UNIVERSITÀ DEGLI STUDI DI MILANO

PhD Course in Molecular and Cellular Biology

XXX Ciclo

MODULATION OF DNA-BINDING IN TRANSCRIPTION FACTORS: THE CASE OF NF-Y AND USF1

Andrea Bernardini

PhD Thesis

Scientific tutor: Dr. Nerina Gnesutta

Academic year: 2016-2017

SSD: BIO/10; BIO/18

Thesis performed at: Dipartimento di Bioscienze, Università degli Studi di Milano

Contents

SUMMARY

Transcription is ultimately driven by DNA-binding events involving regulatory proteins known as transcription factors (TFs). NF-Y is a ubiquitous eukaryotic TF composed by three different subunits, all necessary to specifically recognize and bind CCAAT-box DNA. NF-Y plays a central role in shaping the architecture of promoters and distal regulatory regions, both by influencing local chromatin environment and working as a pioneering binding-platform for several TFs in different genomic contexts. Still, the way NF-Y activity itself is modulated, and the molecular determinants at the basis of NF-Y function as transcriptional organizer pose many open questions.

The aims of my doctoral thesis were, in general, to provide new structure-function correlations of NF-Y activity, and in specific to (I) uncover the molecular role of two NF-YA subunit phosphorylation sites and (II) dissect the molecular interplay between NF-Y and the E-box TF USF1 on different DNA configurations.

NF-YA was shown to be a cell-cycle modulated phospho-protein, with two C-terminal residues (Ser320 and Ser326), located close to the conserved DNA-binding domain of the subunit, targeted as phosphorylation sites by CDK2. Still, their molecular function remained elusive. We designed a panel of NF-YA Ser320 and Ser326 phospho-ablative (Ser to Ala) and phospho-mimicking (Ser to Glu) mutants, in order to assess the phosphorylation state of the protein *in vivo*, and to characterize their behaviour in DNA-binding and transcriptional assays. First, we detected Ser320-P as the major NF-YA phospho-isotype in asynchronous HeLa cells, strongly increasing upon nocodazole-induced mitotic arrest. Second, we found that while none of the phospho-mutant recombinant proteins affected the overall affinity of the NF-Y heterotrimer for DNA, Ser320 (but not Ser326) mutations kinetically destabilized the DNA-bound complex, greatly increasing its off-rate. Third, Ser320 phospho-mutants impaired NF-Y dependent transcriptional activation of the *RHOB* and *MDR1* target promoters in reporter assays. Taken together, our results point at Ser320-phosphorylation as a major regulatory switch for NF-Y activity, directly affecting its residence time on DNA.

In the second project we aimed to dissect the USF1 protein regions which mediate DNA-binding cooperativity with NF-Y, in order to uncover the molecular bases of one of the genome-wide NF-Y-directed regulatory modules. We detected transcriptional synergy in reporter assays using four distinct target promoters characterized by the recurrent 10-12 bp spacing between E-box and CCAAT-box elements. Moreover, we identified the USR, an intrinsically disordered domain of USF1, as the region responsible for cooperative interaction with NF-Y on DNA. Notably, depending on the underlying DNA configuration of the binding sites, distinct portions of this context-specific transactivation domain mediate the protein-protein interactions.

PART I

1.INTRODUCTION

In living organisms as we know them today, cell identity and function are the product of several layers of regulation, now defined by holistic terms, as genome, transcriptome or proteome, and their respective modifications. The genetic information stored in the form of DNA is replicated in a highly efficient and tightly regulated manner by elegant protein machineries, themselves encoded by the same genetic material they are specialized to copy. The very first step by which the genetic information is 'read' and decoded by living cells is transcription: the process in which the genetic information stored in a DNA strand is converted into complementary RNA copies. Transcription gives rise to a wide spectrum of RNA classes, some of which harbour coding sequences that are in turn decoded by the protein synthesis apparatus through the process of translation. This second fundamental step is the source of every protein a cell can build, thus linking through an RNA intermediate the genetic information stored in DNA with protein machineries responsible for the majority of biochemical processes within a living cell.

The entire RNA species content of a cell at a given moment is defined as its transcriptome, a dynamic and heterogeneous layer of control of cell morphology, behaviour and differentiation. Besides the ordered information flux that characterize gene expression, transcriptome actually carries out functions both related to DNA as 'information vehicle', and to proteins as cellular 'effector machineries'. As a matter of fact, within the multitude of RNA species that make up a transcriptome we find molecules specialized to work as information vectors for protein instructions, some that carry out structural functions as flexible scaffolds, others use the information they contain to behave as guides for distinct molecular machineries, and some others are designed to work as biological catalysers.

The crucial role of the transcriptome in a living cell demands an effective control system to manage simultaneously the activity of tens of thousands of genes, and cells have evolved a vast molecular toolkit to achieve this enormously complicated, yet vital task. Indeed, the study of both the molecular details and the interactive networks that characterize gene regulation is fascinating and would allow us to understand how living cells really work at the most basal level. Moreover, it would provide us with an immense range of new opportunities to effectively and precisely prevent or treat the multitude of diseases characterized by deregulated gene expression [1].

1.1 Basic Principles of Transcription

The process of transcription constitutes the first key step in the biological information flow known as Central Dogma of molecular biology: the genetic information stored in DNA is transcribed into RNA intermediates and translated into proteins. Transcription is the synthesis of RNA using DNA as template. This reaction is catalysed by specialised enzymes called RNA polymerases, whose enzymatic activity was first discovered by Weiss and Gladstone (1959).

Classically, a transcription cycle is composed of three main phases: initiation, elongation and termination, each of them subjected to specific regulation. Transcription occurs on a double-stranded DNA molecule which gets partially unwound (12-14 bp) in an ATP-dependent manner within the RNA polymerase complex, resulting in exposed DNA nucleobases. RNA synthesis then proceeds through a series of nucleotide addition cycles, in which the RNA polymerase presents a DNA base belonging to the 'template strand', sterically selects a specific complementary nucleoside triphosphate (NTP) in its active site and catalyzes the formation of a phosphodiester bond between the 3'-OH of the previous nucleotide and the 5'-α-phosphate of the incoming NTP, releasing inorganic pyrophosphate (iPP) (**Fig 1A**). During elongation of the growing RNA molecule, RNA polymerase moves through the template DNA along with the translocating transcription bubble, while displacing the transient DNA/RNA heteroduplex (~ 8 nt) as new NTPs are incorporated [2].

Transcription elongation is a very dynamic process which proceeds with rates of ~ 6-70 nt/sec, involving a plethora of associated proteins which carry out several co-transcriptional functions. In eukaryotes RNA polymerase has to work in a chromatin-dominated DNA landscape, in which nucleosomes render DNA less accessible to the elongating polymerase. In spite of the energetic barrier imposed by nucleosomes in front of the transcription machinery, the average transcription rate measured *in vivo* on transcribed genes closely matches the transcription rate on nucleosome-free DNA *in vitro* [3][4][5][6]. In cells, indeed, RNA polymerase is associated with a series of elongation factors, including histone chaperones, chromatin remodelers and modifiers that assist its progression through arrays of nucleosomes [7].

Eukaryotes, distinctly from bacteria and archea, have acquired a higher degree of specialization for the enzymes that transcribe their genes: they are equipped with three different nuclear RNA polymerases specialized to transcribe distinct classes of genes (see **Table 1**).

RNA polymerase	Localization	Cellular transcripts	Subunits
I	nucleolus	28S, 18S, 5.8S rRNA	14
II	nucleoplasm	mRNA, snRNA, microRNA, lncRNA	12
III	nucleoplasm	tRNA, 5S rRNA, snRNA	17

Table 1. Eukaryotic RNA polymerases localization, dedicated classes of transcribed RNAs and number of subunits.

Despite the sharp differences in gene classes they transcribe, eukaryotic RNA polymerases are big multiprotein complexes that share a conserved core composed of ten subunits (**Fig 1B**). RNA polymerase I (Pol I) and III (Pol III) are devoted to rather housekeeping functions mainly linked to protein synthesis, transcribing the three major classes of ribosomal RNAs (rRNA) and transfer RNAs (tRNA), respectively. RNA polymerase II (Pol II) instead is responsible for transcribing the most dynamic classes of transcripts, mainly protein coding RNAs (mRNA), but including also microRNAs and long non-coding RNAs (lncRNA).

While in prokaryotes transcription is tightly coupled with translation, such that nascent mRNAs start to be translated by ribosomes way before they are completely transcribed, in eukaryotes these two fundamental steps of the Central Dogma (transcription and translation) are kept separated by the nuclear envelope. Moreover, in the eukaryotic system the nascent pre-mRNA undergoes a complex series of co-transcriptional processes that drive its maturation, including 7-methylguanosine 5'-capping, splicing, 3'-polyadenylation and specific types of RNA nucleobase editing. The majority of the co-transcriptional RNA processing machineries are tethered to elongating Pol II through its major subunit carboxy-terminal domain (CTD) (**Fig 1C**). This extended tail is composed by heptad repeats and works as a 'landing platform' that selectively recruits processing factors depending on its phosphorylation state, which in turn is modulated coherently with the progression of the transcription cycle [8].

Co-transcriptional processing of new transcripts is crucial to determine RNA fate, including stability, transport and determination of splicing variants. This regulation level represents a key process to shape the transcriptome of a cell. This is not a trivial task, for instance at least ~95% of human multi-exon genes can undergo alternative splicing [9], with an average of 6.3 isoforms per gene [10], although in each given condition one isoform dominates over the others, usually taking over a large fraction of the total gene output [11][12]. However, the principal step by which a cell manage to regulate gene expression is by controlling transcription initiation. In the next paragraph the main challenges a cell has to face to tightly control transcription initiation are discussed.

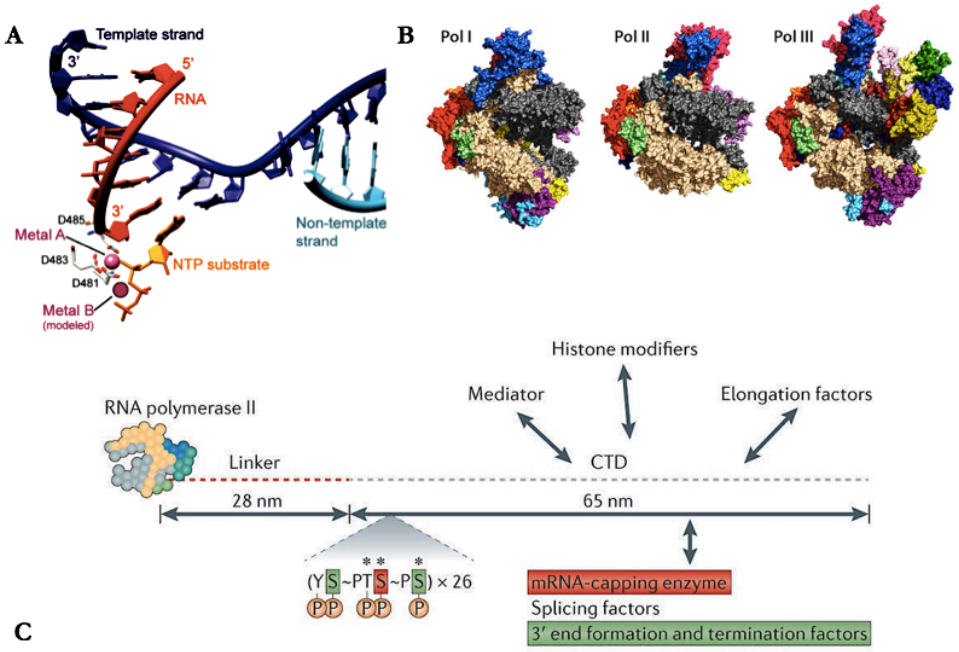


Figure 1. (**A**) Detailed view of Pol II active center occupied by the template DNA and the nascent RNA. Metal A (Mg^{2+}) is permanently bound by three conserved aspartate residues, while Metal B is recruited during catalysis (from [2]). (**B**) Surface view of the three eukaryotic RNA polymerases. Homologous subunits among the three complexes have the same color (from [13]). (**C**) The yeast Pol II C-terminal domain (CTD) constituted of 26 heptad repeats is shown to scale with the core-structure of the complex. Human CTD include 52 heptad repeats. RNA processing and transcription elongation complexes that are recruited by the CTD are indicated (from [8]).

1.2 The Complexity of Transcription Initiation Control

In living cells RNA polymerases are unable to start a productive transcription cycle without assistance by other specialized members of the transcription machinery. This is actually a crucial feature, common to all three domains of life (*Bacteria*, *Archea* and *Eukarya*) to orchestrate the activity of their genetic information. Gene transcription has to be highly responsive to the demanding environmental/developmental conditions of cellular life, entailing the necessity for specific guidance by secondary factors. Indeed, a hypothetical living system where RNA polymerase was intrinsically capable to give rise to any transcript without directed assistance would have been very costly in energetic terms. Exceptions to this living world paradigm come from some viruses, which evolved very specialized, single-polypeptide RNA polymerases capable to start transcription at viral promoters without the aid of any other viral or host-cell specificity proteins [14], such as RNA polymerases from bacteriophages T7, T3 and SP6 [15].

The fundamental unit of transcriptional regulation is itself encoded in the genome in the form of specific recognition elements (REs), short stretches of DNA directly recognized and bound by specialized classes of sequence-specific DNA-binding proteins, known as transcription factors (TFs). At the most basic level, transcription initiation of a given gene can be achieved when

RNA polymerase is recruited by TFs bound to the REs in the promoter region of the correspondent gene (discussed in section 1.3). This event is necessary both for the precise positioning of RNA polymerase on the correct transcription start site (TSS) and the initial DNA melting to generate a stable transcription bubble. The general concept of guided RNA polymerase positioning on the promoter regions by transcription factors applies both to prokaryotic and eukaryotic world, although they substantially differ by some orders of magnitude in terms of the molecular complexity they deploy.

The problem of specificity in gene regulation increases with the increasing size of eukaryotic genomes and their underlying complexity, since the number of stochastic nucleotide combinations giving rise to REs in non-functional regions increases along with genome size. Moreover, the more genes a genome contains the more demanding is their coordinated regulation in response to cell-to-cell signalling and environmental stimuli.

An additional hugely complicated task emerged with the birth of multicellular life: that is development. More specifically, what it demands in terms of gene regulation. The development of an organism is intimately dependent on the capacity of its cells to differentiate in order to acquire specialized functions, at the cost of cellular potency. Cell differentiation is driven by dramatic changes in gene expression at the most basic, transcriptional level [16][17]. Since differentiation is tightly coupled with cell division during development, it is crucial for daughter cells to selectively inherit (or not) the transcriptional arrangement of the parent cell to retain lineage-specific expression patterns.

Eukaryotic cells evolved elegant mechanisms to stably manage and maintain their transcription patterns throughout the dramatic events of DNA replication and cell division, and to cope with their higher genetic complexity.

1.2.2 Chromatin pervasively affects transcription

The first, and drastic, eukaryotic adaptation in DNA metabolism is the protein-complexed form in which DNA is constantly found in eukaryotic cells, named chromatin (Flemming, 1879). Distinctly from its 'naked' prokaryotic counterpart, eukaryotic DNA is intimately associated with nuclear proteins that make up roughly two-thirds of the chromatin content of the nucleus. Chromatin fibres are composed by a repeating substructure of basic units termed nucleosomes [18][19]. Each core nucleosome includes about 147 bp of DNA wrapped around a compact protein core for almost 1.7 left-handed negative superhelical turns, effectively compacting DNA about 6-fold its linear dimension. The histones that compose the nucleosome core are a family of small, basic proteins named H2A, H2B, H3 and H4. They associate forming an octamer constituted by a $(H3/H4)_2$ tetramer and two H2A/H2B dimers [20], presenting an extended positively-charged

interface for DNA binding (**Fig 2**). Core histone proteins are characterized by a conserved Histone-Fold Domain (HFD) composed by a long central α-helix (α2) separated from two short flanking α-helices (α1 and α3) by two loops (L1 and L2) [21]. This domain allows the formation of tight dimers through extensive hydrophobic interactions between each HFD in a 'handshake' fashion. Moreover, individual core nucleosomes are separated by short stretches of linker DNA of variable length (~35 bp in a typical vertebrate nucleus).

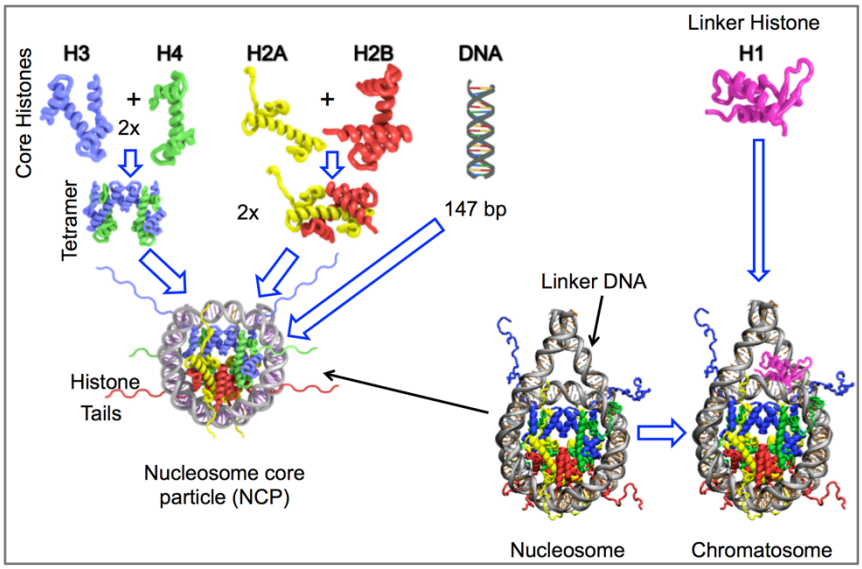


Figure 2. Schematic representation of the assembly of the core histones into the nucleosome, including linker histone H1 forming the so-called chromatosome (image from the background info section of HistoneDB 2.0 database [22]).

The origin of chromatin is still a matter of debate, since there are strong and recent evidences that eukaryotic histones have structural homology with prokaryotic counterparts found in most *Archea*, which would predate the classic octameric core nucleosome [23][24].

Despite its origins, chromatin allowed eukaryotic genomes to expand in size and complexity thanks to two main structural outcomes: the ordered packaging of DNA molecules in space (1) and the establishment of a restrictive "ground state" for transcription (2). The first outcome is quite evident by the fact that a functional higher diploid eukaryotic genome is highly compacted to fit within a defined volume (the nucleus) comparable in size with an entire bacterial cell, yet being more than 1000 times longer than a typical prokaryotic genome. Yet the real task performed by histones is to balance the charge of DNA phosphate backbone and effectively reduce its apparent length in a highly organized fashion. The astonishing DNA compaction achieved by eukaryotic cells (more than 300,000-fold in humans) is achieved by a series of hierarchical higher order chromatin coils and loops, whose structural identity is yet unclear, especially at the 'middle-scale' level between nucleosome arrays and chromosomal domains [25][26][27]. The real complexity concerning chromatin-based genomes goes far beyond the extreme DNA compaction level, it

entails a wide range of regulatory strategies, from highly dynamic gene induction to almost permanent silencing. In the last few years smart technical innovations based on 3C (chromosome conformation capture) coupled with deep sequencing (i.e. 5C, Hi-C and ChIA-PET) allowed unprecedented high-resolution mapping of the 3D structure of genomes, thus prompting exciting correlation studies with gene expression and TFs occupancy [28].

The second basic, yet pervasive function of chromatin is to limit genome accessibility to the transcription apparatus. In general, chromatin has a repressive effect on transcription by impeding the access of macromolecular complexes to DNA, including hiding specific REs against core-histones surface and hindering the initiation or elongation of RNA polymerase through an array of nucleosomes [29][30][31]. The important consequence of this restrictive transcriptional "ground state" is the relieving of the need for active repression, as it is suggested by the sheer prevalence of transcriptional activators over repressors in eukaryotes [32]. Indeed, all genes contain REs for activators, but only some for negative regulators, which are especially used by multicellular eukaryotes to restrict gene expression within different developmental domains or to modulate transcription in response to changes in the microenvironment of the cell [33]. Moreover, the vast majority of active repressing factors in eukaryotes tend to interact with chromatin to exert their function, often resulting in chromatin modification and condensation [34]. The setup of an inherently restrictive substrate for transcription adds additional layers of regulation to control gene expression: for instance, chromatin addresses one of the major issues concerning big genomes with many genes and a large portion of non-coding DNA, that is preventing extensive random transcription driven by interspersed REs or cryptic promoters. Despite the surprisingly increasing evidence for pervasive transcription of the genome both in yeast and mammalian cells [35][11], this inherent "protective" function of chromatin is supported by studies that demonstrate that alterations in nucleosome positioning functions increase spurious transcription output [36][37].

Chromatin stands at the basement of epigenetics, term coined by Waddington in 1942 to describe the phenomena that 'stand on top' of a genotype leading to the respective phenotype, and currently defined as the study of heritable changes in gene function occurring without any changes in the underlying DNA sequence (**Fig 3**). As mentioned above, one of the main challenges in a multicellular organism is coupling differentiation programmes with the maintenance of gene expression profile upon cell division. In this scenario, chromatin works as a semi-stable "memory device" that contributes to define cell identity [38]. This is carried out by a complex molecular toolkit, comprising more than 100 enzymes that transmit to daughter cells a variety of epigenetics marks previously established in their cell lineage [39]. Epigenetic marks include chemical modifications of DNA (methylation and hydroxymethylation) and histones (more than ten different chemical post-translational modifications, including acetylation, methylation,

ubiquitination and phosphorylation), the deposition of specific histone variants (e.g. H3.3 in transcribing regions, H2AZ flanking the nucleosome-free region at poised promoters) and boundaries of maintenance of specific chromatin domains (CTCF binding) [40][41]. These 'gene bookmarking' systems are at the basis of maintenance of cellular memory.

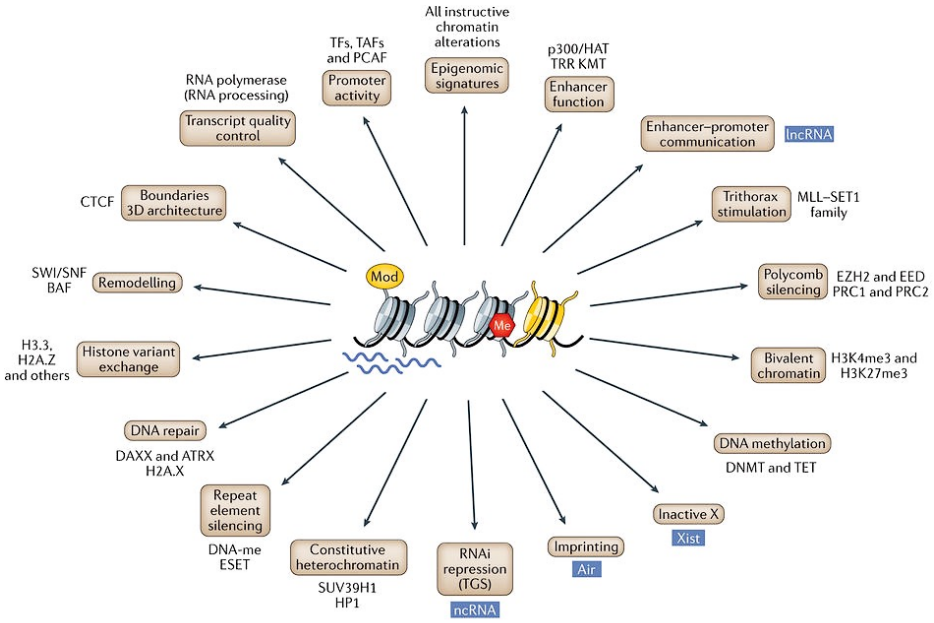


Figure 3. Multiple examples of the central role of chromatin in DNA metabolism, including transcription, replication and repair (from [40]).

Since the publication of the first high-resolution structure of the nucleosome in 1997 [42], the research field in chromatin regulation has exploded both from a structural and functional perspective, especially thanks to the introduction of next generation sequencing (NGS) technologies coupled with potent and selective molecular biology methods to enrich for several chromatin statuses (i.e. nucleosome occupancy, chromatin accessibility, epigenetic marks localization, specific DNA-bound protein enrichment, detection of chromatin-associated non-coding RNAs, mapping of 3D chromatin organization and long-range interactions and others) [43]. Yet the real challenge will be to integrate these immense datasets to have a clearer global picture of how cells manage to control gene expression throughout the epigenetic landscape. The next section will cover the second adaptation that eukaryotic cells adopted to control transcription, the use of vast combinations of sequence specific TFs binding at gene regulatory regions, thus introducing the central topic of this thesis.

1.3 Transcription factors as ultimate controllers of gene expression

The second fundamental adaptation eukaryotic cells have adopted to cope with their increasingly complex genomes was to make their regulatory regions themselves increasingly complex. The human genome, for instance, contains about 1550 genes for sequence specific transcription factors (TFs) [44], needed to regulate 19-20,000 protein-coding genes (and nearly as many of non-coding genes) [45]. The striking observation is that the fraction of sequence specific TF-coding genes in humans is comparable to that in bacteria (5-7%) [46]. Indeed, including splicing variants we obtain more than 2900 different isoforms for human TFs, however the ultimate difference with bacteria is the so called 'combinatorial control of gene regulation', which hugely increases the potential number of unique expression profiles, while reducing robust non-specific transcription initiation events. In this section, the most relevant features of eukaryotic TFs DNA recognition and mode of action are covered.

1.3.1 Core-promoter complexity

Promoters are regulatory DNA regions located at the 5'-end of the gene body directly necessary, but not always sufficient, to initiate transcription of a given gene. Eukaryotic promoters are traditionally composed by a 'core-promoter' and a 'proximal-promoter' depending on the distance from transcription start site (TSS) and their binding partners. The core-promoter is a region of 70-80 bp that includes the TSS, and owns canonical DNA elements sufficient for recognition by key proteins that make up the so called basal transcriptional machinery (**Table 2**).

Core element	Position	Consensus	Frequency		Bound protein
			Flies	Vertebrates	
TATA	-31 to -26	TATAWAAR	33-43%	10-16%	TBP
Inr	-2 to +4	YYANWYY	69%	55%	TAF1/TAF2
DPE	+28 to +32	RGWYV	40%	48%	TAF6/TAF9
BRE^u	-37 to -32	SSRCGCC	-	12-62%	TFIIB
MTE	+18 to +29	CSARCSSAACGS	8.5%	-	TAF6/TAF9
DCE	+6 to +11 +16 to +21 +30 to +34	S_I: CTTC S_{II}: CTGT S_{III}: AGC	-	-	TAF1

Table 2. Features of canonical core-promoter elements, including preferential position from the TSS (+1), consensus sequence in nucleotide IUPAC code, relative frequency in promoters and the effector proteins that recognize each element (adapted from [47]).

It is usually defined as the minimal DNA region sufficient to promote low levels of basal (activator-independent) transcription by Pol II *in vitro*. Although the elements composing the Pol II core-promoter were characterized by biochemical experiments on isolated systems [48], their

simultaneous presence was not confirmed for the vast majority of eukaryotic promoters. For instance, the TATA-box, that was originally considered a universal core-promoter element, was then only found in a minority of metazoan promoters (8-30%) [49] and even less in plants [50]. There is a wide heterogeneity in core-promoter composition which underlies the diversity of promoter regulation and may reflect the assembly of distinct components of transcriptional machinery, therefore the concept of a universal core-promoter clearly does not fit with the genomic evidence. Moreover, most mammalian genes lack canonical core-promoter elements [51] in favour of relatively less characterized non-canonical elements. Recent studies clearly show that more than half of human protein-coding genes promoters are associated with CpG islands [52], DNA regions with a relatively high-density of hypomethylated CpG dinucleotides spanning 0.5-2 kb.

The development of genome-wide transcript mapping techniques with high-sensitivity (i.e. CAGE, small RNA sequencing, GRO-seq) unravelled unprecedented levels of complexity in core-promoter function [53]. First, the concept that transcription occurs from a main or unique TSS does not longer apply to all promoters, indeed over 70% of mammalian promoters do not follow this behaviour [54][55]. Instead, multiple modes of transcription initiation exist (**Fig 4**).

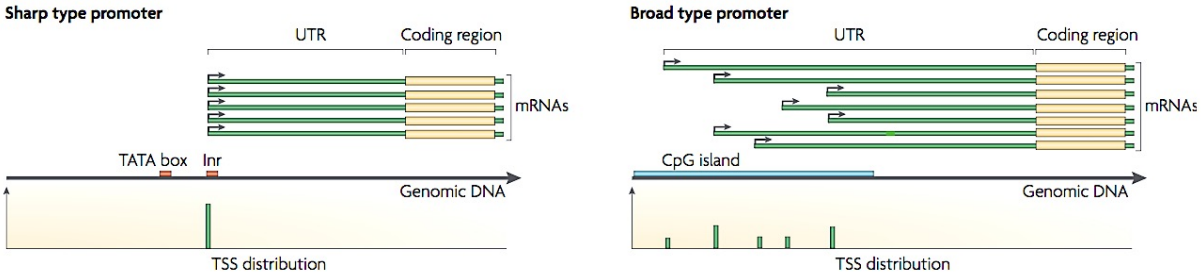


Figure 4. Representation of sharp and broad classes of promoters. Sharp promoters (left panel) give rise to a single peak of TSS distribution and often have TATA-box and Inr elements. Broad promoters (right panel) generate an heterogeneous mRNA population due to a wider TSSs distribution and often are associated with CpG islands (from [51]).

Focused (or 'sharp-peak') promoters contain a single or predominant TSS and are enriched in spatiotemporally regulated and tissue-specific genes. They usually harbour precisely positioned core-promoter elements, such as Inr, TATA-box, DPE or MTE [56]. Dispersed (or 'broad') promoters give rise to several start sites spreading over ~100bp and are especially associated with constitutively expressed genes and CpG islands. The picture gets even more complicated for the presence of 'mixed' promoters, in which a dominant TSS coexists with a dispersed array of minor ones.

Second, a large portion of mammalian promoters give rise to divergent transcription [57][58], given by the recruitment of independent and well-spaced transcription initiation complexes both on the forward and reverse-directed core-promoters [59]. This phenomenon gives

rise to pairs of co-regulated transcripts: the stable pre-mRNA in the forward direction towards the gene body and short, low-abundant species of transcripts (named TSSa-RNAs and uaRNAs) in the reverse direction. The same principle applies to bidirectional promoters, in which oppositely oriented pair of genes gets transcribed from a common promoter region containing intrinsically bidirectional core-promoter elements (**Fig 5**). The latter case is especially common in compact genomes of lower eukaryotes such as yeast. Notably, divergent transcription at promoters shares a common architecture with enhancer regions, where it gives rise to the recently characterized class of non-coding transcripts known as enhancer-RNAs (eRNA) [60]. This widespread divergent activity of regulatory units is the basis of pervasive transcription of non-coding regions in complex genomes and ultimately relies on core-promoters for the assembly of the transcriptional machinery.

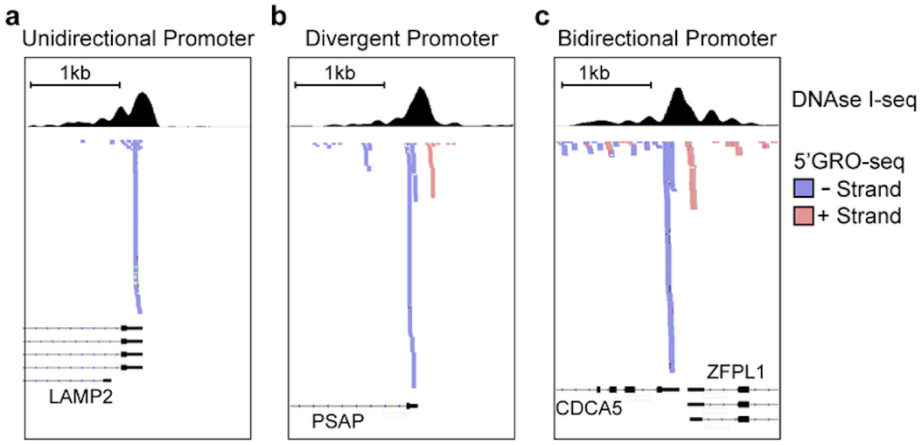


Figure 5. Examples of unidirectional, divergent and bidirectional promoters. It is shown the DNaseI-seq signal (in black) and 5'GRO-seq reads (5' sequencing of cap-protected RNA) from both strands (from [59]).

1.3.2 PIC assembly at core-promoter

Core-promoters are DNA regions on which the so-called transcription pre-initiation complex (PIC) assembles to prime Pol II dependent transcription [61]. The PIC is a huge ~3MDa multiprotein complex composed by several independent sub-complexes, including Pol II, for a total of nearly 50 polypeptides [62]. The assembly of PIC allows Pol II to recognize the core-promoter region and to position the template DNA within the catalytic centre to begin RNA synthesis. The components of the PIC, except Pol II, are known as general (or basal) transcription factors (GTFs), namely TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH, whose specific contribution were intensively investigated through biochemical and structural approaches. The assembly pathway of the PIC on the core-promoter was studied by ordered-addition experiments from purified proteins [63] and led to a canonical stepwise model for PIC assembly depicted in **Fig 6**. Specifically, core-promoter is first recognized by the ~1MDa TFIID complex which harbours the saddle-like TBP subunit (TATA-binding protein) that binds to DNA minor groove, bending it

of about 90°. The rest of TFIID subunits (14-15) are named TAFs (TBP associated factors) and constitute a three-lobes scaffold able to recognize other core-promoter elements, such as Inr or DPE and to interact with general co-factors or specific transcriptional activators to enhance or modulate PIC assembly. The TFIID module is then stabilized by TFIIB and the accessory factor TFIIA that stimulate Pol II recruitment and TBP/DNA interaction, respectively. This TBP/TFIID-TFIIB architecture works as a landing platform for Pol II-TFIIF complex on the promoter, resulting in the core initiation complex. The complete PIC forms when TFIIE and TFIIH are subsequently recruited, allowing the formation of an initiation competent Pol II. The DNA-dependent ATPase activity of TFIIH, stimulated by TFIIE, allows the transition to open-complex by applying a torsion on DNA with respect to the fixed upstream promoter complex, thus inducing upstream double helix melting and template strand loading into Pol II active site [62] (**Fig 7**). Recently, the architecture of human TFIID in complex with TFIIA and core-promoter DNA was solved at sub-nanometre resolution using single-particle cryo-electron microscopy (Cryo-EM) [64].

An alternative holoenzyme model was proposed with the discovery that Pol II could be purified as a preassembled complex with a variable subset of GTFs and other proteins involved in chromatin modification, but devoid of TFIIA and TFIID [65]. Although great advancements in the functions of each GTF have been made thanks to structural characterization and mutational analyses, the mode through which the PIC assembles is still controversial and seems to change depending on the system used and the organism of origin [66]. Moreover, we need to take into account that assembly experiments and structural studies have been performed on artificial super-core promoters (SCP) and in any case always

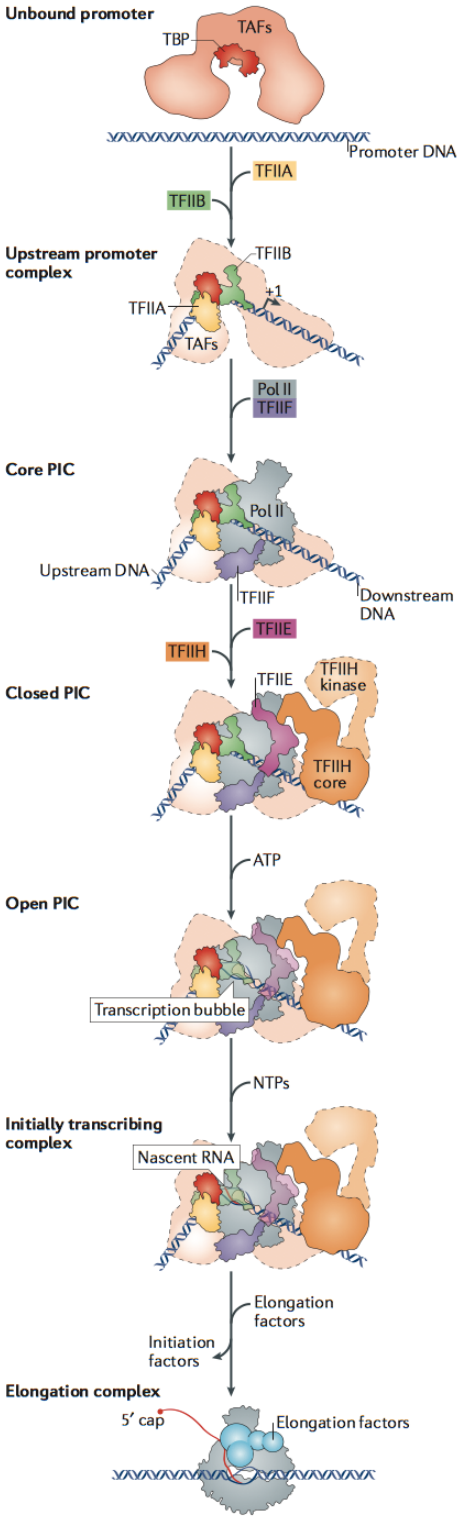


Figure 6. Canonical model of stepwise PIC assembly through sequential addition of the GTFs and Poll II (from [62]).

included a canonical TATA-box, presented only by a minority of eukaryotic promoters. Notably, additional complexity and promoter recognition specificity is given by the existence of TBP-related factors and tissue specific TAF variants [67][68] that challenge the "general" attribute of GTFs.

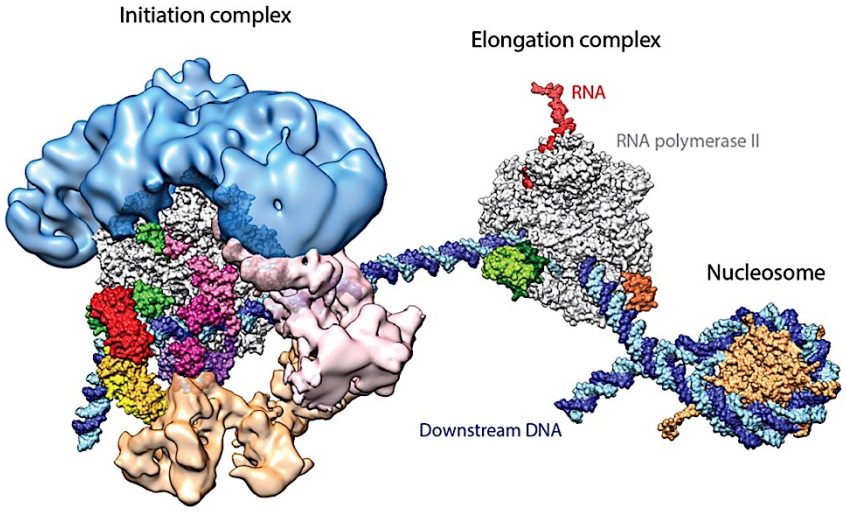


Figure 7. Current view of transcription by Pol II based on results of integrated structural biology. Current structures of the initiation complex (with Mediator in light blue), elongation complex with nascent RNA and a nucleosome on the same DNA molecule (from [69]).

1.3.3 Eukaryotic promoters work as complex information integrators and processors

In the previous two paragraphs, I discussed the inner core mechanisms at the basis of transcription initiation control. Besides the inherent macromolecular complexity of core-promoter recognition and assembly of a functional Pol II complex, the key control gears that efficiently drive transcription initiation are located elsewhere, within proximal-promoter and distal regulatory regions. Differently from core-promoters, proximal-promoter regions do not include the main TSS and are roughly located at -50 to -500 bp from +1 nucleotide. Indeed, core-promoter by itself is not able to support physiological levels of active transcription *in vivo*, both due to chromatin, and the inherent rather low stability it provides to the transcription machinery. The function of proximal-promoter regions is to integrate diverse sources of information and transduce them to the transcription apparatus at the core-promoter, thus altering its function. This complex task is accomplished by a wide variety of *cis*-embedded recognition elements and the corresponding *trans*-acting transcription factors.

Promoters work as substrates of complex molecular dynamics involving sequence-specific DNA binding events, direct protein-protein interactions between co-resident TFs or involving GTFs at core-promoter, nucleosomes modification and remodelling through the recruitment of specific enzymatic complexes, and long-range chromatin rearrangements. The result is the spatiotemporal integration of the following information:

- the unique combination of *cis* transcription factor binding sites (TFBS)
- the presence of specific TFs, their local concentration and their functional state
- the local epigenetic features and dynamics
- triggering of endogenous or external signalling events

The outcome of this complex series of inputs will determine the promoter output in terms of transcription initiation rates. Several conditions are required for transcription initiation to take place and are scattered on different timescales. Chromatin modification and opening for upstream TFs binding and core-promoter nucleosome depletion takes minutes and can actively be maintained for hours. TFs binding instead is quite transient, ranging from milliseconds to few minutes and can be influenced by many local and cellular factors. Nucleation and stabilization of PIC is favoured by direct TFs interaction with GTFs or indirect recruitment of co-activators (e.g. Mediator) by TFs bound to proximal-promoter or distal regulatory regions. The very last, yet often limiting step, is the actual Pol II escape from the open-complex to start transcribing, named promoter clearance. This step is promoted by several phosphorylation events that free Pol II from the very same complex that allowed its loading on DNA and are themselves indirectly regulated by local TFs.

The very same principles of complementarity between REs and corresponding TFs apply also to distal regulatory regions, also known as enhancers (or silencers, if they impart negative control on gene expression). Differently from proximal-promoters, enhancers are DNA regulatory regions not directly adjacent to a single protein-coding gene, instead they can be located thousands of base pairs away from the target gene or even on a distinct chromosome. Since the post-genomic era, enhancers were progressively characterized as non-coding RNA transcription units, effectively working as promoters for a recently discovered class of non-polyadenylated transcripts known as eRNAs [70][71]. For these reasons, the basic principles governing enhancer-associated TFs are fundamentally homologous to proximal-promoter functioning.

1.3.4 Combinatorial control of gene expression

As mentioned above, both proximal promoters and distal regulatory regions work thanks to a series of specific TFBSs. Each gene is functionalized by dozens of currently known TFBSs and by others, still unknown, distributed within its proximal-promoter. This combinatorial arrangement of multiple binding sites opens up the possibility for three crucial features, demanded by eukaryotic gene expression: high control, improved specificity and wide dynamic expression range. Many factors influence these features and include both *cis*-encoded REs and the intrinsic biochemical properties of the corresponding TFs, coupled with signalling events.

The number and type of REs within the promoter are not the only determinants of promoter functionality, but rather their spatial distribution is just as critical. Indeed, spacing,

overlap, order, orientation and distance from TSS of REs are all well proven parameters that can greatly impact how TFs actually 'read' promoter DNA. Literature is plenty of such specific examples [72][73][74][75]. The structural arrangement of REs is intimately coherent with the required structural properties of TFs, coevolving in a complex series of conserved binding patterns. As a matter of fact, the information content enclosed within promoter sequences is processed as a function of local molecular environment through protein-DNA and protein-protein interactions, directed by a dynamic series of transient electrostatic forces, hydrophobic effects, interlocking protein domains and subtle, or even dramatic, conformational changes. So, the (non-random) spatial distribution of REs interlaces with the correspondent TFs mutually exclusive or cooperative binding, and the selective or synergistic recruitment of transcriptional co-activators/repressors.

The increased control over transcription initiation is therefore achieved by increasing the components (REs and TFs) that can simultaneously operate on the same promoter, making it possible to turn on a given gene if the multifactorial conditions that are required for its activation are satisfied (**Fig 8**). This allow cells to promote the selective expression of only a subset of a given TF target genes at any time with a higher degree of control.

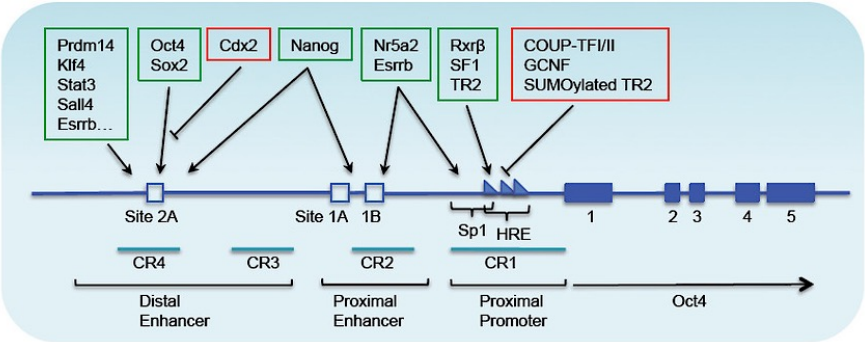


Figure 8. Genomic structure and transcriptional regulation of mouse Oct4 gene, an example of combinatorial transcriptional control. The scheme covers ~24 kb. TFs that bind each region are indicated and have either activating (green) or repressing (red) function (from [76]).

The interactions between pairs of co-resident TFs on DNA can highly improve the recognition specificity or target sequences. For instance, low affinity sites can be selectively bound when a second TF occupies an adjacent site allowing for specific binding cooperativity and promoting the assembly of stereo-specific protein-DNA complexes. Alternatively, binding events can be challenged by competition for distinct overlapping sites or steric hindrance between protein interfaces. Since many TFs bind DNA as homo- or heterodimers, binding specificity can be easily constrained by the underlying sequence spacing, orientation and composition. Well studied examples of these sequence constrains are hormone response elements (HREs) and the variety of nuclear receptor superfamily [77][78], MADS box proteins that synergize with different TFs

partners (e.g. human SRF and SAP-1) [79], POU and SOX proteins partnership (e.g. OCT4/SOX2 heterodimers) [80][81], or NFAT and Fos/Jun heterodimer [82].

The combinatorial array of REs also provides eukaryotic cells with a wide dynamic range of transcription. Transcript abundance can vary significantly of many orders of magnitude (0.001 to 100 copies/cell) [83][84][85]. Moreover, alternative modes of gene expression have been documented, in which initiation events can either be evenly separated in time, such as for some constitutively expressed genes, or happen through transcriptional bursts that underlay clear transitions from OFF to ON states [86].

1.4 How transcription factors read the genome

The essential property of a transcription factor is the capability to recognize a specific DNA sequence and to interact with other transcription-related proteins. In this section I will describe the general principles governing TFs target site recognition and search throughout the genome, along with the highly relevant concepts of binding cooperativity and transcriptional synergy.

1.4.1 The basis of DNA binding and sequence-specific recognition

Achieving DNA binding specificity in a complex eukaryotic genome is not a trivial task. As already pointed out, increasing the genome size also increases the expected number of random, non-functional REs. On top of that, eukaryotic TFs have a very poor intrinsic ability to discriminate cognate sites with high specificity if compared with prokaryotic counterparts; specific core-DNA motifs recognized by eukaryotic TFs are usually short (~6-8 bp) and have a low sequence-based information content. For instance, a 6 bp core element is predicted to occur once every 4 kb, which for the human genome corresponds to roughly 10^5 available sites, even postulating 90% chromatization level [87]. A related observation is that eukaryotic cells have evolved very different DNA-binding domain families, with very few exceptions shared with bacteria, underlying fairly different strategies for gene regulation. The real question then would be whether having TFs with intrinsic lower information content requirement in DNA recognition actually represents a real disadvantage. Probably this is not the case. First, obviously because eukaryotic cells had to take advantage of this change in gene regulatory strategy in order for it to be fixed in evolutionary times. Second, it is reasonable to assume that lower intrinsic specificity can be coupled with the advantage of higher flexibility. The small bacterial genomes contain only a tiny fraction of non-coding regulatory DNA and their genic arrangement in operons facilitates the co-expression of genes belonging to the same network. In this configuration, ideally each single highly-selective TF could directly target its correspondent subset of genes to provoke a very specific response. In order to

employ the same "one TF – one response" strategy for more complex genomes which must direct the development of multicellular life, the number of TF-encoding genes should have increased more than linearly to impose a proper control on gene activity patterns. As already mentioned, however, the proportion of TF encoding genes in eukaryotes is comparable to the prokaryotic one. The strategy employed by eukaryotic cells instead is based on the use of specific subsets of the same TFs repertoire in different combinations. The intrinsic lower DNA sequence requirements allow eukaryotic TFs to take part to a variety of context-specific subcomplexes, still maintaining a good degree of selectivity thanks to protein-protein interactions and distribution of the respective REs, giving rise to cooperativity and transcriptional synergy phenomena (discussed in the next paragraph). Chromatin, on the other hand, helps in limiting non-functional binding events.

Transcription factors are multi-domain proteins, usually including a DNA-binding domain (DBD) and an effector domain, such as a transactivation (TAD) or a transcriptional repressive domain (TRD). TFs are categorized in distinct superclasses on the basis of the structural topology of their DBDs [88][44], thus employing distinct modes for DNA binding.

Although varying in the extent from case to case, the general basic principles of sequence specific DNA recognition rely on the same series of biophysical interactions. These include electrostatic (salt-bridges), hydrogen-bond and hydrophobic interactions. Even though all of them can also occur in non-specific TF/DNA interactions, the latter events are largely mediated by protein basic side-chains and DNA acidic backbone phosphate groups. Protein/DNA interactions are very sensitive to salt concentrations due to the charge-shielding effect of positive counter-ions on the negatively charged DNA phosphate backbone. This is particularly true for non-specific protein-DNA interactions, often mediated by the negatively charged DNA surface. It is important to note that also sequence specific TFs largely employ non-specific contacts with DNA backbone as structural support to properly position protein moieties responsible for sequence specificity; moreover, these interactions provide a substantial contribution to the binding free energy of the protein/DNA complex, boosting the overall TF affinity for DNA.

Specific interactions fall into two non-mutually exclusive categories, namely 'base-readout' and 'shape-readout'. The base-readout mode consists in the direct recognition of DNA sequence through physical interactions between TF amino acids and the functional groups exposed by the DNA nucleobases. The chemical signature of each base-pair is different between the DNA major and minor groove, as the information content of the two grooves is different. While each base-pair combination in the major groove is distinguishable thanks to its unique pattern of hydrogen bond donors and acceptors, this is not the case for the minor groove, where A:T and C:G are indistinguishable from T:A and G:C pairs, respectively (see **Fig 9**) [89]. This degeneration in the minor groove might be one reason for the sheer prevalence of major groove binding domains

among TFs. Base-readout generally involves hydrogen bonding for purine bases, especially using highly selective bidentate hydrogen bonds, and hydrophobic interactions for pyrimidines, differentiating thymine from cytosine. This recognition mode is usually complex, with each amino acid side chain involved in one or more interactions with one or more nucleobases, also including water molecule mediated hydrogen bonds.

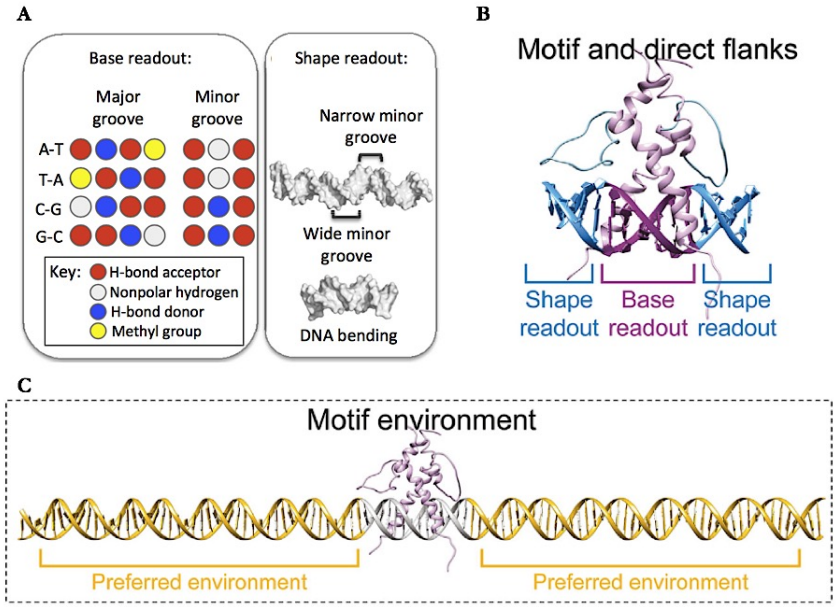


Figure 9. Base and shape readout contribution in DNA-binding specificity. (**A**) Base-readout relies on direct interaction with the nucleobases functional groups; note the different information content between the minor and the major DNA groove. Shape-readout instead makes use of sequence-dependent alterations from the canonical B-DNA (from [90]). (**B**) Often TFs use a combination of base and shape-readout mechanisms. (**C**) A third relevant factor that can direct TF binding is the local motif environment in term of base composition (from [91]).

The second, more subtle mode for sequence-specific DNA recognition is known as shape-readout, that is all sequence-dependent DNA structure deviations from canonical B-DNA [92]. The local sequence composition can indeed alter DNA conformation and properties from its ideal straight double-helix configuration. The shape-readout mechanism can exploit both localized shape deviations, such as narrowing of the minor groove or kinks of the DNA longitudinal axis, and alterations extended to the entire binding site, including DNA flexibility, bending and intrinsic tendency of B- to A-DNA transitions (**Fig 10**). These sequence specific structural properties of DNA can thermodynamically favour its recognition by specific TFs by optimizing their contacts by shape complementarity or electrostatic potential. Indeed, there is evidence that local DNA topography is under selection and well correlates with non-coding DNA regulatory elements [93]. Although conceptually distinct, base- and shape-readout are not exclusive and most TFs use an interplay between these two sequence recognition modes, varying in the extent of the contribution from each of them [94].

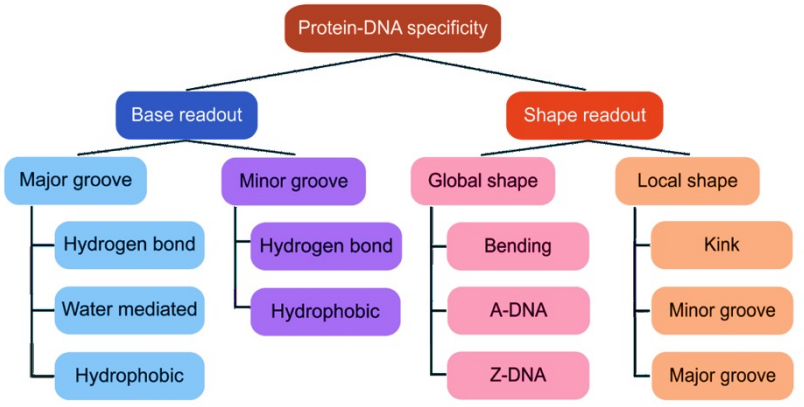


Figure 10. Determinants of base and shape readout in protein-DNA specificity (from [94]).

1.4.2 DNA-binding cooperativity and transcriptional synergy

Cooperativity generically refers to the facilitation of an event given by a previous/separate one. Instead, synergy in general refers to any enhanced effect given by multiple factors acting in concert. Although cooperativity and synergy have slight distinct meanings, in biology they are often used as interchangeable terms to describe higher-than-additive phenomena. However, in the context of transcriptional regulation they can represent two different, although interwoven effects.

In the biology of protein-DNA recognition we refer to cooperativity as the facilitation effect that a TF bound to DNA provides for the subsequent binding of a second TF. This is usually mediated by direct or indirect protein-protein interactions between the two TFs, although there are cases in which the binding of the first TF induces a favourable DNA conformational change for the binding of the second one nearby [95]. Thereby, the effect of cooperativity is to increase the affinity of the second TF for the target DNA driven by the presence of the first one, already bound to DNA. An additional, rather indirect form of cooperativity involves *in vivo* competition between nucleosomes and TFs, whereby the nucleosome displacement driven by the first TF frees adjacent binding sites for other TFs, resulting in co-occupancy [96]. Homotypic cooperativity occurs between two molecules of the same TF following a sequential binding model, while heterotypic cooperativity occurs between different TFs. Note that cooperativity is distinct from DNA binding of pre-assembled homo- or heterodimers, because in the latter case the monomeric form of the protein is not proficient in DNA binding prior to dimerization. Moreover, cooperativity is a DNA-directed phenomenon, meaning that usually the interactions between the two transcription factors are much weaker than their respective interactions with DNA, mainly because the arrangement of the correspondent REs stereo-positions the two proteins, favouring

interaction. Another way to think about cooperative interactions is that a resident TF slows down the release from DNA of a second one.

DNA-binding cooperativity is at the basis of the 'enhanceosome model', derived from the pioneering studies of the interferon-β enhancer [73][97] (**Fig 11**), by which the function of a regulatory module is defined by the cooperative assembly of a handful of TFs that promote each other's binding either through direct protein-protein contacts or by binding-induced changes in DNA conformation. This kind of regulatory module would be very sensitive to changes in spacing and orientation of the single binding-sites, from which cooperativity originates. For this reason, homotypic or heterotypic cooperativity highly increases target site specificity, since enhanced binding will occur only at those sites in which multiple REs are properly arranged. In addition, heterotypic cooperativity is an efficient way to integrate diverse cellular signals using a limited set of effectors. Recent works definitely showed that cooperativity can effectively shape the individual TF's motif in favour of a new composite site occupied by the cooperative TFs pair [98].

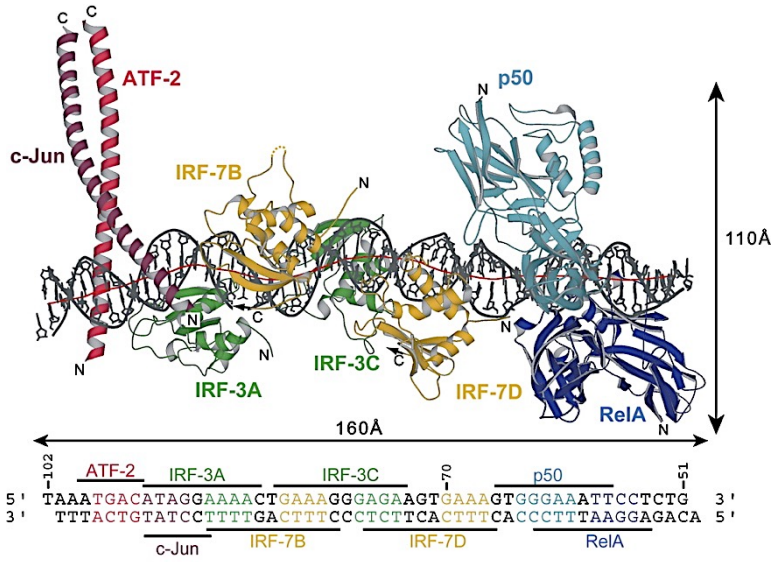


Figure 11. Side view of the overall structure of IFN-β enhanceosome. The red line shows the DNA helical axis highlighting local bending. RelA and p50 are the two NFkB subunits (from [97]).

One of the major potential effects of DNA-binding cooperativity is to facilitate a synergistic transcriptional response. Transcriptional synergy refers to the concerted action of multiple TFs that synchronously contact the basal transcriptional machinery and promote a so called 'sigmoidal transcriptional response' (STR) [99]. The principle underlying transcriptional synergy is that when multiple sites are occupied by the correspondent TFs the extensive contacts made with the transcriptional machinery are disproportionately stronger than the one provided by a single TF. This multiplicative effect gives rise to a STR (**Fig 12A**), characterized by a switch-like ON/OFF expression defined by a relatively narrow threshold width. In this model, a graded concentration

increment in one of the key TF above the threshold value would abruptly trigger gene expression. A classic example of this behaviour is the sharp gene expression boundary of *hunchback (Hb)* in response to the anterior-posterior gradient of the TF Bicoid *(Bcd)* in the developing *Drosophila* embryo [100] (**Fig 12B**). However, it is important to note that DNA-binding cooperativity is not strictly necessary to obtain a STR, as long as different TFs synergize to recruit the basal transcription apparatus. The specific effect of DNA-binding cooperativity, taking into account the simplest case of homotypic cooperativity, is that an increase of the number of properly spaced binding sites would shift the threshold value to lower TF concentrations and would increase the sharpness of the switch [99]. Transcriptional synergy is at the basis of the 'billboard model' for enhancer function, whereby the TFs bound together at the regulatory site do not necessarily work as an all-or-nothing cooperative unit but rather in a combinatorial fashion, underlying a fairly flexible architecture for the regulatory modules [101]. Indeed, these models are not alternative in the absolute sense and different regulatory regions are hardwired with different patterns of REs, which in turn will drive their peculiar transcriptional response.

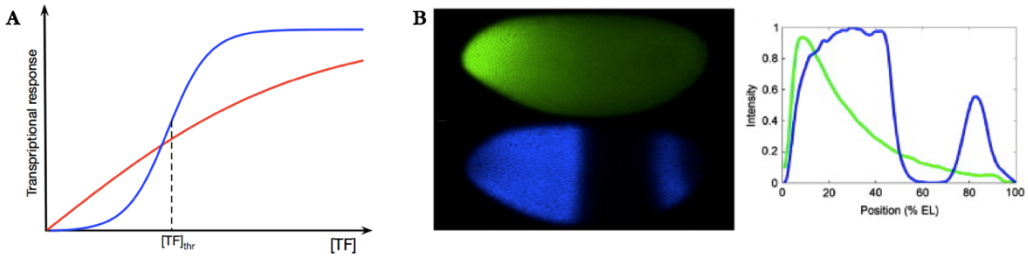


Figure 12. (**A**) Simple comparison between graded (red curve) and sigmoidal transcriptional response (STR, blue curve) as a function of TF concentration; note the switch-like shape of the STR around a specific threshold value (thr). (**B**) Example of the STR of *Hunchback* gene (blue signal) in response to a gradient of the TF Bicoid (green signal) during the syncytial blastoderm stage in *Drosophila* embryo (left panel) and the correspondent fluorescence intensity relative to the anterior-posterior position (right panel, from [102]) Note that the double-peaked pattern for *Hunchback* is due to additional levels of control of its expression.

Beyond the discussed sigmoidal responses, clearly there are reports of graded, or gradual, transcription [103][104][105], in which gene activity can ben modulated in a wide range of expression in response to a wide range of correspondent TF concentrations. While STRs are widely used during development in order to confine gene expression in defined embryonic domains, graded transcription provides a wider responsiveness to a broad spectrum of TF, or its inducer, concentrations. This latter case is especially employed for homeostatic responses, which have to be proportional to the primary stimulus, thanks to their higher sensitivity and accuracy.

1.4.3 Target site search and mobility of transcription factors

Transcription factors are complex objects which operate in a complex environment, the nucleus. A fascinating problem is how transcription factors manage to rapidly find their cognate sites throughout such an excess of non-specific genomic DNA, especially after specific stimuli that demand a fast and punctual transcriptional response.

The main explanation is that TFs are very dynamic and mobile proteins. Although the nucleus is a crowded molecular environment, with macromolecules concentration in the order of 100-400 mg/mL [106], the inter-chromatin space takes a large fraction (40-70%) of the nuclear volume, varying according to the cell type [107]. TFs move rapidly within the nucleoplasm, making it possible to scan the nuclear volume in a matter of few minutes. An interesting observation is that non-chromatin binding proteins (e.g. GFP) have a 10-100 fold higher mobility compared to TFs of the same size in the nucleus [108]. This behaviour is mainly due to transient interactions with the relatively immobile chromatin fibres, and partly to the formation of bigger multi-protein complexes; indeed, impairment of the protein DBD greatly enhances its mobility within the nucleus [109], evidencing the grade of influence of DNA binding in the dynamics of TFs. Although the *in vivo* estimates of the bound fraction for TFs found in literature are very variable, possibly due to different technical assays and discrepancies in the models used for data fitting [110], still they represent a substantial fraction of the protein, ranging from 20% to 99% of the TF present in the nucleus.

Our perception of transcriptional regulation complexity has been hugely increased by the advent of fluorescence-based imaging techniques in living cells (e.g. FRAP, FLIM, FCS, RLSM, SMT) [108][111][112][113][114]. These *in vivo* approaches revealed for the first time how dynamic TFs binding to chromatin is [115]. Most nuclear proteins continuously exchange from chromatin to the inter-chromatin space very rapidly, with residence times on the order of milliseconds to seconds [116]. Exceptions of proteins with much slower exchange rates exist, first of all core histones, which show residence times of hours. Long residence times (>6 min) have been documented also for HSF transcription factor upon heat-shock in *Drosophila* polythene nuclei [117]. Although this dynamic scenario seems to contradict previous biochemical and population-based approaches, recent single-molecule analyses show a remarkable accordance between living cells and *in vitro* kinetic measurements, both pointing at short, frequent binding events [114].

Within the overall dynamic bound fraction, it is possible to distinguish two subpopulations of binding events that differ for their kinetic parameters: a short-lived and a long-lived component, that differ in their respective dwell-times of more than one order of magnitude [114][118]. These two binding-modes respectively account for non-specific DNA/chromatin interactions and specific target sites recognition [119]. The residence time on high-affinity specific DNA REs is

therefore intuitively longer than spurious binding events taking place on non-specific DNA regions.

These observations, coupled with smart *in vitro* systems [120] and theoretical models [121], allowed the formulation of a fast-search mode for target site recognition. TFs would employ a search mode based on 3D diffusion in the inter-chromatin space alternated with transient interactions with chromatin, effectively sampling many sites per time unit, given the short-lived nature of non-specific binding events. This global 3D-jumps-governed sampling behaviour can be complemented by a series of 1D local searching motions, well documented for *in vitro* systems, especially using restriction endonucleases as an easy-readout DNA-binding model [122][123], although examples for TFs are also documented [124][125][126][127]. These local motions might include sliding along DNA, local hopping or intersegmental transfer [128]. Sliding along the DNA is particularly interesting, since this mechanism allows effective scanning of sequences nearby the landing site at a base-pair resolution. Depending on the system, sliding can either be coupled with rotation around the DNA axis, following its helical pitch, or proceed linearly [129][130]. The effectiveness of DNA scanning through sliding motion is limited by its non-directional nature: the movement of the (TF) protein on the DNA surface does not have processive directionality *per-se*, therefore it leads to inefficient sampling of the same region multiple times. Theoretical models and *in vitro* evidences give a sliding length of ~30-200 bp before dissociation followed by local hopping or 3D diffusion [131][132], effectively providing an highly efficient searching strategy. All these local protein-DNA dynamics have to be confirmed by *in vivo* observations, which indeed pose challenging technical problems. Moreover, the chromatin environment could strongly limit sliding length along DNA, favouring instead a higher frequency of local hopping. A second key aspect that can heavily influence TFs dynamics *in vivo* is the role of ATP-dependent chromatin remodelling complexes and chaperones, that can actively mobilize DNA-bound proteins [133][134].

Having reframed the timescale of protein-DNA dynamics in living cells, this scenario of transcription regulation supports a model in which the assembly of multiprotein complexes on DNA happens by random collisions in a probabilistic way. The result is that every step in the formation of a transcriptional competent promoter is a chance event. Control over transcription initiation is therefore exerted by modulating the probability of some of the multiple steps that ultimately lead to promoter clearance. By increasing the stability of each intermediate the probability of assembly of a complete, elongation-competent RNA polymerase, would increase. Indeed, slower exchange rates of the glucocorticoid receptor (GR) at promoters correlate with increased mRNA synthesis [135]. Besides residence time, also local TFs concentrations and DNA-binding cooperativity phenomena greatly contribute to affect transcription initiation probability.

Overall, the dynamic exchange of chromatin-binding proteins opens the possibility for rapid and economic local reorganization and transcriptional outputs responsive to the ever-changing environmental stimuli, ensuring the continuous availability of DNA binding sites and free nuclear regulators.

1.5 The CCAAT-binding transcription factor NF-Y

In the late '80s, the intensive characterization of the murine class II major histocompatibility complex (MHC-II) locus led to the identification of the protein that specifically binds the conserved DNA regulatory module Y-box, therefore named Nuclear Factor Y (NF-Y) [136]. During the last 30 years many aspects of this ubiquitous and peculiar transcription factor have been deeply characterized and represent the main research topic in our laboratory.

In this section I will give a general overview on the biology of NF-Y, describing aspects I find particularly interesting and basal for my thesis project, including its unique structural features, molecular functions, the genomic interplay with a defined set of TFs and its pioneer role in epigenetic and transcriptional regulation.

1.5.1 The CCAAT-box is a widespread eukaryotic Recognition Element

The so called CCAAT-box is one of the first functional DNA elements discovered in eukaryotic promoters, first identified in chicken ovalbumin and human β-globin genes in 1980 [137][138] and thereafter found to be represented in every eukaryotic system studied by far. Even though literature is plenty of DNA-binding proteins that harbour the word CCAAT in their names (i.e. CTF/NF1, C/EBP and CDP), analysis of their consensus binding sequence clearly reveals that none of them strictly requires all five nucleotides of this element; it is indeed true that their binding sites occasionally contain an intervening CCAAT. For example the so called CCAAT-Transcription Factor (CTF/NF1) binds the bipartite consensus TTGGC(N5)GCCAA as a dimer; clearly, a T following the core consensus would give rise to a CCAAT, but it is not strictly necessary for proper binding. Instead, compelling evidences demonstrated that NF-Y, also known as CBF or CP1, strictly counts on the intact CCAAT pentanucleotide to bind DNA. Saturation mutagenesis, surveys on bona-fide NF-Y target promoters and more recent ChIP-on-chip and ChIP-seq approaches univocally defined NF-Y as the prime CCAAT-binding factor and generated consensus sequences with remarkable accordance [139][140][141][142]. These studies also highlighted a strong preference for flanking nucleotides to achieve high affinity interactions, with documented cases of negligible NF-Y DNA-binding on full core CCAAT sequence lacking several preferred flanking nucleotides [140]. At first look, it is clearly detectable a strong prevalence of two purine bases at -2 -1 positions and moderate preference of CAG trinucleotide just downstream the core sequence (**Fig 13A**).

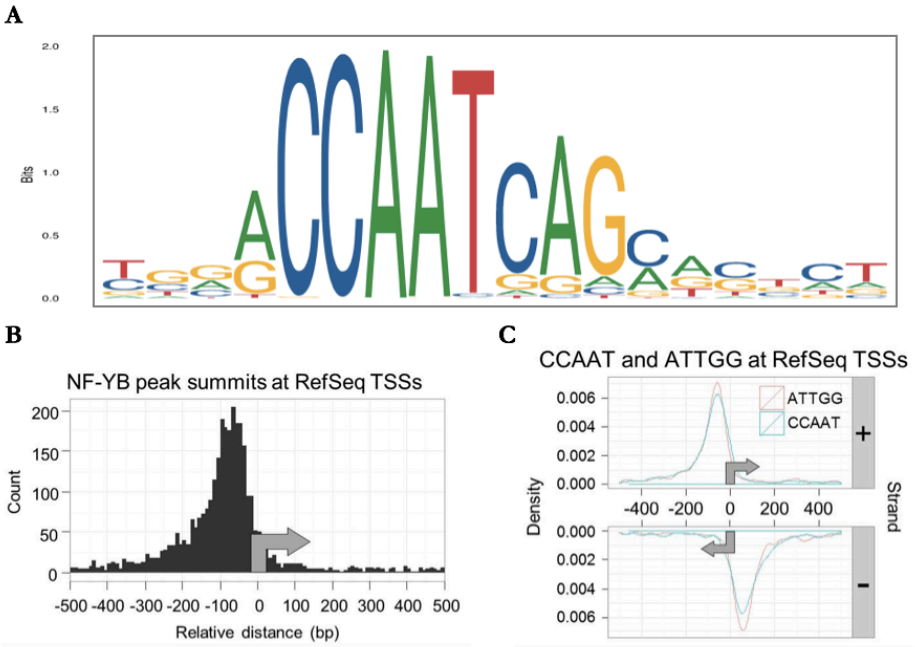


Figure 13. The CCAAT-box. (**A**) Sequence logo of the ChIP-seq derived NF-YA binding motif from JASPAR database [143]. (**B**) Frequency distribution of NF-YB peak summits in K562 cells at RefSeq TSSs. (**C**) Distribution of positive and negative strand 5'-CCAAT-3' and 5'-ATTGG-3' sequences at K562 NF-YB-bound RefSeq TSSs. Grey arrows indicate direction of transcription (from [142]).

The resulting picture of the CCAAT-box is that of a recognition element devoid of any symmetry axis, very common among promoters (>60% human promoters) and occurring on the forward or reverse strand with similar frequencies, often within -600 bp from TSS with a peak centred at -80 bp (**Fig 13B-C**) [144]. The same bioinformatic investigation showed a negative correlation between TATA and CCAAT boxes, since the number of double positive promoters is significantly low considering the number of single positive promoters. Moreover, the CCAAT-box positively correlates with CpG islands and bidirectional promoters [145], both features that negatively correlate with canonical TATA-boxes.

Another striking observation is that CCAAT-boxes often occur multiple times within human promoters [146][141]. Double CCAAT-boxes within a window of 150bp represent a sizable fraction of the genome-wide NF-Y bound sites (35-40% in HeLa cells and more than 50% in K562 cells), with a strong prevalence for the same orientation of this pair of DNA motifs [147]. These findings also revealed a clear preference for few recurrent distances between the two CCAAT-boxes, which vary depending on their relative orientation and whether considering their occurrence within promoters or at genome-wide scale (**Table 3**). These observations fit with detailed functional studies on multiple-CCAAT cell-cycle promoters, that highlighted the strict requirements for precise spacing between these regulatory elements to elicit a proper transcriptional response *in vivo* [148].

CCAAT Orientation	Genomic	Promoters
+/+	19-33-36-43	26-36-47
+/-	29	n.s.
-/+	13-30	n.s.

Table 3. Preference in double CCAAT-box distances depending on their relative orientation (+ forward; - reverse) at genomic scale or in promoters (within 1000bp from a RefSeq TSS). The numbers indicate the peak-summit of the frequency distribution for each category, considering the number of bp that separate two CCAAT pentanucleotides. n.s. indicates a non-significant distribution (summarized from [147]).

1.5.2 The NF-Y complex is ubiquitous and works as a triad

Differently from the majority of TFs that work as monomers or dimers, NF-Y is a heterotrimeric TF. The three subunits, namely NF-YA, NF-YB and NF-YC, are all required for DNA-binding, allowing the NF-Y complex to recognize and avidly bind CCAAT-box sites [149]. Early sequence alignments revealed that each subunit includes an evolutionarily conserved core domain, positioned at the C-terminal, central and N-terminal regions of NF-YA/YB/YC subunits, respectively, and named from the yeast homologues HAP2/3/5 (**Fig 14**).

NF-Y subunits are all encoded by a single copy gene in vertebrate genomes (except some ray-finned fishes due to genome duplication). NF-YB gene does not undergo alternative splicing, at least in mammals, giving rise to only one major protein product of 32 kDa. Instead, NF-YC locus transcriptional regulation shows a much higher complexity, involving both alternative promoters and alternative splicing, giving rise at least to four distinct NF-YC isoforms [150]. The second promoter resides in the first intron and the corresponding product has a different 5'UTR that influences the mRNA stability and its post-transcriptional regulation. Alternative splicing instead mainly occurs within the 3'-end, within the Q-rich region and involves exons 9 and 11, giving rise to three main isoforms identified by their molecular weight: 37 kDa, 48 kDa and 50 kDa. Even though all three splicing isoforms are transcribed in a given cell type, with the 37 kDa being the most abundant, their expression at the protein level is different, with the 37 kDa and 50 kDa isoforms being almost mutually exclusive.

NF-YA gene can give rise to two distinct protein products thanks to alternative splicing of exon-3, which is included in one isoform while the other one is devoid of it [151]. Exon-3 coding sequence is localized within the N-terminal Q-rich domain and accounts only for 28 aa. NF-YA isoform including exon-3 is called NF-YA*long* (NF-YA*l*, 43 kDa) while the isoform lacking exon-3 is called NF-YA*short* (NF-YA*s*, 41 kDa), being 28 residues shorter. Even though no biochemical evidence is actually present that accounts for the molecular function of the additional 28 aa encoded by exon 3, many evidences suggest a well-established divergence between the two isoforms in terms

of cell type, differentiation level and developmental stage at which their prevalent expression occurs. Even though not many cell types exclusively express only one NF-YA isoform, a clear preference towards NF-YA*l* or NF-YA*s* can generally be delineated: at the two extremes fibroblasts predominantly contain the long isoform, while lymphocytes the short one [149].

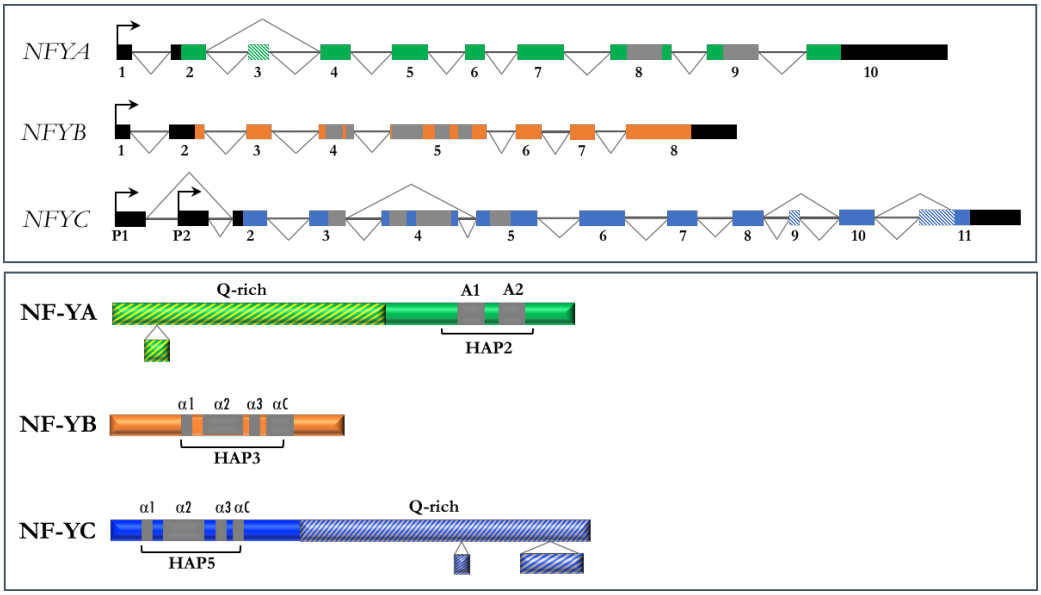


Figure 14. Schematic representation of human NF-YA, NF-YB and NF-YC both at genomic (upper panel) and protein level (lower panel). Arrows indicate the TSS, black regions are 5' and 3' UTRs, main exons undergoing alternative splicing are striped. Grey boxes indicate α-helices included in core domains of each subunit. Core domains are indicated (HAP). Lengths are not in scale.

NF-YA*s* today is widely considered as the stemness-related isoform, thanks to several evidences concerning both adult and embryonic stem cells. Overexpression of NF-YA*s* in hematopoietic stem cells (HSCs) led to an increased expression of several HOX genes as well as other transcription factors notoriously critical for HSCs maintenance. Moreover, retroviral infection of mouse HSCs with NF-YA*s* increased their ability to repopulate the bone marrow of immunocompromised mice [152]. In the embryonic stem cells (ESCs) landscape, NF-Y importance is suggested by the fact that NF-Y sites are highly enriched at active enhancer regions in ESCs [153]. Between the two isoforms, NF-YA*s* predominates in ESCs and has a global role in stemness maintenance. Indeed, a switch between the short and the long isoform occurs during differentiation of mouse ESCs. A major proof of NF-YA*s* involvement in stem cells maintenance came from its ability to maintain stemness features upon TAT-fusion protein transfection after LIF (leukaemia inhibitory factor) withdrawal [154]. In particular it has been depicted a strong relationship between NF-YA*s* and NANOG, a well characterized stemness TF, with 30-50% of NANOG peaks having an underlying NF-Y site, and NF-Y binding being a prerequisite for association of NANOG to

DNA. These findings, together with recent publications (see paragraph 1.5.4), push toward including NF-Y in the restricted circle of TFs that govern embryonic stemness circuitry.

Although no comprehensive phylogenetic study has yet been done that includes all three NF-Y subunits and sequences from all major phyla, the NF-Y complex has indeed ancient origins. Sequence alignments and a consistent mass of experimental evidence demonstrate that clear homologs of its subunits can be found in all metazoa, plants, fungi and some protists, pointing at NF-Y occurrence at the basement of the eukaryotic domain. Moreover, NF-YB/NF-YC core domains show clear homology with prokaryotic archeal histone proteins. The yeast homolog was independently discovered as a regulator of the UAS region of cytochrome genes, where the three subunits were named HAP2/3/4 [155]. In plants universe, NF-Y genes underwent an explosive expansion, with more than ten genes per subunit, giving rise to multiple tissue-specific and stimulus-specific heterotrimeric combinations [156]. In higher plants NF-Y is involved in embryo and in root development, response to drought and nutrient scarcity, plant-microbiome interactions and photoperiodic flowering [157][158][159]. Concerning invertebrates, research on NF-Y is lagging behind, although well-defined roles in developmental and stress response pathways in *Drosophila* and *C. elegans* are well established [160]. Within mammals, the conservation at protein level is extreme, with more than 100 UniProtKB entries from other species for each subunit that share more than 90% sequence identity with the human orthologues. For instance, mouse and human share an ~80 million years old common ancestor, yet their NF-YA proteins are identical, except for a single substitution and an additional glutamine residue within the N-terminal domain in one of the splicing isoforms.

On top of the remarkable evolutionary conservation, NF-Y subunits expression is ubiquitous in the vast majority of human and mouse cell types, suggesting a basal role in tissues homeostasis. Indeed, NF-YA knockout mouse is embryonic lethal due to a block of cell proliferation and induction of apoptosis, thus indicating NF-Y as essential during early developmental stages [161]. Recently, several groups developed murine conditional knockout NF-YA alleles (**Fig 15**); this approach definitely confirmed that post-natal tissues count on NF-Y for their functionality, as evidenced by the subsequent severe abnormalities in all body districts that have been tested, including liver, forebrain, adipose tissue and bone marrow [162][163][164][165].

On the other hand, key exceptions to NF-Y ubiquity exist. In human monocytes, the expression of the NF-YA subunit is undetectable at protein level, but it increases to high levels during maturation to macrophages, along with the expression of MHC-II surface antigens [166]. A second, well characterized system where variation in NF-Y expression accompanies cell differentiation is skeletal and cardiac muscle tissue. During the transition from proliferating myoblasts to differentiated myotubes NF-YA expression is shut down and in adult mouse skeletal

and cardiac cells the protein is definitely absent [167][168]. These specific systems exemplify the central role of the NF-YA subunit, which is commonly in charge as the regulatory component of the trimer, as evidenced in following sections of this thesis.

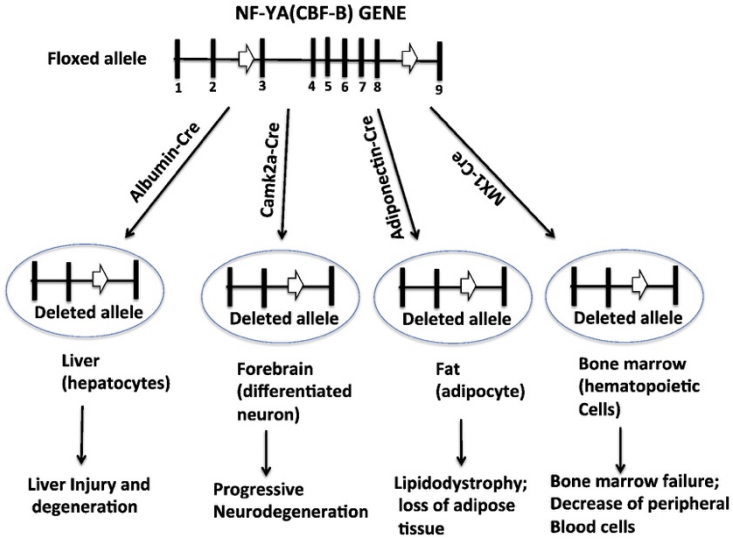


Figure 15. Conditional knockout of the NF-YA gene in four different tissues in mice after birth. A conditional NF-YA allele was designed by inserting loxP sites (white arrows) flaking exon-3 and exon-8; mice with a floxed NF-YA allele were crossed with transgenic mice expressing Cre-enzyme in specific tissues. This setup allowed tissue specific NF-YA deletion at post-natal stage, leading to the indicated phenotypes (from [169]).

1.5.3 The structural basis of the NF-Y unique DNA-binding mode

The recent progresses made in the structural field of NF-Y biology provided valuable advancements both in understanding the fine details of its molecular architecture, and in setting the bases for new and exciting approaches to study the molecular biology and genetics of this TF.

Since the first characterization of the primary sequence of NF-YB and NF-YC subunits, it was evident that their core domains share high similarity to histones H2B and H2A, respectively [170], and that they use these regions to form a tight 'histone-fold' heterodimer. It was also firmly established by biochemical assays that the formation of this NF-YB/NF-YC dimer is a prerequisite for the association of the third subunit, NF-YA, which is devoid of any clear homology with other known proteins [171]. The development of co-expression techniques [172] led to the first crystal structure of the human histone-fold (HFD) NF-YB/NF-YC dimer [173]. This confirmed the histone-like architecture of the dimer and finally set the bases for the subsequent crystallization of the NF-Y core heterotrimeric complex bound to CCAAT-DNA, both for mammalian [174]and fungal CBC homolog models [175]. These crystals were derived from the minimal trimerization/DNA-binding conserved core domains of each subunit in order to maximize crystallization capability, since the inclusion of the flanking unstructured regions would have possibly hindered the formation of an ordered crystal lattice. The human heterotrimer was

complexed with a 25 bp double-stranded DNA oligonucleotide harbouring the CCAAT from the HSP70 promoter.

The structural homology of the NF-YB/NF-YC HFD with histone-folds of H2B/H2A is evident from the superimposition of the respective quaternary structures (**Fig 16A**) and extends to the additional C-terminal loop-α-helix element (named αC) present in both subunits. The NF-YB/NF-YC HFD is further stabilized by intra-chain Arg-Asp bidentate salt bridges between loop 2 (L2) and α-helix 3 (α3) in each subunit. The result is a compact heterodimer stabilised by extensive hydrophobic interactions between the two HFDs arranged in a "hand-shake" assembly. The upper region of the HFD exposes an extended basic surface, responsible for non-specific contacts with the negatively-charged DNA backbone.

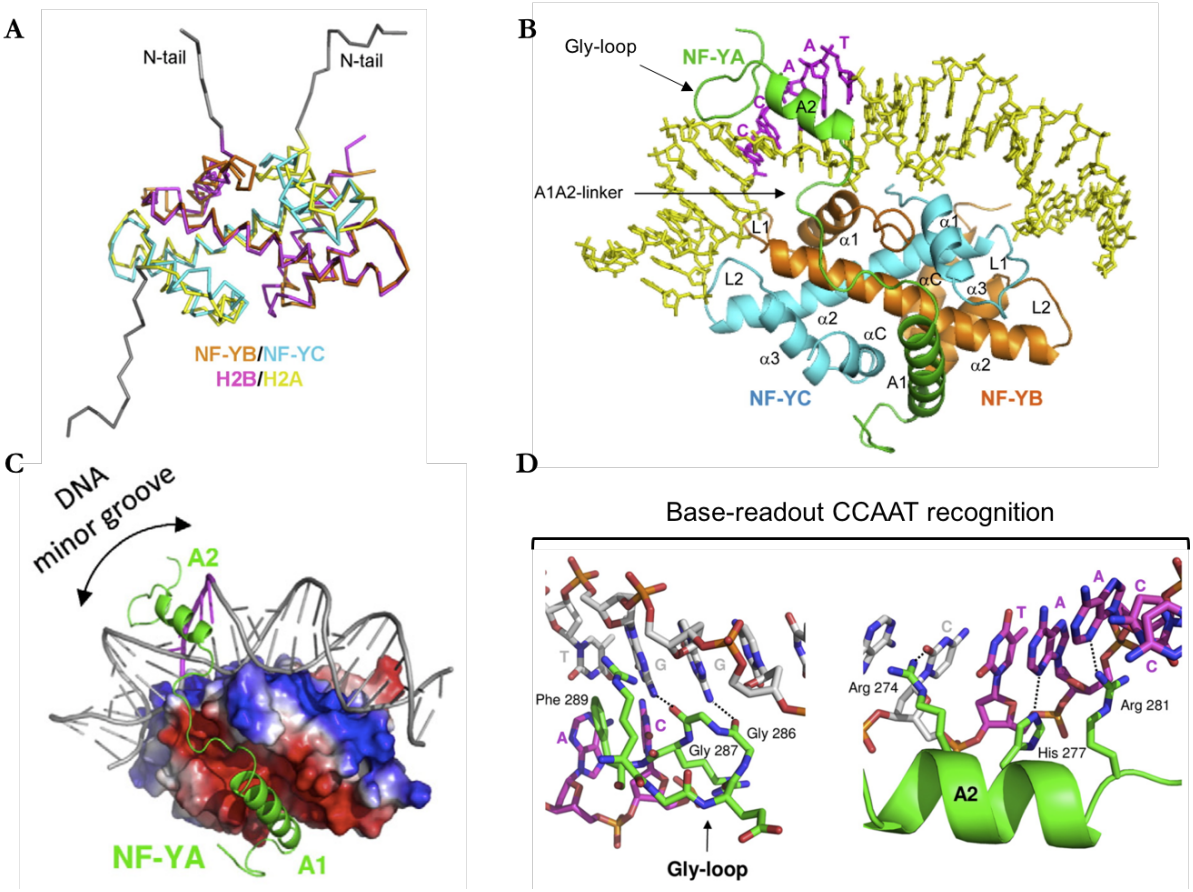


Figure 16. (**A**) Structural superimposition of coil diagrams between NF-YB/NF-YC HFD dimer (PDB ID: 4CSR) and H2B/H2A dimer (PDB ID: 1AOI, chains C and D). (**B**) Ribbon representation of the NF-Y core heterotrimer in complex with a 25 bp oligonucleotide from HSP70 promoter CCAAT-box (PDB ID: 4AWL). (**C**) Electrostatic surface of HFD NF-YB/NF-YC dimer in complex with NF-YA and DNA as in (B). Red and blue colors indicate negatively and positively charged regions, respectively. NF-YA minor groove insertion is highlighted. (**D**) Close-up on the base-readout mechanism employed by NF-YA to specifically recognize CCAAT nucleotides. The CCAAT nucleotides are in magenta and complementary strand bases in grey color; hydrogen bonds interactions involving Gly-loop and A2-helix residues are indicated by dashed lines (from [174][176]).

The structural hallmark of the NF-Y HFD dimer is a negatively charged groove composed by NF-YC α1 and αC, and NF-YB α2. This unique feature provides a docking site for the third subunit, NF-YA (**Fig 16C**). Interestingly, the core HFD of the NF-YB/NF-YC dimer also harbours structural elements responsible for the interaction with other specific TFs: αC helices from NF-YB and NF-YC are responsible for interaction with TBP [177], and NF-YC αC can interact with the tumour suppressor p53 and with the proto-oncogene c-MYC [178][179].

While NF-YB/NF-YC heterodimer provides a wide positive surface for non-specific DNA contacts, NF-YA is devoted to CCAAT recognition, thus accounting for DNA-binding specificity of the trimer. The core-domain of NF-YA subunit is less than 60 amino acids (aa) long and is located at the C-terminal portion of the protein. It is composed by two α-helices, A1 and A2, separated by a 12/13 aa flexible linker, and few residues following helix A2 (**Fig 16B**). A1 α-helix docks in the extended acidic groove displayed by the HFD thanks to specific positively charged side-chains, thus allowing trimerization. Helix A2 and the subsequent Gly-loop motif (Gly-X-Gly-Gly-Arg-Phe, with X=any residue) impose CCAAT-binding specificity by deeply inserting into the DNA minor groove. This results in an extreme widening of the minor groove, up to 19 Å at the middle CCAAT adenine (the canonical width for the minor groove in B-form DNA is 11.7 Å). Moreover, the Phe318 side-chain belonging to the Gly-loop is found inserted between the two consecutive C and A base pairs within the CCAAT, providing stacking interactions and inducing a positive roll of 48° between the base pair. The concerted actions of NF-YB/NF-YC convex DNA-binding surface and minor-groove invasion by NF-YA result in a global DNA bending of ~80°, in accordance with previous circular permutation experiments [180].

The sequence specificity is achieved thanks to Arg and His residues of the A2-helix, that contact specific functional groups of the CCAAT strand nucleobases, and to the kinked backbone of the Gly-loop, that positions the two Gly-Gly main chain carbonyl oxygen atoms in hydrogen-bond contact with the specific exocyclic amino group of the two guanines of the CCAAT complementary strand (**Fig 16D**). The register of the sequence specific contacts is stabilized by several additional NF-YA-mediated interactions with the DNA phosphate backbone.

The resulting NF-Y/DNA complex has a compact aspect, with the bent DNA that faithfully follows the HFD upper convex surface. Note that, however, both NF-YA and NF-YC full-length proteins possess long and disordered N- and C-terminal Gln-rich transactivation domains, respectively. These regions make up a large portion of the full-length complex and they must be kept in mind when looking at the crystal structure of the "minimum-domain" (DNA binding) complex of NF-Y. Nevertheless, years of TFs research clearly showed that in general the DNA-binding and the transcription-effector portions of the protein constitute two distinct and semi-autonomous functional domains [181], and the picture derived from the X-ray crystallography

experiments is in remarkable accordance with previous *in vitro* and *in vivo* observations on the full-length complex.

Besides DNA bending, a striking feature of the complex is the extension of the contacted DNA, spanning at least 25 bp. There are 41 detected protein-DNA contacts, widely distributed along the protein-DNA interface, explaining the high-affinity of the complex for CCAAT DNA. Given the extension of the surface interested by this 'bimolecular' interaction, one can speculate that the binding event is coupled with the release of a significant number of water hydration molecules and counter-ions from both protein and DNA surfaces, which would provide a strong entropic driving force for complex formation.

By comparing the NF-Y DNA-binding logo (**Fig 13A**) with these structural observations it is evident that the sequence preference outside the CCAAT pentanucleotide cannot be explained by a simple base-readout model, since there are not direct and specific interactions with the flanking DNA nucleobases (except NF-YA Arg303, which in the crystal directly contacts the cytosine after the CCAAT). The most direct explanation would be a shape-readout mechanism, possibly due to local DNA bendability or slight differences in the DNA electrostatic potential. Indeed, core histones themselves tend to have a preference for base composition that enhance nucleosome positioning *in vitro* and *in vivo*, still lacking any direct base-readout moiety [182][183].

The combination of HFD non-specific DNA contacts provided by NF-YB/NF-YC dimer and CCAAT recognition provided by NF-YA subunit might suggest a scouting model in which the preassembled heterotrimer transiently contacts the DNA backbone through its 'deviant histone' HFD component. The complex could proceed by local sliding on the DNA surface, while the flexible NF-YA A1A2-linker allows the A2-helix residues to constantly search for high-affinity CCAAT sites, thereby locking the complex in place with the Gly-loop and phenylalanine side chain insertion. Interestingly, the A2 region is unstructured in the complex prior to DNA binding [175].

1.5.4 The 'pioneering' role of NF-Y

The unique DNA-binding strategy adopted by NF-Y might well be related to one of the recently discovered properties of NF-Y function in the nucleus. Genome-wide mapping of NF-Y binding sites from the ENCODE project from distinct cell lines revealed that a significant portion of NF-Y locations fall in repressed chromatin regions or devoid of any positive histone mark [142].

The vast majority of TFs are unable to gain access to repressed or non-modified chromatin domains, even if high-affinity binding sites are present. Exceptions to this paradigm are the so called 'pioneer' TFs, whose capability to colonize neutral or hostile chromatin environments is of paramount importance during cell differentiation, cell reprogramming and establishment of altered transcriptional patterns in cancer cells [184][185], where they have been used as prognostic

biomarkers. Pioneer TFs are grouped according to their distinctive biological function rather than their structural homology. The main function of this subclass of TFs is to establish competence for gene expression, either by promoting the binding of a second, non-pioneer TF through direct cooperativity or by recruiting chromatin remodelling/modifying complexes which in turn physically provide DNA access to other TFs. Thus, the binding of a pioneer TF on a repressed regulatory region precedes, but is not necessarily concomitant with, gene expression.

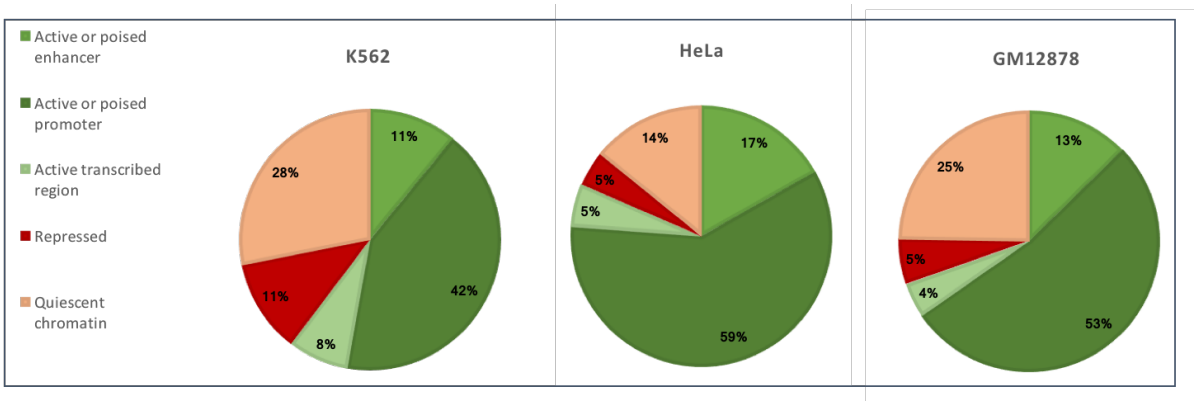


Figure 17. Distribution of NF-YB binding sites according to chromatin state in three ENCODE cell lines (from [147]). Note the consistent fraction of sites located within quiescent or repressed chromatin.

Besides the very first genomic observations, the following studies confirmed the pioneer nature of NF-Y with different methods and in different systems. First, a new unbiased computational approach to analyse genome-wide DNaseI hypersensitivity data at unprecedented detail was used [186]. It uses machine-learning methods to predict TFs binding from the shape and magnitude of DNaseI profile and a list of position weight matrices (PWMs) describing the specific binding motifs of candidate TFs. This method was applied to differentiating mouse embryonic stem cells (mESCs) and led to the identification of a handful of predicted and validated pioneer TFs. Among them, NF-YA was identified as a 'directional-pioneer', since the open-chromatin state was asymmetrical and strongly enriched towards the CCAAT downstream region. A second paper found that NF-Y promotes chromatin accessibility for the key pluripotency factors Oct4 and Sox2 in mESCs, and that binding of these master TFs is dependent on NF-Y, specifically at co-bound sites [187]. A third, recent study aimed to shed light on how the chromatin accessibility landscape is established during the zygotic to pre-implantation stage of mouse embryonic development [188]. They found the CCAAT box as the most enriched motif in DNaseI hypersensitivity sites (DHSs) at the 2-cell stage embryo and that NF-YA is one of the TFs responsible for the zygotic genome activation, that resumes transcriptional activity from the previously inert oocyte and sperm genomes. Taken together, all these evidences disclosed the pioneering role of NF-Y and contributed to define it as an essential component of the core pluripotency TFs network.

The real challenge set by these studies would be to define the molecular mechanism by which NF-Y would be able to gain access to repressed chromatin DNA. Concerning the few other examples of TFs studied in their pioneer activity, they all share the capability to recognize their binding site on nucleosomal DNA [189][190]. All pioneer TFs tested so far likely bind their target sites on the surface of a nucleosome, to then establish and extend a competent chromatin environment for other TFs. However, this nucleosome-surface binding model would be hard to adapt to NF-Y structural features, especially for its extensive DNA contact surface. An appealing model would assume that NF-Y pioneer activity would rely on its HFD subunits and their natural homology with H2A/H2B core histone proteins. Indeed, NF-YB/NF-YC dimer has been reported to interact with histone H3/H4 tetramers during nucleosome assembly *in vitro* and NF-YA can interact to such preformed NF-YB/NF-YC/$(H3/H4)_2$ complex and impart CCAAT specificity [191][192]. These observations led to the intriguing hypothesis of a "hybrid nucleosome", where the association of NF-Y heterotrimer with a H3/H4 tetramer would give rise to a hybrid particle that would recognize CCAAT-box DNA and generate an alternative, sequence-specific nucleosome, potentially working as a bookmark for regulatory regions [176]. This fascinating hypothesis has yet to be thoroughly tested both *in vitro* and *in vivo*.

1.5.5 NF-Y works as a multifunctional context-specific interactor

How does NF-Y exert his molecular functions? The basic, yet puzzling answer is suggested by simply looking at current protein binary interactions repositories for NF-Y subunits interacting partners. For example, we can currently retrieve a total 102 non-redundant interactors for the NF-Y complex from the BioGRID database. Most of them fall in predictable classes of nuclear proteins, mainly other TFs, components of the basal transcriptional machinery, transcriptional co-factors and enzymes, including kinases, methyltransferases, histone-deacetylases, and lysine-acetyltransferases. It is clear that NF-Y, as the majority of sequence specific TFs, works by interacting with other proteins, by recruiting or stabilizing them on CCAAT-box DNA. From a functional point of view NF-Y behaves as an organizer transcription factor. Generally speaking it can be represented as a promoter architectural protein that works by means of its stunning versatility in terms of protein-protein interactions and multiplicity of high affinity binding sites throughout the genome. Therefore, his role as facilitator of other TFs binding is not exclusively due to its intrinsic capacity of accessing to chromatin, bending DNA and exposing other TFs recognition sites, but also to its ability of directly interact with many of them.

To list all of the known NF-Y interacting partners found in literature is beyond the purpose of this thesis, but I will provide few examples to highlight the trend on NF-Y-mediated protein-

protein interactions and then I will introduce a tightly related and fascinating subtopic, that is the existence and organization of NF-Y-directed regulatory modules throughout the genome.

The most direct mode of NF-Y-mediated transcriptional activation is the ability to directly interact and recruit components of the GTFs on promoter DNA, such as TFIID. Specifically, NF-Y can directly bind TBP and several histone-fold TAFs thanks to NF-YB/NF-YC heterodimer, thereby priming the promoter for PIC assembly [177][193].

NF-Y DNA-binding cooperativity with other sequence-specific TFs has been widely demonstrated, starting from the system where it was originally discovered (MHC-II), where it cooperatively interacts with RFX [194]. Since then, other specific NF-Y cooperative systems have been studied, such as the long-standing interplay with SP1 [195][196], cooperativity with C/EBPα [197], direct recruitment of the endoplasmic reticulum stress-activated ATF6 [198] or formation of a ternary complex with USF on HOXB4 promoter [199].

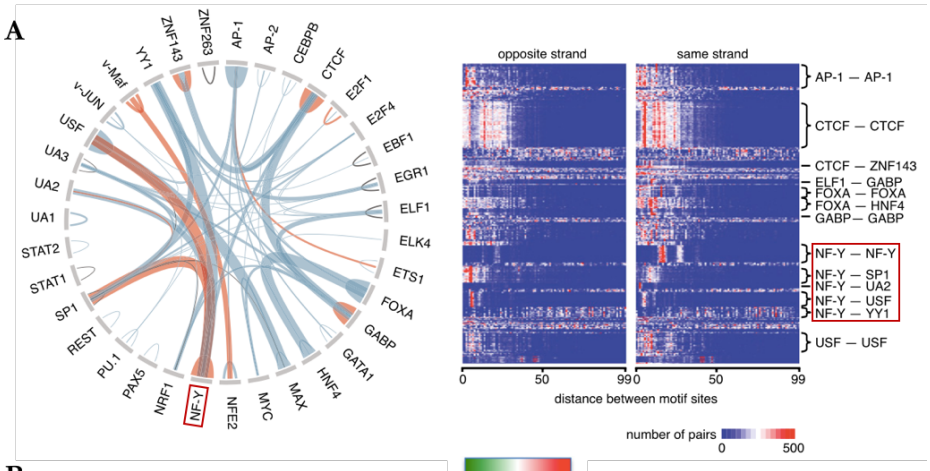
Strikingly, there are many reports in which a TF is found associated to a DNA region that lacks its respective canonical RE, but display a CCAAT-box with NF-Y bound to it, suggesting an either direct or indirect recruitment of the TF by NF-Y. Some notable examples of this "tethering" function of NF-Y are the following:

- the tumour suppressor p53 is found bound on several G2/M transition promoters through multiple NF-Y binding platforms which are indispensable for p53-mediated transcriptional repression of these genes upon DNA damage [178].
- NF-Y is extensively found colocalized with FOS in several genomic contexts in absence of its natural partner JUN and its cognate sequence, the AP-1 motif [142][200].
- IRF3 lacks the expected binding motif for the IRF family, instead its most enriched binding motif upon IFNγ stimulation is found to be the CCAAT-box [201].
- the DBD of Sp2 is completely dispensable for its genomic interactions, rather, its Gln-rich amino-terminal region is sufficient for its recruitment to target promoters *in vivo*, where it overlaps with NF-Y sites, possibly via direct tethering [202].
- the cell-type specific SOX9 TF is found co-associated with NF-Y sites in a DBD-independent fashion onto ubiquitously expressed cell-cycle genes in colorectal cancer cells [203].

Indeed, the CCAAT-box is found to be the most enriched motif in ChIP-seq peaks for TFs different from NF-Y [204], suggesting a role of NF-Y as a transcriptional landing platform. A second, intriguing interpretation of these ChIP-seq peaks on regions devoid of the expected binding motif could be a 3D chromatin looping effect: a TF bound to its cognate site that is involved in a long-range interaction by direct or indirect contact with a second TF, bound elsewhere in the genome.

This scenario could lead to a co-immunoprecipitation of the anchored distal region together with the one actually bound by the TF. Therefore, to distinguish between simple tethering and tethering-with-looping, ChIP-seq experiments could be complemented with 3D-chromatin analysis techniques (e.g. Hi-C or ChIA-PET).

A recent growing line of evidence, coming from genome-wide approaches to study TF binding-sites distribution, is revealing fascinating transcription factors co-association modules, in which specific combinations of TFs are found bound together in defined arrangements, with recurrent distances among their binding sites on DNA [205][201][206]. This higher order "code-above-the-code" is at the heart of the combinatorial nature of transcription regulation in higher eukaryotes, and NF-Y undertakes a central role in this scenario.



B

Family	Factor	DNA binding site	TSS	ENH	TX	Other	Repeats
E2F	E2F4	CCAAT+E2F	1.21	-2	-5.41	-8.39	-3.73
b-HLH	USF1	CCAAT+Ebox	0.22	0.95	-0.58	-0.66	0.09
b-HLH	USF2	CCAAT+Ebox	0.1	0.55	-0.41	-0.24	0.24
GTF	TBP	CCAAT+TATAbox	1.15	-1.23	-2.96	-4.94	-2.07
b-Zip	FOS	Double CCAAT	0.81	-0.12	-1.13	-1.59	-1
ZNF	Sp1	Double CCAAT	1.01	-0.99	-1.32	-2.46	-1.83
ZNF	Sp2	Double CCAAT	1.1	-1.29	-2.79	-3.08	-3.21
Other	RFX5	Double CCAAT	1.03	-0.61	-1.7	-3.09	-2.63
Other	IRF3	Double CCAAT	1.19	-1.95	-4.42	-5.44	-3.68
Cofactors	CHD2	Double CCAAT	1.13	-1.19	-3.05	-3.91	-2.81

Figure 18. NF-Y co-associates with several TFs. (**A**) Circos plot (left) depicting motif pairs presenting significant distance preferences between their sites. Each motif pair is connected by an arch; the thickness of the connection is proportional to the frequency of the pair. Red connections indicate that the pair is found in multiple datasets. The right panel represent a heat-map of the distribution of distances between each motif pair, where each row is a motif pair in a ChIP-seq dataset and each column is the edge-to-edge distance between the two sites. NF-Y is highlighted as a red box (from [206]). (**B**) NF-Y co-associated factors with global enrichment. For each factor, the binding motifs enriched on NF-YB co-bound sites are indicated. The heat-map shows whether the extent of binding overlap with NF-Y sites for each chromatin region is below (green) or above (red) the expected value. The numbers refer to the log2 of the ratio observed/expected number of regions (adapted from [147]).

The first genome-wide observation of NF-Y co-association with other TFs came with an unbiased bioinformatic study based on binding sites mapping from extensive ChIP-seq datasets of human TFs generated by the ENCODE consortium [206]. NF-Y emerged as widely involved in

the tethering process of and co-association with other TFs, showing preferential CCAAT orientation and distance from a nearby site occupied by SP1, PBX3 or USF. This 'motif stereo-positioning' between CCAAT-boxes occupied by NF-Y and associated binding sites for other TFs was confirmed in a second, independent study for the AP-1 motif, TATA-box, E2F element, and MAX or USF-bound E-boxes [142]. A third study aimed to thoroughly characterize NF-Y "neighbour TFs" by specifically assessing the presence of CCAAT-boxes and enrichment for NF-YB underneath the genomic bound loci of 154 human TFs from 380 different ChIP-seq experiments [207]. From here, 33 factors were found with a significant CCAAT-box enrichment within their peaks, and were ranked in three groups based on the relative extent of this enrichment. The first group represents all those factors for which the CCAAT-box was the primary motif that emerged within their binding sites, including FOS, CHD2, IRF3, RFX5, PBX3, SP1 and Sp2. For some of them, as already mentioned, their canonical binding motif is absent in presence of NF-Y, suggesting NF-Y-dependent tethering events. The other two groups include factors for which the CCAAT-box represents the second-most-enriched motif within their binding sites, either at a global (second group) or local scale (third group). By analysing the occurrence of preferential spacing between the respective motifs in the co-bound regions, a series of TFs showing a significant positional-bias was retrieved (**Fig 18**). The most striking are E-boxes bound either by MAX or USF1/2 (but not MYC) and motifs for the zinc-finger protein ZNF143 (the latter found locally enriched in repetitive regions, see also **Fig 19D**), that show almost exclusive distance biases from the CCAAT-box of 12 bp and 36 bp, respectively. A second, interesting observation was that for many of the first-group TFs, namely FOS, IRF3, SP1, SP2, PBX and RFX5, the co-bound sites with NF-Y displayed a double CCAAT-box configuration, usually with a distance of 24-28 or 35-37 bp between the two NF-Y sites. This widespread module formed by a platform of two NF-Y sites associated with the TFs described above can be found in more than 1000 promoters in the cells lines analysed. These observations were then broadened with chromatin state annotations, on one hand confirming that most factors co-associate with NF-Y at active TSSs (e.g. FOS, TBP, E2F4), with the exception of USF1/2 that are found globally enriched also at enhancers [147]. Analogous results were recently obtained by an independent work that employed a new bioinformatic approach, named regulatory module discovery (RMD), applied to K562 cells ChIP-seq dataset to capture TF-TF correlation in a genome-wide scale [208]. Here, the already described NF-Y+FOS association could be further split into NF-Y+FOS+SP module, enriched at TSSs, and NF-Y+FOS+USF module, enriched at distal regions.

1.5.6 NF-Y and the E-box transcription factor USF

As described in the previous paragraph, one of the most striking NF-Y genomic co-associations is the one with the ubiquitous transcription factor USF (Upstream Stimulating Factor). Here I provide a closer examination on the biology of USF, specifically concerning the systems where its relationship with NF-Y was documented, since the interplay between these two TFs is at the heart of the second project I am involved in (see Part III).

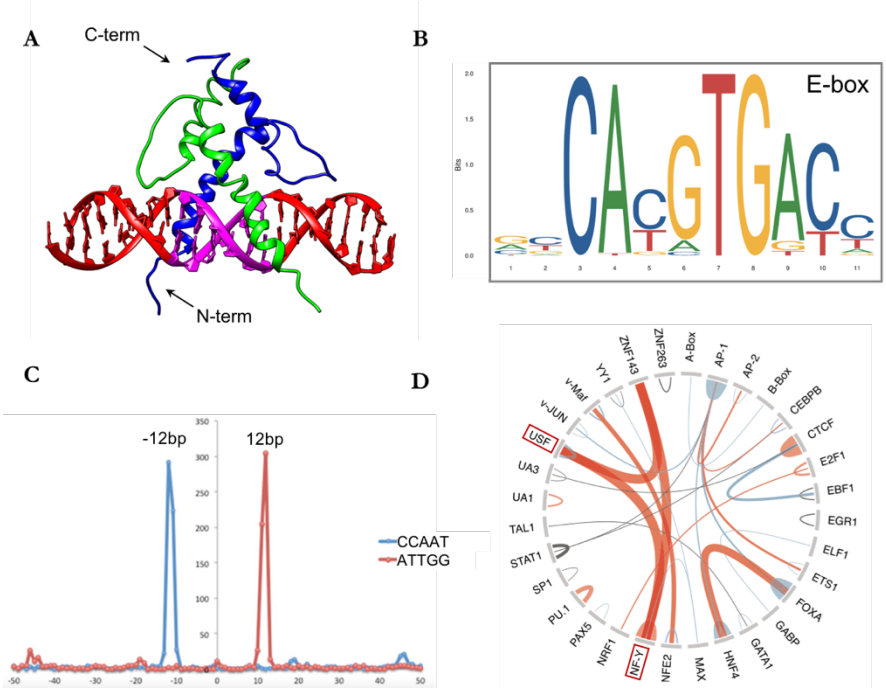


Figure 19. (**A**) Ribbon representation of the crystal structure of USF1 dimer (green and blue) bound to E-box DNA (in magenta), only comprising the bHLH-domain of the protein (PDB ID: 1AN4). (**B**) E-box sequence logo of the ChIP-seq derived USF1 binding motif from JASPAR database. (**C**) Distribution of the relative position between the E-box and the centred CCAAT-box in regions co-bound by USF1 and NF-Y in K562 cells, for both the forward (CCAAT) and reverse strands (ATTGG) (from [147]). (**D**) Circos plot depicting motif pairs presenting significant distance preferences between their sites as in Fig 18A, but referred to motifs discovered in repetitive regions. NF-Y and USF1 are boxed in red (adapted from [206]).

USF was first identified in human cells as a TF involved in the control of the adenovirus major late promoter [209], and later resulted in the isolation of two related genes, *USF1* and *USF2*, located in human chromosomes 1 and 19, respectively [210][211]. In mammals the two paralog genes have a similar genomic organization with 10 exons. The *USF* genes products are transcription factors belonging to the basic helix-loop-helix leucine-zipper (bHLH-ZIP) family of TFs, which among others includes also MAX and MYC. All these proteins work as dimers and share a common C-terminal DBD architecture, composed of a basic helix involved in sequence-specific DNA contacts, connected through a stabilizing loop to the leucine-zipper helix responsible for homo- and hetero-dimerization [212][213]. While the USF1 and USF2 bHLH-ZIP domains share high similarity (70% sequence identity), they are quite divergent in their N-terminal effector

domains, except for a disordered stretch of ~150 aa immediately upstream of the basic region, the so-called USF Specific Region (USR). The USR region, encoded by exon 6, has shown to work as a context specific transcriptional activation domain (TAD) essential for USF-dependent promoter activity [214][215], although the mechanism employed is completely unknown.

The symmetrical E-box (CACGTG) is the preferred binding site for USF, as for many other bHLH-LZ TFs; however, the flanking sequence can affect the binding preference, possibly specifying the *in vivo* genomic distribution among these related TFs, although the promoter-context has indeed a major role in allowing binding competence with other TFs nearby. Conversely, methylation of the CpG site central to the E-box severely impairs DNA-binding [216].

Even before genomic data on TFBSs distribution were available, NF-Y and USF were shown to co-regulate different promoters by independent groups [217][218][219]. Not surprisingly, in several examples the two REs (E-box and CCAAT-box) were separated by 10-12 bp, as would be predicted by the already described genome observations [142][207][147]. For instance, the expression of sodium-vitamin C co-transporter 2 (SVCT2) gene driven by its CpG-poor promoter is dependent on an E-box bound by USF and CCAAT-box bound by NF-Y spaced by 12 bp. It was demonstrated that the spacing between the two sites is critical to achieve DNA-binding cooperativity and a synergistic transcriptional activation [220]. Analogous configurations were found and studied for HOXB4 (10 bp spacing) and HOXB7 (12 bp) promoters [199][221].

A curious observation coming from genomic studies is that USF1 is one of the few TFs that co-associate with NF-Y in non-modified chromatin regions, in particular at endogenous long terminal repeats (LTR) [142]. The precise spacing of the two sites could provide the cooperative binding necessary to access or to cling on to these hostile chromatin environments, perhaps contributing to the active repression of retrotransposons.

1.5.7 The NF-Y connection with the epigenetic landscape

The list of proteins that have shown direct or indirect association to NF-Y is continuing to grow, and besides sequence specific and general TFs it comprises co-activators and co-repressors, including chromatin modifying and remodelling complexes. Among them, we find the histone-acetyltransferases (HATs) P/CAF, GCN5, p300 and CBP [222][223][148][224], the histone-deacetylase HDAC1 [225][226], the ASH2L subunit of the H3K4 histone-methyltransferase complex MLL [227] and the kinases JNK (a stress-activated kinase shown to phosphorylate H3S10) and ERK2 [228][229]; moreover, a tight NF-Y genomic co-association was found with the co-repressors proteins CoREST, SIN3A, KDM5B, HCFC1 and with the CHD2 nucleosome remodeller, member of the SNF2-like family of helicase-related enzymes [207].

Clearly, the association with this wide variety of chromatin modifying factors links NF-Y to epigenetic regulation of gene expression. A direct or indirect NF-Y connection with local epigenetic environment has been found in several studies. For instance, one of the earliest histone marks applied on transcriptional competent promoters, the mono-ubiquitination of Lys120 of histone H2B (H2BK120ub), is NF-Y-dependent. In particular, NF-YB mono-ubiquitination on Lys140, structural homolog of H2B Lys120, is required for H2BK120ub deposition on promoter regions [174]. Other positive histone-marks have shown NF-Y dependency, such as H3K4me3 and H3K79me2 [230] and acetylation of H3K9 and H4 [231]. The picture is complicated by the fact that NF-Y dependency is not restricted only to transcription-associated histone marks on active genes, but also to negative ones at repressed genes, as H4K20me3 and H3K27me3 [146]. Effectively, NF-Y is a bifunctional TF that can work as a molecular switch for either positive or negative histone marks on CCAAT promoters [230]. In this respect, one of the major challenges would be to isolate the molecular determinants that lead NF-Y to recruit either co-activators or co-repressors. Examples of promoters in which the same NF-Y/CCAAT platform sets the stage for a dynamic duet between co-repressors and co-activators exist in literature, and post-translational modifications (PTMs) can surely have a role [232][233][234]. These observations confirm the tight link between NF-Y and the surrounding chromatinic environment.

1.5.8 Post-translational modifications: additional layer of control on NF-Y activity

The similarities of NF-Y with core-histones in terms of evolution, structure, DNA-binding mode and relationship with chromatin can be extended also to the variegate post-translational modifications (PTMs) its subunits undergo. I will give some hints on the few NF-Y PTMs for which a function is currently known, while in the next paragraph I will focus on the background concerning the phosphorylation sites that represent the main topic of my project.

Starting from the NF-YC subunit, currently, no dedicated research can be found in literature that investigates its specific PTMs, except for a paper in which they showed a global reduction of NF-YC phospho-threonine signal upon γ-irradiation in HUVECs [233]. The best candidate residue would be Thr343, since it is found phosphorylated in proteomic analysis of PTMs, as indicated in the PhosphoSitePlus® repository (https://www.phosphosite.org). Additionally, few lysine residues flanking or within the HFD are ubiquitination sites, possibly involved in protein turnover or signalling.

Concerning the NF-YB subunit, it was shown to be acetylated by p300 and GCN5, a modification important for transcriptional induction, although the residues involved are still unknown [235]. Even though two phospho-threonines (Thr2 and Thr112) are retrieved from a proteomic dataset, the best characterized NF-YB PTM is indeed Lys140 ubiquitination and its

involvement in priming the deposition of the activating H2BK120ub mark, as already mentioned in the previous paragraph [174]. An additional modification the NF-YB fungal homolog (HapC) undergoes is oxidation of three cysteine residues in response to unbalanced redox-potential; the outcome is the inactivation of the protein through homodimerization and retention in the cytoplasm. This mechanism has yet to be investigated in higher eukaryotes, even though these sensor cysteine residues are conserved in human NF-YB [236].

Confirming its role as the main regulatory subunit, NF-YA is the most heavily modified component of the trimer. One of the major mechanisms used to control abundance of functional NF-Y heterotrimer in the cell is the degradation of NF-YA subunit through the polyubiquitin-proteasome pathway. At least four lysine residues falling within the conserved region of the protein are sites of polyubiquitination and are responsible for the relatively short half-life of the protein (~2h) [237]. On the other hand, p300 can acetylate six lysines located in the A1 helix and in the linker region of NF-YA core domain; these modifications prolong the relatively short half-life of the subunit by a classic mechanism: prevention of polyubiquitination-mediated protein degradation by the proteasome [237]. This subtle equilibrium driven by the competition between acetylation and ubiquitination impact on the protein half-life. However, it is not known whether NF-YA acetylation can alter other properties of the protein, as the interaction with the other subunits or its transcriptional activity.

The third kind of PTM detected on NF-YA is phosphorylation. At least three sites are well documented phosphorylation substrates: Tyr266, Ser320 and Ser326, all retrieved in unbiased phosphoproteomic studies. Tyr266 lies ahead of helix A1, within the conserved HAP2 region, while Ser320 and Ser326 are positioned within the NF-YA C-terminal tail, just downstream of the CCAAT-recognition Gly-loop (see Figure 1 in PART II). While no hints on the function of Tyr266 phosphorylation are present in literature, Ser320 and Ser326 were originally shown to be phospho-serines in a 2003 paper [238]. In the next paragraph I will provide the background underlying the first characterization of NF-YA as a phospho-protein in the context of cell-cycle regulation, from which my research project originated.

1.5.9 NF-Y dependent cell-cycle regulation is linked with NF-YA phosphorylation

Orderly progression of the cell cycle is effected through the periodic activation and inactivation of a series of cyclin-dependent kinases (CDKs). CDK activities can be regulated through fluctuant accumulation and degradation of their cognate cyclins, their transient nuclear localization and/or phosphorylation, and their association with different CDK inhibitors (CKIs). In turn, CDK/cyclin complexes drive cell-cycle progression by phosphorylating key regulatory proteins with precise timing. It is well established that retinoblastoma (RB) tumour suppressor

proteins coordinate cell-cycle entry and G1/S transition by inhibiting E2F transcription factors, hence transcriptional activation of S-phase genes. The CDK4/Cyclin D complex induced upon mitotic stimuli phosphorylates RB, promoting Cyclin E expression through its E2F-dependent transcriptional activation, culminating in CDK2 kinase activation (CDK2/Cyclin E complex) and S-phase progression. CDK2 associates with cyclin A during late S-phase and early G2, while giving way to CDK1 for cyclin A and cyclin B binding in late G2 and mitosis, respectively. This latter complex promotes G2/M transition [239].

One of the first insights that suggested the involvement of NF-Y in the regulation of cell proliferation came from the observation that the levels of its regulatory subunit, NF-YA, were modulated during the cell cycle, increasing during G1/S, reaching a peak during mid-S phase and dropping at G2/M transition [240]. Even more suggestive was the significant enrichment of CCAAT sites within cell-cycle genes regulatory elements [140], many of which have TATA-less promoters. Moreover, the expression of a dominant-negative NF-YA mutant unable to bind DNA induced a delayed induction of the S-phase in quiescent fibroblast cells after serum stimulation [241]. Many important cell-cycle related genes are transcriptionally regulated by NF-Y, such as CDC2, CDC25C, CCNA2, CCNB1, CCNB2, PLK1, POLA1, E2F1; moreover, and the Gene Ontology term *cell-cycle* is one of the most recurrent among NF-Y target genes [242][243].

Interestingly, it was found that NF-Y is a direct target of CDK2/cyclin A complex, which is able to phosphorylate two serine residues nearby the DNA-binding domain of NF-YA, namely Ser320 and Ser326 [238][244]. Supporting this, the global NF-YA phosphorylation was found to be cell-cycle dependent, peaking at G2/M, and CDK2 was found to be associated with NF-YA both *in vitro* and *in vivo* [238][245].

To assess the role of this PTM, a phosphorylation-deficient NF-YA mutant was employed, in which two target serine residues were mutated in alanine (S320A, S326A), named YA-AA. The overexpression of this mutant form led to the repression of NF-Y cell-cycle target promoters in luciferase assays, reduction of DNA binding both in electromobility shift assay, using a CCAAT-probe from CDC2 promoter, and in ChIP, looking at CCNB1 promoter probed with an α-NF-YC antibody. These deregulations induced by the phosphorylation-defective mutant also led to abnormal cell-cycle progression, with actual accumulation in G1 and G2/M phases, upon adenoviral infection of HCT116 cells. This kind of cell-cycle arrest at both G1 and G2/M has been mainly explained as a consequence of cell-cycle regulatory genes repression due to an impaired NF-Y binding efficiency on their promoters upon concomitant overexpression of YA-AA [244].

Recently, new evidences suggested an additional layer of complexity to the previously depicted scenario [245]. NF-Y has been shown to be directly bound by p21 on the PLK1 promoter (Polo-Like Kinase 1) upon DNA-damage response. PLK1 is S/T kinase that has important roles

in multiple phases during mitosis, involving centrosome maturation, mitotic entry and anaphase progression. Its expression has been shown to be suppressed by p53 and/or p21. In unstressed cells, CDK2 is found associated with the CCAAT-box containing region of PLK1 promoter, whereas its occupancy falls upon DNA damage, with concomitant recruitment of p21. Both *in vivo* and *in vitro* evidences demonstrated the interaction between NF-YA and p21 and depletion of NF-YA suppressed the p53-mediated protection from mitotic death after DNA damage. Notably, this interchange between CDK2/NF-Y and p21/NF-Y has been found also for a second G2/M critical gene, CDC25A. This new scenario does not conceive a DNA-binding decrease of NF-Y upon CDK2 inactivation by p21; instead an active repression effect is more realistic. In unstressed cells CDK2/NF-Y complex bound to the proximal CCAAT element functions as an activator, presumably by recruiting HATs; upon DNA damage, instead, the activation of p53 leads to increased p21 levels which cause CDK2 activity inhibition and its replacement with p21, thus forming a repressive p21/NF-Y complex that actively inhibits gene expression (presumably through HDACs recruitment).

The role of NF-YA phosphorylation in this model has not yet been studied. One of its potential roles could be to prevent p21 association to NF-Y and hence transcriptional repression of target gene, however new biochemical evidences are needed to uncover the hidden roles of this crucial cell-cycle dependent NF-Y PTM. Indeed, the molecular function of (CDK2-dependent) NF-YA phosphorylation is still unknown.

2. AIMS OF THE THESIS

During my PhD, I followed two independent lines of research, unified by a common theme. The first one refers to my own research project, which began with me, and deals with the molecular role of NF-YA phosphorylation sites as described in the submitted manuscript in PART II. The second project instead is about the interplay between NF-Y and E-box transcription factors on DNA, the bases of which derive from the main topic of the doctoral thesis of Mariangela Lorenzo [246]. I took the latter project on since my second PhD year, and I contributed to specifically characterize DNA-binding cooperativity end synergy between NF-Y and the E-box TF USF1, as described in PART III. Here I will briefly introduce the rationale and the aims of each project.

2.1 Molecular characterization of NF-YA Ser320 and Ser326 phosphorylation sites

The aim of my project was to characterize the molecular role of two interesting phosphorylation sites of the NF-YA subunit. As already described in paragraph 1.5.8, NF-YA was found to be a cell-cycle related phospho-protein in a single dedicated paper in 2003 [238]. Based on the framework of the latter, few subsequent papers confirmed the importance of NF-YA Ser320 and Ser326 phosphorylation sites in cell-cycle progression, expression of cell-cycle regulatory genes and G2 checkpoint arrest by p53 [244][247][248].

Two critical questions remained unsolved. First, NF-YA phosphorylation studies were limited in assessing its effect on cell-cycle related promoters due to the direct link with the cell-cycle regulator CDK2. However, there is no reason to exclude a wider phosphorylation-dependent modulation of NF-Y activity beyond target promoters involved in cell-cycle regulation, possibly even by other upstream kinases. If that is the case, this PTM could impinge on many other NF-Y regulated cellular processes; if, on the other hand, NF-YA phosphorylation happens in a tightly regulated and context-dependent fashion, it would provide a potentially new and specific molecular target to control NF-Y activity in living cells.

Second, the original paper concluded that NF-YA phosphorylation was an activating modification, which would be necessary for efficient DNA-binding of NF-Y [238]. The use of the phosphorylation-deficient mutant YA-AA showed an impaired DNA-binding and a correspondent reduction in expression of few NF-Y target genes; moreover, the overexpression of p21 (an endogenous CDK2 inhibitor) was again accompanied with a reduced NF-Y DNA-binding as assessed by EMSA. However, the molecular function of NF-YA phosphorylation remained controversial, since the two target serine residues are close to the conserved DNA-binding domain of the protein and the crystal-structure of the trimer bound to the CCAAT-box [174] confirmed their close relationship with DNA. Consequently, there is no obvious explanation for how the deposition of a bulky and negatively charged chemical group (phosphate) would be necessary to

improve NF-Y affinity for DNA. Therefore, we aimed to fill this structure-function gap to investigate how NF-YA phosphorylation at Ser320 and Ser326 could alter NF-Y function, in particular its DNA-binding properties.

2.2 Dissecting the interplay between NF-Y and USF1 on DNA

The second project stems from genomic observations linking CCAAT-boxes to E-boxes throughout the genome, as described in paragraphs 1.5.5 and 1.5.6. In particular, NF-Y co-associates with USF1/2 and MAX with a recurrent stereo-specific distance of ~12 bp between E-box and CCAAT-box [142]. The same is not true for MYC, although it is often found associated with NF-Y positive promoters. Our laboratory undertook the aim to verify at the molecular scale the basis of these genome-wide observations. In particular, the interplay between NF-Y and the E-box TFs MYC, MAX and USF1 was thoroughly analysed through DNA-binding assays using recombinant proteins, structural studies and ChIP approaches as indicated in the introductory section of PART III [246]. Three were the core outcomes: (1) NF-Y cooperatively binds DNA with USF1 at the specific 10-12 bp distance between CCAAT and E-box, but not with MAX or MYC; (2) the *in vivo* binding of USF1, MAX and MYC was altered for NF-Y positive promoters upon NF-YB inactivation; (3) the structural resolution of the ternary complex of NF-Y and USF1 on DNA by SAXS revealed that the two proteins interact thanks to the N-terminal intrinsically disordered portion of USF1, and the interaction surface depends on DNA configuration (10 vs 12 bp spacing). However, neither the low-resolution SAXS model nor the *in vitro* binding assays provided evidences to unambiguously correlate the USF1 segments responsible for NF-Y interaction with its known functional domains.

My contribution to the project aimed to (1) assess the transcriptional synergy between the two TFs on several natural promoters, (2) dissect the USF1 regions involved in DNA-binding cooperativity with NF-Y in distinct DNA configurations, and (3) set up the conditions to perform ChIP-seq experiments on USF1 upon NF-Y silencing in distinct cell lines to obtain a genome-wide view of the interplay between the two TFs on chromatin.

The unifying theme of this thesis is the study of the DNA-binding activity in transcription factors. Specifically, the common purpose of this work is to offer two independent and innovative examples on how a transcription factor DNA-binding activity can be *directly* modulated, either by a post-translational modification, or by the interaction with a second, stereo-positioned transcription factor.

3. MAIN RESULTS

During my PhD, I contributed to the investigation of two basic, yet poorly understood aspects of the molecular biology of the ubiquitous transcription factors NF-Y and USF1. Here I briefly introduce the main results that we could infer from our experiments, reported in detail in the manuscript in PART II for the first project, and in the form of results report for the second one in PART III.

3.1 The phosphorylatable Ser320 of NF-YA is involved in DNA-binding of the NF-Y trimer

In the first project, we aimed to characterize the molecular function of NF-YA Ser320 and Ser326 phosphorylation sites, the effects of which were originally studied in the context of cell-cycle regulation. Due to the lack of a specific α-phospho-NF-YA antibody, we adopted a site-specific mutational strategy, thus generating a complete panel of either phosphorylation-deficient (Ser to Ala) or a phospho-mimicking (Ser to Glu) mutants for both residues. We employed them to assess the NF-YA phosphorylation state in cells, to perform DNA-binding assays both at equilibrium and in dissociation kinetics and to evaluate their effects on natural NF-Y target promoters in reporter assays.

- *Ser320 is in direct contact with DNA phosphate backbone:* we analysed the published crystal structure of human NF-Y heterotrimer bound to HSP70 CCAAT-box DNA (PDB ID: 4AWL) to verify the 3D-space positioning of the main NF-YA phosphorylation sites. Tyr266 lies ahead of helix A1 and is oriented outwards the complex core, distant from any DNA interacting surface. Ser320 is positioned just downstream the NF-YA Gly-loop, the protein moiety responsible for direct sequence recognition, and its hydroxyl group is hydrogen bonded with the phosphate belonging to the central CCAAT adenine. Ser326 instead, although present in the crystallized construct, belongs to the disordered C-terminal tail of the protein and is not visible in the model. Therefore, we focused our study on the two latter residues, which could potentially affect NF-Y DNA-binding properties directly.
- *Ser320-P is the main NF-YA phospho-isotype in HeLa cells:* in order to assess the phosphorylation state of the protein in asynchronously growing HeLa cells we transfected Ser320-Ser326 double or single Ala mutants of the protein. We analysed cell extracts through Phos-tag™ technology coupled with immunoblotting. Ser320 resulted to be the only detectable phospho-isotype of the protein. Moreover, its abundance highly increased in mitotic, nocodazole treated cells, as it is also evident for the endogenous protein.
- *Ser320 kinetically stabilises NF-Y binding to DNA:* as the crystal structure suggested, we investigated for a direct role for Ser320 or Ser326 in DNA-binding. We exploited electromobility shift assays (EMSA) with recombinant NF-Y trimers harbouring either the wild-type NF-YA construct (YA3) or the respective serine mutant panel. The mutations did not alter the overall

affinity for the CCAAT-containing HSP70 probe. Interestingly, Ser320 Ala and Glu mutants induced a respectively moderate and drastic reduction in the stability of NF-Y/DNA complex over time, as evidenced by competition off-rate assays. Analogous results were obtained using HeLa nuclear extracts expressing full-length versions of the wild-type and Ser mutants NF-YA. These results demonstrate an unprecedented DNA-binding kinetic role for Ser320 (but not Ser326) in the stabilization of the DNA-bound trimer, and suggest a phosphorylation-dependent mechanism to evict NF-Y from its site. Importantly, we demonstrated that Ser320Ala is not a neutral mutation for protein function as previously assumed.

- *Ser320 mutants impair NF-Y dependent transcriptional activation:* in order to evaluate potential transcriptional effects driven by NF-YA Ser320 or Ser326 phosphorylation mutants, we set up luciferase reporter assays using two well established NF-Y natural promoters (RHOB and MDR1). In all cases, Ser320 mutants, but not Ser326, significantly reduced NF-Y dependent transactivation of the reporter promoter, with Glu phospho-mimicking mutant being more severe than the Ala-mutant. This result fits with our DNA-binding experiments, suggesting that a phosphorylation-dependent reduced DNA-binding stability can be translated in hampered promoter activation.
- *Evolutionary conservation of Ser320 in metazoa:* we performed multiple sequence alignments using the C-terminal portion of NF-YA human protein to explore the conservation level of Ser320 and Ser326 sites among other known orthologues. We found that Ser320 belongs to the evolutionary conserved core of the protein, since it is conserved in nearly all metazoa. Ser326 instead is part of a C-terminal amino acid block conserved throughout all vertebrate amniotes.

3.2 USF1 cooperativity with NF-Y is mediated by the USR domain

The second project deals with the interplay between NF-Y and E-box TFs, in particular I contributed to the detailed study of the heterotypic cooperativity between NF-Y and USF1 on distinct DNA configurations (see PART III for detailed data), obtaining the following results.

- *USF1 USR domain is necessary and sufficient for cooperative binding with NF-Y:* preliminary SAXS structural data gathered in our laboratory show that DNA-binding cooperativity between NF-Y and USF1 involves a large, yet not defined portion of USF1 N-terminal region. In order to restrict the USF1 portion responsible for this activity, we designed USF1 constructs with progressive N-terminal deletions and we expressed them as recombinant proteins. We then assayed cooperativity with NF-Y for each deletion construct through competition off-rate EMSAs using probes with different E-box/CCAAT-box spacing. In all cases, the presence of a complete USR domain in the USF1 construct was sufficient to promote full DNA-binding

cooperativity. Instead, the construct devoid of the entire USR domain lost any cooperative behaviour with NF-Y, irrespectively of the DNA configuration.

- *USR portions responsible for cooperativity vary depending on DNA configuration:* the use of an intermediate-deletion USF1 construct, which lacks half of the USR domain, revealed a markedly different behaviour depending on the DNA configuration used. In particular this construct maintained partial cooperativity only on the 12 bp-spacing DNA configuration (LTR and HOXB7), but it was completely lost in the 10 bp-spacing configuration (HOXB4), a clear sign of independent binding of the two proteins. This result validates the structural SAXS models, in which the two DNA configurations (12 bp vs 10 bp) entail different conformations of the USF1 N-terminal regions, suggesting distinct portions of the USR domain responsible for interaction with NF-Y.
- *NF-Y and USF1 synergise in several natural promoters with different configurations:* in order to test whether DNA-binding cooperativity could be coupled with transcriptional synergy between NF-Y and USF1 on natural promoters we used four distinct human promoters with either 10 bp or 12 bp spacing between E-box and CCAAT-box in luciferase-reporter assays. The extent of activation provided by single TFs clearly varied among the promoter tested. The co-transfection with the two TFs gave clear higher-than-additive synergistic activation for HOXB7 promoter (12 bp). For the remaining promoters, the degree of synergy was lower, indicating a strong context-dependent outcome from the interplay between NF-Y and USF1.

4. DISCUSSION AND FUTURE PERSPECTIVES

This thesis aimed to uncover novel molecular aspects of the biology of the two ubiquitous human TFs NF-Y and USF1. Both projects involved a direct characterization of the DNA-binding properties of these proteins, either in function of a PTM or a cooperative interaction. Both kinds of regulation are central to the function of many other known TFs and are dictated either by an upstream stimulus resulting in the activation of a specific TF-targeting kinase, or by the precise spacing of two adjacent REs for cognate TFs, prompting their local physical interaction.

Here I will discuss the key results obtained from each project, and I will comment on the specific points I find more relevant, also providing some conjectures to stimulate further studies. Please refer to PART II and PART III for the actual results reported for the first and second project, respectively.

4.1 Phosphorylation as a direct switch for NF-Y mobility on DNA

Compared to other aspects of the molecular biology of the TF NF-Y, dedicated studies on the specific repertoire of PTMs its subunits undergo are sparse. Filling this gap is not trivial, but it would provide invaluable hints on how the activity of a ubiquitous and pioneer TF that targets a vast portion of human genes can be specifically modulated.

This study contributed to clarify the molecular role of NF-YA Ser320 phosphorylation, providing evidence that this modification could impact on DNA-binding and hence transcriptional activity of the NF-Y heterotrimer. Ser320 side-chain directly interacts with DNA backbone as shown in the crystal structure of the complex; clearly, a phosphorylation at this position would disrupt this positive interaction, providing a net repulsive force against the DNA-phosphate backbone. Still, our results suggest that NF-YA phosphorylation on Ser320 would not evidently reduce the affinity of the NF-Y complex for DNA, as evidenced by our DNA-binding assays performed at equilibrium. Indeed, the Ser320-dependent interaction with DNA is localized at the extreme edge of the DNA-binding domain of the protein, representing the last contact NF-YA makes with DNA; therefore, its ablation is not predicted to alter the overall number of the core direct protein/DNA contacts (at least 41 for the entire trimeric complex). The result would be the loss of a single, external hydrogen bond, maintaining the overall thermodynamic properties of the complex (i.e. affinity) largely unaltered.

Instead, our results point at a drastic kinetic effect of Ser320 phosphorylation (mimicked here by our YA-ES mutant): this modification strongly decreases the stability, hence residence time, of the complex on DNA, uncovering a crucial stabilizing role for Ser320-mediated interaction with DNA. This latter point is confirmed by the milder, but rather significant effects of Ser320Ala substitution, both in impairing DNA-binding stability and transcriptional activation of target

promoters. Our observations can also explain the transcriptional 'dominant-negative' behaviour of the YA-AA mutant already described in the original paper, where it was interpreted as the effect of a merely phosphorylation-deficient mutant. Indeed, our data demonstrate that Ser320Ala is not a 'neutral' mutation, but rather it can strongly affect the behaviour of the protein, both *in vitro* and in transcriptional assays in living cells.

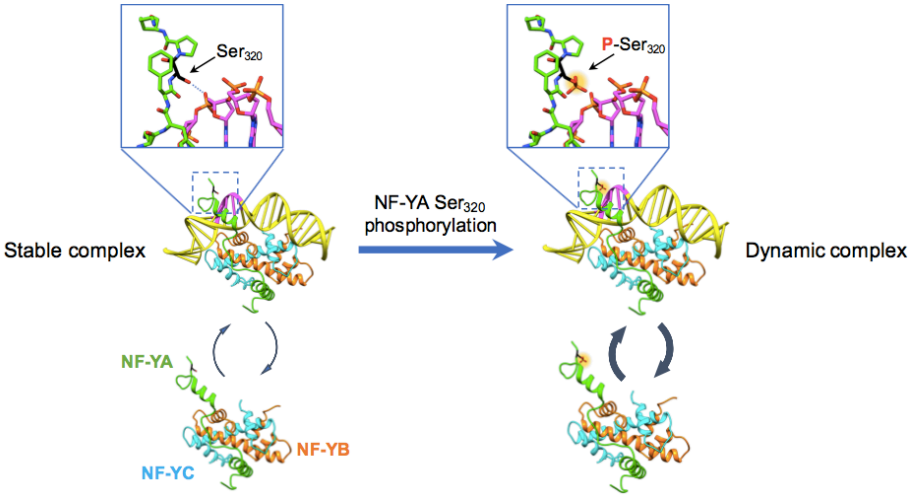


Figure 20. Schematic graphical abstract of the model by which NF-YA phospho-Ser320 might influence the DNA-binding dynamics of the NF-Y heterotrimer. Cycling arrows represent the ON/OFF dynamics on DNA; their thickness is proportional to the exchange kinetic rates.

In our model, NF-YA Ser320 phosphorylation would reduce the residence time of the protein on DNA (**Fig 20**). This could either be simply caused by the concerted ablation of a stabilizing interaction and application of a local repulsive force given by the bulky and negatively charged phosphate, or by a more complex reorganization of the C-terminal tail of the protein upon phosphorylation. This region of the protein is predicted to be intrinsically unstructured, as also evidenced by the lack of detectable electron density from NF-YA C-terminus in the crystallographic model of the NF-Y complex. Moreover, during my first PhD year I purified and isolated a novel NF-Y heterotrimer bound to DNA which harboured the YA3-wt construct (aa 239-347) employed in this study, which includes the complete C-terminus of the protein. We crystallized the complex and subjected it to X-ray diffraction in collaboration with Prof. M. Nardini. Although some refinement statistics we obtained were not fitting the required standards for publication, we could assess that for the new model no interpretable electron density could be retrieved after Ser326. These observations confirm the disordered nature of NF-YA C-terminal tail. Therefore, we cannot exclude that this region could undergo a phosphorylation-directed folding, as demonstrated for other proteins [249].

The effects of our NF-YA phospho-mutants in reporter assays with NF-Y target promoters can be interpreted in the light of the *in vitro* DNA-binding experiments. The reduced DNA-binding stability of the NF-YA Ser320 phospho-mimicking mutant complex would underlie higher on/off (association/dissociation) rates on target promoter. The higher DNA on/off rates for TFs have already been associated with reduced mRNA synthesis from the target gene in literature [135][250]. According to the probabilistic nature of transcription initiation (paragraph 1.4.3), TFs residence time on DNA would affect the probability to build a productive transcriptional event. In other terms, the less time a TF spends bound on DNA, the lower is the probability it would be able to stochastically recruit other factors on target promoter. The importance of TFs mobility on chromatin for gene activity is exemplified by the recent example of Sox2 in mouse 4-cell stage embryos [251]: cell-specific differences in terms of Sox2 long- or short-lived binding to DNA can predict the tendency of each cell to contribute to the inner-cell mass in the developing embryo. Currently, only few reports exist in which variations in site exchange rates are directly linked to a PTM of the TF. Namely, the phosphorylation of Serum Response Factor (SRF) as described in a paper in 1992 [252], and the acetylation of p53, as shown in a paper this year using live-cell single-molecule microscopy [253]. Still, the mechanism by which these modifications altered protein binding kinetics are obscure.

What would be the effects of increasing NF-Y mobility on DNA? At a single promoter level, this would promote the disruption of NF-Y-dependent promoter architecture by reducing the probability of productive interactions with nearby TFs. Due to the bivalent nature of NF-Y, the outcome in terms of transcriptional output would be promoter- and context-dependent. The local association of NF-Y with a specific kinase partner could trigger its release from DNA via Ser320 phosphorylation. However, no straightforward explanation for promoter-specificity can be found, since it might involve local chromatin state, other PTMs or the interaction with other specific TFs. At a global scale, reversible eviction of NF-Y from its bound state could promote availability of the free complex to bind elsewhere in the genome or recycling of the NF-Y-associated factors. This could be particularly important upon certain cellular stimuli to speed up the transcriptional reorganization of the genome. In the same perspective, the NF-YA phosphorylated form could effectively lower the target site search-time throughout the genome by increasing the association rates on DNA: although our data do not provide any direct evidence for this scenario, we can speculate that, given the higher off-rate (k_{off}) for Ser320 phospho-mimicking mutant, the respective on-rate (k_{on}) is also expected to be proportionally higher in order to maintain unaltered the affinity for DNA ($K_D=k_{off}/k_{on}$). Since the on-rates for NF-Y are too fast to be measured by EMSA, other methods should be employed to ascertain this hypothesis. For instance, *in vivo* fluorescence imaging techniques (e.g. FRAP or single molecule tracking) have never been

applied to NF-Y and would indeed give interesting perspectives on its behaviour in the nucleus. An appealing hypothesis stems from an observation coming from previous cell-cycle studies [238] and confirmed in our experimental setup, the fact that NF-YA phosphorylation is regulated throughout the cell-cycle, peaking at G2/M. Therefore, the potential function of Ser320 phosphorylation could be the eviction of a sizable fraction of NF-Y occupied sites to promote metaphase chromatin condensation during mitosis. Indeed, most sequence-specific TFs are displaced from DNA during mitotic chromosome formation, and NF-Y might follow this principle too [254]. Interestingly, NF-Y has been reported to be a gene-specific mitotic 'bookmarking' factor in the distal enhancer of the *DRA* MHC class II gene by recruiting on chromatin the PP2A phosphatase [255]. In this scenario the local association with a phosphatase might specifically prevent NF-Y displacement during mitosis.

Concerning Ser326, it seems to diverge in terms of position, function and conservation compared to Ser320. In our experiments we did not detect NF-YA Ser326 phospho-isotype; in accordance to this observation, both Ser326 phosphorylation-deficient and phospho-mimicking mutants were undistinguishable from the wild-type protein in NF-Y dependent reporter gene activation. This might also imply that phospho-Ser326 function requires the presence of a second, unknown effector element, such as a specific protein interactor. Also, it is not a constitutive phosphorylation site, at least in the cellular system we used, suggesting the requirement for a specific cellular signal. Indeed, among the published phospho-proteomic studies that retrieved Ser326 phospho-peptides (currently 27 distinct studies, including human, mouse and rat model systems), many of them provide specific treatments with inhibitors, hormones, cellular stress or DNA damage. Thereby a thorough analysis of the existing datasets would provide the basement to investigate the role of this phosphorylation site.

As stated in the discussion of the attached manuscript, multiple sequence alignment revealed an astonishing conservation of NF-YA C-terminus among amniotes (reptiles, mammals and birds), including Ser326. Interestingly, some amphibians, many fishes, non-vertebrate chordates and echinoderms display an acidic residue (Glu or Asp) at that position, possibly suggesting an evolution from a constitutive acidic side chain to an acquired conditional acidic phospho-serine. Instead, Ser320 origin is definitely at the basement of metazoa, with few acquired substitutions in some invertebrate species, not surprisingly never presenting an acidic residue at this position. In radial symmetry animals, fungi and some protozoa a threonine (Thr) occupies the very same position, which may well represent the ancestor of the modern serine in metazoa lineage.

4.2 USF1 cooperativity with NF-Y is mediated by the transactivation USR domain

The second topic of this thesis is directly linked to the recent findings of NF-Y directed regulatory modules throughout the genome (discussed in section 1.5.5). Among them, the genomic co-association with USF1/2 is pervasive and occurs with a strikingly defined and stereospecific distance between the E-box and CCAAT-box.

Our laboratory aimed to characterize at the molecular level the potential direct interplay between NF-Y, USF1 and other two E-box TFs, MYC and MAX [246]. While the latter two TFs did not show any enhanced DNA-binding with NF-Y *in vitro*, USF1 cooperatively interacts with NF-Y, forming a highly stable ternary complex on DNA with E-box and CCAAT-box spacing of 10-12 bp. Indeed, formation of the ternary complex was essentially DNA-directed, since the insertion of additional 5 bp between the two elements completely abolished cooperativity.

A second key-point of this study was the determination of the low-resolution structures of ternary complexes between the core NF-Y heterotrimer and USF-*short* on two distinct DNA configurations (LTR and HOXB4). Notably, the direct interaction between the two proteins involved an extended, yet not defined portion of USF1 dimer N-termini, that takes hold on the nearby NF-Y core.

I contributed to define which portions of USF1 N-terminus were mediating this DNA-directed protein-protein interaction responsible for enhanced DNA binding. We found that the intrinsically disordered USF Specific Region (USR) is necessary and sufficient for full cooperativity with NF-Y. This region was already described as a 'context-specific' transactivation domain (TAD) [214][215]. Its integrity was mandatory for USF-dependent transcriptional activation, since its presence was necessary also for the function of exon-5 activation domain. Differently from other activation domains, the USR did not exert its function when fused to the GAL4 DBD, suggesting the need for native surrounding protein domains. Moreover, a correlation with the underlying promoter architecture was described, showing that USR function is dependent on the presence of Inr core-promoter element. It was then shown that USF1 cooperatively interacts with a Inr-binding protein named TFII-I [256][257].

We could extend this 'context-specific' feature of the USR also to its structure. Indeed this region is predicted to be intrinsically disordered as shown in **Fig 21**, similarly to the majority of effector domains in TFs. It is not surprising that TFs are one of the most enriched classes of proteins possessing intrinsically disordered (ID) regions [258], since the latter are characterized by the peculiar combination of high specificity and low affinity for their interactive partners. Both these properties are exploited by TFs to mediate transient protein-protein interactions with multiple binding partners while maintaining high specificity. This allows TFs to behave as multifunctional molecules while limiting non-functional binding events or deleterious aggregation.

As a matter of fact, ID regions can undergo a disorder-to-order transition upon binding to their specific cognate interactor. This process has been shown for several proteins, including TFs such as CREB, HIFα or p53 in complex with different co-activators [259][260][261]. These crystallographic studies employed short peptides derived from the ID activation domains of the respective TF, since the crystallization step poses criticality when dealing with transient interactions involving large disordered proteins. On the contrary, the SAXS structural data on USF1 and NF-Y complexed on DNA provide the first look at a transcriptional activation domain - the USR - involved in a specific protein-protein interaction, at the cost of atomic resolution. With regard to this fascinating point, future aims will be the isolation of ternary complexes harbouring full-length subunits both for USF1 and NF-Y, suited to be subjected to single-particle Cryo-EM. This promising approach could lead to the visualization of complexes at an unprecedented high resolution, possibly defining details on the specific protein interfaces involved in the interaction mediated by the USR.

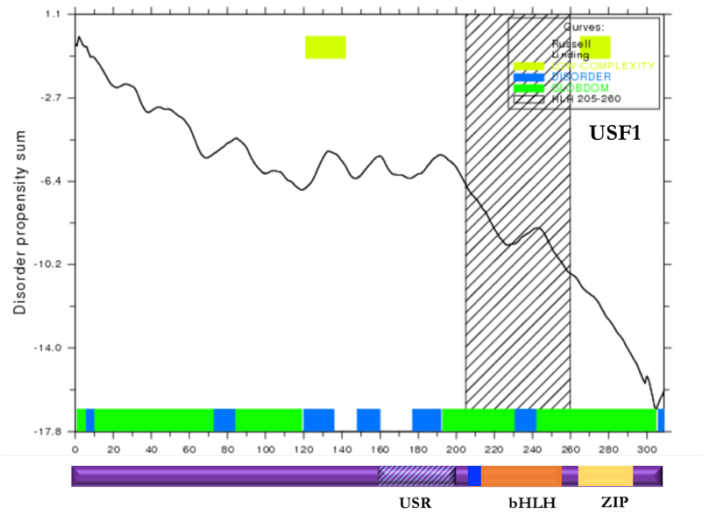


Figure 21. Disorder propensity plot for USF1 generated by GlobPlot v2.3 [262]. Under the plot the organization in domains of USF1 is represented schematically. The downhill regions (in green) represent putative globular domains, while the uphill regions (in blue) are predicted to be disordered. The yellow box highlights low-complexity regions. The well-structured HLH portion of the protein is marked by the dashed pattern.

A second, important observation is that our EMSA-based experimental setup is in accordance to the structural model, pointing at a differential conformation adopted by the USR as a function of the underlying DNA configuration of the binding sites. The relative ~70° rotation of the two proteins along the DNA axis, induced by the two bp difference between the 10 and 12 bp spacing, has indeed a major influence on the USR portions mediating the interaction with NF-Y. A key aspect of the USF1/NF-Y interaction is imposed by the intrinsic asymmetry of the NF-Y heterotrimer. Consequently, each USF1 monomer will interact with distinct portions of the NF-

Y core domains, potentially with markedly different binding affinities dependent on the specific surfaces involved. Therefore, one of the two USF1 monomers might provide the bulk of binding free energy, while the contribution given by the other monomer might be secondary. Moreover, the conformational change USR domains undergo on the 12 bp configuration might be optimal to provide the best register for the binding with NF-Y. The combination of asymmetric interactions and distinct USR portions involved might be at the basis of the difference in DNA-binding cooperativity observed on the 10 bp- *versus* 12 bp-spacing probes.

A fascinating question that stems from these *in vitro* data would be whether the different binding modes adopted by USF1 on the two distinct DNA configurations might impart distinct functional properties to the complex. Concerning this point, our attempts to evaluate transcriptional interactions between the two TFs on natural promoters did not highlight marked differences between the 10 and 12 bp spacing configurations. HOXB7 promoter (12 bp) showed a sharp transcriptional synergy between NF-Y and USF1, since its activation was completely dependent on the co-transfection of both TFs. The other promoters promptly responded to the overexpression of either NF-Y or USF1, with varying extents, although in all cases the co-transfection of both TFs substantially increased promoter output. For HOXB4 promoter it was shown that the precise spacing between E-box and CCAAT-box was crucial for proper gene activity and that NF-Y and USF1/2 positively synergise, similarly to our results [199]. We could extend this observation to other three important tissue-specific and developmentally regulated promoters, namely HOXC4, HOXC9 and HOXB7.

Clearly, natural promoters are complex objects, and their behaviour is influenced by a multitude of distinct REs and corresponding TFs. To potentially obtain a statistically significant view on the functional interplay between NF-Y and USF1 a wider (genomic) approach would be beneficial. In this regard, we are focusing our intentions to functionally correlate the USF1 genome-wide occupancy to NF-Y sites. As already described in [246] NF-YB silencing induced local DNA-binding rearrangements for the tested E-box TFs. I am currently working on the experimental setup to efficiently silence NF-YA or NF-YB transcripts on different cell-lines, including the establishment of an inducible short-hairpin RNA (shRNA) system for large-scale silencing. This would allow to perform ChIP-seq experiments to extend our observations to a genome-wide scale.

PART II

SUBMITTED MANUSCRIPT

Manuscript submitted to ***Oncotarget*** as Research Paper on the 6th September 2017.

The phosphorylatable Ser320 of NF-YA is involved in DNA-binding of the NF-Y trimer

Andrea Bernardini, Mariangela Lorenzo, Valentina Nardone, Marco Nardini, Roberto Mantovani and Nerina Gnesutta*

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

* To whom correspondence should be addressed:
Nerina Gnesutta, PhD
Dipartimento di Bioscienze
Università degli Studi di Milano
Via Celoria 26, 20133 Milano, Italy
nerina.gnesutta@unimi.it
Tel + 39 02 5031 4917
Fax + 39 02 5031 4925

Total No. of Figures and Tables: 6 main text figures; 3 supplemental data figures.

ABSTRACT

NF-Y is a trimeric transcription factor binding to the functionally important CCAAT-box, present in promoters of growth-promoting and cell cycle-regulated genes. The regulatory NF-YA subunit confers sequence-specificity to the NF-YB/NF-YC dimer. It harbors two Serines -Ser320 and Ser326- shown to be phosphorylated by CDK2. High-throughput proteomics data indicate that they are phosphorylated *in vivo*. Specifically, Ser320 makes structural contacts with the DNA phosphate backbone; Ser320-P is the major NF-YA phosphorylation isoform following overexpression in cells, increasing upon mitotic arrest. EMSAs with recombinant Ala and Glu mutants confirm a role of Ser320, but not Ser326, in stabilization of DNA-binding. Transactivation assays of the CCAAT-dependent MDR1 and RHOB promoters show loss in transcription function for Ser320Glu and Ser320Ala NF-YA mutants. Phylogenetic analysis of NF-YA proteins indicates that Ser320 is indeed evolutionarily conserved. We conclude that this residue belongs to the core mechanisms of DNA-binding control, possibly driven by the necessity to unfasten binding of, or to evict NF-Y from CCAAT sites under specific conditions of growth regulation.

KEYWORDS

Transcription Factor; DNA-protein interactions; Histone-Fold-Domain; Post-Translational-Modification; Phosphorylation

INTRODUCTION

The transcriptome of each living cell is ultimately determined by sequence-specific binding of transcription factors (TFs) to DNA elements within regulatory regions of genomes. Neoplastic transformation entails profound changes in gene expression patterns, and it has long been known that driving events are often started by oncogenic TFs, in their "altered" version, either because of overexpression, mutation or fusion with unrelated proteins. NF-Y is a TF that binds to the CCAAT element [1]. It is a heterotrimer composed of NF-YA, NF-YB and NF-YC subunits [2]. NF-YB/NF-YC have a Histone Fold Domain (HFD) related to core histones H2A/H2B [3]. NF-YA has an α-helical domain consisting of helices A1 and A2 involved in NF-YB/NF-YC association and sequence-specific DNA binding, respectively. DNA-binding of the trimer is exerted through contacting backbone phosphates by the HFD subunits and CCAAT nucleotides by NF-YA residues. Mechanistically, different types of results suggest that NF-Y plays a "pioneer" role in transcriptional activation [4-7]. Apparently, its subunits are not vastly overexpressed in cancer cells, nor are they frequently mutated or fused to other proteins, as it is often the case for oncogenic TFs. Yet, a recently growing body of evidence links NF-Y to cellular transformation [8]. At least four lines of experiments converge on this idea.

First, the target CCAAT box is enriched in "cancer" genes. The analysis of RNA profiling in different types of tumor cells *versus* their normal tissue counterparts identified signature genes whose expression is increased in cancer samples: promoters of these genes are analyzed and many times the CCAAT box is found statistically enriched, often with E2F sites, including by *de novo* motif discovery in large Oncomine cancer datasets (Reviewed in [2]; see also [9-15]). In some cases, overexpression of NF-Y subunits has been found as prognostic of patients survival [16, 17]. Particularly explicative GO categories of genes enriched in the NF-Y regulome are *cell-cycle* and *metabolism* genes [4]. It has long been known that NF-Y targets the majority of cell-cycle regulated genes, and, in particular, nearly all G2/M genes. Such genes are notoriously common hallmarks of cancer progression [18]. As for metabolic genes, NF-Y activates specifically genes of the glycolytic pathway, fueling the anaerobic energy production typical of cancer cells, of lipid anabolism, of SOG -Serine, One carbon, Glycine- and glutamine metabolisms and nucleotide biosynthesis [19]. In all such pathways, NF-Y is crucial in activation of rate-limiting genes, whose expression is frequently altered in cancer cells, and often *bona fide* "driver" oncogenes.

Second, independent reports recently pointed to NF-YB and NF-YC are "driver" oncogenes for specific cancers: (i) NF-YB is among a handful TFs driving Diffuse Large B-cell Lymphoma (DLBCL), the aggressive evolution of follicular lymphomas [20]; (ii) Mouse/human genomic screenings of Choroid Plexus Carcinomas identified the synthenic NF-YC, TAF12 and

RAD54L genes, with the formers having the greatest impact as drivers of tumor development [21].

Third, overexpression of NF-YA has been associated to increased proliferation and/or aggressiveness of cancer cells [22-24]. Finally, identification of genomic targets of oncoproteins showed a predominance of CCAAT in the peaks: PRAME, an E3-ligase originally identified as enriched in metastasis, the cell-cycle promoting FoxM1, PBX/Prep, SOX9, a powerful oncogene in colorectal cancer [25-29], and two members of the MAP kinase family, JNKs, a stress-responsive branch, and ERK2 [30, 31]. This indicates that NF-Y sites might be part of a landing platform for such oncogenes, a model originally suggested as a mechanism driving transcriptional activation of mutant p53 [8].

Compared to the knowledge available on the structure, genomic locations, TFs partnerships and functions of NF-Y, much less is known on the post-translational modifications -PTMs- its subunits undergo, and on their effects on the basic functions of the trimer. Specifically, NF-YA was shown to be phosphorylated [32,33]; two serines -Ser320 and Ser326- lying outside of the evolutionarily conserved domain (also termed HAP2 domain) are targeted by CDK2, serving an important function in cell-cycle progression [32,34]. The presence of a Ser-Pro motif in the NF-YA CDK2 targeted residues supports this. Overexpression of a Ser-Ala double mutant -YA-AA - behaved as dominant negative on cdc2 promoter activity, and blocked cell-cycle progression of HCT116 cells in G1 and G2. This result suggested that Ser320 and/or Ser326 are CDK2 targets of paramount importance for cell-cycle progression. However, the role of phosphorylation in NF-YA regulation is not well understood: lack of phosphorylation in the alanine mutants was proposed to entail decreased DNA-binding by the NF-Y trimer and inactivation of the cdc2 promoter [32], but recombinant, full length NF-YA proteins produced in *E. coli*, hence presumably not phosphorylated, and deletion mutants lacking these serines are proficient in DNA-binding [35-39]. Furthermore, from the 3D crystallographic analysis of the trimer, it was not obvious to discern why phosphorylation would impact positively on DNA-binding [40].

We thus decided to study the functional roles of Ser320 and Ser326 modifications by *in vitro* (EMSAs with recombinant proteins) and *in vivo* (transient transfections in mammalian cells) approaches.

RESULTS

Position of phosphorylated residues in the NF-Y/CCAAT quaternary structure.

It has been shown that NF-YA is phosphorylated by CDK2 *in vivo* [32]. Several datasets of phosphorylated mammalian proteins derived from high-throughput phosphoproteomics studies are available, organized by the PhosphoSitePlus (PSP) repository (http://www.phosphosite.org; [41]). We searched this resource for phosphorylations of NF-YA: three sites indeed emerged in mammals: Ser320-P (3 curated datasets [42, 43] and [44]), Ser326-P (26 datasets), and Tyr266-P (>100 datasets) (**Fig. 1A**). A fourth site -Ser302-P- is also annotated in PSP, but considering that the corresponding phospho-peptide was retrieved in a single tumor sample of the original curated dataset [44], we decided not to focus on this residue. Thus, the PSP repository datasets confirm that Ser320 and Ser326 are phosphorylated *in vivo*.

Note that two NF-YA splice variants exist in mammals [2], with 28-29 amino acid difference in length: Ser320 and 326 correspond to residues 291 and 297 of the short isoform (NF-YAs); in this paper, to conform with PSP annotation of the phosphorylated residues, we will use the long isoform numbering (**Fig. 1A**).

The crystal structure of the mammalian NF-Y/DNA complex (PDB-code: 4AWL) [40] shows that residues following the GxGGRF motif, at the C-terminus of NF-YA, have an extended conformation up to Lys322 (Lys293 in 4AWL and NF-YAs numbering) (**Fig. 1A** and **1B**). Consequently, Ser320 is visible in the electron density map, while Ser326 is not, being part of a disordered C-terminal region of the crystallized protein. Ser320 is positioned in close proximity to the DNA chain, with its side-chain hydroxyl group hydrogen bonded to the phosphate group of the first CCAAT adenine A_{10} (**Fig. 1B**). Therefore, the position of Ser320 within the NF-Y/DNA complex suggests that its phosphorylation could cause a direct impairment in the ability of NF-YA to bind the CCAAT box, because of the consequent charge repulsion at the protein-DNA interface. On the other hand, Tyr266 (Tyr237 in 4AWL) lies within the evolutionarily conserved HAP2 part (**Fig. 1A**), just ahead of helix A1, responsible for HFD interactions. However, this residue points outwards the NF-Y trimerization surface, and it is far from the DNA-contacting region (**Fig. 1B**); thus, its modification is not predicted to alter or affect directly either trimerization or DNA-binding. Indeed, a triple mutant (Pro-Leu-Tyr to Ala-Ala-Ala) was reported to associate the NF-YB/NF-YC HFD dimer and bind DNA normally [36], and a double mutant including the corresponding Tyr162 in yeast HAP2 -Tyr162Cys/Tyr168Asp -, had negligible effects on subunits interaction and DNA-binding [45].

Analysis of Ser320 and Ser326 phosphorylation *in vivo*.

To further verify phosphorylation of Ser320 and Ser326 *in vivo*, we overexpressed wt NF-YA in HeLa cells, together with a mutant in which both serine residues were mutated to alanine (NF-YA-AA), in the short isoform configuration: the proteins from nuclear extracts were separated by SDS-electrophoresis on a polyacrylamide gel added of Phos-tag™ -to detect phosphorylated proteins by mobility shift (see Methods section) - and overexpression monitored by Western Blot analysis. **Fig. 2A** shows that, while slower migrating bands are visible in both samples, in addition to the overexpressed NF-YAs, the most prominent retarded protein in the wt sample is absent in the NF-YA-AA mutant, indeed indicating that phosphorylation of NF-YA occurs on either (or both) Ser residue(s). To ascertain which of the two serines is predominantly phosphorylated, we generated single mutants by reverse mutagenesis of each alanine of the NF-YA-AA double mutant to (wt) serine (NF-YA-SA and NF-YA-AS): upon the same transfection protocol and Phos-tag Western blot analysis, mutant NF-YA-AS still showed an almost complete lack of the band corresponding to the specific phosphorylation of wt NF-YA, whereas in NF-YA-SA the band was almost as intense as the wild-type protein (**Fig. 2B**).

It was previously shown that phosphorylation of endogenous NF-YA increases over cell cycle progression, peaking at late-S, G2/M [32]. We therefore analyzed serine NF-YA phosphorylation in nocodazole treated HeLa cells to enrich for the mitotic population, after transfection with wt and mutant NF-YAs. NF-YAs was chosen because it is the predominant isoform expressed in HeLa cells [46]. **Fig. 2C** shows the results of such experiment. In the upper panel, the global levels of the overexpressed NF-YAs were monitored in standard SDS-PAGE analysis by Western blot, indeed showing that all proteins were produced at similar –high– levels under growing conditions in asynchronous cells. Note, however, that in nocodazole treated cells, the levels of all overexpressed proteins were markedly reduced. In the bottom panels, Western blots of Phos-tag gels illustrates that the phosphorylated form of NF-YAs –corresponding to Ser320 by analysis of the mutants- is increased in G2/M arrested cells: in relative terms, this form increases significantly its abundance compared to the non-phosphorylated form (from 3% to 35% of the total protein levels). Longer acquisition times allow the detection of the endogenous NF-YA isoforms: the increase in the phosphorylated form of NF-YA upon G2/M arrest is also evident, reaching 45-50% of the total protein levels. In summary, the major phosphorylated residue of NF-YA is Ser320, which is regulated during the cell-cycle, increasing in G2/M.

NF-YA Ser320 is involved in CCAAT-binding.

To ascertain the role of Ser320 and Ser326 in DNA-binding of NF-Y and verify whether phosphorylation of either one could influence it, we produced in *E. coli* and purified a recombinant,

N-terminally truncated, NF-YA construct (**Fig. S1A**) which includes the complete C-terminal portion of the protein (from Met239 to Ser347, termed YA3 here, see **Fig. 1**). In parallel, we produced the double alanine mutant described above, in the YA3 configuration, as well as mutants containing double (YA3-EE) or single (YA3-ES, YA3-SE) substitutions to glutamate of serines 320 and 326. Recombinant YA3 proteins were produced in *E. coli* and purified separately (**Fig. S1**). In an initial set of EMSAs performed at equilibrium, increasing amounts of wt and mutant YA3s were incubated with recombinant NF-YB/NF-YC HFD subunits (YB/YCmd) and with the labelled high affinity HSP70 oligo containing a canonical CCAAT: **Fig. 3A** shows that overall affinities of the mutant trimers are very similar to the (YA3) wt NF-Y. Regarding binding kinetics, on-rates experiments were not informative with such assays, since the equilibrium of DNA-binding complexes is reached at rates too fast to measure, even with immediate gel loading and electrophoresis of the binding reactions (data not shown). Nonetheless, off-rates can be measured by competition assays, in which pre-formed complexes are challenged with excess unlabelled oligos. We performed such experiments with the YA3-wt and mutants, and the resulting dissociation curves are shown in **Fig. 3B**: YA3-wt and YA3-SE complexes showed a slow off rate: with half-lives ($t_{1/2}$) of 17.4 and 13.5 min, respectively, and dissociation rates (k_{off}) around 0.04-0.05 min^{-1}; YA3-EE and YA3-ES complexes were clearly more unstable (k_{off}: 0.59 and 0.47 min^{-1}; $t_{1/2}$: 1.5 and 1.2 min, respectively). YA3-AA had an intermediate stability (k_{off} : 0.12 min^{-1}), forming DNA complexes with half-life of 5.8 min, longer than YA3-EE/YA3-ES, but shorter than YA3-wt. Overall, we conclude that Ser320 is involved in stabilizing the DNA-bound complex and that a phospho-mimicking modification to glutamate is capable to destabilize the quaternary NF-Y/CCAAT complex *in vitro*, increasing the mobility of the trimer on DNA.

Mutation of Ser320 leads to loss of NF-Y transcriptional activity *in vivo*.

NF-Y possesses two *bona fide* trans-activation domains, located in NF-YA and NF-YC, both rich in glutamine and hydrophobic residues [47,48]. NF-Y was shown to activate some, but not all promoters in transfection assays. We initially tested 10 CCAAT-dependent promoters in luciferase gene reporter assays, alone and upon cotransfection of the three subunits in human HeLa cells: some promoter units -TK, CCNB2, HSP40, HBP1- showed paltry activation of 3-fold or less, one -TGFβRII- repression of basal activity, whereas NF-YC2, CYP3A4, TopoIIα, MDR1 and RHOB (See below) were more robustly -4 to 12-fold- activated by the NF-Y trimer (**Fig. S2**). We decided to pursue analysis with the MDR1 and RHOB promoters, as indicative of this latter group. **Fig. 4A** shows cotransfection of NF-YB, NF-YC (37 kDa isoform, see [46] and NF-YAs (short isoform) in activation of MDR1. As previously mentioned, note that the short isoform of human NF-YA lacks 29 amino acids in the N-terminal trans-activation domain because of alternative

splicing of exon 3 [49] (see **Fig. 1**). For all subsequent experiments with mutant NF-Y, the HFD dimer was kept constant, and the different NF-YA Ser mutants described above were tested as full-length proteins in the short isoform configuration. As a control, we also used a Dominant Negative mutant -NF-YAm29-, which contains mutations in the A2 helix abolishing DNA-binding ([36], reviewed by [1]): this mutant was indeed very inefficient in NF-Y transactivation. Interestingly, all constructs containing a mutated Ser320 showed decreased activity: drastic for NF-YA-EE, NF-YA-ES, less so for NF-YA-AA and NF-YA-AS, whereas constructs where only the Ser326 was mutated, i.e. NF-YA-SE and NF-YA-SA, were as efficient as wt NF-YA. Control of NF-Y subunits levels produced in HeLa cells was carried out by Western blot, and showed similar expressions (**Fig. S3**). These experiments were replicated on the RHOB-Luc promoter which led to very similar results, bar differences in the extent of activation by NF-YA-AS: NF-YA-AA, -AS and -ES were lower than, and statistically distinguishable from, wt NF-YA, while NF-YA-SE and -SA did not differ significantly from the wild type (**Fig. 4B**). In conclusion, mutations affecting Ser320, especially introducing a phospho-mimetic glutamate, but not Ser326, impair trans-activation of the NF-Y trimer.

To verify the effective DNA-binding proficiency of the different NF-YAs mutants produced *in vivo*, we transfected HeLa cells with the same wt and mutant vectors, prepared nuclear extracts and assayed them in EMSAs. By addition of NF-YB/NF-YC HFD dimer (short constructs used in EMSAs), the complexes formed with overexpressed NF-YAs could be discriminated from the endogenous NF-Y trimer present in extracts, by virtue of the faster electrophoretic mobility. The results are shown in **Fig. 5A**: a faster migrating trimer complex was clearly visible, and highly increased, compared to non-transfected samples, when the HFD dimer was incubated with nuclear extracts containing the overexpressed NF-YA, whether wt or mutant. Again, off-rates experiments performed with NF-YA-ES or NF-YA-AA mutant containing extracts showed a less stable behaviour of these complexes compared to wt NF-YA (**Fig. 5B**). Western blot analysis indicated production of similar levels of the transfected NF-YA proteins (**Fig. 5C**). These data confirm for *in vivo* produced proteins the decreased DNA-binding properties of the Ser320 mutants observed with recombinant proteins, and are in accordance with the functional results.

DISCUSSION

Compared to other structural, functional and genomic features, our knowledge of post-translational modifications of the NF-Y trimer is lagging behind. We report here novel findings concerning a role of NF-YA Ser320 impacting on the DNA-binding function of the NF-Y trimer.

Effect of phosphorylation on NF-Y activity.

Phosphorylation is a common PTM in proteins, and TFs are no exception to this rule: whenever searched, this modification was found on most, if not all TFs. Consequent with the specific location(s) on the protein, the effects of phosphorylation are the most diverse, ranging from regulation of stability, to nuclear targeting, chromatin binding, protein-protein interactions, DNA target recognition and affinity, and tuning of transactivation [50]. Our study stems from the previous identification of a DNA-binding role of phosphorylation of Ser320 and Ser326, pinpointing CDK2 as a potential kinase [32, 34]. The observed cell-cycle blocking phenotype of the Ser to Ala double mutant was interpreted as a lack of DNA-binding to cell-cycle promoters such as cdc2, because of decreased phosphorylation [34]. These reports did not address the question whether the two serines would fulfill an identical role, and the interpretation of the results were at odds with biochemical and further structural data.

Unbiased high throughput mass spectrometry datasets confirm that NF-YA is indeed a phosphoprotein (PSP, http://www.phosphosite.org); although quantitative approaches did not report on differential modifications to occur on this protein (e.g. in asynchronous vs. mitotic cells), possibly arguing on the modulatory role of these events, we notice that the specific NF-YA phosphopeptide (which includes Ser320 and Ser326) is differentially enriched by treatments that perturb the cell cycle [42]. Our new data are in line with the direct evidences provided by independent groups [32-34, 51].

We find here that Ser320 and Ser326, and their respective modifications, serve different roles and that only the first is involved in direct control of DNA-binding. In particular, we show that the Ser to Ala mutation could be intrinsically deleterious for stable DNA-binding of the trimer, as observed *in vitro* by the higher k_{off} of the double alanine mutant. We therefore propose an alternative explanation for the phenotype of overexpressing the double Ser320/Ser326 Ala (AA) NF-YA: a mutant with diminished DNA-binding potential and transcriptional activity, which would be competing with the endogenous NF-YA, depleting endogenous NF-Y trimers of stable CCAAT recognition. The lack of stabilization would be provided by the absence of the Ser320 hydroxyl group and the potential repulsion of the Ala methyl side chain in a region involved in critical NF-YA/DNA interactions. Indeed, the negative charges provided by a phosphorylation event on

Ser320, mimicked *in vitro* by our YA3-ES mutant and full-length NF-YA-ES expressed in cells, destabilize the DNA bound complex, by increasing the off-rate (hence mobility) of the trimer, and affecting its transcriptional activity in vivo. This is clearly not the case for the YA3-SE mutant, reinforcing the idea of a fundamentally different role of the two serines.

What would be the scope of inducing a decrease in the stability of CCAAT-binding by NF-Y? On one side, reducing residence-time might directly affect the transcriptional output. Several studies indicate that NF-Y serves a cooperative role in promoter access and specific DNA-binding for a number of activating TFs. In addition, some TFs apparently associate to a platform represented by –at least– two well-spaced CCAAT boxes bound by NF-Y trimers [52] (and References therein), even in the absence of direct sequence-recognition, as elegantly shown for Sp2 [53]. In general, a weakening of NF-Y binding might be instrumental to displace multi-TFs complexes which rely on robust CCAAT-binding, at large: this could happen, for example, in cell-cycle promoters in phases in which the specific gene is not supposed to be active (G1/S genes in G2/M and G2/M genes in G1), or in other transiently inducible systems, once the stimulus is eliminated and transcription returns to low, "basal" levels. On the other, considering the unaltered affinity of our phospho-mimetic NF-YA mutants, another possible consequence is that phosphorylation of NF-YA Ser320 (or both residues of the C-term tail) might impact on the general 3D architecture of such multi-CCAAT promoters, due to different DNA bending capacity of the modified protein, with consequent effects on NF-Y spatial relationships with its partners.

Serine 320 is evolutionarily conserved

As two fundamental features of NF-YA –HFD interaction and CCAAT-binding– are evolutionarily extremely conserved, we wondered whether Ser320 would equally be conserved in NF-YA from different species. In addition to published sequences, a set of NF-YA genes were identified in representative species –from *fungi* to mammals- and encoded proteins aligned in **Fig. 6**. Note that plants were excluded from this analysis, as the vast expansion of NF-YA and of the related CCT genes warrants a separate analysis. It is apparent that Ser320 is present in all bilateral animals, and absent in *fungi* and protozoa. However, these organisms do display a Thr residue at the same position, which is also phosphorylatable. As for the succeeding Pro residue, it is conserved in mammals, birds, fishes and amphibians, but not in flatworms. On the other hand, Ser326 is strictly limited to vertebrates, and more specifically to mammals, birds and certain amphibians, although it is true that many, but not all, species show Ser residues within the very C-terminal region, at various distances from Ser320 position. It is also interesting to note the presence of acidic amino acids at residue 326 in more ancient orthologs (sea urchin and invertebrates, also conserved in fishes; see **Fig. 6**). This suggests a critical role of a negatively charged residue at this position,

which emerged later as a phosphorylatable Ser, and possibly turning it into a functional switch [54]. In summary, our phylogenetic analysis concords with the idea that a potentially modifiable residue (Ser/Thr) is present at the very end of the DNA-binding HAP2 subdomain, as part of the core structural features of NF-YA.

As to which kinase(s) impact on Ser320 (and Ser326), CDK2 was first shown to be a candidate [32], and the presence of Ser-Pro motifs at both phosphorylated sites point to Pro-directed serine/threonine kinases at large, such as other cyclin-dependent protein kinases (CDKs), and members of the MAPK families, and/or GSK-3 signaling kinases, as potential players in NF-YA regulation [55]. CDKs in general, and CDK2 in particular, are known to target many TFs [50], many of which are notorious NF-Y genomic partners in cell cycle control, such as E2F1 (G1/S progression), MYC (senescence, repression), B-MYB and FOXM1 (G2/M progression) [52]. JNKs, a subfamily within MAPKs, were also associated to NF-Y in ChIP-Seq experiments in mES cells, before and after differentiation; this result was corroborated by the lack of (activating) JNK kinase on promoters when cells were transfected with the NF-YAm29 mutant, which removes the endogenous trimer from DNA [30]. A NF-Y-JNK connection was also reported in Drosophila with genetic experiments [56]. In addition, p38 was involved in NF-Y activity, both by activating RHOB [57], and by repressing expression of Annexin A [58]. Thus, different members of this class could be targeting Ser320 and/or Ser326. The data shown here of a substantial increase in Ser320 phosphorylation in G2/M warrant future studies concerning the specific role of this modification on the expression of genes specifically regulated in this phase, which are, for the vast majority, NF-Y-dependent.

In conclusion, we detail independent functional roles of phosphorylated Ser-Pro motifs in the C-terminal portion of the DNA-recognition subunit NF-YA. Biochemical and functional evidences underline a direct role in the control of stability of NF-Y/DNA complexes for Ser320, while phylogenetic considerations assign to the conditional negative charge of residue Ser326 other possible roles, such as the control of interactions with potential protein partners.

MATERIALS AND METHODS

Expression vectors and reporter constructs

Wild-type and mutant cDNA sequences encoding aa 239-347 of human NF-YA (long isoform numbering) (YA3-wt), or bearing substitutions of Ser320 and Ser326 to Ala (YA3-AA), Glu (YA3-EE), or single residue substitutions to Glu (YA3-ES or YA3-SE, respectively) were obtained by gene-synthesis (Eurofins) with the addition of 5' NdeI and 3' BglII restriction enzyme sites and subcloned through restriction ends ligation into pmcnEATCH vector [59] at NdeI-BamHI sites, allowing *E. coli* expression of fusion proteins with a C-terminal 6His-tag preceded by a thrombin cleavage site. NF-YC expression construct was obtained by subcloning the cDNA encoding the HFD (aa 27-120) [40] in pmcnYC [59]; NF-YB HFD cDNA (encoding aa 49-141) [40] was subcloned into pET15b to generate an N-terminal 6His-tag fusion construct.

To obtain mammalian expression vectors of full-length mutant NF-YA (YA-AA, YA-EE, YA-SE, YA-ES), the KpnI-BamHI fragment of wild-type NF-YA was replaced with the corresponding fragment of each mutant sequence (harboring Ser320 and Ser326 substitutions) by subcloning into pSG5-NF-YAs full-length construct (short isoform) [49, 46]. To generate the NF-YA-SA and NF-YA-AS mutant constructs, pSG5-NF-YA-AA was used as a template to perform site-directed reverse mutagenesis (Quick Change™ Site-Directed Mutagenesis Kit, Stratagene). Presence of desired mutations in resulting plasmids was checked by sequencing. Mammalian expression vectors of Flag-tagged NF-YC and NF-YB were described in [40, 46]. Gene promoter-driven Luciferase expression plasmids used in reporter assays in transient cotransfections with NF-Y MDR1-Luc and RHOB-Luc were described before [60, 61] (see also Supplemental Information).

Protein Expression and Purification

The YA3 wild-type and mutant proteins were produced in *E. coli* BL21(DE3). Cells were grown in LB medium (additioned with 100 μg/mL ampicillin) at 37 °C to 0.3 OD_{600} units. The culture was induced with 0.4 mM IPTG and further grown for 2 hours at 30 °C. Cells were collected, resuspended in buffer A (10 mM Tris·Cl pH 8.0, 400 mM NaCl, 2 mM $MgCl_2$, 5 mM imidazole) and lysed by sonication. The soluble YA3-6His proteins were purified by nickel ion metal affinity chromatography (IMAC) (His-Select® Nickel Affinity Gel, Sigma-Aldrich) and eluted in buffer A containing 100 mM imidazole. Protein samples used in EMSA were then subjected to a dialysis step in buffer B (10 mM Tris·Cl pH 8.0, 400 mM NaCl, 2 mM DTT) containing 10% glycerol. 6His-NF-YB/NF-YC HFD soluble dimers (YB/YCmd) were produced by co-expression in *E. coli* BL21(DE3) and purified by IMAC in buffer A as described [62], followed by dialysis in buffer B with 10% glycerol.

Electrophoretic Mobility Shift Assays (EMSA)

For EMSAs, recombinant proteins were added at the indicated molar concentrations to a Binding Mix in 16 µL reaction volume, containing a 31 bp Cy5-labelled DNA probe derived from human HSP70 promoter (5'-[Cy5]-CTTCTGAGCCAATCACCGAGCTCGATGAGGC-3'). The final reaction composition is the following: 20 nM HSP70 probe, 20 mM Tris·Cl pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$, 0.5 mM EDTA, 6.5% glycerol, 2.5 mM DTT, 0.1 mg/mL BSA. The reactions were assembled on ice and then incubated at 30 °C for 30 minutes in the dark. An aliquot of each reaction was loaded on a 6% polyacrylamide non-denaturing gel and run in 0.25X TBE at 80 V at 4 °C. In EMSAs performed with full-length NF-YA constructs expressed in HeLa cells, 4 µg of nuclear extracts [46] in buffer C (20 mM HEPES pH 7.9, 700 mM NaCl, 0.1 mM EDTA, 0.1 mM EGTA, 1 mM DTT, protease inhibitor cocktail, 1 mM PMSF) were added to each reaction and 200 ng of poly-dIdC were included in the Binding Mix. In off-rate assays, after the 30 minutes incubation time, an excess (50X) of unlabelled 25 bp HSP70 CCAAT DNA competitor (5'-TTCTGAGCCAATCACCGAGCTCGAT-3') was added to each binding reaction. Aliquots of the reaction were loaded for electrophoresis at different time points after the addition of the competitor. The reactions performed with nuclear extracts or recombinant proteins were kept at 4 °C and 30 °C, respectively, for all the duration of the assay. Fluorescence signals were detected with Chemidoc™ MP apparatus (Bio-Rad). Band intensity quantification was performed with ImageLab™ software (Bio-Rad) as band percentage within each lane. The average of three independent off-rate experiments was used to fit the experimental data in a simple one-phase exponential decay model, setting the 'no competitor' point as 100% initial binding in the following equation:

$$Y = (Y_0 - plateau)\, e^{-kt} + plateau$$

Where Y is the normalized binding percentage at any time point t by the addition of the competitor DNA; Y_0 is the initial Y value in absence of competitor DNA; k is the off-rate constant; *plateau* is the Y value at infinite times (off-set of the dissociation reaction). The non-linear regression analysis was performed using Prism (v5.0).

Mammalian Cell Culture Transfection and Luciferase Assays

HeLa cells were cultured in DMEM High Glucose with L-glutamine (EuroClone) supplemented with 10% FBS and antibiotics (100 U/ml penicillin, 100 µg/ml streptomycin). For phosphorylation analyses, the day before transfection $9x10^5$ cells were plated in 6 cm plates in medium without antibiotics. Cells were transfected with 3 µg total plasmid DNA (0.5 µg of NF-YA expression vector) using 9 µL of transfection reagent (Lipofectamine® 2000, Invitrogen). 24 hours post-transfection, cells were harvested and nuclear extracts were prepared as described above

in presence of phosphatase inhibitors (PhosSTOP®, Roche). For the enrichment of the mitotic population, 6 hours post-transfection cells were treated with 0.1 μg/mL nocodazole, and harvested after additional 16 hours incubation. Total protein extracts were prepared in RIPA buffer additioned with phosphatase inhibitors.

For luciferase assays the transfection protocol described above was modified as follows: $2x10^5$ cells were seeded in 12-wells plates and transfected with 0.6 μg total plasmid DNA (p-Luc reporter 50 ng, and 100 ng of each NF-Y subunit plasmids, or empty vector, to a total of 0.6 μg) using 1.8 μL of transfection reagent. 40 hours post-transfection, cells were harvested for luciferase assay in PLB buffer, according to the manufacturer instructions (Promega). For experiments of promoter-Luc testing of **Fig. S2**, a minimum of two biological replicates in duplicate were performed. For the experiments of **Fig. 4**, results of three independent experiments with the complete sets of NF-YA proteins, each in duplicate, were performed. Luminescence was measured with a Lumat LB9501 luminometer using 10 μL of cell extracts in 50 μL of Luciferase assay buffer (25 mM glycyl-glycine pH 7.8, 7.5 mM potassium phosphate pH 7.8, 7.5 mM $MgSO_4$, 2 mM EGTA, 1 mM ATP, 0.25 mM D-luciferin, 5 mM DTT) and 12 sec integration time; raw data counts of each sample were normalized to protein concentration determined by Bradford protein assay (Bio-Rad).

Phos-tag SDS-PAGE and Immunoblot

For mobility shift detection of phosphorylated proteins, the 9% acrylamide SDS-PAGE running gel solution was additioned with 30 μM Phos-tag™ acrylamide (AAL-107M, Wako) and 60 μM $MnCl_2$. Protein extracts used in Phos-tag applications were prepared in extraction buffers devoid of EDTA. The run was performed in Tris-Glycine buffer. After electrophoresis, the gel was washed twice with transfer buffer additioned with 10 mM EDTA for 30 minutes to remove Mn^{2+} ions and once in regular transfer buffer for 15 minutes. Proteins were transferred to nitrocellulose membrane for immunoblotting with the following antibodies: rabbit anti-NF-YA (H-209) and goat anti-Lamin B (M-20) (Santa Cruz Biotechnology); rabbit anti-NF-YB (GeneSpin); mouse anti-FLAG (M2) and anti-vinculin (V284) (Sigma-Aldrich) and HRP-conjugated secondary antibodies. Immunoblots were developed with the ECL protein detection system (GeneSpin) and acquired with a Chemidoc™ MP Imaging system with the ImageLab Software (Bio-Rad).

ABBREVIATIONS

TF: Transcription Factor; NF-Y: Nuclear Factor Y; CDK: Cyclin dependent kinase; EMSA Electromobility-shift assay; HFD Histone-fold domain; PTM: Post-translational modification; PSP: PhosphoSitePlus; PDB: Protein Data Bank

AUTHOR CONTRIBUTIONS

AB and ML performed the experiments; MN and VN perfomed the analysis of the X-ray structure; NG, AB and RM conceived the experiments; RM and NG supervised the work, analyzed the data and wrote the manuscript.

ACKNOWLEDGMENTS

We thank D. Kardassis for gift of the RHOB promoter construct.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest with the content of this article.

FUNDING

This work was supported by the Italian Association for Cancer Research (AIRC) Investigator Grants IG 2014 to MN [grant number IG-15267] and IG 2013 to RM [grant number IG-14130].

MAIN FIGURES

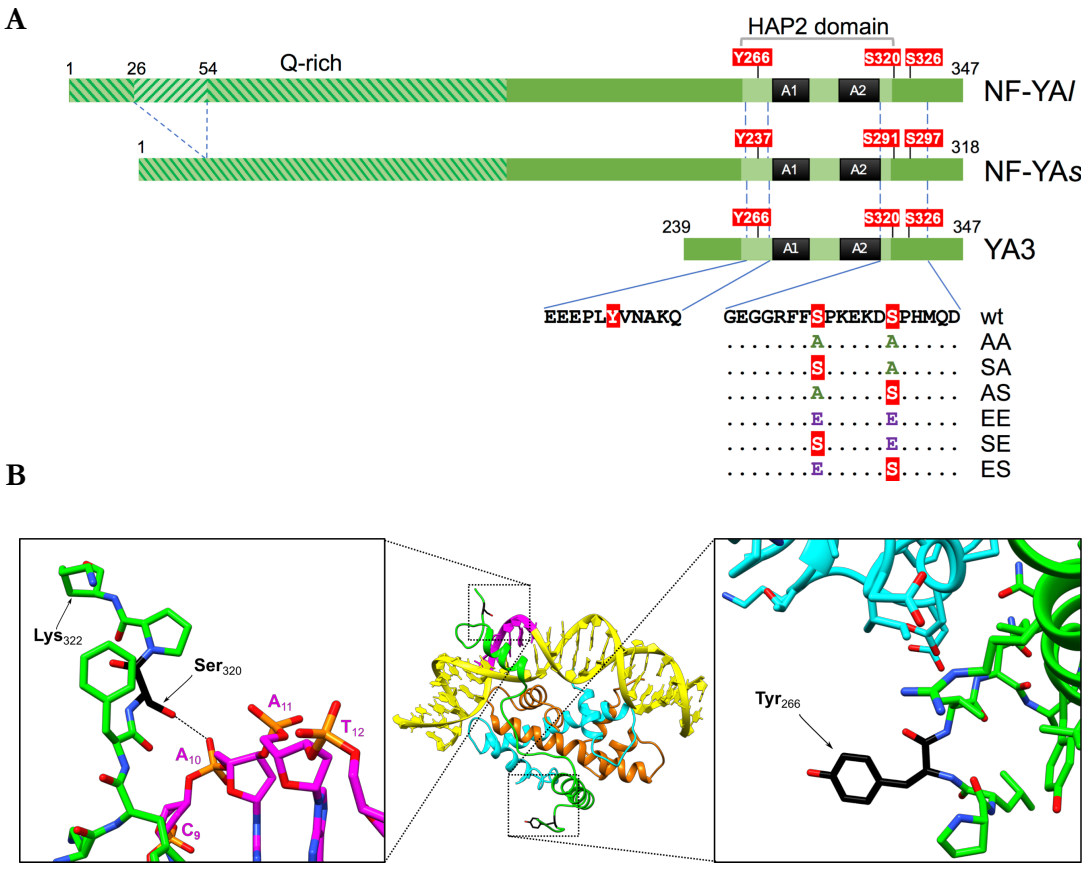


Figure 1. Scheme of NF-YA subunit isoforms, recombinant protein constructs, and structural arrangement of phosphorylated residues.

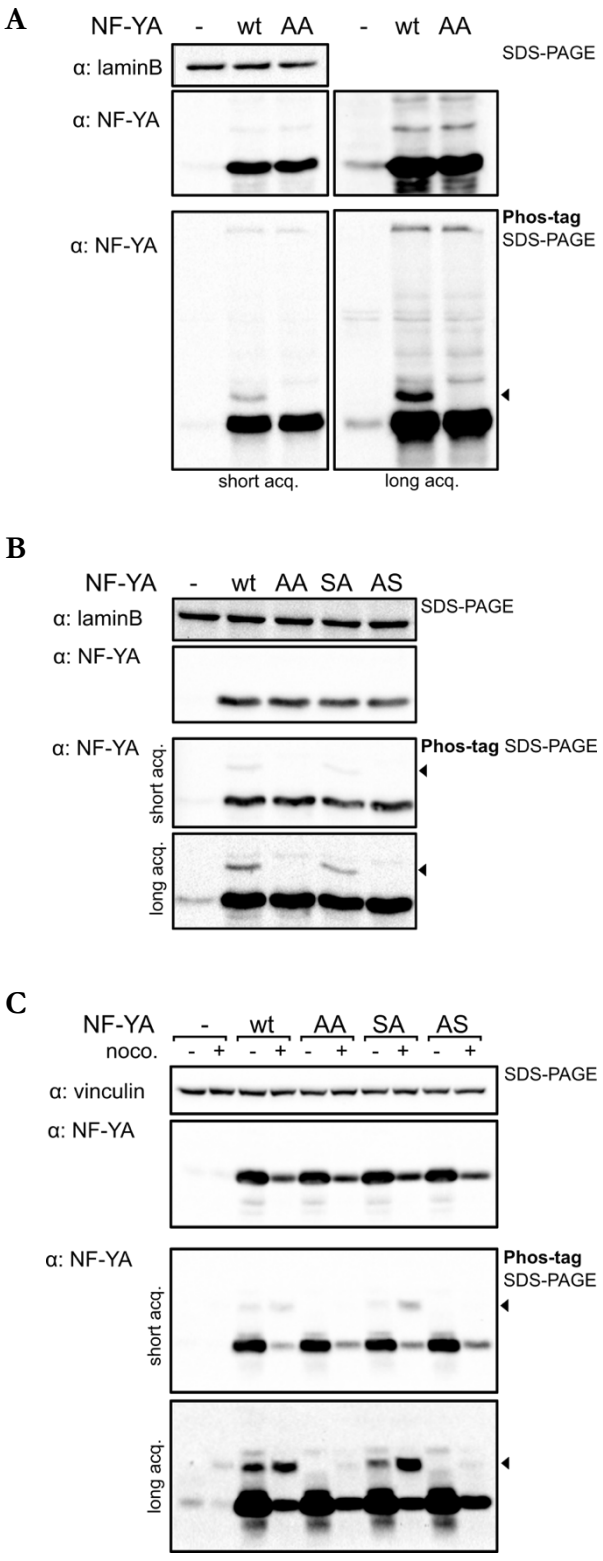
A. Position of the described phosphorylation sites found in PSP (in red) within NF-YA long and short isoforms with their respective amino acid sequence numbering. YA3 recombinant construct used in subsequent biochemical studies is also represented. The main functional domains of the protein are illustrated, including the HAP2 conserved protein core, containing A1 and A2 α-helices, and the N-terminal glutamine-rich transactivation domain (Q-rich). The amino acid context nearby the phosphorylated residues is displayed below, along with the designed panel of phospho-mutants described in the text. **B.** 3D organization of NF-YA phosphorylation sites within the NF-Y/DNA complex. Close-ups of Ser320 (left panel) and Tyr266 (right panel) positions within the 4AWL crystal structure. The regions depicted are indicated in the ribbon representation of the overall NF-Y heterotrimer in complex with the 25 bp HSP70 oligonucleotide (PDB-code: 4AWL; central panel). Dashed line indicates hydrogen-bond interaction between Ser320 and the DNA phosphate backbone. 3D structure chains and elements are depicted with the following color code: NF-YA (green), NF-YB (orange), NF-YC (cyan), DNA (yellow), CCAAT pentanucleotide (magenta), Ser320 and Tyr266 (black). Within the insets, distal protein segments obstructing the close-up views were hidden for clarity. Images were obtained with UCSF Chimera.

Figure 2. NF-YA Ser320 is phosphorylated *in vivo.*

A. Western blot (WB) analysis of HeLa cells nuclear extracts transfected, or not (-), with expression vectors for wild-type (wt) or double alanine mutant (AA) full-length NF-YA. Proteins were separated by standard SDS-PAGE and Phos-tag gel electrophoresis (see Methods), as indicated on the right side of the figure, and transferred to nitrocellulose membranes for immunodetection with anti-NF-YA. Anti-LaminB was used as loading control. In the Phos-tag gel blot, a retarded migration band is visible in wt NF-YA transfected sample, and absent in the mutant sample (arrowhead), as highlighted in the long acquisition time panels of the figure (right).

B. Nuclear extracts of HeLa cells transfected with wt, AA, or single alanine mutant (SA or AS) full-length NF-YA were separated by standard SDS-PAGE or Phos-tag gel electrophoresis and analyzed by WB as described in (A). The NF-YA slower migrating band in the Phos-tag gel (arrowhead) is detected in wt and SA mutant NF-YA transfected samples, and absent in AS or AA samples.

C. Same as B, except that transfected cells were split and treated (+) with 0.1 μg/mL nocodazole (noco.), to enrich for the G2/M population (+), or grown in complete medium (-), for 16 hours.



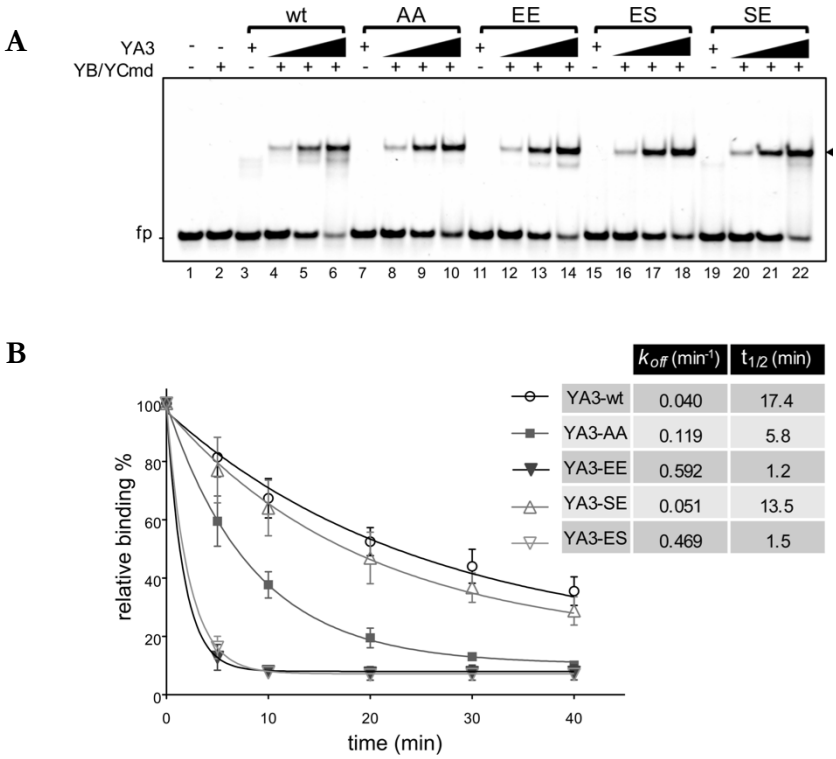


Figure 3. *In vitro* analysis of DNA binding properties of NF-YA Ser mutants.

A. Dose-response EMSA with recombinant YA3 mutant constructs. The amounts of labelled 31 bp HSP70 probe (20 nM) and NF-YB/NF-YC HFD dimer (YB/YCmd; 60 nM) were kept constant. The binding reactions were incubated with increasing concentrations of each YA3 subunit (20, 60, 120 nM). As controls, YA3 proteins (120 nM) or YB/YCmd (60 nM) were incubated alone with the probe. fp: free probe; arrowhead: NF-Y complex. **B.** Off-rate EMSAs quantification. DNA-bound complexes of NF-Y containing wt or mutant YA3 were challenged by addition of excess unlabeled competitor and loaded for electrophoresis at indicated time points. The bound complex was quantified as described in the Methods section. The data are plotted setting at 100% the binding percentage of the sample devoid of competitor DNA. Each data-point is the mean of three independent experiments +/- SD (error bars). The fitting curves for each mutant are overlaid on the plot and the derived off-rate kinetic constants (k_{off}) and complex half-lives ($t_{1/2}$) are reported in the table.

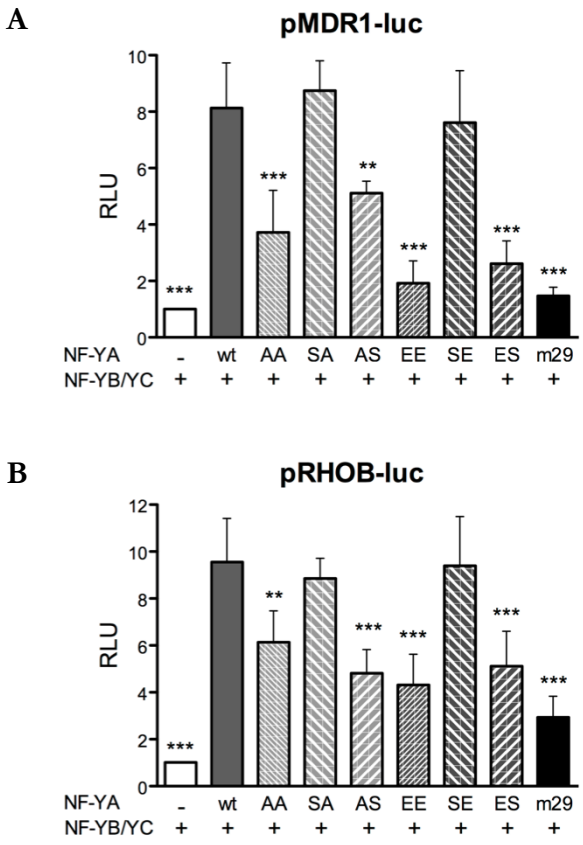


Figure 4. Functional analysis of NF-YA Ser mutants in transcription.

Luciferase-reporter assays performed with MDR1 [60] (**A**) and RHOB [61] (**B**) human promoter constructs. HeLa cells were co-transfected with the reporter and each NF-Y subunit expression vectors, assaying all NF-YAs phospho-mutants panel. Data are reported setting to 1 the Relative Luminescence Units (RLU) of the p-Luc reporter co-transfected only with NF-YB/YC dimer expression vectors. Error bars represent SD. (** $p<0.01$; *** $p<0.001$ according to one-way ANOVA followed by Dunnet's post-test using NF-YA-wt as control column).

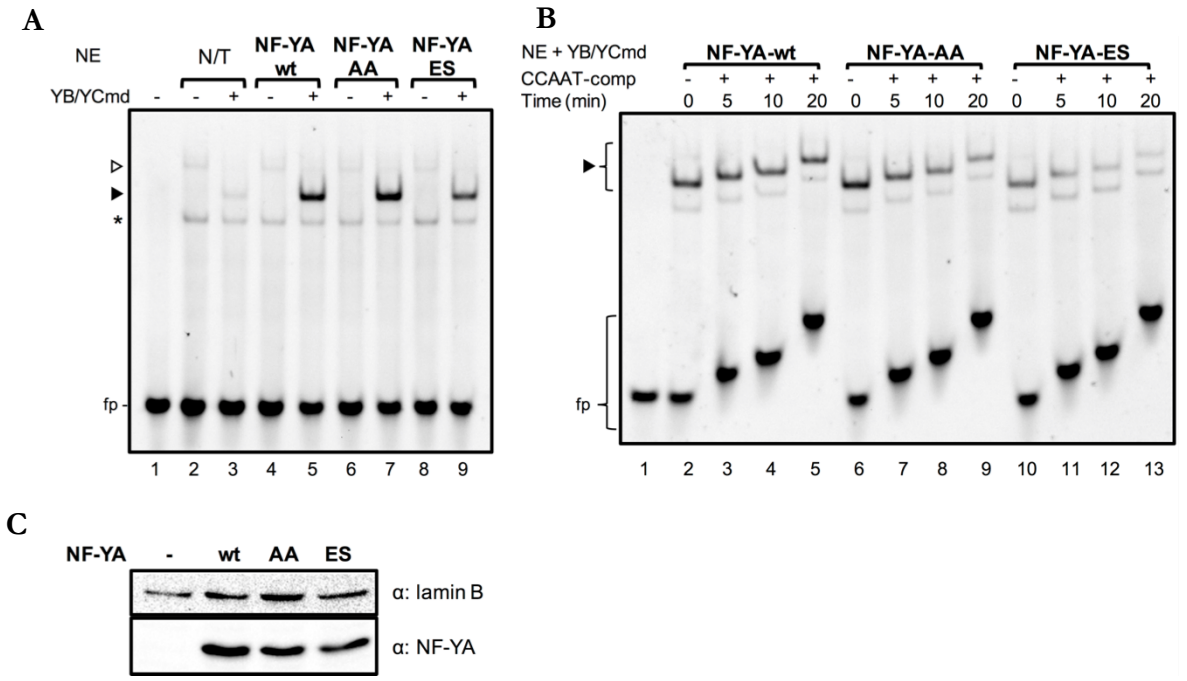


Figure 5. DNA binding properties of full-length wt and mutant NF-YAs expressed *in vivo*.

A. EMSAs with nuclear extracts (NE) of HeLa cells transfected with full-length NF-YA and its phospho-mutant versions (NF-YA-AA and NF-YA-ES). Recombinant NF-YB/NF-YC HFD dimer (YB/YCmd; 20 nM) was included in reaction to allow DNA binding and visualization of the overexpressed NF-YA subunit. An open arrow-head indicates endogenous NF-Y/DNA complex; closed arrow-head: NF-Y complex containing recombinant YB/YCmd; asterisk: non-specific band; fp: free probe. N/T: mock-transfected cells nuclear extract control. **B**. Off-rate EMSA performed in the same conditions as in (A) in presence of recombinant YB/YCmd. The competition reaction has been carried out at 4°C using 50-fold excess of unlabelled CCAAT-competitor DNA. **C**. Immunoblot control of NEs used in EMSA.

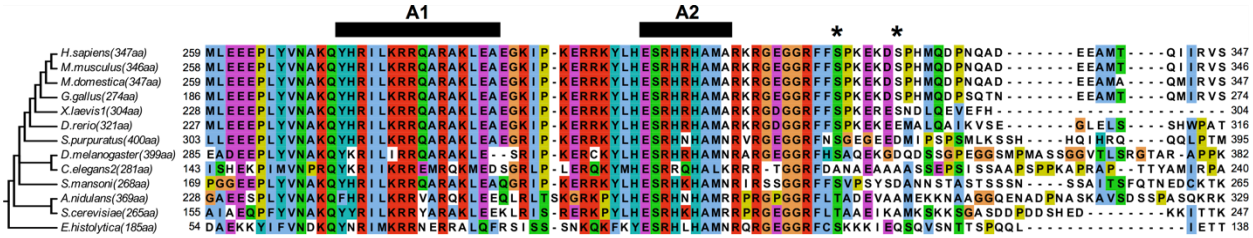


Figure 6. Multiple alignment of NF-YA C-terminal region from different species.

Multiple alignment of NF-YA protein sequence from representative species showing the C-terminal region including A1 and A2 α-helices, obtained with Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo). The position of human Ser320 and Ser326 is indicated by asterisks. The cladogram summarizes the phylogenetic relationships among the species reported. The length of the proteins and the range of the amino acids displayed within the alignment are indicated. Clustalx color scheme has been used. NCBI protein sequence IDs: human NF-YA *(H. sapiens)*: NP_002496; mouse *(M. musculus)*: NP_001334330; opossum *(M. domestica)*: XP_001370287; chicken *(G. gallus)*: NP_001006325; clawed frog *(X. laevis)*: XP_018104815; zebrafish *(D. rerio)*: AAI33846; sea urchin *(S. purpuratus)*: NP_999822; fruit fly *(D. melanogaster)*: NP_648313; nematode round worm *(C. elegans)*: Q5ZEP9; flat worm *(S. mansoni)*: Q26561; aspergillus *(A. nidulans)*: CAA74100; budding yeast *(S. cerevisiae)*: CAA96955; entamoeba *(E. histolytica)*: N9TB85.

MANUSCRIPT REFERENCES

1. Dolfini D, Zambelli F, Pavesi G, Mantovani R. A perspective of promoter architecture from the CCAAT box. Cell Cycle. 2009; 8: 4127-37.
2. Dolfini D, Gatta R, Mantovani R. NF-Y and the transcriptional activation of CCAAT promoters. Crit Rev Biochem. Mol Biol. 2012; 47: 29-49.
3. Nardone V, Chaves-Sanjuan A, Nardini M. Structural determinants for NF-Y/DNA interaction at the CCAAT box. BBA-Gene Regulatory Mechanisms. 2017; 1860: 571-80.
4. Fleming JD, Pavesi G, Benatti P, Imbriano C, Mantovani R, Struhl K. NF-Y coassociates with FOS at promoters, enhancers, repetitive elements, and inactive chromatin regions, and is stereo-positioned with growth-controlling transcription factors. Genome Res. 2013; 23: 1195-209.
5. Sherwood RI, Hashimoto T, O'Donnell CW, Lewis S, Barkal AA, van Hoff JP, Karun V, Jaakkola T, Gifford DK. Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. Nat. Biotechnol. 2014; 32: 171-78.
6. Oldfield AJ, Yang P, Conway AE, Cinghu S, Freudenberg JM, Yellaboina S, Jothi R. Histone-Fold Domain Protein NF-Y Promotes Chromatin Accessibility for Cell Type-Specific Master Transcription Factors. Mol. Cell. 2014; 55: 708-722.
7. Lu F, Liu Y, Inoue A, Suzuki T, Zhao K, Zhang Y. Establishing Chromatin Regulatory Landscape during Mouse Preimplantation Development. Cell. 2016; 165: 1375-88.
8. Gurtner A, Manni I, Piaggio G. NF-Y in cancer: Impact on cell transformation of a gene essential for proliferation. Biochim Biophys Acta. 2017; 1860: 604-16.
9. Gusev Y, Riggins RB, Bhuvaneshwar K, Gauba R, Sheahan L, Clarke R, Madhavan S. In silico discovery of mitosis regulation networks associated with early distant metastases in estrogen receptor positive breast cancers. Cancer Inform. 2013; 12: 31-51.
10. Ung MH, Varn FS, Lou S, Cheng C. Regulators associated with clinical outcomes revealed by DNA methylation data in breast cancer. PLoS Comput Biol. 2015; 11: e1004269.
11. Yang C, Zhao X, Cui N, Liang Y. Cadherins Associate with Distinct Stem Cell-Related Transcription Factors to Coordinate the Maintenance of Stemness in Triple-Negative Breast Cancer. Stem Cells Int. 2017; 2017: 5091541.
12. Andrews E, Wang Y, Xia T, Cheng W, Cheng C. Contextual Refinement of Regulatory Targets Reveals Effects on Breast Cancer Prognosis of the Regulome. PLoS Comput Biol. 2017; 13: e1005340.
13. Zuo ZG, Zhang XF, Ye XZ, Zhou ZH, Wu XB, Ni SC, Song HY. Bioinformatic analysis of RNA-seq data unveiled critical genes in rectal adenocarcinoma. Eur Rev Med Pharmacol Sci. 2016; 20: 3017-25.
14. Pappa KI, Polyzos A, Jacob-Hirsch J, Amariglio N, Vlachos GD, Loutradis D, Anagnou NP. Profiling of Discrete Gynecological Cancers Reveals Novel Transcriptional Modules and Common Features Shared by Other Cancer Types and Embryonic Stem Cells. PLoS One. 2015; 10: e0142229.
15. Shi C, Zhang Z. Screening of potentially crucial genes and regulatory factors involved in epithelial ovarian cancer using microarray analysis. Oncol Lett. 2017; 14: 725-32.
16. Cicchillitti L, Corrado G, Carosi M, Dabrowska ME, Loria R, Falcioni R, Cutillo G, Piaggio G, Vizza E. Prognostic role of NF-YA splicing isoforms and Lamin A status in low grade endometrial cancer. Oncotarget. 2017; 8: 7935-45.

17. Cui H, Zhang M, Wang Y, Wang Y. NF-YC in glioma cell proliferation and tumor growth and its role as an independent predictor of patient survival. Neurosci Lett. 2016; 631: 40-9.
18. Whitfield ML, George LK, Grant GD, Perou CM. Common markers of proliferation. Nat Rev Cancer. 2006; 6: 99-106.
19. Dolfini D, Zambelli F, Pedrazzoli M, Mantovani R, Pavesi G. A high definition look at the NF-Y regulome reveals genome-wide associations with selected transcription factors. Nucleic Acids Res. 2016; 44: 4684-702.
20. Bisikirska B, Bansal M, Shen Y, Teruya-Feldstein J, Chaganti R, Califano A. Elucidation and Pharmacological Targeting of Novel Molecular Drivers of Follicular Lymphoma Progression. Cancer Res. 2016; 76: 664-74.
21. Tong Y, Merino D, Nimmervoll B, Gupta K, Wang YD, Finkelstein D, Dalton J, Ellison DW, Ma X, Zhang J, Malkin D, Gilbertson RJ. Cross-Species Genomics Identifies TAF12, NFYC, and RAD54L as Choroid Plexus Carcinoma Oncogenes. Cancer Cell. 2015; 27: 712-27.
22. Garipov A, Li H, Bitler BG, Thapa RJ, Balachandran S, Zhang R. NF-YA underlies EZH2 upregulation and is essential for proliferation of human epithelial ovarian cancer cells. Mol Cancer Res. 2013; 11: 360-9.
23. Xu Z, Sun Y, Guo Y, Qin G, Mu S, Fan R, Wang B, Gao W, Wu H, Wang G, Zhang Z. NF-YA promotes invasion and angiogenesis by upregulating EZH2-STAT3 signaling in human melanoma cells. Oncol Rep. 2016; 35: 3630-8.
24. Guo J, Kong LM, Peng AF, Long XH, Zhou Y, Shu Y. Transcription factor NF-YA promotes a malignant phenotype by upregulating fatty acid synthase expression. Mol Med Rep. 2016; 14: 5007-14.
25. Costessi A, Mahrour N, Tijchon E, Stunnenberg R, Stoel MA, Jansen PW, Sela D, Martin-Brown S, Washburn MP, Florens L, Conaway JW, Conaway RC, Stunnenberg HG. The tumour antigen PRAME is a subunit of a Cul2 ubiquitin ligase and associates with active NFY promoters. EMBO J. 2011; 30: 3786-98.
26. Chen X, Müller GA, Quaas M, Fischer M, Han N, Stutchbury B, Sharrocks AD, Engeland K. The forkhead transcription factor FOXM1 controls cell cycle-dependent gene expression through an atypical chromatin binding mechanism. Mol Cell Biol. 2013; 33: 227-36.
27. Penkov D, Mateos San Martín D, Fernandez-Díaz LC, Rosselló CA, Torroja C, Sánchez-Cabo F, Warnatz HJ, Sultan M, Yaspo ML, Gabrieli A, Tkachuk V, Brendolan A, Blasi F, Torres M. Analysis of the DNA-binding profile and function of TALE homeoproteins reveals their specialization and specific interactions with Hox genes/proteins. Cell Rep. 2013; 3: 1321-33.
28. Dardaei L, Penkov D, Mathiasen L, Bora P, Morelli MJ, Blasi F. Tumorigenesis by Meis1 overexpression is accompanied by a change of DNA target-sequence specificity which allows binding to the AP-1 element. Oncotarget. 2015; 6: 25175-87.
29. Shi Z, Chiang CI, Labhart P, Zhao Y, Yang J, Mistretta TA, Henning SJ, Maity SN, Mori-Akiyama Y. Context-specific role of SOX9 in NF-Y mediated gene regulation in colorectal cancer cells. Nucleic Acids Res. 2015; 43: 6257-69.
30. Tiwari VK, Stadler MB, Wirbelauer C, Paro R, Schübeler D, Beisel C. A chromatin-modifying function of JNK during stem cell differentiation. Nat. Genet. 2011; 44: 94-100.
31. Göke J, Chan YS, Yan J, Vingron M, Ng HH. Genome-wide kinase-chromatin interactions reveal the regulatory network of ERK signaling in human embryonic stem cells. Mol Cell. 2013; 50: 844-55.
32. Yun J, Chae HD, Choi TS, Kim EH, Bang YJ, Chung J, Choi KS, Mantovani R, Shin DY. Cdk2-dependent phosphorylation of the NF-Y transcription factor and its involvement in the p53-p21 signaling pathway. J Biol Chem. 2003; 278: 36966-72.

33. Chan QK, Lam HM, Ng CF, Lee AY, Chan ES, Ng HK, Ho SM, Lau KM. Activation of GPR30 inhibits the growth of prostate cancer cells through sustained activation of Erk1/2, c-jun/c-fos-dependent upregulation of p21, and induction of G(2) cell-cycle arrest. Cell Death Differ. 2010; 17: 1511-23.
34. Chae HD, Yun J, Bang YJ, Shin DY. Cdk2-dependent phosphorylation of the NF-Y transcription factor is essential for the expression of the cell cycle-regulatory genes and cell cycle G1/S and G2/M transitions. Oncogene. 2004; 23: 4084-88.
35. Maity SN, de Crombrugghe B. 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. 1992; 267: 8286-92.
36. Mantovani R, Li XY, Pessara U, Hooft van Huisjduijnen R, Benoist C, Mathis D. Dominant negative analogs of NF-YA. J Biol Chem. 1994; 269: 20340-46.
37. Sinha S, Kim IS, Sohn KY, de Crombrugghe B, Maity SN. 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. 1996; 16: 328-37.
38. Bellorini M, Zemzoumi K, Farina A, Berthelsen J, Piaggio G, Mantovani R. Cloning and expression of human NF-YC. Gene. 1997; 193: 119-25.
39. Bi W, Wu L, Coustry F, de Crombrugghe B, Maity SN. DNA binding specificity of the CCAAT-binding factor CBF/NF-Y. J Biol Chem. 1997; 272: 26562-72.
40. Nardini M, Gnesutta N, Donati G, Gatta R, Forni C, Fossati A, Vonrhein C, Moras D, Romier C, Bolognesi M, Mantovani R. Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination. Cell. 2013; 152: 132-43.
41. Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res. 2015; 43: D512-20.
42. Kettenbach AN, Schweppe DK, Faherty BK, Pechenick D, Pletnev AA, Gerber SA. Quantitative phosphoproteomics identifies substrates and functional modules of aurora and polo-like kinase activities in mitotic cells. Sci. Signal. 2011; 4: rs5
43. Mertins P, Qiao JW, Patel J, Udeshi ND, Clauser KR, Mani DR, Burgess MW, Gillette MA, Jaffe JD, Carr SA. Integrated proteomic analysis of post-translational modifications by serial enrichment. Nat. Methods. 2013; 10: 634-7.
44. Mertins P, Yang F, Liu T, Mani DR, Petyuk VA, Gillette MA, Clauser KR, Qiao JW, Gritsenko MA, Moore RJ, Levine DA, Townsend R, Erdmann-Gilmore P, et al. Ischemia in tumors induces early and sustained phosphorylation changes in stress kinase pathways but does not affect global protein levels. Mol. Cell. Proteomics. 2014; 13: 1690-704.
45. Xing Y, Fikes JD, Guarente L. Mutations in yeast HAP2/HAP3 define a hybrid CCAAT box binding domain. EMBO J. 1993; 12: 4647-55.
46. Ceribelli M, Benatti P, Imbriano C, Mantovani R. NF-YC complexity is generated by dual promoters and alternative splicing. J Biol Chem. 2009; 284: 34189-200.
47. Coustry F, Maity SN, Sinha S, de Crombrugghe B. 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. 1996; 271: 14485-91.
48. de Silvio A, Imbriano C, Mantovani R. Dissection of the NF-Y transcriptional activation potential. Nucleic Acids Res. 1999; 27: 2578-84.

49. Li XY, Hooft van Huijsduijnen R, Mantovani R, Benoist C, Mathis D. Intron-exon organization of the NF-Y genes. Tissue-specific splicing modifies an activation domain. J. Biol. Chem. 1992; 267: 8984-90.
50. Hydbring P, Malumbres M, Sicinski P. Non-canonical functions of cell cycle cyclins and cyclin-dependent kinases. Nat. Rev. Mol. Cell. Biol. 2016; 17: 280-92.
51. Yun UJ, Park HD, Shin DY. p53 prevents immature escaping from cell cycle G2 checkpoint arrest through inhibiting cdk2-dependent NF-Y phosphorylation. Cancer Res. Treat. 2006; 38: 224-8.
52. Zambelli F, Pavesi G. Genome wide features, distribution and correlations of NF-Y binding sites. Biochim Biophys Acta. 2017; 1860:581-9.
53. Suske G. NF-Y and SP transcription factors - New insights in a long-standing liaison. Biochim Biophys Acta. 2017; 1860: 590-7.
54. Pearlman SM, Serber Z, Ferrell JE Jr. A mechanism for the evolution of phosphorylation sites. Cell. 2011; 147: 934-46.
55. Ubersax JA, Ferrell JE Jr. Mechanisms of specificity in protein phosphorylation. Nat. Rev. Mol. Cell. Biol. 2007; 8: 530-41.
56. Yoshioka Y, Suyari O, Yamaguchi M. Transcription factor NF-Y is involved in regulation of the JNK pathway during Drosophila thorax development. Genes Cells. 2008; 13: 117-30.
57. Ahn J, Choi JH, Won M, Kang CM, Gyun MR, Park HM, Kim CH, Chung KS. The activation of p38 MAPK primarily contributes to UV-induced RhoB expression by recruiting the c-Jun and p300 to the distal CCAAT box of the RhoB promoter. Biochem. Biophys. Res. Commun. 2011; 409: 211-6.
58. Lecona E, Barrasa JI, Olmo N, Llorente B, Turnay J, Lizarbe MA. Upregulation of annexin A1 expression by butyrate in human colon adenocarcinoma cells: role of p53, NF-Y, and p38 mitogen-activated protein kinase. Mol. Cell. Biol. 2008; 28: 4665-74.
59. Diebold ML, Fribourg S, Koch M, Metzger T, Romier C. Deciphering correct strategies for multiprotein complex assembly by co-expression: application to complexes as large as the histone octamer. J. Struct. Biol. 2011; 175: 178-88.
60. Hu Z, Jin S, Scotto KW. Transcriptional activation of the MDR1 gene by UV irradiation. Role of NF-Y and Sp1. J. Biol. Chem. 2000; 275: 2979-85.
61. Vasilaki E, Papadimitriou E, Tajadura V, Ridley AJ, Stournaras C, Kardassis D. Transcriptional regulation of the small GTPase RhoB gene by TGF{beta}-induced signaling pathways. FASEB J. 2010; 24: 891-905.
62. Romier C, Cocchiarella F, Mantovani R, Moras, D. The NF-YB/NF-YC structure gives insight into DNA binding and transcription regulation by CCAAT factor NF-Y. J. Biol. Chem. 2003; 278: 1336-45.

Supplemental Information

SUPPLEMENTAL REFERENCES

63. Magan N, Szremska AP, Isaacs RJ, Stowell KM. Modulation of DNA topoisomerase II alpha promoter activity by members of the Sp (specificity protein) and NF-Y (nuclear factor Y) families of transcription factors. Biochem. J. 2003; 374: 723-9.
64. Sorensen P, Wintersberger E. Sp1 and NF-Y are necessary and sufficient for growth- dependent regulation of the hamster thymidine kinase promoter. J. Biol. Chem. 1999; 274: 30943-9.
65. Bolognese F, Wasner M, Dohna CL, Gurtner A, Ronchi A, Muller H, Manni I, Mossner J, Piaggio G, Mantovani R, Engeland K. The cyclin B2 promoter depends on NF-Y, a trimer whose CCAAT-binding activity is cell-cycle regulated. Oncogene. 1999; 18: 1845-53.
66. Hata M, Ohtsuka K. Characterization of HSE sequences in human Hsp40 gene: structural and promoter analysis. Biochim Biophys Acta. 1998; 1397: 43-55.
67. Borrelli S, Candi E, Hu B, Dolfini D, Ravo M, Grober OM, Weisz A, Dotto GP, Melino G, Viganò MA, Mantovani R. The p63 target HBP1 is required for skin differentiation and stratification. Cell Death Differ. 2010; 17: 1896-907.
68. Rodríguez-Antona C, Bort R, Jover R, Tindberg N, Ingelman-Sundberg M, Gómez- Lechón MJ, Castell JV. Transcriptional regulation of human CYP3A4 basal expression by CCAAT enhancer-binding protein alpha and hepatocyte nuclear factor- 3 gamma. Mol Pharmacol. 2003; 63: 1180-9.
69. Park SH, Lee SR, Kim BC, Cho EA, Patel SP, Kang HB, Sausville EA, Nakanishi O, Trepel JB, Lee BI, Kim SJ. Transcriptional regulation of the transforming growth factor beta type II receptor gene by histone acetyltransferase and deacetylase is mediated by NF-Y in human breast cancer cells. J. Biol. Chem. 2002; 277: 5168-74.

SUPPLEMENTAL FIGURES

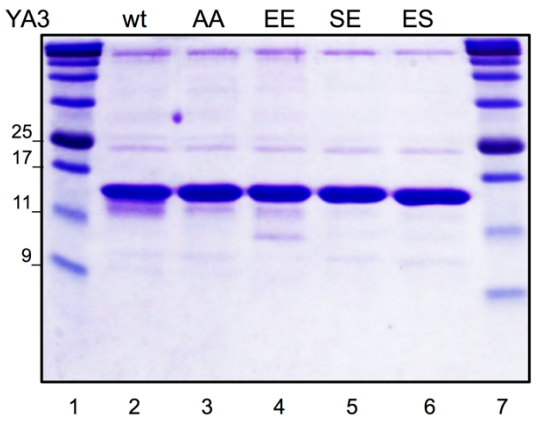


Figure S1. Coomassie stained gel of purified recombinant YA3 proteins.
Coomassie stained gel of IMAC purified recombinant YA3 wild-type and mutant constructs used in EMSAs. 4.5 μg of each recombinant protein preparation were loaded. MW markers are indicated on the left (kDa).

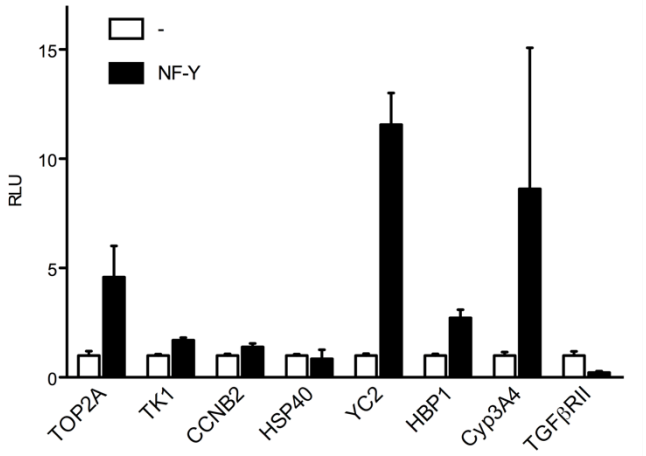


Figure S2. Analysis of NF-Y transactivation of different promoters in Luc reporter assays.
Screening of several NF-Y target luciferase-reporter promoters in response to NF-Y co- transfection in HeLa cells. 100 ng of each reporter alone (-) (open bars) or together with 100 ng of each NF-Y subunit expression vectors (closed bars) were used in each transfection. Data represent the mean of two independent experiments with each sample transfected in duplicate. Data are reported setting to 1 RLU the basal reporter activity. TOP2A-Luc, TK1- Luc, CCNB2-Luc, HSP40-Luc, NF-YC2-Luc, HBP1-Luc, CYP4A4-Luc and TGFβIIR-Luc reporter plasmids were described in: [63-66; 46; 67-69].

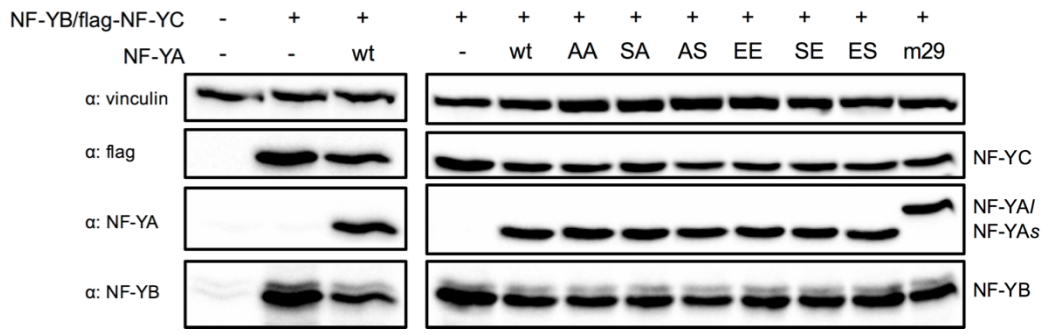


Figure S3. NF-Y subunits and NF-YA mutants overexpression controls.
Western blot analysis of NF-Y subunits expression levels was performed by immunoblot with indicated antibodies on cell extracts after co-transfection of HeLa cells in conditions used for luciferase assays. Vinculin was used as loading control.

PART III

In this final section of the thesis, first I will provide a brief background on the results published in the doctoral thesis of Mariangela Lorenzo (*The interplay among NF-Y and E-BOX Transcription Factors: MYC, MAX and USF1* [246]), representing the basement for the second research topic of my thesis. For more details and complete set of experiments please refer to the above cited publication. Second, I will report the results obtained during my research on this topic.

Introductory overview on the framework of the project

As briefly described in section 2 (*Aims of the thesis*), a detailed analysis of the molecular interplay between NF-Y and E-box TFs was previously carried out in our laboratory. Recombinant versions of NF-Y, MYC, MAX and USF1 were successfully expressed in *E. coli* and purified to perform DNA-binding assays. Among the available E-box TFs, USF1 showed a clear DNA-binding cooperativity with NF-Y on the following natural probes characterized by the recurrent E-box/CCAAT-box genomic spacing of 12-10 bp.

- **LTR** probe (12bp), derived from the LTR *MLT1K* of the ERVL-MaLR family (chr 5):
  Cy5-5'-GGACACCAT**CACGTG**ATCCAAGCCCGG**CCAAT**CAGGTTCTTTCTTTCTCA**-3'**
- **HOXB7** probe (12bp), derived from the human *HOXB7* promoter:
  Cy5-5'-GACCCCGCC**CACGTG**ACGTCCCCTCCG**CCAAT**GGCCGGGCCGTCTCCCCA**-3'**
- **HOXB4** probe (10bp), derived from the human *HOXB4* promoter:
  Cy5-5'-TTAGGCGCC**CACGTG**ATCCTCCGAG**CCAAT**GGCCGCCCCGCCTGCGAT**-3'**

In **Fig 22** an example of off-rate competition EMSA used to evaluate the cooperativity between two proteins is shown, together with the relative quantification derived for the three probes. In this kind of experiment, the pre-assembled protein-DNA complex is challenged with an excess of unlabelled oligonucleotide harbouring the specific DNA element recognized by one of the proteins that compose the complex object of the study. The dissociation curve is followed by loading aliquots of the reaction on a running gel at different time points. A cooperative DNA-binding is characterized by a slower dissociation rate of the protein from the ternary complex (protein-A/protein-B/DNA) if compared to the dissociation rate in absence of the second protein (protein-A/DNA). Note that both E-box and CCAAT-box competitor oligonucleotides have been used, obtaining the same qualitative results. For the quantification, the intensity of the shifted band at the time of competitor addition was set to 100% and the intensities from each time point were scaled accordingly.

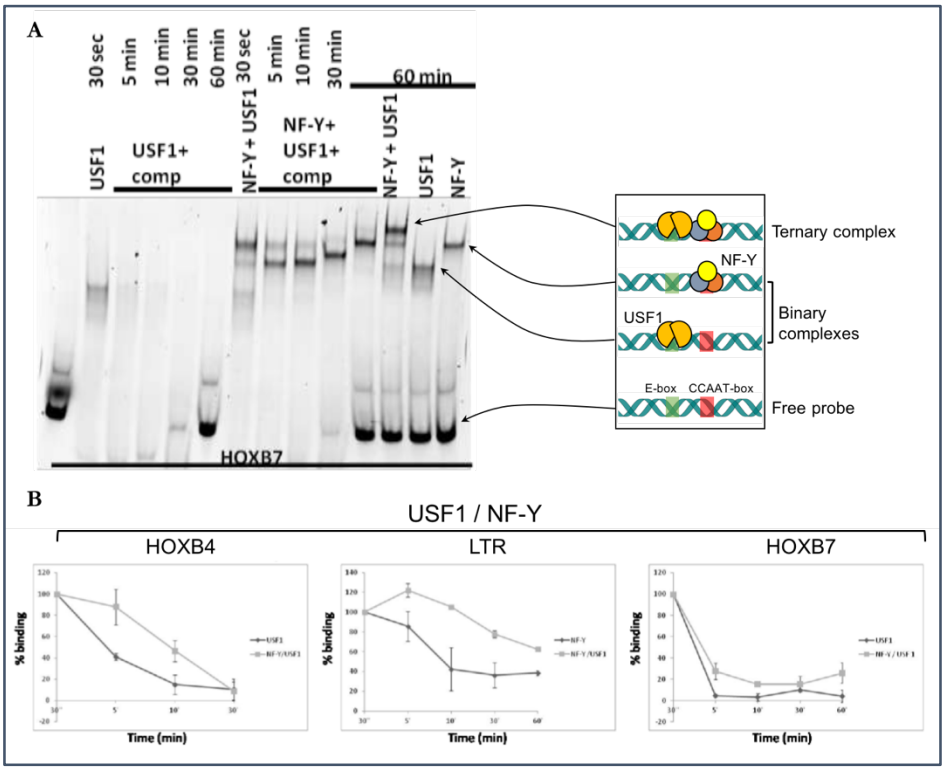


Figure 22. (**A**) Representative example of off-rate competition EMSA. Cooperative binding between USF1 and NF-Y on HOXB7 probe is evaluated by assessing the stability of ternary versus binary complex in parallel reactions over time. In this case an E-box competitor was used. On the right, a cartoon links each molecular species to the correspondent band in the gel. (**B**) Summary of quantifications from each probe in experimental setup as in (A). The dissociation curve for the ternary complex (light grey) is slower than the one for the binary complex (dark grey), indicating a cooperative behaviour between USF1 and NF-Y on these DNA configurations (adapted from [246]).

On the other hand, cooperativity was completely lost in favour of independent binding when 5 additional bp were inserted between the two REs, obtaining the two following probes, with spacing of 17-15bp, respectively:

- **LTR+5** probe (17bp):

  Cy5-5'-GGACACCAT<u>CACGTG</u>AT*<u>AAGAT</u>*CCAAGCCCGG<u>CCAAT</u>CAGGTTCTTTCTTTCTCA-3'

- **HOXB4+5** probe (15bp):

  Cy5-5'-TTAGGCGCC<u>CACGTG</u>AT*<u>AAGAT</u>*CCTCCGAG<u>CCAAT</u>GGCCGCCCCGCCTGCGAT-3'

This observation highlighted the importance of precise spacing between the two DNA elements to elicit any degree of cooperativity between the two proteins, in accordance with the genomic observations (**Fig 23A-B**).

A second, important passage was the use of shorter protein constructs, both for NF-Y and for USF1, in order to check whether cooperativity could be affected by the lack of these regions. Specifically, for NF-Y the same trimer used in the published crystallized complex was used [174], named here NF-Y minimal-domain (NF-Ymd), which only includes the core DNA-binding domains. USF-*short* construct (USF*s* – aa 106-310) lacks a large portion of the N-terminal region

of the protein and was already described in literature as DNA-binding proficient [263]. The resulting behaviour of these shorter constructs replicated the cooperativity described for the full-length proteins for all three target probes (**Fig 23C**).

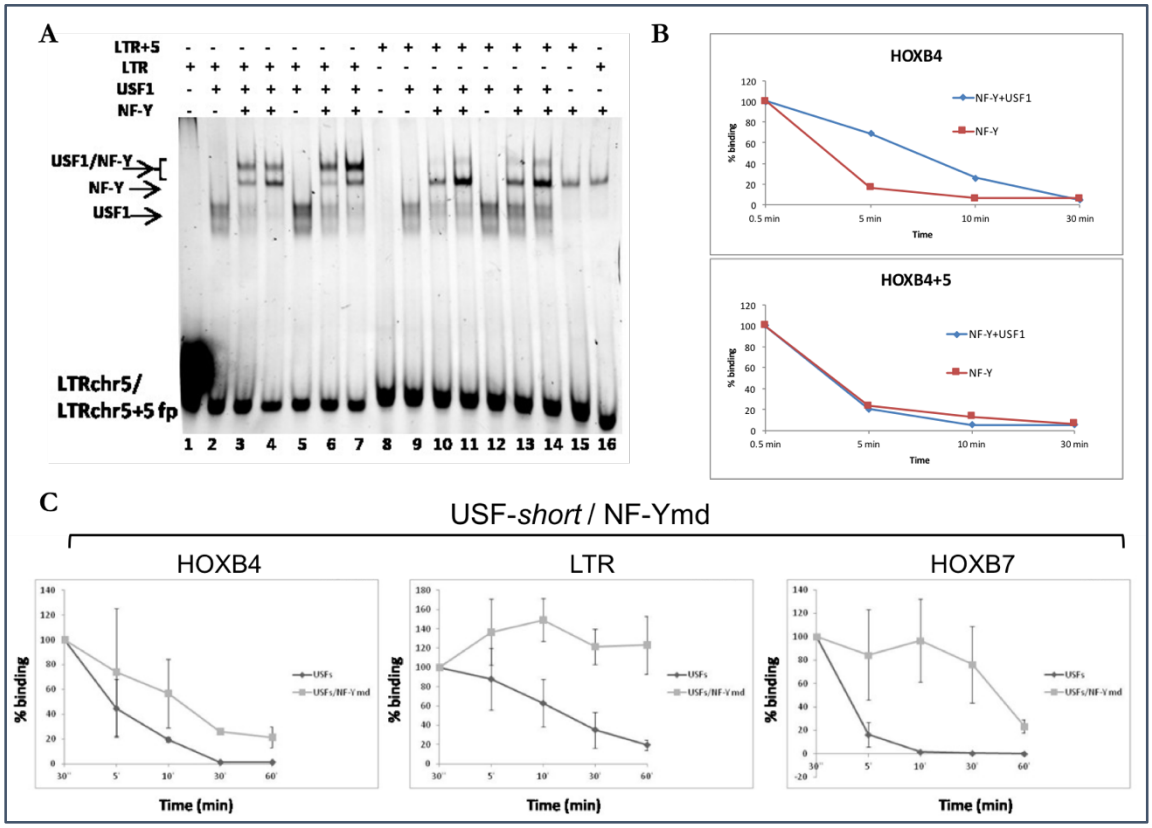


Figure 23. (**A**) Comparative dose-response EMSA with USF1 and NF-Y on LTR (left side) or LTR+5 probes (right side). For each probe, two concentrations of each protein were used, giving a total of four distinct ternary reactions (i.e. including both proteins and DNA). Each protein concentration was also tested in binary reactions (devoid of the second protein). Note that the single proteins bound DNA equally well, irrespectively of the probe configuration. Instead, the tendency to form a ternary complex (upper-shifted band) was dramatically impaired on the LTR+5 DNA configuration. (**B**) Example of off-rate competition EMSA quantification for HOXB4 versus HOXB4+5 probes. Cooperative binding is disrupted by the insertion of 5 bp between E-box and CCAAT-box. (**C**) USF*s* and NF-Ymd constructs completely preserve DNA-binding cooperativity. Quantification of dissociation curves is represented as in Fig 22B (adapted from [246]).

Considering the good production yield of USF-*short* construct and its strong cooperativity with NF-Y, it has been used to reconstitute and isolate a ternary complex with NF-Ymd on either LTR (12 bp spacing) or HOXB4 (10 bp spacing) dsDNA oligonucleotides. The resulting complexes were subjected to SAXS (Small Angle X-ray Scattering - in collaboration with the EMBL-Outstation, Hamburg), a low-resolution X-ray diffraction technique which allows the determination of the size and shape of monodispersed macromolecules in a concentrated solution. Protein domains with known atomic structure can be fitted within the resulting molecular envelope to obtain a clearer picture of the complex in space. From this analysis, performed in collaboration with Prof. M. Nardini, two were the most striking observations: (1) the interaction between the

two proteins does not seem to involve the core bHLH-ZIP domain of USF1, instead the structural data suggest it is mediated by USF1 N-terminal region. This portion of the protein is predicted to be intrinsically disordered, still in these complexes it seems to occupy a well-defined position by interacting with NF-Y core complex. Interestingly, (2) depending on the underlying spacing between the two elements, the structural models point at potentially different NF-Y/USF1 protein regions involved in the interaction (**Fig 24A**). Indeed, USF1 N-termini adopt different conformations in the two models, influenced by a complex combination of distance between the two sites and reciprocal rotation of the two proteins along the DNA axis of about 70°. By estimating the distance covered by USF1 regions linking its core domain with the NF-Y-interactive surface, we can notice a consistent asymmetry in the HOXB4 model (~20Å difference between the two USF1 monomers bridging regions), contrary to the LTR model conformation (**Fig 24B**).

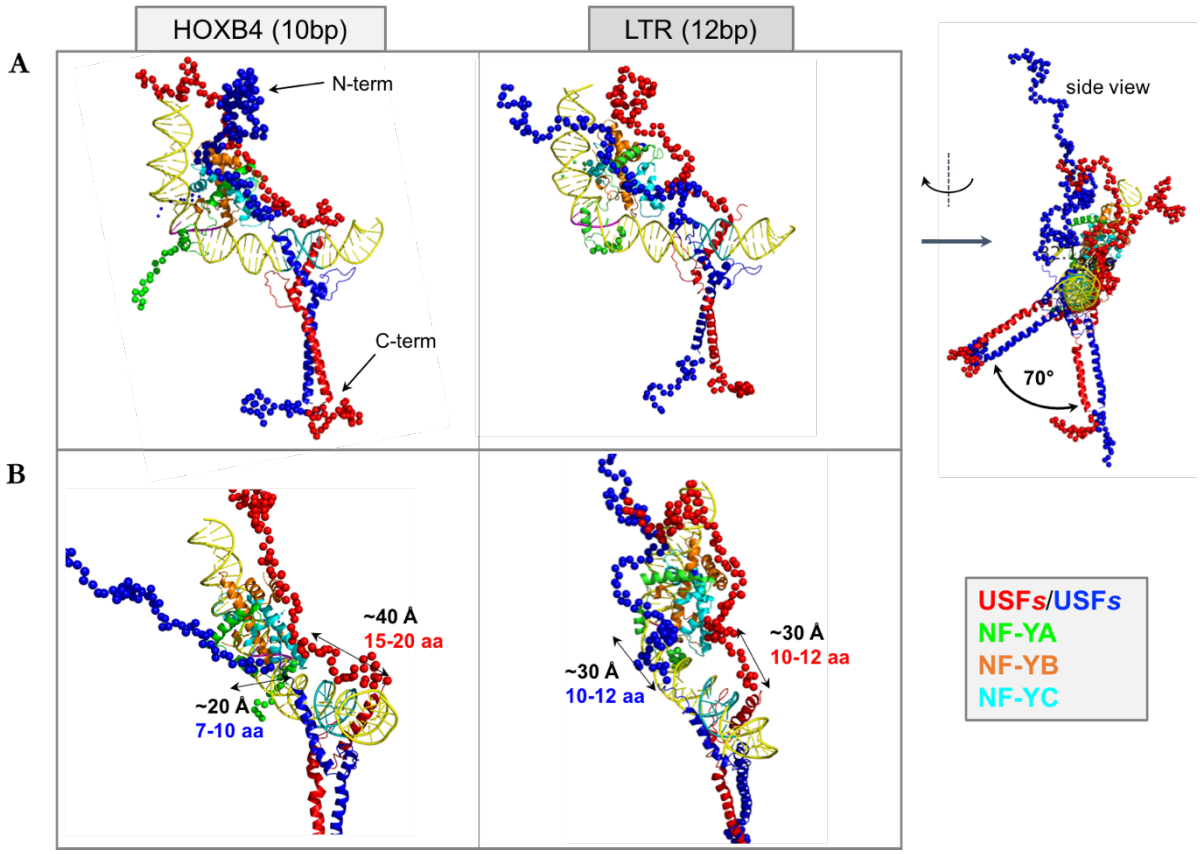


Figure 24. Preliminary SAXS models for USF*s*/NF-Ymd complexes on HOXB4 (48bp DNA, 10bp spacing) or LTR (50bp DNA, 12bp spacing) oligonucleotides. (**A**) Overview of the two complexes with available crystal structures for NF-Y (4AWL) and USF1 (1AN4) fitted within the SAXS model. Protein regions not included in crystal structures are represented as a coarse-grained sphere model. USF*s* N- and C-terminal regions are indicated by arrows. In the right panel the two models (10 vs 12bp spacing) are superimposed on the side view, highlighting the ~70° rotation along the DNA axis imposed by the 2bp difference between the two DNA molecules. (**B**) Estimation of the lengths of USF1 regions involved in bridging the interactive surface with NF-Y. Note the different conformations adopted by USF1 N-termini between the two DNA configurations.

Results report: USF1 cooperativity with NF-Y is mediated by the USR domain

Given the results reported above, three were the main open questions regarding the interplay between USF1 and NF-Y: (I) do the two TFs provide transcriptional synergy on natural promoters harbouring the 10-12 bp spacing signature between E-box and CCAAT-box? (II) Given the structural data, which region of USF1 N-terminus is mediating the cooperative interaction with NF-Y? (III) Can we provide a functional evidence of the interdependency of USF1 and NF-Y co-association at a genome-wide scale? Here I will present the results obtained so far, aimed to account for each of the questions stated above.

NF-Y and USF1 can synergise on natural promoters with different DNA configurations

To assess the question concerning the potential transcriptional synergy, we set up luciferase-assays using reporter constructs harbouring the following human target promoters:

- pHOXB7: [-264/+106] region from human *HOXB7* gene. We subcloned this region from the pKAT-CAT construct, obtained by A. Carè and described in [221], by PCR amplification, T/A cloning and subsequent restriction enzyme subcloning into pGL3-*basic*.
- pHOXC9: [-305/+16] region from human *HOXC9* gene, designed by gene-synthesis and passed into pGL3-*basic* by restriction enzyme based cloning.
- pHOXB4: [-90 bp] region from human *HOXB4* gene in pGL3-*basic*, gently provided by SG. Emerson and described in [199].
- pHOXC4: [-775/+40] region from human *HOXC4* gene in pGL3-*basic* plasmid, gently provided by P. Casali.

We checked for NF-Y and USF1 co-occupancy on these promoters in at least one of the ENCODE cell-lines for which ChIP-seq data is available. They all resulted positive for both TFs. We chose these target promoters since they present either 12 bp (*HOXB7, HOXC9*) or 10 bp (*HOXB4, HOXC4*) canonical spacing between E-box and CCAAT-box. HeLa cells were co-transfected with each reporter plasmid and expression vectors for NF-Y subunits, USF1 or both.

The results, displayed in **Fig 25**, show that the extent of activation dependent on NF-Y or USF1 varied among the promoters, absent for *HOXB7* and high for *HOXB4* or *HOXC9*. By combining NF-Y and USF1 transfection, the degree of synergy is variable and depends on the promoter considered. The two TFs promote a clear-cut synergy on *HOXB7* promoter, in which the co-transfection of both NF-Y and USF1 triggers a clear more-than-additive transcriptional response. Concerning the other promoters, the synergistic response is less sharp, but we obtained in all cases a clear increase in reporter activity. This observation excludes a potential negative interaction between the two TFs, at least in this system. Notably, the two DNA configurations (12-10 bp spacing) did not seem to correlate with any differential elicited response.

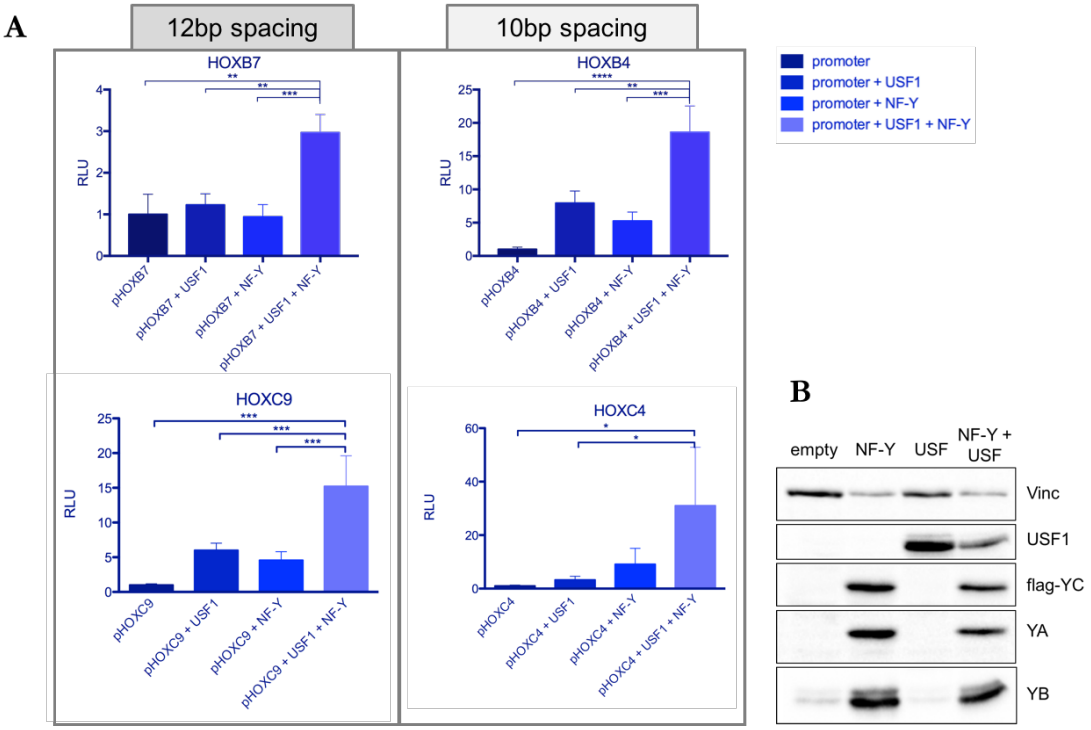


Figure 25. Transcriptional synergy between USF1 and NF-Y on different human natural promoters. (**A**) Luciferase assays were performed on HeLa cells co-transfected with the reporter and expression vectors for either USF1, NF-Y or both. Cells were harvested 40 hours later. Data are reported setting to 1 the normalized relative luminescence unit (RLU) of the 'only-promoter' sample. Error bars represent SD. (* $p<0.05$; ** $p<0.01$; *** $p<0.001$ according to one-way ANOVA followed by Tukey's multiple comparison test. Only statistical significance against the USF1+NF-Y column is shown). (**B**) Western-blot analysis performed on HeLa cells total extracts in conditions used for luciferase assay is shown. Vinculin (Vinc) is used as loading control.

USF1 cooperativity with NF-Y is mediated by the USR domain

In order to fully characterize the USF1 interaction regions with NF-Y and validate the structural models of their arrangement on DNA, we employed a progressive deletion strategy to assess either maintenance or ablation of DNA-binding cooperativity. We designed three additional USF1 constructs, represented in **Fig 26**, progressively deleting the N-terminal portion of the protein. The SAXS structural data provided a putative USF1 density in direct contact with NF-Y core complex, compatible in length with the USR (USF-specific region), a 40 aa transactivation domain. Therefore, the first designed construct (USF-USR) includes the complete USR domain (aa 158-310). The subsequent one presents an additional 19 aa N-terminal deletion, displaying a half-truncated USR (USF-dUSR, aa 177-310). The last and shorter construct only includes the C-terminal region of the protein harbouring the entire DNA-binding and dimerization domain, hence named USF minimal-domain (USF-md, aa 197-310). The constructs were obtained by gene-synthesis, subcloned in prokaryotic expression vectors and purified as soluble C-terminal His-tagged recombinant proteins.

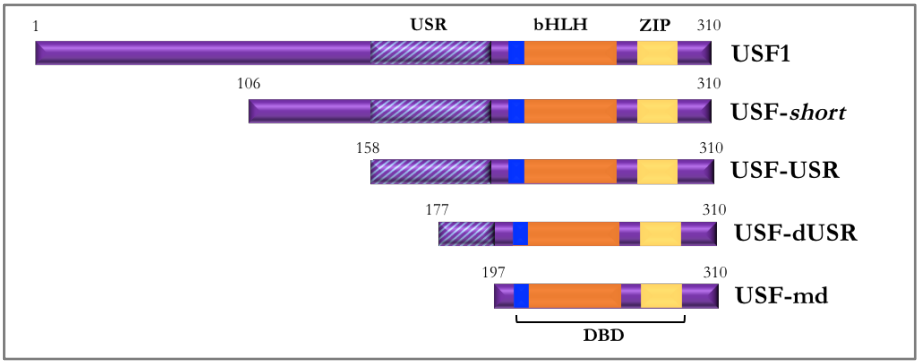


Figure 26. Schematic representation of the complete series of USF1 constructs used in the project. The main functional domains of the protein are indicated as coloured boxes. The latter three deletion constructs (USR, dUSR and md) were designed as synthetic genes.

First, we used these recombinant constructs to assess DNA-binding and complex formation with NF-Ymd in simple dose-response EMSAs as shown in **Fig 27**. All USF1 deletion forms are proficient in DNA binding with the similar affinities for all E-box containing probes. We set up the conditions in terms of protein concentrations allowing the proper detection of the ternary complex for subsequent off-rate competition assays.

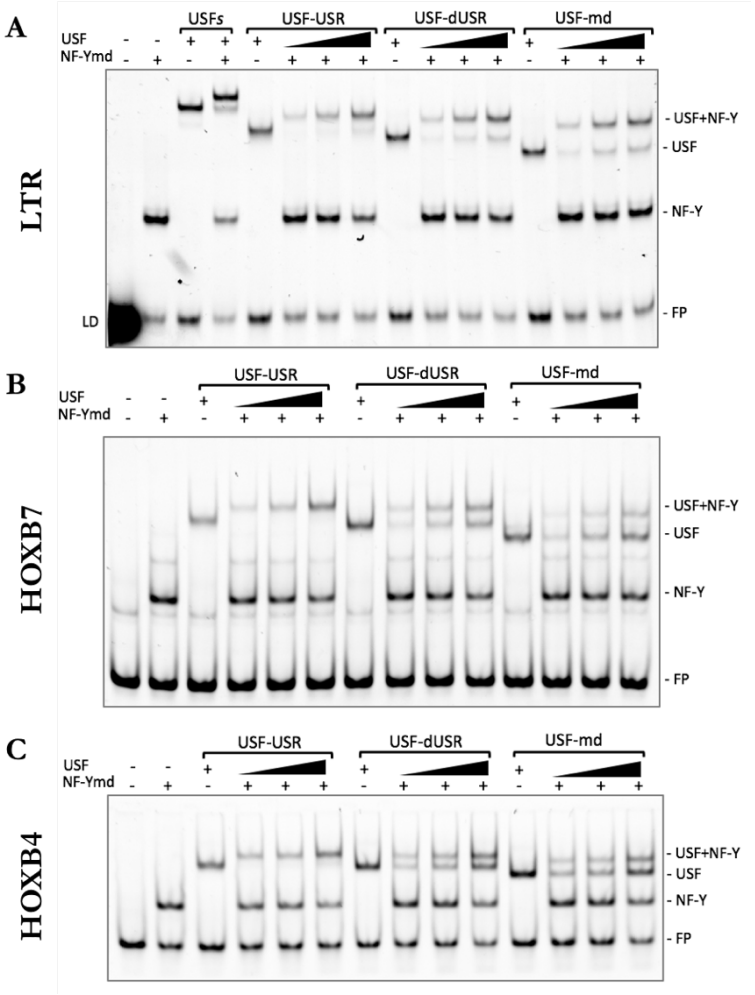


Figure 27. Dose-response EMSA experiments performed with the three USF1 N-term deletion constructs USF-USR, USF-dUSR and USF-md on the three distinct probes LTR (**A**), HOXB7 (**B**) and HOXB4 (**C**). The probe concentration was kept constant at 20 nM. NF-Ymd 2:1 (protein:DNA) in (A) and 1.5:1 in (B) and (C). USF concentration were 1:1, 2:1, 3:1 in (A) and (B), and 2:1, 3:1, 4:1 in (C). In (A) also a USF-*short* control was included (wells 3-4). FP: free probe; LD: loading dye. Bands identity are indicated on the right side of each gel.

Next, cooperativity between NF-Ymd and each USF1 deletion construct was evaluated by off-rate competition assays, as described in the introductory paragraph, using an E-box competitor. A reaction devoid of NF-Ymd was carried out in parallel and loaded in the same gel. The same kind of experiment was carried out on each of the three selected probes, for a total of nine construct/probe combinations. The results of **Fig 28** show the quantification of dissociation curves derived from of each experiment performed in triplicate, while an example of each gel is show in **Fig 29**.

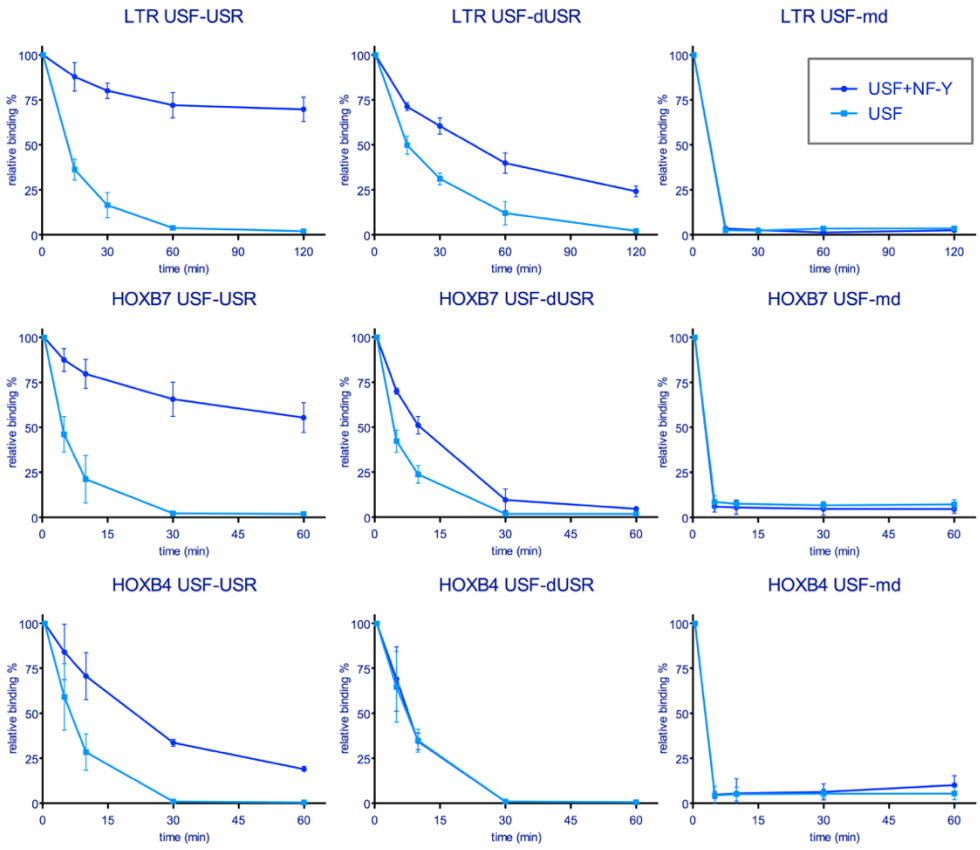


Figure 28. Off-rate competition EMSAs quantification summary. Each plot displays the average of three independent experimental replicates. Error bars represent SD. The shifted-band intensity in each experiment at the time of E-box competitor addition (100x for LTR, 50x for HOXB7 and HOXB4) is set to 100. The two curves derive from the dissociation reactions of USF/DNA complex (light blue) and USF/NF-Ymd/DNA complex (dark blue) run in parallel and loaded on the same gel at the indicated time points.

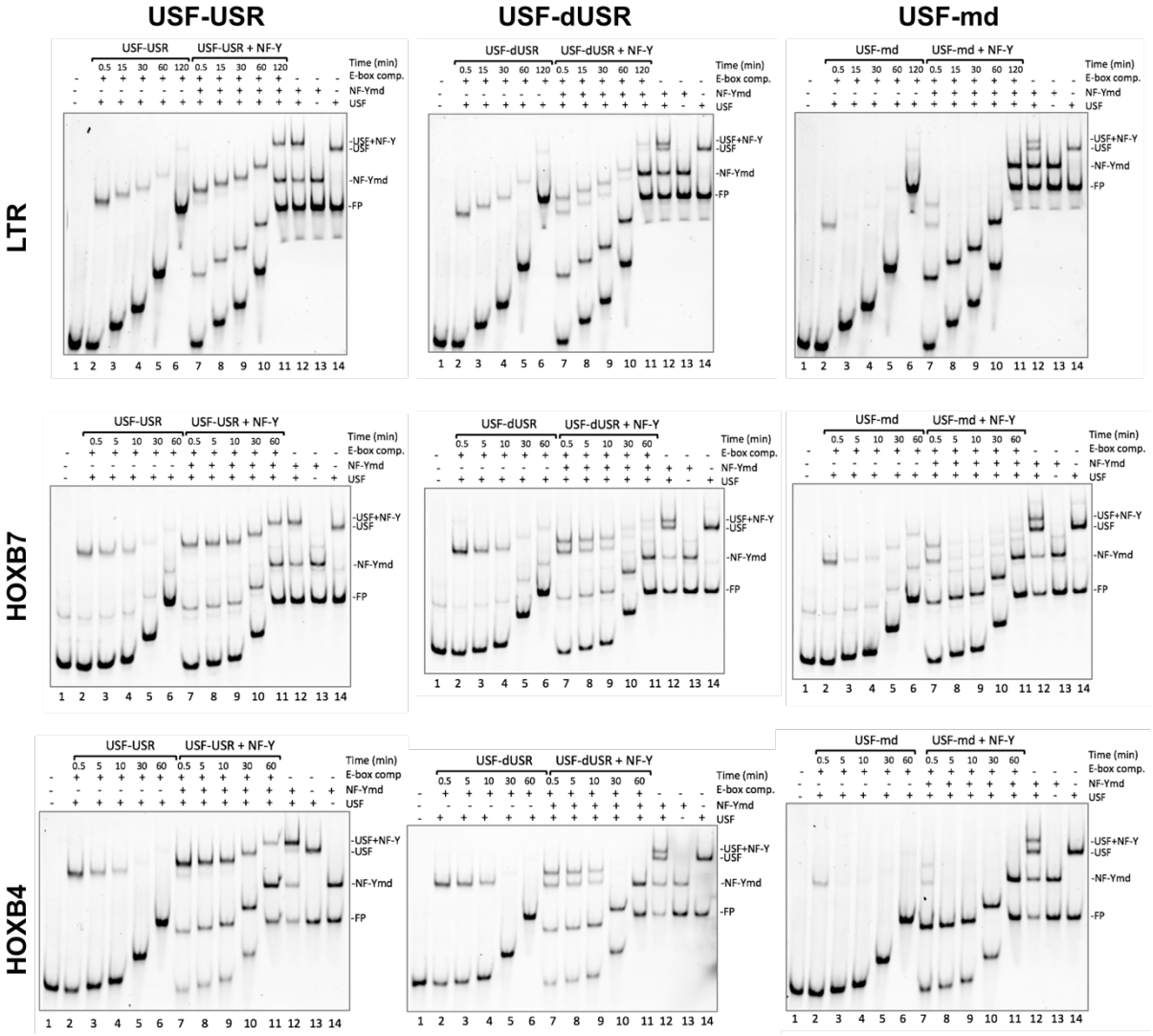


Figure 29. Representative off-rate competition EMSAs for each combination of USF1 deletion-constructs (columns) and probes (rows). In each gel, both the reaction devoid (left series) and including (right series) NF-Ymd is present. Time points indicate the time elapsed from competitor (E-box comp.) addition to each reaction. Wells 12-13-14 include controls devoid of E-box competitor and loaded at the last time-point. FP: free probe.

From these DNA-binding kinetic data, we can make the following observations:

- USF-USR maintains cooperative binding with NF-Ymd on all probes (i.e. the dissociation of the ternary complex is slower than the dissociation of the binary complex), similarly to USF*s* (confront with **Fig 23C**), demonstrating that the USR region is fully sufficient to mediate the interaction with NF-Y on 10-12 bp DNA configurations. Moreover, the degree of cooperativity is markedly reduced in the HOXB4 (10 bp) configuration, as previously described with USF*s*.
- USF-md does not cooperate with NF-Ymd, regardless of the probe used, showing clear independent binding. Moreover, USF-md binding to DNA is intrinsically unstable, quickly

reaching the final equilibrium upon competitor addition (<5 min). These results confirm that the USF1 DNA-binding/dimerization domain is not responsible for direct DNA-binding cooperativity with NF-Y, as suggested by the structural models.

- USF-dUSR construct, which lacks half-USR, reveals marked differences between the 12 and 10 bp spacing configurations. It displays cooperative behaviour on LTR and HOXB7, albeit reduced in respect to USF-USR. On the contrary, cooperativity is clearly absent on the 10 bp spacing configuration (HOXB4), where the two dissociation curves collapse, indicating independent binding.

This line of evidence strongly points at the USR as the USF1 portion in charge of specifically mediating DNA-binding cooperativity with NF-Y, both in 10 and 12 bp spaced sites. Moreover, different stretches of the USR region seem to be involved in this protein-protein interaction. Overall, these EMSA-based observations are in accordance with the structural data.

Bibliography

[1] T. I. Lee and R. A. Young, "Transcriptional regulation and its misregulation in disease," *Cell*, vol. 152, no. 6, pp. 1237–1251, 2013.

[2] M. Hantsche and P. Cramer, "The Structural Basis of Transcription: 10 Years After the Nobel Prize in Chemistry.," *Angew. Chem. Int. Ed. Engl.*, vol. 55, no. 52, pp. 15972–15981, 2016.

[3] C. N. Tennyson, "The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionaly spliced," *Nat. Genet.*, vol. 10, pp. 196–201, 1995.

[4] X. Darzacq *et al.*, "In vivo dynamics of RNA polymerase II transcription," *Nat. Struct. Mol. Biol.*, vol. 14, no. 9, pp. 796–806, 2007.

[5] M. H. Larson *et al.*, "Trigger loop dynamics mediate the balance between the transcriptional fidelity and speed of RNA polymerase II," *Proc. Natl. Acad. Sci.*, vol. 109, no. 17, pp. 6555–6560, 2012.

[6] M. Palangat, M. Larson, X. Hu, A. Gnatt, S. Block, and R. Landick, "Efficient reconstitution of transcription elongation complexes for single-molecule studies of eukaryotic RNA polymerase II," *Transcription*, vol. 3, no. 3, pp. 146–153, 2012.

[7] S. Venkatesh and J. L. Workman, "Histone exchange, chromatin structure and the regulation of transcription," *Nat. Rev. Mol. Cell Biol.*, vol. 16, no. 3, pp. 178–189, 2015.

[8] D. L. Bentley, "Coupling mRNA processing with transcription in time and space," *Nat. Rev. Genet.*, vol. 15, no. 3, pp. 163–175, 2014.

[9] E. T. Wang *et al.*, "Alternative isoform regulation in human tissue transcriptomes," *Nature*, vol. 456, no. 7221, pp. 470–476, 2008.

[10] I. Dunham *et al.*, "An integrated encyclopedia of DNA elements in the human genome," *Nature*, vol. 489, no. 7414, pp. 57–74, 2012.

[11] S. Djebali *et al.*, "Landscape of transcription in human cells," *Nature*, vol. 489, no. 7414, pp. 101–108, 2012.

[12] M. Gonzàlez-Porta, A. Frankish, J. Rung, J. Harrow, and A. Brazma, "Transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene," *Genome Biol.*, vol. 14, no. 7, p. R70, 2013.

[13] N. A. Hoffmann *et al.*, "Molecular structures of unbound and transcribing RNA polymerase III," *Nature*, vol. 528, no. 7581, pp. 231–236, 2015.

[14] K. H. Choi, "Viral polymerases," *Adv. Exp. Med. Biol.*, vol. 726, pp. 267–304, 2012.

[15] M. Rong, B. He, W. T. McAllister, and R. K. Durbin, "Promoter specificity determinants of T7 RNA polymerase.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 95, no. 2, pp. 515–519, 1998.

[16] K. P. White *et al.*, "Microarray analysis of Drosophila development during metamorphosis.," *Science*, vol. 286, no. 5447, pp. 2179–84, 1999.

[17] N. D. L. Owens *et al.*, "Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development," *Cell Rep.*, vol. 14, no. 3, pp. 632–647, 2016.

[18] J. J. O'Malley and J. L. Weaver, "Subunit Structure of Glucose Oxidase from Aspergillus niger," *Biochemistry*, vol. 11, no. 19, pp. 3527–3532, 1972.

[19] P. Oudet, M. Gross-Bellard, and P. Chambon, "Electron microscopic and biochemical evidence that chromatin structure is a repeating unit," *Cell*, vol. 4, no. 4, pp. 281–300, 1975.

[20] J. 0 Thomas and R. D. Kornbergt, "An octamer of histones in chromatin and free in solution," *Biochemistry*, vol. 72, no. 7, pp. 2626–2630, 1975.

[21] G. Arents and E. N. Moudrianakis, "The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 92, no. 24, pp. 11170–4, 1995.

[22] E. J. Draizen, A. K. Shaytan, L. Mariño-Ramírez, P. B. Talbert, D. Landsman, and A. R. Panchenko, "HistoneDB 2.0: A histone database with variants - An integrated resource to explore histones and their variants," *Database*, vol. 2016, 2016.

[23] K. Sandman and J. N. Reeve, "Archaeal histones and the origin of the histone fold," *Curr. Opin. Microbiol.*, vol. 9, no. 5, pp. 520–525, 2006.

[24] F. Mattiroli *et al.*, "Structure of histone-based chromatin in Archaea," *Science (80-.).*, vol. 357, no. 6351, pp. 609–612, 2017.

[25] T. Schalch, S. Duda, D. F. Sargent, and T. J. Richmond, "X-ray structure of a tetranucleosome and its implications for the chromatin fibre," *Nature*, vol. 436, no. 7047, pp. 138–141, 2005.

[26] F. Song *et al.*, "Cryo-EM Study of the Chromatin Fiber Reveals a Double Helix Twisted by Tetranucleosomal Units," *Science (80-.).*, vol. 344, no. 6182, pp. 376–380, 2014.

[27] S. A. Sewitz, Z. Fahmi, and K. Lipkow, "Higher order assembly: folding the chromosome," *Curr. Opin. Struct. Biol.*, vol. 42, pp. 162–168, 2017.

[28] B. Bonev and G. Cavalli, "Organization and function of the 3D genome," *Nat. Rev. Genet.*, vol. 17, no. 12, pp. 772–772, 2016.

[29] J. J. Hayes and A. P. Wolffe, "Histones H2A/H2B inhibit the interaction of transcription factor IIIA with the Xenopus borealis somatic 5S RNA gene in a nucleosome," *Proc Natl Acad Sci U S A*, vol. 89, no. 4, pp. 1229–1233, 1992.

[30] M. G. Izban and D. S. Luse, "Transcription on nucleosomal templates by RNA polymerase II in vitro: Inhibition of elongation with enhancement of sequence-specific pausing," *Genes Dev.*, vol. 5, no. 4, pp. 683–696, 1991.

[31] G. A. Hartzog, J. L. Speer, and D. L. Lindstrom, "Transcript elongation on a nucleoprotein template," *Biochim. Biophys. Acta - Gene Struct. Expr.*, vol. 1577, no. 2, pp. 276–286, 2002.

[32] K. Struhl, "Fundamentally different logic of gene regulation in eukaryotes and prokaryotes," *Cell*, vol. 98, no. 1, pp. 1–4, 1999.

[33] M. I. Arnone and E. H. Davidson, "The hardwiring of development: organization and function of genomic regulatory systems.," *Development*, vol. 124, no. 10, pp. 1851–1864, 1997.

[34] S. Payankaulam, L. M. Li, and D. N. Arnosti, "Transcriptional repression: Conserved and evolved features," *Curr. Biol.*, vol. 20, no. 17, 2010.

[35] T. Jensen, A. Jacquier, and D. Libri, "Dealing with pervasive transcription," *Mol. Cell*, vol. 52, no. 4, pp. 473–484, 2013.

[36] C. D. Kaplan, "Transcription Elongation Factors Repress Transcription Initiation from Cryptic Sites," *Science (80-.).*, vol. 301, no. 5636, pp. 1096–1099, 2003.

[37] A. N. Yadon, D. Van de Mark, R. Basom, J. Delrow, I. Whitehouse, and T. Tsukiyama, "Chromatin Remodeling around Nucleosome-Free Regions Leads to Repression of Noncoding RNA Transcription," *Mol. Cell. Biol.*, vol. 30, no. 21, pp. 5110–5122, 2010.

[38] Y. Atlasi and H. G. Stunnenberg, "The interplay of epigenetic marks during stem cell differentiation and development," *Nat. Rev. Genet.*, 2017.

[39] A. V. Probst, E. Dunleavy, and G. Almouzni, "Epigenetic inheritance during the cell cycle," *Nat. Rev. Mol. Cell Biol.*, vol. 10, no. 3, pp. 192–206, 2009.

[40] C. D. Allis and T. Jenuwein, "The molecular hallmarks of epigenetic control," *Nat. Rev. Genet.*, vol. 17, no. 8, pp. 487–500, 2016.

[41] B. Li, M. Carey, and J. L. Workman, "The Role of Chromatin during Transcription," *Cell*, vol. 128, no. 4, pp. 707–719, 2007.

[42] K. Luger, a W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond, "Crystal structure of the nucleosome core particle at 2.8 A resolution.," *Nature*, vol. 389, no. 6648, pp. 251–260, 1997.

[43] C. M. Rivera and B. Ren, "Mapping human epigenomes," *Cell*, vol. 155, no. 1, pp. 39–55, 2013.

[44] E. Wingender, T. Schoeps, and J. Dönitz, "TFClass: An expandable hierarchical classification of human transcription factors," *Nucleic Acids Res.*, vol. 41, no. D1, 2013.

[45] C.-C. Hon *et al.*, "An atlas of human long non-coding RNAs with accurate 5' ends," *Nature*, vol. 543, no. 7644, pp. 199–204, 2017.

[46] M. Madan Babu, S. A. Teichmann, and L. Aravind, "Evolutionary Dynamics of Prokaryotic Transcriptional Regulatory Networks," *J. Mol. Biol.*, vol. 358, no. 2, pp. 614–633, 2006.

[47] N. D. Heintzman and B. Ren, "The gateway to transcription: Identifying, characterizing and understanding promoters in the eukaryotic genome," *Cell. Mol. Life Sci.*, vol. 64, no. 4, pp. 386–400, 2007.

[48] M. C. Thomas and C. M. Chiang, "The general transcription machinery and general cofactors.," *Crit. Rev. Biochem. Mol. Biol.*, vol. 41, no. 3, pp. 105–178, 2006.

[49] N. I. Gershenzon and I. P. Ioshikhes, "Synergy of human Pol II core promoter elements revealed by statistical sequence analysis," *Bioinformatics*, vol. 21, no. 8, pp. 1295–1300, 2005.

[50] T. Morton *et al.*, "Paired-End Analysis of Transcription Start Sites in *Arabidopsis* Reveals Plant-Specific Promoter Signatures," *Plant Cell*, vol. 26, no. 7, pp. 2746–2760, 2014.

[51] A. Sandelin, P. Carninci, B. Lenhard, J. Ponjavic, Y. Hayashizaki, and D. A. Hume, "Mammalian RNA polymerase II core promoters: insights from genome-wide studies," *Nat. Rev. Genet.*, vol. 8, no. 6, pp. 424–436, 2007.

[52] A. M. Deaton and A. Bird, "CpG islands and the regulation of transcription," *Genes Dev.*, vol. 25, no. 10, pp. 1010–1022, 2011.

[53] B. Lenhard, A. Sandelin, and P. Carninci, "Metazoan promoters: emerging characteristics and insights into transcriptional regulation," *Nat. Rev. Genet.*, 2012.

[54] P. Carninci *et al.*, "Genome-wide analysis of mammalian promoter architecture and evolution," *Nat. Genet.*, vol. 38, no. 6, pp. 626–635, 2006.

[55] "The Transcriptional Landscape of the Mammalian Genome," *Science (80-.).*, vol. 309, no. 5740, pp. 1559–1563, 2005.

[56] E. A. Rach, H.-Y. Yuan, W. H. Majoros, P. Tomancak, and U. Ohler, "Motif composition, conservation and condition-specificity of single and alternative transcription start sites in the Drosophila genome," *Genome Biol.*, vol. 10, no. 7, p. R73, 2009.

[57] A. C. Seila *et al.*, "Divergent Transcription from Active Promoters," *Science (80-.).*, vol. 322, no. 5909, pp. 1849–1851, 2008.

[58] L. J. Core, J. J. Waterfall, and J. T. Lis, "Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters," *Science (80-.).*, vol. 322, no. 5909, pp. 1845–1848, 2008.

[59] S. H. C. Duttke *et al.*, "Human promoters are intrinsically directional," *Mol. Cell*, vol. 57, no. 4, pp. 674–684, 2015.

[60] W. Li, D. Notani, and M. G. Rosenfeld, "Enhancers as non-coding RNA transcription units: recent insights and future perspectives," *Nat. Rev. Genet.*, vol. 17, no. 4, pp. 207–223, 2016.

[61] T. Matsui, J. Segall, P. A. Weil, and R. G. Roeder, "Multiple factors required for accurate initiation of

transcription by purified RNA polymerase II.," *J. Biol. Chem.*, vol. 255, no. 24, pp. 11992–11996, 1980.

[62] S. Sainsbury, C. Bernecky, and P. Cramer, "Structural basis of transcription initiation by RNA polymerase II," *Nat. Rev. Mol. Cell Biol.*, vol. 16, no. 3, pp. 129–143, 2015.

[63] S. Buratowski, S. Hahn, L. Guarente, and P. A. Sharp, "Five intermediate complexes in transcription initiation by RNA polymerase II," *Cell*, vol. 56, no. 4, pp. 549–561, 1989.

[64] R. K. Louder, Y. He, J. R. López-Blanco, J. Fang, P. Chacón, and E. Nogales, "Structure of promoter-bound TFIID and model of human pre-initiation complex assembly," *Nature*, vol. 531, no. 7596, pp. 604–609, 2016.

[65] S. Y. Wu, M. C. Thomas, S. Y. Hou, V. Likhite, and C. M. Chiang, "Isolation of mouse TFIID and functional characterization of TBP and TFIID in mediating estrogen receptor and chromatin transcription," *J. Biol. Chem.*, vol. 274, no. 33, pp. 23480–23490, 1999.

[66] D. S. Luse, "The RNA polymerase II preinitiation complex.," *Transcription*, vol. 5, no. 1, p. e27050, 2014.

[67] M. D. E. Deato and R. Tjian, "An unexpected role of TAFs and TRFs in skeletal muscle differentiation: Switching core promoter complexes," *Cold Spring Harb. Symp. Quant. Biol.*, vol. 73, pp. 217–225, 2008.

[68] F. Müller, A. Zaucker, and L. Tora, "Developmental regulation of transcription initiation: More than just changing the actors," *Curr. Opin. Genet. Dev.*, vol. 20, no. 5, pp. 533–540, 2010.

[69] P. Cramer, "Structural Molecular Biology—A Personal Reflection on the Occasion of John Kendrew's 100th Birthday," *J. Mol. Biol.*, vol. 429, no. 17, pp. 2603–2610, 2017.

[70] J. S??nchez and R. Olivares, "Problem solving and collaboration using mobile serious games," *Comput. Educ.*, vol. 57, no. 3, pp. 1943–1952, 2011.

[71] T.-K. Kim *et al.*, "Widespread transcription at neuronal activity-regulated enhancers," *Nature*, vol. 465, no. 7295, pp. 182–187, 2010.

[72] J. F. Leblanc, L. Cohen, M. Rodrigues, and J. Hiscott, "Synergism between distinct enhanson domains in viral induction of the human beta interferon gene.," *Mol. Cell. Biol.*, vol. 10, no. 8, pp. 3987–93, 1990.

[73] D. Thanos and T. Maniatis, "Virus induction of human IFN?? gene expression requires the assembly of an enhanceosome," *Cell*, vol. 83, no. 7, pp. 1091–1100, 1995.

[74] Y. Pan and R. Nussinov, "The role of response elements organization in transcription factor selectivity: The ifn-β enhanceosome example," *PLoS Comput. Biol.*, vol. 7, no. 6, 2011.

[75] R. Chatterjee *et al.*, "Overlapping ETS and CRE Motifs (^{G} / $_{C}$ CGGAAGTGACGTCA) Preferentially Bound by GABPα and CREB Proteins," *G3: Genes|Genomes|Genetics*, vol. 2, no. 10, pp. 1243–1256, 2012.

[76] G. Wu and H. R. Schöler, "Role of Oct4 in the early embryo development," *Cell Regen.*, vol. 3, no. 1, p. 3:7, 2014.

[77] F. Rastinejad, T. Perlmann, R. M. Evans, and P. B. Sigler, "Structural determinants of nuclear receptor assembly on DNA direct repeats.," *Nature*, vol. 375, no. 6528, pp. 203–211, 1995.

[78] D. J. Mangelsdorf and R. M. Evans, "The RXR heterodimers and orphan receptors," *Cell*, vol. 83, no. 6, pp. 841–850, 1995.

[79] M. Hassler and T. J. Richmond, "The B-box dominates SAP-1-SRF interactions in the structure of the ternary complex," *EMBO J.*, vol. 20, no. 12, pp. 3018–3028, 2001.

[80] D. C. Ambrosetti, C. Basilico, and L. Dailey, "Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites.," *Mol. Cell. Biol.*, vol. 17, no. 11, pp. 6321–9, 1997.

[81] A. Reményi, K. Lins, L. J. Nissen, R. Reinbold, H. R. Schöler, and M. Wilmanns, "Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers," *Genes Dev.*, vol. 17, no. 16, pp. 2048–2059, 2003.

[82] L. Chen, J. N. Glover, P. G. Hogan, A. Rao, and S. C. Harrison, "Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA.," *Nature*, vol. 392, no. 6671, pp. 42–48, 1998.

[83] M. J. Holland, "Transcript abundance in yeast varies over six orders of magnitude," *J. Biol. Chem.*, vol. 277, no. 17, pp. 14363–14366, 2002.

[84] M. G. Carter *et al.*, "Transcript copy number estimation using a mouse whole-genome oligonucleotide microarray.," *Genome Biol.*, vol. 6, no. 7, p. R61, 2005.

[85] S. Islam *et al.*, "Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq," *Genome Res.*, vol. 21, no. 7, pp. 1160–1167, 2011.

[86] D. Zenklusen, D. R. Larson, and R. H. Singer, "Single-RNA counting reveals alternative modes of gene expression in yeast," *Nat. Struct. Mol. Biol.*, vol. 15, no. 12, pp. 1263–1271, 2008.

[87] Z. Wunderlich and L. A. Mirny, "Different gene regulation strategies revealed by analysis of binding motifs," *Trends Genet.*, vol. 25, no. 10, pp. 434–440, 2009.

[88] P. Stegmaier, A. E. Kel, and E. Wingender, "Systematic DNA-binding domain classification of transcription factors.," *Genome Inform.*, vol. 15, no. 2, pp. 276–286, 2004.

[89] N. C. Seeman, J. M. Rosenberg, and A. Rich, "Sequence-specific recognition of double helical nucleic acids by proteins.," *Proc. Natl. Acad. Sci.*, vol. 73, no. 3, pp. 804–808, 1976.

[90] M. Slattery, T. Zhou, L. Yang, A. C. Dantas Machado, R. Gordân, and R. Rohs, "Absence of a simple code: How transcription factors read the genome," *Trends Biochem. Sci.*, vol. 39, no. 9, pp. 381–399, 2014.

[91] I. Dror, R. Rohs, and Y. Mandel-Gutfreund, "How motif environment influences transcription factor search dynamics: Finding a needle in a haystack," *BioEssays*, vol. 38, no. 7, pp. 605–612, 2016.

[92] a a Travers, "DNA conformation and protein binding.," *Annu. Rev. Biochem.*, vol. 58, pp. 427–452, 1989.

[93] S. C. J. Parker, L. Hansen, H. O. Abaan, T. D. Tullius, and E. H. Margulies, "Local DNA Topography Correlates with Functional Noncoding Regions of the Human Genome," *Science (80-.).*, vol. 324, no. 5925, pp. 389–392, 2009.

[94] R. Rohs, X. Jin, S. M. West, R. Joshi, B. Honig, and R. S. Mann, "Origins of Specificity in Protein-DNA Recognition," *Annu. Rev. Biochem.*, vol. 79, no. 1, pp. 233–269, 2010.

[95] J. Yie, M. Merika, N. Munshi, G. Chen, and D. Thanos, "The role of HMG I(Y) in the assembly and function of the IFN-?? enhanceosome," *EMBO J.*, vol. 18, no. 11, pp. 3074–3089, 1999.

[96] E. Morgunova and J. Taipale, "Structural perspective of cooperative transcription factor binding," *Curr. Opin. Struct. Biol.*, vol. 47, pp. 1–8, 2017.

[97] D. Panne, T. Maniatis, and S. C. Harrison, "An Atomic Model of the Interferon-?? Enhanceosome," *Cell*, vol. 129, no. 6, pp. 1111–1123, 2007.

[98] A. Jolma *et al.*, "DNA-dependent formation of transcription factor pairs alters their binding specificity," *Nature*, vol. 527, no. 7578, pp. 384–388, 2015.

[99] R. a Veitia, "A sigmoidal transcriptional response: cooperativity, synergy and dosage effects.," *Biol. Rev. Camb. Philos. Soc.*, vol. 78, no. 1, pp. 149–170, 2003.

[100] D. Lebrecht *et al.*, "Bicoid cooperative DNA binding is critical for embryonic patterning in Drosophila.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 37, pp. 13176–81, 2005.

[101] R. P. Smith *et al.*, "Massively parallel decoding of mammalian regulatory sequences supports a flexible

organizational model," *Nat. Genet.*, vol. 45, no. 9, pp. 1021–1028, 2013.

[102] D. M. Holloway *et al.*, "Gene expression noise in spatial patterning: Hunchback promoter structure affects noise amplitude and distribution in Drosophila segmentation," *PLoS Comput. Biol.*, vol. 7, no. 2, 2011.

[103] a M. Kringstein, F. M. Rossi, a Hofmann, and H. M. Blau, "Graded transcriptional response to different concentrations of a single transactivator.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 95, no. 23, pp. 13670–13675, 1998.

[104] S. R. Biggar and G. R. Crabtree, "Cell signaling can direct either binary or graded transcriptional responses," *EMBO J.*, vol. 20, no. 12, pp. 3167–3176, 2001.

[105] L. Giorgetti *et al.*, "Noncooperative Interactions between Transcription Factors and Clustered DNA Binding Sites Enable Graded Transcriptional Responses to Environmental Inputs," *Mol. Cell*, vol. 37, no. 3, pp. 418–428, 2010.

[106] R. Hancock, "A role for macromolecular crowding effects in the assembly and function of compartments in the nucleus," *J. Struct. Biol.*, vol. 146, no. 3, pp. 281–290, 2004.

[107] J. Rouquette, C. Genoud, G. H. Vazquez-Nin, B. Kraus, T. Cremer, and S. Fakan, "Revealing the high-resolution three-dimensional network of chromatin and interchromatin space: A novel electron-microscopic approach to reconstructing nuclear architecture," *Chromosom. Res.*, vol. 17, no. 6, pp. 801–810, 2009.

[108] R. D. Phair *et al.*, "Global Nature of Dynamic Protein-Chromatin Interactions In Vivo: Three-Dimensional Genome Scanning and Dynamic Interaction Networks of Chromatin Proteins," *Mol. Cell. Biol.*, vol. 24, no. 14, pp. 6393–6402, 2004.

[109] N. Kawamata, M. A. Pennella, J. L. Woo, A. J. Berk, and H. P. Koeffler, "Dominant-negative mechanism of leukemogenic PAX5 fusions," *Oncogene*, vol. 31, no. 8, pp. 966–977, 2012.

[110] D. Mazza, A. Abernathy, N. Golob, T. Morisaki, and J. G. McNally, "A benchmark for chromatin binding measurements in live cells," *Nucleic Acids Res.*, vol. 40, no. 15, 2012.

[111] Y. Sun, N. M. Hays, A. Periasamy, M. W. Davidson, and R. N. Day, "Monitoring protein interactions in living cells with fluorescence lifetime imaging microscopy," *Methods Enzymol.*, vol. 504, pp. 371–391, 2012.

[112] G. Kaur *et al.*, "Probing transcription factor diffusion dynamics in the living mammalian embryo with photoactivatable fluorescence correlation spectroscopy," *Nat. Commun.*, vol. 4, p. 1637, 2013.

[113] J. C. M. Gebhardt *et al.*, "Single-molecule imaging of transcription factor binding to DNA in live mammalian cells," *Nat. Methods*, vol. 10, no. 5, pp. 421–426, 2013.

[114] J. Chen *et al.*, "Single-molecule dynamics of enhanceosome assembly in embryonic stem cells," *Cell*, vol. 156, no. 6, pp. 1274–1285, 2014.

[115] J. G. McNally, "The Glucocorticoid Receptor: Rapid Exchange with Regulatory Sites in Living Cells," *Science (80-.).*, vol. 287, no. 5456, pp. 1262–1265, 2000.

[116] G. L. Hager, J. G. McNally, and T. Misteli, "Transcription Dynamics," *Mol. Cell*, vol. 35, no. 6, pp. 741–753, 2009.

[117] J. Yao, K. M. Munson, W. W. Webb, and J. T. Lis, "Dynamics of heat shock factor association with native gene loci in living cells," *Nature*, vol. 442, no. 7106, pp. 1050–1053, 2006.

[118] V. Paakinaho *et al.*, "Single-molecule analysis of steroid receptor and cofactor action in living cells," *Nat. Commun.*, vol. 8, p. 15896, 2017.

[119] T. Morisaki, W. G. Müller, N. Golob, D. Mazza, and J. G. McNally, "Single-molecule analysis of transcription factor binding at transcription sites in live cells," *Nat. Commun.*, vol. 5, 2014.

[120] Y. M. Wang, R. H. Austin, and E. C. Cox, "Single molecule measurements of repressor protein 1D diffusion

on DNA," *Phys. Rev. Lett.*, vol. 97, no. 4, 2006.

[121] S. E. Halford and J. F. Marko, "How do site-specific DNA-binding proteins find their targets?," *Nucleic Acids Res.*, vol. 32, no. 10, pp. 3040–3052, 2004.

[122] D. M. Gowers, G. G. Wilson, and S. E. Halford, "Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 44, pp. 15883–8, 2005.

[123] A. Jeltsch, C. Wenz, F. Stahl, and A. Pingoud, "Linear diffusion of the restriction endonuclease EcoRV on DNA is essential for the in vivo function of the enzyme," *EMBO J.*, vol. 15, no. 18, pp. 5104–5111, 1996.

[124] K. McKinney, M. Mattia, V. Gottifredi, and C. Prives, "p53 linear diffusion along DNA requires its C terminus," *Mol. Cell*, vol. 16, no. 3, pp. 413–424, 2004.

[125] A. Esadze, C. A. Kemme, A. B. Kolomeisky, and J. Iwahara, "Positive and negative impacts of nonspecific sites during target location by a sequence-specific DNA-binding protein: Origin of the optimal search at physiological ionic strength," *Nucleic Acids Res.*, vol. 42, no. 11, pp. 7039–7046, 2014.

[126] Y. Takayama and G. M. Clore, "Interplay between minor and major groove-binding transcription factors Sox2 and Oct1 in translocation on DNA studied by paramagnetic and diamagnetic NMR," *J. Biol. Chem.*, vol. 287, no. 18, pp. 14349–14363, 2012.

[127] A. Murata *et al.*, "One-Dimensional Search Dynamics of Tumor Suppressor p53 Regulated by a Disordered C-Terminal Domain," *Biophys. J.*, vol. 112, no. 11, pp. 2301–2314, 2017.

[128] J. Gorman and E. C. Greene, "Visualizing one-dimensional diffusion of proteins along DNA.," *Nat. Struct. Mol. Biol.*, vol. 15, no. 8, pp. 768–774, 2008.

[129] Y. Takayama and G. M. Clore, "Intra- and intermolecular translocation of the bi-domain transcription factor Oct1 characterized by liquid crystal and paramagnetic NMR.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. 22, pp. E169-76, 2011.

[130] T. Terakawa, H. Kenzaki, and S. Takada, "P53 searches on DNA by rotation-uncoupled sliding at C-terminal tails and restricted hopping of core domains," *J. Am. Chem. Soc.*, vol. 134, no. 35, pp. 14555–14562, 2012.

[131] S. E. Halford, "An end to 40 years of mistakes in DNA–protein association kinetics?," *Biochem. Soc. Trans.*, vol. 37, no. 2, pp. 343–348, 2009.

[132] A. Murata *et al.*, "One-Dimensional Sliding of p53 Along DNA Is Accelerated in the Presence of Ca^{2+} or Mg^{2+} at Millimolar Concentrations," *J. Mol. Biol.*, vol. 427, no. 16, pp. 2663–2678, 2015.

[133] A. K. Nagaich, D. A. Walker, R. Wolford, and G. L. Hager, "Rapid periodic binding and displacement of the glucocorticoid receptor during chromatin remodeling," *Mol. Cell*, vol. 14, no. 2, pp. 163–174, 2004.

[134] B. C. Freeman *et al.*, "Disassembly of transcriptional regulatory complexes by molecular chaperones.," *Science*, vol. 296, no. 5576, pp. 2232–5, 2002.

[135] D. a Stavreva, W. G. Müller, G. L. Hager, C. L. Smith, and J. G. McNally, "Rapid glucocorticoid receptor exchange at a promoter is coupled to transcription and regulated by chaperones and proteasomes.," *Mol. Cell. Biol.*, vol. 24, no. 7, pp. 2682–2697, 2004.

[136] a Dorn, B. Durand, C. Marfing, M. Le Meur, C. Benoist, and D. Mathis, "Conserved major histocompatibility complex class II boxes--X and Y--are transcriptional control elements and specifically bind nuclear proteins.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 84, no. 17, pp. 6249–6253, 1987.

[137] C. Benoist, K. O'Hare, R. Breathnach, and P. Chambon, "The ovalbumin gene-sequence of putative control regions.," *Nucleic Acids Res.*, vol. 8, no. 1, pp. 127–42, 1980.

[138] A. Efstratiadis, "The structure and evolution of the human β-globin gene family," *Cell*, vol. 21, no. 3, pp. 653–668, 1980.

[139] A. Dorn, J. Bollekens, A. Staub, C. Benoist, and D. Mathis, "A multiplicity of CCAAT box-binding proteins," *Cell*, vol. 50, no. 6, pp. 863–872, 1987.

[140] R. Mantovani, "A survey of 178 NF-Y binding CCAAT boxes," *Nucleic Acids Res.*, vol. 26, no. 5, pp. 1135–1143, 1998.

[141] D. Dolfini, F. Zambelli, G. Pavesi, and R. Mantovani, "A perspective of promoter architecture from the CCAAT box," *Cell Cycle*, vol. 8, no. 24, pp. 4127–4137, 2009.

[142] J. D. Fleming, G. Pavesi, P. Benatti, C. Imbriano, R. Mantovani, and 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.*, vol. 23, no. 8, pp. 1195–1209, 2013.

[143] A. Mathelier *et al.*, "JASPAR 2016: A major expansion and update of the open-access database of transcription factor binding profiles," *Nucleic Acids Res.*, vol. 44, no. D1, pp. D110–D115, 2016.

[144] L. Mariño-Ramírez, J. L. Spouge, G. C. Kanga, and D. Landsman, "Statistical analysis of over-represented words in human promoter sequences," *Nucleic Acids Res.*, vol. 32, no. 3, pp. 949–958, 2004.

[145] C. Wakano, J. S. Byun, L. J. Di, and K. Gardner, "The dual lives of bidirectional promoters," *Biochim. Biophys. Acta - Gene Regul. Mech.*, vol. 1819, no. 7, pp. 688–693, 2012.

[146] M. Ceribelli *et al.*, "The Histone-Like NF-Y Is a Bifunctional Transcription Factor," *Mol. Cell. Biol.*, vol. 28, no. 6, pp. 2047–2058, 2008.

[147] F. Zambelli and G. Pavesi, "Genome wide features, distribution and correlations of NF-Y binding sites," *Biochim. Biophys. Acta - Gene Regul. Mech.*, vol. 1860, no. 5, pp. 581–589, 2017.

[148] V. Salsi *et al.*, "Interactions between p300 and multiple NF-Y trimers govern cyclin B2 promoter function," *J. Biol. Chem.*, vol. 278, no. 9, pp. 6642–6650, 2003.

[149] R. Mantovani, "The molecular biology of the CCAAT-binding factor NF-Y," *Gene*, vol. 239, no. 1, pp. 15–27, 1999.

[150] M. Ceribelli, P. Benatti, C. Imbriano, and R. Mantovani, "NF-YC complexity is generated by dual promoters and alternative splicing," *J. Biol. Chem.*, vol. 284, no. 49, pp. 34189–34200, 2009.

[151] X. Y. Li, R. Hooft Van Huijsduijnen, R. Mantovani, C. Benoist, and D. Mathis, "Intron-exon organization of the NF-Y genes: Tissue-specific splicing modifies an activation domain," *J. Biol. Chem.*, vol. 267, no. 13, pp. 8984–8990, 1992.

[152] J. Zhu, Y. Zhang, G. J. Joe, R. Pompetti, and S. G. Emerson, "NF-Ya activates multiple hematopoietic stem cell (HSC) regulatory genes and promotes HSC self-renewal.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 33, pp. 11728–11733, 2005.

[153] M. Grskovic, C. Chaivorapol, A. Gaspar-Maia, H. Li, and M. Ramalho-Santos, "Systematic identification of cis-regulatory sequences active in mouse and human embryonic stem cells," *PLoS Genet.*, vol. 3, no. 8, pp. 1524–1540, 2007.

[154] D. Dolfini, M. Minuzzo, G. Pavesi, and R. Mantovani, "The short isoform of NF-YA belongs to the embryonic stem cell transcription factor circuitry," *Stem Cells*, vol. 30, no. 11, pp. 2450–2459, 2012.

[155] S. L. Forsburg and L. Guarente, "Mutational analysis of upstream activation sequence 2 of the CYC1 gene of Saccharomyces cerevisiae: a HAP2-HAP3-responsive site.," *Mol. Cell. Biol.*, vol. 8, no. 2, pp. 647–54, 1988.

[156] N. Siefers, K. K. Dang, R. W. Kumimoto, W. E. Bynum, G. Tayrose, and B. F. Holt, "Tissue-Specific Expression Patterns of Arabidopsis NF-Y Transcription Factors Suggest Potential for Extensive

Combinatorial Complexity," *Plant Physiol.*, vol. 149, no. 2, pp. 625–641, 2008.

[157] S. Swain, Z. A. Myers, C. L. Siriwardana, and B. F. Holt, "The multifaceted roles of NUCLEAR FACTOR-Y in Arabidopsis thaliana development and stress responses," *Biochim. Biophys. Acta - Gene Regul. Mech.*, vol. 1860, no. 5, pp. 636–644, 2017.

[158] M. E. Zanetti, C. Rípodas, and A. Niebel, "Plant NF-Y transcription factors: Key players in plant-microbe interactions, root development and adaptation to stress," *Biochim. Biophys. Acta - Gene Regul. Mech.*, vol. 1860, no. 5, pp. 645–654, 2017.

[159] V. Brambilla and F. Fornara, "Y flowering? Regulation and activity of CONSTANS and CCT-domain proteins in Arabidopsis and crop species," *Biochim. Biophys. Acta - Gene Regul. Mech.*, vol. 1860, no. 5, pp. 655–660, 2017.

[160] M. Yamaguchi, M. S. Ali, Y. Yoshioka, L. L. Ly, and H. Yoshida, "NF-Y in invertebrates," *Biochim. Biophys. Acta - Gene Regul. Mech.*, vol. 1860, no. 5, pp. 630–635, 2017.

[161] A. Bhattacharya, J. M. Deng, Z. Zhang, R. Behringer, B. De Crombrugghe, and 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.*, vol. 63, no. 23, pp. 8167–8172, 2003.

[162] T. Yamanaka *et al.*, "NF-Y inactivation causes atypical neurodegeneration characterized by ubiquitin and p62 accumulation and endoplasmic reticulum disorganization.," *Nat. Commun.*, vol. 5, p. 3354, 2014.

[163] R. Luo, S. A. Klumpp, M. J. Finegold, and S. N. Maity, "Inactivation of CBF/NF-Y in postnatal liver causes hepatocellular degeneration, lipid deposition and endoplasmic reticulum stress," *Sci. Rep.*, vol. 1, no. 1, p. 136, 2011.

[164] Y. H. Lu, O. S. Dallner, K. Birsoy, G. Fayzikhodjaeva, and J. M. Friedman, "Nuclear Factor-Y is an adipogenic factor that regulates leptin gene expression," *Mol. Metab.*, vol. 4, no. 5, pp. 392–405, 2015.

[165] G. Bungartz, H. Land, D. T. Scadden, and S. G. Emerson, "NF-Y is necessary for hematopoietic stem cell proliferation and survival," *Blood*, vol. 119, no. 6, pp. 1380–1389, 2012.

[166] G. Marziali *et al.*, "The activity of the CCAAT-box binding factor NF-Y is modulated through the regulated expression of its A subunit during monocyte to macrophage differentiation: regulation of tissue-specific genes through a ubiquitous transcription factor," *Blood*, vol. 93, no. 2, pp. 519–526, 1999.

[167] A. Farina *et al.*, "Down-regulation of cyclin B1 gene transcription in terminally differentiated skeletal muscle cells is associated with loss of functional CCAAT-binding NF-Y complex," *Oncogene*, vol. 18, no. 18, pp. 2818–2827, 1999.

[168] A. Gurtner *et al.*, "Requirement for down-regulation of the CCAAT-binding activity of the NF-Y transcription factor during skeletal muscle differentiation.," *Mol. Biol. Cell*, vol. 14, no. 7, pp. 2706–15, 2003.

[169] S. N. Maity, "NF-Y (CBF) regulation in specific cell types and mouse models," *Biochim. Biophys. Acta - Gene Regul. Mech.*, vol. 1860, no. 5, pp. 598–603, 2017.

[170] S. N. Maity and B. De Crombrugghe, "Role of the CCAAT-binding protein CBF/NF-Y in transcription," *Trends Biochem. Sci.*, vol. 23, no. 5, pp. 174–178, 1998.

[171] I. S. Kim, S. Sinha, B. de Crombrugghe, and 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.*, vol. 16, no. 8, pp. 4003–4013, 1996.

[172] S. Fribourg, C. Romier, S. Werten, Y. G. Gangloff, a Poterszman, and D. Moras, "Dissecting the interaction network of multiprotein complexes by pairwise coexpression of subunits in E. coli.," *J. Mol. Biol.*, vol. 306,

no. 2, pp. 363–373, 2001.

[173] C. Romier, F. Cocchiarella, R. Mantovani, and D. Moras, "The NF-YB/NF-YC structure gives insight into DNA binding and transcription regulation by CCAAT factor NF-Y," *J. Biol. Chem.*, vol. 278, no. 2, pp. 1336–1345, 2003.

[174] M. Nardini *et al.*, "Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination," *Cell*, vol. 152, no. 1–2, pp. 132–143, 2013.

[175] E. M. Huber, D. H. Scharf, P. Hortschansky, M. Groll, and A. A. Brakhage, "DNA minor groove sensing and widening by the ccaat-binding complex," *Structure*, vol. 20, no. 10, pp. 1757–1768, 2012.

[176] V. Nardone, A. Chaves-Sanjuan, and M. Nardini, "Structural determinants for NF-Y/DNA interaction at the CCAAT box," *Biochim. Biophys. Acta - Gene Regul. Mech.*, vol. 1860, no. 5, pp. 571–580, 2017.

[177] M. Bellorini *et al.*, "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.*, vol. 25, no. 11, pp. 2174–2181, 1997.

[178] C. Imbriano *et al.*, "Direct p53 transcriptional repression: in vivo analysis of CCAAT-containing G2/M promoters.," *Mol. Cell. Biol.*, vol. 25, no. 9, pp. 3737–51, 2005.

[179] T. Taira, M. Sawai, M. Ikeda, K. Tamai, S. M. M. Iguchi-Ariga, and H. Ariga, "Cell cycle-dependent switch of up- and down-regulation of human hsp70 gene expression by interaction between c-Myc and CBF/NF-Y," *J. Biol. Chem.*, vol. 274, no. 34, pp. 24270–24279, 1999.

[180] C. Liberati, a di Silvio, S. Ottolenghi, and R. Mantovani, "NF-Y binding to twin CCAAT boxes: role of Q-rich domains and histone fold helices.," *J. Mol. Biol.*, vol. 285, no. 4, pp. 1441–55, 1999.

[181] F. Coustry, S. N. Maity, S. Sinha, and 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.*, vol. 271, no. 24, pp. 14485–14491, 1996.

[182] H. R. Drew and A. A. Travers, "DNA bending and its relation to nucleosome positioning," *J. Mol. Biol.*, vol. 186, no. 4, pp. 773–790, 1985.

[183] C. Jiang and B. F. Pugh, "Nucleosome positioning and gene regulation: advances through genomics," *Nat. Rev. Genet.*, vol. 10, no. 3, pp. 161–172, 2009.

[184] L. Magnani, J. Eeckhoute, and M. Lupien, "Pioneer factors: Directing transcriptional regulators within the chromatin environment," *Trends Genet.*, vol. 27, no. 11, pp. 465–474, 2011.

[185] S. A. Morris, "Direct lineage reprogramming via pioneer factors; a detour through developmental gene regulatory networks," *Development*, vol. 143, no. 15, pp. 2696–2705, 2016.

[186] R. I. Sherwood *et al.*, "Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape," *Nat. Biotechnol.*, vol. 32, no. 2, pp. 171–178, 2014.

[187] A. J. Oldfield *et al.*, "Histone-Fold Domain Protein NF-Y Promotes Chromatin Accessibility for Cell Type-Specific Master Transcription Factors," *Mol. Cell*, vol. 55, no. 5, pp. 708–722, 2014.

[188] F. Lu, Y. Liu, A. Inoue, T. Suzuki, K. Zhao, and Y. Zhang, "Establishing chromatin regulatory landscape during mouse preimplantation development," *Cell*, vol. 165, no. 6, pp. 1375–1388, 2016.

[189] T. Sekiya, U. M. Muthurajan, K. Luger, A. V. Tulin, and K. S. Zaret, "Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA," *Genes Dev.*, vol. 23, no. 7, pp. 804–809, 2009.

[190] A. Soufi, M. F. Garcia, A. Jaroszewicz, N. Osman, M. Pellegrini, and K. S. Zaret, "Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming," *Cell*, vol. 161, no. 3, pp. 555–

568, 2015.

[191] M. C. Motta, G. Caretti, G. F. Badaracco, and R. Mantovani, "Interactions of the CCAAT-binding trimer NF-Y with nucleosomes," *J. Biol. Chem.*, vol. 274, no. 3, pp. 1326–1333, 1999.

[192] G. Caretti, M. C. Motta, and R. Mantovani, "NF-Y associates with H3-H4 tetramers and octamers by multiple mechanisms.," *Mol. Cell. Biol.*, vol. 19, no. 12, pp. 8591–8603, 1999.

[193] M. Frontini *et al.*, "NF-Y recruitment of TFIID, multiple interactions with histone fold TAFIIs," *J. Biol. Chem.*, vol. 277, no. 8, pp. 5841–5848, 2002.

[194] G. Caretti *et al.*, "Dissection of functional NF-Y-RFX cooperative interactions on the MHC class II Ea promoter.," *J. Mol. Biol.*, vol. 302, no. 3, pp. 539–52, 2000.

[195] K. Roder, S. S. Wolf, K. F. Beck, and M. Schweizer, "Cooperative binding of NF-Y and Sp1 at the DNase I-hypersensitive site, fatty acid synthase insulin-responsive element 1, located at -500 in the rat fatty acid synthase promoter," *J. Biol. Chem.*, vol. 272, no. 34, pp. 21616–21624, 1997.

[196] G. Suske, "NF-Y and SP transcription factors — New insights in a long-standing liaison," *Biochim. Biophys. Acta - Gene Regul. Mech.*, vol. 1860, no. 5, pp. 590–597, 2017.

[197] Q.-S. Zhu, B. Qian, and D. Levy, "CCAAT/enhancer-binding protein alpha (C/EBPalpha) activates transcription of the human microsomal epoxide hydrolase gene (EPHX1) through the interaction with DNA-bound NF-Y.," *J. Biol. Chem.*, vol. 279, no. 29, pp. 29902–10, 2004.

[198] H. Yoshida *et al.*, "ATF6 activated by proteolysis binds in the presence of NF-Y (CBF) directly to the cis-acting element responsible for the mammalian unfolded protein response.," *Mol. Cell. Biol.*, vol. 20, no. 18, pp. 6755–67, 2000.

[199] J. Zhu, D. M. Giannola, Y. Zhang, A. J. Rivera, and S. G. Emerson, "NF-Y cooperates with USF1/2 to induce the hematopoietic expression of HOXB4," *Blood*, vol. 102, no. 7, pp. 2420–2427, 2003.

[200] M. Haubrock, F. Hartmann, and E. Wingender, "NF-Y binding site architecture defines a C-Fos targeted promoter class," *PLoS One*, vol. 11, no. 8, 2016.

[201] P. Kheradpour and M. Kellis, "Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments," *Nucleic Acids Res.*, vol. 42, no. 5, pp. 2976–2987, 2014.

[202] S. Völkel, B. Stielow, F. Finkernagel, T. Stiewe, A. Nist, and G. Suske, "Zinc Finger Independent Genome-Wide Binding of Sp2 Potentiates Recruitment of Histone-Fold Protein Nf-y Distinguishing It from Sp1 and Sp3," *PLoS Genet.*, vol. 11, no. 3, 2015.

[203] Z. Shi *et al.*, "Context-specific role of SOX9 in NF-Y mediated gene regulation in colorectal cancer cells," *Nucleic Acids Res.*, vol. 43, no. 13, pp. 6257–6269, 2015.

[204] J. Wang *et al.*, "Factorbook.org: A Wiki-based database for transcription factor-binding data generated by the ENCODE consortium," *Nucleic Acids Res.*, vol. 41, no. D1, 2013.

[205] D. Xie, A. P. Boyle, L. Wu, J. Zhai, T. Kawli, and M. Snyder, "XDynamic trans-acting factor colocalization in human cells," *Cell*, vol. 155, no. 3, 2013.

[206] J. Wang *et al.*, "Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors," *Genome Res.*, vol. 22, no. 9, pp. 1798–1812, 2012.

[207] D. Dolfini, F. Zambelli, M. Pedrazzoli, R. Mantovani, and G. Pavesi, "A high definition look at the NF-Y regulome reveals genome-wide associations with selected transcription factors," *Nucleic Acids Res.*, vol. 44, no. 10, pp. 4684–4702, 2016.

[208] Y. Guo and D. K. Gifford, "Modular combinatorial binding among human trans-acting factors reveals direct and indirect factor binding," *BMC Genomics*, vol. 18, no. 1, p. 45, 2017.

[209] M. Sawadogo and R. G. Roeder, "Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region," *Cell*, vol. 43, no. 1, pp. 165–175, 1985.

[210] B. H. Shieh, R. S. Sparkes, R. B. Gaynor, and A. J. Lusis, "Localization of the gene-encoding upstream stimulatory factor (USF) to human chromosome 1q22-q23.," *Genomics*, vol. 16, no. 1, pp. 266–8, 1993.

[211] P. M. Groenen, E. Garcia, P. Debeer, K. Devriendt, J. P. Fryns, and W. J. Van de Ven, "Structure, sequence, and chromosome 19 localization of human USF2 and its rearrangement in a patient with multicystic renal dysplasia," *Genomics*, vol. 38, no. 2, pp. 141–148, 1996.

[212] A. R. Ferre-D'Amare *et al.*, "Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain," *Nature*, vol. 363, no. 6424, pp. 38–45, 1993.

[213] A. R. Ferré-D'Amaré, P. Pognonec, R. G. Roeder, and S. K. Burley, "Structure and function of the b/HLH/Z domain of USF.," *EMBO J.*, vol. 13, no. 1, pp. 180–9, 1994.

[214] X. Luo and M. Sawadogo, "Functional domains of the transcription factor USF2: atypical nuclear localization signals and context-dependent transcriptional activation domains.," *Mol. Cell. Biol.*, vol. 16, no. 4, pp. 1367–75, 1996.

[215] Y. Qyang *et al.*, "Cell-Type-Dependent Activity of the Ubiquitous Transcription Factor USF in Cellular Proliferation and Transcriptional Activation," *Mol. Cell. Biol.*, vol. 19, no. 2, pp. 1508–1517, 1999.

[216] F. Watt and P. L. Molloy, "Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter.," *Genes Dev.*, vol. 2, no. 9, pp. 1136–1143, 1988.

[217] S. P. Scherrer, D. A. Rice, and L. L. Heckert, "Expression of Steroidogenic Factor 1 in the Testis Requires an Interactive Array of Elements Within Its Proximal Promoter," *Biol. Reprod.*, vol. 67, no. 5, pp. 1509–1521, 2002.

[218] N. Kovacevic Grujicic, M. Mojsin, A. Krstic, and M. Stevanovic, "Functional characterization of the human SOX3 promoter: Identification of transcription factors implicated in basal promoter activity," *Gene*, vol. 344, pp. 287–297, 2005.

[219] Y. Ito, Y. Zhang, S. Dangaria, X. Luan, and T. G. H. Diekwisch, "NF-Y and USF1 transcription factor binding to CCAAT-box and E-box elements activates the CP27 promoter," *Gene*, vol. 473, no. 2, pp. 92–99, 2011.

[220] H. Qiao and J. M. May, "CpG methylation at the USF-binding site mediates cell-specific transcription of human ascorbate transporter SVCT2 exon 1a.," *Biochem. J.*, vol. 440, no. 1, pp. 73–84, 2011.

[221] E. Meccia, L. Bottero, F. Felicetti, C. Peschle, M. P. Colombo, and A. Carè, "HOXB7 expression is regulated by the transcription factors NF-Y, YY1, Sp1 and USF-1," *Biochim. Biophys. Acta - Gene Struct. Expr.*, vol. 1626, no. 1–3, pp. 1–9, 2003.

[222] S. Jin and K. W. Scotto, "Transcriptional regulation of the MDR1 gene by histone acetyltransferase and deacetylase is mediated by NF-Y.," *Mol. Cell. Biol.*, vol. 18, no. 7, pp. 4377–4384, 1998.

[223] R. A. Currie, "NF-Y is associated with the histone acetyltransferases GCN5 and P/CAF," *J. Biol. Chem.*, vol. 273, no. 3, pp. 1430–1434, 1998.

[224] R. Hughes, M. Kristiansen, I. Lassot, S. Desagher, R. Mantovani, and J. Ham, "NF-Y is essential for expression of the proapoptotic bim gene in sympathetic neurons," *Cell Death Differ.*, vol. 18, no. 6, pp. 937–947, 2011.

[225] G. J. M. Leclerc, C. Mou, G. J. M. Leclerc, A. M. Mian, and J. C. Barredo, "Histone deacetylase inhibitors induce FPGS mRNA expression and intracellular accumulation of long-chain methotrexate polyglutamates

in childhood acute lymphoblastic leukemia: implications for combination therapy.," *Leukemia*, vol. 24, no. 3, pp. 552–62, 2010.

[226] T. Kato *et al.*, "Characterization of the HDAC1 complex that regulates the sensitivity of cancer cells to oxidative stress," *Cancer Res.*, vol. 69, no. 8, pp. 3597–3604, 2009.

[227] A. Fossati, D. Dolfini, G. Donati, and R. Mantovani, "NF-Y recruits Ash2L to impart H3K4 trimethylation on CCAAT promoters," *PLoS One*, vol. 6, no. 3, 2011.

[228] V. K. Tiwari, M. B. Stadler, C. Wirbelauer, R. Paro, D. Schübeler, and C. Beisel, "A chromatin-modifying function of JNK during stem cell differentiation," *Nat. Genet.*, vol. 44, no. 1, pp. 94–100, 2011.

[229] J. Göke, Y.-S. Chan, J. Yan, M. Vingron, and H.-H. Ng, "Genome-wide Kinase-Chromatin Interactions Reveal the Regulatory Network of ERK Signaling in Human Embryonic Stem Cells," *Mol. Cell*, vol. 50, no. 6, pp. 844–855, 2013.

[230] G. Donati, R. Gatta, D. Dolfini, A. Fossati, M. Ceribelli, and R. Mantovani, "An NF-Y-dependent switch of positive and negative histone methyl marks on CCAAT promoters," *PLoS One*, vol. 3, no. 4, 2008.

[231] A. Gurtner *et al.*, "NF-Y dependent epigenetic modifications discriminate between proliferating and postmitotic tissue," *PLoS One*, vol. 3, no. 4, 2008.

[232] G. Caretti, V. Salsi, C. Vecchi, C. Imbriano, and R. Mantovani, "Dynamic recruitment of NF-Y and histone acetyltransferases on cell-cycle promoters," *J. Biol. Chem.*, vol. 278, no. 33, pp. 30435–30440, 2003.

[233] Y. Peng *et al.*, "Irradiation modulates association of NF-Y with histone-modifying cofactors PCAF and HDAC," *Oncogene*, vol. 26, no. 54, pp. 7576–7583, 2007.

[234] X. Zhu, Y. Wang, W. Pi, H. Liu, A. Wickrema, and D. Tuan, "NF-Y Recruits Both Transcription Activator and Repressor to Modulate Tissue- and Developmental Stage-Specific Expression of Human ??-Globin Gene," *PLoS One*, vol. 7, no. 10, 2012.

[235] Q. Li *et al.*, "Xenopus NF-Y pre-sets chromatin to potentiate p300 and acetylation-responsive transcription from the Xenopus hsp70 promoter in vivo," *EMBO J.*, vol. 17, no. 21, pp. 6300–6315, 1998.

[236] M. Thön *et al.*, "The CCAAT-binding complex coordinates the oxidative stress response in eukaryotes," *Nucleic Acids Res.*, vol. 38, no. 4, pp. 1098–1113, 2009.

[237] I. Manni *et al.*, "Posttranslational regulation of NF-YA modulates NF-Y transcriptional activity.," *Mol. Biol. Cell*, vol. 19, no. 12, pp. 5203–13, 2008.

[238] J. Yun *et al.*, "Cdk2-dependent phosphorylation of the NF-Y transcription factor and its involvement in the p53-p21 signaling pathway," *J. Biol. Chem.*, vol. 278, no. 38, pp. 36966–36972, 2003.

[239] J. V Harper and G. Brooks, "The mammalian cell cycle: an overview.," *Methods Mol. Biol.*, vol. 296, pp. 113–153, 2005.

[240] F. Bolognese *et al.*, "The cyclin B2 promoter depends on NF-Y, a trimer whose CCAAT-binding activity is cell-cycle regulated.," *Oncogene*, vol. 18, no. 10, pp. 1845–1853, 1999.

[241] Q. Hu and S. N. Maity, "Stable expression of a dominant negative mutant of CCAAT binding factor/NF-Y in mouse fibroblast cells resulting in retardation of cell growth and inhibition of transcription of various cellular genes," *J. Biol. Chem.*, vol. 275, no. 6, pp. 4435–4444, 2000.

[242] K. Matuoka and K. Yu Chen, "Nuclear factor Y (NF-Y) and cellular senescence.," *Exp. Cell Res.*, vol. 253, no. 2, pp. 365–371, 1999.

[243] S. Di Agostino *et al.*, "Gain of function of mutant p53: The mutant p53/NF-Y protein complex reveals an aberrant transcriptional mechanism of cell cycle regulation," *Cancer Cell*, vol. 10, no. 3, pp. 191–202, 2006.

[244] H.-D. Chae, J. Yun, Y.-J. Bang, and D. Y. Shin, "Cdk2-dependent phosphorylation of the NF-Y

transcription factor is essential for the expression of the cell cycle-regulatory genes and cell cycle G1/S and G2/M transitions," *Oncogene*, vol. 23, no. 23, pp. 4084–4088, 2004.

[245] Y.-C. Lin, Y.-N. Chen, K.-F. Lin, F.-F. Wang, T.-Y. Chou, and M.-Y. Chen, "Association of p21 with NF-YA suppresses the expression of Polo-like kinase 1 and prevents mitotic death in response to DNA damage," *Cell Death Dis.*, vol. 5, no. 1, p. e987, 2014.

[246] M. Lorenzo, R. Mantovani, and N. Gnesutta, "THE INTERPLAY AMONG NF-Y AND E-BOX TRANSCRIPTION FACTORS: MAX, MYC AND USF1," University of Milan, 2015.

[247] H. D. Chae, J. Kim, and D. Y. Shin, "NF-Y binds to both G1- and G2-specific cyclin promoters; a possible role in linking CDK2/cyclin a to CDK1/cyclin B," *BMB Rep.*, vol. 44, no. 8, pp. 553–557, 2011.

[248] U.-J. Yun, H.-D. Park, and D. Y. Shin, "p53 prevents immature escaping from cell cycle G2 checkpoint arrest through inhibiting cdk2-dependent NF-Y phosphorylation.," *Cancer Res. Treat.*, vol. 38, no. 4, pp. 224–8, 2006.

[249] A. Bah *et al.*, "Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch," *Nature*, vol. 519, no. 7541, pp. 106–109, 2014.

[250] C. R. Lickwar, F. Mueller, S. E. Hanlon, J. G. McNally, and J. D. Lieb, "Genome-wide protein–DNA binding dynamics suggest a molecular clutch for transcription factor function," *Nature*, vol. 484, no. 7393, pp. 251–255, 2012.

[251] M. D. White *et al.*, "Long-Lived Binding of Sox2 to DNA Predicts Cell Fate in the Four-Cell Mouse Embryo," *Cell*, vol. 165, no. 1, pp. 75–87, 2016.

[252] R. Janknecht, R. a Hipskind, T. Houthaeve, a Nordheim, and H. G. Stunnenberg, "Identification of multiple SRF N-terminal phosphorylation sites affecting DNA binding properties.," *EMBO J.*, vol. 11, no. 3, pp. 1045–1054, 1992.

[253] A. Loffreda *et al.*, "Live-cell p53 single-molecule binding is modulated by C-terminal acetylation and correlates with transcriptional activity," *Nat. Commun.*, vol. 8, no. 1, p. 313, 2017.

[254] M. A. Martínez-Balbás, A. Dey, S. K. Rabindran, K. Ozato, and C. Wu, "Displacement of sequence-specific transcription factors from mitotic chromatin," *Cell*, vol. 83, no. 1, pp. 29–38, 1995.

[255] P. Arampatzi, M. Gialitakis, T. Makatounakis, and J. Papamatheakis, "Gene-specific factors determine mitotic expression and bookmarking via alternate regulatory elements," *Nucleic Acids Res.*, vol. 41, no. 4, pp. 2202–2215, 2013.

[256] A. L. Roy, M. Meisterernst, P. Pognonec, and R. G. Roeder, "Cooperative interaction of an initiator-binding transcription initiation factor and the helix-loop-helix activator USF," *Nature*, vol. 354, no. 6350, pp. 245–248, 1991.

[257] a L. Roy, H. Du, P. D. Gregor, C. D. Novina, E. Martinez, and R. G. Roeder, "Cloning of an Inr- and E-box-binding protein, TFII-I, that interacts physically and functionally with USF1.," *EMBO J.*, vol. 16, no. 23, pp. 7091–104, 1997.

[258] J. Liu, N. B. Perumal, C. J. Oldfield, E. W. Su, V. N. Uversky, and A. K. Dunker, "Intrinsic disorder in transcription factors," *Biochemistry*, vol. 45, no. 22, pp. 6873–6888, 2006.

[259] T. Zor, R. N. De Guzman, H. J. Dyson, and P. E. Wright, "Solution Structure of the KIX Domain of CBP Bound to the Transactivation Domain of c-Myb," *J. Mol. Biol.*, vol. 337, no. 3, pp. 521–534, 2004.

[260] S. A. Dames, M. Martinez-Yamout, R. N. De Guzman, H. J. Dyson, and P. E. Wright, "Structural basis for Hif-1 /CBP recognition in the cellular hypoxic response," *Proc. Natl. Acad. Sci.*, vol. 99, no. 8, pp. 5271–5276, 2002.

[261] R. B. Berlow, H. J. Dyson, and P. E. Wright, "Functional advantages of dynamic protein disorder," *FEBS Lett.*, vol. 589, no. 19, pp. 2433–2440, 2015.

[262] R. Linding, R. B. Russell, V. Neduva, and T. J. Gibson, "GlobPlot: Exploring protein sequences for globularity and disorder," *Nucleic Acids Res.*, vol. 31, no. 13, pp. 3701–3708, 2003.

[263] T. Lu and M. Sawadogo, "Role of the leucine zipper in the kinetics of DNA binding by transcription factor USF.," *J. Biol. Chem.*, 1994.