2012; Nardini *et al.*, 2013). Despite these overall similarities, relevant plant-specific differences emerge that include contributions of (1) the region that follows AtNF-YA A2 helix, also termed "Gly-loop" (Gly1-X-Gly2-Gly3-Arg-Phe) and (2) the N-terminal part of AtNF-YC3.

CCAAT specificity is provided by five AtNF-YA6 residues situated in the A2 helix (His214) and the following loop region (Arg218, Arg220, Gly224, Phe226). Arg211 provides additional contacts at the +1 flanking position (Figures 3, 4, and S2). As expected, this pattern is reminiscent of that found in mammalian and *Aspergillus* NF-Y/DNA complexes, yet with an important difference. In AtNF-YA6 the carbonyl O atom of Gly224 (Gly3 of the Gly-loop) contacts the N2 of both *G*(-8) and *G*(-9) of the *CCAAT*-complementary strand (Figure 4), while Gly223 (Gly2) is dispensable. On the contrary, in mammalian and *Aspergillus* NF-YA the carbonyl O of both Gly2 and Gly3 are involved in base interactions, with the N2 of *G*(-8) and of *G*(-9), respectively. This is relevant since AtNF-YA6 might represent a good prototype concerning the DNA-binding mode by CCTs, where only Gly3 is present (Figure 3). Still, the stem Arg-X-Arg and the second part of the Gly-loop (Gly-Arg/Gln-Phe) motifs are conserved and contain key residues for specific contacts to the first three nucleotides of *CCAAT* (Figure 4). Thus, we predict that the differences in the Gly-loop between CCTs and AtNF-YAs will not affect NF-CO sequence-specific contacts to the *CCA* bases of the bound *CORE* matrix (*CCACA*). We remark that AtNF-YA6 His214, which does bind the second *A* of the *CCAAT* (a *C* in *CORE*) is not conserved in CCTs (Figure 3).

A second notable difference in the DNA-binding contacts of plant NF-Y relates to the N-terminal part of the AtNF-YC3 structure, where residues 65-70 are close to the DNA, with Lys67 and His69 side chains providing two potential additional DNA-phosphate contacts (Figure 7a). To check whether the structure of this region is driven by the interaction with DNA, we compared it with the crystal structure of the isolated AtNF-YB2/NF-YC3 (Table 1). In the isolated AtNF-

YB2/NF-YC3 dimer, only AtNF-YC3 His69 is visible, projecting its side chain away from the expected DNA site (Figure 7b). In *Arabidopsis* L1L/AtNF-YC3 HFD dimer (PDB-code 5G49), both NF-YC3 Lys67 and His69 are facing the putative DNA-binding region, but with a different backbone conformation relative to the NF-Y/DNA complex (Figure 7c). Thus, it appears that the presence of the bound DNA alters the AtNF-YC3 Lys67 and His69 position, favouring their orientation towards the phosphate backbone of the DNA. Both residues are conserved in *Arabidopsis* and in most plant NF-YCs, while in animals a basic residue and a Gln are present. In the mammalian NF-Y/DNA structure the corresponding Arg is not visible, and the Gln residue is facing away from DNA (Nardini et al., 2013). Lys and His residues are also conserved in *Aspergillu*s NF-YC (HapE Lys83 and His85): however, the crystallized *Aspergillus* NF-YC further extends N-terminally, being differently folded from plants as an additional αN helix that runs antiparallel to α2 helix of the HFD. Nevertheless, Lys83 is also facing the DNA phosphate backbone, albeit toward the following turn of the DNA helix (Figure 7d). In summary, the AtNF-YC N-terminus provides a ~~new~~ plant-specific DNA-binding surface, leading to almost complete wrapping of the NF-Y trimer around the bound DNA.

A final important aspect of the NF-Y/DNA interaction is related to regulation. A phosphorylation-dependent regulatory mechanism on NF-YA Ser292 has been demonstrated for mammalian NF-Y functional activity (Yun *et al.*, 2003; Bernardini *et al.*, 2019). In the mammalian NF-Y structure, NF-YA Ser292 directly contacts the DNA phosphate backbone at the first adenine of the *CCAAT* box (Nardini et al., 2013). Ser292 phosphorylation reduces the stability of the NF-Y/*CCAAT* complex and, consequently, its functional activity at *CCAAT*-dependent promoters (Yun *et al.*, 2003; Bernardini *et al.*, 2019). In AtNF-YA6, the corresponding residue is the Asn228 (Figure 3), which marginally interacts with DNA. However, in the AtNF-Y/DNA structure the following residue, the phosphorylatable Thr229, H-bonds to the phosphate group of the second adenine of the *CCAAT* box (Figures 4, and S2). Most AtNF-YAs include a Thr/Ser at this position, except for AtNF-YA4 and AtNF-YA7 (Petroni *et al.*, 2012). Thus, the AtNF-Y/DNA structure suggests that the regulatory mechanism demonstrated for the mammalian NF-YA Ser292 phosphorylation might also account in plants.

DISCUSSION

Fundamental questions focus on the regulation of multi-subunit TFs in plants, given their genetic expansion with respect to other kingdoms. Three regulation layers are explored at structural and biochemical level in this work: (i) the combination of the different NF-Y HFD pairs, (ii) the trimerization of NF-YB/NF-YC dimers with different AtNF-YAs and with CCTs partners, and (iii) the details of the DNA/NF-Y interaction. Our structural analysis of the intermolecular contacts of NF-YB/NF-YC plant structures, and comparison with the mammalian and *Aspergillus* cases, shows that the regions with the higher impact on dimer stabilization correspond to the most conserved in terms of amino acid sequence and backbone shape (Figure S1). These results are the ~~first~~ structural evidence to support a paradigm for NF-Y HFD assembly whereby essentially all combinations are possible, driven by shape complementarity and sequence similarity, suggesting that specific amino acid composition at the dimeric interface might, in general, only slightly tune affinity between specific subunits, balancing dimer proportions of coexpressed subunits within the cell.

Flexibility of NF-Y trimerization to host different DNA readers

The increased size of NF-Y families in plants reflects a dramatic increase of regulatory complexity by HFD containing complexes, not only because of hundreds of potential "canonical" NF-Y trimers, but also because of the thousands potential NF-CO (or NF-CCTs) complexes. The CCT domain is found in 39 *Arabidopsis* proteins including CO, CO-like, TOC1 and PRR sub-families that are phylogenetically classified in five clades (Cockram *et al.*, 2012). The DNA-binding A2 helix and the following Gly-rich loop of NF-YA and CCT shares a certain sequence homology, while the trimerization A1 helix seems to diverge (Figure 3). Therefore, we focused our research on the trimerization interface to deepen the knowledge of NF-Y subunits combinatorial possibilities. On one hand, the AtNF-YA6 structure permits to hypothesize the molecular determinants for NF-YAs and CCTs A1 helix interaction with the NF-YB/NF-YC dimer. On the other hand, our analysis of the HFD mutants points out key structural features of the NF-YB/NF-YC dimer to provide flexible interaction properties to the trimer.

At the trimerization interface, AtNF-YA6 Arg187 and Arg190 are conserved in mammalian NF-YA and AnHapB, as well as in CO, CO-likes (COLs), TOC1 and PRRs as Arg/Lys (Figure 3). These two Arg residues face the HFD negative surface and contact the NF-YC subunit (Figure 5) and likely play a similar role in CCTs. AtNF-Y6 Ala182, and the equivalent Arg residue in NF-YA and AnHapB (Figure 3), is the key residue that determines the whole trimer interaction

network that can be categorized in two binding modes: *à la* AtNF-YA6 and *à la* NF-YA/AnHapB. An Ala in such a position results in a missing electrostatic interaction with acidic residues from the HFD. In AtNF-YA6, an Arg/Arg dyad (Arg185 and Arg189) compensates that with other extra salt bridges (Figure 5), while the equivalent Lys/Ala dyad is dispensable in NF-YA and AnHapB (Figure 3) (Huber *et al.*, 2012; Nardini *et al.*, 2013). CCTs of clades III and IV, including AtCOL9 and PRRs, present an Ala similar to AtNF-YA6 (Figure 3). CO and COLs accommodate an Arg/Lys dyad, whereas TOC1/PRRs a Lys/Lys (Figure 3). Thus, we propose a similar stabilizing role for the basic/basic residue dyad in CCTs with a binding *à la* AtNF-YA6. Instead, CO and COLs of the CCT family clades I and II that contain an Arg (Figure 3) and, therefore, are expected to show an HFD-binding pattern *à la* mammalian NF-YA/AnHapB.

The residue located at an equivalent position of AtNF-YA6 Leu193 could represent another key point for CCTs. All NF-YAs typically contain a hydrophobic Leu/Ala, which is substituted by Arg in all CCTs (Figure 4). AtNF-YA6 Leu193 faces AtNF-YC3 Ala82, which is followed by two absolutely conserved residues, Asp83 and Glu84 (Figure 5). This conserved acidic stretch has been somewhat a mystery in the previous mammalian NF-Y/DNA or AnHapB/C/E/DNA structures because no intersubunit contacts were shown. In the plant AtNF-Y trimer, there is an Arg189-Asp83 contact between AtNF-YA6 and AtNF-YC3 (Figure 5). The CCT Arg residue could nicely reproduce, together with the Arg189-corresponding basic residue conserved in CCTs (see below), the same electrostatic interaction with this acidic stretch provided by all plant NF-YCs.

In summary, sequence and structure analyses indicate that the promiscuous NF-Y trimerization behavior in plants depends on its conserved, large, HFD acidic patch which can provide alternative interaction options for the various Arg distribution at the A1 helix of the sequence-specific subunit of the trimer. Such a conclusion, that can be extended to NF-CO complexes, is supported by the results of experiments on HFD mutants to test their impact on the binding of two different NF-YAs and one CCT protein, CO.

Based on previous mutagenesis studies in mammalian NF-YA and AtNF-YA2 (Sinha *et al.*, 1996; Tiwari *et al.*, 2010), we selected AtNF-YB2-E65R as a good candidate to challenge trimer formation. EMSA experiments showed that AtNF-YA6 retains substantial binding, less so CO and absolutely no trimerization with AtNF-YA2 (Figure 6a; see also Figure S5a). The loss of the Arg186-Glu65 contact prevents AtNF-YA2 trimerization, while AtNF-YA6 compensates with its specific Arg189-Asp83 interaction (Figure 5). This compensation is, instead, not possible in

AtNF-YA2 where a Ser substitutes AtNF-YA6 Arg189 (Figure 3). These data indicate that the loss-of-function flowering timing observed in complementing *nf-yb2 nf-yb3* mutants (Siriwardana *et al.*, 2016) cannot rely on the activity of AtNF-YA6. The AtNF-YC3-D83R-D85R mutant, instead, shows the most dramatic effect on CO, whose binding is completely obliterated, with residual binding for either AtNF-YAs tested (Figure 6b). AtNF-YA6 Arg189 interacts with AtNF-YC3 Asp83 or AtNF-YB2 Glu73 or Asp76 (Figure 5). It has to be noted that among NF-YAs only NF-YA6 and NF-YA5 host a basic residue at this position, while CCTs display a conserved Lys. It is not yet possible to rationalize the contacts of CO with the HFD, however, this AtNF-YC3 L1 mutant confirms the importance of basic contacts of the A1 involving the conserved Lys (corresponding to Arg189) or Arg (substituting Leu193) in the CCT family (Figure 3). If this is indeed the case, targeting these Asp residues on the NF-YC subunit with amino acids less invasive for NF-YAs interactions than arginines, might prove crucial to discriminate the NF-CO *vs* NF-Y legs of HFD *in vivo* activities.

DNA-binding

Animal kingdom species have one or two genes encoding each NF-Y subunit, limiting their trimeric combinatorial possibilities. Isoforms generated by alternative splicing represent an exception, however the DNA-binding domain always remains intact. In mammals, a 10-bp DNA matrix centred on the *CCAAT* pentanucleotide with well-defined flanking sequences has been reported in genetic studies and verified by affinity of NF-Y *in vitro* and its binding sites *in vivo* (Fleming *et al.*, 2013; Dolfini *et al.*, 2009). Plants have evolved differently. Genome-wide analysis on plant promoters showed a statistical enrichment of the *CCAAT* pentanucleotide without any flanking sequence (Kumimoto *et al.*, 2008). Recently, we have demonstrated an *in vitro* DNA matrix for the same AtNF-Y complex (AtNF-YA6/NF-YB2/NF-YC3) studied here, where the sequence-specificity is mostly limited to the *CCAAT* box, with only few preferences at the flankings (Gnesutta *et al.*, 2019). The AtNF-Y/*FT* DNA structure reported here confirms contacts with the *CCAAT* box, but also with flanking nucleotides. AtNF-YA6 Arg211 is responsible for the strong *C* preference at +1 position. Upstream of *CCAAT*, the NF-YA Gly-rich loop kinks allowing the two Gly-Gly carbonyl O atoms approach to bases of the *CCAAT* complementary strand. In mammalian NF-YA and AnHapB, both consecutive Gly residues H-bond to the N2 of the *G(-8)* and *G(-9)*. In AtNF-YA6 only the second Gly224 contacts *G(-9)*, while the Gly223 carbonyl O atom locates between the *A(-7)* and the *G(-8)* and is free from interactions (Figures 4, lower

panel). This explains the *in vitro* data indicating that the presence of *G* at the -1 position increases the affinity (Gnesutta *et al.*, 2019), because the corresponding *C* in the complementary DNA strand would bring the O2 atom of *C(-7)* and the Gly223 carbonyl O atom within H-bonding distance.

In CCTs, the NF-YA Gly-rich loop is one residue shorter, so it cannot be defined as "Gly-rich". Still, the first part of the loop (Arg-Pro-Arg) and the second (Gly-Arg/Gln-Phe) are conserved in sequence (Figure 3). Thus, we predict that the differences in the Gly-rich loop between CCTs and AtNF-YAs will not affect sequence-specific contacts for the first three bases of the *CORE* (*CCACA*), which are indeed shared with *CCAAT*. In summary, the joint analysis of the structural data presented here and the data in the literature (Gnesutta *et al.*, 2019) demonstrate the flexible behaviour of the plant NF-Y DNA-binding rules to adapt to a more diverse context of genomic motifs as compared to the mammalian requirements for *CCAAT* binding.

The region following the Gly-rich loop contains also Thr229, which is H-bonded to the phosphate group of the second adenine of the *CCAAT* box (Figure 4, upper panel). This phosphorylatable residue is highly conserved (Thr/Ser) in AtNF-YAs (Petroni et al., 2012) and it structurally matches the position of mammalian NF-YA Ser292, whose phosphorylation has been demonstrated to decrease the stability of the NF-Y/*CCAAT* complex and thus the functional activity of *CCAAT*-dependent promoters (Bernardini et al., 2019). A similar regulatory mechanism might be proposed to occur in plants, possibly through different effector kinases targeting Thr residues. Indeed, based on developmental adaptation to stress, phosphorylation has been hypothesized to occur on NF-YA1 subunits in legume symbiotic interactions. (Zanetti et al., 2017)

Taken together, the structural and mutational data support a model of increased flexibility for subunit interactions and DNA-binding in plants, serving the scopes of adapting HFDs to a large number of sequence-specific partners, such as NF-YA and CCT subunits, and to a more diverse set of genomic motifs.

EXPERIMENTAL PROCEDURES

Protein expression, purification

Expression vectors for AtNF-YA6-6His (aa 170-237), AtNF-YA2-6His (aa 134-207), and CO-6His (aa 290-352) were previously described (Gnesutta *et al.*, 2017a; Siriwardana *et al.*,

2016). NF-YB and NF-YC co-expression vectors of WT *Arabidopsis* 6His-NF-YB2/NF-YC3 (AtNF-YB2 aa 24-116; AtNF-YC3 aa 55-148), rice 6His-Ghd8/OsNF-YC7 (Ghd8 aa 54-147; Os-NF-YC7 aa 4-97), and of mutant 6His-AtNF-YB2-E65R/AtNF-YC3 were described (Goretti *et al.*, 2017; Gnesutta *et al.*, 2017a; Siriwardana *et al.*, 2016). Mutant AtNF-YC3 D83R-D85R cDNA coding for aa 55-148 and containing Asp to Arg substitutions in residues 83 and 85 was obtained by gene synthesis (Eurofins Genomics) and subcloned in pmcnYC expression vector. Proteins were produced in *E. coli* and purified by nickel-ion metal affinity chromatography (IMAC) as described (Calvenzani *et al.*, 2012; Romier *et al.*, 2003). HFD dimers were produced by co-expression of 6His-tagged NF-YB subunits with untagged NF-YCs. In all WT and mutant constructs, the NF-YC subunit co-purified from the soluble fraction with the respective His-tagged NF-YB partner during the IMAC procedure, indicating correct folding and dimerization of the HFD. After protease cleavage with thrombin to remove the 6His-tag, proteins were further purified by size exclusion chromatography in a buffer containing 10 mM Tris-HCl pH 8.0, 400 mM NaCl, 2 mM DTT (Buffer B). 6His-tagged IMAC purified proteins used in EMSA were dialysed against the same buffer including 10% glycerol, frozen in liquid nitrogen and stored at -80 °C.

Crystallization, and 3D determination

Ghd8/OsNF-YC7 crystals grew in *I*2 space group, while two crystal forms of AtNF-YB2/NF-YC3, belonging to *P*212121 and *I*2, were obtained. Ghd8/OsNF-YC7 crystals were obtained using the vapor diffusion method with 20% PEG 4000, 0.1 M sodium acetate pH 5.0, 0.2 M ammonium acetate at 4 °C. AtNF-YB2/NF-YC3 *P*212121 crystals were grown in vapor diffusion experiments using 30% PEG 3350, 0.1 M MES pH 6.5 and 0.3 M lithium citrate as precipitant at 4 °C. AtNF-YB2/NF-YC3 *I*2 crystals grew using the salting-in method in which the protein (in 150 mM NaCl, 50 mM Tris-HCl pH 8.0 and 2 mM DTT buffer) was dialyzed against a buffer containing 100 mM NaCl, 50 mM Tris-HCl pH 8.0 and 2 mM DTT at 4 °C.

Equimolar amounts of AtNF-YA6, AtNF-YB2/NF-YC3 and a lyophilized 25bp *CCAAT* box derived from the *FT* enhancer were mixed in a buffer containing 10 mM Tris-HCl pH 8.0, 400 mM NaCl, 2 mM DTT. Then, buffer salt concentration was decreased to 50 mM NaCl, 50 mM Tris-HCl pH 8.0 and 2 mM DTT by dilution with a buffer without NaCl. Proteins were concentrated to 7 mg/ml and vapor diffusion experiments were set up yielding the best crystals in 10% PEG 8000, 5% ethylene glycol, 0.1M MES pH 6.5 at 21 °C.

*P*212121 and *I*2 crystals of AtNF-YB2/NF-YC3 were cryo-protected in 30% PEG 3350, 0.1 M MES pH 6.5 and 25% ethylene glycol and 100 mM NaCl, 50 mM Tris-HCl pH 8.0, 2 mM DTT, 25% ethylene glycol, respectively. Cryo-protection of Ghd8/OsNF-YC7 crystals were carried out with mother liquor solutions supplemented with 10% glycerol. Crystals of the AtNF-Y/DNA complex were cryo-protected in 10% PEG 8000, 25% ethylene glycol, 0.1M MES pH 6.5. All crystals were then frozen in liquid nitrogen and tested at the ESRF synchrotron (Grenoble).

All collected data were reduced using XDS (Kabsch, 2010). Dimer data sets were scaled with AIMLESS, from the CCP4 package (Winn *et al.*, 2011), while AtNF-Y/DNA complex were anisotropically scaled with STARANISO (Tickle *et al.*, 2018). The three dimeric and the AtNF-Y/DNA structures were solved by molecular replacement using the coordinates of AtNF-YB6/NF-YC3 (PDB: 5G49) for the dimers and the mammalian NF-Y/DNA complex for the plant complex (PDB: 4AWL). Iterative cycles of model building with Coot (Emsley and Cowtan, 2004) and refinement with Phenix, Refmac5 and Buster (Adams *et al.*, 2010; Murshudov *et al.*, 1997; Bricogne *et al.*, 2010) were carried out to produce the final models (Figure S6). Stereochemical parameters of the final models were checked with Molprobity (Chen *et al.*, 2010). Data processing and refinement statistics are summarized in Table 1.

Electrophoretic Mobility Shift Assays

For dose-response EMSAs, DNA binding reactions were assembled with the indicated IMAC purified 6His-tagged dimers, and IMAC purified AtNF-YA6, AtNF-YA2 or CO (see Figure S4), in a binding mix containing the Cy5-labeled probe (20 nM) with the following final composition: 12 mM Tris-HCl pH 8.0, 50 mM NaCl, 50 mM KCl, 5 mM MgCl2, 0.5 mM EDTA, 12% glycerol, 0.2 mg/mL BSA, 2.5 mM DTT, additioned with 100 ng of poly (*dA:dT*). Binding reactions were incubated at 30 °C for 30', and separated by electrophoresis on 6% polyacrylamide gels in 0.25X TBE. Prior to reaction assembly, serial dilutions of proteins were prepared in Dilution Buffer (Buffer B additioned with 10% glycerol and 0.1 mg/ml BSA). The Cy5-labeled 31 bp *CAB2 CCAAT* probe ([Cy5]*CTTAAAATCCAATGAATGAACAGATAAAGAT*) derived from the Arabidopsis *CAB2* (*lhcb1*1*) (AT1G29920) promoter, the -5.3 kb *CCAAT* box *FT CCAAT* ([Cy5]*GCACTCATCCAATCCTTTATGGAATCTTCTT*) and the *FT* promoter *CORE2* ([Cy5]*AAGAAAAAGATTGTGGTTATGATTTCACCGA*) probes were previously described (Gnesutta et al, 2019, Gnesutta et al., 2017a; Cao et al 2014). Fluorescence gel images were

acquired and analysed with a ChemidocMP imaging system with the ImageLab Software (Bio-Rad Laboratories).

NOTE ADDED IN PROOF

During the reviewing process, Shen et al. (Plant Cell. 2020 Aug 25;tpc.00067.2020. doi: 10.1105/tpc.20.00067) reported on the structure of the HD1/GHD8/OsNF-YC2/CORE complex, detailing interactions largely anticipated and discussed in the present work.

DATA STATEMENT

Structural data that support the findings of this work have been deposited to Protein Data Bank with the following accession numbers: 6R0L (Ghd8/OsNF-YC7), 6R0M and 6R0N (At NF-YB2/NF-YC3), and 6R2V (At NF-YA6/NF-YB2/NF-YC3/*FT-CCAAT*).

ACKNOWLEDGEMENTS

We appreciate access to ID-29 and ID-30B beamlines at the ESRF (Grenoble, France). This work was supported by the Italian Association for Cancer Research (AIRC) Investigator Grants IG 2014 to M.N. [grant number IG-15267] and MIUR Progetto di Rilevante Interesse Nazionale (PRIN) 2017 grant no. 2017SBFHLH to R.M.

AUTHOR CONTRIBUTIONS

A.C.S., A.G. and N.G. purified the AtNF-Y trimer and the AtNF-Y dimers and mutants; D.M. produced the Ghd8/OsNF-YC7 dimer; A.C.S. and A.G. crystallized the proteins and prepared the atomic models. A.C.S. and M.N. analyzed the structures. A.B. and N.G. performed the EMSAs; A.C.S., F.F., M.N. and R.M. planned the experiments. A.C.S., M.N., and R.M. wrote the manuscript. All authors have read and approved the manuscript.

CONFLICTS OF INTEREST STATEMENT

The Authors declare no competing financial interests.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Structure-based sequence alignment and dimerization interface residues of the NF-YBs and the NF-YCs.

Figure S2. AtNF-Y/DNA contacts.

Figure S3. Structural comparison of *Arabidopsis* NF-Y/DNA complex with the mammalian and *Aspergillus* counterparts.

Figure S4. Recombinant proteins used in EMSA experiments.

Figure S5. Quantification of HDF trimerization mutants EMSAs.

Figure S6. Electron density at the HFD dimer and NF-Y-DNA interfaces.

REFERENCES

Adams, P.D., Afonine, P.V., Bunkóczi, G., et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.

Adrian, J., Farrona, S., Reimer, J.J., Albani, M.C., Coupland, G. and Turck, F. (2010) cis-Regulatory elements and chromatin state coordinately control temporal and spatial expression of FLOWERING LOCUS T in Arabidopsis. *Plant Cell*, **22**, 1425–1440.

Bernardini, A., Lorenzo, M., Nardini, M., Mantovani, R. and Gnesutta, N. (2019) The phosphorylatable Ser320 of NF-YA is involved in DNA binding of the NF-Y trimer. *FASEB J.*, **33**, 4790–4801.

Bricogne G, Blanc E, Brandl M, et al. (2010) *BUSTER*, Cambridge, United Kingdom: Global Phasing Ltd.

Calvenzani, V., Testoni, B., Gusmaroli, G., Lorenzo, M., Gnesutta, N., Petroni, K., Mantovani, R. and Tonelli, C. (2012) Interactions and CCAAT-binding of *Arabidopsis thaliana* NF-Y subunits. *PLoS ONE*, **7**, e42902.

Cao, S., Kumimoto, R.W., Gnesutta, N., Calogero, A.M., Mantovani, R. and Holt, B.F. (2014) A distal CCAAT/NUCLEAR FACTOR Y complex promotes chromatin looping at the FLOWERING LOCUS T promoter and regulates the timing of flowering in *Arabidopsis*. *Plant Cell*, **26**, 1009–1017.

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S. and Richardson, D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 12–21

Cockram, J., Thiel, T., Steuernagel, B., Stein, N., Taudien, S., Bailey, P.C. and O'Sullivan, D.M. (2012) Genome dynamics explain the evolution of flowering time CCT domain gene families in the Poaceae. *PLoS ONE*, **7**, e45307.

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

Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.

Fleming, J.D., Pavesi, G., Benatti, P., Imbriano, C., Mantovani, R. and Struhl, K. (2013) 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**, 1195–1209.

Gnesutta, N., Chiara, M., Bernardini, A., Balestra, M., Horner, D.S. and Mantovani, R. (2019) The Plant NF-Y DNA Matrix *In Vitro* and *In Vivo*. *Plants (Basel)*, **8**. https://www.mdpi.com/2223-7747/8/10/406

Gnesutta, N., Kumimoto, R.W., Swain, S., Chiara, M., Siriwardana, C., Horner, D.S., Holt, B.F. and Mantovani, R. (2017a) CONSTANS Imparts DNA Sequence Specificity to the Histone Fold NF-YB/NF-YC Dimer. *Plant Cell*, **29**, 1516–1532.

Gnesutta, N., Mantovani, R. and Fornara, F. (2018) Plant Flowering: Imposing DNA Specificity on Histone-Fold Subunits. *Trends Plant Sci.*, **23**, 293–301.

Gnesutta, N., Saad, D., Chaves-Sanjuan, A., Mantovani, R. and Nardini, M. (2017b) 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*, **10**, 645–648.

Goretti, D., Martignago, D., Landini, M., et al. (2017) Transcriptional and Post-transcriptional Mechanisms Limit Heading Date 1 (Hd1) Function to Adapt Rice to High Latitudes. *PLOS Genetics*, **13**, e1006530.

Gusmaroli, G., Tonelli, C. and Mantovani, R. (2001) Regulation of the CCAAT-Binding NF-Y subunits in *Arabidopsis thaliana*. *Gene*, **264**, 173–185.

Hackenberg, D., Wu, Y., Voigt, A., Adams, R., Schramm, P. and Grimm, B. (2012) Studies on differential nuclear translocation mechanism and assembly of the three subunits of the *Arabidopsis thaliana* transcription factor NF-Y. *Mol Plant*, **5**, 876-888.

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

Hyun, K., Jeon, J., Park, K. and Kim, J. (2017) Writing, erasing and reading histone lysine methylations. *Exp. Mol. Med.*, **49**, e324.

Kabsch, W. (2010) XDS. *Acta Crystallogr D Biol Crystallogr*, **66**, 125–132.

Kumimoto, R.W., Adam, L., Hymus, G.J., Repetti, P.P., Reuber, T.L., Marion, C.M., Hempel, F.D. and Ratcliffe, O.J. (2008) The Nuclear Factor Y subunits NF-YB2 and NF-YB3 play additive roles in the promotion of flowering by inductive long-day photoperiods in *Arabidopsis*. *Planta*, **228**, 709–723.

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

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

Levine, M., Cattoglio, C. and Tjian, R. (2014) Looping Back to Leap Forward: Transcription Enters a New Era. *Cell*, **157**, 13–25.

Lu, F., Liu, Y., Inoue, A., Suzuki, T., Zhao, K. and Zhang, Y. (2016) Establishing Chromatin Regulatory Landscape during Mouse Preimplantation Development. *Cell*, **165**, 1375–1388.

Luo, X., Gao, Z., Wang, Y., Chen, Z., Zhang, W., Huang, J., Yu, H. and He, Y. (2018) The NUCLEAR FACTOR-CONSTANS complex antagonizes Polycomb repression to de-repress FLOWERING LOCUS T expression in response to inductive long days in *Arabidopsis*. *Plant J.*, **95**, 17–29.

Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.*, **53**, 240–255.

Nardini, M., Gnesutta, N., Donati, G., et al. (2013) Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination. *Cell*, **152**, 132–143.

Nardone, V., Chaves-Sanjuan, A., Nardini, M. (2017) Structural determinants for NF-Y/DNA interaction at the CCAAT box. *Biochim. Biophys. Acta Gene Regul. Mech*., **1860,** 571–580.

Oldfield, A.J., Yang, P., Conway, A.E., Cinghu, S., Freudenberg, J.M., Yellaboina, S. and Jothi, R. (2014) Histone-fold domain protein NF-Y promotes chromatin accessibility for cell type-specific master transcription factors. *Mol. Cell*, **55**, 708–722.

Petroni, K., Kumimoto, R.W., Gnesutta, N., Calvenzani, V., Fornari, M., Tonelli, C., Holt, B.F. and Mantovani, R. (2012) The Promiscuous Life of Plant NUCLEAR FACTOR Y Transcription Factors. *Plant Cell*, **24**, 4777–4792.

Romier, C., Cocchiarella, F., Mantovani, R. and Moras, D. (2003) The NF-YB/NF-YC structure gives insight into DNA binding and transcription regulation by CCAAT factor NF-Y. *J. Biol. Chem.*, **278**, 1336–1345.

Siefers, N., Dang, K.K., Kumimoto, R.W., Bynum, W.E., Tayrose, G. and Holt, B.F. (2009) Tissue-Specific Expression Patterns of Arabidopsis NF-Y Transcription Factors Suggest Potential for Extensive Combinatorial Complexity. *Plant Physiol.*, **149**, 625–641.

Sinha, S., Kim, I.S., Sohn, K.Y., Crombrugghe, B. de and Maity, S.N. (1996) Three classes of mutations in the A subunit of the CCAAT-binding factor CBF delineate functional domains involved in the three-step assembly of the CBF-DNA complex. *Mol. Cell. Biol.*, **16**, 328–337.

Siriwardana, C.L., Gnesutta, N., Kumimoto, R.W., Jones, D.S., Myers, Z.A., Mantovani, R. and Holt, B.F. (2016) NUCLEAR FACTOR Y, Subunit A (NF-YA) Proteins Positively Regulate Flowering and Act Through FLOWERING LOCUS T. *PLoS Genet.*, **12**, e1006496.

Swain, S., Myers, Z.A., Siriwardana, C.L. and Holt, B.F. (2017) The multifaceted roles of NUCLEAR FACTOR-Y in *Arabidopsis thaliana* development and stress responses. *Biochim. Biophys. Acta Gene Regul. Mech.*, **1860**, 636–644.

Tao, Z., Shen, L., Gu, X., Wang, Y., Yu, H. and He, Y. (2017) Embryonic epigenetic reprogramming by a pioneer transcription factor in plants. *Nature*, **551**, 124–128.

Tickle, I.J., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C. and Bricogne, G. (2018) *STARANISO*, Cambridge, United Kingdom: Global Phasing Ltd.

Tiwari, S.B., Shen, Y., Chang, H.-C., et al. (2010) The flowering time regulator CONSTANS is recruited to the FLOWERING LOCUS T promoter via a unique cis-element. *New Phytol.*, **187**, 57–66.

Trigg, S.A., Garza, R.M., MacWilliams, A., et al. (2017) CrY2H-seq: a massively multiplexed assay for deep-coverage interactome mapping. *Nat. Methods*, **14**, 819–825.

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

Winn, M.D., Ballard, C.C., Cowtan, K.D., et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 235–242.

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

Yun, J., Chae, H.-D., Choi, T.-S., Kim, E.-H., Bang, Y.-J., Chung, J., Choi, K.-S., Mantovani, R. and Shin, D.Y. (2003) Cdk2-dependent phosphorylation of the NF-Y transcription factor and its involvement in the p53-p21 signaling pathway. *J. Biol. Chem.*, **278**, 36966–36972.

Zanetti, M.E., Rípodas, C., Niebel, A. (2017) Plant NF-Y transcription factors: Key players in plant-microbe interactions, root development and adaptation to stress. *Biochim. Biophys. Acta Gene Regul. Mech.*, **1860,** 645-654.

Zaret, K.S. and Carroll, J.S. (2011) Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.*, **25**, 2227–2241.

FIGURES LEGENDS

Figure 1. Structure of plant NF-YB/NF-YC HFD dimers. (**a**) Ribbon diagram showing the HFD dimer of AtNF-YB2/NF-YC3 (PDB-code 6R0N), Ghd8/OsNF-YC7 (PDB-code 6R0L), and comparison with L1L(AtNF-YB6)/AtNF-YC3 (PDB-code 5G49). (**b**) The electrostatic surface of the three plant HFD heterodimers is shown in the same orientation as panel (**a**). Blue and red colors indicate positively and negatively charged surface regions, corresponding to the DNA and NF-YA binding regions, respectively. The N-terminal parts of the NF-YB and NF-YC subunits were omitted in panel (**b**) to facilitate the electrostatic surface comparison.

Figure 2. Structure of the plant NF-Y/DNA complex. (**a**) Ribbon representations of plant AtNF-YA6/NF-YB2/NF-YC3 heterotrimer in complex with the *FT* 25 bp-*CCAAT* box oligonucleotide. Secondary structure elements of AtNF-YA6 are labeled and the *CCAAT* nucleotides are highlighted in magenta.

Figure 3. Structure-based sequence alignment of NF-YAs and CCTs. Structure-based sequence alignment of AtNF-YA6 core domain with mammalian (NF-YA) and *Aspergillus* (AnHapB) NF-YAs, AtNF-YA2, and with five representative *Arabidopsis* CCT domain proteins of each sub-family. Residues at relevant positions for trimerization are highlighted depending on their role: in blue and cyan Arg and Lys residues, in red those not-involved in the interaction. Residues involved in *CCAAT* box recognition and phosphorylatable residues are displayed in orange and yellow, respectively. The phosphorylatable residues contacting DNA backbone are highlighted in yellow.

Figure 4. AtNF-YA6-DNA interactions at the *CCAAT* box. AtNF-YA6 residues (green) from the A2 helix and the Gly-rich loop contacting nucleotide bases at the *CCAAT* box. Nucleotides and sequence-specific residue side chains were represented as sticks. In the upper panel, AtNF-YA6 interactions with the *CCAAT* strand, colored in violet. The *CCAAT* complementary DNA strand was omitted to facilitate the view. In the lower panel, AtNF-YA6 contacts with the complementary strand, colored in yellow. Residues and nucleotides were labelled and H-bonds were shown with dashed lines. The H-bond between the phosphorylatable Thr229 and the DNA phosphate backbone is also shown.

Figure 5. Trimerization interactions between AtNF-YA6 and the AtNF-YB2/NF-YC3 dimer. Detailed view of the interaction network within the AtNF-Y trimeric complex. AtNF-YA6, AtNF-YB2, and AtNF-YC3 are shown in green, yellow, and cyan, respectively. Positively charged residues contacting acidic residues of the HFD, as well as the substituted key residues in mammalian and *Aspergillus* NF-Y, are represented as sticks. The main chains are displayed as a semitransparent ribbon. Salt bridges and H-bonds are marked as dashed lines. Residues mutated in this work are highlighted with red labels.

Figure 6. DNA-binding of HFD trimerization mutants. The AtNF-YB2 E65R mutant **(a)**, or the AtNF-YC3-D83R-D85R double mutant **(b)**, co-purified as dimers with the relevant AtNF-YC3 or AtNF-YB2 wild-type partner, were tested in dose-response EMSAs. The *FT CCAAT* (top panels) or *CAB2 CCAAT* (bottom panels) probes were used to assess trimerization and DNA-binding of AtNF-Y mutants; in parallel, the *FT* CORE2 probe was used to assess NF-CO trimerization. The indicated HFD mutant dimers (mut; 40 nM) were incubated with increasing amounts of AtNF-YA2, AtNF-YA6 or CO (60, 120, 240 nM), together with the relevant probe (20 nM). WT dimers were used as positive controls. Single subunits were incubated alone with the probe (-), at the highest concentration (240 nM) for AtNF-YAs and CO. NF-Y or NF-CO/DNA shifted complexes are indicated by closed, or open, arrowheads. fp: free probe. In (a) an asterisk denotes a non-specific DNA(CAB2 CCAAT) complex observed with the mutant dimers alone. NF-CO EMSAs shown in top and bottom gels in each panel represent equivalent independent DNA binding experiments performed in parallel with NF-Y reactions loaded on the same gel.

Figure 7. NF-YC N-terminus structural comparison. Comparison between the NF-YC N-terminal region within **(a)** the AtNF-Y/DNA (AtNF-YA6/NF-YB2/NF-YC3/DNA) complex, **(b)** the isolated AtNF-YB2/NF-YC3 dimer, **(c)** the isolated L1L/AtNF-YC3 dimer, and **(d)** the AnHapB/C/E/DNA complex. DNAs were represented as ribbon and proteins were represented as a grey surface, with the exception of the NF-YCs N-termini, highlighted as semitransparent colored surface (AtNF-YC3 in cyan and AnHapE in red). The conserved His and Lys were represented as spheres. Protein-DNA interactions and residue labels were included in the inlets.

Table 1: Diffraction data collection and refinement statistics

Data set	Ghd8/OsNF-YC7	At NF-YB2/	At NF-YB2/	At NF-YA6/

	(PDB: 6R0L)	**NF-YC3** (PDB: 6R0M)	**NF-YC3** (PDB: 6R0N)	**NF-YB2/NF-YC3/** ***FT-CCAAT*** (PDB: 6R2V)
Data Collection				
ESRF beamline	ID-30A	ID-30B	ID-29	ID-29
Space Group	*I*2	$P2_12_12_1$	*I*2	$P2_12_12_1$
Cell dimensions				
a, b, c (Å)	65.84, 34.10, 67.08	70.02, 74.98, 86.65	37.19, 57.42 97.83	53.45, 85.53, 130.31
α, β, γ (º)	90.0, 95.4, 90.0	90.0, 90.0, 90.0	90.0, 99.4, 90.0	90.0, 90.0, 90.0
Wavelength (Å)	0.967700	1.010876	0.999996	0.999998
Resolution range (Å)				
upper limit along orthogonal reciprocal axis	2.70, 2.70, 2.70 (2.84-2.70)	2.30, 2.30, 2.30 (2.38-2.30)	2.10, 2.10, 2.10 (2.18-2.10)	3.21, 2.50, 3.21 [3.26-3.21] (2.90-2.50)‡
lower limit	44.73	44.07	49.35	41.32
#R_{pim}	0.104 (0.257)	0.035 (0.215)	0.018 (0.351)	0.079 [0.315] (0.484)‡
+$CC_{1/2}$	0.970 (0.867)	0.998 (0.575)	1.000 (0.860)	0.990 [0.847] (0.678)‡
$\langle I/\sigma(I)\rangle$	8.5 (5.2)	12.5 (3.6)	13.9 (2.3)	7.6 [2.6] (1.7)‡
Redundancy	12.2 (12.3)	12.7 (13.5)	6.7 (7.0)	7.7 [7.9] (8.2)‡
Completeness (%)				
spherical	99.8 (99.8)	99.9 (100.0)	99.3 (98.8)	56.1 [100] (8.5) ‡
ellipsoidal				90.6 [100] (62.4) ‡
Refinement				
Resolution (Å)	44.73-2.70 (2.80-2.70)	44.07-2.30 (2.40-2.30)	49.40-2.10 (2.18-2.10)	41.32- 2.50 (2.75-2.50)
Number of reflections	4172 (424)	20843 (2054)	11899 (1183)	11940 (271)
R_{work}/R_{free}	0.2530.268 (0.299/0.318)	0.204/0.239 (0.317/0.332)	0.193/0.207 (0.333/0.362)	0.200/0.247 (0.353/0.327)
Number of molecules in the AU	1	2	1	1
Protein residues	165	351	171	250
Nucleotides				50
GOL molecules	2		3	
PO4 molecules				
Water molecules	42	120	58	140
Average B factors	19.93	63.91	70.11	54.69

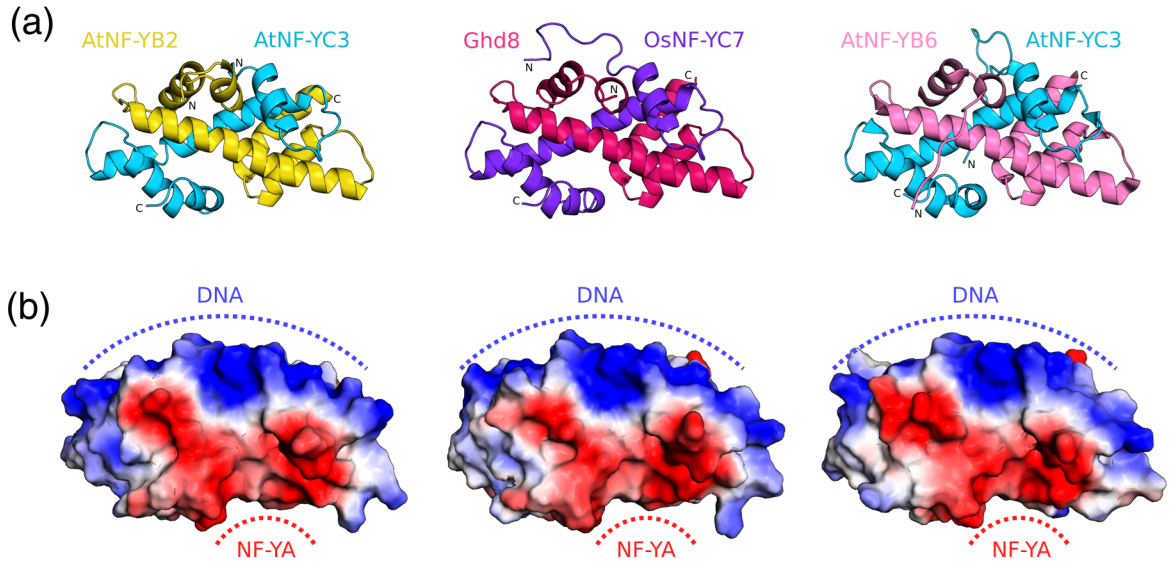
(Å^2)				
Rmsd				
Bond lengths (Å)	0.012	0.012	0.014	0.007
Bond angles (°)	1.46	1.17	1.35	0.89
Ramachandran plot statistics	95 % in favoured	99 % in favoured	97 % in favoured	97 % in favoured
	0 % outliers	0 % outliers	0 % outliers	0 % outliers

The highest-resolution shell is shown in parentheses. The highest-resolution shell for the 6R2V data, truncated to 3.21 Å, is shown in brackets.

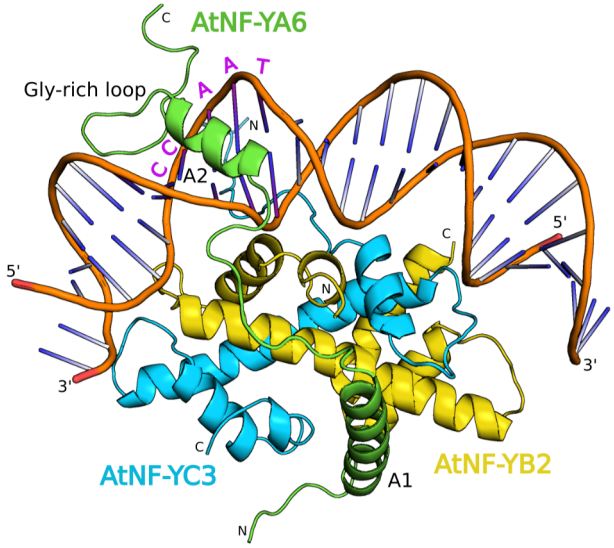
‡Anisotropicallly truncated data

$^{\#}R_{p.i.m.} = \sum_{hkl}\sqrt{1/n-1}\sum_{j=1}^{n}\left|I_{hkl}-\langle I_{hkl}\rangle\right| \Big/ \sum_{hkl}\sum_{j} I_{hkl,j}$

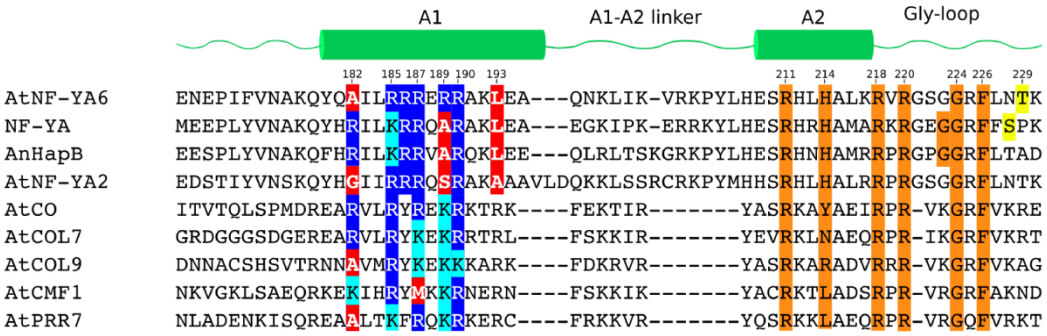
$^{+}CC_{1/2}$ is the correlation coefficient of the mean intensities between two random half-sets of data.



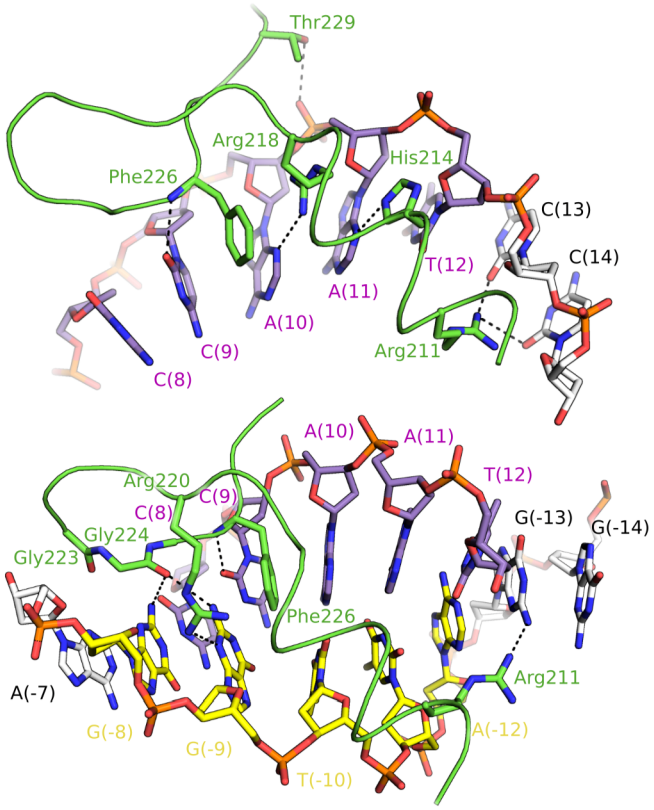
tpj_15038_f1.tiff



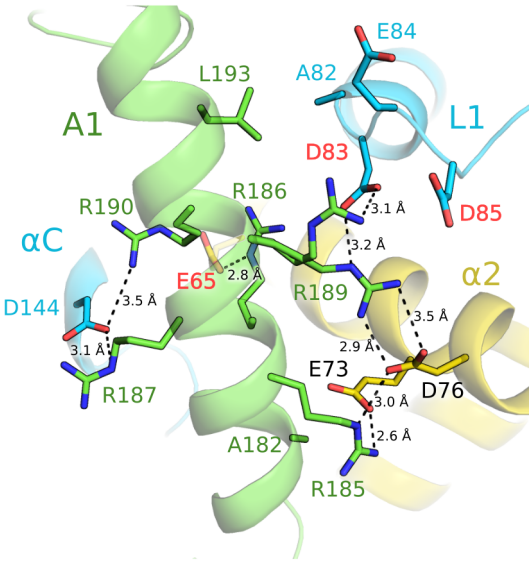
tpj_15038_f2.tiff



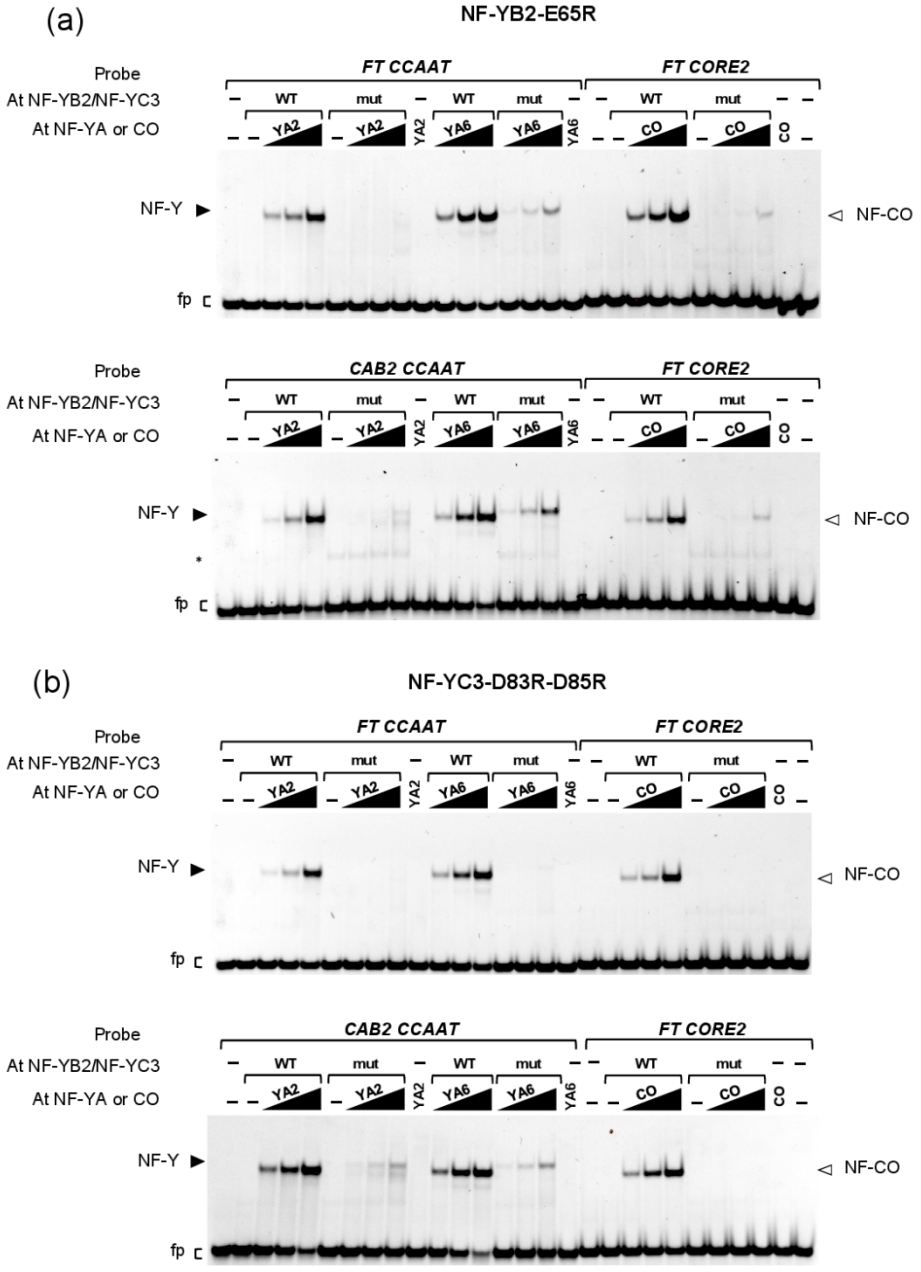
tpj_15038_f3.tiff



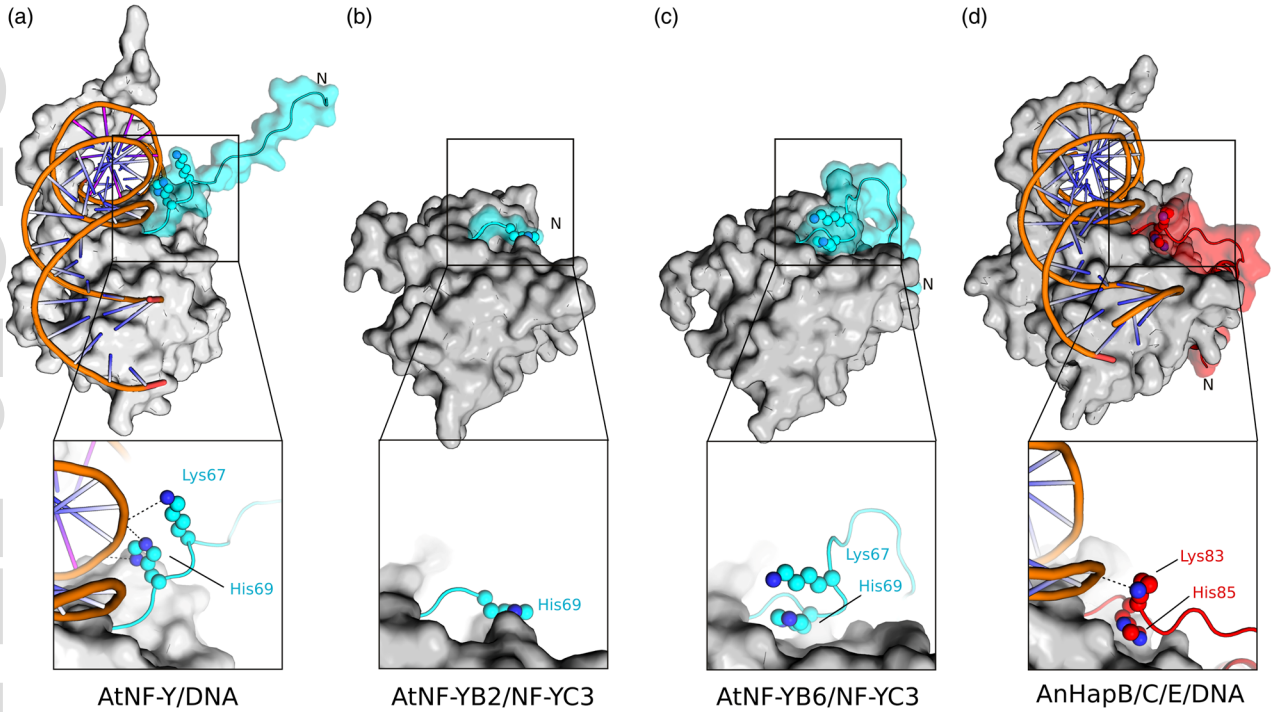
tpj_15038_f4.tiff



tpj_15038_f5.png



tpj_15038_f6.tiff



tpj_15038_f7.tiff