# High Conformational Flexibility of the E2F1/DP1/DNA complex.

Dana Saad[1], Cristina Paissoni[1], Antonio Chaves-Sanjuan[1], Marco Nardini[1], Roberto Mantovani[1], Nerina Gnesutta[1,*] and Carlo Camilloni[1,*].

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

* To whom correspondence should be addressed. Email: carlo.camilloni@unimi.it Correspondence may also be addressed to nerina.gnesutta@unimi.it

Graphical Abstract
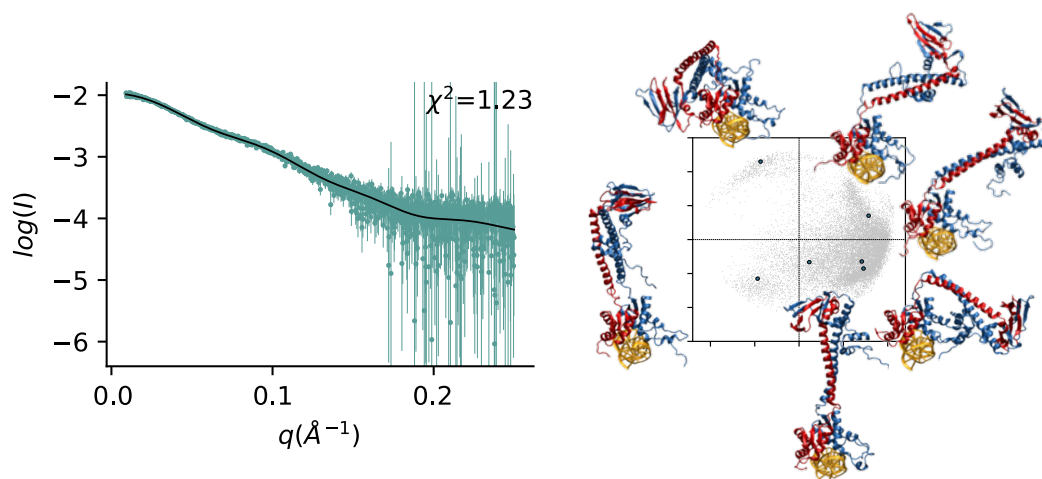
Highlights

- The structure and dynamics of the E2F1/DP1/DNA complex is modelled
- E2F1 and DP1 play a different role in the dynamics of the system
- DP1 flexibility may be relevant for protein-protein and protein-DNA interactions.

ABBREVIATIONS: TF: Transcription Factor; DBD: DNA-binding Domain; TAD: Trans-Activation Domain; CC: coiled-coil; MB: marked box; DCM: DBD-CC-MB; SEC: size exclusion chromatography; SAXS: Small-Angle X-ray Scattering; MD: Molecular Dynamics; RMSF: Root mean square fluctuations; IMAC: ion-metal affinity chromatography

ABSTRACT

The E2F1 transcription factor is a master regulator of cell-cycle progression whose uncontrolled activation contributes to tumor cells growth. E2F1 binds DNA as a heterodimer with DP partners, resulting in a multi-domain quaternary-structure complex composed of DNA binding domains, a coiled coil domain and a marked box domain separated by short linkers. Building on the 3D knowledge of the single domains of E2F and DPs, we characterized the structure and dynamics of the complete E2F1/DP1/DNA complex by a combination of small-angle X-ray scattering and molecular dynamics simulations. It shows an asymmetric contribution of the dynamics of the two proteins. Namely, the coiled-coil domain leans toward the DP1 side of the complex; the DP1 loop between $\alpha2$ and $\alpha3$ of the DBD partially populates a helical structure leaning far from the DNA and in the same direction of the coiled-coil domain; and the N-terminal disordered region of DP1, rich in basic residues, contributes to DNA binding stabilization. Intriguingly, tumor mutations in the flexible regions of the complex suggest that perturbation of protein dynamics could affect protein function in a context-dependent way. Our data suggest fundamental contributions of DP proteins in distinct aspects of E2F biology.

KEYWORDS

Transcription Factor; E2F; DP1; SAXS; Molecular Dynamics

INTRODUCTION

RNA production is regulated by a plethora of sequence-specific Transcription Factors -TFs- that recognize short DNA sequences in enhancers and promoters of genes, forming modules dictated by the arrangement of the DNA sites. Binding of TFs to regulatory regions leads to the recruitment of cofactors endowed with chromatin remodeling and histone modification activities, which make chromatin accessible for RNA Polymerase processivity [1]. TFs typically require a minimum of a sequence-specific DNA-binding domain (DBD) and a Trans-Activation domain (TAD).

E2Fs constitute a family of eight evolutionarily conserved TFs, whose founding member -E2F1- was originally identified as a master regulator of cell-cycle progression [2,3]. In particular, E2F1 is essential for the G1/S transition and S-phase progression [4]; E2F4, as part of the DREAM complex, is also involved in repression of G2/M genes in G1/S, and cell-cycle exit during terminal differentiation [5], but it is an activator in mES cells [6]. In addition to the regulation of cell-cycle genes, E2F1 has a global role in the regulation of apoptosis, the DNA-damage response, metabolism, and differentiation [7]. E2F1 is usually catalogued as an activator, but depending on the cellular context, it can be also a repressor in differentiation [8]. E2F1, among others, is targeted by the Retinoblastoma -Rb- and by other "pocket" proteins, as part of repressive complexes that include other corepressors: genetic mutations in Rb, as found in tumors, eliminate or weaken interaction, leading to uncontrolled E2F1 activation, cell cycle progression and cell growth [9]. Somatic mutations in the E2F1 gene are relatively rare in cancer, but, as many other family members, the gene expression is often upregulated in transformed cells [10].

E2Fs bind DNA as heterodimers with the structurally related DP1/2/3 subunits. In addition to DBDs, E2Fs have a coiled-coil domain (CC) involved in dimerization with the partner DPs, a marked-box (MB) domain responsible for the interaction with pocket proteins, and a C-terminal TAD. A similar structural scheme is shared by DP proteins, except that they lack a TAD. The DBD and CC domains are typically connected by a short linker [3]. From the structural viewpoint, two parts have been characterized by crystallography: the minimal DBDs of E2F4/DP2, and of E2F8 in complex with "canonical" DNA sites [11,12]; the CC-MB domains of E2F1/DP1 and of E2F5/DP1, including in complex with Rb pocket proteins peptides [13,14]. Such structures refer to DBDs separated from CC-MB: a coherent view of the relative topology of the linked domains is not available to date (Fig. 1). For this reason, we set out to produce and characterize, both biochemically and structurally, the ensembles of E2F1 and DP1 DBD, CC, and MB domains. In so doing, we provide evidence for a high level of flexibility of the heterodimer, including a newly identified region stabilizing DNA contacts.
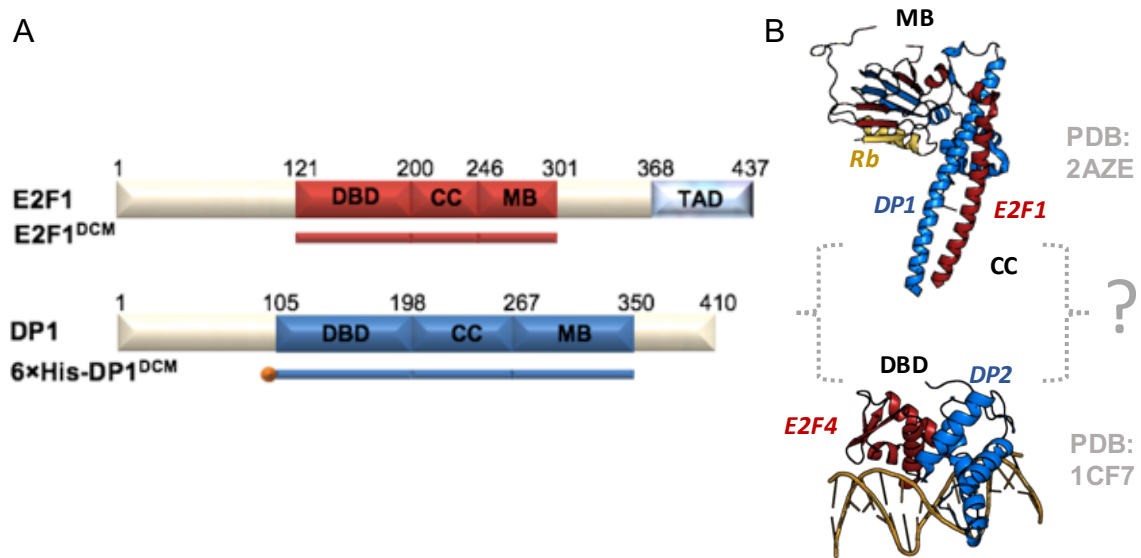
Figure 1. (A) Schematic representation of E2F1 and DP1 domain organization with their respective amino acid sequence numbering, and recombinant protein constructs used in this work. The conserved domains of E2F1 and DP1 are depicted in red and blue, respectively. The 6×His-tag is indicated as an orange sphere. **(B)** Cartoon representation of E2F/DP homologous proteins from available PDBs: 2AZE (top), corresponding to the marked-box and coiled-coil domains of E2F1/DP1, in complex with the Rb peptide; and 1CF7 (bottom), including the DBD of E2F4/DP2 bound to 15-bp DNA.

RESULTS

The E2F1/DP1/DNA complex displays a spectrum of flexible conformations

The structural hallmarks of E2F and DP proteins consist of a DBD connected through a short linker to the CC-MB domains. As the available 3D structures from distinct E2F-DP family members only refer to the separated domains, to have a global view of the DNA-bound full-length heterodimers, we produced recombinant proteins including the full conserved regions of E2F1 and DP1 (DBD-CC-MB, here-after E2F1^{DCM}/DP1^{DCM}) (**Fig. 1A**). Singularly, DP1^{DCM} was insoluble and E2F1^{DCM} soluble, possibly binding DNA as a homodimer [12,15]; proteins were co-expressed and purified by the 6×His-tag of DP1, thus obtaining soluble and homogenous heterodimer (**Fig. S1A**). The purified E2F1^{DCM}/DP1^{DCM} was functionally characterized by EMSA on the canonical E2F site of the human E2F1 promoter. It efficiently binds DNA at the expected nanomolar concentrations, while E2F1^{DCM} alone displayed extremely low DNA-binding activity (**Fig. S1B**).

The purified E2F1^{DCM}/DP1^{DCM} heterodimer was incubated on a target 15-mer DNA oligo (**Table S1**) of the same length as the one used to obtain the E2F4/DP2/DNA crystals [11], and further purified by size exclusion chromatography (SEC). As attempts to obtain protein crystals failed, also considering the potential intrinsic flexibility of the DBD-CC-MB module, we set out to analyze the E2F^{DCM}/DP^{DCM}/DNA complexes by solution-based Small-Angle X-ray Scattering (SAXS). SAXS is an

established method that utilizes 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 provides useful information on flexible regions. To overcome sample polydispersity, which can easily lead to data misinterpretation in SAXS data analysis, we performed SEC in-line with X-ray data collection (SEC-SAXS). Two data sets were collected from E2F1^{DCM}/DP1^{DCM}/DNA with and without the RbC peptide [13] at the Diamond Synchrotron. The data collection and the structural parameters, including the values of the radius of gyration (R_g), the maximum diameter (D_{max}), and the molecular mass of the complexes, are shown in **Table S2**. The R_g and the forward scattering intensity I(0) trace plotted as a function of measurement frame shows that the E2F1^{DCM}/DP1^{DCM}/DNA sample is indeed monodispersed (**Fig. S2A**). Data frames under the elution peak, centered on the maximum in I(0) where the R_g had plateaued, were selected and averaged for subsequent analysis. **Fig. 2A** shows the log I(q) vs. scattering angle (q) plot which provides a clear representation of the buffer-subtracted averaged data over the entire q range, with the Guinier plot (ln(I(q)) vs. (q^2)) [16] shown as an inset. The Guinier analysis shows linearity in the small q region (qR$_g$ < 1.3), suggesting that the sample, within the selected frames, is aggregation free.

To qualitatively evaluate the protein compactness and the presence of disordered regions, Kratky plot analysis was employed [17]. The dimensionless Kratky plot [(qRg)^{2}I(q)/I(0) vs. qR$_g$], where q is normalized to the R_g of the protein, and the intensity I(q) normalized to the I(0) is most useful to discriminate between different degrees of folding. Typically, for a compact and well-folded single-domain protein, the normalized Kratky plot displays a bell-shaped curve, with a well-defined maximum of ~ 1.1 at qR$_g$ around 1.75 [18]: The generated plot obtained from the E2F1^{DCM}/DP1^{DCM}/DNA SEC-SAXS data (**Fig. 2B**) has instead a distinct profile, where the maximum is shifted upwards and right, indicating an elongated dumbbell-shaped structure of the complex. The shallow oscillation at 2.5 < qRg < 4 reflects the well-folded two-domain structure connected by a flexible linker [19,20]. At large q values, the intensity rises increasingly indicating significant flexibility. The P(r) distance distribution function vs. r profile (**Fig. S2B**) is well behaved, showing a peak at low r values, corresponding to the intradomain distances, and a tail with pronounced shoulder corresponding to the interdomain distances, expected for a multidomain protein with flexible linker [21]. E2F1^{DCM}/DP1^{DCM}/RbC/DNA SEC-SAXS data analysis provides similar, essentially superimposable, results (**Table S2** and **Fig. S3**) on the overall shape, folding and flexibility of the protein/DNA complex.
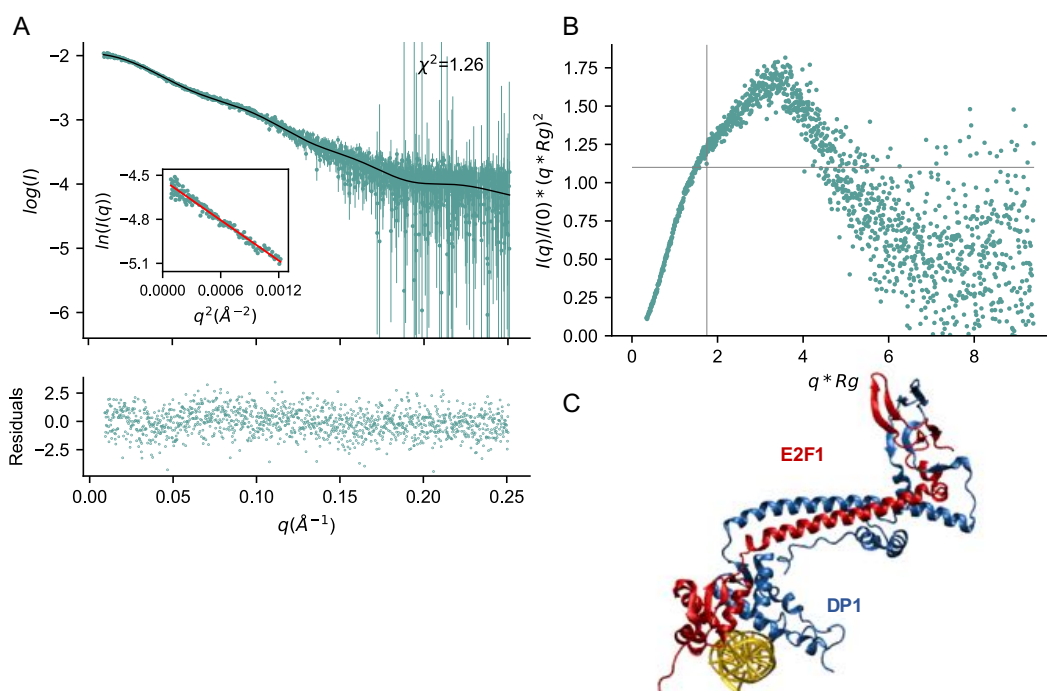
Figure 2. Experimental SEC-SAXS profiles and atomistic model for E2F1[DCM]/DP1[DCM]/DNA complex. (A) I(q) vs. q as semi-logarithmic representation, with Guinier fit (red line) for $qR_g < 1.3$ is shown as inset. The black line represents the model SAXS profile, back-calculated using CRYSOL, displaying a $\chi^2 = 1.26$ in the range 0.009-0.25 Å^{-1}. In the lower panel residuals are reported. **(B)** Normalized Kratky plot [$(qR_g)^2 I(q)/I(0)$ vs. qR_g] for the data in (A). **(C)** Representation of the E2F1[DCM]/DP1[DCM]/DNA model refined via single-replica SAXS-driven MD.

Building a structural model of the E2F1/DP1/DNA assembly

To build a structural model of the E2F1[DCM]/DP1[DCM]/DNA complex, we resorted to homology modelling followed by SAXS-driven Molecular Dynamics (MD) simulations. Our procedure consisted of five steps (**Fig. S4**): 1) building a tentative initial model with homology modelling; 2) minimization and equilibration with MD; 3) exhaustive exploration of the conformational space with Metadynamics [22]; 4) selection of a starting model via comparison with experimental SAXS data; 5) refinement of the model with SAXS-driven MD simulations.

The homology model was based on available PDBs: 1CF7 [11] contains the DBDs of E2F4/DP2 in complex with DNA, while 2AZE [13] is the complex of the CC-MB domains of E2F1/DP1 with the RbC peptide (**Fig. S5**). As the relative orientation of the CC-MB domains with respect to the DBDs couldn't be deduced from the starting PDBs, we guessed a possible initial conformation of the complex and then we explored a wide variety of possible CC-MB-DBDs orientations by Metadynamics [22], a technique that allows an exhaustive exploration of the conformational space by enhancing the sampling along specified collective variables. The wide ensemble of conformations collected with Metadynamics was analyzed in terms of agreement with the experimental SAXS data using CRYSOL [23], allowing to identify and select a preliminary model with a reasonable agreement with experimental data ($\chi^2 < 2$, see Materials and Methods for details).

The preliminary E2F1[DCM]/DP1[DCM]/DNA model was used as a starting point for further refinement via 120 ns long SAXS-driven MD simulation. Eventually, a reference model was extracted from this

simulation, showing both good agreement with SAXS data (χ^2 =1.26 in the range 0.009-0.25 Å^{-1}, see **Fig. 2**) and good quality according to Molprobity [24] (Molprobity-score of 1.73, clash-score of 1.16, 1.2 % Ramachandran outliers). The model (**Fig. 2C**) displays a gyration radius of 37.7 Å (well compatible with the one determined from SAXS experiments, see **Table S2**) and suggests that there is a preferential orientation of the CC-MB domains with respect to the DBDs, with the coiled-coil inclined towards the part of DNA occupied by DP1.

## E2F1^{DCM}/DP1^{DCM}/DNA dynamics

SAXS data as well as preliminary simulations indicated that the E2F1^{DCM}/DP1^{DCM}/DNA complex is dynamic, and, therefore, the system should be better described by a conformational ensemble rather than by a single model. We performed multi-replica metainference MD simulation [25] in which the average SAXS intensities, computed over the replicas, are forced to fit with the experimental data within an accuracy determined on-the-fly. The resulting conformational ensemble shows good agreement with experiments (χ^2 =1.23 in the range 0.009-0.25 Å^{-1}) and highlights a significant inter-domain flexibility, with the CC domains adopting a continuum of possible orientations with respect to the DBDs (**Fig. 3** and **Fig. S6**). Despite this large conformational freedom, a preferential orientation of the CC domains, pointing toward the DNA side occupied by DP1, could be identified (**Fig. 3**), in line with the previous observation on the SAXS-refined single model. The described inter-domain movement is due to the flexibility of the linker regions between the DBDs and the CC domains, that include E2F1 residues 192-201 (GSHTTVGVGG) and DP1 residues 193-198 (GLPTNS), both showing a poor propensity to stabilize secondary structure according to our simulation (**Fig. S7A**). The presence of a preferential orientation is likely the result of the different length and amino-acidic composition of the two linkers, with the E2F1 linker rich in Glycine.
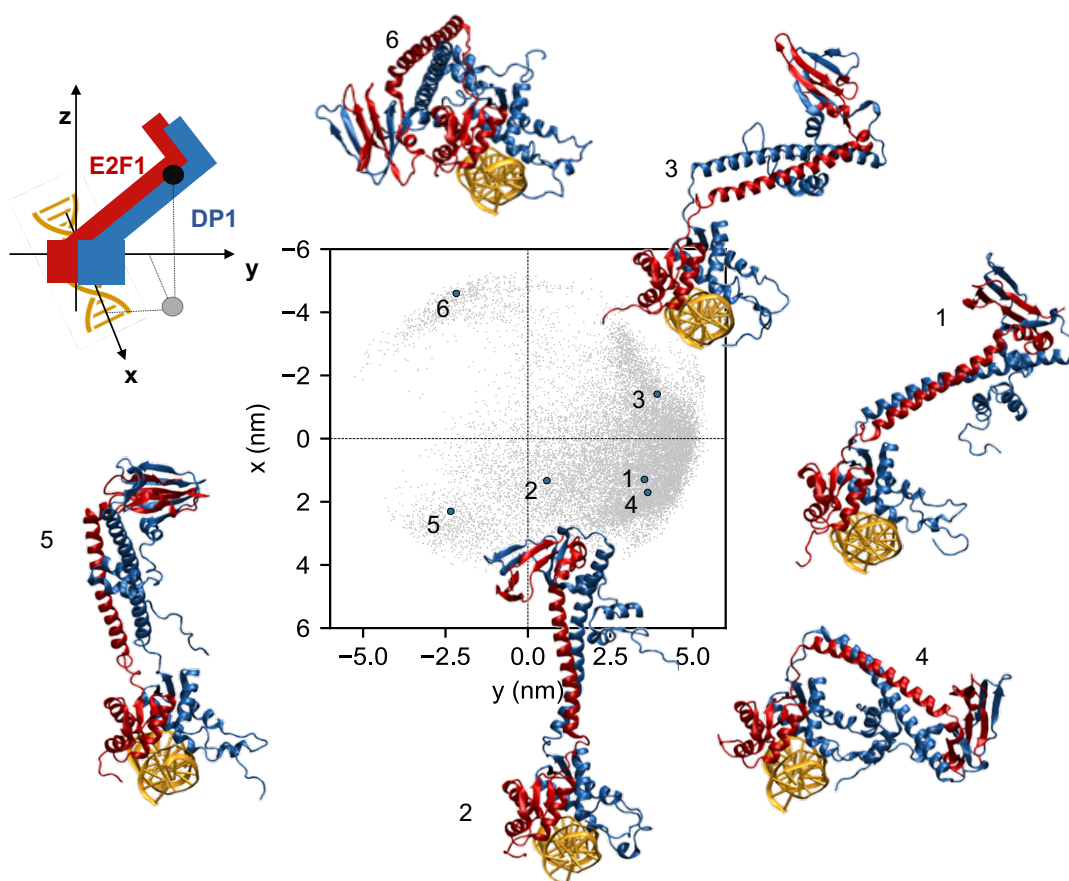
Figure 3. Schematic representation of the orientation adopted by the CC domain with respect to the DNA/DBDs. The DNA axis is aligned to the x axis, while the z axis was aligned to the axis connecting the DNA and DBDs geometric centers. In the plot are represented the x-y projections of the positions adopted by the last turns of the coiled-coil domain geometric center (black sphere in the model), once fixed the DNA and DBDs: here the observer is looking from the top of the z axis. Few representative structures, selected as the centers of the main six clusters, based on backbone-RMSD, are illustrated (in this case the observer is looking from the top of the x axis).

In addition to the inter-domain movement, few regions within the DBDs display poor secondary structure content and high flexibility (**Figs. 4, S7A**, and **Movie 1**). These regions include the N-terminal tails (**Fig. 4**) of both E2F1 (residues 120-127, MSPGEKSR) and DP1 (residues 105-113, RNRKGEKNG, plus four residues derived from thrombin cleavage): herein, we noted that the long flexible DP1 N-terminal tail contains a high number of basic amino-acids, which could potentially favor binding to longer DNA constructs, aiding DNA recognition and binding outside the usual RRXYD motif. The DP1 loop between helices α2 and α3, comprising residues 148-161 (AADNHILPNESAYD) was also identified as a highly dynamic region according to Root Mean Square Fluctuation and secondary-structure content analysis (**Figs. 4**, **S7A**, and **Movie 1**).
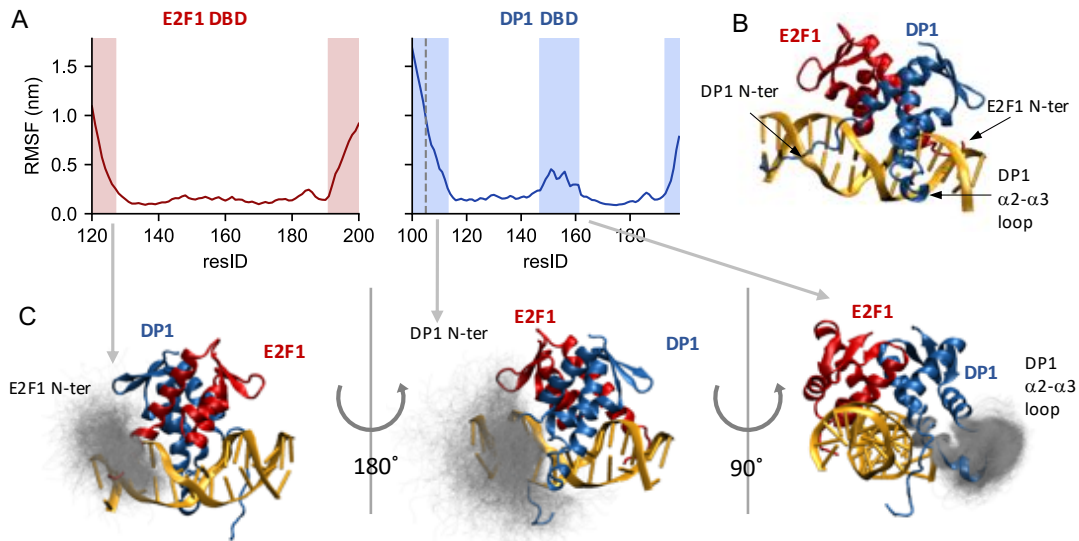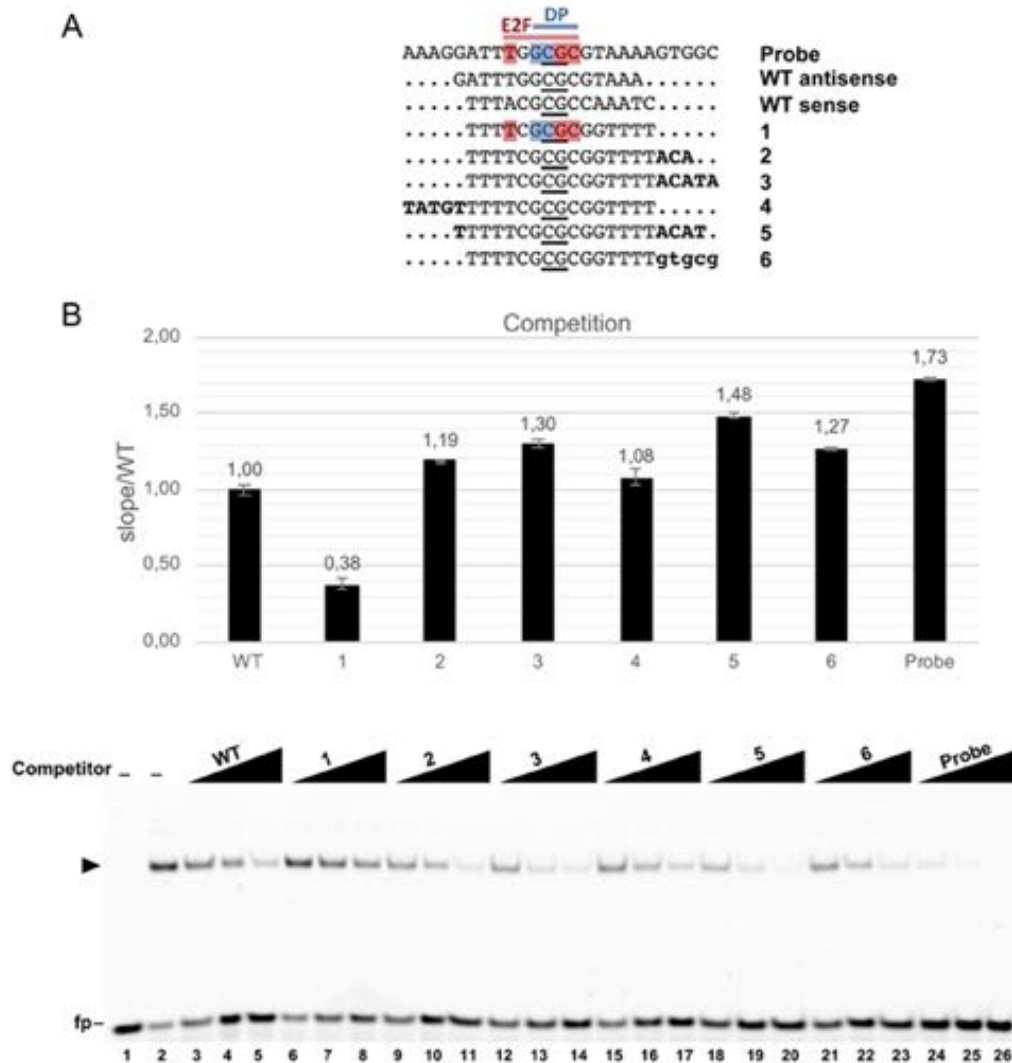
Figure 4. E2F1/DP1 dynamics. (A) Per-residue Root mean square fluctuations (RMSF) of the E2F1 and DP1 DBDs. Regions showing RMSF greater than 0.25 nm, are highlighted with shades. **(B)** Cartoon representation of a DNA/DBDs structure, where the position of the identified flexible regions is indicated. **(C)** Highlights of the flexible DBDs regions, including: E2F1 N-terminal (left); DP1 N-terminal (center); DP1 $\alpha2$-$\alpha3$ loop, comprising residues 148-161. For each cartoon representation the dynamics regions are highlighted with transparent grey tubes, representing all the possible conformations of the simulated ensemble.

DNA contacts on the DP1 side of the E2F motif improved affinity

The data provided above prompted us to further analyze DNA requirements and to define whether the extension of the DNA oligo on the DP1 area may support the predictions of MD simulations. To this aim, we used a 25-mer labelled probe to analyze the affinity of E2F1[DCM]/DP1[DCM] for DNA oligos of different length in EMSA competition assays. Given the symmetry of the E2F1 promoter E2F core element, to discriminate between the E2F1 and DP1 contributions, we started with the defined non-symmetrical E2F DNA element of 15 bp (oligo 1; **Fig. 5A** and **Table S1**) used for the E2F4/DP2/DNA crystals [11]; (PDB entry 1CF7). Five bp extensions with a random sequence on either side of the oligo were used as competitors to challenge binding (**Fig. 5A**). The competition rate of the 15-mer WT oligo derived from the probe was set as 1. The unlabeled 25-mer oligo, used as a positive control, showed higher competition efficiency, indicating that regions outside the 15-mer oligo indeed contribute to stabilize binding to DNA. With oligo 1 (15-mer) a substantial decrease in competition rates was observed (**Fig. 5B**). However, both (20-mer) 5bp-extended versions of this oligo, on either DP1 or E2F1 sides (oligos 3 and 4, respectively) compete well, showing an increased competition rate as compared to the WT oligo, which was more pronounced using oligo 3, on the DP1 side. To discriminate the contribution of E2F1 and DP1 to the DNA affinity, we designed optimized competitors where one and four bp were added at the E2F1 and DP1 sides, respectively: enhanced affinity was observed compared to the previous extensions, indicating an impact of nucleotides flanking the core DNA element, with an asymmetric effect, stronger for the sequence flanking the DP1 side (oligo 5 vs oligo 4).

Next, we started to evaluate a possible sequence-selectivity. First, a 20-mer oligo in which transitions from A/T to G/C were introduced to the five additional bases on the DP1 side competed similarly to oligo 3 (1.27 and 1.30 competition rates, respectively). To further substantiate this point, we used a set of unlabeled oligos in which the three nucleotides at the 3' flanking of the core motif were mutated (oligo A-D; **Fig. S8**), based on the 18-mer oligo (oligo 2) that shows high competition rates: all mutations show similar competition rates, indicating no sequence preference (**Fig. S8**).

Figure 5. *In vitro* **validation of DP1 stabilization by competition EMSA experiments. (A)** Sequence details and the alignment of the probe and the designed unlabeled oligonucleotides - centered 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 E2F1 and DP1 proteins [11], and corresponding positions on the E2F1 probe, are highlighted in red and blue, respectively. Mutant nucleotides are indicated in lowercase. **(B)** Direct competition EMSAs quantification. The indicated unlabeled competitors were incubated at increasing concentrations (5x, 25x, 100x fold excess) with the E2F1-25 bp probe (20 nM), with the addition of recombinant E2F1^{DCM}/DP1^{DCM} 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 (15-mer) oligonucleotide competition slope. Representative gel of direct competition EMSA experiment is illustrated in the lower panel. As a positive control, the wild-type 25-mer E2F1 unlabeled oligo (Probe) was used as a 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.

DISCUSSION

Sequence-specific TFs are proteins with minimally two essential parts: a DNA-binding domain (DBD) and a transcriptional activation domain (TAD). Most of them feature additional parts, which regulate their activities. In general, DBDs are well structured and indeed many of them have been detailed in complex with DNA by 3D studies over the last 30 years; other parts, including TADs, often characterized by a large content of disordered regions [26], remain less understood. Even less is known about how different domains of TFs connect to their DBDs. The E2F/DP family is paradigmatic in this regard: although crystal structures of the heterodimeric DBD and CC-MB domains are available, the relative connections and overall picture of the large, central portions of the heterodimers are missing. We used our recently developed integrative modelling approach, in which molecular dynamics is driven by SEC-SAXS data and coupled to homology modelling, to characterize the structural dynamics of a E2F1/DP1/DNA complex, composed by the three structured domains -DBD-CC-MB- connected by disordered linkers (**Fig. 1**). The long, presumably disordered C- and the N-terminal regions of the proteins were not included. The TF/DNA complex results in a very flexible system and relevant results are reported: (i) the CC domains are preferentially oriented toward the DNA side occupied by DP1 (**Fig. 3**); (ii) additional dynamics are observed within the $\alpha2$-$\alpha3$ loop -residues 148-161- of the DP1 DBD (**Fig. 4**); (iii) the N-terminal regions of the DBDs are very flexible (**Fig. 4**), with DP1 showing a contribution to stabilization of DNA-binding. Intriguingly, these features highlight novel properties of DP1.

The CC and MB domains are mobile with respect to the DBD heterodimer in complex with DNA and the flexible linkers are responsible for this. Remarkably, the movement of some 80° is not circular, but lateral and flexing along the DNA axis, with a precise directionality: specifically, the CC leans toward the DP1 side of the complex (**Fig. 3**). Sequence alignments of E2F and DP proteins show that the linker region is short and highly conserved in DP proteins, spanning 7aa with 100% identity: therefore, DPs appear to limit the flexibility, and all three members (DP1-3) are predicted to do it in the same way and to the same extent. Instead, canonical E2Fs -E2F1-6- have linkers whose length and amino acids composition are not conserved, ranging from the Gly-rich 11 aa of E2F1 to 15 aa of E2F4/5. Such variability points to potential differences in the degree of flexibility of the DNA-bound heterodimers, dictated by the specific E2F partner. As for interactions with pocket proteins, SAXS data with and without the RbC peptide are essentially superimposable, meaning that it has marginal influence on DNA-binding or on the flexibility of the full CC-MB portion of the complex.

The second dynamic area is the DP1 loop between $\alpha2$ and $\alpha3$ of the DBD: according to the simulations, this stretch, partially populating a helical structure, is considered a highly dynamic region,

displaying poor secondary structure content (**Figs. 4 and S7A**). This is in striking contrast to the corresponding area of E2F1: as shown in the E2F4/DP2/DNA crystals, in fact, this area is robustly structured on the E2F4 side, neither do we detect a flexible area within E2F1 in our analysis. The E2F1 loop is much shorter and sequence alignments indicate high sequence conservation among E2F1-6 (**Fig. S7B**): thus, the prediction would be of little flexibility in all E2F members in this area. In contrast, the corresponding DP1 region shows a relatively low degree of sequence similarity to other DP proteins, particularly DP2, with only 6 conserved residues over 14 (43%), whereas our DP1 construct has an overall identity with DP2 of 79% (**Fig. S7B**). In the E2F4/DP2/DNA structure, the DP2 loop is in a stretched configuration, but the presence of crystallographic contacts prevents us to gather conclusions about its status. Apparently, despite its flexibility, the DP1 loop is close to, but not in contact with, DNA: we are tempted to speculate that this adaptable area might be a flexible -and asymmetric- docking spot for protein-protein interactions with other TFs binding close to the E2F motif, potentially specific for the single DP proteins.

The third highly dynamic area is the N-terminal region of DP1 -aa 105-RNRKGEKNG-113- containing numerous basic residues, which could make additional contacts to DNA, thus increasing overall DNA-binding affinity (**Fig. 4**). This was experimentally verified by EMSAs (**Fig. 5**): DNA oligos extended of 3 bps on the DP1 side stabilize DNA-binding. Note that the corresponding area in E2F1 is also flexible, although to a less extent, and that the addition of nucleotides on this side had less effect on DNA-binding stabilization. The DNA area of the extra DP1 contacts lies outside of the conventional motif contacted in a sequence-specific way by the DBDs. We thus further investigated whether the binding of DP1 to this extended area is sequence-specific: no significant preference was observed. It was important to establish this, since the exact limit of the core DNA region contacted by E2Fs/DPs has been a matter of some debate. In fact, the matrix derived from in vivo ChIP-seq data [27–29] appears to be a symmetric portion -CGCGC- of the larger, asymmetric TTTSSCGC matrix originally derived by extended mutagenesis and SELEX analysis with recombinant proteins *in vitro* [30] and subsequent alignment of target sites [28,31].

In general, despite the apparent symmetry emerging from the available E2F4/DP2/DNA partial structures, in our complete SAXS structure of the E2F1^{DCM}/DP1^{DCM}/DNA complex we identified different levels of asymmetry, at the level of DNA-binding -extra domain of DP1- of the $\alpha2$-$\alpha3$ loop, and on the mobility of the CC-MB domains, toward the DP1 side.

DNA-binding sites are known to be highly "packed" in regulatory regions, and it is intuitive how the E2F1/DP1 flexible properties might guide/adapt specific contacts with TFs neighboring E2F sites. Illustrative of this are data on the interactions between E2F1/DP1 and the bZip C/EBPα: DP1 interferes with DNA-binding and transactivation of C/EBP/E2F co-targeted genes upon direct protein-protein binding, independently from repressive pocket proteins interactions. Specifically, the DP1 residues 105-127 were implicated in the interactions [32]. The Authors showed that a C/EBPα bearing a mutation in the basic region (BRM2) has an increased affinity for DP1, but a reduced affinity for DNA; importantly, DNA-binding -and transactivation- of this mutant on co-targeted promoters was restored upon knockdown of DP1, and E2F1, establishing a functional link between the basic region of C/EBPα and the N-terminal flexible region of DP1. Interestingly, BRM2 resembles disease-causing

mutations frequently found in acute myeloid leukemia (AML) patients [32]. As we provide data suggesting that the DP1 N-terminal residues are involved in DNA-binding stabilization, the two DBDs might physiologically compete for DNA-binding: the less basic BRM2, or other similar tumor mutants, could "lose" in the competition with the flexible DP1 N-terminal region, letting the pro-growth E2F/DP complex functionally prevail on neighboring, pro-differentiation C/EBP sites.

Finally, our data have potential implications for the interpretation of E2F and DP cancer mutations. The E2F pathway is frequently deregulated in cancer through multiple mechanisms, but mutations in the E2F1 or DP1 genes are not frequently found in tumor cells [10,33]. We mined the Catalogue of Somatic Mutation in Cancer (COSMIC) and Cancer Cell Line Encyclopedia (CCLE) public databases [34,35] and we updated the list of E2F1 and DP1 missense mutations (**Fig. 6**). As shown before, mutations in DP1 are more numerous than those in E2F1, probably reflecting the inability of cells to tolerate loss- or gain-of-function mutants of the latter. Nevertheless, mutations in DP1 could impact on and deregulate E2F activity [36]. The locations of the mutations are throughout DP1 and E2F1, but there is an increased abundance located within, or flanking, the flexible and unstructured regions described here, particularly in the DBDs (**Fig. 6**). This could be because their effect is null, or because they have a specific function -either gain or loss of- on DP1, or on the E2F partner. 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 [37]. To the best of our knowledge, no systematic study was done on E2F1 tumor mutants and there are no data on E2F1 mutations in the flexible linker (T195I, V199A, G201V) nor DBD N-terminal tail (S121F, P122L, E124Q, R127H), although this region was shown to be important for DNA-binding *in vitro* [38]. As for DP1, tumor mutants are found in the flexible regions of the DBD N-terminal end (R105H, K108N, G113D/S, K114R, G115D, R117W), the $\alpha2$-$\alpha3$ loop (A149V, D150N, I153T, P155Q) and the linker (T196A), among which G113S, G115D and D150N were systematically assayed [36] and showed normal E2F1 heterodimerization. As for the flexible areas identified here, G115D, but not D150N, is defective in binding DNA targets *in vivo* (ChIPs), which is consistent with the former, not the latter, being DNA-binding defective. Thus, the reduced activation of reporter genes relying on E2F sites is justified by the phenotype of decreased DNA binding of G113S, and G115D, but not for the $\alpha2$-$\alpha3$ loop D150N mutant. However, loss of transactivation was selective for certain promoters, suggesting mechanisms more subtle than simple on/off DNA-binding proficiency. The presence of pathological mutations in these regions could lead to distorted protein dynamics that affect protein function in a context-dependent way, including the above-mentioned docking spot(s) for binding of neighboring TFs. In addition to paving the road for further analysis of other E2F/DP family members, the strategies employed here could be used to study the relative dynamics of separate domains of other TFs whose structures are partially known.
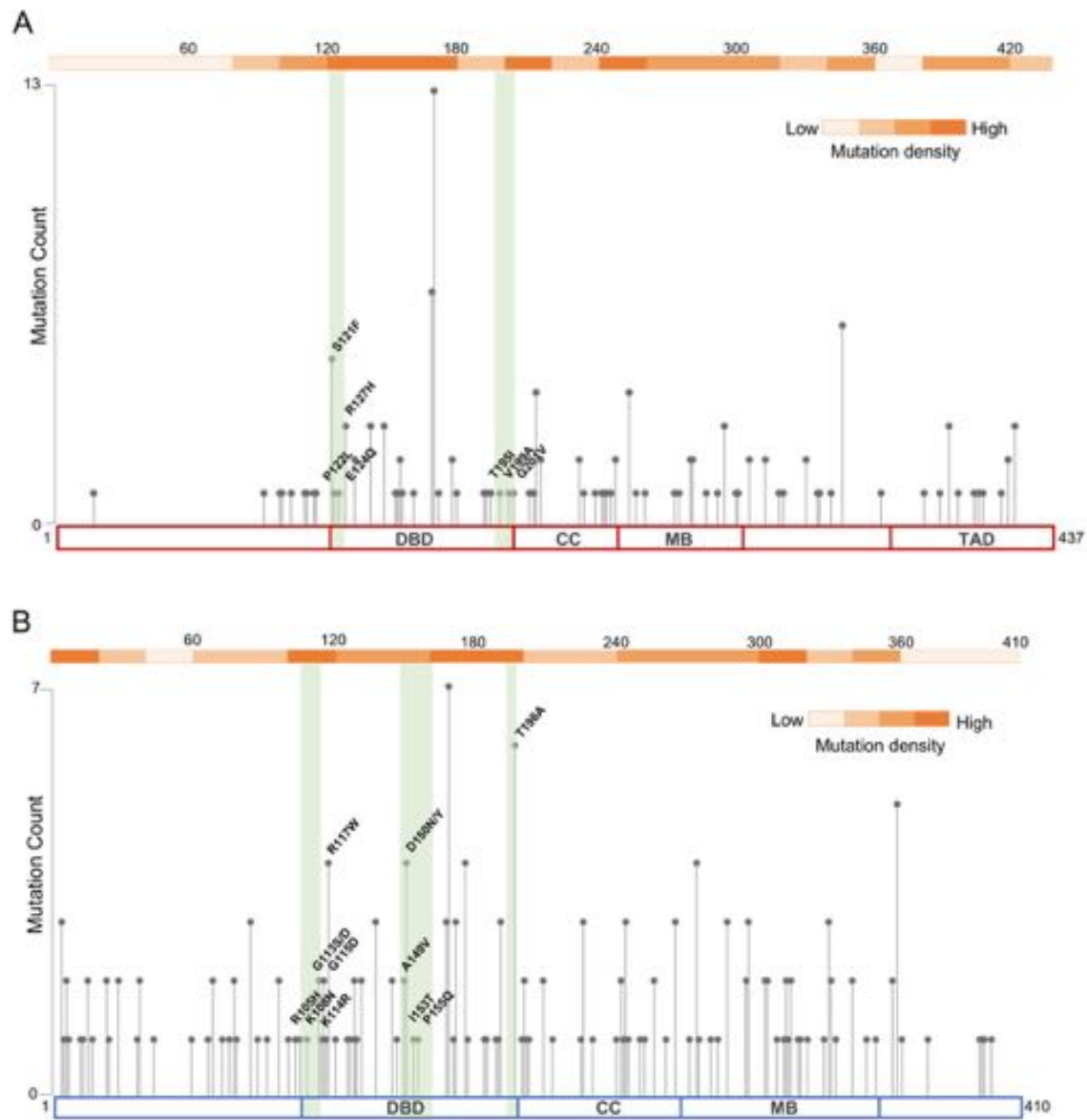
Figure 6. Schematic diagrams showing the somatic mutations in E2F1 and DP1. Data of somatic mutations in E2F1 **(A)** and DP1 **(B)** were retrieved from COSMIC and CCLE database and mapped onto the corresponding domains. The orange shaded windows of 20 residues each (on top of each diagram) highlight the mutation density (total unique samples/ aa) throughout the proteins. Mutations within or flanking the described flexible regions are labelled. The green shaded areas highlight the flexible regions in the DBDs according to the MD simulations. Height of the bars denotes the mutation count according to COSMIC and CCLE.

MATERIALS AND METHODS

Protein co-expression and purification

E2F1 was co-produced with DP1, both containing the DBD, the linker region, the coiled-coil (CC) heterodimerization domain, and the marked-box (MB) domain (DBD-CC-MB; E2F1^{DCM}/DP1^{DCM})(**Fig. 1A**). The cDNA sequence encoding residues Arg105-Thr353 (DP1^{DCM}) of human DP1 was amplified by PCR from pCMV-Neo-Bam-Dp1 (Addgene plasmid # 37968, a kind gift of Jaqueline Lees) [39] and

14

subcloned into the pmcnEATNH vector [40], allowing *E. coli* expression of the fusion protein with a N-terminal 6×His-tag, followed by a thrombin cleavage site. Human E2F1 cDNA (a kind gift of L. Vandel) [41] encoding residues Ser121-Glu301 (E2F1^{DCM}) was amplified by PCR and subcloned into the pmcnYC vector [40]. PCR primers are listed in **Table S1**. The constructs were verified by sequencing and co-expressed in *E. coli* BL21(DE3). Co-transformed cells were grown in LB medium with double antibiotic selection (100 µg mL^{-1} ampicillin and spectinomycin) at 37 °C to an OD600 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 (50 mM Tris-Cl pH 8.0, 400 mM NaCl, 2 mM MgCl$_2$, 10 mM imidazole, 10% glycerol, 2 mM DTT) added of PIC (Protease Inhibitor Cocktail, Sigma-Aldrich), and lysed by sonication. The resulting soluble fraction was recovered by centrifugation at 13,000 x g for 30 min. E2F1^{DCM}/6xHis DP1^{DCM} was purified using nickel ion metal affinity chromatography (IMAC) on an ÄKTA pure system with 1 mL His-Trap FF column (GE Healthcare) and eluted in Buffer A containing 100 mM imidazole. The pooled protein fractions were then incubated overnight at 4 °C with thrombin (Sigma-Aldrich) to cleave the N-terminal 6×His-tag. The cleaved protein was concentrated and further purified by gel filtration chromatography using a HiLoad 10/300 Superdex 200 column (GE Healthcare) equilibrated in Buffer B (50 mM Tris-Cl pH 8.0, 400 mM NaCl, 5% glycerol, 2 mM DTT). The purity of the protein was examined on SDS-PAGE gel stained with Coomassie brilliant blue (**Fig. S1**).

Electrophoretic Mobility Shift Assay

Recombinant E2F1^{DCM}/DP1^{DCM} heterodimer (**Fig. 1A and S1**) was added to a binding reaction mix containing a 25 bp Cy5-labelled probe derived from human E2F1 promoter (E2F1 Probe; 5′-[Cy5]-AAAGGATTTGGCGCGTAAAAGTGGC-3′). The final reaction composition was the following: 20 nM probe, 20 mM Tris-Cl pH 7.5, 20 mM NaCl, 5 mM MgCl$_2$, 0.5 mM EDTA, 5% glycerol, 0.2 mg mL^{-1} BSA and 2.5 mM DTT. The reactions were incubated at 30 °C for 30 min in the dark, loaded onto a 6% native polyacrylamide gel in 0.25x TBE, and run at 100 V for electrophoresis. A Chemidoc MP imaging system (Bio-Rad, Hercules, CA, USA) coupled with the ImageLab software (Bio-Rad) was used in the acquisition and analysis of fluorescence signals. For competition assays, increasing molar excess (5x, 25x, 100x fold) of unlabeled competitor oligos were added to the mix, containing 15 nM protein. The oligonucleotides used in competition assays are listed in **Table S1**. In each dose curve data point, the bound DNA was quantified and plotted in function of the molar excess of the competitor (fold competitor, expressed as the ratio of the competitor vs. total oligo concentration). For each experiment, the regression line slope obtained through the 0, 5x, 25x, 100x competition data points vs. the WT oligo slope represents the competitor efficiency (slope/WT). The shown data represent the means ± standard error of at least three independent competition experiments.

Sample preparation for SEC-SAXS analysis

The 15-mer DNA oligo of the human E2F1 promoter (5′-GATTTGGCGCGTAAA-3′) used for the assembly of protein/DNA complexes was obtained from Eurofins Genomics as single-stranded oligos

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 at a molar ratio of 2:1 and subjected to step dialysis to gradually decrease salt concentration to 150 mM NaCl, thus allowing DNA binding. The ternary complex was isolated using Superdex 200 gel filtration chromatography and concentrated to 4 mg mL^{-1} in Buffer C (50 mM Tris-Cl pH 8.0, 150 mM NaCl and 2 mM DTT). The RbC peptide (residues 829-874) [13] 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[DCM]/DP1[DCM]/DNA complex at a molar ratio of 2:1 and subjected to SEC-SAXS analysis.

Online SEC-SAXS Measurements and Data Processing

X-ray scattering patterns of E2F1[DCM]/DP1[DCM] complexes were recorded using the online 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 S2** for details). For each sample, 45 μL of a 4 mg mL^{-1} was injected onto a pre-equilibrated column in Buffer C, and roughly 600 sequential images were recorded at a flow rate of 0.075 mL min^{-1}. Primary data reduction was performed using ScÅtter [42] 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). Using the same software, the two-dimensional scattering images were radially processed, the buffer scattering intensities were subtracted and the R_g, D_{max}, and I(0) were computed. The molecular weight was estimated using SAXSMoW program available at http://saxs.ifsc.usp.br/ [43].

Building a structural model of the E2F1[DCM]/DP1[DCM] heterodimer.

A tentative initial model of the E2F1[DCM]/DP1[DCM]/DNA construct was built starting from the available PDB structures of homologous proteins: 1CF7 [11], containing the DBDs of E2F4/DP2, and 2AZE [13], comprising the CC and MB domains of E2F1/DP1 in complex with the Rb peptide. The model was built using Modeller [44], adding the missing residues at the termini and in the linker-regions connecting the DBDs to the CC domains. The 15 base-pairs DNA was built starting from the one in PDB 1CF7 replacing the different nucleotides using Coot [45].

The generated model of the DNA-bound E2F1[DCM]/DP1[DCM] heterodimer was minimized and equilibrated with GROMACS 2018 [46] and the PLUMED2 software [47], using amber99sb-disp force field [48], combined with Parmbsc1 for DNA [49], and TIP4P-D water [50]. The system was solvated in a dodecahedron box, initially 1.6 nm larger than the complex in each direction, neutralized with Na and Cl ions to reach a salt concentration of 10 mM, then minimized and equilibrated at the temperature of 298 K and pressure of 1 atm using the Berendsen thermostat and barostat [51]. Production simulations were run in the NVT ensemble maintaining the temperature at 298 K with the Bussi thermostat [52]; the time step was of 2 fs, bonds involving hydrogen atoms were constrained with the LINCS algorithm [53], the electrostatic was treated by using the particle mesh Ewald scheme [54] with a short-range cut-off of 0.9 nm and van der Waals interaction cut-off was set to 0.9 nm. In all the simulations, harmonic distance restraints were applied with PLUMED to maintain the hydrogen

bonds between the first/last DNA base-pairs and between the Arginine side chains of E2F1 (R165-R166) and DP1 (R167-R168), belonging to the DNA-recognition motif, with DNA nucleotides G9', G8, G6, and G7', respectively. To obtain an exhaustive exploration of the conformational space, particularly focusing on the relative orientation of the CC domains and DBDs, we performed a long multi-replica Metadynamics [22,55] simulation (see Supplementary Materials for more details). The conformations explored were analyzed in terms of agreement with the experimental SAXS data using CRYSOL [23], to identify a representative conformation to be used as a starting model for further refinement. The selection was performed saving a snapshot for each ns of simulation (total of ~1600 conformations) and calculating the χ^2 with respect to experimental data (in the range 0-0.3 Å^{-1}) using CRYSOL [23]. The conformations were then clustered based on protein-backbone plus DNA RMSD with a cutoff of 9 Å, and the average χ^2 was computed for all the clusters containing at least 30 conformations: eventually, a representative structure was extracted from the most populated clusters with a $<\chi^2>$ lower than 2.

SAXS-driven MD simulations of the E2F1^{DCM}/DP1^{DCM}/DNA

The selected conformation of DNA-bound E2F1^{DCM}/DP1^{DCM} heterodimer was used as the starting model for a new iteration of minimization and equilibration, followed by SAXS-restrained MD simulations. We followed two different protocols for SAXS-driven simulations, both based on metainference [25,56], a technique to integrate ensemble-averaged experimental data in MD simulations exploiting Bayesian inference to deal with both random and systematic errors. As the calculation of SAXS intensities is computationally expensive, we made use of the hybrid approach described in [57] to integrate SAXS restraints, using PLUMED: herein SAXS intensities are computed via a coarse-grained method based on the Martini force field, while the simulations are run with full atomistic details. In the first simulation, we used a single replica, i.e. we forced a single structure to fit the experimental restraints: this simulation was evolved for 120 ns and the model with a lower metainference score was extracted as reference. In the second simulation, we adopted a multi-replica approach (as in [56]), computing the SAXS intensities for each replica and forcing the average over the replica to fit the experimental restraints, thus allowing us to take into account molecular flexibility. Here 20 replicas were used, each one evolved for 400 ns, resulting in a total simulation time of 8 μs. In both cases a set of 12 representative SAXS intensities at different scattering angles, ranging between 0.009 Å^{-1} and 0.20 Å^{-1} and equally spaced, was used as restraints. These representative intensities were extracted from experimental data, after regularization was performed with the Distance Distribution tool of Primus, based on Gnom [58,59]. Metainference was applied every 6 steps, using a single Gaussian noise per data point and sampling a scaling factor between experimental and calculated SAXS intensities with a flat prior between 0.7 and 1.3. For these SAXS-driven simulations, the hydrogen mass repartitioning scheme [60] was used to reduce the computational cost: the mass of heavy atoms was repartitioned into the bonded hydrogen atoms using the *heavyh* flag in the pdb2gmx tool, the LINCS algorithm [53] was used to constraint all bonds, eventually allowing to use a time step of 5 fs.

RESEARCH DATA

The input files used to generate the structural models, as well as the bundle of structures shown in **Fig. 3**, are deposited in PLUMED-NEST [61] as plumID:20.029. The SAXS data and the single model shown in **Fig. 2** are deposited in the SASBDB [62] as SASDKH2.

ACCESSION NUMBERS

RefSeq protein IDs: NP_009042 (transcription factor Dp-1 [Homo sapiens]); NP_005216 (transcription factor E2F1 [Homo sapiens]). PDB ID: 2AZE_C (Rb C-terminal domain); PDB ID: 1CF7; PDB ID: 2AZE

SUPPLEMENTAL DATA

Supplemental Data are available online.

- Supplemental Data PDF file

- Supplemental Movie file: Movie 1. Video displaying the dynamics of the DNA-binding domain.

ACKNOWLEDGEMENTS

FUNDING

CONFLICT OF INTEREST

Authors declare no conflicts of interest

REFERENCES

[1]   S.A. Lambert, A. Jolma, L.F. Campitelli, P.K. Das, Y. Yin, M. Albu, X. Chen, J. Taipale, T.R. Hughes, M.T. Weirauch, The Human Transcription Factors, Cell. 172 (2018) 650–665.

https://doi.org/10.1016/j.cell.2018.01.029.

[2]   J. DeGregori, D. Johnson, Distinct and Overlapping Roles for E2F Family Members in Transcription, Proliferation and Apoptosis, Curr. Mol. Med. 6 (2006) 739–748. https://doi.org/10.2174/1566524010606070739.

[3]   T. Lammens, J. Li, G. Leone, L. De Veylder, Atypical E2Fs: new players in the E2F transcription factor family, Trends Cell Biol. 19 (2009) 111–118. https://doi.org/10.1016/j.tcb.2009.01.002.

[4]   W. Zhu, P.H. Giangrande, J.R. Nevins, E2Fs link the control of G1/S and G2/M transcription, EMBO J. 23 (2004) 4615–4626. https://doi.org/10.1038/sj.emboj.7600459.

[5]   S. Sadasivam, J.A. DeCaprio, The DREAM complex: master coordinator of cell cycle-dependent gene expression, Nat. Rev. Cancer. 13 (2013) 585–595. https://doi.org/10.1038/nrc3556.

[6]   J. Hsu, J. Arand, A. Chaikovsky, N.A. Mooney, J. Demeter, C.M. Brison, R. Oliverio, H. Vogel, S.M. Rubin, P.K. Jackson, J. Sage, E2F4 regulates transcriptional activation in mouse embryonic stem cells independently of the RB family, Nat. Commun. 10 (2019) 2939. https://doi.org/10.1038/s41467-019-10901-x.

[7]   A. Poppy Roworth, F. Ghari, N.B. La Thangue, To live or let die – complexity within the E2F1 pathway, Mol. Cell. Oncol. 2 (2015) e970480. https://doi.org/10.4161/23723548.2014.970480.

[8]   J.-L. Chong, P.L. Wenzel, M.T. Sáenz-Robles, V. Nair, A. Ferrey, J.P. Hagan, Y.M. Gomez, N. Sharma, H.-Z. Chen, M. Ouseph, S.-H. Wang, P. Trikha, B. Culp, L. Mezache, D.J. Winton, O.J. Sansom, D. Chen, R. Bremner, P.G. Cantalupo, M.L. Robinson, J.M. Pipas, G. Leone, E2f1–3 switch from activators in progenitor cells to repressors in differentiating cells, Nature. 462 (2009) 930–934. https://doi.org/10.1038/nature08677.

[9]   T.C. Hallstrom, J.R. Nevins, Balancing the decision of cell proliferation and cell fate, Cell Cycle. 8 (2009) 532–535. https://doi.org/10.4161/cc.8.4.7609.

[10]   H.-Z. Chen, S.-Y. Tsai, G. Leone, Emerging roles of E2Fs in cancer: an exit from cell cycle control, Nat. Rev. Cancer. 9 (2009) 785–797. https://doi.org/10.1038/nrc2696.

[11]   N. Zheng, E. Fraenkel, C.O. Pabo, N.P. Pavletich, Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP, Genes Dev. 13 (1999) 666–674. https://doi.org/10.1101/gad.13.6.666.

[12]   E. Morgunova, Y. Yin, A. Jolma, K. Dave, B. Schmierer, A. Popov, N. Eremina, L. Nilsson, J. Taipale, Structural insights into the DNA-binding specificity of E2F family transcription factors, Nat. Commun. 6 (2015) 10050. https://doi.org/10.1038/ncomms10050.

[13]   S.M. Rubin, A.-L. Gall, N. Zheng, N.P. Pavletich, Structure of the Rb C-Terminal Domain Bound to E2F1-DP1: A Mechanism for Phosphorylation-Induced E2F Release, Cell. 123 (2005) 1093–1106. https://doi.org/10.1016/j.cell.2005.09.044.

[14]   T.J. Liban, E.M. Medina, S. Tripathi, S. Sengupta, R.W. Henry, N.E. Buchler, S.M. Rubin, Conservation and divergence of C-terminal domain structure in the retinoblastoma protein family, Proc. Natl. Acad. Sci. 114 (2017) 4942–4947. https://doi.org/10.1073/pnas.1619170114.

[15]   H.E. Huber, G. Edwards, P.J. Goodhart, D.R. Patrick, P.S. Huang, M. Ivey-Hoyle, S.F. Barnett, A. Oliff, D.C. Heimbrook, Transcription factor E2F binds DNA as a heterodimer., Proc. Natl. Acad. Sci. 90 (1993) 3525–3529. https://doi.org/10.1073/pnas.90.8.3525.

[16]   A. Guinier, Diffraction of x-rays of very small angles-application to the study of ultramicroscopic phenomenon, Ann. Phys. (Paris). (1939).

[17]   O. Glatter, O. Kratky, General theory X- Ray scattering, Small-Angle X-Ray Scatt. (1982). https://doi.org/10.1002/actp.1985.010360520.

[18]   D. Durand, C. Vivès, 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. 169 (2010) 45–53. https://doi.org/10.1016/j.jsb.2009.08.009.

[19]   P. Bernadó, D.I. Svergun, Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering, Mol. BioSyst. 8 (2012) 151–167. https://doi.org/10.1039/C1MB05275F.

[20]   J. Trewhella, A.P. Duff, D. Durand, F. Gabel, J.M. Guss, W.A. Hendrickson, G.L. Hura, D.A. Jacques, N.M. Kirby, A.H. Kwan, J. Pérez, L. Pollack, T.M. Ryan, A. Sali, D. Schneidman-Duhovny, T. Schwede, D.I. Svergun, M. Sugiyama, J.A. Tainer, P. Vachette, J. Westbrook, A.E. Whitten, 2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution: an update, Acta Crystallogr. Sect. D Struct. Biol. 73 (2017) 710–728. https://doi.org/10.1107/S2059798317011597.

[21]   V. Receveur-Brechot, D. Durand, How Random are Intrinsically Disordered Proteins? A Small Angle Scattering Perspective, Curr. Protein Pept. Sci. (2012). https://doi.org/10.2174/138920312799277901.

[22]   A. Laio, M. Parrinello, Escaping free-energy minima, Proc. Natl. Acad. Sci. 99 (2002) 12562–12566. https://doi.org/10.1073/pnas.202427399.

[23]   D. Svergun, C. Barberato, M.H.J. Koch, CRYSOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates, J. Appl. Crystallogr. 28 (1995) 768–773. https://doi.org/10.1107/S0021889895007047.

[24]   I.W. Davis, A. Leaver-Fay, V.B. Chen, J.N. Block, G.J. Kapral, X. Wang, L.W. Murray, W.B. Arendall, J. Snoeyink, J.S. Richardson, D.C. Richardson, MolProbity: all-atom contacts and structure validation for proteins and nucleic acids, Nucleic Acids Res. 35 (2007) W375–W383. https://doi.org/10.1093/nar/gkm216.

[25]   M. Bonomi, C. Camilloni, A. Cavalli, M. Vendruscolo, Metainference: A Bayesian inference method for heterogeneous systems, Sci. Adv. 2 (2016) e1501177. https://doi.org/10.1126/sciadv.1501177.

[26]   L. Staby, C. O'Shea, M. Willemoës, F. Theisen, B.B. Kragelund, K. Skriver, Eukaryotic transcription factors: paradigms of protein intrinsic disorder, Biochem. J. 474 (2017) 2509–2532. https://doi.org/10.1042/BCJ20160631.

[27]   A. Rabinovich, V.X. Jin, R. Rabinovich, X. Xu, P.J. Farnham, E2F in vivo binding specificity: Comparison of consensus versus nonconsensus binding sites, Genome Res. 18 (2008) 1763–1777. https://doi.org/10.1101/gr.080622.108.

[28]   A.R. Cao, R. Rabinovich, M. Xu, X. Xu, V.X. Jin, 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. 286 (2011) 11985–11996. https://doi.org/10.1074/jbc.M110.217158.

[29]   M. Bieda, Unbiased location analysis of E2F1-binding sites suggests a widespread role for E2F1 in the human genome, Genome Res. 16 (2006) 595–605. https://doi.org/10.1101/gr.4887606.

[30]   Y. Tao, R.F. Kassatly, W.D. Cress, J.M. Horowitz, Subunit composition determines E2F DNA-binding site specificity., Mol. Cell. Biol. (1997). https://doi.org/10.1128/mcb.17.12.6994.

[31]   A.E. Kel, O. V. Kel-Margoulis, P.J. Farnham, S.M. Bartley, E. Wingender, M.Q. Zhang, Computer-assisted identification of cell cycle-related genes: new targets for E2F transcription factors, J. Mol. Biol. 309 (2001) 99–120. https://doi.org/10.1006/jmbi.2001.4650.

[32]   K. Zaragoza, V. Bégay, A. Schuetz, U. Heinemann, A. Leutz, Repression of Transcriptional Activity of C/EBPα by E2F-Dimerization Partner Complexes, Mol. Cell. Biol. 30 (2010) 2293–2304. https://doi.org/10.1128/MCB.01619-09.

[33]   H.M. Dingerdissen, J. Torcivia-Rodriguez, Y. Hu, T.C. Chang, R. Mazumder, R. Kahsay, BioMuta and BioXpress: Mutation and expression knowledgebases for cancer biomarker discovery, Nucleic Acids Res. (2018). https://doi.org/10.1093/nar/gkx907.

[34]   S.A. Forbes, N. Bindal, S. Bamford, C. Cole, C.Y. Kok, D. Beare, M. Jia, R. Shepherd, K. Leung, A. Menzies, J.W. Teague, P.J. Campbell, M.R. Stratton, P.A. Futreal, COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer, Nucleic Acids Res. 39 (2011) D945–D950. https://doi.org/10.1093/nar/gkq929.

[35]   J. Barretina, G. Caponigro, N. Stransky, K. Venkatesan, A.A. Margolin, S. Kim, C.J. Wilson, J. Lehár, G. V. Kryukov, D. Sonkin, A. Reddy, M. Liu, L. Murray, M.F. Berger, J.E. Monahan, P. Morais, J. Meltzer, A. Korejwa, J. Jané-Valbuena, F.A. Mapa, J. Thibault, E. Bric-Furlong, P. Raman, A. Shipway, I.H. Engels, J. Cheng, G.K. Yu, J. Yu, P. Aspesi, M. de Silva, K. Jagtap, M.D. Jones, L. Wang, C. Hatton, E. Palescandolo, S. Gupta, S. Mahan, C. Sougnez, R.C. Onofrio, T. Liefeld, L. MacConaill, W. Winckler, M. Reich, N. Li, J.P. Mesirov, S.B. Gabriel, G. Getz, K. Ardlie, V. Chan, V.E. Myer, B.L. Weber, J. Porter, M. Warmuth, P. Finan, J.L. Harris, M. Meyerson, T.R. Golub, M.P. Morrissey, W.R. Sellers, R. Schlegel, L.A. Garraway, The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity, Nature. 483 (2012) 603–607. https://doi.org/10.1038/nature11003.

[36]   S. Munro, U. Oppermann, N.B. La Thangue, Pleiotropic effect of somatic mutations in the E2F subunit DP-1 gene in human cancer, Oncogene. 33 (2014) 3594–3603. https://doi.org/10.1038/onc.2013.316.

[37]   J.F. Sayılgan, T. Haliloğlu, M. Gönen, Protein dynamics analysis reveals that missense mutations in cancer-related genes appear frequently on hinge-neighboring residues, Proteins Struct. Funct. Bioinforma. (2019). https://doi.org/10.1002/prot.25673.

[38]   M. Ivey-Hoyle, R. Conroy, H.E. Huber, P.J. Goodhart, A. Oliff, D.C. Heimbrook, Cloning and characterization of E2F-2, a novel protein with the biochemical properties of transcription factor E2F., Mol. Cell. Biol. 13 (1993) 7802–7812. https://doi.org/10.1128/MCB.13.12.7802.

[39] K. Moberg, M.A. Starz, J.A. Lees, E2F-4 switches from p130 to p107 and pRB in response to cell cycle reentry., Mol. Cell. Biol. 16 (1996) 1436–1449. https://doi.org/10.1128/MCB.16.4.1436.

[40] M.-L. Diebold, S. Fribourg, M. Koch, T. Metzger, C. Romier, Deciphering correct strategies for multiprotein complex assembly by co-expression: Application to complexes as large as the histone octamer, J. Struct. Biol. 175 (2011) 178–188. https://doi.org/10.1016/j.jsb.2011.02.001.

[41] L. Vandel, Residues phosphorylated by TFIIH are required for E2F-1 degradation during S-phase, EMBO J. 18 (1999) 4280–4291. https://doi.org/10.1093/emboj/18.15.4280.

[42] R.P. Rambo, ScÅtter, Bioisis.Net. (2017). http://www.bioisis.net/tutorials.

[43] V. Piiadov, E. Ares de Araújo, M. Oliveira Neto, A.F. Craievich, 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. 28 (2019) 454–463. https://doi.org/10.1002/pro.3528.

[44] A. Šali, T.L. Blundell, Comparative Protein Modelling by Satisfaction of Spatial Restraints, J. Mol. Biol. 234 (1993) 779–815. https://doi.org/10.1006/jmbi.1993.1626.

[45] P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Features and development of Coot, Acta Crystallogr. Sect. D Biol. Crystallogr. 66 (2010) 486–501. https://doi.org/10.1107/S0907444910007493.

[46] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, SoftwareX. 1–2 (2015) 19–25. https://doi.org/10.1016/j.softx.2015.06.001.

[47] G.A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, PLUMED 2: New feathers for an old bird, Comput. Phys. Commun. 185 (2014) 604–613. https://doi.org/10.1016/j.cpc.2013.09.018.

[48] P. Robustelli, S. Piana, D.E. Shaw, Developing a molecular dynamics force field for both folded and disordered protein states, Proc. Natl. Acad. Sci. U. S. A. (2018). https://doi.org/10.1073/pnas.1800690115.

[49] I. Ivani, P.D. Dans, A. Noy, A. Pérez, I. Faustino, A. Hospital, J. Walther, P. Andrio, R. Goñi, A. Balaceanu, G. Portella, F. Battistini, J.L. Gelpí, C. González, M. Vendruscolo, C.A. Laughton, S.A. Harris, D.A. Case, M. Orozco, Parmbsc1: a refined force field for DNA simulations, Nat. Methods. 13 (2016) 55–58. https://doi.org/10.1038/nmeth.3658.

[50] S. Piana, A.G. Donchev, P. Robustelli, D.E. Shaw, Water dispersion interactions strongly influence simulated structural properties of disordered protein states, J. Phys. Chem. B. (2015). https://doi.org/10.1021/jp508971m.

[51] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, Molecular dynamics with coupling to an external bath, J. Chem. Phys. 81 (1984) 3684–3690. https://doi.org/10.1063/1.448118.

[52] G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling, J. Chem. Phys. (2007). https://doi.org/10.1063/1.2408420.

[53] B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, LINCS: A linear constraint solver for

molecular simulations, J. Comput. Chem. 18 (1997) 1463–1472.
https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

[54]   U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh Ewald method, J. Chem. Phys. 103 (1995) 8577–8593. https://doi.org/10.1063/1.470117.

[55]   J. Pfaendtner, M. Bonomi, Efficient Sampling of High-Dimensional Free-Energy Landscapes with Parallel Bias Metadynamics, J. Chem. Theory Comput. 11 (2015) 5062–5067. https://doi.org/10.1021/acs.jctc.5b00846.

[56]   C. Paissoni, A. Jussupow, C. Camilloni, Determination of Protein Structural Ensembles by Hybrid-Resolution SAXS Restrained Molecular Dynamics, J. Chem. Theory Comput. 16 (2020) 2825–2834. https://doi.org/10.1021/acs.jctc.9b01181.

[57]   C. Paissoni, A. Jussupow, C. Camilloni, Martini bead form factors for nucleic acids and their application in the refinement of protein–nucleic acid complexes against SAXS data, J. Appl. Crystallogr. 52 (2019) 394–402. https://doi.org/10.1107/S1600576719002450.

[58]   P. V. Konarev, V. V. Volkov, A. V. Sokolova, M.H.J. Koch, D.I. Svergun, PRIMUS : a Windows PC-based system for small-angle scattering data analysis, J. Appl. Crystallogr. 36 (2003) 1277–1282. https://doi.org/10.1107/S0021889803012779.

[59]   D.I. Svergun, Determination of the regularization parameter in indirect-transform methods using perceptual criteria, J. Appl. Crystallogr. 25 (1992) 495–503. https://doi.org/10.1107/S0021889892001663.

[60]   C.W. Hopkins, S. Le Grand, R.C. Walker, A.E. Roitberg, Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning, J. Chem. Theory Comput. 11 (2015) 1864–1874. https://doi.org/10.1021/ct5010406.

[61]   M. Bonomi, G. Bussi, C. Camilloni, G.A. Tribello, P. Banáš, A. Barducci, M. Bernetti, P.G. Bolhuis, S. Bottaro, D. Branduardi, R. Capelli, P. Carloni, M. Ceriotti, A. Cesari, H. Chen, W. Chen, F. Colizzi, S. De, M. De La Pierre, D. Donadio, V. Drobot, B. Ensing, A.L. Ferguson, M. Filizola, J.S. Fraser, H. Fu, P. Gasparotto, F.L. Gervasio, F. Giberti, A. Gil-Ley, T. Giorgino, G.T. Heller, G.M. Hocky, M. Iannuzzi, M. Invernizzi, K.E. Jelfs, A. Jussupow, E. Kirilin, A. Laio, V. Limongelli, K. Lindorff-Larsen, T. Löhr, F. Marinelli, L. Martin-Samos, M. Masetti, R. Meyer, A. Michaelides, C. Molteni, T. Morishita, M. Nava, C. Paissoni, E. Papaleo, M. Parrinello, J. Pfaendtner, P. Piaggi, G.M. Piccini, A. Pietropaolo, F. Pietrucci, S. Pipolo, D. Provasi, D. Quigley, P. Raiteri, S. Raniolo, J. Rydzewski, M. Salvalaglio, G.C. Sosso, V. Spiwok, J. Šponer, D.W.H. Swenson, P. Tiwary, O. Valsson, M. Vendruscolo, G.A. Voth, A. White, Promoting transparency and reproducibility in enhanced molecular simulations, Nat. Methods. (2019). https://doi.org/10.1038/s41592-019-0506-8.

[62]   E. Valentini, A.G. Kikhney, G. Previtali, C.M. Jeffries, D.I. Svergun, SASBDB, a repository for biological small-angle scattering data, Nucleic Acids Res. 43 (2015) D357–D363. https://doi.org/10.1093/nar/gku1047.

FIGURES LEGENDS

Figure 1. (A) Schematic representation of E2F1 and DP1 domain organization 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 red and blue, respectively. The 6×His-tag is indicated as an orange sphere. (B) Cartoon representation of E2F/DP homologous proteins from available PDBs: 2AZE (top), corresponding to the marked-box and coiled-coil domains of E2F1/DP1, in complex with the Rb peptide; and 1CF7 (bottom), including the DBD of E2F4/DP2 bound to 15-bp DNA.

Figure 2. Experimental SEC-SAXS profiles and atomistic model for E2F1^{DCM}/DP1^{DCM}/DNA complex. (A) I(q) vs. q as semi-logarithmic representation, with Guinier fit (red line) for qR$_g$ < 1.3 is shown as inset. The black line represents the model SAXS profile, back-calculated using CRYSOL, displaying a χ^2 = 1.26 in the range 0.009-0.25 Å^{-1}. In the lower panel residuals are reported. (B) Normalized Kratky plot [(qR$_g$)^{2}I(q)/I(0) vs. qR$_g$] for the data in (A). (C) Representation of the E2F1^{DCM}/DP1^{DCM}/DNA model refined via single-replica SAXS-driven MD.

Figure 3. Schematic representation of the orientation adopted by the CC domain with respect to the DNA/DBDs. The DNA axis is aligned to the x axis, while the z axis was aligned to the axis connecting the DNA and DBDs geometric centers. In the plot are represented the x-y projections of the positions adopted by the last turns of the coiled-coil domain geometric center (black sphere in the model), once fixed the DNA and DBDs: here the observer is looking from the top of the z axis. Few representative structures, selected as the centers of the main six clusters, based on backbone-RMSD, are illustrated (in this case the observer is looking from the top of the x axis

Figure 4. E2F1/DP1 dynamics. (A) Per-residue Root mean square fluctuations (RMSF) of the E2F1 and DP1 DBDs. Regions showing RMSF greater than 0.25 nm, are highlighted with shades. (B) Cartoon representation of a DNA/DBDs structure, where the position of the identified flexible regions is indicated. (C) Highlights of the flexible DBDs regions, including: E2F1 N-terminal (left); DP1 N-terminal (center); DP1 α2-α3 loop, comprising residues 148-161. For each cartoon representation the dynamics regions are highlighted with transparent grey tubes, representing all the possible conformations of the simulated ensemble.

Figure 5. *In vitro* validation of DP1 stabilization by competition EMSA experiments. (A) Sequence details and the alignment of the probe and the designed unlabeled oligonucleotides -centered 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 E2F1 and DP1 proteins [11], and corresponding positions on the E2F1 probe, are highlighted in red and blue, respectively. Mutant nucleotides are indicated in lowercase. (B) Direct competition EMSAs quantification. The indicated unlabeled competitors were incubated at increasing concentrations (5x,

25x, 100x fold excess) with the E2F1-25 bp probe (20 nM), with the addition of recombinant E2F1^{DCM}/DP1^{DCM} 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 (15-mer) oligonucleotide competition slope. Representative gel of direct competition EMSA experiment is illustrated in the lower panel. As a positive control, the wild-type 25-mer E2F1 unlabeled oligo (Probe) was used as a 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 6. Schematic diagrams showing the somatic mutations in E2F1 and DP1. Data of somatic mutations in E2F1 (A) and DP1 (B) were retrieved from COSMIC and CCLE database and mapped onto the corresponding domains. The orange shaded windows of 20 residues each (on top of each diagram) highlight the mutation density (total unique samples/ aa) throughout the proteins. Mutations within or flanking the described flexible regions are labelled. The green shaded areas highlight the flexible regions in the DBDs according to the MD simulations. Height of the bars denotes the mutation count according to COSMIC and CCLE.