UNIVERSITÀ DEGLI STUDI DI MILANO

PhD Course in Molecular and Cellular Biology

XXXIII Ciclo

Structural Determinants of NF-Y Interaction with Partner Transcription Factors on DNA

Dana Saad

PhD Thesis

Scientific Tutors: Prof. Nerina Gnesutta & Prof. Roberto Mantovani

Academic year: 2019-2020

Table of Contents

ABSTRACT

Gene expression is a complex set of events coordinated by a plethora of sequence-specific transcription factors (TFs) that bind and recognise regulatory elements of genes in a combinatorial fashion, and interface with chromatin structures. NF-Y is an ubiquitous eukaryotic TF composed by evolutionary conserved subunits (NF-YA, NF-YB and NF-YC), all necessary for specific binding of the CCAAT box, a widespread element occurring in 30% of genes promoters. NF-Y binding is critical in transcriptional regulation, establishing a permissive chromatin conformation and enabling master TF binding. Genome-wide analyses of TF locations *in vivo* unveiled NF-Y partnership with a defined set of TFs, some of them with precise binding-site (TFBS) arrangements with respect to the CCAAT box. Analysis of promoters of cancer "signature" genes pointed toward NF-Y and E2F TFBSs as the most enriched motifs in genes overexpressed in human tumours. Their significant co-localization *in vivo* further suggests NF-Y and E2Fs might functionally interact by building a landing platform for co-activators. With the long-term goal of controlling such TF interactions in cancer, in this thesis we initially focused on E2F structure, and then extended our studies to NF-Y interactions on DNA.

The E2F family of TFs, originally identified with E2F1 as master regulators of cell-cycle progression, bind DNA as heterodimers with the structurally related DP subunits. Their hallmark features consist of a DNA Binding Domain (DBD) connected through a short linker to the Coiled-Coil-Marked Box (CC-MB) dimerization domain, also involved in Rb tumour suppressor proteins interactions. As structural information is only available for the separated domains, we set out to obtain a coherent view of the linked domains on DNA, and produced the ensembles of DBD-CC-MB portions of E2F1/DP1. To structurally characterise the $E2F1^{DCM}/DP1^{DCM}$/DNA complex we applied an integrated approach that combines in-solution (SAXS) analyses with Molecular Dynamics simulations. We provide evidence for a high level of flexibility in the overall relative topology of the heterodimer, including in a newly identified part stabilizing DNA contacts.

Next, we applied EMSA experiments to characterise DNA binding of the CDC2 cell-cycle promoter, shown to be functionally dependent from both E2F and NF-Y. While results obtained indicate non-cooperative DNA association of NF-Y and E2F1, suggesting alternative mechanisms to elicit transcriptional synergy, we observed efficient assembly of stable ternary complexes. Together, our results and tools obtained in this work will be instrumental for future structural studies involving multiprotein complex assembly on DNA, and for defining the molecular bases of TFs interactions involved in architectural organization of gene promoters.

1. INTRODUCTION

Multi-cellular organism begins as a single fertilised cell that repeatedly divides and proliferates in a final pattern of spectacular precision and complexity to produce different cell types with the same genetic template. During development, the genetic information contained within genomic DNA determines the cell identity and function through selective gene expression programs regulated by thousands of transcription factors (TF), chromatin regulators and cofactors. To maintain the integrity of all eukaryotic cells and organisms, gene expression must be finely regulated during different biological processes, including cell differentiation and metabolism, and its mis-regulation causes various human diseases and disorders. The basis of the central dogma of molecular biology explains the biological information flow, whereby the DNA is transcribed into an intermediate RNA and translated into protein. Transcription is the first step in decoding the DNA sequence, in which genetic information from a gene is used to synthesize a functional gene product, namely protein-coding and noncoding RNA species. The gene expression pattern of a specific cell state, also defined as its transcriptome, includes the entire RNA species from genes that are expressed constantly in all cells and genes that are active predominantly in a particular developmental pathway. The transcription is regulated at two interconnected levels: the first involves TFs and the transcription machinery; and the second involves chromatin and its regulators. Human genome contains ~35,000 genetic loci among which more than 2,000 loci are likely to encode DNA-binding regulatory factors. TFs typically regulate gene expression by specifically recognising and binding short DNA sequences within regulatory regions, and recruiting the transcription apparatus to target genes [1]. The cooperation of multiple cis-acting regulatory elements is required to achieve maximal levels of transcription. In complex genomes, transcriptional regulatory elements are key players in the dynamic usage of the genome during responses to external stimuli, development, cell and tissue homeostasis and disease [2].

1.1 The functional basis of transcription

In eukaryotes, the temporal and special expression of ~20,000 protein-coding genes is primarily regulated at the transcriptional level and is central to many biological processes such as differentiation, development and homeostasis [3]. Three structurally and functionally distinct nuclear RNA polymerases (Pol) were identified that independently control the synthesis levels of the major classes of RNAs. While Pol I and Pol III specifically transcribe large ribosomal subunits rRNA and transfer tRNA, respectively, required for protein synthesis, Pol II is responsible for transcribing all protein-encoding and many non-coding genes [4]. Because Pol II-transcribed genes are highly heterogenous with respect to context specificity and expression level, transcriptional control need to be highly dynamic and specialised [5].

Typically, transcription starts at a defined position, the transcription start site (TSS), embedded within a core promoter at the 5' start of genes (**Figure 1**). The core promoter is a short sequence of ~100 bps encompassing several interchangeable sequence elements surrounding the TSS, which assists the assembly of the pre-initiation complex (PIC), consisting of Pol II and its accessory general transcription factors (GTFs) and guides transcription initiation from accurate position. Several core-promoter motifs have been identified with fixed positioning relative to a well-defined TSS, including the conserved TATA box positioned ~30 bps upstream the TSS [6] and the Initiator (Inr) motif which directly overlap the TSS [7]. The Inr motif is not universal but is more abundant than the TATA-box found only in a minority of core promoters [8]. In TATA-less promoters, the Inr element is often accompanied by the downstream promoter element (DPE) located downstream of the TSS [9]. Core promoters possess low basal activity, which can be activated either by TFs and cofactors bound at proximal elements within several hundred base pairs of the TSS or at distal regulatory elements located thousands of base pairs away, termed enhancers [5][10][11]. Independent of their relative distance and orientation, enhancers can increase transcription from a core promoter by binding regulatory proteins and recruiting transcription cofactors [10].

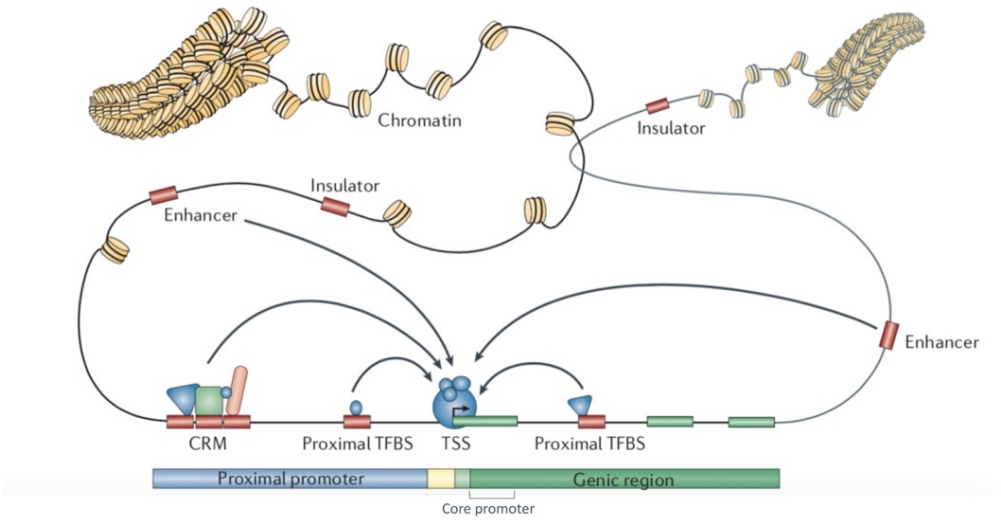


Figure 1. Gene promoter architecture. Nucleosome is the basic unit of chromatin, consisting of a histone core around which DNA is wrapped. Chromatin states (tightly wrapped or accessible) may be marked by insulators. The core promoter is a small region surrounding the TSS, while the larger proximal promoter is present upstream TSS. Sequence-specific TFs bind to specific sites, either in the promoter region (namely, transcription factor binding sites TFBSs) or further away at distal enhancers. Cis-regulator modules (CRMs) are clusters of TFBSs (adapted from [5]).

1.1.1 Nucleosome landscape and its dynamic regulation in human genome

To fit within the small confines of a nucleus, negatively charged DNA polymer is neutralised and compacted through the binding of highly basic histone proteins, forming chromatin. Although wrapping DNA around histones limits access to most genomic DNA, the ubiquity of the histones bound at all chromosomal regions direct enzymes that read, replicate and repair DNA to the appropriate entry sites.

The DNA is wrapped around histone octamers core forming nucleosomes, which constitute the basic unit of chromatin. Each core nucleosome includes approximately 147 bps DNA wrapped 1.65 times around compact protein core composed of two copies of each histone proteins, H2A, H2B, H3 and H4 in a left-handed toroid (**Figure 2A**).

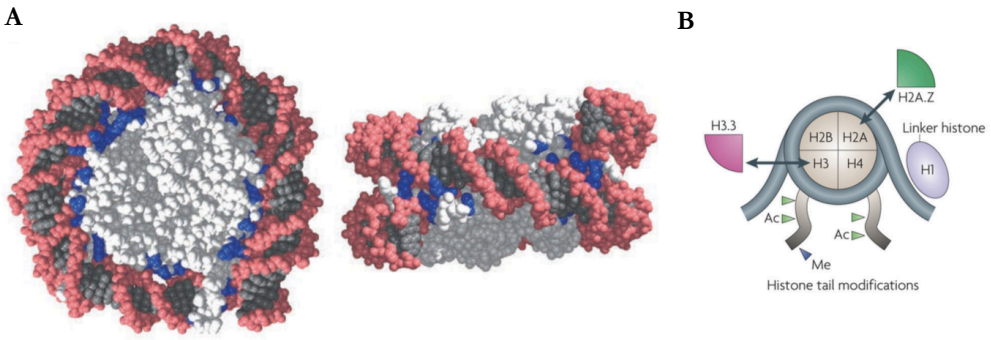


Figure 2. Nucleosome structure. (A) Front and side view of nucleosome core particle structure [12][13]. Histones, DNA helix and backbone are depicted in light grey, dark grey and pink respectively. Electrostatic contacts between the DNA phosphates and histones mediated by basic amino acids (lysine and arginine) are highlighted in blue. **(B)** Schematic representation of DNA wrapped around a nucleosome. Histone tail modifications are indicated (Me, methylation; Ac, acetylation). Histone variants (H2A.Z and H3.3) that replace canonical histones H2A and H3 at active genes are shown (adapted from [14]).

The protein octamer is constituted by two H3/H4 pairs and two H2A/H2B dimers, exposing an extended positively charged interface for DNA binding. Core histone proteins are characterised by a conserved Histone-Fold Domains (HFD) that share a highly similar structural motif composed by three α-helices connected by two loops, forming compact dimers through extensive hydrophobic interactions between HFDs that assemble in a head-to-tail fashion. The N-terminal histone tails that emanate from the nucleosome core are subjected to covalent modifications, including methylation and acetylation (**Figure 2B**) [12][14]. The core nucleosomes are arranged as "beads on a string" along the DNA spaced with short stretches of linker DNA of variable length among species (~38 bps linker in human) [15][16] (**Figure 1**).

Through specific positioning of nucleosome core particles, they can limit DNA accessibility to cellular machinery [17]. Nucleosome positioning (NP) is dynamically influenced by TFs bound to specific DNA sequences, chemical and compositional modifications, as well as chromatin remodelling or spacing complexes. Chromatin remodelling complexes that selectively organise and position nucleosomes are required to regulate DNA accessibility to gene promoters and regulatory regions, influencing gene expression. Since many PIC components, including TFIID and SAGA complex possess nucleosome-binding subunits, specific positioning of nucleosomes might determine the TSS location by positioning the PIC [14]. Many models have been proposed to explain how positioned nucleosome can be disrupted or displaced to regulate DNA accessibility to transcription machinery, and therefore gene expression.

The model proposed by *Widom* and colleagues is based on the intrinsic dynamic structures of nucleosomes that transiently expose DNA-binding sites for transcriptional regulators through thermal fluctuation of DNA on the nucleosome surface (**Figure 3A**) [18][19]. Binding of one regulatory factor to its cognate site might stabilise a partially disassembled state, alleviating the nucleosome-mediated inhibition of other TFs binding to internal sites on the same nucleosome [20]. Transcription might be regulated by ATP-dependent chromatin remodelling complexes and histone modifying enzymes. ATP-dependent remodelling complexes, including SWI/SNF use the energy of ATP hydrolysis to create super-helical torsion into DNA leading to translational repositioning (nucleosome sliding) and exposing DNA regulatory sites [14][21] (**Figure 3B**). In addition to nucleosome sliding, eviction of a nucleosome from a particular genomic location promotes TSS access and gene activation (**Figure 3C**).

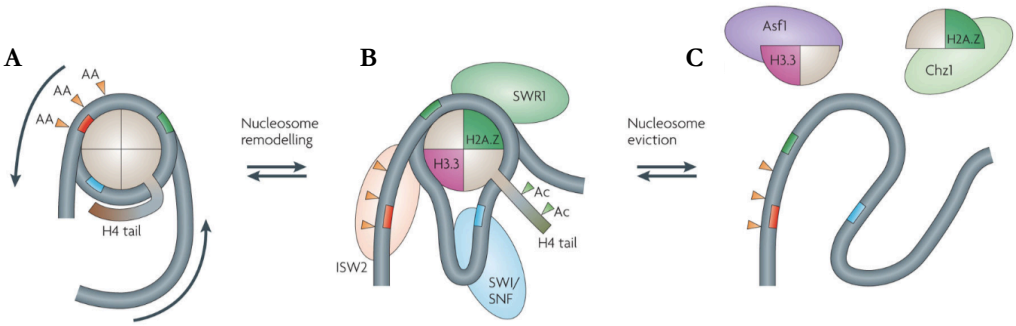


Figure 3. Mechanisms that control DNA accessibility. (A) Nucleosome sliding; **(B)** Nucleosome remodelling; **(C)** Nucleosome eviction. Red, green and blue indicate three TFBSs. The green site is always accessible, while the red and blue sites become accessible during nucleosome remodelling. The nucleosome sliding is indicated by the arrows. Ac, acetylation (adapted from [14]).

1.1.2 Transcription initiation at gene promoters

Transcription from gene core promoters is a stepwise process mainly composed of three phases: initiation, elongation and termination, resulting in a defined transcriptional output. Despite the enormous complexity, RNA polymerases are unable to accurately initiate transcription on defined core promoter elements without their accessory factors. The C-terminal domain (CTD) of Pol II integrates dynamic and regulated interactions with initiation, capping, elongation, chromatin remodelling, termination and mRNA export regulatory factors [22]. Indeed, for individual RNA polymerases, specific general initiation factors were identified, including SL1/TIF-1B for Pol I, TFIIIC and Pol III, and several general transcription factors (GTFs) for Pol II, including TFIID, TFIIA, TFIIB, TFIIE and TFIIH [23][24]. RNA polymerase and these specific initiation factors jointly form a large functional pre-initiation complex (PIC) that assembles at the core promoter. Particularly, the recruitment of TFIID, TFIIA and TFIIB to the core promoter promotes correct loading of the Pol II-TFIIF complex, and subsequent assembly of the remaining cognate factors (TFIIE and TFIIH) into the PIC (**Figure 4A**) [25][26]. After PIC assembly, the melted single-stranded DNA template is inserted into the Pol II-active site forming an open complex,

allowing Pol II to initiate transcription at the TSS. To continue transcribing, Pol II and the associated GTFs are released (promoter escape) from the core promoter through phosphorylation of Pol II-CTD by TFIIH (**Figure 4B**) [27]. After escaping from the TSS, Pol II transcribes a short stretch of nascent RNA (30-50 nucleotides) and then undergoes promoter-proximal pausing downstream of the TSS. Pol II pausing is promoted by the binding of DRB sensitivity inducing factor (DSIF) and negative elongation factor (NELF) to Pol II and nascent RNA. Paused Pol II is rapidly released into productive elongation by cyclin-dependent kinase 9 (CDK9), subunit of the positive transcription elongation factor b (P-TEFb), that phosphorylates NELF, DSIF and Pol II CTD (**Figure 4C**) [28]–[30].

Active promoter is in spatial proximity to transcriptional enhancer engaged in physical contacts with their target-gene promoters, often bridging considerable distances. Enhancers promote PIC assembly, promoter escape and Pol II promoter-proximal pausing by recruiting TFs and cofactors (COFs) such as Mediator complex (MED) that bridges enhancer-bound regulatory factors and the general transcription machinery bound at the core promoter (**Figure 4**; right panel) [31].

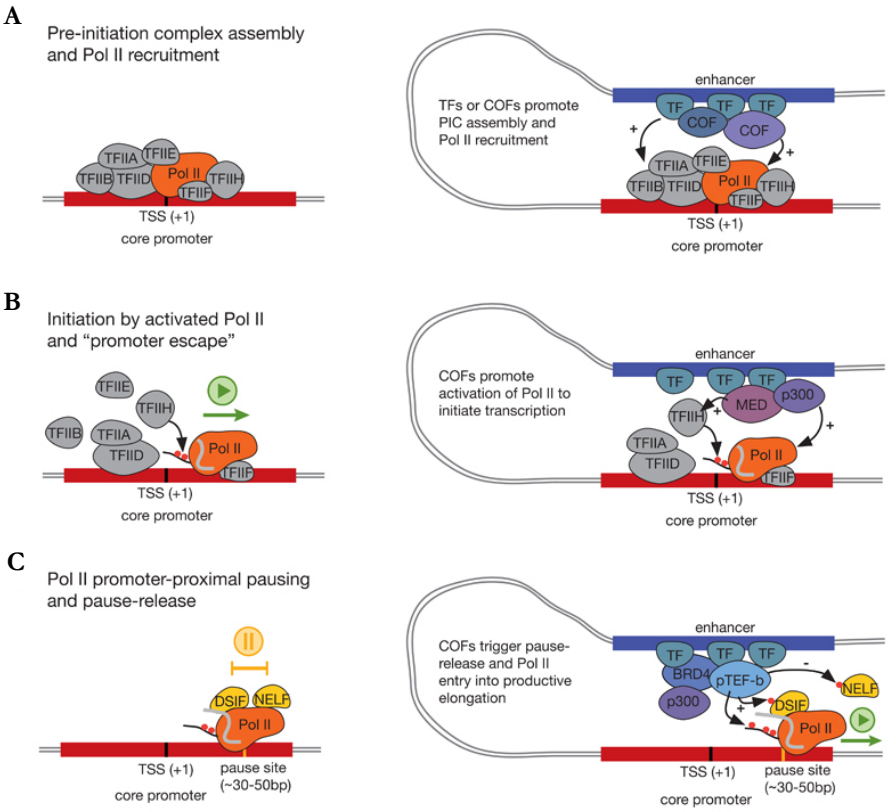


Figure 4. Regulation of different steps of transcription from core promoters (left panel) and enhancers (right panel). (A) PIC assembly and Pol II recruitment; **(B)** Initiation by Pol II and promoter escape. **(C)** Pol II promoter-proximal pausing (Adapted from [10]).

1.2 The molecular basis of recognition of specific DNA motifs by TFs

In contrast to prokaryotic TFs that regulate gene expression through recognising long DNA motifs, eukaryotic TF repertoire recognise short core-DNA motifs with low sequence-based information content and require clustering of sites in regulatory region to achieve specificity. In eukaryotic genomes, the low intrinsic information content required in DNA recognition and the degeneracy of the motifs have suggested low intrinsic specificity of eukaryotic TFs to bind to cognate sites exclusively, instead TFs can partially share their binding motifs due to their common binding preferences.

Cooperativity, which is an inherent feature of mammalian TFs, occurs promiscuously between different TF structural families and is usually mediated by direct or indirect protein-protein interactions. In the biology of protein-DNA recognition, cooperativity is referred as the facilitation effect that a TF bound to DNA provides for the subsequent binding of a second TF. Active cooperative binding of TFs is important in defining the mammalian gene regulatory lexicon [32]. The significant effect of such cooperativity is to facilitate a synergistic transcriptional response. Transcriptional synergy refers to a multiplicative effect given by concerted action of multiple factors that synchronously bind the transcriptional machinery, promoting a strengthened transcriptional response [32][33].

Different mechanisms of cooperativity with specific features have been documented [32][34]. Some modes of cooperativity act locally, while others act more distally. In addition, some modes are very sensitive to changes in spacing and orientation of the single binding sites, whereas others allow more promiscuous pairings between TFs in a space-independent manner. The simplest mechanism is mediated by direct protein-protein interaction between TFs in the absence of DNA, and together bind DNA with enhanced affinity than separately. In a related phenomenon, cooperativity can be directed by DNA, meaning that the interaction between two TFs is not sufficient to form stable dimer in solution. Binding induces changes in DNA conformation, triggering favourable arrangement of the corresponding TFBSs and bringing the TFs close together. In the absence of interactions between TFs, cooperativity can be achieved entirely by DNA. Such cooperativity can be achieved by binding of one TF that induce DNA bending and exposing other TFs recognition sites [32].

The classical mechanism of TF cooperativity, where direct protein-protein interaction is required for stabilization and transactivation potential of the DNA-bound complex, is constrained by the precise arrangement of the cognate TFBSs in gene regulatory regions, thus requiring the correct orientation and the precise spacing between TFBSs. However, genomic and functional studies have reported a significant degree of flexibility and promiscuity in the arrangements of TFBSs, while retaining regulatory functionality *in vivo* [34]. An alternative mechanism of cooperativity that allows significant flexibility in arrangements of TFBSs is based on synergistic binding of non-interacting TFs mediated by a nucleosome [19] [35]. Nucleosome-mediated cooperativity could enhance the binding of disparate TFs to promoter elements, requiring only the minimal DBD of the proteins and that a sufficient number of binding sites

are deeply located in the nucleosomal core. Whereby, the binding of an activator to its cognate site on a nucleosome can alleviate the nucleosome-mediated inhibition of an adjacent internal site by disrupting or displacing the nucleosome structure, creating a nucleosome-depletion region, thus making the site of the second TF more accessible, resulting in co-occupancy [32][34][35].

1.3 The CCAAT-Binding Factor NF-Y

NF-Y has long been a topic of interest since it is first identified in mice in 1987 as a nuclear protein that specifically binds the conserved Y-box transcriptional regulatory element in the major histocompatibility complex class II (MHC-II) gene, thus named Nuclear Factor-Y [36]. Over the past three decades, many aspects of this ubiquitous TF have been extensively scrutinised and multiple lines of evidence have established the crucial role of NF-Y in many biological processes in yeast, plants and animals. Particularly, NF-Y represent the major asset of our focused research programme, which aims at characterising the molecular mechanisms of this striking TF.

In this chapter, I mainly summarise the research progress related to the biology of NF-Y, describing features that was the starting point of my thesis project, including the unique structural characteristics, the potentially diverse biological features and the co-association pattern with a defined set of TFs that make NF-Y a fundamental TF for transcriptional activation.

1.3.1 The CCAAT-box is a widespread feature of eukaryotic promoters

Regulation of gene expression is achieved by gene-specific *trans*-acting factors that recognise a combinatorial puzzle of short *cis*-acting elements within regulatory regions of genome and operating in the context of high order chromatin structures. Some, such as CCAAT, GC and TATA elements, are frequently encountered. The first identified CCAAT box, is a widely distributed regulatory element located at a conserved distance of around -80 bps away from the TSS in either forward or reverse orientation, and occurs at ~30% of eukaryotic promoters and enhancers [37][38]. This occurrence, particularly in the reverse ATTGG orientation appears to be more frequent in TATA-less promoters [37]. Beside the binding to proximal promoter elements, CCAAT motifs are found distally at inactive chromatin and repetitive loci and to a lesser extent at active enhancers [39][40]. A survey of 178 NF-Y binding sites confirms that most, if not all, of the CCAAT boxes are regulated by NF-Y [37]. *In vitro* saturation mutagenesis and SELEX studies have clearly demonstrated that NF-Y strictly requires all invariably conserved five nucleotides and pointed also at flanking nucleotides as important too for high affinity interactions, as further confirmed by genome-wide studies of *in vivo* bound NF-Y loci (**Figure 5**) [37], [39]–[41]. In particular, at positions -2 and -1 there is a clear prevalence for purines, while at the 3-end CAG nucleotides are overrepresented [37][39]. The resulting optimal recognition site encompasses 13 nucleotides devoid of any recognisable symmetry axis [37]. Interestingly, the presence of multiple *bona*

fide CCAAT boxes with precise spacing and organisation within human promoters has been previously reported [38][42]. Bioinformatic analyses of ChIP-seq data evidenced a co-association pattern of NF-Y and other sequence-specific TFs that show global enrichment of CCAAT and significant overlap with NF-Y, including FOS, IRF3, Sp1 and Sp2 bind together with NF-Y specific promoter regions characterised by twin CCAAT boxes spaced by 24-28 and 35-37 bps [40].

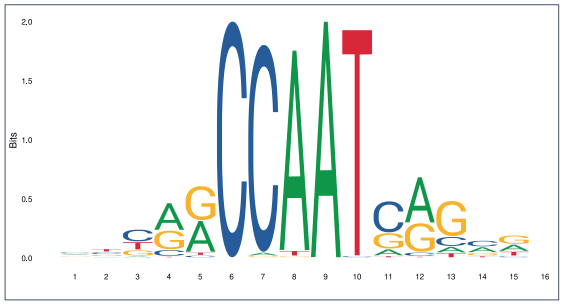


Figure 5. Sequence Logo of NF-YA consensus motif optimised with ChIP-seq data from JASPAR database (Matrix ID: MA0060.1) [43].

1.3.2 The ubiquitous CCAAT-binding NF-Y transcription factor

Nuclear factor Y (NF-Y), also known as the CCAAT-binding factor (CBF) exists in almost all higher eukaryotes. It is a ubiquitously expressed heterotrimeric TF composed of three subunits, namely NF-YA, NF-YB and NF-YC all required for stable DNA binding and share high sequence homology to the yeast HAP2/3/5, respectively. The formed complex recognises and avidly binds the functionally important CCAAT element, frequently present in promoters of cell cycle and growth-regulated genes [38]. *In vivo*, the HFD subunits are more abundant and NF-YA is considered the limiting subunit of the trimer [44][45].

In eukaryotic genomes, NF-Y subunits are usually encoded by a single copy gene, usually with alternative splicing events that generate NF-Y variants (**Figure 6**) [44][46]. NF-YC locus involves the combinatorial usage of alternative promoter and alternative splicing, increasing the post-transcriptional regulation complexity. Particularly, alternative splicing of exons 9 and 11 which reside within the C-terminal glutamine-rich (Q-rich) domain give rise to three nonredundant splicing isoforms that are translated into different proteins of 37, 48 and 50 kDa, with the 37 kDa isoform being the most abundant. Instead, alternative promoter located within the first intron downstream to the canonical promoter position generates a product with different 5'-UTR, affecting the mRNA stability. Noteworthy, NF-YB gene does not involve alternative splicing events, generating a unique translated mRNA with a product of 32 kDa [47]. Instead, NF-YA gene undergoes alternative splicing of exon-3 giving rise to two major protein isoforms, the long isoform (NF-YA*l*, 43 kDa) and the short isoform (NF-YA*s*, 41 kDa), lacking a 28-amino acid encoding exon within the N-terminal Q-rich transactivation domain [48]. Although the two isoforms are considered functionally equivalent, growing evidences suggest different expression

levels that depend on the cell type, developmental stage and differentiation level. Particularly, the NF-YA*l* and NF-YA*s* isoforms are preferentially expressed in fibroblasts and lymphocytes, respectively. Several evidences concerning mouse embryonic (mESCs) and human hematopoietic stem cells (hHSCs) have considered the NF-YA*s* isoform as a regulator of stemness and proliferation [49]–[51]. Overexpression of NF-YA*s* in HSCs led to an increased expression of many HOX genes, as well as multiple TFs notably crucial in HSC maintenance [50]. In addition, retroviral infections of the short isoform in HSCs enhanced their ability to repopulate the bone marrow of immunocompromised mice. Previous study has shown that the NF-YA alternative splicing differentially affects muscle cell fate, where the overexpression of NF-YA*s* boosts cell proliferation, while the long isoform enhances differentiation. They further shown that NF-Y specifically binds active enhancer regions in ESCs where its binding site is overrepresented and is essential for ESC proliferation [48], [51]–[54]

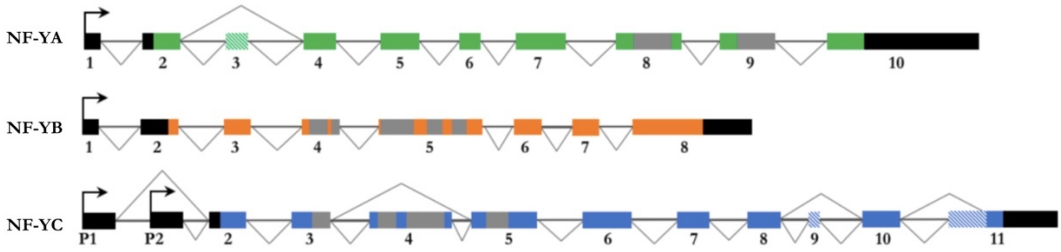


Figure 6. Genomic structure of human NF-Y subunit genes. Arrows indicate the TSS and exons that undergo alternative splicing are striped. Grey boxes indicate the position of the α-helices of the core domains while black boxes indicate the 5'- and 3'-UTRs [44] (adapted from [55]).

Because of its ubiquitous expression, NF-Y has been implicated in regulating a myriad of critical cellular and organismal functions including cell cycle progression, differentiation, development and tumorigenesis. Indeed, knockout of NF-YA is lethal during early mouse embryonic development due to a severe defects of cell proliferation and massive apoptosis [56]. Recently, conditional inactivation of NF-YA alleles after birth in a tissue-specific manner led to a complete block in cell cycle progression, confirming the essential role of NF-Y in maintaining basal tissue homeostasis, metabolic function and cell survival [57]. Functional silencing of NF-Y transcripts by an inducible short-hairpin RNA (shRNA) system revealed that while inactivating NF-YA caused a delay in S-phase progression followed by apoptosis in a p53-dependent manner, targeting the histonic subunits induced G_2/M-specific depletion and subsequent caspase-dependent apoptosis [45][58].

Although the ubiquitous tissue expression pattern of NF-Y subunits in most human and murine cell types, certain type of post-mitotic cells lack NF-YA. Particularly, in circulating monocytes, the NF-YA binding activity is barely detectable at the protein level on the CCAAT-containing gene promoters including class II MHC, mig and fibronectin, however, an increased expression levels is observed on these promoters during macrophages differentiation, with the expression of surface antigens [59]. A

similar pattern of NF-YA involvement has been reported in the muscle system. They have shown that the levels of NF-YA is high in proliferating cells, and drops after differentiation; inactivation of NF-YA in mouse myoblasts drastically affects cells proliferation and differentiation [51].

In vivo findings from *Drosophila* models have indicated an established roles of NF-Y in various human diseases, including neurological and developmental disorders, cardiovascular, metabolic and storage diseases and cancer [60]. Indeed, several observations highlighted the role of NF-Y in the regulation of cell cycle and apoptosis in many tumour cell lines, mainly by regulating the expression of cancer-related genes [61]–[63]. Recent studies revealed that ectopic expression of NF-YA promotes a malignant phenotype in osteosarcoma cells [64]. It was previously reported that suppression of NF-YA in mouse brain neurons induced progressive neurodegeneration characterised by abnormal co-accumulation of p62/ubiquitin accompanied with drastic ER disorganisation [65]. Recently, inactivation of NF-Y in different sets of neurons caused a cell type-specific neuropathology and gene downregulation in mouse Central Nervous System (CNS), suggesting a key roles of NF-Y in neuronal maintenance and protein homeostasis in the CNS [66].

1.3.3 The structure of NF-Y subunits

In this paragraph we will address the considerable progress made towards understanding the structural basis for the fascinating biology of NF-Y during the past years.

Sequence alignments revealed that each subunit holds an evolutionary conserved core domain sufficient for subunit trimerization and CCAAT-binding (**Figure 7**). The core domain of NF-YA subunit is <60 amino acids (aa) long and is located at the C-terminal region of the protein. It is composed by two extended α-helices (A1 and A2) separated by a ~12-aa linker loop (A1A2). The 20-aa N-terminus A1 contacts extensively the NF-YB/NF-YC dimer, whereas the 21-aa C-terminus A2 provides exquisite sequence-specificity for recognition and binding of the CCAAT *cis*-element. The A1A2 linker adopts an elongated structure that contributes to hetero-trimerization by providing the conformational flexibility necessary to direct the A2-helix toward DNA and optimizing the electrostatic interactions at the created subunits interface [67][68]. NF-YB and NF-YC core regions consist of histone-fold domains (HFDs) which closely resemble the H2B and H2A core histones, respectively, and function in protein-protein and protein-DNA interactions. Beside these conserved core domains, NF-YA and NF-YC subunits hold a Q-rich domain at the N- and C-terminal ends, respectively, with transactivation potential [42]. The activation function in yeast and fungi where no Q-rich domains are present is encoded by an additional subunit, with no clear homologues in other species [69]–[71].

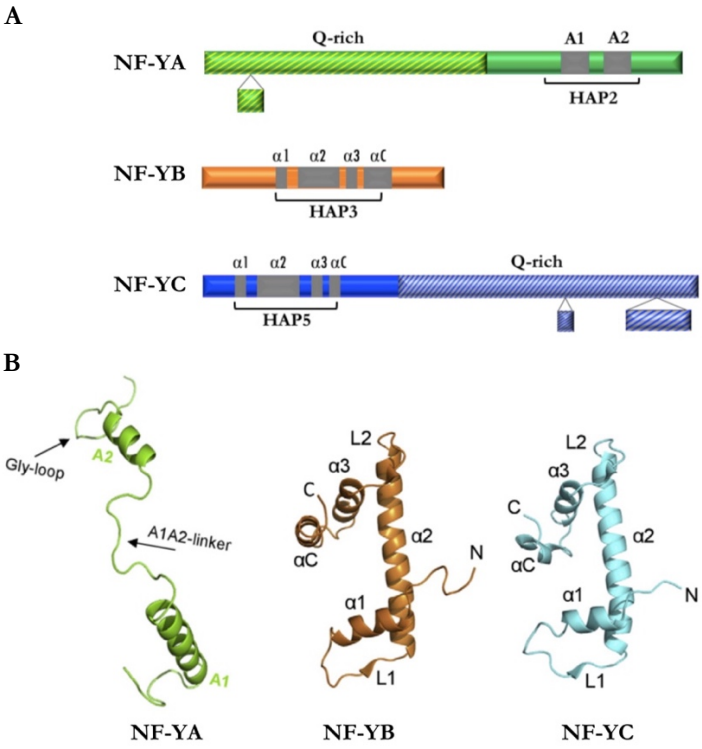


Figure 7. NF-Y subunits. (A) Schematic representation of NF-Y subunits. Green, orange and blue-coloured boxes highlight the evolutionary conserved core domains of NF-YA, NF-YB and NF-YC subunits, respectively. Grey-shaded boxes indicate the secondary structure composition included in the core domains of each subunit. The yeast homologous regions (HAP) and Q-rich domains are indicated. **(B)** Ribbon representations of the tertiary structure of human NF-Y subunits. The key secondary structure elements are labelled. The arrows point to the flexible A1A2-linker and the Gly-loop motif of NF-YA subunit [68].

The first crystal structure of the conserved core regions of human NF-YB/NF-YC subcomplex [72] and the subsequent crystallisation of human and fungal heterotrimeric NF-Y/CCAAT complexes firmly confirmed the histone-like architecture of the HFD-pair with common and unique features [67][71]. Similar to the typical H2B/H2A and NC2β/NC2α subunits, NF-YB/NF-YC HFD is minimally formed by three α-helices (namely α1, α2 and α3) separated by linker loops (L1 and L2), with α1 and α3 are orthogonal to α2. The structural homology of the HFDs of NF-YB/NF-YC and H2B/H2A is clear from the superimposition of the respective structures (**Figure 8A**). Outside the HFD, the similarity extends to an additional C-terminal loop-short helix secondary structure element (called αC) found in both pair. Instead, the N-termini markedly differ, with an additional extension present in H2A and H2B, that is absent in NF-YB and NF-YC HFDs [72]. In fungi, the HapE subunit (homologous to NF-YC) holds also an extended N-terminal region, named αN, that adopts different orientation compared to the core histones. A characteristic intra-chain Arg-Asp bidentate salt bridges linking L2 to α3 in each subunit, further stabilise the NF-YB/NF-YC HFD [68][71]. In their tertiary structure, the HFD modules of NF-YB and NF-YC typically assemble in a head-to-tail fashion, creating a compact heterodimer stabilised by extensive hydrophobic interactions and exposing an extended basic surface, allowing favourable non-specific contacts with the negatively charged DNA sugar-phosphate backbone.

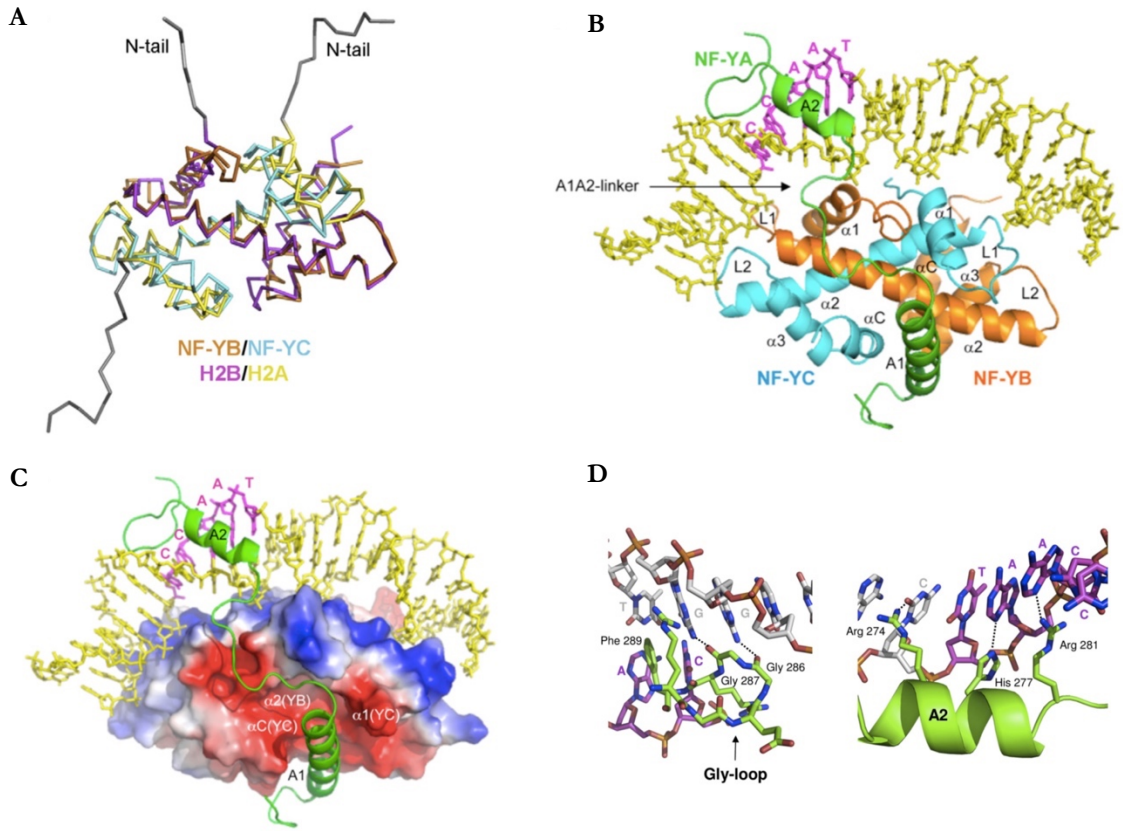


Figure 8. **Three-dimensional structure of NF-Y. (A)** Structural superimposition of coil representations between NF-YB and NF-YC (PDB ID 4CSR) and the core histones H2B and H2A (PDB ID 1AO1, chains C/D, respectively). The N-terminal histone tails are highlighted in grey. **(B)** Ribbon representation of NF-Y heterotrimer in complex with a 25 bp-CCAAT box oligo derived from HSP70 promoter (PDB ID 4AWL). The secondary structure elements are labelled. **(C)** Electrostatic surface of HFD NF-YB/NF-YC heterodimer in complex with NF-YA and DNA illustrated as ribbon and stick diagrams. Red and blue colours indicate negatively and positively charged regions, respectively. The insertion of NF-YA A2-helix in the minor groove is highlighted. **(D)** Sequence-specific interactions in NF-YA/CCAAT box recognition. The CCAAT DNA and the complementary strands are coloured in magenta and grey, respectively. Dashed lines indicate the hydrogen bond interactions between the NF-YA residues belonging to the Gly-loop and the A2-helix with the DNA. Important residues are labelled [67][68].

The unique structural feature of NF-YB/NF-YC HFD dimer compared to other HFD-containing proteins is a wide negatively charged surface groove formed by NF-YC αC and α1, and NF-YB α2 offering docking sites for NF-YA, the regulatory subunit that confer sequence-specificity to the formed histone-like pair. Interestingly, the additional αC of NF-YB and NF-YC have an established role in NF-Y trimerization and targeting regulatory proteins. Particularly, the αC helices of NF-YB/NF-YC have been shown to interact with the TATA-binding protein (TBP), while NF-YC αC can associate with the proto-oncogene c-MYC and the tumour suppressor p53 [72].

The crystal structure analysis of mammalian NF-Y/CCAAT complex highlights important structural features that account for specific DNA recognition of NF-Y at the CCAAT box (**Figure 8B**). Unlike HFD subunits which interact with DNA *via* nonspecific-DNA contacts, NF-YA docks into the groove on the NF-YB/NF-YC subcomplex and deeply inserts an α-helix in the DNA minor groove and

specifically contacts the CCAAT nucleotides, thus providing NF-Y a sequence-specific targeting capability. Trimerization occurs mainly through the docking of the positively charged side-chains of A1-helix in the extended acidic groove of the HFD-pair. NF-Y binding results in a nucleosome-like global DNA bending of about 80°. The CCAAT-binding specificity is achieved by inserting deeply into the DNA minor groove the A2-helix and the following Gly-loop motif (GxGGRF with x = any residue), resulting in a strong widening of the minor groove of about 19 Å at the first adenine of the CCAAT box, providing a means to induce a significant bend in the DNA (**Figure 8C**). Particularly, the Arginine and Histidine residues of the A2-helix that contact specific atoms in the CCAAT box, together with the kinked backbone of the Gly-loop that allows close proximity of the two Gly-Gly carbonyl oxygen atoms to the two guanines of the CCAAT complementary DNA strand, are extremely crucial for CCAAT-sequence specificity. Importantly, the binding of NF-YA does not affect the adjacent major grooves, promoting the potential binding of other TFs that can synergize with NF-Y, whose recognition motifs remain accessible. The NF-YA-mediated DNA interactions are further stabilized by several additional contacts with the DNA phosphate backbone and the insertion of a Phe residue belonging to the Gly-loop between the two consecutive C and A bps within the CCAAT box (**Figure 8D**). As a result, NF-Y combines the ability of the HFD to form compact but non-specific complexes with DNA, together with the unique DNA-reading mode of NF-YA to form stable and highly selective complex [67], [68], [72].

These structural observations offer useful starting points to rationalise the fine molecular crosstalk associated with the wide combinatorial association of NF-Y and other TFs that act in concert for transcriptional activation and unravel the 3D structure of such higher order complexes using integrative structural biology techniques.

1.3.4 The Pioneering role of NF-Y in transcription activation

In view of its unique DNA-binding mode, NF-Y is an established paradigm of pioneer TF. Through its ability to bend DNA together with the sequence-specific recognition by NF-YA, NF-Y acts as an organiser TF, shaping the promoter architecture and distal regulatory regions [67][68]. Additionally, genome-wide mapping of NF-Y binding landscape from the ENCODE (Encyclopaedia of DNA Elements) consortium revealed high degree of complexity, were NF-Y has been shown to be associated with regions containing negative histone marks [73].

Pioneer TFs possess a unique property of binding specific DNA motifs within compacted or "closed" heterochromatin where genomic DNA is not readily accessible and elicit remodelling or "opening" of the chromatin, thus providing accessibility to non-pioneer TFs for cellular reprogramming implementation [74]–[76]. Pioneer's binding is distinguished by a higher residency period, suggesting stable chromatin-TF interactions that lead to a long-term maintenance of cell identity [77]. In light of their chromatin remodelling properties, pioneers have the potential for high epigenetic perturbations that

dysregulate the genome, thus altering the transcriptional pattern as seen in cancer cells. Therefore, pioneer factors can be used as prognostic biomarker and drug targets in cancer [73][76].

To orchestrate nucleosome removal/displacement, NF-Y can recruit a bona fide ATP-dependent chromatin remodeler to the promoter of target genes. Mechanistically, through its nucleosome-like structural features and its distinct DNA-binding mode which induces an ~80º bend in the DNA , NF-Y can independently facilitates a permissive chromatin conformation, promoting master TF binding [78]–[80]. Several lines of evidence have pointed toward a dual role of NF-Y in regulating both genes with housekeeping functions and genes required for cell-cycle identity through its binding to cell type-invariant proximal promoters [80] and cell type-specific enhancers [79], respectively.

Besides the first genomic and structural observations, further evidences supported the pioneering role of NF-Y: (i) *in vivo* functional assays coupled with unbiased machine-learning techniques were applied to differentiating mESCs to analyse genome-wide DNaseI-hypersensitivity sites (DHSs) profile to predict TF-binding sites from the magnitude and shape of DNase footprints [81]; (ii) genomic profiling in mESCs revealed that NF-Y facilitates a permissive chromatin conformation at distal regulatory elements of developmentally regulated genes for the master pluripotency factors including Sox2, NANOG and Oct4 [79]; (iii) genome-wide mapping of DHSs was applied to understand how the chromatin regulatory landscape is dynamically established during the zygotic to preimplantation development of mouse embryos. Consistent with the fact that DHSs are often occupied by TFs, TF-binding motif enrichment analysis was performed using 2-cell DHS map and identified the CCAAT box as the most overrepresented motif. In this study, they firmly identified NF-YA as one of the key TFs responsible for the zygotic genome activation [82]; (iv) siRNA-mediated knockdown of one of NF-Y subunits or overexpression of a dominant-negative NF-YA mutant unable to bind DNA is deleterious for recruitment of several TFs and cofactors, suggesting NF-Y-dependent tethering events [40][83]; (v) in the latest study aimed to investigate NF-Y's function and mechanism of action at proximal promoter regions in mESCs, they used siRNA to generate NF-YA null mice coupled with DHSs-profiling. Examination of the nucleosomal landscape at NF-Y-bound promoter regions revealed a striking well-defined nucleosome-depleted regions (NDRs), suggesting a steric anti-correlation between NF-Y and nucleosome occupancy. NF-YA knockdown exhibits a reduced promoter accessibility through the accumulation of ectopic nucleosomes over the TSS. They conclude that NF-Y has an established role in safeguarding the integrity of the NDR, correct TSS selection and PIC localisation by occluding nucleosome encroachment at its binding sites, thus ensuring faithful transcription initiation [80].

1.4 NF-Y and fellow TFs

Owning to its intrinsic ability of accessing to chromatin, influencing local chromatin environment, bending DNA and physically interacting with many TFs, NF-Y behaves as an "architectural organiser"

that shapes the three-dimensional arrangement of the promoter, as well as a "facilitator" of transcription that brings distal TFs closer the transcription machinery [40][45]. NF-Y collaborates and acts in concert with several TFs to exert accurate regulation of their target genes.

Over the past years many individual examples of NF-Y-mediated DNA binding cooperativity with several TFs have been established. Starting by the MHC-II system where NF-Y was originally characterised, NF-Y bound to the Y-box and RFX bound to an adjacent X-box cooperatively interact to form a functional unit in enhancers [84]. A subsequent well-characterised paradigms are represented by NF-Y combined with SREBP [85], SP1 [86], C/EBPα [87], ATF6 [88] and USF [89] to elicit synergistic transcriptional activation of specific set of promoters, typically through adjacent binding sites. Interestingly, several TFs are found to bind DNA regions devoid of their respective consensus binding motifs but possess a CCAAT-box bound by NF-Y, suggesting NF-Y-mediated recruitment. Numerous examples have been previously reported, including the tumour suppressor p53 [90], FOS [73], IRF3 [40] and SP2 [91].

TF Class	Group 1 Global CCAAT enrichment only Peaks overlap >20%	Group 2 Global or Local CCAAT enrichment Peaks overlap < 20%	Group 3 No CCAAT enrichment Peaks overlap > 10%
b-Zip	CFOS	JUND ATF1 ATF3 MAFF MAFK CEBPB NFE2	NRF1
b-HLH	USF1 USF2	MAX	MXI1 MYC
ZNF	SP1 SP2	ZNF143	EGR1 MAZ
HD	SIX5 PBX3		POU2F2
E2F	E2F4		E2F1 E2F6
ETS		ELF1 GABP SRF	ELK1 ELK4 ETS1
Other	RFX5 IRF3	RUNX3 ARID3A	TCF7L2 THAP1 YY1
GTFs	TBP		GTF2B GTF2F TAF1
Cofactors	HCFC1 CHD2	COREST BRCA1	KDM5B SIN3A PML

Figure 9. Classification of factors that co-associate with NF-Y locations in three groups, considering the combination of CCAAT enrichment and peak overlap analysis. Group 1 TFs showing global enrichment and genome-wide overlap with the CCAAT motif; Group 2 limited enrichment and overlap with NF-Y; Group 3 TFs with significant overlap with NF-Y, but with no relevant CCAAT enrichment [40].

Starting by a comparison of NF-Y ChIP-Seq dataset derived from the ENCODE consortium with available experimental data for other TFs, several TFs that show genome-wide co-localisation with NF-Y have been identified. Particularly, a notable co-association peaks were found for FOS, USF2 and E2F4 [73]. Additionally, several bioinformatic studies have shown that the functional CCAAT elements bound by NF-Y are among the most frequent motifs enriched in ChIP-seq peaks for TFs different from NF-Y [92][93]. Further studies aimed to characterise those TFs binding DNA in the neighbourhood of NF-Y, by assessing the enrichment for the CCAAT box and the overlap with NF-YB peaks in the genomic loci

bound by 154 TFs from a complete set of 380 ENCODE ChIP-seq experiments. For sequence-specific TFs, they were able to identify three main different levels of co-association with NF-Y (summarised in **Figure 9**). The first group, is the most relevant one that represents TFs that show a significant genome-wide overlap and enrichment for CCAAT boxes in their DNA binding loci, including FOS, RFX5, IRF3, SP1/2, PBX3 and CHD2. Some of these factors show precise arrangement of their DNA-binding sites with respect to the CCAAT box, while others appear to be bound to genomic regions that lack their canonical target motif, suggesting NF-Y-dependent recruitment. The second group includes TFs with limited enrichment and overlap, while the third group represents TFs for which the overlap is still significant, but with no significant CCAAT enrichment [40][93].

A

Family	Factor	DNA binding site	TSS	ENH	TX	Other	TOT	TSS	ENH	TX	Other	Repeats
E2F	E2F4	CCAAT+E2F	1638	46	3	2	1689	1.21	-2	-5.41	-8.39	-3.73
b-HLH	USF1	CCAAT+Ebox	1164	416	148	871	2599	0.22	0.95	-0.58	-0.66	0.09
b-HLH	USF2	CCAAT+Ebox	465	200	48	238	951	0.1	0.55	-0.41	-0.24	0.24
GTF	TBP	CCAAT+TATAbox	2308	115	24	32	2479	1.15	-1.23	-2.96	-4.94	-2.07
b-Zip	FOS	Double CCAAT	2712	371	128	486	3697	0.81	-0.12	-1.13	-1.59	-1
ZNF	Sp1	Double CCAAT	2205	144	79	188	2616	1.01	-0.99	-1.32	-2.46	-1.83
ZNF	Sp2	Double CCAAT	1725	86	21	90	1922	1.1	-1.29	-2.79	-3.08	-3.21
Other	RFX5	Double CCAAT	443	37	12	24	516	1.03	-0.61	-1.7	-3.09	-2.63
Other	IRF3	Double CCAAT	1362	40	5	13	1420	1.19	-1.95	-4.42	-5.44	-3.68
Cofactors	CHD2	Double CCAAT	1911	100	19	55	2085	1.13	-1.19	-3.05	-3.91	-2.81

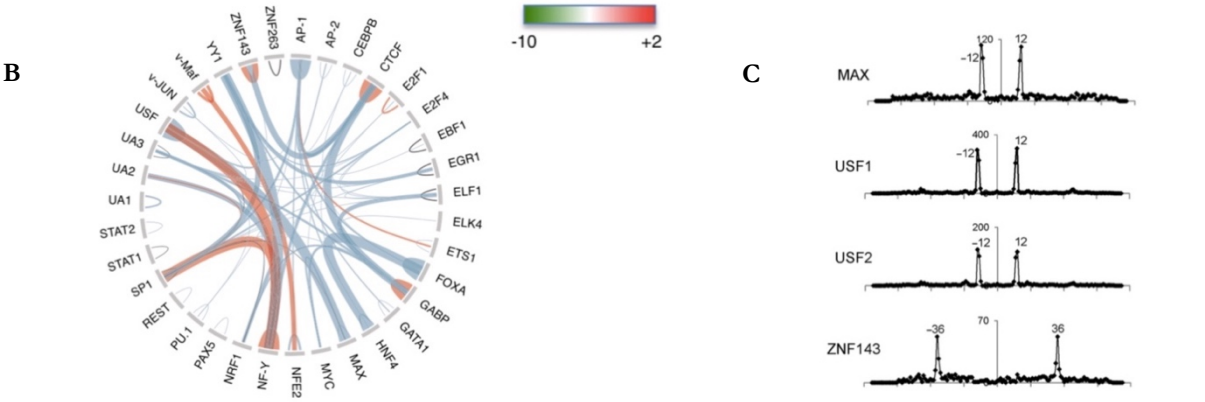


Figure 10. **NF-Y global interactors. (A)** The TF, its family, the binding motifs enriched on NF-YB loci, the number of co-bound sites in active regions (TSS), enhancer (ENH), transcribed (TX) and other genomic location are indicated from left to right. For each chromatin region, the heatmap shows if the overlap with NF-Y sites is above (red) or below (green) the expected value. The numbers indicate the log2 of the observed/expected number of regions ratio [93]. **(B)** Circos plot illustrates motif pairs connected by an arch with significant preferential distances between their binding sites. The thickness of the connection is proportional to the frequency of the pair. Red, black and blue connections indicate that the motif pair is discovered in multiple, same or different datasets, respectively [94] **(C)** Distribution of the relative position of the binding sites for TFs that show a positional bias with respect to the CCAAT motif (MAX, USF1/2 and ZNF143) in co-bound regions in K562 cell line [40].

TFs belonging to the group 1, considered as "global" interactors, are further classified by analysing at the single base pair level the occurrence of a preferential positioning on DNA of their binding sites with respect to the CCAAT motif (**Figure 10A**). An interesting observation is that the regions bound by NF-Y together with the TFs that lack their canonical binding sites contain a twin CCAAT-box configuration, usually with 24-28 and 35-37 bps distance. On the other hand, the E-box binding TFs (USF1/2 and MAX, but not MYC) and Znf143 show predominant and almost exclusive distance biases

of 12 and 36 bps, respectively in the NF-Y loci with respect to the CCAAT box as shown in **Figure 10B, C** [40][93][94]. Chromatin state annotations were further investigated to assess possible bias for given genomic location or chromatin states. Beside USF1 and USF2 found enriched at distal enhancer and repetitive regions, most TFs show a bias towards binding with NF-Y at active TSS (**Figure 10A**) [93]. E2F family members belonging to the global NF-Y interactors without a positional bias of its canonical binding site, correlate with NF-Y to regulate genes with different functional outcomes, suggesting possible NF-Y-mediated tethering on DNA. The most striking co-localisation was found for E2F4, that depends on NF-Y for efficient promoter association [40][93].

Genome-wide co-association of NF-Y with the most global partners on human promoters characterised by twin CCAAT boxes with preferred distances is summarised in **Figure 11**, showing a gapless alignment of zinc finger genes promoters bound by NF-Y in K562 cell line and containing two CCAAT elements separated by 24 bps [93].

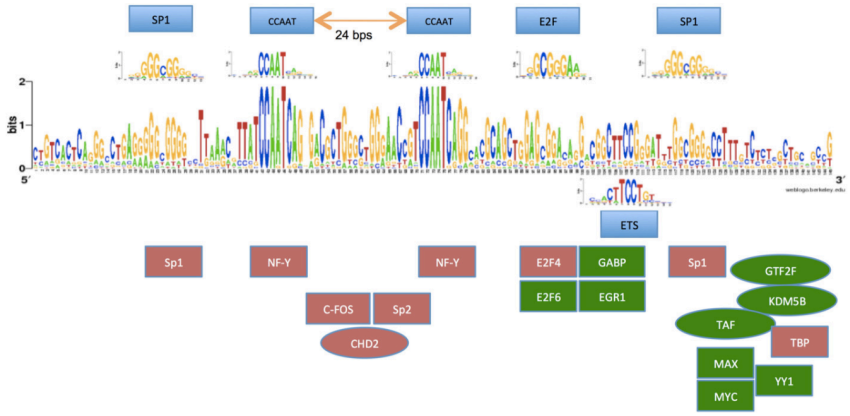


Figure 11. Logo illustrating the gapless multiple alignment of human zinc finger genes promoters that contain two CCAAT boxes separated by 24 bps. Light blue boxes indicate the localisation of the recognisable binding motifs. According to ENCODE ChIP-Seq dataset, green and red ovals (GTFs and co-factors) and boxes (TFs) indicate the approximate localisation, along the promoters, of the TFs bound in K562 cell line. Factors belonging to group 1 are depicted in red [93].

1.4.1 NF-Y and E-box transcription factor USF1

The Upstream Stimulatory Factor (USF) 1 and 2 belong to the basic helix-loop-helix leucine-zipper (bHLH-ZIP) family of TFs that bind to the E-box (CACGTG); this family includes also MAX and MYC. The structural hallmark of this family is an evolutionary conserved bHLH C-terminal domain divided into two functionally distinct regions: the basic region that directly contacts DNA connected through a stabilising loop to the leucine-zipper region that provides the potential for homo-and heterodimerization. Upstream of the basic region, a predicted disordered region termed USF-Specific Region or USR essential for USF-dependent promoter activity [95]–[97]. In addition to these conserved regions, these factors often contain a leucine-zipper region involved in protein-protein interaction (**Figure 12**) [98]. NF-Y and USF were shown to co-regulate different gene promoters in a synergistic fashion [99][100]. It was previously demonstrated that the precise spacing between the E-box and the CCAAT-box of 10-12 bps

is critical to achieve DNA-binding cooperativity and a synergistic transcriptional activation [101][102]. The ENCODE analysis previously described confirmed the co-localisation of USF1/2 and MAX (but not MYC) with NF-Y [40][73]. Interestingly, USF1 is one of the few TFs that co-associates with NF-Y in the long terminal repeats (LTR)/nonmodified repressed chromatin (**Figure 10**) [73].

Figure 12. Schematic representation of USF1 protein. The main functional domains are indicated as coloured boxes (adapted from [55]).

1.5 The adenoviral Early region 2 binding Factor

In the late '80s, characterisation of the adenovirus E2 promoter led to the identification of cellular factor with functions important for early events in viral replication, named the adenoviral early region 2 binding factor E2F [103][104]. In mammals, the E2F family of TFs consists of eight members (E2F1-E2F8) with unique and overlapping functions, each with the characteristic hallmark of a conserved DBD that directly binds the consensus E2F-response elements ($TTT^{C}/_{G}{}^{G}/_{C}CGC$) [105]. The complexity of this family has further increased by the discovery of numerous E2F isoforms that arise from alternative splicing and multiple TSSs [106]. In particular, the E2F3a and E2F3b are different isoforms of the E2F3 gene locus that share similar DBD, dimerization and TADs but are regulated by distinct promoters [107]. E2F proteins are generally sorted into two different classes on the basis of sequence homology (**Figure 13A**) [108]. Typical E2Fs (E2F1-E2F6) contain a conserved dimerization domains and associate with the differentiation-regulated transcription factor-1 polypeptide (DP1-3) family of genes, which are distantly related to the E2F family [109]. Interestingly, the heterodimeric E2F/DP1 and E2F/DP2 complexes bind DNA in a sequence-specific manner and regulate E2F target genes similarly, while E2F/DP3 complexes fail to bind efficiently the E2F consensus site [110][111]. While DP1 is ubiquitously expressed, DP2 has a restricted expression level and many isoforms [112]. E2F6 protein has been isolated in complexes with polycomb-group (PcG) proteins and seems to function exclusively as a repressor [113]. The "atypical" E2F7 and E2F8 contain two tandem repeats of the E2F DBD that associate to form a unique DNA binding surface that binds the E2F consensus sequence in a DP-independent manner [114][115]. Homodimers and heterodimers of E2F7 and E2F8 negatively regulate cell cycle progression by directly suppressing the transcription of a subset of E2F-activated target genes [116][117].

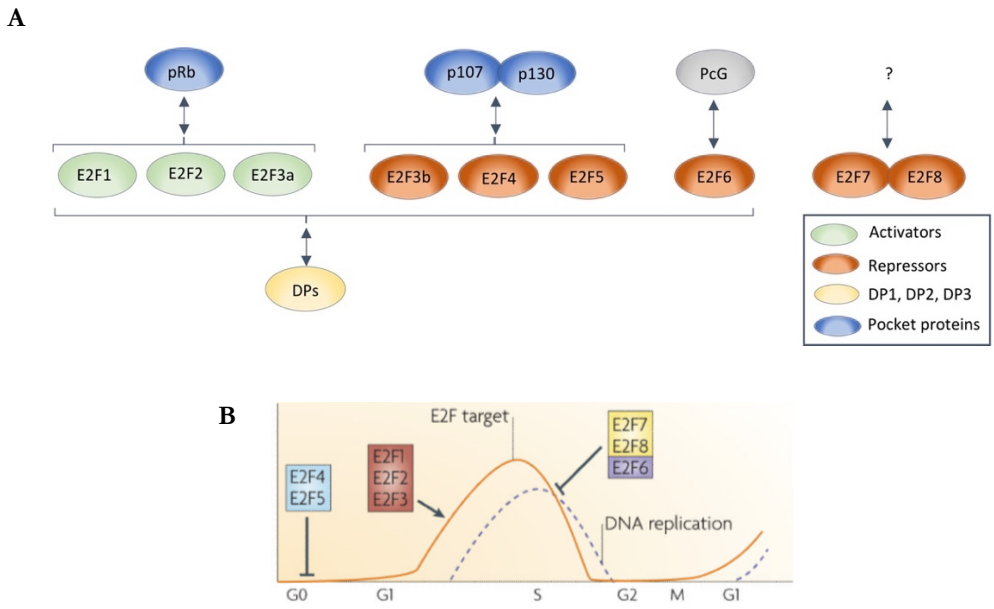


Figure 13. The mammalian E2F family members. **(A)** E2F and pRb protein families. E2F family is divided into two subclasses: activators (E2F1-3a) and repressors (E2F3b-8). Typical E2F1-6 bind DNA as heterodimers with the structurally related DP1-3 subunits. The pocket proteins (pRb, p107 and p130) associate and inhibit different subset of E2F proteins [117]. **(B)** The expression of E2F proteins throughout the cell cycle [118].

E2F proteins ensure correct timing expression of cell-cycle genes by sequential binding of E2F activators and repressors to target promoters. Therefore, individual E2F shows distinct activity and expression level during the cell cycle (**Figure 13B**). While the canonical repressors (E2F3b-E2F6) are constitutively expressed throughout the cell cycle, activator protein (E2F1-E2F3a) levels increase at the G_1-S phase transition, and the atypical E2F7 and E2F8 repressor levels peak in the late S-phase [118]. Recent study has found that in differentiating cells, activator E2Fs can switch roles from activators to repressors [119]. Acting through association with other TFs, E2F4 may stimulate proliferation during development [120][121]. Beyond cell cycle regulation, E2Fs play roles in apoptosis, metabolism, angiogenesis and maintenance of genomic stability. Activator E2Fs act redundantly to promote cell proliferation, increasing the complexity of this family. Several studies have shown that while loss of one E2F is generally tolerated in mouse, combinatorial loss of all E2F activators is embryonic lethal by abolishing progression through the cell cycle and proliferation [122][123].

Within the contexts of the cell cycle and oncogenesis, E2F activity is tightly regulated at multiple levels, including expression, mRNA stability, PTMs, protein degradation, subcellular localisation and binding to regulatory proteins [122]. The best mode of regulating E2F activity is by the retinoblastoma protein (pRb) and the homologous pocket proteins p107 and p130, which directly bind and inhibit different subsets of E2F proteins by masking their TADs and recruit transcriptional repressive factors to E2F-responsive promoters [117][122]. Although the C-terminal binding domain of pocket proteins is highly conserved, they show clear preferences for distinct E2Fs. While the activating E2Fs associate exclusively with pRb, repressors are found associated primarily with p107/p130, suggesting important

distinct functions. Structural, biochemical and phylogenetic analysis reveal that pRb evolved specific structural motifs that make it uniquely able to bind and regulate the activating E2Fs [124]. Consistent with these observations, sequence alignment revealed that E2F1-3 are highly related to each other compared to E2F4-5. Pocket protein-E2F dissociation are tightly regulated by cyclin-dependent kinase (CDK)-mediated phosphorylation or binding of viral oncoproteins including the SV40 T-antigen [125], releasing of active E2Fs [126]. Alterations in the E2F-Rb-CDK axis either by deletion or mutation of pRb, or by perturbations in the upstream CDK pathways lead to aberrant proliferation that occurs in virtually all cancers, underscoring the vital role of this pathway in the determination of cell fate [122]. Interestingly, the transcriptional potential of E2F is tightly regulated by cell cycle-dependent changes in subcellular localisation. While E2F1-3 contain nuclear localisation signals and are localised exclusively in the nucleus, the nuclear localisation of E2F4 and E2F5 is restricted to certain cell cycle stages and dependent upon specific association with a subset of nuclear factors, including their DP partners or the pocket proteins [127].

E2F1 is known as the founding member of this family, that is endowed with contradictory functions, being able to promote cell-cycle progression and therefore is potentially oncogenic and induce apoptosis. DNA damage and mitogenic signals trigger multiple signalling pathways that direct E2F1 to proliferation or cell death (**Figure 14**) [128]. Consistent with the dual role of E2F1 as an activator and a repressor of transcription, knockout of E2F1 in mice lead to dysplasia and tumours in some tissues, and to atrophy in other tissues [129][130]. Most E2F genes, specifically E2F1, are significantly upregulated in different types of cancer [131]. In a mouse model of metastatic breast cancer, recent work has uncovered new function for the activator E2Fs in promoting tumour development and metastasis by regulating the expression of genes critical for angiogenesis, extracellular matrix remodelling, tumour cell interactions with endothelial cells and tumour cell survival [132].

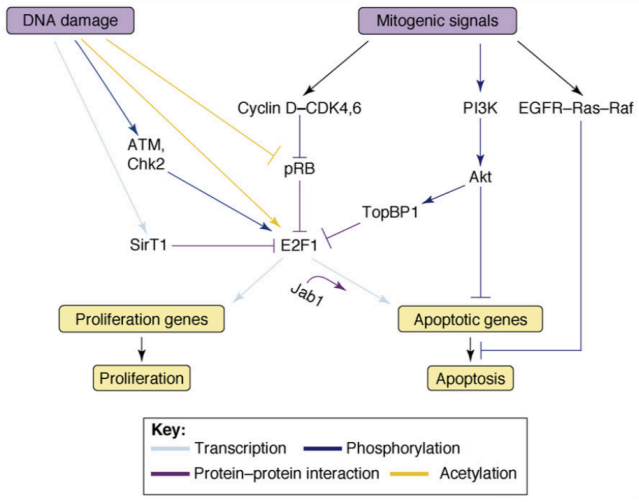


Figure 14. Upstream signals that direct E2F1 to apoptosis or proliferation. DNA damage and mitogenic signals activate a cascade of signalling pathways that induce E2F1-mediated activation of apoptotic or proliferative target genes. In the presence of external signals for growth and survival, the apoptotic activity of E2F is suppressed. When cells experience DNA damage, the apoptotic potential of E2F1 is unleashed [128].

1.5.1 The structural basis of E2F and DP proteins

From the structural viewpoint, the canonical E2Fs have similar arrangement of functional domains and share 20%-55% sequence identity. The DP proteins share high sequence homology (70%) but are distantly related to the E2F proteins [133]. Similar to the DP proteins, E2F1 through E2F6 are characterised by a conserved DNA-binding domain (DBD), coiled-coil (CC) and marked box (MB) domains (**Figure 15**). The DBDs of E2F and DP proteins are homologous and consist of ~70 and 90-residue domains, respectively that bind as heterodimer similar DNA sequences [133]. In particular, the DBD of DP proteins contains a 30-residues region that resembles that of E2F, termed the DEF box [134]. The subsequent CC-MB domains of E2F and DP heterodimerise, and the formed heterodimers bind other TFs as a mechanism that explains how specific E2Fs that recognise the same consensus motif can activate distinct genes [135][136]. The transactivation domain (TAD) is unique for E2F1-5 and consists of ~60 amino acids located near the C-terminus, harbouring the binding sites of positive regulators including TBP, p300/CBP and the MDM2 oncoprotein and negatively acting proteins such as p53 and pRb proteins [133]. Interestingly, in E2F1 promoters, E2F recognition motif possesses an intrinsic DNA bend of ~40° that is reserved upon binding of E2F1 but in the opposite orientation. They also found that this DNA-bending is critical in the activation of the E2F1 promoter. Mutational analysis identified the MB domain as critical to reverse the inherent bending of E2F site. On the other hand, the binding of Rb inhibits the ability of E2F1 to reverse the bent DNA [137].

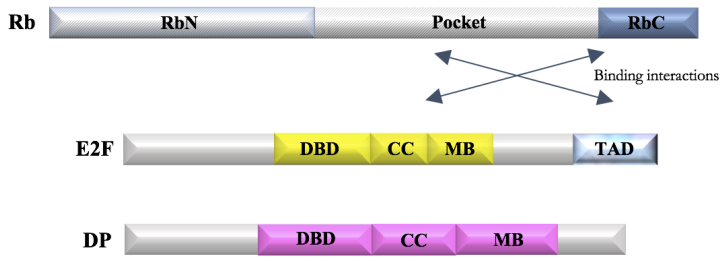


Figure 15. Schematic representation of E2F, DP and Rb domains. The conserved domains are indicated as coloured boxes. The E2F TAD interacts with the Rb pocket domain, and the CC-MB domains of E2F and DP bind the RbC [133].

The crystal structure of the ternary complex harbouring the DBDs of E2F4 and DP2 bound to a 15 bps DNA derived from the adenovirus E2 promoter revealed a winged-helix DNA-binding motif composed by three α helices (α1, α2 and α3) and three β sheets (β1, β2 and β3) allowing the formation of a compact hydrophobic core through extensive protein-protein hydrophobic interactions (**Figure 16A**) [133]. The DBDs are structurally homologous and bind to and transactivate at the consensus TTT$^{C}/_{G}{}^{G}/_{C}$CGC E2F-site [124][133]. The structural homology of E2F4 and DP2 is evident from the superimposition of the respective quaternary structures, except that E2F4 is characterised by a shorter α2 and α3 and an extended N-terminal helix (αN) (**Figure 16B**). The recognition of the palindromic CGCGCG sequence is fundamentally a symmetrical arrangement, where E2F4 and DP2 each contact

half of the sequence through the conserved RRXYD DNA recognition motif on their α3-helices. Interestingly, the αN helix is deeply inserted in the minor groove near the T-rich stretch to contact an invariant arginine residue (Arg17) among E2F family members, orienting the E2F-DP heterodimer on the promoter (**Figure 16C**). This heterodimer is further stabilised by an extensive hydrophobic interaction formed by the α1-α3 helix packing from both E2F4 and DP2. Although the DBDs of E2F and DP proteins have limited homology across-family, most of the E2F4 and DP2 residues that contact DNA and that create the heterodimerisation interface are invariant within their respective families, suggesting that other combinations of E2F and DP proteins would create similar contacts and structural folding [133].

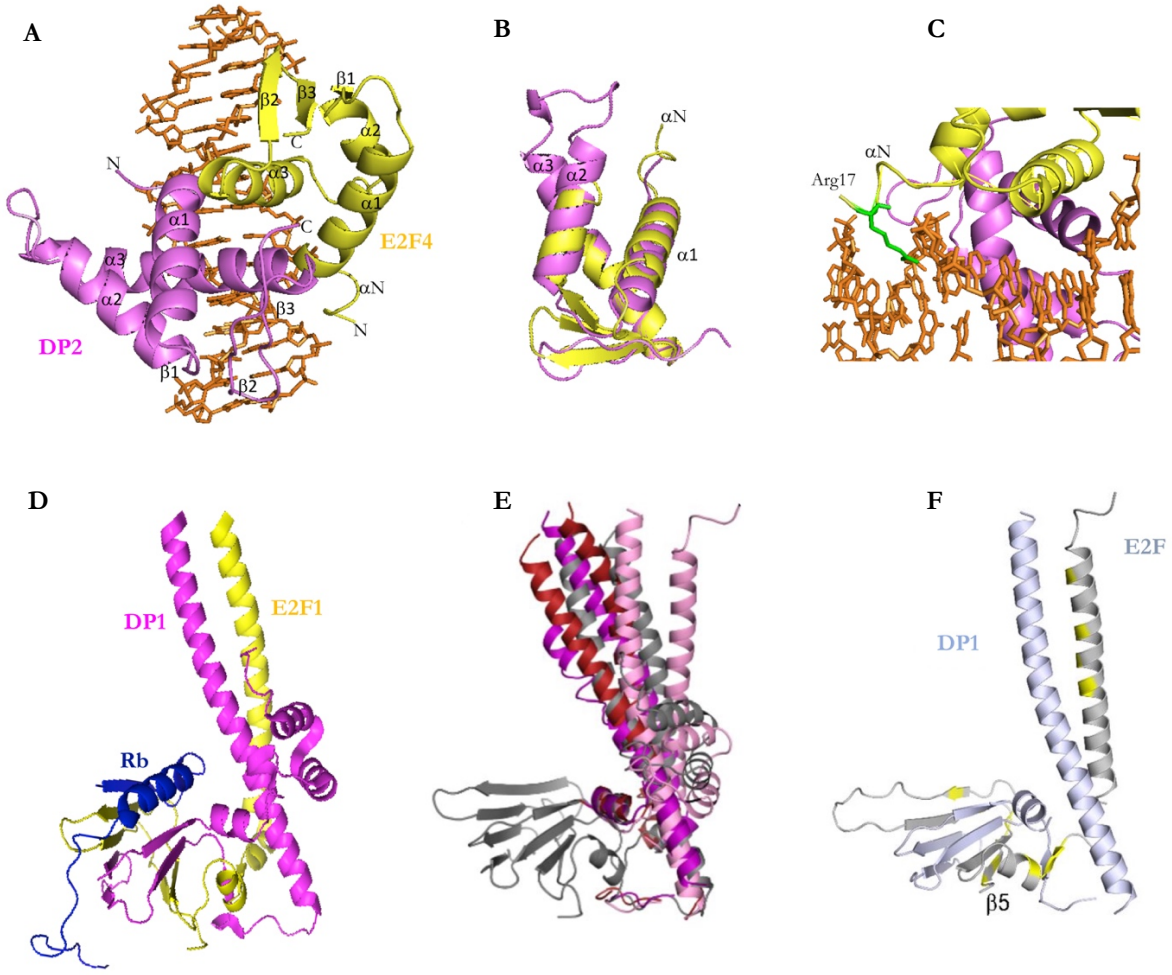


Figure 16. Ribbon representation of human E2F/DP three dimensional structures. **(A)** DBDs of E2F4/DP2 heterodimer in complex with a 15 bps oligonucleotide from the adenovirus E2 promoter (PDB-code: 1CF7 [133]). E2Fs and DPs are depicted in yellow and violet, respectively, with DNA nucleotides highlighted in orange. The secondary structure elements are indicated. **(B)** Structural superimposition of E2F4 and DP2 winged helix DBDs. **(C)** Close-up on the E2F-αN helix inserted in the DNA minor groove near the T-rich stretch. The conserved Arg17 is indicated in green. **(D)** $E2F1^{CM}/DP1^{CM}$ in complex with RbC peptide (blue) (PDB-code: 2AZE [138]). **(E)** Overlay of the CC-MB domain structures of E2F1/DP1 (grey; PDB-code: 2AZE), E2F5/DP1 (red and purple; PDB-code: 5TUV), and E2F4/DP1 (pink; PDB-code: 5TUU). **(F)** Conserved residues among E2F paralogs are mapped and highlighted in yellow on $E2F4^{CM}/DP1^{CM}$ structure. Most of these residues form the last β5 strand and the preceding loop [124].

The retinoblastoma tumour suppressor (Rb) is a large and multifunctional protein that regulates diverse cellular pathways, including cell cycle and apoptosis through the activation/inhibition of interacting proteins. The structure of Rb is organised in multiple domains, referred as RbN, pocket and RbC, connected by intrinsically disordered regions (**Figure 15**) [139]. The pocket domain is highly conserved throughout Rb paralogs and is necessary for E2F binding, but is not sufficient for maximal growth suppression, which in addition requires RbC. Different crystal structures of various domains have been solved [138], [140]–[142]. In particular, the crystal structure of the CC-MB domains of E2F1/DP1 (hereafter $E2F1^{CM}/DP1^{CM}$) in complex with the $RbC^{826-874}$ peptide demonstrated that the MB domains of the formed heterodimer associates with RbC and the involved contacts are conserved across E2F and DP families. Although the RbC domain is intrinsically disordered, ~30 residues undergo disorder-to-order transition, adopting a strand-turn-helix conformation upon binding to $E2F1^{CM}/DP1^{CM}$ heterodimer. The structure of the overall complex is asymmetric and reveals an intertwined heterodimer composed of an intermolecular CC and β-sandwich subdomains bridged by two small α-helices and two-stranded short β-sheets (**Figure 16D**). The requirement for DP to complete the hydrophobic core shed light on the importance of heterodimerisation for E2F stability, DNA binding and transcription activation [124][138].

The structures of $E2F^{CM}/DP^{CM}$ are conserved, where the crystal structures of $E2F4^{CM}/DP1^{CM}$ alone and $E2F5^{CM}/DP1^{CM}$ bound to C-terminal domain (CTD) of p107 reveal similar overall secondary structure topology compared to the previously determined $E2F1^{CM}/DP1^{CM}$ structure with a notable variation in the orientation of the CC domain relative to the β-sandwich domain. Structural alignment of the overall structures supports the conclusion that the CC domain is flexible and can rotate around a fixed contact point made with α2-helix in DP1 bridging the interaction with DNA and with other TFs bound to the MB domain or C-terminal regions in E2F (**Figure 16E**). Sequence and structural comparisons clarify the generality of the pocket protein CTDs to associate with E2F CC-MB domains and reveal molecular basis that explain the respective preferences of Rb for activating E2Fs and p107/p130 for repressive E2Fs. Particularly, sequence alignment of the human $E2F^{CM}$ reveals a potential protein interaction surface common to all E2F paralogs and formed by the last E2F strand (β5) and the preceding loop. The most notable divergent region which account for differences in specificity for different pocket proteins includes the end of β3 and the preceding loop (**Figure 16F**) [124].

1.6 Interplay between NF-Y and E2F1

It is well established that NF-Y plays a pivotal role in the activation of "cancer" genes. Transcriptome profiling of commonly up-regulated genes in tumour vs. normal cells and *de novo* motif discovery identified "signature" genes whose CCAAT, GC-rich and E2F-motifs are statistically enriched

in their promoters [143]. RNA profiling of patients with Diffuse Large B-cell Lymphoma (DLBCL), an aggressive progression of Follicular Lymphomas (FL), identified functional drivers of the FL transformation, including NF-YB and DP1 [144]. *In vivo* independent studies and ChIP-seq analysis using tiled ENCODE arrays identified frequent co-occupancy of E2F1/4 and NF-Y on gene promoters with wide-ranging functions in proliferation, cell cycle and metabolism, suggesting possible cooperative binding in the activation of the aforementioned promoters. Despite co-occupancy, genome-wide unbiased search for binding sites of NF-Y and E2Fs *in vivo* has not revealed a preferential positional-bias between the respective binding-motifs [40][73][145]. Genome-wide mapping of E2F1-binding sites from the ENCODE project from distinct cell lines revealed that ~90% of bound sites lack the consensus *in vitro*-derived motif, often absent, or retrieved at tens of bps distance from peak summits, suggesting tethered/assisted recruitment of E2F1 to core promoters lacking the E2F-consensus site [146].

Strikingly, the E2F1 promoter is regulated by NF-Y [147], and E2F1 regulates the NF-YB promoter [148]. Two independent studies have investigated the functional interplay between E2F1 and NF-Y. In the first study, ectopic expression of NF-YA in mouse quiescent fibroblasts induces an increase in E2F1 levels leading to apoptosis in an E2F1-and p53-dependent manner, impaired in cells lacking E2F1 and p53. In particular, unrestricted NF-Y activity utilises the E2F1-mediated phosphorylation of p53 pathway to trigger p53 activation and promotes NF-Y-dependent apoptosis [149]. In the second study, they identified an inverse relationship, where E2F1 increases the NF-YB levels, and jointly regulate a large group of anti-apoptotic genes that are usually overexpressed in sarcomas and are associated with chemotherapy resistance, limiting the pro-apoptotic functions of E2F1. They also provided evidence that NF-YB inactivation enhances E2F1-mediated apoptosis [148].

Notwithstanding, better understanding of the interplay between NF-Y and E2Fs at the molecular and structural level which represent the main topic of my thesis possibly will explain the coordinate overexpression patterns and the co-occupancy observed *in vivo*.

2. AIMS OF THE THESIS

My doctoral thesis aims, in general, to characterise at the biochemical, structural and functional levels the co-association patterns of NF-Y and some of its partners observed *in vivo*, including E2F1 and USF1. The unifying theme of this thesis is to uncover the molecular details of the crosstalk between NF-Y and *bona fide* interacting partners on DNA. With the long-term goal of controlling such TF interactions, we first focused on characterising single TF, and then extending our study to NF-Y interactions to construct a complete picture of the promoter architecture.

2.1 Interplay between NF-Y and E2F1

Given the abundance and functional importance of NF-Y and E2F sites in promoter of cancer signature genes, and the overlap of NF-Y and E2Fs genomic locations suggest that they potentially mutually interact [45][143]. Instead, no distance bias between CCAAT and E2F sites are found [40]. With the aim of controlling such TF interactions, we focused first on E2Fs, in particular E2F1 and its functional partner DP1, and then we extended our understanding to NF-Y interactions on DNA.

- **Characterisation of E2F1/DP1 using an integrated biochemical/structural approach**

The structural hallmarks of E2F and DP proteins consist of a DBD connected through a short linker to the CC-MB dimerization domain. The available 3D structures refer to the separated domains from distinct E2F and DP family members. To have a coherent view of the relative topology of the linked domains, the full conserved regions of E2F1 and DP1 are co-produced as soluble heterodimer, and characterised using integrative structural biology techniques, including X-Ray crystallography, SAXS and Molecular Dynamics simulations.

- **Dissect synergistic and functional interactions between NF-Y and E2F1 on DNA**

With the aim of dissecting synergistic binding between NF-Y and E2F1 on DNA , we utilised the human CDC2 promoter, shown to be functionally dependent from both CCAAT and E2F-binding sites [150]. We applied EMSA experiments using purified recombinant proteins to (1) quantitatively and qualitatively measure the binding of single proteins and complexes, (2) determine synergistic binding by comparing the stability of single TF vs. multi-TFs complexes, (3) identify stable higher-order complexes that could be assembled and isolated *in vitro* for further structural characterisation to analyse TFs-contacting interfaces.

2.2 Uncovering novel structural information of NF-Y using X-Ray crystallography and Cryo-EM

- **Design optimised NF-Y constructs for structural and TFs partnership analyses**

Beside the previously crystallised NF-YB/NF-YC minimal core domains [67][72], and to possibly gain novel structural information on additional portions, we setup co-expression screening tests with different

combination of NF-YB and NF-YC deletions maintaining the conserved regions to improve protein folding and solubility. Among the designed and produced constructs, we have chosen the largest proteins suitable for Cryo-EM analysis, holding the TADs that could be involved in possible protein-protein interaction and DNA-binding cooperativity, while constructs encompassing predicted structured and ordered regions were used in crystallisation screening. As attempts to obtain protein crystals using traditional crystallisation trials failed, we applied *in situ* proteolysis, a rescue strategy based on crystallising protease-resistant polypeptide fragments on the timescale of crystal formation. We started by a pilot screening to identify the optimal conditions to yield a stably truncated form of the protein. To increase the likelihood of successful structure determination, we have combined limited proteolysis and mass spectroscopy to generate optimal constructs for subsequent crystallisation trials.

- **Analysis of NF-Y/USF1 contacting surfaces using Cryo-EM**

Preliminary biochemical, structural and functional data on USF1 and NF-Y from our lab provide the first look at the region involved in the protein-protein interaction [151][152]. To possibly gain defining details on the specific protein interfaces involved in this interaction, we set up new approach to assemble and isolate ternary complexes on different DNA configurations harbouring full-length proteins, suited to be visualised using single-particle Cryo-EM. My contribution to the project aimed to (1) co-produce soluble full-length NF-YB and NF-YC subunits, (2) set-up and optimise new approach to assemble and isolate multimeric complexes on DNA together with *A. Bernardini*. This approach consists of defining an efficient co-expression strategy for high throughput recombinant protein production in *E. coli*, and then passing through sequential chromatography steps to ensure high degree of sample purity and homogeneity, followed by functional characterisation using EMSA assays.

3. RESULTS, DISCUSSION AND FUTURE PERSPECTIVES

PART I

Interplay between NF-Y and E2F1

The main project consists of an integrated biochemical/structural approach targeting the interplay between NF-Y and E2F1. We started by characterising E2F1 and its functional partner DP1 and then we extended our approach to NF-Y and E2F1 partnership on natural promoter. The main findings of this part are included in the attached manuscript which has been recently submitted for publication.

1. Characterisation of E2F1/DP1 using integrative structural biology techniques

The evolutionary conserved domains of the canonical E2F and DP family members consist of a winged-helix DNA Binding Domain (DBD) and a Coiled-Coil-Marked-Box (CC-MB) heterodimerization domain tethered by a short linker (**Figure 17A**). So far, the structural information of the separated domains from distinct E2F and DP family members have been determined by X-ray crystallography [133][138]. To further characterise the global structure, dynamics and the relative orientations of the linked system, we applied an integrated approach that combines SAXS experiment and molecular dynamics (MD) simulations.

1.1 Main Results

1.1.1 Co-expression of E2F1 and DP1 yields a soluble heterodimer that binds efficiently the DNA

We started to produce recombinant E2F1 and DP1 proteins that include the full conserved DBD-CC-MB domains (here-after $E2F1^{DCM}$ and $DP1^{DCM}$). Because singularly, $DP1^{DCM}$ is insoluble and $E2F1^{DCM}$ is soluble (data not shown), proteins were co-expressed in *E. coli* and purified to homogeneity by exploiting a N-terminal 6xHis-tag fused to the DP1 subunit, yielding soluble heterodimer (**Figure 17B**). As observed in **Figure 17C**, the purified $E2F1^{DCM}/DP1^{DCM}$ binds efficiently DNA as assessed by EMSA on the canonical E2F site of the human E2F1 promoter, while $E2F1^{DCM}$ alone binds DNA, probably as homodimer [114][153], with extremely low affinity.

Next, a set of preliminary crystallisation trials were conducted on ternary complexes with a 15-mer DNA oligo of the same length as the one used to obtain the $E2F4^{DBD}/DP2^{DBD}$/DNA crystals [133]. The trials were not successful, also with the use of longer DNA and in the presence of RbC peptide used to crystalize the $E2F1^{CM}/DP1^{CM}$/RbC complex [138]: quasi crystals that are not suitable for X-Ray diffraction were obtained. Because using single structural technique to characterise the structure of macromolecule complexes is a very arduous task, we tried to combine experimental SAXS analysis with computational modelling to characterise such challenging complexes.

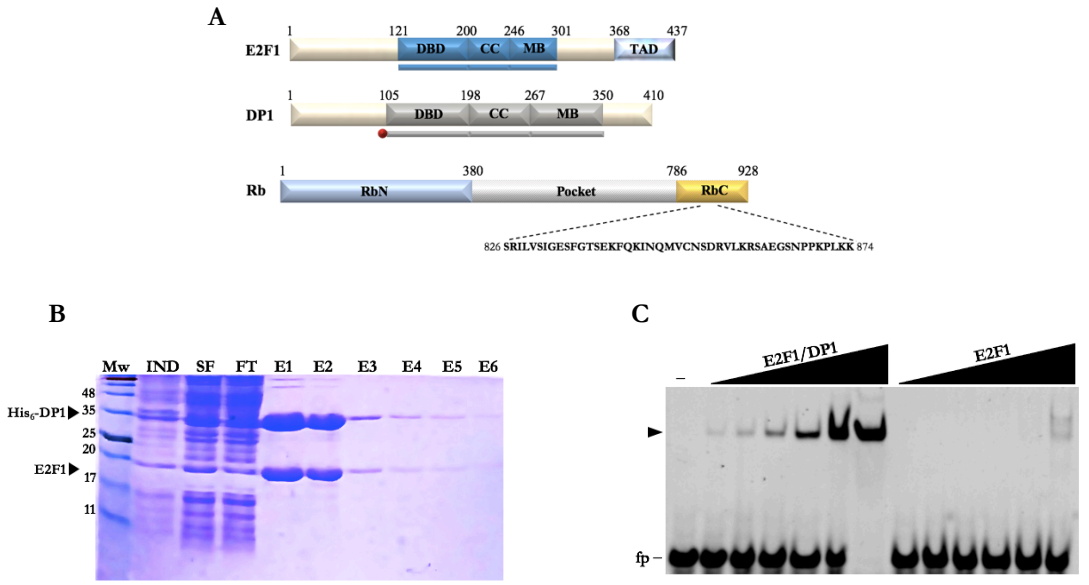


Figure 17. **Characterisation of purified recombinant E2F1^{DCM} and DP1^{DCM} proteins. (A)** Schematic representation of E2F1 and DP1 domain organisation with their respective amino acid sequence numbering, and recombinant protein constructs used in this work. The conserved domains of E2F1 and DP1 are depicted in blue and grey, respectively. The 6xHis-tag is indicated as a red sphere. The RbC peptide used is indicated. **(B)** Coomassie stained gel of co-expression and IMAC purification of recombinant E2F1^{DCM}/DP1^{DCM} heterodimer. Total extracts of the induced culture (IND), the soluble fraction (SF), Flow through (FT) and elution (E1-6) samples were loaded. Molecular weight (Mw) markers (kDa) and bands identity are indicated on the left side of the gel. **(C)** Dose response EMSA experiment performed with recombinant E2F1^{DCM} alone and in the presence of DP1^{DCM} on the E2F1 probe. The proteins were incubated at increasing concentrations (4, 8, 16, 26, 64, 128 nM) with the probe (20 nM). Arrowhead: E2F1^{DCM}/DP1^{DCM} and E2F1^{DCM} complexes; fp: Free probe. (-) probe alone, without protein additions.

1.1.2 Solution structural analysis of E2F1/DP1 complexes reveals high conformational flexibility

Solution-based Small-Angle X-Ray Scattering (SAXS) is an established method that utilises the scattering of X-Ray beams to investigate the overall dimension, low-resolution shape and dynamics of biomolecules in solution at the nanometric scale. Expanding the investigation range of high-resolution structural techniques, SAXS analysis directly provide useful information on flexible regions. To overcome sample polydispersity in SAXS data analysis that can easily lead to data misinterpretation, we performed Size-Exclusion Chromatography (SEC) in-line with X-Ray data collection, providing both a more homogenous sample and a continuous set of data as the concentration is extrapolated to zero. Two data sets were collected of E2F1^{DCM}/DP1^{DCM}/DNA with and without the RbC peptide [138] in collaboration with *M. Nardini* and *A. Chaves* at the Diamond Synchrotron. The values of the radius of gyration (R_g), the maximum diameter (D_{max}) and the molecular weight (MW) derived from the scattering curves are represented in **Table 1**. The R_g and the forward scattering intensity I(0) traces as a function of measurement frame show that the samples are monodispersed (**Figure 18A, B**). Data frames under the elution peaks centred on the maximum in I(0) where the R_g values were the same within error and statistically identical as assessed using CorMap goodness-of-fit test and reduction χ^2-test (p-value > 0.01) [154] were selected and averaged for subsequent analyses. The log I(q) vs. scattering angle (q) plots

provide clear representation of the buffer-subtracted data over the entire q range, with the Guinier plots (ln I(q) vs. q^2; [155]) shown as inset (**Figure 18C, D**). The minimum q measured (0.008 $Å^{-1}$) is well below the recommended q of $\simeq \pi/D_{max}$ for accurate assessment of the particle size and shape. Indeed, the Guinier analysis show linearity in the small q regions ($qR_g < 1.3$) and the double logarithmic plots (**Figure 19C**) show zero slopes at low q values, indicating that the solutions contained monodisperse particles with similar size. Overall, these measurements provide confidence that the selected frames are free of significant aggregation.

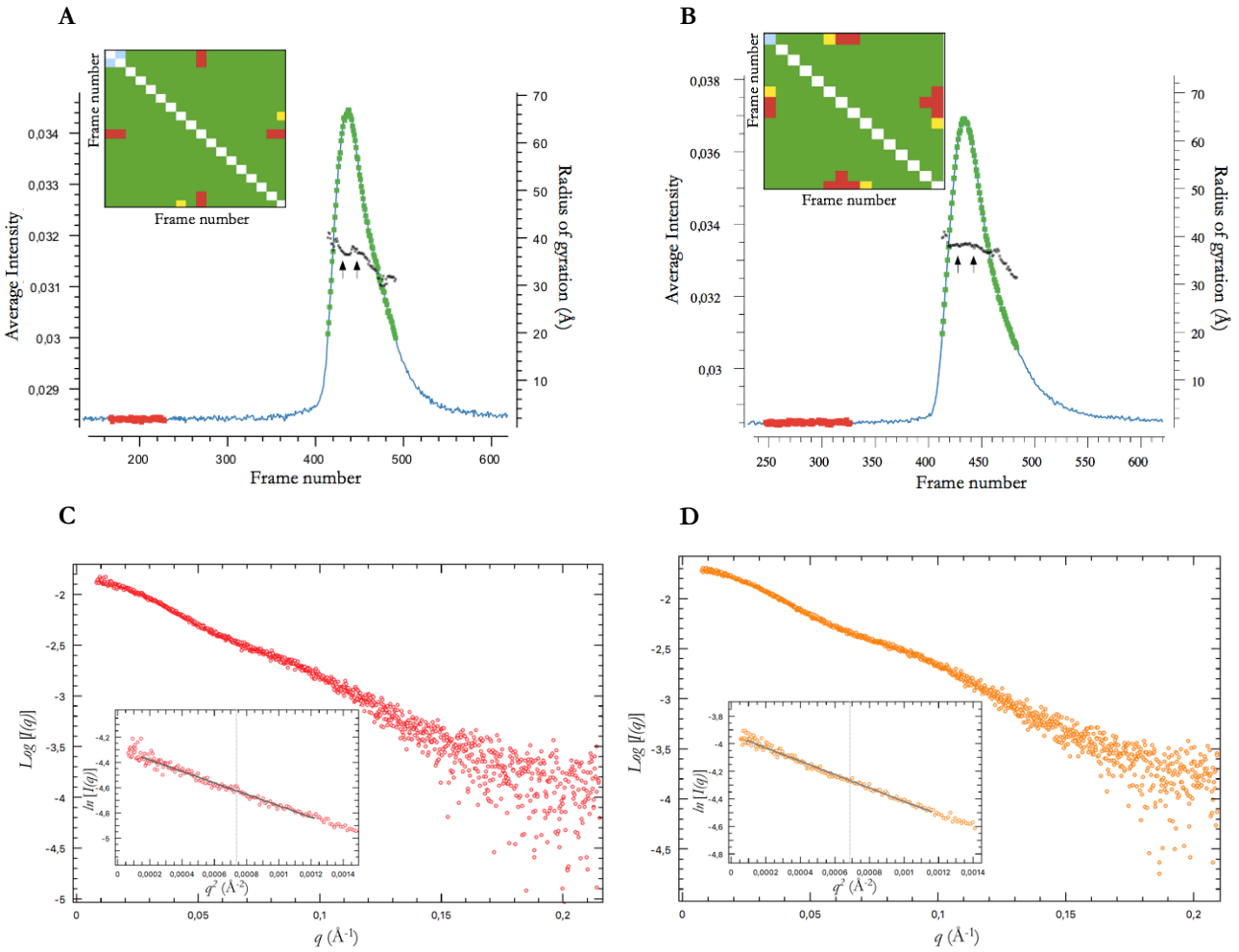


Figure 18. SEC-SAXS data reduction. SEC-SAXS chromatograms of E2F1/DP1/DNA complexes in the absence **(A)** and presence **(B)** of RbC peptide, showing the average scattering intensity (blue line) and the radius of gyration (black squares) as a function of measurement frame for the SEC-SAXS run. The frames corresponding to the buffer and sample are marked by red and green, respectively. Data frames between the two arrows were selected and averaged for subsequent analyses with the insets showing the results of two-dimensional correlation maps (CorMap) for frames in the region of interest where green and red squares stand for matching ($p \geq 0.9$) and non-matching frames ($p \leq 0.1$), respectively, and yellow squares correspond to all other cases [154]. **(C)** and **(D)** are plots showing I(q) vs. q as semi-logarithmic scale for E2F1/DP1/DNA complexes in the absence (red curve) and presence (orange curve) of the RbC peptide, respectively, with Guinier fits (grey line) for $qR_g < 1.3$ are shown as insets. For better visualisation, the profiles have been displaced along the ordinate.

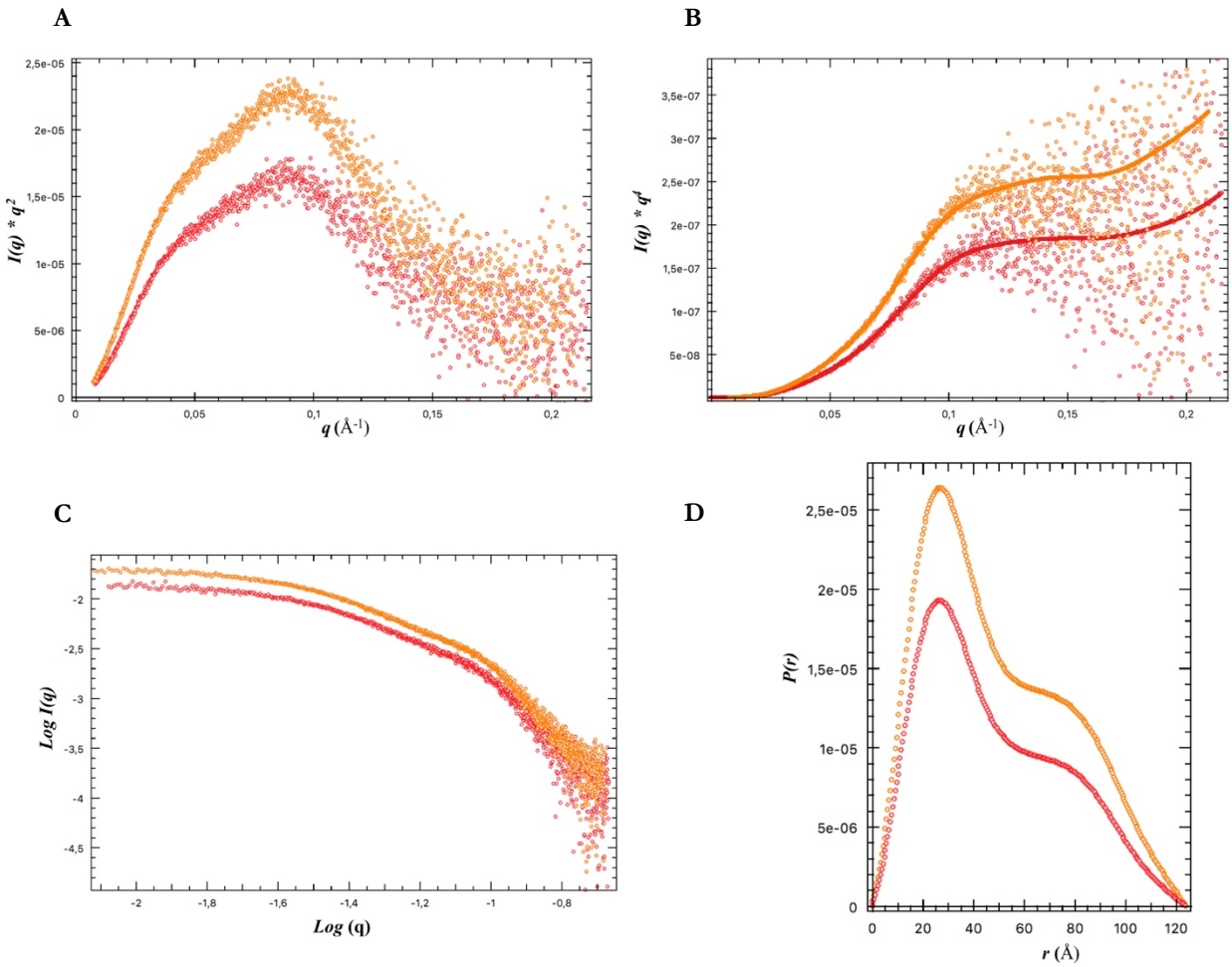


Figure 19. Experimental SEC-SAXS profiles for E2F1/DP1/DNA complexes. **(A)** Kratky plots for the SAXS data represented as I(q) * q^2 vs. q (Å^{-1}). **(B)** Porod plots of SAXS data transformed as I(q)* q^4 vs. q (Å^{-1}). Trend lines are given as a guide to the eye. **(C)** Double logarithmic scale showing near zero slope at low angle. **(D)** P(r) distance distribution functions vs. r profiles. Colour key is as in figure: +RbC (orange) and -RbC (red).

To qualitatively evaluate the protein compactness and the presence of disordered regions, Kratky plot analysis was employed [156]. Typically, for a compact and well-folded single-domain protein, the Kratky plot displays a bell-shaped curve, with a well-defined maximum [157]. The generated plots obtained from SEC-SAXS data (**Figure 19A**) have distinct profiles, where the maximum is shifted upwards and right indicating an anisotropic shape. The shallow oscillation reflects the well-folded two-domain structure connected by a flexible linker [158]. At large q values, the intensity rises increasingly indicating significant flexibility. In SAXS experiment, the Porod-Debye low is an alternative powerful tool for assessing macromolecule flexibility [159]. While Kratky analysis relies on the measurements at high q which are exponentially more sensitive to noise and buffer subtraction, the Porod-Debye region resides near the Guinier region that is not prone to buffer subtraction issues [160]. For most folded proteins, Porod curve displays a defined plateau indicating that the scattering particles exist with an homogenous electron density contrast between the solvent and particle [159]. As observed in **Figure 19B**, for both states, Porod-Debye plots (q^4 * I(q) vs. q) exhibit shape-dependent oscillation features. Significant flexibility is indicated by the continuous increase in intensity at high q values that is similar in the presence of the RbC peptide, indicating that the peptide used has no impact on protein flexibility.

Using both Kratky and Porod analyses, similar conclusions can be drawn regarding flexibility. The P(r) distance distribution function vs. r profile are well behaved (**Figure 19D**), showing a peak at low r values, corresponding to the intradomain distances, and a long trailing tail approaches zero gently with pronounced shoulder corresponding to the interdomain distances, expected for an elongated dumbbell-shaped structure [161]. For both complexes, the estimated MW values [162]–[164] are within the expected range based on the chemical composition (**Table 1**).

The scattering profiles of E2F1/DP1/DNA and E2F1/DP1/RbC/DNA complexes are superimposable, suggesting a marginal influence of the RbC peptide on the overall shape and flexibility of the E2F1/DP1 heterodimer (**Figure 18, 19 and Table 1**).

Table 1. SEC-SAXS data collection and analyses.

Data collection parameters	
Beamline	Diamond, B21
SEC system/column	Agilent 1200 HPLC system/Superdex 200 Increase 3.2/300 (GE Healthcare Life Sciences)
Sample Volume (µL)	45
Sample concentration (mg mL^{-1})	4
Sample buffer	50 mM Tris-Cl pH 8.0, 150 mM NaCl and 2 mM DTT
Flow rate (mL min^{-1})	0.075
Beam defining slits size (mm)	4.6 x 1.1
Sample- Detector Distance (mm)	3718.9376
Wavelength (Å)	0.94644433
q measurement range ($Å^{-1}$)	0.0032-0.3800
Beam size	1 x 0.25 mm (at sample)/50 x 50 µm (at detector)
Exposure time	20 exposures per second
Temperature (K)	293
Sample cell size (quartz capillary)	1.5 mm in diameter (10 µm thick walls)

Structural parameters [a]	**E2F1/DP1/DNA**	**E2F1/DP1/DNA/RbC**
Guinier analysis		
I(0) (cm^{-1})	0.014 ± $6.2e^{-05}$	0.02 ± $6.2e^{-05}$
R_g (Å)	37.15 ± 0.25	38.18 ± 0.18
q_{min} ($Å^{-1}$)	0.0083	0.0075
qR_g range	0.31 – 1.28	0.29 – 1.30
p(r) analysis		
I(0) (cm^{-1})	0.01392	0.01976
R_g (Å)	38.85	39.86
D_{max} (Å)	123.81	123.86
Total-quality-estimate (GNOM estimate)	0.7273	0.79
Volume, shape, and molecular weight (MW) analysis		
Porod volume ($Å^{-3}$)	77314	80648
MW calculated from chemical composition (Da)	59150	64290
MW from Vc (Da) (ratio to predicted)[b]	58334 (0.98)	61097 (0.95)
MW from SAXSMoW (Da) (ratio to predicted) [b]	62400 (1.05)	66400 (1.03)
MW from consensus Bayesian assessment	59500	62350
Credibility Interval [Da]	[57500, 63100]	[61600, 69650]
Credibility Interval Probability [%]	94.28	95.46

[a] ATSAS-3.0.1-1 version [165] employed for SEC-SAXS data reduction and analyses.
[b] The experimental MW ratio relative to the calculated MW from the amino acid sequences.

1.1.3 SAXS-restrained MD simulations generate reliable conformational ensembles

SAXS experiments together with preliminary MD simulations highlight the dynamic nature of this system. As conformationally flexible protein cannot be represented as single protein structure, SAXS-guided MD simulations were performed in collaboration with *C. Paissoni* and *C. Camilloni* to generate a conformational ensemble that better describe this dynamic system. The method and the strategy

employed are detailed in the attached manuscript. Briefly, using homology modelling, initial tentative models of $E2F1^{DCM}/DP1^{DCM}$ complexes were built from the available separate structures of $E2F4^{DBD}/DP2^{DBD}$ on DNA (PDB ID: 1CF7 [133]) and $E2F1^{CM}/DP1^{CM}$/RbC (PDB ID: 2AZE [138]) (**Figure 20A**). Representative models with good agreement with experimental data were obtained by employing SAXS-guided Metadynamics, an extensive and enhanced conformational sampling. The generated models were then refined against SAXS data through SAXS-driven MD simulations (**Figure 20B, 21A**). Remarkably, models suggest a preferential orientation of the CC-MB domains with respect to the DBDs, with the CC bent toward the DP1-bound side of the DNA.

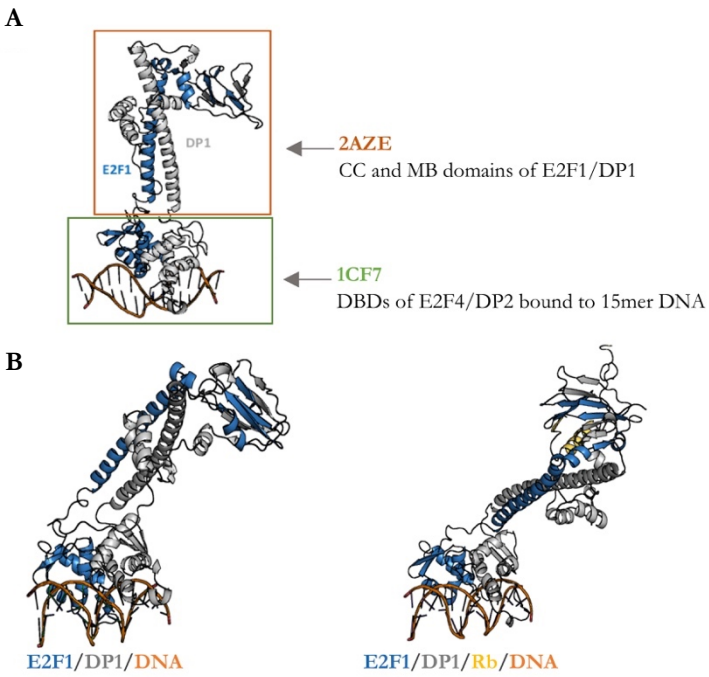


Figure 20. Building a representative structural models of E2F1/DP1/DNA complexes. (A) Building a tentative starting model with homology modelling based on the available PDBs (1CF7 [133] and 2AZE [138]). **(B)** Refined models with (right panel) and without the RbC peptide (left panel) obtained with SAXS-driven MD simulations. E2F1, DP1 DNA and RbC peptide are depicted in blue, grey, orange and yellow, respectively.

Representative structural ensembles with reasonable agreement with experiments were generated by performing metainference MD simulations (**Figure 21A**). Simulations unveil the inherently flexible and heterogenous nature of this system and point toward multiple regions with high conformational freedom (**Figure 21B**): the inter-domain linkers, the DP1 α2-α3 loop (aa 148-161), the DP1 N-terminal tail and to lesser extent the E2F1 N-terminus. Particularly, the unstructured linker regions confer great flexibility between domains, allowing the CCs to adopt various relative orientations with respect to the DBDs. In spite of this significant inter-domain flexibility, the CC-MB domains adopt a preferential spatial orientation on DNA toward the region occupied by DP1. As observed in **Figure 21A**, SAXS profiles of the generated ensembles agree with the experimental data and SAXS-refined single structure within the estimated experimental errors. Additionally, the N-terminal tail of the DP1 DBD which displays poor secondary structure content exhibits increased mobility. Considering the numerous basic residues, longer DNA could induce a stronger binding, facilitating DNA recognition outside the conventional RRXYD motif.

All these procedures were applied in parallel for the E2F1^{DCM}/DP1^{DCM}/RbC/DNA complex, yielding superimposable conclusions regarding flexibility (**Figure 20B and 21C**). In line with the first SAXS observation, MD simulations confirm that the RbC peptide has negligible influence on the overall structure, the efficiency of DNA-binding, or on the flexibility of the CC-MB portion of the complex.

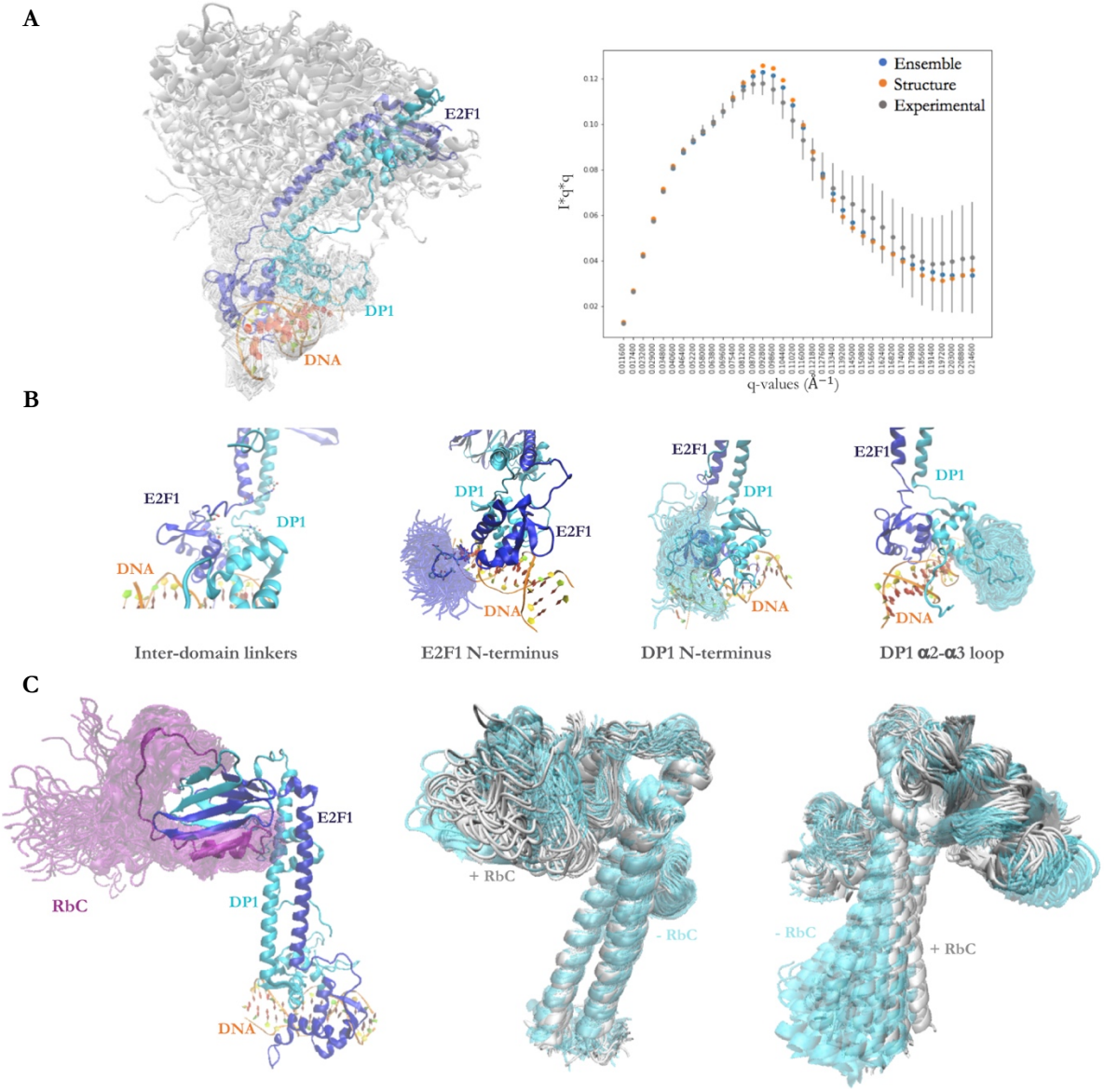


Figure 21. SAXS-driven MD simulations of E2F1/DP1/DNA complexes. (A) Conformational ensembles of E2F1/DP1/DNA complex (left panel), with SAXS curves (right panel) showing the agreement of the ensembles (blue dots) with the experimental SAXS data (grey dots) and SAXS-refined single model (orange dots) within the experimental errors. **(B)** Zoomed-in view showing the flexible regions of E2F1 and DP1 according to simulations. **(C)** The E2F1/DP1/RbC/DNA model (left panel), with the volume occupied by the RbC peptide shown as violet cloud. The negligible role of the RbC peptide on the structure and dynamics of the dimer is illustrated in the overlaid models (middle and right panels).

1.1.4 DNA length defines extended contacts on the DP1 side of the bound DNA motif

Simulations analyses indicate that the N-terminal tail of DP1 is highly flexible and disordered. In view of the numerous exposed basic residues, an extended DNA on the DP1 side could induce stronger binding by facilitating DNA recognition and binding beyond the conventionally defined RRXYD motif.

To define whether extension of DNA may support the predictions of MD simulations, we assayed by *in vitro* competition EMSA experiments the affinity of E2F1/DP1 for unlabelled oligos of different length. In particular, we assessed the ability of each oligo to compete for protein binding to a 25-mer labelled probe derived from the human E2F1 promoter, as compared to the WT 15-mer E2F1 oligo. Given that the probe used is essentially symmetrical, in order to discriminate among the E2F1 and DP1 contributions, we started with the defined non-symmetrical E2F DNA element of 15 bps (oligo 1; **Figure 22A**) used to crystallise the E2F4/DP2/DNA complex (PDB entry 1CF7; [133]). Five bps extensions with a random sequence on either side of the oligo were designed, and unlabelled competitors were used to challenge E2F1/DP1 binding on the 25-mer probe (**Figure 22A**). The competition rate of the 15-mer WT oligo derived from the probe was set as 1. The unlabelled 25-mer oligo, used as a positive control, showed higher competition rates, i.e., higher affinity, indicating that regions outside the 15-mer oligo indeed contribute to stabilize E2F1/DP1 binding on DNA. When the Oligo 1 (15-mer) was used, possibly also due to the asymmetrical nature of competitor, a substantial decrease in competition rates were observed (**Figure 22B**). However, both (20-mer) 5 bps-extended versions of this oligo, on either DP1 or E2F1 sides (oligos 3 and 4, respectively) compete extremely well, showing an increased competition rate as compared to the WT oligo, which was more pronounced using oligo 3 (on the DP1 side), and also evident for the 18-mer oligo with a 3 bps extension (oligo 2). To evidence the contribution of both E2F1 and DP1 to the DNA binding affinity, we designed an optimised competitor (oligo 5) where one and four bps were added at the E2F1 and DP1 sides, respectively. As observed, an enhanced binding affinity was gained compared to the previous extensions, indicating that the base pairs flanking the core DNA element modulate the affinity of the protein-DNA complex. In particular, the effect of the flanking sequences is asymmetric, with a stronger effect for the bases flanking the DP1 side. The DNA sequence on the 3' extension did not influence competition rates, as a 20-mer oligo in which the five additional bases on the DP1-side were all substituted to introduce transitions from A/T to G/C competed in levels similar to that obtained for the "wild-type" oligo 3 (1.27 and 1.30 competition rates, respectively).

To further investigate whether the binding of DP1 to the core element flanking region displays sequence-specificity, we designed a set of unlabelled oligos in which the three nucleotides at the 3' flanking core motif were mutated (oligo A-D; **Figure 23**), based on the 18-mer oligo (oligo 2) that was sufficient to retain high competition rates. All mutations in the downstream flank show a negligible effect on complex formation (similar competition rates), indicating no sequence preference (**Figure 23**).

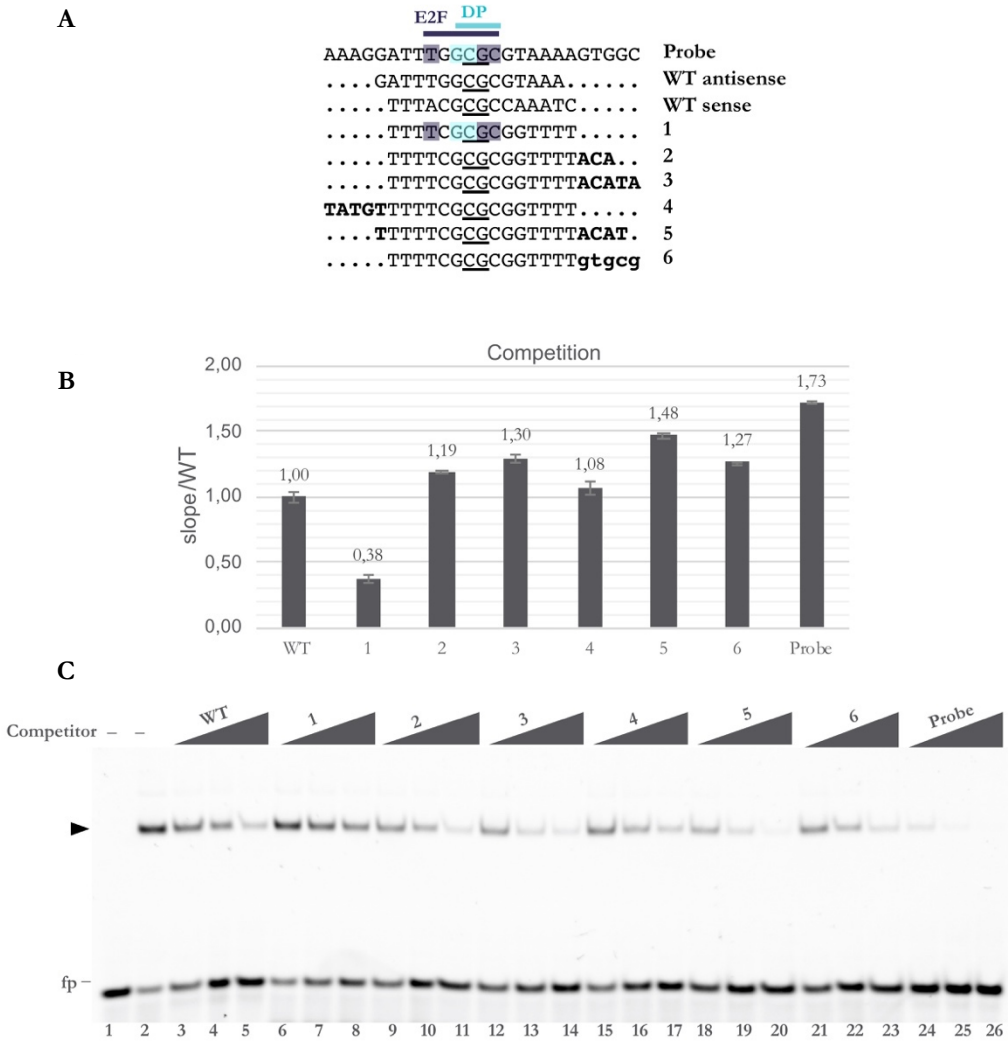


Figure 22. ***In vitro*** **validation of DP1 stabilisation by competition EMSA experiments. (A)** Sequence details and the alignment of the probe and the designed unlabelled oligonucleotides -centred on the core CG bases (underlined)- annealed with their complementary sequences and used in competition EMSA. The two-potential binding-orientations for the WT oligo (sense and antisense strands) are shown. The 5' or 3' sequence extensions are highlighted in bold letters. The DNA bases contacted by E2F and DP proteins [133], and corresponding positions on the E2F1 probe, are highlighted in violet and cyan, respectively. Mutant nucleotides are indicated in lowercase. **(B)** Direct competition EMSAs quantification. The indicated unlabelled competitors were incubated at increasing concentrations (5x, 25x, 100x fold excess) with the E2F1-25 bps probe (20 nM), with the addition of recombinant E2F1/DP1 heterodimer (15 nM) and loaded for electrophoresis. The bound DNA was quantified in each dose curve data point as described in Materials and Methods. Competitor efficiencies (competition) displayed in the bar graph are the mean of at least three independent experiments ± SD (error bar). The data are plotted setting as 1 the WT (15mer) oligonucleotide competition slope. **(C)** Representative gel of direct competition EMSA experiment. As positive control, the wild-type 25-mer E2F1 unlabelled oligo (Probe) was used as competitor. Lane 2: $E2F1^{DCM}/DP1^{DCM}$ was incubated alone with the probe. Lane 1: probe alone, in the absence of $E2F1^{DCM}/DP1^{DCM}$ addition. On the left side of the gel, an arrowhead indicates the $E2F1^{DCM}/DP1^{DCM}$/DNA complex migration; fp: Free probe.

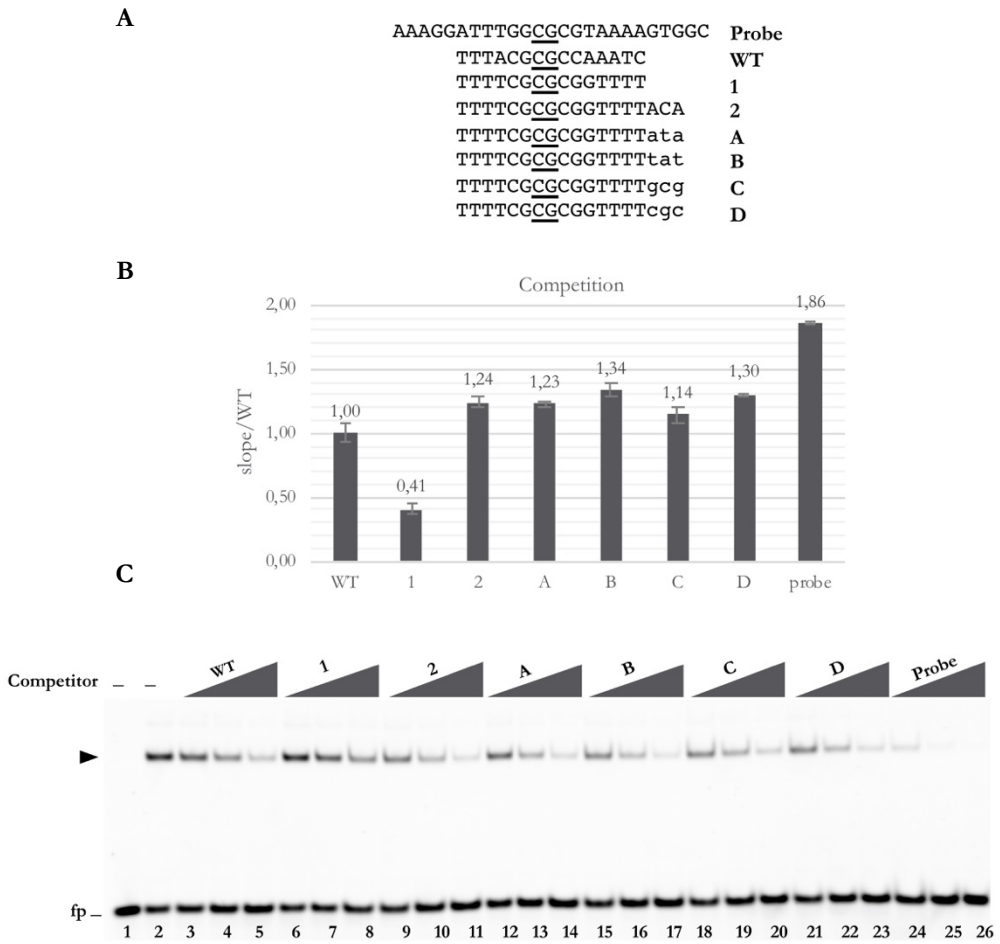


Figure 23. **Competition EMSA to dissect a preference for specific sequences. (A)** The probe and unlabelled oligonucleotide sequences centred on the core CG bases (underlined) used in competition EMSA with mutated bases are denoted by lower case letters. **(B)** Direct competition EMSAs quantification. **(C)** Representative gel of direct competition EMSA experiment, refer to **Figure 22B** for more details.

1.2 Discussion and Future Perspectives

In cells, proteins frequently consist of multiple domains joined with flexible inter-domain linkers, whose composition, length and conformational preferences allow them to execute tailored and specific biological functions [166][167]. Due to the inherent freedom of interdomain linkers, deciphering the structural characteristics of these macromolecule assemblies has been challenging by traditional structural biology approaches. Instead, integrative structural biology techniques, which combine low-resolution data and sophisticated computational procedures represent powerful tools to reveal, at high resolution, the conformational dynamics of such challenging macromolecules [168]. Illustrative of this are the E2F and DP family members that share an evolutionary conserved DBD bridged to the CC-MB heterodimerization domains through short disordered linkers. While the available structures refer to the separated domains, the three-dimensional organisation of the tethered module has remained elusive. In this work, we have relied in a combination of homology modelling, enhanced sampling MD and an

integrative modelling approach based on SAXS recently developed by [169] to characterise the structural dynamics of E2F1^{DCM}/DP1^{DCM}/DNA complex. MD-refined structural ensembles, fitting SAXS data, highlight high conformational freedom with pronounced conformational fluctuations in the following regions: (1) the inter-domain regions, (2) the region encompassing residues 148-161 of DP1 α2- α3 loop and (3) the N-terminal tails of the DBDs. This dynamic system behaved evenly in the presence of the RbC peptide, giving essentially superimposable results, indicating marginal impact of the RbC peptide on the overall topology, flexibility and the relative orientation.

Despite the high inter-domain freedom, likely due to the flexibility of the linker regions, we found a precise directionality and a preferential orientation of the CC domains relative to the DBDs toward the DNA side occupied by DP1, potentially facilitating interactions with neighbouring binding partners and/or coactivators. Regarding potential interaction partners of EF/DP, bioinformatic analysis of E2F1 RNA-seq data from ENCODE unveiled a global enrichment of E2F1 sites with AP1 sites of b-ZIP TFs and E-box motif of b-HLH TFs [170]. These observations fit with detailed functional studies that have investigated the functional interplay between the proto-oncogene Myc, belonging to the b-HLH family members, and E2F1-3 in regulating cell cycle progression [171]. Indeed, the Myc promoter is regulated by E2F1-3 [172], and Myc binds and regulates the E2F1-3 promoters [106]. Nevertheless, the precise mechanism of the E2F-Myc functional interplay awaits further studies, providing new opportunity for cancer therapies.

Our data have shown that the lateral flexibility of the CC-MB moiety of the dimer is centred upon the flexible linkers joining the structured domains. Remarkably, sequence alignments of human E2F and DP proteins (**Figure 24**) show that the linker region is highly conserved throughout the DP proteins. Instead, for canonical E2Fs, the linker region is divergent in both length and amino acid content, suggesting that flexibility is possibly governed by the E2F protein partner while DPs appear to limit this flexibility. To investigate the functional implications of this aspect, we will pursue studies on E2F/DP mutants, carrying relevant point mutations in the linker regions (**Figure 24, 25**), and we will evaluate whether the E2F subunit (i.e. E2F4) may differently contribute on the overall flexibility of the complex in future SAXS-structural studies.

In addition to the inter-domain movement, the DP1 loop between $\alpha 2$ and $\alpha 3$ is considered a highly dynamic region, displaying poor propensity to stabilise secondary structure according to simulations (aa 148-161, **Figure 25B**, lower panel). Herein, we noted that this region shows a relatively low degree of sequence similarity to the homologous DP2 protein (**Figure 24**), suggesting possible docking sites for DP1-specific partners. We noticed that the corresponding area in E2Fs is shorter and highly conserved (aa 162-166, **Figure 24A**, lower panel), with no detectable flexibility according to our simulations, suggesting little flexibility in all E2F members.

Figure 24. Alignment of the canonical E2F and DP sequences, including the DBD (blue), linker region (violet) and the CC-MB domains (green). The α2-α3 loops and the flexible N-terminal tails, belonging to the DBDs are highlighted in yellow and red, respectively. Dark shading indicates identical residues.

Additional dynamics are observed at the DP1 N-terminal region of the DBD, and to lesser extent the E2F1 counterpart, showing poor secondary structure content (**Figure 25**, lower panels). We noticed that the highly dynamic N-terminal tail of DP1 (residues 105-113; RNRKGEKNG, plus five residues derived from thrombin cleavage) contains numerous basic amino acids, which could make additional DNA contacts to longer DNA, boosting the affinity of E2F1/DP1 for DNA. This was experimentally verified by competition EMSAs, showing that DNA oligos extended on the DP1 side substantially facilitate DNA recognition and binding outside the conventional motif, and indeed addition of nucleotides on the E2F1 side had less effect on DNA-binding stabilisation. We further employed binding assays to investigate whether the binding of DP1 to the region adjacent to the core-recognition site (DP1 side) is sequence specific: in these experiments, no significant sequence preference was observed. Moreover, we could identify an optimal 20-mer DNA sequence that can be used to increase the likelihood of obtaining diffraction-quality crystals. Notably, the CC-MB domains incline in the same direction of the flexible DP1 N-terminus and α2-α3 loop, suggesting that this arrangement could facilitate intermolecular interactions on this side of the complex.

The dynamic flexibility has emerged as a compelling strategy that confer greater adaptability for proteins to interact with numerous molecular partners and to exert their manifold cellular functions in a strictly regulated manner. Typical of this is the intricate relationship between E2F/DP and the CCAAT/enhancer-binding protein α (C/EBPα), family of b-ZIP TFs and a key regulator of cell proliferation and arrest differentiation of neutrophils and adipocytes [173]–[175]. Previous studies have shown that the C/EBPα interacts with the conserved N-terminal region contained in DP1 and DP2

(corresponding to aa 105 to 127), abrogating the binding of C/EBPα to DNA and transactivation of C/EBP−target genes [175]. Strikingly, as both the DP1 N-terminal residues (105-127) and the bZip C/EBPα are highly basic, we could suggest that binding of C/EBPα and E2F/DP to their cognate sites is competitive and mutually exclusive. This suggestion is in agreement with previous findings, showing that the reduced affinity toward DNA observed for the C/EBPα BRM2 mutant, associated with acute myeloid leukemia (in which residues of the basic region are mutated) was restored upon E2F1 or DP1 knockdown [175].

Although the E2F pathway is frequently deregulated in cancer, E2F1 and DP1 are rarely mutated in tumour cells, and few somatic mutations have been described in literature [176][177]. Here, we explored and mined the currently available databases derived from large-scale exome sequencing of human tumours that enable the identification of putative cancer drivers. In particular, somatic mutations in the evolutionary conserved regions of E2F1 and DP1 were retrieved from BioMuta v4.0, a cancer-associated single-nucleotide variations public database compiled from various sources, including COSMIC, CCLE, TCGA and ICGC [178]–[182]. As observed in **Figure 25**, mutations in DP1 are more numerous than those of E2F1, reflecting the vital role of E2F1, and the inability of cells to tolerate either loss or gain-of-function mutants. By virtue of its role as a predominant E2F partner under both normal physiological conditions and in tumours, mutations in DP1 could impact on and deregulate E2F activity, contributing to human cancer [177]. The location of the mutations occurred throughout $DP1^{DCM}$ and $E2F1^{DCM}$, although were particularly prevalent in the DBDs. Notably, many of the affected residues are located within or flanking the described flexible and unstructured regions, including residues belonging to the flexible linkers (T195I and G201V in E2F1 and T196 in DP1); the N-terminal tails (S121F, E124Q, R127H/L in E2F1 and R105C/H, K108N, G113D/S, G115D, R117W in DP1); and the flexible DP1 α2-α3 loop (D150N/Y, I153T, P155Q, A159S). In light of the essential role of dynamics in protein function, any molecular change caused by mutations might drastically distort protein's intrinsic dynamics, therefore the protein functional activity [183]. As flexible regions in E2F1/DP1 contain various motives, allowing multiple interactions with TF partners (some of them are listed in **Figure 25**) and contributing to DNA binding outside the common RRXYD recognition motif, the presence of pathological mutations in these regions could lead to distorted protein dynamics that plausibly affect protein function. This observation highlights and rationalises the functional importance of somatic mutations on dynamics-neighbouring residues in cancer. In order to investigate the molecular mechanisms underlining these mutations, we will analyse the impact of putative cancer-driver mutations in dynamics-neighbouring residues on the properties of E2F/DP heterodimers, including stability, interaction, and conformational dynamics.

Taken together, our results provide insight into the full-length structure of the conserved E2F and DP regions, a crucial step toward elucidating their functions and exploiting druggable protein sites to regulate these functions.

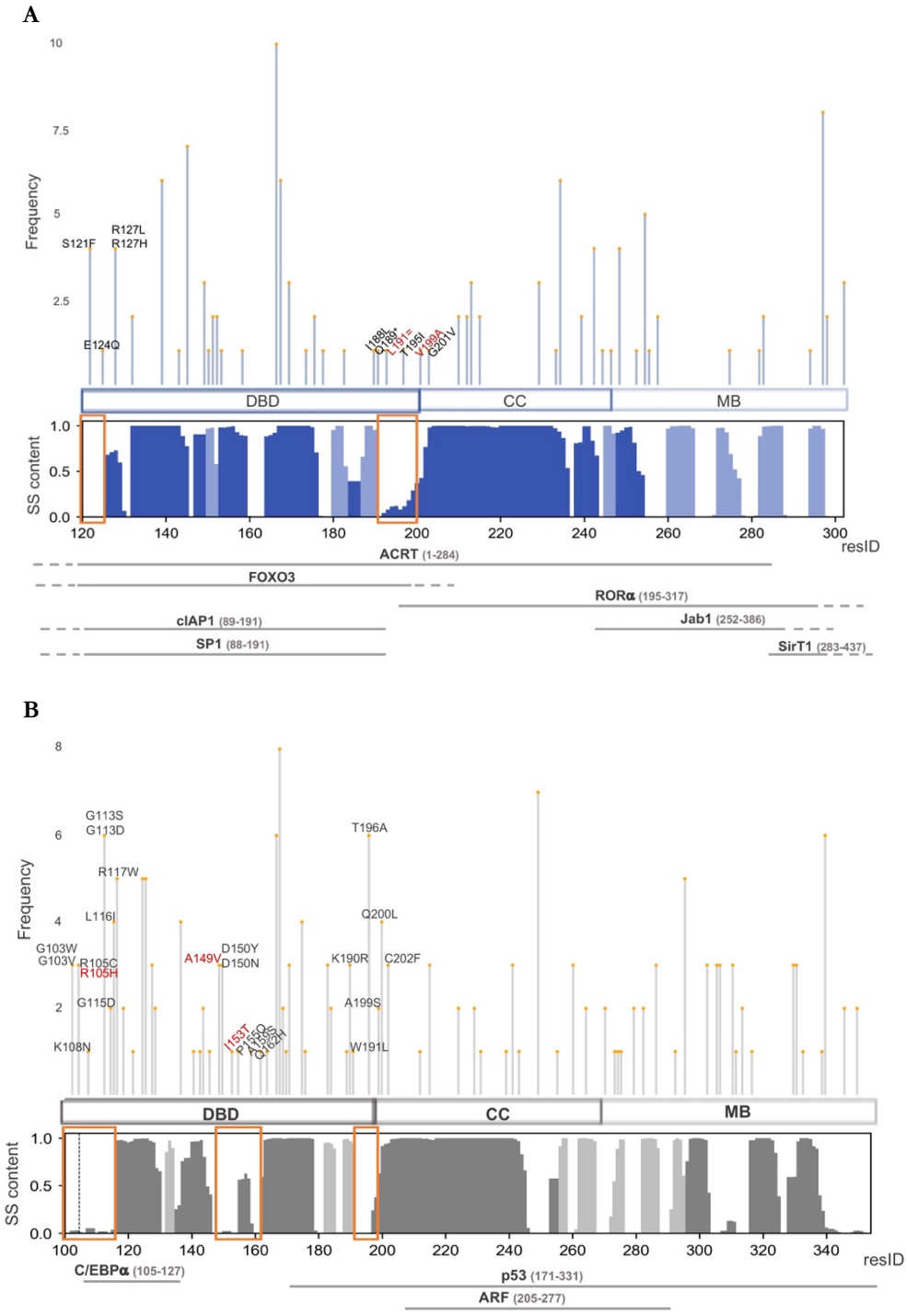


Figure 25. Schematic diagrams showing the somatic mutations in E2F1^{DCM} **(A)** and DP1^{DCM} **(B)** identified from the BioMuta database and mapped onto the corresponding domains. Mutations in the interested flexible regions are labelled, with black and red colours indicate probably damaging and likely benign mutations, respectively, according to the PolyPhen score. The length of the bars indicates the mutation frequency according to BioMuta. In the lower panels, Per-residue secondary structure (SS) content of E2F1^{DCM} (blue, A) and DP1^{DCM} (grey, B) according to the SAXS-driven multi-replica simulation are illustrated, with light and dark shades indicate β-sheet and α-helical content, respectively. Orange rectangles highlight the flexible DBDs regions. The locations of protein interactors reported in literatures are shown [175], [184]–[191].

2. Characterisation of NF-Y and E2F1 partnership on human CDC2 promoter

In this chapter, first I will provide a brief background on results published in literature that represent a starting point of this part of my research. Second, I will present the main results obtained, aimed to assess the question concerning possible DNA-binding cooperativity between NF-Y and E2F1 on the CDC2 promoter, shown to be functionally dependent both from E2F and CCAAT sites [150].

2.1 Introductory overview on the framework of the project

The eukaryotic cell cycle is a precisely regulated process that relies on alternating phases that must be carefully controlled to ensure faithful duplication and maintenance of the genomic integrity. Cell cycle progression is mainly regulated by a set of cyclin-dependent kinases (CDKs), a family of serine/threonine protein kinases that associate with a regulatory subunit known as cyclin that provides essential domain for enzymatic activity and specificity [192]. The CDKs have an established role in promoting and coordinating transitions through the cell cycle with the periodic expression of critical gene clusters required for the transition. The evolutionary conserved CDK1 (also termed CDC2) is the predominant member of this kinases family which plays a central role in cell cycle progression [193][194]. The CDC2 promoter is extremely regulated in a growth-dependent manner. In particular, the Rb gene product has an established role in the repression of this promoter through direct interaction with E2F1 protein. Mutational analyses have clearly shown that intact E2F sites are required for optimal Rb-mediated repression [195].

To further investigate CDC2 transcriptional regulation throughout the cell cycle, the human gene has been sequenced in its entirety along with the 5'-flanking sequences in the late 80s [196][195]. Next, genomic footprinting along 800 bps of the promoter coupled with functional analysis have identified potential upstream cell-cycle regulatory elements, including Myb binding sites, two Sp1 sites, two inverted CCAAT boxes, ETS-2 site, CFR element and both positive-and negative-acting E2F-elements. These findings are consistent with subsequent ChIP assays that have demonstrated that the CCAAT, SP1 and ETS-2 elements are occupied constitutively throughout the cell cycle, while the E2Fs interactions are highly dynamic. ChIP approaches have identified that activator E2Fs bind the distal positive-acting element located at -128 bps, while a subset of E2F4-p130 complexes are bound at the -20 element in proximity of the TSS, suppressing the promoter activity at inappropriate phases of the cell cycle [197][150].

Interestingly, sequences recognised by the activator and repressor E2Fs are quite similar, suggesting that other aspects, beyond a simple DNA sequence recognition, determine the specificity of promoter interaction and explain the dynamic nature of the E2F interactions. The specificity of the activator and repressor E2Fs to interact with distinct E2F-binding sites in the CDC2 promoter would be dictated by the presence of nearby TFs that could stabilise the binding, either directly through physical

interaction or indirectly by shaping the promoter architecture. This observation well fits with detailed functional studies on CDC2 promoter that highlighted the strict requirement of a nearby CCAAT element bound by NF-Y for the binding of E2F1, while the binding of E2F4 is dependent on an adjacent CHR element, suggesting a role for cooperative interactions to elicit both transcriptional activation and repression. Particularly, they have shown that mutation of the distal CCAAT element inhibits the binding of E2F1 to the CDC2 promoter and completely abolishes promoter activation [150]. Indeed, compelling evidences demonstrated that an intact E2F site is required for optimal B-Myb-mediated transcriptional activation of G2-regulated genes. Given the apparent cooperative role of NF-Y and E2F1 in the activation of CDC2 transcription, we have investigated whether this synergy observed *in vivo* is mediated through "assisted DNA-binding". To evaluate possible DNA-binding cooperativity, we further characterised *in vitro* this promoter by gel shift assays using purified recombinant proteins.

2.2 Main Results

2.2.1 E2F1/DP1 and NF-Y bind *in vitro* the CDC2 promoter elements with sequence-specificity

To investigate whether the transcriptional synergy observed *in vivo* between NF-Y and E2F1 on the CDC2 promoter is mediated through "assisted binding" and to identify stable higher order complexes that could be assembled and isolated *in vitro* for further structural characterisation, we purified a 100 bps probe derived from the CDC2 promoter, holding the two inverted CCAAT elements and the canonical E2F-site bound by E2F1 (**Figure 26A**).

First, we used the purified recombinant proteins: NF-Y minimal domain (NF-Ymd), which only includes the core DNA-binding domains used in the published crystallised complex [67] and $E2F1^{DCM}/DP1^{DCM}$ (previously described) to assess DNA-binding and complex formation in simple dose-response EMSAs (**Figure 26B**). Indeed, these assays show binding to the Cy5-labeled CDC2-100 bps probe, confirming that either TF efficiently binds the purified probe, forming retarded mobility DNA complexes. In particular, at low molar ratio with respect to DNA, one complex is observed consistent with the occupancy of one (high affinity) binding site. Increasing the molar ratio, a super-shifted complex is also observed, indicating co-binding on a second (low affinity) site. We used the mammalian Position Specific Frequency Matrix (PSFM) derived from the analysis of NF-Y locations from ChiP on ChiP data to verify the predicted "quality" of the two CCAAT elements within the CDC2 probe in order to explain the different DNA-binding affinities of NF-Y to these sites [38]. Interestingly, the distal CCAAT box to TSS is considered optimal in nucleotides flanking CCAAT pentanucleotides with respect to the mammalian matrix compared to the proximal CCAAT (see paragraph 1.3.1).

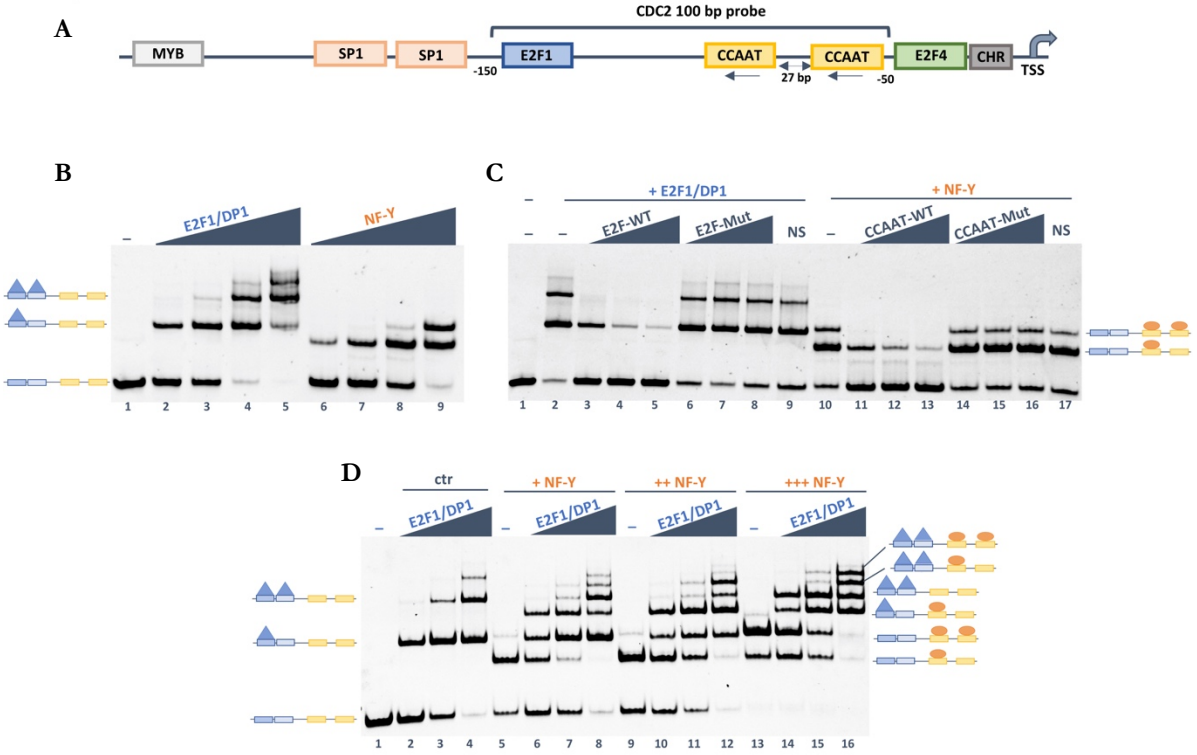


Figure 26. The identity of the protein-DNA complexes at the human CDC2 promoter. (A) Schematic of human CDC2 promoter indicating CCAAT elements, CHR element, E2F, SP1 and MYB binding sites previously identified [197]. The purified 100 bps probe is indicated. **(B)** Dose-response EMSA experiment was performed with the recombinant NF-Ymd and E2F1^{DCM}/DP1^{DCM} proteins where the probe concentration was kept constant at 20 nM. The proteins were incubated at increasing concentrations (10, 20, 40, 60 nM) with the probe. **(C)** Representative gel of direct competition EMSA. The indicated unlabelled competitors were incubated at increasing concentrations (5x, 25x, 50x fold excess) with the CDC2-100 bps probe (20 nM), with the addition of E2F1/DP1 (30 nM) or NF-Ymd (40 nM). As negative controls mutated and non-specific (NS) oligos were used. As positive control, the proteins were incubated alone with the probe (Lane 2 and 10). **(D)** Fixed NF-Ymd concentrations (20 nM; lanes 5-8, 40 nM; lanes 9-12, 60 nM; lanes 13-16) were incubated with the probe (20 nM) with the addition of an increased concentrations of E2F1/DP1 (20, 40, 60 nM). Lane 1: free probe. Bands identity are schematised on either sides of each gel, where high and low affinity E2F1 sites are indicated by dark and light blue boxes, respectively, and CCAAT elements by yellow boxes. Occupied sites by E2F1/DP1 and NF-Y are depicted by blue triangles and orange ovals, respectively.

Next, the DNA binding specificity of the observed NF-Y and E2F1/DP1 complexes was assessed by direct competition assays. The specificity of the formed complexes was verified by challenging with unlabelled oligos. Particularly, E2F1/DP1/DNA and NF-Y/DNA complexes were challenged with the respective *bona fide* E2F or NF-Y consensus site (22 bps and 31 bps competitors, respectively) derived from the human CDC2 promoter (oligo sequences are listed in **Table 2**). **Figure 26C** shows that either specific oligo competes TF-binding efficiently. The specificity of the formed complexes was further confirmed by challenging with competitors containing mutations in the central bps of the canonical E2F and NF-Y sites. As expected for sequence-specific binding of each of the NF-Y or E2F-bound shifted complexes, with mutated oligos no competitions were observed (lanes 6-8 and lanes 14-16). We can conclude that the human CDC2 promoter contains at least two E2F sites and two CCAAT elements specifically bound by the relative TF with different affinities.

Having established that NF-Y and E2F1/DP1 bind specifically the CDC2 promoter *in vitro*, we set out to dissect whether the two proteins can bind DNA in a cooperative manner. To this aim, we assessed the binding affinity of E2F1/DP1 in the presence of NF-Y. As shown in **Figure 26D**, at low molar ratio, the high affinity sites are equally occupied (lane 6), while at higher molar ratio all possible combinations of higher order complexes are formed (lane 16). Although the pattern of banding is too difficult to interpret, we can observe somewhat reassuringly that concomitant binding of E2F and NF-Y is possible at multiple sites. Indeed, multiprotein complexes are stable and we can exclude in this case (at least for the employed protein constructs) possible negative interactions between these TFs on the CDC2 promoter.

2.2.2 Dimerization of E2F1 and DP1 is required for efficient binding to the CDC2 promoter

Considering its potential homodimeric DNA-binding properties, we wanted to establish whether the E2F1 protein can bind the CDC2 promoter in the absence of its partner. We assayed a purified recombinant E2F1^{DCM} and compared it to E2F1^{DCM}/DP1^{DCM} heterodimer in a dose-response EMSA (**Figure 27A**). Under standard gel shift conditions, i.e. with 30 min binding reactions incubated at 30°C (see Methods section), E2F1 alone was not sufficient to bind efficiently the CDC2 promoter (lanes 2-7). In contrast, interaction of E2F1 with the DNA was significantly increased when dimerised with DP1 (lanes 8-13). Next, we assessed whether the presence of nearby bound NF-Y could enhance the binding of E2F1 in the absence of DP1 by dose response EMSA (**Figure 27B**). Also, with two CCAAT elements occupied by NF-Y, E2F1 alone was not able to bind efficiently the DNA, as no further shifted complexes are formed by addition of the E2F1 protein. We can conclude that, DP1 is an obligate partner of E2F1 for efficient DNA binding to the CDC2 promoter.

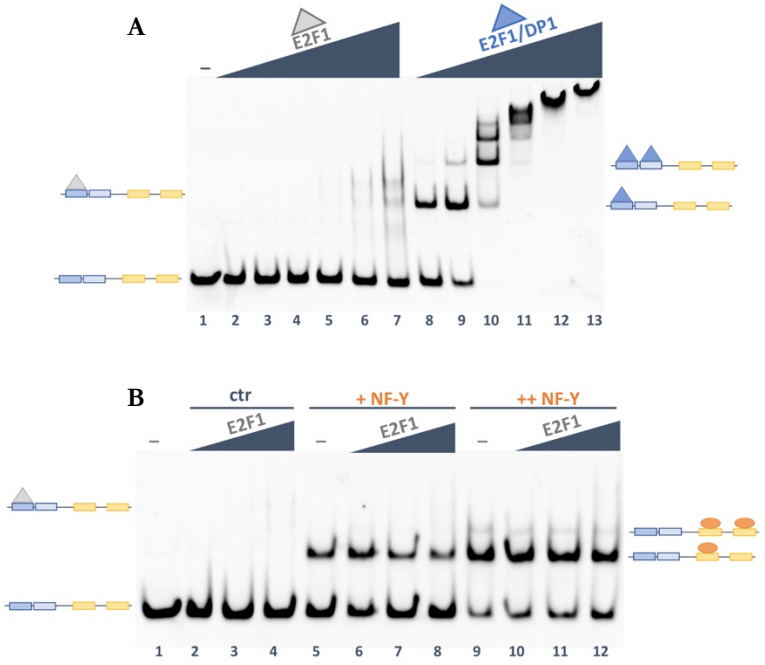


Figure 27. E2F1 binds efficiently the CDC2 promoter as heterodimer. (A) E2F1 in the absence and presence of the DP1 subunit were incubated at increasing concentrations (10, 20, 60, 120, 240, 480 nM) with the CDC2-100 bps probe (20 nM). **(B)** E2F1 was incubated at increasing concentrations (20, 60, 120 nM) with the probe (20 nM) in the absence and presence of NF-Ymd (20 nM; lanes 5-8 or 40 nM; lanes 9-12). Lane 1: free probe. Bands identity are indicated on the sides of each gel.

2.2.3 E2F1/DP1 and NF-Y bind the CDC2 promoter simultaneously, forming stable higher order complexes

Due to the complexity of the banding pattern using a 100-bps probe, we therefore trimmed down the target DNA, to one (distal) CCAAT and the E2F sites (**Figure 28A**). As illustrated in **Figure 28B**, dose-response EMSA confirms the presence of at least two specific E2F-sites with different degree of affinity and one CCAAT box bound by NF-Y.

The synergistic activation of the CDC2 promoter by E2F1 and NF-Y observed *in vivo*, combined with the vicinity of the two binding sites, suggested that the two TFs might simultaneously bind the promoter to cooperatively activate transcription. To test this hypothesis, EMSAs were performed in which increasing amounts of E2F1/DP1 were added to a constant amount of NF-Y and DNA. **Figure 28C** shows that E2F1/DP1 binds the DNA in the absence of NF-Y (lanes 2-5). Addition of NF-Y to the binding reactions resulted in the presence of DNA bound solely by NF-Y or E2F1/DP1 and in the formation of a complexes of E2F1/DP1, NF-Y and DNA. The slower mobility bands suggest that NF-Y and E2F1/DP1 form a ternary and quaternary complex that require both proteins and DNA. These complexes increase in abundance with increasing quantities of NF-Y relative to E2F1/DP1 (lanes 12-15). These results support the notion that NF-Y and E2F1 co-occupy the CDC2 promoter simultaneously and suggest that both proteins may function together as part of multimeric complex to synergistically activate transcription. In addition, we could select promising candidates corresponding to co-binding of the two TFs (lane 15) for subsequent stability/cooperativity analyses and *in vitro* multimeric complex assembly.

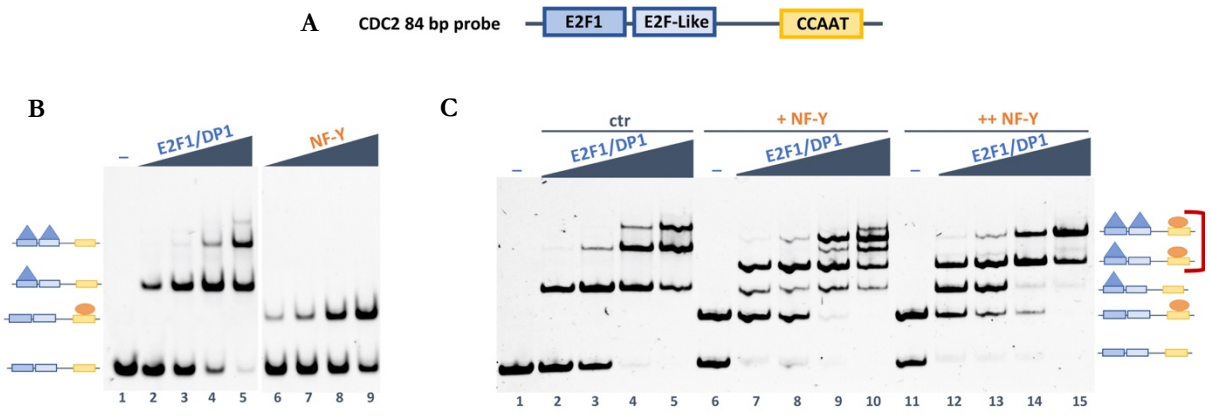


Figure 28. E2F1/DP1 and NF-Y form stable higher order complexes on the purified CDC2-84 bps probe. (A) Schematic of the purified probe encompassing one CCAAT element and at least two E2F-sites. Dose-response EMSAs were performed to assess DNA binding and complex formation of recombinant proteins on the CDC2-84 bps probe. **(B)** E2F1/DP1 and NF-Y were incubated at increasing concentrations (10, 20, 40 nm.) with the probe (20 nM). **(C)** Fixed NF-Y concentrations (40 nM; lanes 5-8, 60 nM; lanes 9-12) were incubated with the probe (20 nM) with the addition of an increased concentrations of E2F1/DP1 (20, 40, 60 nM). Lane 1: free probe. Bands identity are indicated on the side of each gel. Bands corresponding to co-binding of the two TFs are highlighted by a red bracket.

2.2.4 Characterisation of an E2F-like binding site on the CDC2 promoter

To localise the additional E2F binding site(s) that were revealed by our dose-response analyses on the CDC2 promoter, we designed a set of oligos spanning the probe, centred upon sequences containing the minimal E2F consensus site that were used in direct competition EMSA (**Figure 29**). As expected, only one of the designed oligos (E2F-Like1), whose sequence is located in the vicinity of the canonical binding site could compete E2F binding on the CDC2 probe. Although the putative E2F-like site partially deviates from the canonical motif (ACC**CG**GGAA), E2F1/DP1 is bound (thus compete) with lower affinity as compared to the canonical site (lanes 6-8 vs. lanes 3-6).

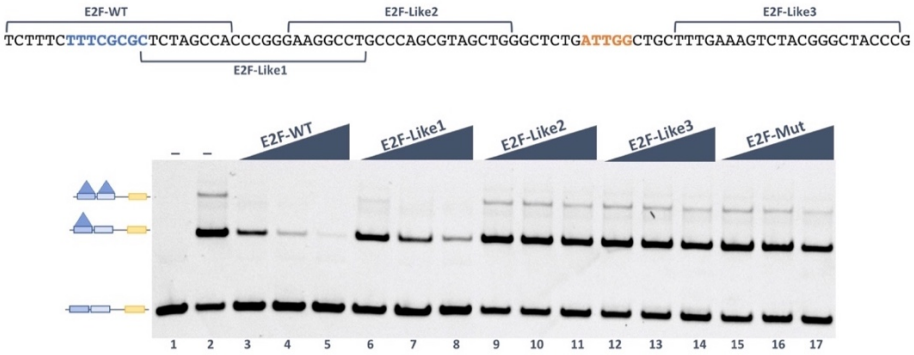


Figure 29. **Mapping the E2F-like binding sites on the CDC2 promoter by EMSA competition analysis.** The 22-mer E2F oligonucleotides' sequences spanning the CDC2-100 bps DNA are indicated on the top of the gel. The indicated competitors were incubated at increasing concentrations (5x, 25x, 50x fold excess) with the CDC2-84 bps probe (20 nM), with the addition of E2F1/DP1 (15 nM). The wild-type canonical E2F1 site (E2F-WT) was used as a positive control. Mutated E2F (E2F-Mut) oligo was used as negative control. Lane 1: free probe. Bands identity are indicated on the left side of the gel.

To investigate whether the presence of NF-Y bound nearby could stabilise E2F1/DP1 at the low-affinity site and to better characterise the E2F-like site, we purified an 84 bps probe mutating the canonical E2F site (**Figure 30A**). From a comparison between WT and mutated probes, we can make the following observation: (i) mutating the canonical site, E2F-like site is bound at higher affinity, suggesting negative interaction with the WT (canonical) site (lanes 3 vs. 13); (ii) E2F1/DP1 prefers to bind the E2F-like site when DNA already occupied by NF-Y compared to free DNA (compare the abundance of the binary and ternary complexes in lanes 17 and 18).

In conclusion, with these set of experiments we could set up the conditions in terms of protein concentrations allowing the proper detection of the selected higher order complexes for subsequent stability assays. We further confirmed the location of the E2F-like site by direct competition EMSA using mutated E2F-like1 oligo (**Figure 30B**).

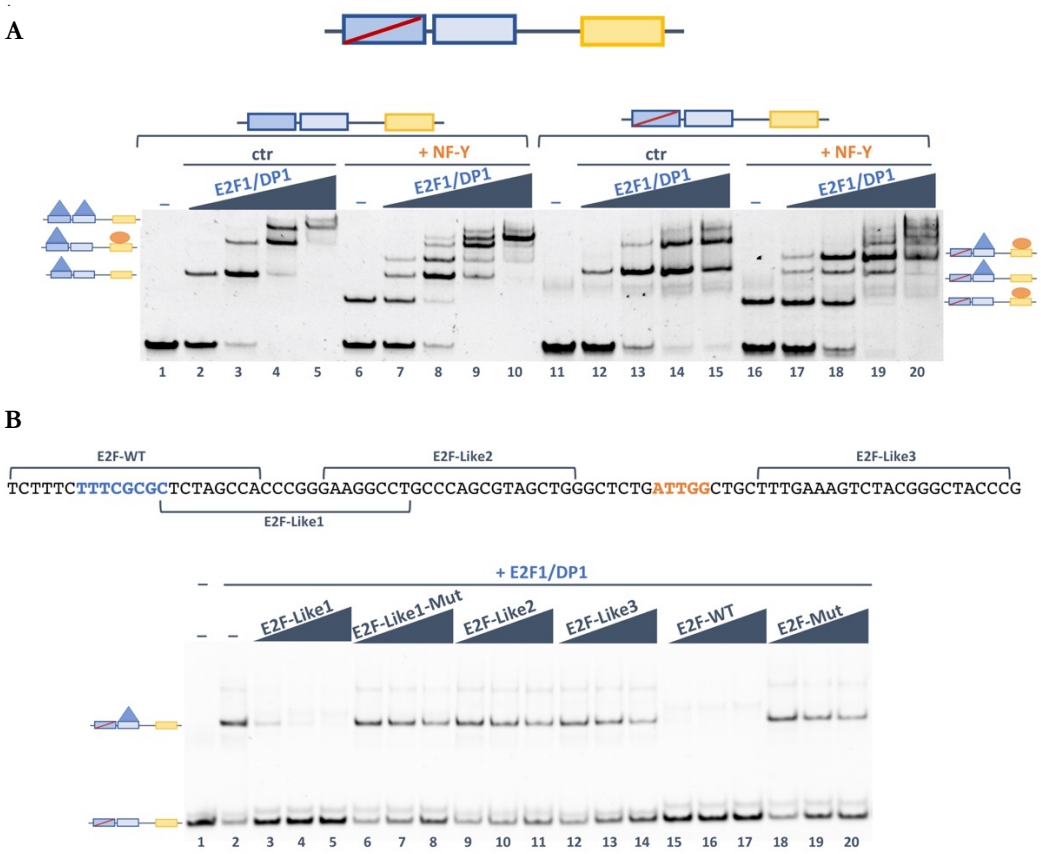


Figure 30. Characterisation of the E2F-like site on the CDC2 promoter. (A) Comparison between the WT and mutated CDC2-84 bps probes. The probe concentration was kept constant at 20 nM. Increased E2F1/DP1 concentrations (2, 10, 20, 30 nM) were added to the reaction mix contained fixed amount of NF-Ymd (10 nM) or dilution buffer used as control (ctr). **(B)** The indicated unlabelled oligos were incubated at increasing concentrations (5x, 25x, 50x fold excess) with the mutated CDC2-84 bps probe (20 nM), with the addition of E2F1/DP1 (15 nM). The wild-type canonical E2F1 site (E2F-WT) was used as a positive control. Mutated E2F (E2F-Mut) oligo was used as negative control. Lane 1: free probe. Bands identity of relevant complexes are indicated on the left side of the gel.

2.2.5 E2F and NF-Y do not display cooperative DNA binding on the CDC2 promoter *in vitro*

Next, considering that our analyses were performed on the NF-Y construct only composed of the homology regions of the protein, an important passage was the use of larger protein constructs in order to check whether stability/cooperativity of complexes could be enhanced or modified by additional regions of the protein. In particular, we used the assembled NF-YA/NF-YB ΔQ/NF-YC, NF-YA/NF-YB ΔQ/NF-YC ΔQ, NF-YAmd/NF-YB ΔQ/NF-YC ΔQ and NF-Ymd heterotrimers (ΔQ: truncated constructs lacking the Q-rich domain; md: minimal domain, for more details refer to part II, chapter 1.1.1). As shown in **Figure 31A**, all NF-Y forms are proficient in DNA binding with relatively similar affinities for the CCAAT-containing probe. To simplify data analyses of NF-Y/DNA complexes with E2F, we started with the mutated CDC2-84 bps probe and we selected the best protein concentrations allowing efficient detection of the ternary complexes for subsequent stability analysis (**Figure 31B**).

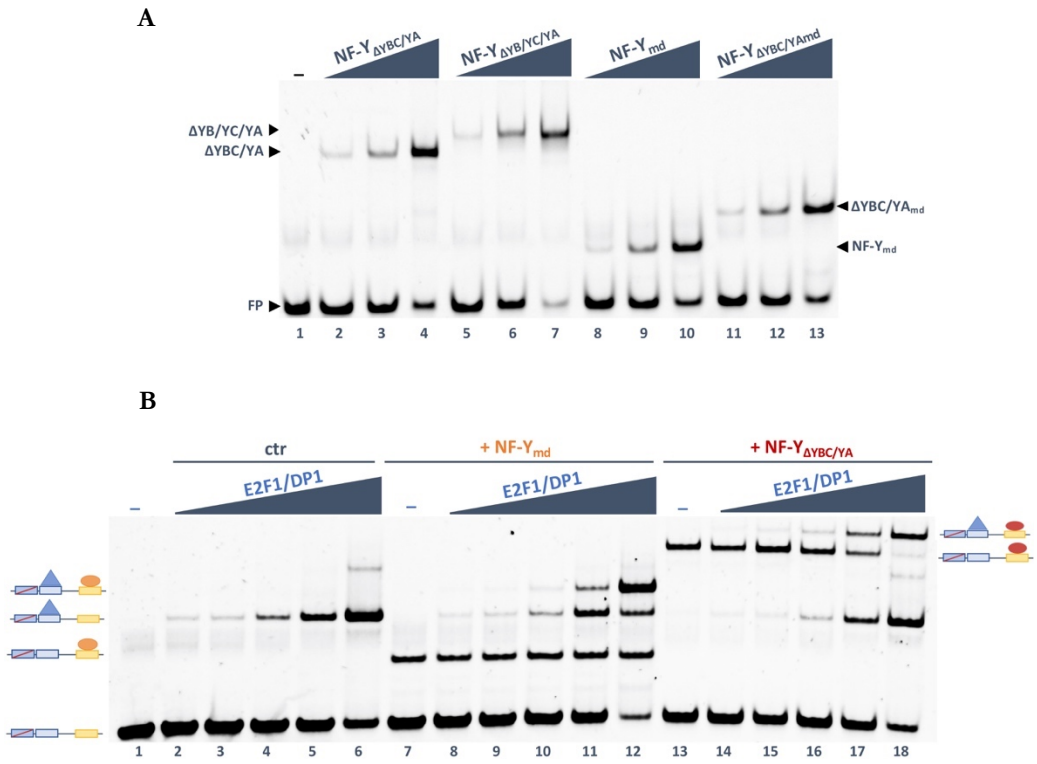


Figure 31. Different NF-Y construct combinations bind DNA with similar affinity. (A) The indicated assembled NF-Y trimers (NF-Ymd in orange while ΔYBC/YA in red) were incubated at increasing concentrations (2, 10, 20 nM) with the mutated CDC2-84 bps probe (20 nM). On the side of the gel, an arrowhead indicates the formed NF-Y/DNA complex. **(B)** Increasing amounts of E2F1/DP1 (2, 4, 8, 16, 32 nM) were added to a constant amount of NF-Y (15 nM) incubated with the probe (20 nM). Lane 1: free probe. Bands identity are indicated on the side of the gel. FP: free probe.

We then assayed cooperativity with E2F1/DP1 for both NF-Ymd and NF-YA/NF-YB ΔQ/NF-YC ΔQ by assessing the stability of the ternary complexes vs. binary complexes in parallel reactions over time using E2F and CCAAT competitors in off-rate EMSAs (**Figure 32A, B**). Quantitatively, cooperative DNA-binding is characterised by slower dissociation rate of the protein from the multimeric complexes compared to the binary complexes bound by a single TF. Regarding the analysis of the NF-Y complexes, the high affinity binding observed on the CDC2 probe was not informative as the equilibrium of DNA binding complexes was reached at rates too slow using different CCAAT competitors with different length and origin (HSP70-25, HSP70-31 and CDC2-31) (data not shown). Concerning binding kinetics on the E2F side, off-rate experiments revealed instead high dissociation rates, as the equilibrium of DNA-binding complexes was reached at rates too fast to measure using the E2F competitor (even with immediate gel loading, decreasing the working temperature and using different E2F competitors with different origin, length and composition).

Off-rate experiments performed in the presence of two different NF-Y constructs showed similar behaviour, indicating that neither the homology domains nor the regions included in the ΔYBC/YA protein could enhance stability of E2F on DNA. Using this strategy, we could not detect an evident

DNA binding cooperativity *in vitro*, suggesting that further mechanisms underlie the transcriptional synergy observed *in vivo*.

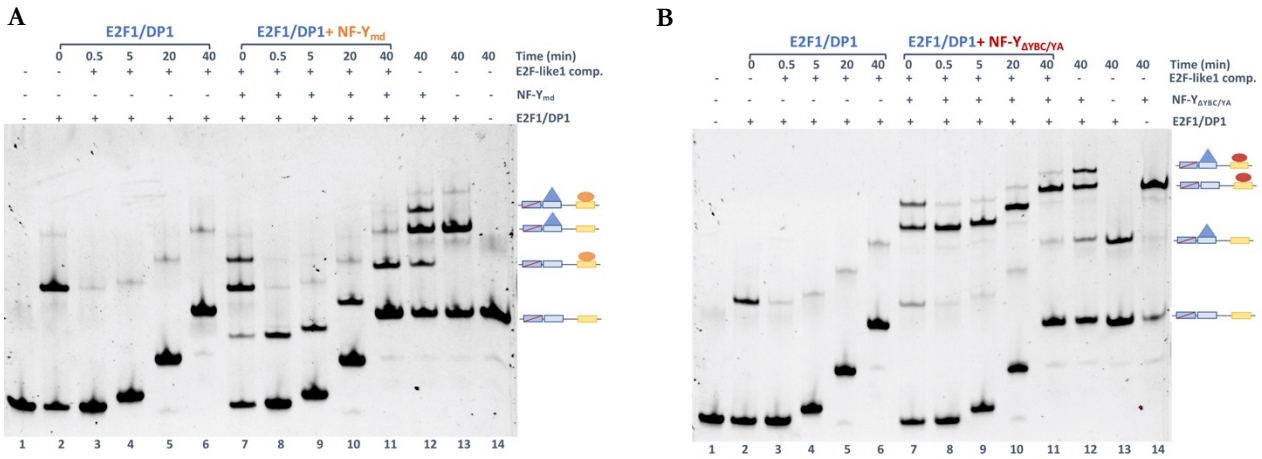


Figure 32. Representative off-rate competition EMSAs used to evaluate the cooperativity between E2F1 and NF-Ymd **(A)** or NF-YA/NF-YB ΔQ/NF-YC ΔQ **(B)** on the mutated CDC2-84 bps probe. In each gel, the reaction devoid (left) and including (right) NF-Y construct is present. DNA-bound complexes were challenged with an excess (25-fold) of unlabelled E2F-like1 competitor and loaded on a running acrylamide gel at different time points for electrophoresis. Loading time (after addition of the unlabelled competitor) is indicated.

Table 2. EMSA probes and unlabelled competitors used in this project. The canonical motifs are indicated in bold; mutant nucleotides are highlighted in lowercase; unmodified oligo (na).

Name	Sequence (5'-3')	Label
CDC2-100 bps	TCTTTC**TTTCGCGC**TCTAGCCACCCGGGAAGGCCTGCCCAGCGTAGCTGG GCTCTG**ATTGG**CTGCTTTGAAAGTCTACGGGCTACCCG**ATTGG**TGAATCC	5'-Cy5
CDC2-84 bps WT	TCTTTC**TTTCGCGC**TCTAGCCACCCGGGAAGGCCTGCCCAGCGTAGCTGG GCTCTG**ATTGG**CTGCTTTGAAAGTCTACGGGCTA	5'-Cy5
CDC2-84 bps Mut	TCTTTC**TTTCGtaC**TCTAGCCACCCGGGAAGGCCTGCCCAGCGTAGCTGG GCTCTG**ATTGG**CTGCTTTGAAAGTCTACGGGCTA	5'-Cy5
CCAAT-WT	CAGCGTAGCTGGGCTCTG**ATTGG**CTGCTTTG	na
E2F-WT	TCTTTC**TTTCGCGC**TCTAGCCA	na
E2F-Mut	TCTTTC**TTTCGtaC**TCTAGCCA	na
E2F-Like1	CTCTAGCC**ACCCGGGAA**GGCCT	na
E2F-Like2	GAAGGCCTGCCCAGCGTAGCTG	na
E2F-Like3	TTGAAAGTCTACGGGCTACCCG	na
HSP70 CCAAT-25 bps	TTCTGAG**CCAAT**CACCGAGCTCGAT	na
HSP70 CCAAT-31 bps	CTTCTGAG**CCAAT**CACCGAGCTCGATGAGGC	na

2.3 Discussion and Future Perspectives

Despite the ability of TFs to execute distinct functions *in vivo*, often tend to bind similar DNA sequences. To solve this specificity paradox, eukaryotes have exploit combinatorial strategies to better

distinguish between similar TFs, increasing genomic target specificity. These strategies include the cooperative binding between TFs and the use of functional low affinity sites [198].

In this part of my work we aimed to analyse whether NF-Y and E2F1 have synergistic interaction on DNA that might provide important cues on different unsolved aspects of their functions, including: (1) the transcriptional synergy observed *in vivo* on the CDC2 promoter [150]; (2) the sequence-specificity paradox between E2F-activators and repressors; (3) the flexible behaviour of E2F1/DP1/DNA complex; (4) the abundance of CCAAT and E2F-binding sites in the promoter of cancer genes with a robust overlap between NF-Y and E2F genomic location, despite the absence of precise distance between their respective sites [45]. For this aim, we used the well-characterised CDC2 promoter, shown to be functionally dependent from both E2F and CCAAT sites [150]. Starting by better characterising the promoter *in vitro* using different probes derived from the CDC2 promoter, we have shown that recombinant E2F1/DP1 and NF-Y proteins bind the CDC2 promoter simultaneously with high specificity, forming stable higher order complexes.

Our EMSA experiments reveal that E2F1/DP1 binds with lower affinity to an E2F-like site that deviates from the canonical motif and is composed solely of GC repeats, lacking the T-rich extension. This observation fit with the genome-wide mapping studies of E2F1 bound loci, unveiling low enrichment for the *in vitro*-derived consensus motif (TTTSSCGC), but rather E2F1 binds to a GC-rich motif that is common to most human promoters [199]. From this work, beyond a simple DNA recognition code, it become apparent that E2Fs use a variety of mechanisms to recognize their cognate sites in the genome. Three models were proposed for binding of E2Fs to core promoters lacking a canonical motif (**Figure 33**). The first mechanism is through indirect recruitment, without contacting DNA directly. This view has been excluded through ChIP-Seq assays using mutated E2F1 in the DBD, confirming that DBD is indeed indispensable. The second mechanism is *via* looping to the core promoter mediated by an E2F bound to distal motif. However, no evidences were obtained which could support this hypothesis. The third model includes TF-mediated tethering and cofactor assisted/facilitated DNA association, where the presence of other site-specific TF bound nearby could stabilise E2F-binding to the DNA either directly through protein-protein interactions or indirectly by shaping the architecture of promoters [200]. Indeed, previous studies have shown that "tethered" recruitment of E2F1 to genome could be mediated by the occupancy of an adjacent SP1-consensus site [201]. In addition, the binding of an activator E2F to the CDC2 promoter is dependent on a nearby CCAAT motif bound by NF-Y, while the binding of a repressor E2F is dependent on an adjacent CHR element [150]. Genome-wide observation of NF-Y co-association with other TFs unveiled the involvement of NF-Y in the tethering recruitment of many TFs to a DNA regions that lack their respective canonical sites, but display nearby a CCAAT-box occupied by NF-Y, including FOS, IRF3, p53, Sp2 and SOX9 [73]. Additionally, E2F family members belonging to the NF-Y interactors without a positional bias of its canonical binding site,

correlate with NF-Y to regulate genes with different functional outcomes, suggesting a possible NF-Y-mediated tethering on DNA. The most striking co-localisation was found for E2F4, that depends on NF-Y for efficient promoter association [40][93].

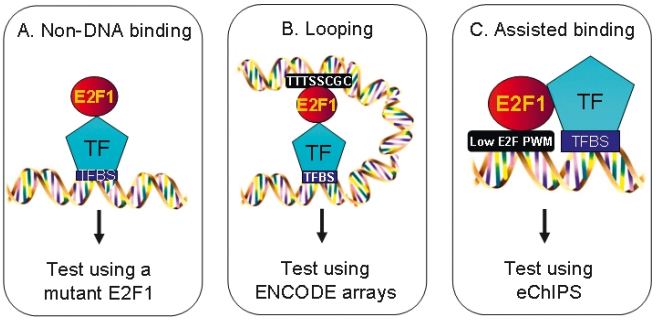


Figure 33. A speculative model for possible mechanisms of E2F1 recruitment to the core promoter that lacks the consensus E2F-motif. **(A)** *Via* inverted recruitment, E2F1 is recruited to promoters through direct interaction with other TF independently of its DBD; **(B)** via looping to the core promoter; **(C)** via assisted binding, E2F1 binds cooperatively with other TF bound nearby [200].

Herein we noted that by mutating the canonical site, the E2F-like site is bound more efficiently. Several hypotheses can be proposed: (1) the binding of E2F1/DP1 to both sites is competitive but not mutually exclusive; (2) the binding of E2F1/DP1 to the E2F-like site depends on the regions flanking the recognition site and changing the flanked nucleotides could alter E2F1 selectivity; (3) the binding of E2F1/DP1 to the canonical site could alter the local DNA topography, hindering the binding to the E2F1-like site. Previous studies have demonstrated the importance of low-affinity sites to finetuning the expression of important developmental genes both spatially and temporally [202], and to distinguish between activation and repression [203]. However, the functional relevance of the E2F-like element remained elusive and requires further investigation.

The results obtained from off-rate EMSA experiments point toward a non-cooperative DNA binding between NF-Y and E2F1, were the binary and ternary complexes behave similarly over time (also using larger NF-Y constructs). Beside the classical model of cooperativity that requires direct protein-protein interactions and precise arrangement of TFBSs in the regulatory regions, cooperativity between TFs can be indirectly manifested in the context of chromatin [20]. Given the NF-Y's distinct DNA-binding mode that promotes chromatin accessibility, one possible mechanism that could explain the transcriptional synergy between NF-Y and E2F is that the binding of NF-Y leads to nucleosome destabilisation or repositioning, facilitating the binding of E2F1 to a nearby site. One of the prominent examples is the role of NF-Y to promote the binding at enhancers of key pluripotency factors, including Sox2 and Oct4 by facilitating a permissive chromatin conformation [79]. Alternatively, cooperativity between E2F1 and NF-Y may involve additional co-factors that are present in the cell context.

Given that the binding kinetics (K_{on} and K_{off}) for E2F1/DP1 and NF-Y are too fast to be measured by EMSA, other quantitative methods must be applied to ascertain this hypothesis. One method could be the use of Reflective Phantom Interface (RPI), a novel technology for label-free biosensing to measure

equilibrium and kinetics constants [204]. Alternatively, single-molecule fluorescence imaging (FRAP) can be applied to capture the binding dynamics of these TFs in living cells [205].

PART II

In this part, I briefly introduced the side/collaborative projects that I was involved in, reporting the main results obtained from our experiments.

Uncovering novel structural information of NF-Y using X-Ray crystallography and Cryo-EM

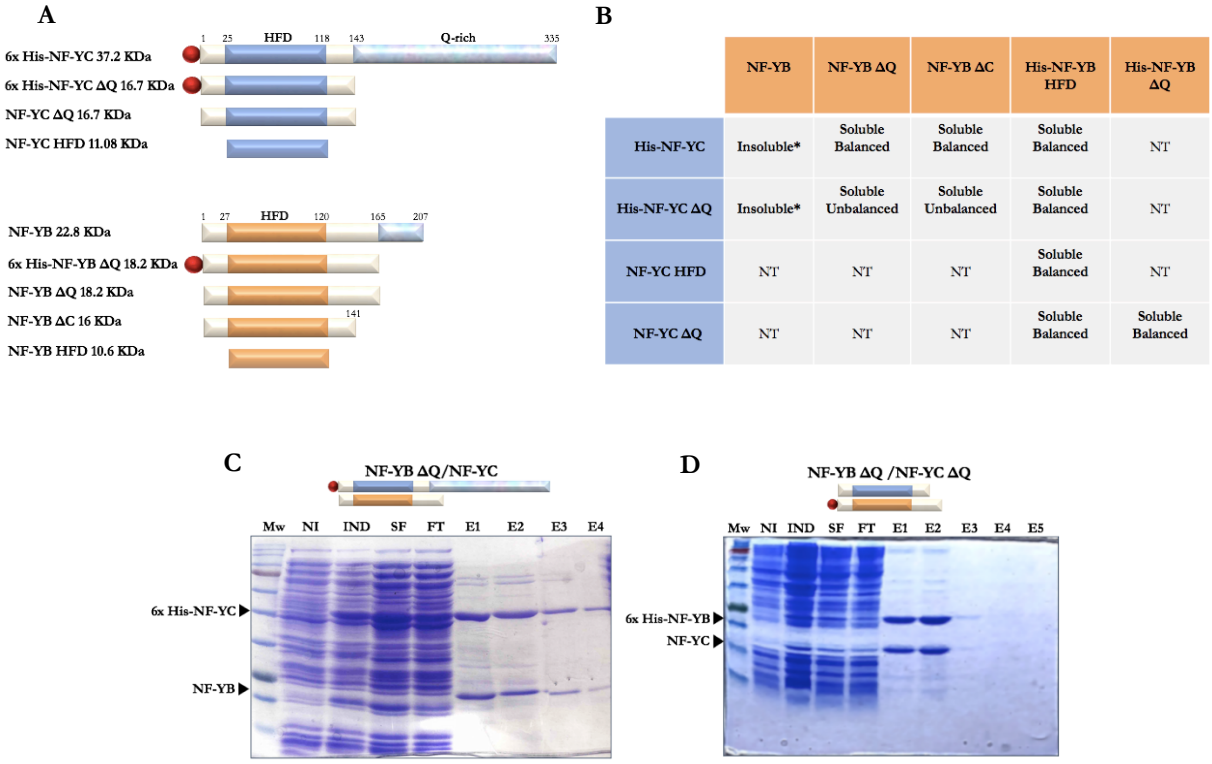
So far, the only structural information of NF-Y refers to the evolutionary conserved minimal core domains [67][71][72]. To gain novel structural information on the additional structured regions, we applied X-ray crystallography on NF-Y constructs lacking the predicted disordered regions, while full-length proteins were used in multimeric complexes and analysed by single particle Cryo-EM for both structural and USF1 partnership analyses.

1. Design optimised NF-Y constructs for structural and TFs partnership analyses

1.1 Main Results

1.1.1 Production of NF-Y: HFD recombinant heterodimers

Because of the size of the complexes using minimal domains apparently limit the use of Cryo-EM, we set out to produce the full NF-Y trimer. NF-YA (45 kDa) subunit is successfully produced as soluble protein by *A. Bernardini*. For NF-YB and NF-YC subunits, a serial deletion analysis is performed maintaining the conserved regions previously crystallised [67][72] (**Figure 34A**) . Because the subunits are insoluble singularly, we set up co-expression screening tests with the full-length proteins and with different combination of deletions to possibly improve solubility and protein folding using different *E. coli* strains (**Figure 34B**). In particular, we co-produced NF-YC full length protein (37 kDa), together with NF-YB construct lacking the last 42 amino acids (19 kDa, hereafter NF-YB ΔQ) and NF-YC construct lacking the predicted disordered Q-rich domain (16.5 kDa, hereafter NF-YC ΔQ) together with NF-YB ΔQ. We optimised co-expression and purification steps: the process was efficient, and the obtained yields are summarised in **Table 3**. Co-expression and purification of the selected NF-YB/NF-YC heterodimers during the various steps are evaluated by Coomassie-stained gels as shown in **Figure 34C, D**. We then started two types of experiments: NF-YB ΔQ/NF-YC ΔQ was subjected to preliminary crystallisation tests/trials; NF-YB ΔQ/NF-YC was used to assemble and isolate multimeric complexes that were then analysed by Cryo-EM. Additionally, these constructs were used in different EMSA experiments conducted to evaluate the effect of the additional NF-Y portions on the DNA-binding stability/cooperativity (Part I, chapter 2.2.5).



	NF-YB	NF-YB ΔQ	NF-YB ΔC	His-NF-YB HFD	His-NF-YB ΔQ
His-NF-YC	Insoluble*	Soluble Balanced	Soluble Balanced	Soluble Balanced	NT
His-NF-YC ΔQ	Insoluble*	Soluble Unbalanced	Soluble Unbalanced	Soluble Balanced	NT
NF-YC HFD	NT	NT	NT	Soluble Balanced	NT
NF-YC ΔQ	NT	NT	NT	Soluble Balanced	Soluble Balanced

Figure 34. Co-production of NF-YB and NF-YC recombinant heterodimers. (A) Schematic representation of the series of NF-YB and NF-YC constructs used in this project. Evolutionary conserved core HFDs are indicated as orange and blue boxes for NF-YB and NF-YC, respectively. The Q-rich domains are highlighted in grey. The 6xHis-tag is represented as a red sphere. **(B)** Summary of co-expression screening tests conducted using full length recombinant HFD proteins and different construct combinations. NT: Not Tested, and * tested in different *E.coli* stains, including ROSETTA, PlySE, GroEL/GroES and BL21(DE3). Coomassie-stained gels of IMAC purifications of NF-YB ΔQ/NF-YC used to assemble multimeric complexes visualised by Cryo-EM **(C)**, and NF-YB ΔQ/NF-YC ΔQ heterodimer used in crystallisation trials (**D**). The arrowhead indicates the identity of each band.

Table 3. overview of recombinant protein constructs used in this thesis.

Construct	Molecular weight (KDa)	Yield (mg/L culture)	Comments
6x His-NF-YC/NF-YB ΔQ	41.6 + 19.8	~4.5 mg	ΔQ: deletion of small Q-rich domain at the C-terminus
6x His-NF-YC/NF-YB ΔC	41.6 + 17	~7 mg	ΔC: deletion of the entire C-terminus
6x His-NF-YC ΔQ/NF-YB ΔQ	21 + 19.8	>10 mg	Subunits seem unbalanced
6x His-NF-YC ΔQ/NF-YB ΔC	21 + 17	>10 mg	Subunits seem unbalanced
6x His-NF-YC ΔQ/NF-YB HFD	21 + 10	>10 mg	-
NF-YC ΔQ/6x His-NF-YB ΔQ	16.7 + 21	~8 mg	Balanced subunits
NF-YC ΔQ/6x His-NF-YB ΔC	16.7 + 18.2	~9 mg	Balanced subunits
E2F1^{DCM}/6x His-DP1^{DCM}	20.45 + 29.4	>10 mg	Produced in large-scale by fermentation
6x His-E2F1^{DCM}	21.13	>10 mg	-

1.1.2 Crystallisation trials of NF-YB/NF-YC heterodimer

To gain novel structural information on the additional structured regions of NF-Y using X-Ray crystallography, we co-produced the NF-YB (aa 1 to 165) and NF-YC (aa 1 to 143) subunits lacking the predicted disordered Q-rich domains (for more details, refer to **Figure 34**). We started by applying large

crystal screening using commercial kits and in-house prepared crystallisation solutions. Quasi-crystals not suitable for X-Ray diffraction were obtained, indicating high flexibility and conformational heterogeneity. As attempts to generate crystals suitable for X-Ray diffraction using the entire N-terminal regions of NF-YB and NF-YC (excluding the Q-rich domains) were unsuccessful and to identify a sub-construct amenable to crystallisation, we performed limited proteolysis. Limited digestion of NF-YB ΔQ/NF-YC ΔQ (**Figure 35A**) using proteinase K, a broad-spectrum serine protease, generated a homogenous and stable fragment (lanes 4-7) with an intermediate size between the previously crystallised minimal HFD dimer (lane 2) and the ΔQ constructs (lane 3). To investigate whether the presence of NF-YA subunit could stabilise additional portions of the dimer, we assembled the NF-Y trimer (lane 10) composed of NF-YAmd and NF-YB ΔQ/NF-YC ΔQ on a 25-mer DNA oligo harbouring the CCAAT box from the HSP70 promoter used to crystalize the mammalian NF-Y/DNA complex [67] followed by PK treatment. The results show clearly that, while NF-YA is highly unstable after PK treatment (lane 9), the presence of the HFD subunits stabilize NF-YA in the presence of DNA (lanes 11-14).

Next, we employed *in situ* limited proteolysis, where the purified heterodimer is contaminated with trace amounts of broad-spectrum protease. On the timescale of crystal formation, serendipitous proteolysis occurs to remove floppy termini, generating a stable proteolytic digestion product. We started by a pilot screening to identify promising stable degradation pattern at different temperature (4°C and 19°C) on a time-scale comparable to the duration of crystallisation as detected by denaturating gel (**Figure 35B**). Using this strategy, liquid-liquid separation is often seen in the plates with different amounts of proteases compared to the control/untreated protein plates (data not shown).

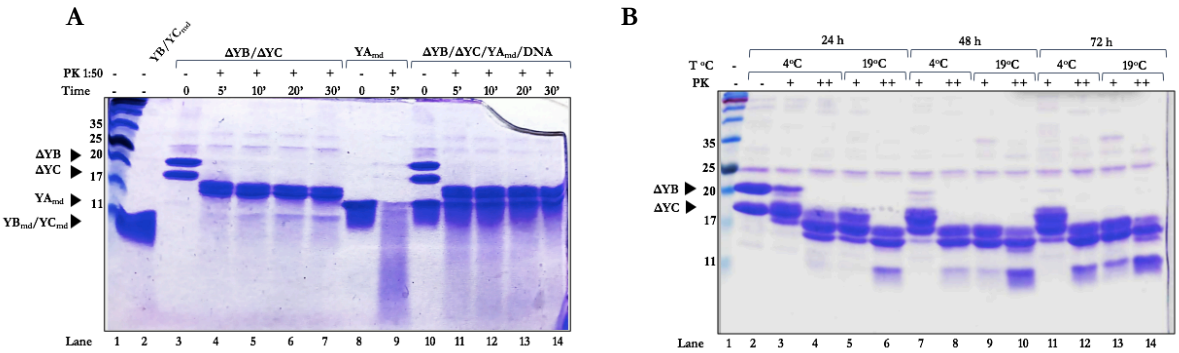


Figure 35. Limited proteolysis yields stable products over time. (A) Time course digestion of NF-YB ΔQ/NF-YC ΔQ heterodimer alone (lanes 4-7) and in complex with NF-YAmd and DNA (lanes 11-14) using constant proteinase K ratio of 1:50 with respect to the protein at 25°C. Untreated samples (lanes 1, 2, 8 and 10) are used as positive control. In the absence on NF-YB/NF-YC and DNA, NF-YAmd subunit show high proteolysis rate after 5 min of treatment (lane 9). **(B)** NF-YB ΔQ/NF-YC ΔQ was subjected to PK proteolysis [1:50,000 (+) and 1:10,000 (++)] and incubated at 4°C and 19°C for various time periods 24, 48 and 72 h. 6 µg of each preparation were loaded. Lane 1 indicates the protein marker, and the sizes (in kDa) are indicated on the left side of the gel. Arrowhead indicates band identity.

As strategy for increasing the efficiency of protein crystallisation, we integrated proteolysis experiments and Mass Spectroscopy (MS) to define the N- and C-terminal boundaries of the proteolytic products. In-gel acid/enzymatic hydrolysis followed by MS conducted at the EMBL Proteomics core

facility (Heidelberg) identified polypeptides encompassing residues 251-342 for NF-YA, 27-155 for NF-YB and 20-127 for NF-YC, indicating trimming from both termini (**Figure 36**). The genes of the refined constructs were synthesised (Eurofins) and cloned into bacterial expression vectors. Co-expression and purification tests will be conducted to optimise an efficient strategy to isolate soluble NF-Y heterotrimer that binds efficiently the DNA. The purified NF-Y will be structurally characterised by X-Ray crystallography and used in multimeric complexes with nucleosome and/or partner TF in Cryo-EM analysis.

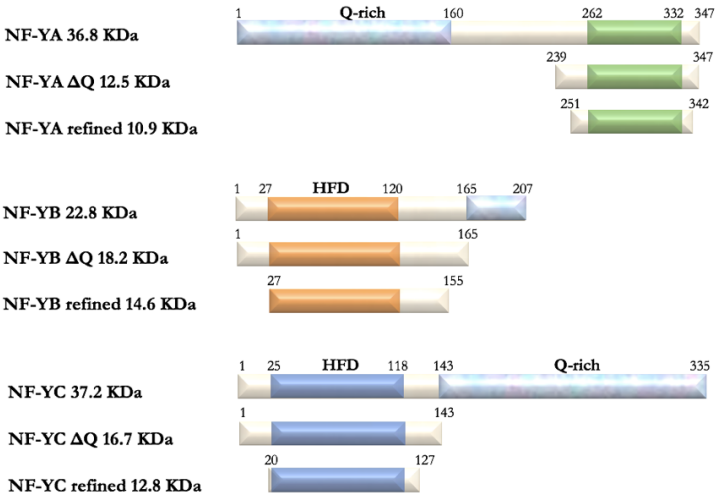


Figure 36. Refined NF-Y construct design. The evolutionary conserved core domains are indicated as green, orange and blue boxes for NF-YA, NF-YB and NF-YC, respectively. The glutamine-rich domains (Q-rich) are highlighted in grey.

2. Analysis of NF-Y/USF1 contacting surfaces using Cryo-EM

This project began by a genome-wide observation linking the CCAAT-box to E-box, recognised by NF-Y and b-HLH TFs, respectively. Particularly, NF-Y co-associates with USF1/2 and MAX, but not MYC, with a recurrent stereo-specific distance of 10-12 bps between their respective binding sites [73]. Preliminary data from our laboratory aimed to verify at the molecular level these genomic observations through DNA-binding approaches using recombinant proteins, structural studies and ChIP assays, and they determined that (1) the binding of NF-Y and USF1 is synergistic, and the distance of 10-12 bps is critical for cooperative binding, while binding of MAX and MYC are independent. (2) Inactivation of NF-YB *in vivo* leads to global decrease of USF1 association to positive promoters, but not to LTR repetitive sequences. (3) NF-Y in complex with USF1 on different DNA configurations visualised by SAXS revealed that NF-Y/USF1 complexes show extended contacts between the NF-Y core and the USR region of USF1, predicted to be intrinsically disordered. Depending on the distance between the E-box and the CCAAT-box (10 vs. 12 bps), the USR adopts different conformations and reciprocal rotation of the two proteins along the DNA axis of about 70°. This was confirmed *in vitro* by deletion mutants of

the USR. (4) The minimal DBD of NF-Y and the USR portion of USF1 are sufficient to mediate synergism on natural promoters with different DNA configurations, including HOXB4 and LTR genes characterised by the recurrent E-box/CCAAT-box genomic spacing of 10 and 12 bps, respectively [151][152].

Structural, biochemical and functional data gathered in our laboratory provide the first look at the domain -USR- involved in a specific protein-protein interaction. However, neither the low-resolution SAXS models nor the *in vitro* assays provided evidences on the specific protein interfaces involved in the interaction mediated by USR. Therefore, we turned to single-particle Cryo-EM to possibly fill this structure gape at near-atomic resolution.

2.1 Main Results

2.1.1 NF-Y and USF1 are proficient in DNA binding separately and combined

First, we used purified recombinant NF-YA, NF-YB ΔQ and NF-YC proteins to assess NF-Y trimerization, DNA-binding and complex formation with USF1 full homodimer (70 kDa) produced by *N. Gnesutta* in simple dose-response EMSA (**Figure 37**). USF1 has clearly shown a cooperative DNA-binding and functional synergism with NF-Y on several natural gene promoters that harbour the 10/12 bps spacing between E-box and CCAAT-box [151][152]. In particular, we used HOXB4 (10 bps) and LTR (12 bps) probes derived from the human *HOXB4* promoter and the LTR *MLT1K* of the ERVL-MaLR family (chr 5), respectively. As expected, NF-Y and USF1 demonstrated strong binding affinities toward both probes, separately and combined, in a sequence-specific manner. We set up the conditions in terms of protein/protein and protein/DNA ratios, allowing the proper formation of stable ternary complexes for subsequent large-scale complex assembly and isolation.

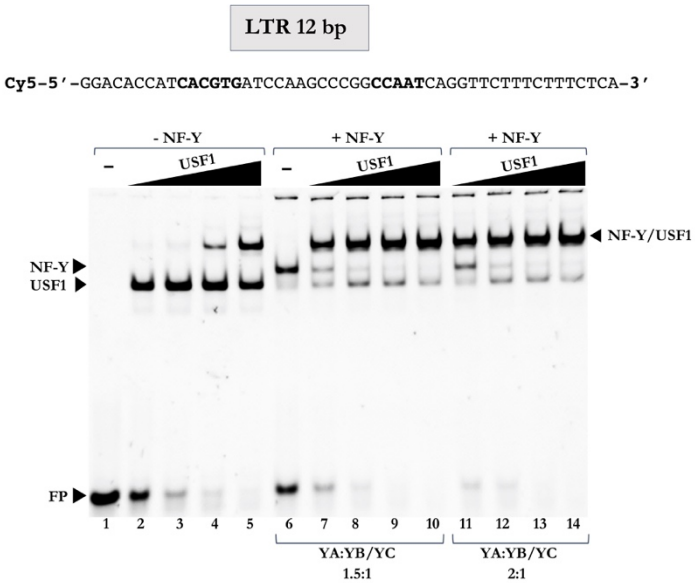


Figure 37. Dose-response EMSA experiment performed with assembled NF-Y and USF1 on LTR (12 bps spacing between sites) probe. NF-Y trimer composed of NF-YA and NF-YB ΔQ/NF-YC dimer (1.5:1 or 2:1, fixed molar ratio) were used in binding reaction at fixed concentration (30 nM) with the probe (20 nM). USF1 concentration were 20, 40, 60 and 80 nM. The 50-mer (LTR) E-CCAAT oligonucleotides' sequences used is indicated. The E-box and CCAAT-box are highlighted in boldface. On the side of the gel, a arrowheads indicate the NF-Y, USF1 and NF-Y/USF1 DNA complexes and FP: free probe.

2.1.2 Ternary complexes are efficiently assembled and isolated *in vitro*

The likelihood of successful structural and biochemical characterisations directly depends on sample quality, and thus sample preparation. Besides improving the expression yields, we attempted to develop an optimal approach to assemble and isolate stable and monodispersed ternary complexes suited to be visualised by Cryo-EM (**Figure 38**). Our optimised strategy consists of first, express and produce the protein in a derivative of the *E. coli* strain. To obtain proteins with high purity, we employed two consecutive chromatographic steps (IMAC and SEC) in high salts, and optimised various parameters in each step for each protein. The purified recombinant proteins are first functionally characterised to bind efficiently DNA, forming stable complexes as assessed by EMSA experiments, and then used to reconstitute multiple ternary complexes on either LTR (12 bps) or HOXB4 (10 bps) dsDNA oligonucleotides. The resulting complexes are subjected to step dialysis decreasing salt concentration, thus favouring DNA binding. The complexes are isolated by gel filtration, proved monodispersed by EMSA assays and SDS-PAGE followed by Coomassie-staining, and finally subjected to Cryo-EM (**Figure 39A**). Our strategy provides an efficient method to produce and isolate monodispersed samples with a purity of >99%, ideal for structural studies.

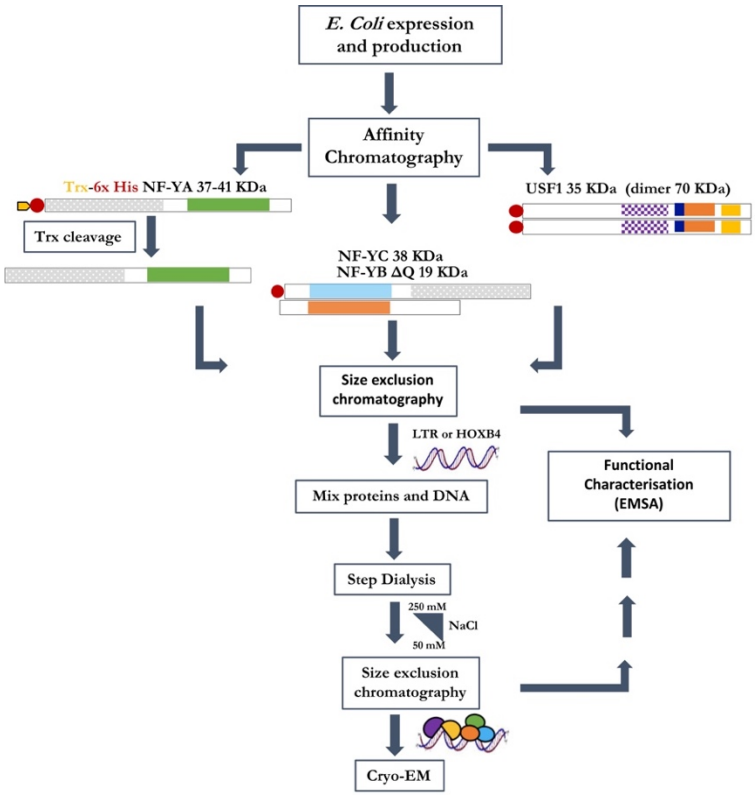


Figure 38. Flowchart illustrating the experimental strategy used to assemble and isolate multimeric complexes.

2.1.3 Single-particle Cryo-EM analysis

In collaboration with our Cryo-EM facility, three independent datasets of ~700 images are collected under cryogenic conditions: NF-Y/LTR, NF-Y/USF1/LTR and NF-Y/USF1/HOXB4. After initial data processing, ~20,000 particles datasets are extracted, aligned and classified yielding 2D class average images (**Figure 39B**). The 2D class averages show strong density for the conserved and structured parts of the NF-Y trimer, producing DNA bending with an angle of 80° consistent with what observed in the crystallised NF-Ymd/DNA complex [67]. Moreover, there are clear additional densities on the flanking E-box, corresponding to the USF1 bHLH-Leu zipper region on both DNA configurations. The density in this region is generally weaker and more diffuse, indicative of flexibility.

However, the large, apparently unstructured TADs of USF1, NF-YA and NF-YC are not visible. These results validate previous structural and biochemical data, in which USF1-intrinsically disordered N-terminal region adopts different conformations adapting to E-CCAAT distances. These data lend support to the hypothesis that the vast majority of intrinsically disordered TADs adapt to highly ordered structures. On the other hand, the presence of high levels of intrinsic disorder and high conformational flexibility limit further processing and single particle reconstitution.

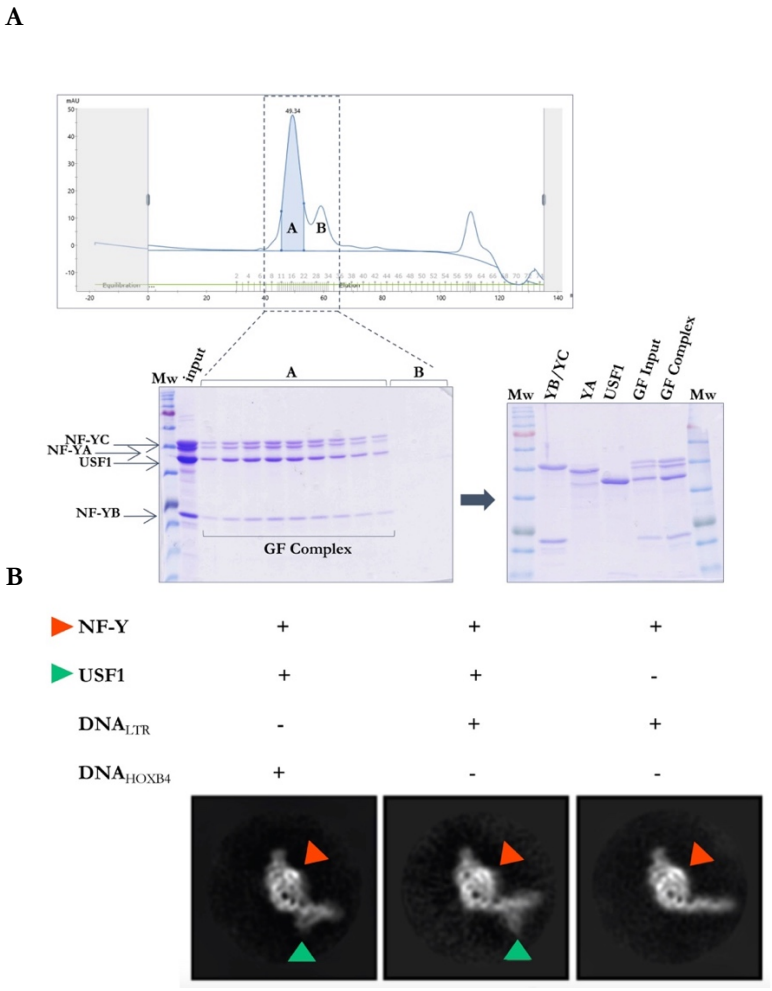


Figure 39. The isolated complexes are subjected to Cryo-EM analysis. **(A)** Example of ternary complex isolation using size exclusion chromatography. Chromatogram derived from purification of *in vitro* assembled NF-Y/USF1/LTR ternary complex on a Superdex 300 10/300 size exclusion column (GE Healthcare). **(B)** Coomassie-stained SDS-PAGE analysis of elution fractions (marked by dashed lines) verified the presence of the complex (left); and soluble purified individual subunits, and pooled reconstituted complex peak fractions (right). Bands identity are indicated on each gel. **(C)** Representative 2D class averages of binary (NF-Y/HOXB4, 135 kDa) and ternary complexes (NF-Y/USF1/LTR and NF-Y/USF1/HOXB4, 200 kDa) subjected to single-particle Cryo-EM analysis.

4. MATERIALS AND METHODS

4.1 Cloning, expression and purification of NF-YB and NF-YC constructs

NF-YB and NF-YC constructs were subcloned and purified exploiting the co-expression system previously described [206]. All NF-YC constructs were inserted in the pmcnYC vector generating an N-terminal 6xHis-tag fusion protein preceded by a thrombin cleavage site, whereas all NF-YB constructs were subcloned into pmcnEATCH vector, through restriction ends ligation. Co-expression tests were performed according to the method of *Fribourg et al.* [207]. Large scale expression was carried out using the previously reported protocol [72].

4.2 Limited proteolysis

20 µL of 0.3 mg mL^{-1} protein was incubated with 5 µL of proteinase K, a broad specificity protease, at different protease-to-protein ratio (1:50, 1:100, 1:1000, 1:10,000, 1:20,000 and 50,000, diluted in PBS storage buffer) at 19°C over 6 days. Reactions were quenched by the addition of 7.5 µL of SDS-Coomassie sample loading buffer (50 mM Tris-Cl pH 6.8, 100 mM DTT, 2% SDS, 10% glycerol, 0,1% BBF, 5.2 mM EDTA and 5.2 mM PMSF), boiled at 95°C for 10 min and loaded onto SDS-PAGE gel.

4.3 E2F1/DP1 co-expression and purification

E2F1 is co-produced with its structurally homologous partner DP1, both contain the DBD, the linker region, the coil-coil (CC) heterodimerisation domain and the marked-box (MB) domain. The cDNA sequence encoding residues Arg_{105}-Thr_{353} of human DP1 obtained by PCR with synthetic flanking RE sites was subcloned into the pmcnEATNH vector, allowing *E. coli* expression of fusion protein with a N-terminal 6×His-tag followed by a thrombin cleavage site. Human E2F1 fragment encoding residues Ser_{121}-Glu_{301} obtained by PCR with synthetic flanking RE sites was subcloned into the pmcnYC vector [206]. The constructs were verified by sequencing and co-expressed in *E. coli* BL21 (DE3). Co-transformed cells with both vectors were grown in LB medium with double antibiotic selection (100 µg mL^{-1} ampicillin and spectinomycin) at 37°C to an OD_{600} of 0.3 and induced by adding a final concentration of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), with protein expression taking place overnight at 25°C. Cells were pelleted by centrifugation at 4,000 × g for 30 min, resuspended in lysis buffer A containing 50 mM Tris-Cl pH 8.0, 400 mM NaCl, 2 mM $MgCl_2$, 10 mM imidazole, 10% glycerol, 2 mM DTT and PIC (Protease Inhibitor Cocktail, Sigma-Aldrich) and lysed by sonication. The resulting soluble fraction recovered by centrifugation at 13,000 x g for 30 min was purified using nickel ion metal affinity chromatography (IMAC) on an ÄKTA Pure system with His-Trap™ FF 1 mL column (GE Healthcare) and eluted in buffer A containing 100 mM imidazole. The pooled protein fractions were incubated overnight at 4°C with the thrombin protease (Thrombin CleanCleave Kit, Sigma-Aldrich) to

cleave the N-terminal 6×His-tag. The resulted cleaved protein was concentrated and purified by gel filtration chromatography step using HiLoad 10/300 Superdex™ 200 column (GE Healthcare) equilibrated in buffer B (50 mM Tris-Cl pH 8.0, 400 mM NaCl, 2 mM DTT). Purity of the protein was examined on SDS-PAGE gel stained with Coomassie brilliant blue.

4.4 Sample preparation for SAXS analysis

The 15-mer DNA fragment used in SAXS was obtained from Eurofins Genomics as single-stranded oligo and annealed in 10 mM Tris-Cl (pH 8.0) containing 50 mM NaCl and 1 mM EDTA. The purified $E2F1^{DCM}/DP1^{DCM}$ heterodimer was mixed with the DNA duplex at a molar ratio of 2:1 and subjected to step dialysis to gradually decrease salt concentration to 150 mM NaCl, thus favouring DNA binding. The ternary complex was isolated using Superdex™ 200 gel filtration chromatography and concentrated to ~4 mg mL^{-1} in a buffer of 50 mM Tris-Cl pH 8.0, 150 mM NaCl and 2 mM DTT. The $Rb^{829-874}$ peptide was synthesized (ProteoGenixSAS, France) and dissolved in 10 mM Tris-Cl pH 8.0 and 50 mM NaCl. The peptide was mixed with purified E2F1/DP1/DNA complex at a molar ratio of 2:1 and subjected to SEC-SAXS analysis.

4.5 In-line SEC-SAXS measurements and data processing

X-ray scattering patterns of $E2F1^{DCM}/DP1^{DCM}$ complexes were recorded using the on-line Size Exclusion Chromatography using Superdex 200 increase 3.2/300 (GE Healthcare) column available at the B21 BioSAXS beamline of the Diamond Synchrotron (Didcot, Oxfordshire, UK) (refer to table for details). For each sample, 45 µL of a 4 mg mL^{-1} was injected onto a pre-equilibrated column in 50 mM Tris-Cl pH 8.0, 150 mM NaCl and 2 mM DTT, and roughly 600 sequential images were recorded at a flow rate of 0.075 mL min^{-1}. Profiles collected prior to the elution peak were averaged and used as a buffer scattering for subtraction. Buffer subtraction, radius of gyration (R_g), forward scattering I(0) and P(r) were performed with programs from the ATSAS package [165]. Primary data reduction were performed using CHROMIX [208] yielding averaged scattering intensities I(q), where q is the momentum transfer ($q = 4\pi\sin\theta/\lambda$, with 2θ being the scattering angle and λ is the wavelength of the incident radiation). Data frames under the elution peak centred on the maximum in I(0) where the R_g values were the same within error and statistically identical as assessed using CorMap goodness-of-fit test and reduction χ^2-test (p-value > 0.01) [154] were selected and averaged for subsequent analysis. The two-dimensional scattering images were radially processed and the buffer scattering intensities subtracted using the program package PRIMUS [209]. The same program together with AUTORG were used to compute the R_g and I(0) by means of the Guinier approximation that automatically identify a suitable Guinier interval [155]. To remove minor parasitic interference at low angles resulting in a slight upturn in the Guinier plot, 30 and

26 data points were truncated from E2F1^{DCM}/DP1^{DCM}/DNA and E2F1^{DCM}/DP1^{DCM}/DNA/Rb dataset, respectively. GNOM, an indirect Fourier transform program, was used for automatic determination of the particle distance distribution function P(r) [210]. The estimated I(0)-based molecular weight (MW)[162]–[164] is in excellent agreement with MW obtained using SAXSMoW program, which is independent of the sample concentration and is based on the Porod volume determination (available at http://saxs.ifsc.usp.br).

4.6 Purification of Cy5-labeled Probes by anion-exchange chromatography

DNA probes derived from the human CDC2 promoter (nt -150 to -50, and nt -150 to -67 from TSS) were amplified by PCR from a cdc2-Pst plasmid [211] using the following set of primers:

Forward: 5'-[Cy5] TCTTTCTTTCGCGCTCTAGC-3'

Reverse: 5'-TAGCCCGTAGACTTTCAAAGC-3' (84 bps probe)

Reverse: 5'-GCTACCCGATTGGTGAATC-3' (100 bps probe)

PCR reactions were performed using Pfu DNA polymerase (GeneSpin). The PCR products were precipitated overnight at -20°C, resuspended in TE (10 mM Tris-Cl pH 8.0 and 1 mM EDTA), loaded into a His-Trap^{TM} 1 mL Q FF column (GE Healthcare) and purified by Ion Exchange chromatography over a linear NaCl gradient (up to 2 M NaCl) in TE. The pooled DNA peak was precipitated overnight and dissolved in TE. The total yield was efficient; 0.4 mg/10 mL of PCR product.

Site directed mutagenesis of the canonical E2F-recognition site was carried out by PCR mismatched primer using the following set of primers:

Forward: 5'-[Cy5] TCTTTCTTTCG**ta**CTCTAGCCACC-3'

Reverse: 5'-TAGCCCGTAGACTTTCAAAGC-3'

The 84-mer mutated probe was amplified by standard PCR procedure and purified by anion exchange chromatography (as described above).

4.7 Electrophoretic Mobility Shift Assay (EMSA) and *in vitro* competition analysis

For dose-response EMSAs, purified recombinant proteins were added to a binding mix, containing a fixed amount of Cy5-labeled probe with the following final reaction composition: 20 nM DNA probe, 20 mM Tris-HCl pH 8.0, 50 mM NaCl, 5 mM $MgCl_2$, 0.5 mM EDTA, 10% glycerol, 0.1 mg mL^{-1} BSA, 2.5 mM DTT. The reactions were incubated at 30°C for 30 minutes and loaded onto a 6% polyacrylamide gel in 0.25x TBE at 100 V for electrophoresis. In direct competition EMSAs, proteins were added to a binding mix containing the probe at a fixed concentration (20 nM), together with increasing amounts of unlabelled competitors at the indicated molar concentrations. In off-rate experiments, the binding mix supplemented with the protein was incubated at 30°C. After 30 minutes of incubation, 50-fold excess of

unlabelled competitor was added to the reaction mixture and aliquots were electrophoresed at different time points. Chemidoc™ MP imaging system (Bio-Rad) coupled with the ImageLab™ software were used in the acquisition and analysis of fluorescence signals. Mutant and WT oligos used in competition assays are listed in Table 2.

4.8 DNA complex assembly and sample preparation for Cryo-EM analysis

Trx-6xHis-NF-YA, NF-YB ΔQ/6xHis-NF-YC and 6xHis-USF1 were first purified separately by IMAC followed by gel filtration chromatography in high salt using HiLoad 10/300 Superdex™ 75 column (GE Healthcare) equilibrated in 10 mM Tris-Cl pH 8.0, 400 mM NaCl, 2 mM DTT. Trx-6xHis-NF-YA and NF-YB ΔQ/6xHis-NF-YC subunits were mixed at a molar ratio of 1.2:1. The assembled NF-Y was incubated overnight at 4°C with the thrombin protease (Thrombin CleanCleave Kit, Sigma-Aldrich) to cleave the N-terminal Trx-6×His-tag and 6×His-tag of NF-YA and NF-YC, respectively. The cleaved NF-Y was mixed with USF1 and DNA duplex (LTR or HOXB4) at a molar ratio of 1.25:1:1 and subjected to step dialysis to gradually decrease salt concentration to 50 mM NaCl, thus favouring DNA binding. The complex is then isolated by gel filtration using Hiprep 16/60 Sepharyl™ S-300 column (GE Healthcare) in a final buffer containing 10 mM Tris-Cl pH 8.0, 50 mM NaCl, 2 mM DTT. Purity and homogeneity of the sample were examined on SDS-PAGE gel stained with Coomassie brilliant blue. The pooled reconstituted complex peak fractions was subjected to Cryo-EM analysis.

5. 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] R. Andersson and A. Sandelin, "Determinants of enhancer and promoter activities of regulatory elements," *Nat. Rev. Genet.*, vol. 21, no. 2, pp. 71–87, 2020.

[3] R. G. Roeder, "50+ Years of Eukaryotic Transcription: an Expanding Universe of Factors and Mechanisms," *Nat. Struct. Mol. Biol.*, vol. 26, no. 9, pp. 783–791, 2019.

[4] R. Weinmann, H. J. Raskas, and R. G. Roeder, "Role of DNA dependent RNA polymerases II and III in transcription of the adenovirus genome late in productive infection," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 71, no. 9, pp. 3426–3430, 1974.

[5] B. Lenhard, A. Sandelin, and P. Carninci, "Metazoan promoters: Emerging characteristics and insights into transcriptional regulation," *Nat. Rev. Genet.*, vol. 13, no. 4, pp. 233–245, 2012.

[6] J. Ponjavic *et al.*, "Transcriptional and structural impact of TATA-initiation site spacing in mammalian core promoters," *Genome Biol.*, vol. 7, no. 8, 2006.

[7] S. T. Smale and D. Baltimore, "The 'initiator' as a transcription control element," *Cell*, vol. 57, no. 1, pp. 103–113, 1989.

[8] P. C. FitzGerald, D. Sturgill, A. Shyakhtenko, B. Oliver, and C. Vinson, "Comparative genomics of Drosophila and human core promoters," *Genome Biol.*, vol. 7, no. 7, 2006.

[9] T. W. Burke and J. T. Kadonaga, "The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila," *Genes Dev.*, vol. 11, no. 22, p. 3020—3031, 1997.

[10] V. Haberle and A. Stark, "Eukaryotic core promoters and the functional basis of transcription initiation," *Nat. Rev. Mol. Cell Biol.*, vol. 19, no. 10, pp. 621–637, 2018.

[11] H. M, "Molecular Genetics of the RNA polymerase II General Transcriptional Machinery," *Microbiol. Mol. Biol. Rev.*, vol. 62, no. 2, p. 39, 1998.

[12] 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 Å resolution," *Nature*, vol. 389, no. 6648, pp. 251–260, 1997.

[13] C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder, and T. J. Richmond, "Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution," *J. Mol. Biol.*, vol. 319, no. 5, pp. 1097–1113, 2002.

[14] S. Collection, "Nucleosome positioning and gene regulation: advances through genomics Cizhong," vol. 8, no. 5, pp. 583–592, 2016.

[15] Heather J. Szerlong and Jeffrey C. Hansen, "Nucleosome distribution and linker DNA: connecting nuclear function to dynamic chromatin structure," *Biochem Cell Biol.*, vol. 89, no. 1, pp. 24–34, 2011.

[16] D. E. Schones *et al.*, "Dynamic Regulation of Nucleosome Positioning in the Human Genome," *Cell*, vol. 132, no. 5, pp. 887–898, 2008.

[17] A. Sadakierska-Chudy and M. Filip, "A Comprehensive View of the Epigenetic Landscape. Part II: Histone Post-translational Modification, Nucleosome Level, and Chromatin Regulation by ncRNAs," *Neurotox. Res.*, vol. 27, no. 2, pp. 172–197, 2014.

[18] K. J. Polach and J. Widom, "Mechanism of protein access to specific DNA sequences in chromatin: A dynamic equilibrium model for gene regulation," *J. Mol. Biol.*, vol. 254, no. 2, pp. 130–149, 1995.

[19] K. J. Polach and J. Widom, "A model for the cooperative binding of eukaryotic regulatory proteins

to nucleosomal target sites," *J. Mol. Biol.*, vol. 258, no. 5, pp. 800–812, 1996.

[20] C. C. Adams and J. L. Workman, "Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative.," *Mol. Cell. Biol.*, vol. 15, no. 3, pp. 1405–1421, 1995.

[21] P. Varga-Weisz, "ATP-dependent chromatin remodeling factors: Nucleosome shufflers with many missions," *Oncogene*, vol. 20, no. 24 REV. ISS. 3, pp. 3076–3085, 2001.

[22] and Z. F. B. Maria L. Kireeva, Mikhail Kashlev, "RNA polymerase structure, function, regulation, dynamics, fidelity and roles in GENE EXPRESSION," *Chem Rev*, vol. 135, no. 2, pp. 612–615, 2013.

[23] J. Segall, T. Matsui, and R. G. Roeder, "Multiple factors are required for the accurate transcription of purified genes by RNA polymerase III.," *J. Biol. Chem.*, vol. 255, no. 24, pp. 11986–11991, 1980.

[24] R. G. Roeder, "The role of general initiation factors in transcription by RNA polymerase II," *Trends Biochem. Sci.*, vol. 21, no. 9, pp. 327–335, 1996.

[25] R. G. Roeder, "The eukaryotic transcriptional machinery : complexities and mechanisms unforeseen," *Nat. Med.*, vol. 9, no. 10, pp. 1239–1244, 2003.

[26] Z. Zhang *et al.*, "Rapid dynamics of general transcription factor TFIIB binding during preinitiation complex assembly revealed by single-molecule analysis," *Genes Dev.*, vol. 30, no. 18, pp. 2106–2118, 2016.

[27] and E. N. Yuan He, Chunli Yan, Jie Fang, Carla Inouye, Robert Tjian, Ivaylo Ivanov, "Near-atomic resolution visualization of human transcription promoter openin," *Nature*, vol. 176, no. 1, pp. 139–148, 2016.

[28] and K. A. Ginger W Muse, Daniel A Gilchrist, Sergei Nechaev, Ruchir Shah, Joel S Parker, Sherry F Grissom, Julia Zeitlinger, "RNA polymerase is poised for activation across the genome," *Nat. Genet.*, vol. 22, no. 5, p. 629, 2007.

[29] J. T. Lis, P. Mason, J. Peng, D. H. Price, and J. Werner, "P-TEFb kinase recruitment and function at heat shock loci," *Genes Dev.*, vol. 14, no. 7, pp. 792–803, 2000.

[30] and K. A. Telmo Henriques, Daniel A. Gilchrist, Sergei Nechaev, Michael Bern1, Ginger W. Muse, Adam Burkholder, David C. Fargo, "Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals," *Mol. Cell*, vol. 23, no. 1, pp. 1–25, 2013.

[31] O. Symmons *et al.*, "Functional and topological characteristics of mammalian regulatory domains," *Genome Res.*, vol. 24, no. 3, pp. 390–400, 2014.

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

[33] K. Ellwood, W. Huang, R. Johnson, and M. Carey, "Multiple Layers of Cooperativity Regulate Enhanceosome-Responsive RNA Polymerase II Transcription Complex Assembly," *Mol. Cell. Biol.*, vol. 19, no. 4, pp. 2613–2623, 1999.

[34] L. A. Mirny, "Nucleosome-mediated cooperativity between transcription factors," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 52, pp. 22534–22539, 2010.

[35] S. Vashee, K. Melchert, W. V. Ding, S. A. Johnstont, and T. Kodadek, "Evidence for two modes of cooperative DNA binding in vivo that do not involve, direct protein-protein interactions," *Curr. Biol.*, vol. 8, no. 8, pp. 452–458, 1998.

[36] 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.

[37] R. Mantovani, "A survey of 178 NF-Y binding CCAAT boxes" *Nucleic Acids Res.*, vol. 26, no. 5,

pp. 1135–1143, 1998.

[38] 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.

[39] 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.

[40] 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.

[41] W. Bi, L. Wu, F. Coustry, B. De Crombrugghe, and S. N. Maity, "DNA binding specificity of the CCAAT-binding factor CBF/NF-Y," *J. Biol. Chem.*, vol. 272, no. 42, pp. 26562–26572, 1997.

[42] 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. 1441–1455, 1999.

[43] O. Fornes *et al.*, "JASPAR 2020: update of the open-access database of transcription factor binding profiles," *Nucleic Acids Res.*, vol. 48, no. November 2019, pp. 87–92, 2019.

[44] R. Mantovani, "The molecular biology of the CCAAT-binding factor NF-Y," vol. 239, pp. 15–27, 1999.

[45] D. Dolfini, R. Gatta, and R. Mantovani, "NF-Y and the transcriptional activation of CCAAT promoters," *Crit. Rev. Biochem. Mol. Biol.*, vol. 47, no. 1, pp. 29–49, 2012.

[46] G. Li, H. Zhao, L. Wang, Y. Wang, X. Guo, and B. Xu, "The animal nuclear factor Y: an enigmatic and important heterotrimeric transcription factor.," *Am. J. Cancer Res.*, vol. 8, no. 7, pp. 1106–1125, 2018.

[47] 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.

[48] D. Dolfini, V. Andrioletti, and R. Mantovani, "Overexpression and alternative splicing of NF-YA in breast cancer," *Sci. Rep.*, vol. 9, no. 1, pp. 1–12, 2019.

[49] A. Di Silvio, C. Imbriano, and R. Mantovani, "Dissection of the NF-Y transcriptional activation potential," *Nucleic Acids Res.*, vol. 27, no. 13, pp. 2578–2584, 1999.

[50] D. Dolfini, R. Gatta, and R. Mantovani, "NF-Y and the transcriptional activation of CCAAT promoters," *Crit. Rev. Biochem. Mol. Biol.*, vol. 47, no. 1, pp. 29–49, 2012.

[51] V. Basile *et al.*, "NF-YA splice variants have different roles on muscle differentiation," *Biochim. Biophys. Acta - Gene Regul. Mech.*, vol. 1859, no. 4, pp. 627–638, 2016.

[52] 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.

[53] D. Libetti, A. Bernardini, S. Sertic, G. Messina, D. Dolfini, and R. Mantovani, "The Switch from NF-YAl to NF-YAs Isoform Impairs Myotubes Formation," *Cells*, vol. 9, no. 3, p. 789, 2020.

[54] 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.

[55] A. Bernardini, "Modulation of DNA-binding in transcription factors: the case of NF-Y and USF," *PhD dissertation thesi*s, 2017.

[56] 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.

[57] S. N. Maity, "NF-Y (CBF) regulation in specific cell types and mouse models," *Biochim Biophys Acta*, vol. 176, no. 5, pp. 139–148, 2017.

[58] P. Benatti *et al.*, "Direct non transcriptional role of NF-Y in DNA replication," *Biochim. Biophys. Acta - Mol. Cell Res.*, vol. 1863, no. 4, pp. 673–685, 2016.

[59] 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.

[60] L. L. Ly, H. Yoshida, and M. Yamaguchi, "Nuclear transcription factor Y and its roles in cellular processes related to human disease.," *Am. J. Cancer Res.*, vol. 3, no. 4, pp. 339–46, 2013.

[61] C. Imbriano, N. Gnesutta, and R. Mantovani, "The NF-Y/p53 liaison: Well beyond repression," *Biochim. Biophys. Acta - Rev. Cancer*, vol. 1825, no. 2, pp. 131–139, 2012.

[62] 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," no. September, pp. 191–202, 2006.

[63] P. Benatti, D. Dolfini, A. Vigan, M. Ravo, A. Weisz, and C. Imbriano, "Specific inhibition of NF-Y subunits triggers different cell proliferation defects," *Nucleic Acids Res.*, vol. 39, no. 13, pp. 5356–5368, 2011.

[64] J. Guo, L. M. Kong, A. F. Peng, X. H. Long, Y. Zhou, and Y. Shu, "Transcription factor NF-YA promotes a malignant phenotype by upregulating fatty acid synthase expression," *Mol. Med. Rep.*, vol. 14, no. 6, pp. 5007–5014, 2016.

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

[66] T. Yamanaka *et al.*, "Differential roles of NF-Y transcription factor in ER chaperone expression and neuronal maintenance in the CNS," *Sci. Rep.*, vol. 6, no. September, pp. 1–14, 2016.

[67] M. Nardini, N. Gnesutta, G. Donati, R. Gatta, C. Forni, A. Fossati, C. Vonrhein, D, Moras, C. Romier, M. Bolognesi, and R. Mantovani, "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.

[68] 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.

[69] S. L. Forsburg and L. Guarente, "Identification and characterization of HAP4: a third component of the CCAAT-bound HAP2/HAP3 heteromer.," *Genes Dev.*, vol. 3, no. 8, pp. 1166–1178, 1989.

[70] D. S. McNabb, K. A. Tseng, and L. Guarente, "The Saccharomyces cerevisiae Hap5p homolog from fission yeast reveals two conserved domains that are essential for assembly of heterotetrameric CCAAT-binding factor.," *Mol. Cell. Biol.*, vol. 17, no. 12, pp. 7008–7018, 1997.

[71] 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.

[72] 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.

[73] 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.

[74] A. Mayran and J. Drouin, "Pioneer transcription factors shape the epigenetic landscape," *J. Biol. Chem.*, vol. 293, no. 36, pp. 13795–13804, 2018.

[75] 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.

[76] A. Mayran *et al.*, "Pioneer and nonpioneer factor cooperation drives lineage specific chromatin opening," *Nat. Commun.*, vol. 10, no. 1, 2019.

[77] K. S. Zaret and J. S. Carroll, "Pioneer transcription factors : establishing competence for gene expression Parameters affecting transcription factor access to target sites in chromatin Initiating events in chromatin : pioneer factors bind first," *Genes Dev.*, pp. 2227–2241, 2011.

[78] 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.

[79] A. J. Oldfield, P. Yang, A. E. Conway, S. Cinghu, J. M. Freudenberg, S. Yellaboina, and R. Jothi, "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.

[80] A. J. Oldfield *et al.*, "NF-Y controls fidelity of transcription initiation at gene promoters through maintenance of the nucleosome-depleted region," *Nat. Commun.*, vol. 10, no. 1, pp. 1–12, 2019.

[81] 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.

[82] 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.

[83] 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.

[84] G. Caretti, F. Cocchiarella, C. Sidoli, J. Villaed, M. Peretti, W. Reith, and R. Mantovani, "Dissection of functional NF-Y-RFX cooperative interactions on the MHC class II Ea promoter," *J. Mol. Biol.*, vol. 302, no. 3, pp. 539–552, 2000.

[85] M. M. Magaña, S. H. Koo, H. C. Towle, and T. F. Osborne, "Different sterol regulatory element-binding protein-1 isoforms utilize distinct co-regulatory factors to activate the promoter for fatty acid synthase," *J. Biol. Chem.*, vol. 275, no. 7, pp. 4726–4733, 2000.

[86] 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.

[87] Q. S. Zhu, B. Qian, and D. Levy, "CCAAT/enhancer-binding protein α (C/EBPα) 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. 29902–29910, 2004.

[88] H. Yoshida *et al.*, "Endoplasmic Reticulum Stress-Induced Formation of Transcription Factor Complex ERSF Including NF-Y (CBF) and Activating Transcription Factors 6 That Activates the Mammalian Unfolded Protein Response," vol. 21, no. 4, pp. 1239–1248, 2001.

[89] 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.

[90] 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. 3737–3751, 2005.

[91] 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, pp. 1–25, 2015.

[92] 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, pp. 171–176, 2013.

[93] 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.

[94] 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.

[95] A. R. Ferré-D'Amaré, G. C. Prendergast, E. B. Ziff, and S. K. Burley, "Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain," *Nature*, vol. 363, no. 6424, pp. 38–45, 1993.

[96] 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. 180–189, 1994.

[97] 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. 1367–1375, 1996.

[98] C. Craik, "Assembly of b/HLH/z proteins cMyc, Max and Mad1 with Cognate DNA: Importance of Protein-Protein and Protein-DNA Interactions," *Jianzhong Hu, Anamika Banerjee, Dixie J. Goss*, vol. 23, no. 1, pp. 1–7, 2005.

[99] 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.

[100] T. G. H. D. Yoshihiro Ito, Y. Zhang, Smit Dangaria, and X. Luan, "NF-Y and USF1 transcription factor binding to CCAAT box and E-box elements activates the CP27 promoter," *Gene*, vol. 23, no. 1, pp. 1–15, 2010.

[101] 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, pp. 1–9, 2003.

[102] H. Q. 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. 23, no. 1, pp. 1–7, 2011.

[103] I. Kovesdi, R. Reichel, and J. R. Nevins, "Role of an adenovirus E2 promoter binding factor in ElA-mediated coordinate gene control," vol. 84, no., pp. 2180–2184, 1987.

[104] A. S. Yee, R. Reichel, I. Kovesdi, and J. R. Nevins, "Promoter interaction of the E1A-inducible factor E2F and its potential role in the formation of a multi-component complex.," *EMBO J.*, vol. 6, no. 7, pp. 2061–2068, 1987.

[105] D. K. Dimova and N. J. Dyson, "The E2F transcriptional network: Old acquaintances with new faces," *Oncogene*, vol. 24, no. 17, pp. 2810–2826, 2005.

[106] M. R. Adams, R. Sears, F. Nuckolls, G. Leone, and J. R. Nevins, "Complex Transcriptional Regulatory Mechanisms Control Expression of the E2F3 Locus," *Mol. Cell. Biol.*, vol. 20, no. 10, pp. 3633–3639, 2000.

[107] G. Leone *et al.*, "Identification of a Novel E2F3 Product Suggests a Mechanism for Determining Specificity of Repression by Rb Proteins," *Mol. Cell. Biol.*, vol. 20, no. 10, pp. 3626–3632, 2000.

[108] J. DeGregori, "The genetics of the E2F family of transcription factors: Shared functions and unique roles," *Biochim. Biophys. Acta - Rev. Cancer*, vol. 1602, no. 2, pp. 131–150, 2002.

[109] J. J. H. & N. B. L. T. Rowena Girling, Janet F. Partridge, Lasantha R. Bandara, Neil Burden,

Nicholas F. Totty, "A new component of the transcription factor DRTF1/E2F," *Nature*, vol. 365, no., p. 72310, 1993.

[110] H. Qiao *et al.*, "Human TFDP3, a novel DP protein, inhibits DNA binding and transactivation by E2F," *J. Biol. Chem.*, vol. 282, no. 1, pp. 454–466, 2007.

[111] A. Milton *et al.*, "A functionally distinct member of the DP family of E2F subunits," *Oncogene*, vol. 25, no. 22, pp. 3212–3218, 2006.

[112] E. Ormondroyd, S. de la Luna, and N. B. La Thangue, "A new member of the DP family, DP-3, with distinct protein products suggests a regulatory role for alternative splicing in the cell cycle transcription factor DRTF1/E2F ," *Oncogene* , vol. 11, no. 8, p. 1437, 1995.

[113] C. Attwooll *et al.*, "A novel repressive E2F6 complex containing the polycomb group protein, EPC1, that interacts with EZH2 in a proliferation-specific manner," *J. Biol. Chem.*, vol. 280, no. 2, pp. 1199–1208, 2005.

[114] E. Morgunova *et al.*, "Structural insights into the DNA-binding specificity of E2F family transcription factors," *Nat. Commun.*, vol. 6, pp. 4–11, 2015.

[115] C. Chen and A. D. Wells, "Comparative analysis of E2F family member oncogenic activity," *PLoS One*, vol. 2, no. 9, pp. 1–7, 2007.

[116] N. S. Moon and N. Dyson, "E2F7 and E2F8 Keep the E2F Family in Balance," *Dev. Cell*, vol. 14, no. 1, pp. 1–3, 2008.

[117] S. Van Den Heuvel and N. J. Dyson, "Conserved functions of the pRB and E2F families," *Nat. Rev. Mol. Cell Biol.*, vol. 9, no. 9, pp. 713–724, 2008.

[118] G. L. Hui-Zi Chen, Shih-Yin Tsai, "Emerging roles of E2Fs in cancer: an exit from cell cycle control," *Nat. Rev. Cancer*, vol. 135, no. 2, pp. 612–615, 2009.

[119] J. L. Chong *et al.*, "E2f1-3 switch from activators in progenitor cells to repressors in differentiating cells," *Nature*, vol. 462, no. 7275, pp. 930–934, 2009.

[120] J. Hsu and J. Sage, "Novel functions for the transcription factor E2F4 in development and disease," *Cell Cycle*, vol. 15, no. 23, pp. 3183–3190, 2016.

[121] J. Hsu *et al.*, "E2F4 regulates transcriptional activation in mouse embryonic stem cells independently of the RB family," *Nat. Commun.*, vol. 10, no. 1, 2019.

[122] L. N. Kent and G. Leone, "The broken cycle: E2F dysfunction in cancer," *Nat. Rev. Cancer*, vol. 19, no. 6, pp. 326–338, 2019.

[123] M. A. Ciemerych and P. Sicinski, "Cell cycle in mouse development," *Oncogene*, vol. 24, no. 17, pp. 2877–2898, 2005.

[124] T. J. Liban, E. M. Medina, S. Tripathi, S. Sengupta, R. W. Henry, N. E. Buchler, and S. M. Rubin, "Conservation and divergence of C-terminal domain structure in the retinoblastoma protein family," *Proc. Natl. Acad. Sci.*, vol. 114, no. 19, pp. 4942–4947, 2017.

[125] J. A. DeCaprio, "How the Rb tumor suppressor structure and function was revealed by the study of Adenovirus and SV40," *Virology*, vol. 384, no. 2, pp. 274–284, 2009.

[126] S. A. Henley and F. A. Dick, "The retinoblastoma family of proteins and their regulatory functions in the mammalian cell division cycle," *Cell Div.*, vol. 7, no. 1, p. 10, 2012.

[127] R. Verona, K. Moberg, S. Estes, M. Starz, J. P. Vernon, and J. A. Lees, "E2F activity is regulated by cell cycle-dependent changes in subcellular localization.," *Mol. Cell. Biol.*, vol. 17, no. 12, pp. 7268–7282, 1997.

[128] S. Polager and D. Ginsberg, "E2F - at the crossroads of life and death," *Trends Cell Biol.*, vol. 18, no. 11, pp. 528–535, 2008.

[129] S. J. Field, F. Y. Tsai, F. Kuo, A. M. Zubiaga, W. G. Kaelin, J. David, M. Livingston, S. H. Orkin

and M. E. Greenberg, "E2F-1 Functions in mice to promote apoptosis and suppress proliferation," *Cell*, vol. 85, no. 4, pp. 549–561, 1996.

[130] L. Yamasaki, T. Jacks, R. Bronson, E. Goillot, E. Harlow, and N. J. Dyson, "Tumor induction and tissue atrophy in mice lacking E2F-1," *Cell*, vol. 85, no. 4, pp. 537–548, 1996.

[131] H. Sierra, M. Cordova, C.-S. J. Chen, and M. Rajadhyaksha, "Confocal Imaging–Guided Laser Ablation of Basal Cell Carcinomas: An Ex Vivo Study," *J. Invest. Dermatol.*, vol. 135, no. 2, pp. 612–615, 2015.

[132] D. P. Hollern, J. Honeysett, R. D. Cardiff, and E. R. Andrechek, "The E2F Transcription Factors Regulate Tumor Development and Metastasis in a Mouse Model of Metastatic Breast Cancer," *Mol. Cell. Biol.*, vol. 34, no. 17, pp. 3229–3243, 2014.

[133] N. Zheng, E. Fraenkel, C. O. Pabo, and N. P. Pavletich, "Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP," *Genes Dev.*, vol. 13, no. 6, pp. 666–674, 1999.

[134] R. Girling *et al.*, "Correction: A new component of the transcription factor DRTF1/E2F," *Nature*, vol. 365, no. 6445, p. 468, 1993.

[135] P. H. Giangrande, T. C. Hallstrom, C. Tunyaplin, K. Calame, and J. R. Nevins, "Identification of E-Box Factor TFE3 as a Functional Partner for the E2F3 Transcription Factor," *Mol. Cell. Biol.*, vol. 23, no. 11, pp. 3707–3720, 2003.

[136] T. C. Hallstrom and J. R. Nevins, "Specificity in the activation and control of transcription factor E2F-dependent apoptosis," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 19, pp. 10848–10853, 2003.

[137] W. D. Cress and J. R. Nevins, "A role for a bent DNA structure in E2F-mediated transcription activation.," *Mol. Cell. Biol.*, vol. 16, no. 5, pp. 2119–2127, 1996.

[138] S. M. Rubin, A. L. Gall, N. Zheng, and N. P. Pavletich, "Structure of the Rb C-terminal domain bound to E2F1-DP1: A mechanism for phosphorylation-induced E2F release," *Cell*, vol. 123, no. 6, pp. 1093–1106, 2005.

[139] F. A. Dick and S. M. Rubin, "Molecular mechanisms underlying RB protein function," *Nat. Rev. Mol. Cell Biol.*, vol. 14, no. 5, pp. 297–306, 2013.

[140] M. Hassler *et al.*, "Crystal Structure of the Retinoblastoma Protein N Domain Provides Insight into Tumor Suppression, Ligand Interaction, and Holoprotein Architecture," *Mol. Cell*, vol. 28, no. 3, pp. 371–385, 2007.

[141] J. O. Lee, A. A. Russo, and N. P. Pavletich, "Structure of the retinoblastoma tumour-suppressor pocket domain bound to a peptide from HPV E7," *Nature*, vol. 391, no. 6670, pp. 859–865, 1998.

[142] C. Lee, J. H. Chang, H. S. Lee, and Y. Cho, "Structural basis for the recognition of the E2F transactivation domain by the retinoblastoma tumor suppressor," *Genes Dev.*, vol. 16, no. 24, pp. 3199–3212, 2002.

[143] S. T. Hani Goodarzi, Olivier Elemento, "Revealing global regulatory perturbations across human cancers," *Mol. Cell*, vol. 23, no. 1, pp. 1–7, 2009.

[144] A. C. Brygida Bisikirska, Mukesh Bansal, Yao Shen, Julie Teruya-Feldstein, Raju Chaganti, "Elucidation and pharmacological targeting of novel molecular drivers of follicular lymphoma progression," *Cancer Res.*, vol. 176, no. 10, pp. 139–148, 2016.

[145] B. K. Lee, A. A. Bhinge, and V. R. Iyer, "Wide-ranging functions of E2F4 in transcriptional activation and repression revealed by genome-wide analysis," *Nucleic Acids Res.*, vol. 39, no. 9, pp. 3558–3573, 2011.

[146] M. Bieda, X. Xu, M. A. Singer, R. Green, and P. J. Farnham, "Unbiased location analysis of E2F1-binding sites suggests a widespread role for E2F1 in the human genome," *Genome Res.*, vol. 16, no. 5, pp. 595–605, 2006.

[147] P. R. Van Ginkel, K. M. Hsiao, H. Schjerven, and P. J. Farnham, "E2F-mediated growth regulation requires transcription factor cooperation," *J. Biol. Chem.*, vol. 272, no. 29, pp. 18367–18374, 1997.

[148] X. Jiang, J. R. Nevins, I. Shats, and J. T. Chi, "E2F1-mediated induction of NFYB attenuates apoptosis via joint regulation of a pro-survival transcriptional program," *PLoS One*, vol. 10, no. 6, pp. 1–16, 2015.

[149] A. Gurtner *et al.*, "Transcription factor NF-Y induces apoptosis in cells expressing wild-type p53 through E2F1 upregulation and p53 activation," *Cancer Res.*, vol. 70, no. 23, pp. 9711–9720, 2010.

[150] W. Zhu, P. H. Giangrande, and J. R. Nevins, "E2Fs link the control of G1/S and G2/M transcription," *EMBO J.*, vol. 23, no. 23, pp. 4615–4626, 2004.

[151] M. Lorenzo, N. Gnesutta, R. Mantovani, "The interplay among NF-Y and E-BOX Transcription Factors: Myc, MAX and USF," *PhD dissertation thesi*s, 2014.

[152] A. Bernardini, N. Gnesutta, R. Mantovani, "Modulation of DNA-binding in transcription factors: the case of NF-Y and USF," *PhD dissertation thesi*s, 2017.

[153] H. E. Huber, G Edwards, P J Goodhart, D R Patrick, P S Husang, M Ivey-Hoyle, S F Barnett, A Oliff and D C Heimbrook, "Transcription factor E2F binds DNA as a heterodimer," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 90, no. 8, pp. 3525–3529, 1993.

[154] D. Franke, C. M. Jeffries, and D. I. Svergun, "Correlation Map, a goodness-of-fit test for one-dimensional X-ray scattering spectra," *Nat. Methods*, vol. 12, no. 5, pp. 419–422, 2015.

[155] A. Guinier, "Diffraction of X-rays of very small angles – application to the study of ultramicroscopic phenomenon," *Ann. Phys*, vol. 58, no. 7, 1939.

[156] O. Glatter and O. Kratky, Small angle x-ray scattering. *Academic Press*, 1982.

[157] D. Durand, C. Vivè, D. Cannella, J. Pérez, E. Pebay-Peyroula, P. Vachette, F. Fieschi, "NADPH oxidase activator p67phox behaves in solution as a multidomain protein with semi-flexible linkers," *J. Struct. Biol.*, vol. 169, no. 1, pp. 45–53, 2010.

[158] J. Trewhella *et al.*, "2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution: An update," *Acta Crystallogr. Sect. D Struct. Biol.*, vol. 73, no. 9, pp. 710–728, 2017.

[159] R. P. Rambo and J. A. Tainer, "Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law," *Biopolymers*, vol. 95, no. 8, pp. 559–571, 2011.

[160] M. Hammel, "Validation of macromolecular flexibility in solution by small-angle X-ray scattering (SAXS)," *Eur. Biophys. J.*, vol. 41, no. 10, pp. 789–799, 2012.

[161] V. Receveur-Brechot and D. Durand, "How Random are Intrinsically Disordered Proteins? A Small Angle Scattering Perspective," *Current Protein & Peptide Science*, vol. 13, no. 1. pp. 55–75, 2012.

[162] N. R. Hajizadeh, D. Franke, C. M. Jeffries, and D. I. Svergun, "Consensus Bayesian assessment of protein molecular mass from solution X-ray scattering data," *Sci. Rep.*, vol. 8, no. 1, pp. 1–13, 2018.

[163] V. Piiadov, E. Ares de Araújo, M. Oliveira Neto, A. F. Craievich, and I. Polikarpov, "SAXSMoW 2.0: Online calculator of the molecular weight of proteins in dilute solution from experimental SAXS data measured on a relative scale," *Protein Sci.*, vol. 28, no. 2, pp. 454–463, 2019.

[164] H. Fischer, M. De Oliveira Neto, H. B. Napolitano, I. Polikarpov, and A. F. Craievich, "Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale," *J. Appl. Crystallogr.*, vol. 43, no. 1, pp. 101–109, 2010.

[165] D. Franke *et al.*, "ATSAS 2.8: A comprehensive data analysis suite for small-angle scattering from

macromolecular solutions," *J. Appl. Crystallogr.*, vol. 50, pp. 1212–1225, 2017.

[166] M. Levitt, "Nature of the protein universe," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 27, pp. 11079–11084, 2009.

[167] E. Papaleo, G. Saladino, M. Lambrughi, K. Lindorff-Larsen, F. L. Gervasio, and R. Nussinov, "The Role of Protein Loops and Linkers in Conformational Dynamics and Allostery," *Chem. Rev.*, vol. 116, no. 11, pp. 6391–6423, 2016.

[168] H. Van Den Bedem and J. S. Fraser, "Integrative, dynamic structural biology at atomic resolution - It's about time," *Nat. Methods*, vol. 12, no. 4, pp. 307–318, 2015.

[169] C. Paissoni, A. Jussupow, and C. Camilloni, "Determination of Protein Structural Ensembles by Hybrid-Resolution SAXS Restrained Molecular Dynamics," *J. Chem. Theory Comput.*, vol. 16, no. 4, pp. 2825–2834, 2020.

[170] M. Ronzio and D. Dolfini, A. Bernardini, G. Pavesi, and R. Mantovani, "On the NF-Y regulome as in ENCODE (2019)," 2020.

[171] H. Liu *et al.*, "Redeployment of Myc and E2f1-3 drives Rb deficient cell cycles," *Nat. Cell Biol.*, vol. 17, no. 8, pp. 1036–1048, 2015.

[172] F. Oswald, H. Lovec, T. Möröy, and M. Lipp, "E2F-dependent regulation of human MYC: trans-activation by cyclins D1 and A overrides tumour suppressor protein functions," *Oncogene*, vol. 9, no. 7, p. 2029—2036, 1994.

[173] B. A. Slomiany, K. L. D'Arigo, M. M. Kelly, and D. T. Kurtz, "C/EBPα Inhibits Cell Growth via Direct Repression of E2F-DP-Mediated Transcription," *Mol. Cell. Biol.*, vol. 20, no. 16, pp. 5986–5997, 2000.

[174] E. Kowenz-Leutz, A. Schuetz, Q. Liu, M. Knoblich, U. Heinemann, and A. Leutz, "Functional interaction of CCAAT/enhancer-binding-protein-α basic region mutants with E2F transcription factors and DNA," *Biochim. Biophys. Acta - Gene Regul. Mech.*, vol. 1859, no. 7, pp. 841–847, 2016.

[175] K. Zaragoza, V. Bégay, A. Schuetz, U. Heinemann, and A. Leutz, "Repression of Transcriptional Activity of C/EBPα by E2F-Dimerization Partner Complexes," *Mol. Cell. Biol.*, vol. 30, no. 9, pp. 2293–2304, 2010.

[176] W. Yu *et al.*, "First somatic mutation of E2F1 in a critical DNA binding residue discovered in well-differentiated papillary mesothelioma of the peritoneum.," *Genome Biol.*, vol. 12, no. 9, p. R96, 2011.

[177] S. Munro, U. Oppermann, and N. B. La Thangue, "Pleiotropic effect of somatic mutations in the E2F subunit DP-1 gene in human cancer," *Oncogene*, vol. 33, no. 27, pp. 3594–3603, 2014.

[178] H. M. Dingerdissen, J. Torcivia-Rodriguez, Y. Hu, T. C. Chang, R. Mazumder, and R. Kahsay, "BioMuta and BioXpress: Mutation and expression knowledgebases for cancer biomarker discovery," *Nucleic Acids Res.*, vol. 46, no. D1, pp. D1128–D1136, 2018.

[179] S. Bamford *et al.*, "The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website," *Br. J. Cancer*, vol. 91, no. 2, pp. 355–358, 2004.

[180] J. Barretina *et al.*, "The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity," *Nature*, vol. 483, no. 7391, pp. 603–607, 2012.

[181] C. G. A. R. N. John N. Weinstein, Eric A. Collisson, Gordon B. Mills, Kenna M. Shaw, Brad A. Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, Joshua M. Stuart, "The Cancer Genome Atlas Pan-Cancer Analysis Project," *Nat. Genet.*, pp. 1113–1120, 2013.

[182] The International Cancer Genome Consortium, "International network of cancer genome projects," *Nature*, vol. 464(7291), pp. 993–998, 2010.

[183] J. F. Sayılgan, T. Haliloğlu, and M. Gönen, "Protein dynamics analysis reveals that missense

mutations in cancer-related genes appear frequently on hinge-neighboring residues," *Proteins Struct. Funct. Bioinforma.*, vol. 87, no. 6, pp. 512–519, 2019.

[184] M. C. Louie, J. X. Zou, A. Rabinovich, and H.-W. Chen, "ACTR/AIB1 Functions as an E2F1 Coactivator To Promote Breast Cancer Cell Proliferation and Antiestrogen Resistance," *Mol. Cell. Biol.*, vol. 24, no. 12, pp. 5157–5171, 2004.

[185] Q. Xie *et al.*, "E2F transcription factor 1 regulates cellular and organismal senescence by inhibiting forkhead box O transcription factors," *J. Biol. Chem.*, vol. 289, no. 49, pp. 34205–34213, 2014.

[186] G. Xiong and R. Xu, "RORα Binds to E2F1 To Inhibit Cell Proliferation and Regulate Mammary Gland Branching Morphogenesis," *Mol. Cell. Biol.*, vol. 34, no. 16, pp. 3066–3075, 2014.

[187] S. Y. Lin, A. R. Black, D. Kostic, S. Pajovic, C. N. Hoover, and J. C. Azizkhan, "Cell cycle-regulated association of E2F1 and Sp1 is related to their functional interaction.," *Mol. Cell. Biol.*, vol. 16, no. 4, pp. 1668–1675, 1996.

[188] J. Cartier *et al.*, "Cellular inhibitor of apoptosis protein-1 (cIAP1) can regulate E2F1 transcription factor-mediated control of cyclin transcription," *J. Biol. Chem.*, vol. 286, no. 30, pp. 26406–26417, 2011.

[189] C. Wang *et al.*, "Interactions between E2F1 and SirT1 regulate apoptotic response to DNA damage," *Nat. Cell Biol.*, vol. 8, no. 9, pp. 1025–1031, 2006.

[190] A. Datta *et al.*, "ARF Directly Binds DP1: Interaction with DP1 Coincides with the G1 Arrest Function of ARF," *Mol. Cell. Biol.*, vol. 25, no. 18, pp. 8024–8036, 2005.

[191] T. S. Sørensen, R. Girling, C. W. Lee, J. Gannon, L. R. Bandara, and N. B. La Thangue, "Functional interaction between DP-1 and p53.," *Mol. Cell. Biol.*, vol. 16, no. 10, pp. 5888–5895, 1996.

[192] M. Malumbres, "Cyclin-dependent kinases," *Genome Biol.*, vol. 15, no. 6, pp. 1–10, 2014.

[193] G. Banyai, F. Baïdi, D. Coudreuse, and Z. Szilagyi, "Cdk1 activity acts as a quantitative platform for coordinating cell cycle progression with periodic transcription," *Nat. Commun.*, vol. 7, 2016.

[194] A. Satyanarayana and P. Kaldis, "Mammalian cell-cycle regulation: Several cdks, numerous cyclins and diverse compensatory mechanisms," *Oncogene*, vol. 28, no. 33, pp. 2925–2939, 2009.

[195] S. Dalton, "Cell cycle regulation of the human cdc2 gene.," *EMBO J.*, vol. 11, no. 5, pp. 1797–1804, 1992.

[196] M. G. Lee and P. Nurse, "Complementation used to clone a human homologue of the fission yeast cell cycle control gene cdc2," *Nature*, vol. 327, no. 6117, pp. 31–35, 1987.

[197] S. Tommasi and G. P. Pfeifer, "In vivo structure of the human cdc2 promoter: release of a p130-E2F-4 complex from sequences immediately upstream of the transcription initiation site coincides with induction of cdc2 expression.," *Mol. Cell. Biol.*, vol. 15, no. 12, pp. 6901–6913, 1995.

[198] R. S. M. Judith F. Kribelbauer, Chaitanya Rastogi, Harmen J. Bussemaker, "Low-Affinity Binding Sites and the Transcription Factor Specificity Paradox in Eukaryotes," *Annu Rev Cell Dev Biol.*, Nov. 2019.

[199] A. R. Cao, R. Rabinovich, M. Xu, X. Xu, V. X. Jin, and P. J. Farnham, "Genome-wide analysis of transcription factor E2F1 mutant proteins reveals that N- and C-terminal protein interaction domains do not participate in targeting E2F1 to the human genome," *J. Biol. Chem.*, vol. 286, no. 14, pp. 11985–11996, 2011.

[200] A. Rabinovich, V. X. Jin, R. Rabinovich, X. Xu, and P. J. Farnham, "E2F in vivo binding specificity: Comparison of consensus versus nonconsensus binding sites," *Genome Res.*, vol. 18, no. 11, pp. 1763–1777, 2008.

[201] W. R. Lee, C. C. Chen, S. Liu, and S. Safe, "17β-estradiol (E2) induces cdc25A gene expression in

breast cancer cells by genomic and non-genomic pathways," *J. Cell. Biochem.*, vol. 99, no. 1, pp. 209–220, 2006.

[202] J. Gaudet and S. E. Mango, "Regulation of Organogenesis by the Caenorhabditis elegans FoxA Protein PHA-4," *Science (80-.).*, vol. 295, no. 5556, pp. 821 LP – 825, 2002.

[203] A. Zandvakili, I. Campbell, L. M. Gutzwiller, M. T. Weirauch, and B. Gebelein, "Degenerate Pax2 and Senseless binding motifs improve detection of low-affinity sites required for enhancer specificity," *PLoS Genet.*, vol. 14, no. 4, pp. 1–25, 2018.

[204] F. Giavazzi *et al.*, "Multispot, label-free biodetection at a phantom plastic-water interface," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 23, pp. 9350–9355, 2013.

[205] Z. Liu and R. Tjian, "Visualizing transcription factor dynamics in living cells," *J. Cell Biol.*, vol. 217, no. 4, pp. 1181–1191, 2018.

[206] M. L. Diebold, S. Fribourg, M. Koch, T. Metzger, and C. Romier, "Deciphering correct strategies for multiprotein complex assembly by co-expression: Application to complexes as large as the histone octamer," *J. Struct. Biol.*, vol. 175, no. 2, pp. 178–188, 2011.

[207] 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.

[208] A. Panjkovich and D. I. Svergun, "CHROMIXS: Automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data," *Bioinformatics*, vol. 34, no. 11, pp. 1944–1946, 2018.

[209] P. V. Konarev, V. V. Volkov, A. V. Sokolova, M. H. J. Koch, and D. I. Svergun, "PRIMUS: A Windows PC-based system for small-angle scattering data analysis," *J. Appl. Crystallogr.*, vol. 36, no. 5, pp. 1277–1282, 2003.

[210] D. I. Svergun, "Determination of the regularization parameter in indirect-transform methods using perceptual criteria," *J. Appl. Crystallogr.*, vol. 25, no. pt 4, pp. 495–503, 1992.

[211] M. Minuzzo *et al.*, "Selective Effects of the Anticancer Drug Yondelis (ET-743) on Cell-Cycle Promoters," *Mol. Pharmacol.*, vol. 68, no. 5, pp. 1496 LP – 1503, 2005.