Tyr120Asp mutation alters domain flexibility and dynamics of MeCP2 DNA Binding Domain leading to impaired DNA interaction: atomistic characterization of a Rett syndrome causing mutation.

Ilda D'Annessa[1], Anna Gandaglia[2], Elena Brivio[2], Gilda Stefanelli[2], Angelisa Frasca[3]
Nicoletta Landsberger[2,3*], Daniele Di Marino[4*]

[1]Istituto di Chimica del Riconoscimento Molecolare, CNR, Milan, Italy, [2]San Raffaele Rett Research Unit, San Raffaele Scientific Institute, Milan, Italy, [3]Department of Medical Biotechnology and Translational Medicine, University of Milan, Milan, Italy, [4]Department of Informatics, Institute of Computational Science, Università della Svizzera Italiana, Lugano, Switzerland

ABSTRACT

Mutations in the X-linked *MECP2* gene represent the main origin of Rett syndrome, causing a profound intellectual disability in females. MeCP2 is an epigenetic transcriptional regulator containing two main functional domains: a methyl-CpG binding domain (MBD) and a transcription repression domain (TRD). Over 600 pathogenic mutations were reported to affect the whole protein; almost half of missense mutations affect the MBD. Understanding the impact of these mutations on the MBD structure and interaction with DNA will foster the comprehension of their pathogenicity and possibly genotype/phenotype correlation studies. Herein, we use molecular dynamics simulations to obtain a detailed view of the dynamics of WT and mutated MBD in the presence and absence of DNA. The pathogenic mutation Y120D is used as paradigm for our studies. Further, since the Y120 residue was previously found to be a phosphorylation site, we characterize the dynamic profile of the MBD also in the presence of Y120 phosphorylation (pY120). We found that addition of a phosphate group to Y120 or mutation in aspartic acid affect domain mobility that samples an alternative conformational space with respect to the WT, leading to impaired ability to interact with DNA. Experimental assays showing a significant reduction in the binding affinity between the mutated MBD and the DNA confirmed our predictions.

*** to whom correspondence should be addressed**

Daniele Di Marino,
Università della Svizzera Italiana.
Via Giuseppe Buffi 13, Lugano. Switzerland.
Email: daniele.di.marino@usi.ch

Nicoletta Landsberger,
Univeristy of Milan, 20090 Segrate Milano, Italy.
Email: nicoletta.landsberger@unimi.it

1 INTRODUCTION

Gene transcription is finely regulated by epigenetic modifications; in particular, DNA methylation of cytosine residues, mainly present in CpG or CpA dinucleotides, inhibits gene expression either directly by impeding the binding of transcription factors or indirectly by promoting the binding of Methyl-CpG-binding proteins (MBPs). MeCP2 (Methyl-CpG-binding protein 2) represents the founder member of the MBPs family and was originally discovered as a nuclear factor capable of binding *in vitro* DNA probes containing at least one symmetrically methylated CpG dinucleotide (1). *In vivo* the protein accumulates in mouse cells at pericentromeric heterochromatin, which contains highly methylated satellite DNA; this localization is methylation dependent. The *Mecp2* gene was cloned by Bird and colleagues and defined as a 486 residues long protein containing different domains, among which the most characterized being the methyl-CpG binding domain (MBD) and the transcription repression domain (TRD) (2). The importance of this X-linked gene emerged in 1999, when several patients affected by Rett syndrome (RTT) were found mutated in *MECP2* (3). Rett syndrome represents a devastating pediatric neurological disorder that, because of its incidence, is considered the main cause of severe intellectual disabilities in girls (3). The high relevance of MeCP2 in human health and the recent awareness that epigenetics plays a prominent role in the onset of intellectual disabilities have certainly accelerated the research efforts to unveil the structural properties of MeCP2 and to decipher its role in numerous cellular processes. At the functional level, MeCP2 was initially proposed to repress transcription by recruiting on methylated DNA co-repressor complexes such as Sin3A and N-CoR that contain histone deacetylase activities, thereby promoting chromatin compaction . The link between MeCP2 and chromatin structure has been then reinforced by the identification of several other interacting partners owing chromatin remodelling activities or the capacity to methylate histones . It was then proposed that MeCP2 also works as an architectural chromatin protein able to independently organize a highly compacted nucleosomal structure (9). In possible accordance with that, Skene and colleagues have proposed that in mature neurons, where MeCP2 reaches its maximal abundance, the protein functions as an alternative linker histone, organizing a specialized chromatin structure that dampens overall transcriptional noise (10). Subsequent studies have revealed additional molecular roles for MeCP2 including transcriptional activation, mRNA splicing regulation, and modulation of protein synthesis and microRNA processing (11). The MeCP2 ability to carry out so many diverse functions is probably permitted by its highly disorganized structure (12) that can be modulated through a set of post-translational modifications among which phosphorylation is the most frequent . Experimental evidence and computational predictions have detected about 30 phosphorylatable sites on MeCP2, comprising threonine, serines and tyrosines . Interestingly, some of these modifications are tissue specific, as in the case of S421 or S164, whose phosphorylation occurs mainly in neural tissues , while the modification of some residues (such as S421 phosphorylation and S80 dephosphorylation) depends on neuronal activity.

At the functional level, MeCP2 post-translational modifications seem to affect its binding to euchromatic or heterochromatic DNA, its interaction with its partners and probably also MeCP2 stability .

At the structural level, proteolysis analyses of MeCP2 revealed the presence of six main distinct domains corresponding to HMGD1 (aa 1-77), MBD (Methyl-CpG-binding domain; aa 78-163), HMGD2 (aa 164-204), TRD (Transcriptional repressor domain; aa 207-310), CTDα (C-terminal domain α; aa 311-354) and CTDβ (C-terminal domain β; aa 355-486) (16) (Figure 1A). The two HMGD domains share a modest degree of sequence similarity with the HMGA2 protein and are rapidly subjected to trypsin digestion, suggesting a poor degree of structuring. Similarly, the CTDs and a portion of the TRD are poorly structured .

Indeed, Circular Dichroism experiments and computational predictions have identified MeCP2 as an intrinsically disordered protein, displaying 60% of unstructured regions along the sequence (12). The advantage of this class of proteins is the ability to interact with a considerable number of molecular partners, both proteins and nucleic acids, thanks to their structural flexibility. Accordingly, it has been demonstrated that the TRD especially, but also the HMGD1 and CTDs domains, can undergo coil-to-helix transitions that probably permit the already mentioned capability of MeCP2 to interact with many different partners and to acquire multiple and diversified functions .

The MBD is the only MeCP2 domain for which the three-dimensional (3D) organization of the secondary structure elements is known. One NMR structure of the unbound MBD and several X-ray structures at different atomic resolutions of the domain bound to a methylated double stranded DNA are available . All the solved structures show a conserved 3D arrangement of the domain that is composed by three stranded anti-parallel β-sheet comprising residues W104-F132, with the third strand connected to a short α-helix. The N- and C-terminal regions of the domain are unstructured. A huge network of intra-domain hydrogen bonds and salt bridges holds the MBD fold. Moreover, residues D156-G161 comprise a peculiar Asx-ST motif anchored by hydrogen bonds occurring between the main chain of T158 and D156 and contributing to stabilize the domain for a proper interaction with DNA.

The importance of the MBD structure and function is also highlighted by results obtained through molecular genetics approaches, since it is known that almost half of the disease-causing missense mutations in *MECP2* occurs within this domain, including the most frequent T158M mutation. Few studies have proved that some RTT syndrome mutations within the MBD impair MeCP2 binding to methylated DNA and its capability to accumulate on highly methylated pericentromeric heterochromatin in mouse cells . Computational studies applied to the solved MBD structures have also been performed to explain the structural consequences of RTT mutations affecting the MBD ability to bind the double stranded DNA . In detail, it has been demonstrated that some mutations, such as R133C, R111G and D121G, alter MeCP2 binding properties abolishing several direct interactions with methylated or unmethylated DNA. Analogously other point mutations alter the stability of the domain fold, thus indirectly perturbing the interaction with DNA. This is the case for example for L100V, R106W, S134C, A140V, P152R, F155S, T158M and D156E mutations .

Recently, our attention has been focused on the functional characterization of Tyrosine 120 (Y120), located at the central portion of the MBD of MeCP2. This residue is highly conserved among all mammals and has been detected as a phosphorylation site . Notably, Y120 was related to RTT syndrome in 2001 when it was found mutated into an aspartic acid in a patient (32). In accordance with its location, previous studies demonstrated that the Y120D mutation causes a decrease in the binding affinity between MeCP2 and heterochromatin . In addition, the study of MeCP2 phosphorylation at Y120 permitted to reveal a novel localization of the protein to the centrosome, leading to a functional association of MeCP2 to cell growth, spindle geometry and microtubule nucleation . Mutations in aspartic acid of residues that can be phosphorylated are often considered as mimicking a constitutive state of phosphorylation: a possibility that, however, has always to be experimentally tested. Indeed, the differences in size and chemical properties between aspartate and tyrosine make this molecular mimicry more challenging. A deep comprehension of the structural and dynamical properties of the MBD in the presence of the Y120D mutation or its phosphorylation is thus needed to reveal differences and similarities between the two events. Thus, in order to describe at the atomic level the effects on the MBD structure and function of the pathogenic Y120D mutation or the addition of a phosphate group to the tyrosine 120, we have used a combined computational and experimental approach. In particular, we used molecular dynamics simulations to provide a detailed view of the MDB dynamics in the presence and absence of phosphorylation (pY120 and WT) and of the pathological Y120D mutant. Using

our computational approach, we were able to detect the effects of the two events (phosphorylation/mutation) on the internal motions of the domain and to rely them to its (mal)function. In particular, we could detect that both pY120 and Y120D alter the internal dynamics of the domain, although leading to a diverse structural output. In both cases, we predicted that these structural alterations could cause a significant impairment in the capability of the MBD to interact with DNA. Our predictions were confirmed by experimental assays showing a reduced affinity for the nucleic acid of the mutated domain, thus endorsing the quality of our model.

2 MATERIALS AND METHODS

2.1 Computational protocol.

The starting configuration of human MBD of MeCP2 spanning residues 91-168 was extracted from the X-ray structure deposited with Protein Data Bank (PDB) code 3C2I (18), upon elimination of the double stranded DNA co-crystallized with the domain (Figure 1B). We are aware that the more recent X-ray structure from Chia and colleagues (PDB ID: 5BT2) has been solved with a higher resolution (2.2 Å vs 2.5 Å) and a better metrics but nonetheless it harbors a pathogenic non-classical RTT mutation A140V. Using this structure, upon reverting the mutation to the wild type condition, as starting configuration would have affected all the data obtained through the simulations, and then we preferred to use the structure with PDB code 3C2I.

Three simulative systems were built, containing either the wild type domain (WT), the domain phosphorylated in position 120 (pY120) and the Y120D mutant domain. The topology of each system was obtained using tleap with Amber14 all atom force field (33). The triclinic simulative box, filled with TIP3P (34) water molecules and rendered electroneutral by addition of chloride counterions, consists of a final number of atoms of ~30.000 for each system. Simulations were performed using Amber14 pmemd.CUDA (33). The systems were first subjected to a round of minimization of 10000 steps of steepest descend followed by 10000 steps of conjugate gradient. Relaxation of water molecules and thermalization in NPT environment were carried out for 1.2 ns at 1 fs time-step. In particular, 6 runs of 200 ps each were carried out increasing the temperature of 50 K at each step, starting from 50 K to 300 K.

The systems were then simulated with a 2 fs time-step for 300 ns in periodic boundary conditions, using a cut-off of 8 Å for the evaluation of short-range non-bonded interactions and the Particle Mesh Ewald method for the long-range electrostatic interactions (35) . The temperature was kept constant at 300 K with Langevin dynamics (36) whereas pressure fixed at 1 Atmosphere through the Langevin piston method (37) . The bond lengths of solute and water molecules were restrained with the SHAKE (38) and SETTLE (39) algorithms, respectively. The atomic positions were saved every 250 steps (i.e. 0.5 ps) for the analysis with the Gromacs 4.6 package (40) , or with code written in-house.

Principal components analysis (PCA) was carried out on the 3Nx3N Cartesian displacement matrix whose elements are calculated as:

$$C_{ij} = \left(r_i^2 q_i q_j\right))$$

where N is the number of Cα atoms of the three systems and q_i is the (mass-weighted) displacement of the i-th Cα atoms from the reference value (after removal of rotational and translational degrees of freedom). The first few eigenvectors of the diagonalized covariance matrix usually account for a major fraction of the total variance and projection of atomic trajectories over the corresponding eigenvectors represents large collective atomic motions.

The dynamic cross-correlation (DCC) maps were built considering only the coordinates of the C-alpha atoms, enough to describe the largest system motions. The elements of the DCC map (*Cij*) are calculated as:

$$C_{ij} = \left(\Delta r_i \cdot \Delta r_j\right) / \sqrt{r_i^2} \cdot \sqrt{r_j^2}$$

where Δri is the displacement from the mean position of the *i*-th atom and the ⟨⟩ represents the time average over the whole trajectory. Positive *Cij* values represent a correlated motion between residues *i* and *j* (i.e. the residues move in the same direction). Negative values of *Cij* represent an anti-correlated motion between residues *i* and *j* (i.e. they move in opposite directions).

Clustering analysis was performed by fitting the structure of the protein extracted from the trajectories relative to the Cα atoms and then clustered using a Root Mean Squared Deviation (RMSD) cut-off of 0.25 nm. The representative configuration of the most populated cluster of each system (i.e. centroid) was selected for further analyses.

The three centroids (i.e. one for each system) were used to calculate the electrostatic potential surface using continuum solvation methods with the Adaptive Poisson-Boltzmann Solver (APBS) tools integrated in PyMOL (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.). The results were visualized by coloring the molecular surfaces of the three centroids in a blue to red gradient to represent electrostatic values.

The centroids were also used for the docking runs with double stranded DNA. The 3D structure of the nucleic acid was taken from the original 3C2I crystal structure (18).

The complexes obtained through the docking runs were compared to the original human MeCP2 MBD-DNA X-ray structure (PDB ID: 3C2I) as a reference structure to validate the results. The predictions were carried out using the Haddock web server (41) , particularly suitable for docking with nucleic acids, using as active residues (AIRs), i.e. the residues composing the binding site, amino acids K109, R111, K112, R115, S116, T160, R167, R168 and bases Ade6, mCyt7, Gua8, Gua9, Thy10, Thy15. The complexes obtained are classified based on the Haddock score, which is a sum of different contributions for the binding. The lowest is the Haddock score the highest is the affinity between ligand and receptor (42) .

The residues for the AIRs construction were assumed from the analysis of the hydrogen bonds between the protein and the DNA as a function of time, calculated along the trajectory of the MeCP2-MBD with methyl-cytosine containing DNA from the *BDNF* promoter (PDB ID: 3C2I) (mCyt-WT). Simulation of this protein-DNA complex was carried out following the simulative protocol previously described for the domain alone in solution. The only exception is the addition of sodium instead of chloride ions as counterions to neutralize the system, for a final number of atoms of 43562.

2.2 Plasmids.

The Y120D/F-MeCP2 fragments (aa 74-172) were PCR amplified from the corresponding pEGFPC1-MeCP2 derivatives already described in and cloned into the pTYB1 vector (NEB). The WT pTYB1-MeCP2 (74-172) vector was described in . All constructs were verified by sequencing.

2.3 Purification of recombinant MeCP2 fragments.

An overnight pre-culture of ER2566 cells transformed with an MeCP2(74–172)TYB1 plasmid was diluted 100 times in LB (100 μg/ml ampicillin) and allowed to grow at 37°C till 0.6–0.8 OD at 595 nm. The expression of the recombinant fragments was induced by adding IPTG (1mM) to the medium and incubating at 30°C for 5 hours. Bacteria were collected and lysed in lysis buffer (750 mM NaCl, 20 mM Tris-HCl pH 8, 1 mM EDTA pH 8, 0.1% Triton X100) followed by sonication (15 min 100% amplitude). Supernatant lysates were collected by centrifugation and passed 5 times through a chitin resin (NEB) column. Proteins were eluted by incubating over night at 4°C the columns in lysis buffer containing 50 mM DTT, and collecting fractions of 200 μl. Eluted polypeptides were dialyzed against 20 mM Hepes, 100 mM NaCl, 10% Glycerol, 0.250 mM EDTA pH 8 and quantified using nanodrop.

2.4 Electrophoretic mobility shift assay.

A 20bp DNA fragment corresponding to the same portion of the *BDNF* promoter used to determine the X-ray structure of the MBD of MeCP2-DNA complex was chosen as probe for our experiments. Oligonucleotides sequences were the following:

Forward 5'-TCTGGAA**C**GGAATTCTTCTA-3'

Reverse 5'-ATAGAAGAATTC**C**GTTCCAG-3'

In bolds are highlighted residues that could be either methylated (methyl CG probe) or not methylated (unmethylated probe). Increasing amounts of purified MeCP2 fragments (0, 125, 250, 500 ng) were incubated with 0.1 mg/ml BSA and 200 ng of annealed probe in EMSA buffer (10 mM Tris HCl, 50 mM KCl, 0.5 mM MgCl2, 0.1 mM EDTA, 5% glycerol) at RT for 30 min. Samples were loaded on a 6% acrylamide gel in 1X TAE. Electrophoresis was developed at 100/120 V for 40 min/1 h. Gels were stained with ethidium bromide. The same amounts of purified MeCP2 peptides used for the EMSA assay (125, 250, 500 ng) were loaded on a 15% acrylamide gel and then subjected to Coomassie staining. For K_d evaluation, Coomassie stained bands were quantified using the ImageJ software in order to obtain the best estimate of the peptide concentration [P_{tot}] in the EMSA itself. For each lane/reaction of the EMSA gels the amount of total DNA loaded, that of the shifted (i.e. peptide-bound) [DP] and of the free [D_{free}] DNA were quantified using the ImageJ software. By applying the equation $K_d = ([D_{free}][P_{free}]/[DP]) \rightarrow K_d = [D_{free}][P_{tot}]/[DP] - [D_{free}]$, K_d was calculated for each lane, then, for each gel, the value obtained at the different protein concentrations were mediated and used for statistical analysis.

3 RESULTS

3.1 DNA restricts the mobility of the methyl-binding domain of MeCP2 in solution.

The first X-ray structure of the MBD of MeCP2 in complex with a methylated DNA segment of 20 bp from the *BDNF* promoter was released in 2008 (18) . The structure shows that the MBD embraces the DNA at the level of the major groove establishing a large network of hydrophilic interactions, either direct or water mediated. The interactions involve amino acids from the loop L1, composed by residues R111-G118, and from the C-terminal region (Figure 1B and Supplementary Table T1). However, most of these interactions occur with backbone phosphates of the nucleic acids and only three residues establish base-specific hydrogen bonds, R111, D121 and R133 (Supplementary Table T1). In particular, D121 interacts with the methyl group of one mCyt, while the two arginines, which share a symmetric position, interact with the guanines of the two mCpG pairs. This interaction is assisted by two salt bridges, R111-D121 and R133-E137, that fix the position of the side chains of the two arginines toward the guanines (43) .

The X-ray crystal structures of proteins or macromolecular complexes represent single conformations in crystalline environments and this causes the loss of information about the motions of the amino acids side chains, secondary structure elements and, more generally, about the entire macromolecules. MD simulation is a solid approach to study the correlation between the structure and dynamic of biological molecules such as proteins, nucleic acids and macromolecular complexes (44).

In order to achieve a wider picture of the internal motions of the MeCP2 MBD-DNA complex we have monitored the complex for 300 ns of classical molecular dynamics (MD) simulation. This allowed to better understand which are the most important structural and dynamical features that make stable and, consequently, functional the MeCP2-DNA complex. First, we evaluated the amplitude of the motion of the MeCP2 MBD domain and we found that its dynamics is strictly restricted by the presence of the DNA, as reported in the profile of the per-residue Root Mean Square Fluctuation (RMSF) that barely

reaches values of 0.3 nm in the unstructured C-terminal region (Supplementary Figure S1A). Indeed, the structure of the domain does not deviate considerably from its starting configuration. The RMSD of the Cα atoms plotted as a function of time reaches the plateau at around 0.2 nm already after 50 ns of simulation (Supplementary Figure S1B). As expected, the reason for the high stability of the protein domain is the strong interaction with the DNA, as observed in X-ray structure. All the interactions involving the loop L1 that were described in the crystal are maintained along the trajectory (K109, R111, K112, S113, A115, S116, K119), as well as the one with R133, while those with residues D121, D134 and V159 are lost (Figure 1B, left vs right panel). On the other hand, a series of new interactions appears along the trajectory, in particular at the level of the C-terminal portion of the MBD, involving residues T160, R167 and R168, that cooperate to tighten the MBD-DNA binding (Figure 1B and Supplementary Table T1). The restriction in the motion of the domain can be also appreciated by clustering the structures sampled along the trajectory upon fixing the conformation of the DNA. This helps to establish at which extent the nucleic acid influences domain flexibility. Using a cut-off of 1.6 Å a quite large spread of structures could be expected. On the contrary, we find that the first two clusters account for 90% of total conformations sampled, confirming that the presence of the DNA induces some rigidity in the domain. Our findings on the restricted mobility of the domain in the presence of methylated DNA are in line with previous studies in which mass spectrometry-based hydrogen/deuterium exchange have established that, despite the intrinsically disordered nature of MeCP2, the MBD global flexibility is restricted by the DNA (27) .

We then proceeded by focusing our attention on the behavior of Y120 along the simulation. Since the tyrosine is not directly involved in DNA binding, we hypothesized that its phosphorylation or mutation can affect the internal stability of the MBD structure. In the crystal structure this residue establishes a direct hydrogen bond with the phenolic oxygen of Y95 that, by the way, is lost soon upon the domain starts to move in solution. Thus, Y120 is solvent exposed along the entire trajectory and does not engage any further interaction.

To elucidate the molecular consequences of the Y120 phosphorylation or mutation in aspartic acid we performed two additional MD simulations of 300 ns each, of pY120 and Y120D bound to the same molecule of DNA used for the wild type MeCP2 MBD-DNA complex simulation. Curiously, we couldn't find any relevant structural or dynamical difference comparing the results coming from the pY120 and Y120D simulation with the wild type (data not shown). We then hypothesized that these two modifications can have an effect on the structure and the dynamics of the domain in the apo state, thus affecting its ability to achieve a functional conformation able to bind the DNA substrate. We have thus carried out simulations of the MBD in the absence of the DNA for the WT, pY120 and Y120D and further we have analysed their ability to bind the double helix.

3.2 Phosphorylation or mutation of residue Tyr120 affects the dynamics of loop L1

Simulation of the wild type MBD of MeCP2 free in solution shows that the domain's 3D arrangement is highly stable. The RMSD of the Cα atoms reaches the plateau early after 100 ns of simulative time and this behaviour is also observed in the presence of the phosphorylation (Supplementary Figure S2A). The introduction of the Y120D mutation seems to have a slightly deeper effect, with the domain displaying a larger deviation from its starting configuration, reaching a stable RMSD trend only after 180 ns. However, since the jump in the RMSD observed between 100 and 180 ns in the Y120D system is less than 0.2 nm, we performed all the analysis considering the last 200 ns of simulative time for all the three systems. The WT, pY120 and Y120D domains share a common profile of fluctuation, with the loop L1 and the C-terminal portion of the MBD being the most flexible regions (Supplementary Figure S2B). Interestingly, these regions are those deputed to interact with the DNA, as they appear in the MBD-DNA crystal structure, or in the 300 ns simulation of this complex here

performed (Supplementary Table T1). As expected, the degree of fluctuation of all the domains, including the wild type, is much higher than that observed for the domain in the presence of DNA, confirming the effect of the nucleic acid on the domain dynamics (Supplementary Figures S1A and S2A).

Notwithstanding the comparable profile of fluctuation of the three systems, decomposition of the motion in its principal components highlights that each domain samples a peculiar conformational space when compared among them. In all the cases the 50% of total motion is covered by the first three eigenvectors, or even by the first two in the case of WT and pY120 (Supplementary Figure S3). More differences are detected by analysing the projection of the motion along each of these eigenvectors. In particular, we can clearly see how the fluctuation of loop L1 in the WT domain is equally distributed along the first three eigenvectors, while it is mainly described by eigenvectors 2-3 and 1-2 in the case of pY120 and Y120D, respectively (Figure 2A). This behaviour can be better appreciated by the three-dimensional representation of the motion along eigenvectors 1-3, showing that the direction of the motion, and so the space sampled, of the three MBD isoforms (WT, pY120 and Y120D) is completely different (Figure 2B). This is reflected by differences in the correlation of the motion of the diverse regions composing the domain. The Dynamic Cross Correlation maps (DDC) reported in Figure 3A and B, which gives a rationale to the correlated motion among the residues, highlights this behaviour, particularly evident when comparing the Y120D MBD to the WT one (Figure 3B). In the WT, we can observe a high degree of correlation of residues 92-102 with residues 102-108 and 158-162 (Figure 3A and B, upper triangle, rectangles A1 and B1) and a high degree of anti-correlation of residues 164-168 with amino acids 92-102 and 154-160 (Figure 3A and B, upper triangle, rectangles C1 and D1). In the case of loop L1, it shows an internal anti-correlation between residues 116-118 and residues 112-114 and a positive correlation with residues 120-128 (Figure 3A and B, upper triangle, rectangles E1). Concerning pY120, we observe conservation in the anti-correlation of residues 164-168 with 92-102 and 154-160 (Figure 3A lower triangle, rectangles C2 and D2), while the correlation between 92-102 and 158-162 is lower and that with 102-116 abolished (Figure 3A lower triangle, rectangles B2). Small differences are displayed also at the level of loop L1 (Figure 3A, lower triangle, rectangles E2). When considering Y120D, we can see how the whole correlation profile is smoothed and rectangles A-D completely disappear (Figure 3B, lower triangle). Interestingly, high degree of anti-correlation appears at the level of residues 110-116 of loop L1 with residues 122-126, 131-139 and 156-160 (Figure 3B, lower triangle, rectangle F3).

We can then argue that the introduction of a phosphorylation in position Y120 or its mutation in aspartic acid change the profile of the motion of the protein in solution and the communication between residues shown to be important for the DNA binding, thus probably affecting its ability to establish a functional interaction with DNA. To further test this hypothesis, we have clustered the structures sampled along each of the three trajectories in order to extract the most representative configurations to be used for the docking predictions. The docking calculations are useful to integrate the data coming from the MD simulations providing structural and energetic information not directly detectable from the solely analyses of single trajectories.

3.3 Y120 affects the dynamic profile of the MBD of MeCP2 and its ability to bind DNA

We started by clustering the trajectories of the domains in solution as reported in the Materials and Methods section. For the WT, pY120 and Y120D domains we obtained 10, 12 and 12 clusters respectively, with the first cluster accounting for more than 90% of total structures sampled in the last 200 ns of simulations (Supplementary Figure S4). The representative structure of the most populated cluster (i.e. centroid) of each system was chosen to perform the docking calculations. Since the centroids represent the most frequently sampled structure of the MBD in each system it is reasonable to consider them as the most suitable conformations to be prone to interact with the DNA. Based on this hypothesis these structures were also

used to map the surface electrostatic potential in order to detect whether the effect of the modification on residue Y120 could be propagated over the entire domain's surface. Although the chemical structure of phospho-tyrosine differs significantly from the side chain of the aspartic acid the effect of the charge distribution on the molecular surface is similar in both systems when compared to the WT (Figure 3C-E), thus possibly influencing the ability of the MBD to interact with the DNA in a related manner.

Even if the structure of the MBD-DNA wild type complex is available in the PDB (18), we performed redocking calculations for the WT. This has been done both to verify the reliability of our predictions and to prove that, when alone in solution, the MBD samples a conformation that allows it to properly interact with the DNA.

In order to establish to which extent the change in the mobility profile of the MBD affects its binding to DNA, we performed docking predictions using the structure of the *BDNF* promoter co-crystallized with wild type MBD, i.e. the same substrate used for the MBD-DNA complex simulation (18). For each system, the complexes resulting from the docking runs have been divided into families, based on the RMSD between the structures, and these families have been further classified based on the Haddock score (see the Methods section). The predictions gave rise to 10 families for the WT, 11 for pY120 and 10 for Y120D, with the best family having a Haddock score of -110.1, -75.2 and -78.4, respectively. Figure 4 shows the superposition on the DNA atoms of the best complex obtained for each system with the DNA atoms of the X-ray reference structure (PDB ID 3C2I). Remarkably, the best WT complex is highly superimposable with the X-ray (Figure 4A), displaying an RMSD between the proteins of 0.26 nm upon fitting on the DNA atoms. In the case of pY120 and Y120D, the displacement between the docked MBD domains and the reference X-ray structure is much larger so that the binding interface and profile of interaction is completely different when compared to the crystal structure. It is important to remark that the DNA binding regions are extremely positively charged in the three structures used for the docking calculations (Figure 3C-E) making them electrostatically complementary with the DNA backbone that is negatively charged. Notwithstanding this conservation of the electrostatic properties of the binding surface, only the WT structure can properly interact with the substrate (Figure 4A). This suggests that pY120 and Y120D explore a conformational space in solution that induces them to achieve a stable configuration with a reduced affinity for the DNA and due to a different dynamic pattern the two domains are not able to establish a functional interaction with the substrate. It should also be underlined that among all the families of complexes obtained for pY120 and Y120D, none resembles the correct MBD-DNA binding mode (Supplementary Figure S5).

We proceeded by experimentally confirming our prediction through electrophoretic mobility shift assays (EMSAs) in which the ability of three diverse MeCP2-MBD fragments (aa 74-162; see also ref. 15) to bind methylated and unmethylated DNA probes was compared. In particular, as protein fragments we used increasing amounts of the WT and Y120D polypeptides, together with a Y120F fragment corresponding to a non phosphorylatable derivative, herein used to test a diverse and not negatively charged mutation of the Y120 residue. As modified or unmodified probes, we used the same portion of the *BDNF* promoter used for the computational studies (this paper) and by Ho and colleagues for structural studies (18) . By observing not only the shifted DNA but particularly focusing on the unbound probe, it appears quite evident that both WT and Y120F polypeptides exhibit overlapping DNA binding properties characterized by a subtle but consistently higher affinity for methylated DNA (Figure 5A). On the contrary, and in perfect accordance with our computational studies, the Y120D polypeptide manifests a dramatic reduction in the capability to form a discrete shifted band (Figure 5A). The significance of the obtained results was verified by comparing through Coomassie staining the effective amounts of polypeptides present in the assay (Figure 5B). We repeated the same procedure with different WT and Y120D protein

purifications. Then, the gels and Coomassie resulting from three experiments were used to estimate DNA affinity through the evaluation of K_d. Based on the Coomassie staining we selected the three best experiments (including the one shown in panel A) and estimated DNA affinity through the evaluation of the K_d. (see Materials and Methods). We confirmed that WT and Y120F polypeptides share a substantially similar affinity for DNA, while Y120D fragment exhibits an increased K_d of approximately 7 folds for the methylated DNA and a trend towards increase (4 times) for the unmethylated probe. therefore these data testify a reduced affinity for the double helix independently of its modification state (Figure 5C).

4 DISCUSSION

Since mutations in the *MECP2* gene represent the most relevant cause of Rett syndrome, the MeCP2 protein has been subjected to intense studies. These include the analysis of the structural and functional effects of human pathogenic mutations, in order to detect which are the molecular forces that trigger the interaction of MeCP2 with methylated DNA.

A large fraction of RTT causing mutations is located in the MBD domain and affects the nature of residues that bind the nucleic acid or that are critical for the correct folding of the domain. Previous studies attempted to explain the structural consequences of the most frequent MBD mutations in order to gain deeper knowledge on the MBD-DNA binding process and on protein function. For example, using fluorescence spectroscopy and circular dichroism, Ghosh and colleagues established that the RTT hotspot mutations R106W, F155S and T158M, disrupt protein function by affecting the thermal stability and secondary structure content of the domain (45) . Later on, computational studies performed on a vast number of mutant isoforms evaluated the structural effect of these variants to shed light on the functional impairment caused by mutations of residues not directly involved in DNA binding and thus difficult to be directly related to functional impairment .

In the case of our study, the highly conserved Y120 residue is located downstream loop L1 and it is not directly involved in DNA binding. However, Y120 mutation in aspartic acid severely compromises the functionality of MeCP2, as demonstrated by its isolation in a RTT patient (32) . Further, it has recently been demonstrated that Y120 is a site of phosphorylation, thus probably regulating MeCP2 activity . As mentioned, MeCP2 phosphorylation is a well-known cellular strategy to modulate the affinity of the protein for chromatin or its protein partners, therefore modulating its function . Since the structural consequences on the MBD of MeCP2 Y120 phosphorylation or mutation remain unknown, we have engaged into a study aimed at using computational biophysics approaches to compare the structural properties of the WT MBD with respect to its pY120 or Y120D isoforms.

As observed by X-ray crystallography, the residue in position 120 is solvent exposed and engages a unique interaction with Y95 in the crystal structure but it does not bind any of the residues composing the DNA bind site or directly the DNA. For this reason, solely from the structural analysis it is not evident how phosphorylation or mutation might affect DNA binding affinity. Here, using MD simulations we are able to follow the time evolution of the domain structure and give a rationale to the effect of the phosphorylation/mutation. We show here that none of these two events destabilizes the three-dimensional arrangement or structural stability of the domain. However, its flexibility and internal dynamics (Figure 2 and supplementary Figure S2) are altered upon mutation or phosphorylation of Y120 and this is reflected by loss in the correlation profile of the motion between different regions (Figure 3A and B). As a consequence, the domain samples a conformational space not suitable for its interaction with the substrate (Figures 4 and 5). Y120 is surrounded by hydrophilic (Q110) or positively charged (K109, R111, K119) residues. In the WT system it does not interact directly with any of them but the addition of a phosphate group in position 120 creates a different electrostatic potential nearby (Figure 3C-E), thus

inducing these residues to rearrange their lateral chains in the space and to experience a different motion. This effect is more pronounced when Y120 is replaced by a charged carboxylic group. The lateral chain of D120, less bulky than pY120, allows the residue to start interacting with Q110 and R111 at the end of the simulation (data not shown). In particular, an altered mobility of loop L1 is the basis of the impaired affinity of the domain for DNA, as shown by PCA analysis (Figure 2), thus suggesting the central role of this region in guiding protein-DNA recognition and interaction. Indeed, docking predictions showed that both pY120 and Y120D form a non-functional complex with the *BDNF* promoter, characterized by a distorted protein-DNA interface and lowered binding affinity score. Importantly, our predictions for the MBD containing the Y120D mutation have been fully supported by EMSA assays, therefore indicating the strength of our computational studies. However, for obvious practical reason, we could not test DNA binding of the pY120-MBD but only demonstrate that the substitution with a so-called phospho-defective residue (Y120F) does not affect the MBD affinity for the double helix. This result further emphasizes that the inability of the MBD domain to recognize the DNA is exclusively due to the presence of a negative charge (i.e. aspartic acid) in position 120.

Our data suggest that the missense Y120D mutation more profoundly affects the MBD structure and its capability to bind DNA with respect to its post-translational modification (phosphorylation), giving a rationale for the pathogenicity of the mutation compared to the post-translational control of the phosphorylation. This can be due to the lower steric hindrance of the carboxyl group with respect to the phosphorylated tyrosine, allowing D120 to stabilize to a higher extent the new conformation of the domain. Thus, we speculate that Y120 phosphorylation might be used by cells to modulate but not abolish MeCP2 binding to chromatin and/or to influence MeCP2 subcellular localization (see Bergo et al., 2015).

We need to say that our data would increase their significance if we could use as starting model for our simulations the apo unbound state of the MBD. Indeed the conformation of the domain, and in particular of the loop L1 and of residue Y120 could differ from those observed in the DNA-bound state but unfortunately the unbound structure is not available at the moment.

To conclude, our combined computational and experimental approaches have demonstrated the drastic reduction in affinity of the Y120D-MBD for DNA, therefore providing a molecular explanation for the pathological consequences of this mutation. These results have proved the reliability of our computational studies, opening the possibility to apply them to other *MECP2* variants. By predicting at the atomic detail the effects of pathogenic mutations on the MBD structure of MeCP2 and on its capability to bind DNA, these *in silico* studies might contribute to understand MeCP2 activity, predict functional modulation by PTMs and also be useful instrument for genotype/phenotype correlation studies.

ACKNOWLEDGMENTS

We thank the granting agencies that have permitted these studies. In particular, NL and her group are indebted to proRETT ricerca, an association of Italian parents of Rett girls whose support not only has permitted these studies but also represent our major source of drive. ID was funded by the PERTNET project granted to Prof. Giorgio Colombo by the Ministero degli Affari Esteri e della Cooperazione Internazionale. Computational resources were provided by the Swiss National Supercomputing Centre (Project ID s687) We thank Dr. Giorgio Binelli for statistical analyses.

FIGURE LEGENDS

Figure 1. Structure of the MBD-DNA complex. (**A**) Upper panel shows the domains subdivision of the human MeCP2 sequence; lower panel shows the MBD-DNA X-Ray structure from Ho et al. (18). The MBD and DNA are shown in red and

grey ribbon respectively and the van der waals surface of the molecules is also reported. (**B**) The left and right panels show the MBD-DNA profile of interaction in the X-ray structure and along the simulation, respectively. The side chains of the MBD residues interacting with the DNA in X-ray structure (left panel) and along the simulation (right panel) are shown in red stick. The stick representation of the two methyl-Cytosines is also depicted and residue Y120 object of this study is shown. Residues for which the interaction with the DNA is not conserved between the two systems (i.e. are present only in the X-ray or appear only in the simulation) are underlined.

Figure 2. Principal Component Analysis. (A) Fluctuation along the first three eigenvectors for the WT (black line), pY120 (red line) and Y120D (green line) of the MBD. **(B)** Projection of the motion along the first three eigenvectors mapped on the structure of the three MBD domains showing the amplitude of the motion along each eigenvector.

Figure 3. Cross Correlation Map and Electrostatic Surface Ptential. Comparison of the Dynamic Cross Correlation (DCC) map between **(A)** WT (upper triangle) and pY120 (lower triangle) and **(B)** WT (upper triangle) and Y120D (lower triangle). The colour scale defining the degree of correlation is reported at the bottom of the panels. Values comprised between -0.5 and -1 defines high degree of anticorrelation (residues move along the same vector but in opposite directions); values between 0.5 and 1 define high correlation (residues move along the same vector in same directions); values between -0.5 and 0.5 display no correlation. The regions of higher negative or positive correlation commented in text are defined by black rectangles.

Electrostatic potential mapped on the molecular surface of the MeCP2 MBD in the WT **(C)**, pY120 **(D)** and Y120D **(E)**. The MBD is coloured according to the electrostatic potential, blue is positive, white is neutral and red is negative. The three lower inserts report the zoom of the regions where the Y120, pY120 and the mutant D120 are located in the MBD with respect to the DNA binding site. The three residues Y120, pY120 and D120 are also reported as stick.

Figure 4. MBD-DNA docking results. Superposition on the DNA atoms between the MeCP2 MBD-DNA X-ray structure and the best complex predicted by docking experiments for the WT **(A)**, pY120-WT **(B)** and Y120D **(C)** domains. In all three panels the X-ray complex is shown in grey while the domain is shown in orange, green and red for the docking results of WT, pY120 and Y120D, respectively.

Figure 5. MBD-DNA binding affinity assay. (A) Representative image of an EMSA assay, showing the binding of the WT-, Y120F- and Y120D-MeCP2 MBD fragments (0, 125, 250, 500 ng) to a CpG methylated (upper panel) and unmethylated (lower panel) portion of the *BDNF* promoter sequence (n=5). **(B)** Representative Coomassie staining (corresponding to the experiment of panel A), showing the effective amounts of MeCP2 polypeptides used for the EMSA assays. **(C)** Graph showing the Kd of the MeCP2 fragments, estimated from the EMSA gels. Data are represented as mean±s.e.m. (n=3). Data were analyzed by randomized block ANOVA to test for the effect of the gels, considered as blocks. Once this effect was ruled out, significance between experimental groups was evaluated by Duncan's post-hoc test, with $\alpha = 0.05$.

REFERENCES

1. Meehan, R.R., Lewis, J.D. and Bird, A.P. (1992) Characterization of MeCP2, a vertebrate DNA binding protein

with affinity for methylated DNA. *Nucleic acids research*, **20**, 5085-5092.
2. Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurerfogy, I., Jeppesen, P., Klein, F. and Bird, A. (1992) Purification, Sequence, and Cellular-Localization of a Novel Chromosomal Protein That Binds to Methylated DNA. *Cell*, **69**, 905-914.
3. Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U. and Zoghbi, H.Y. (1999) Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat Genet*, **23**, 185-188.
4. Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N. and Bird, A. (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*, **393**, 386-389.
5. Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J. and Wolffe, A.P. (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet*, **19**, 187-191.
6. Lyst, M.J., Ekiert, R., Ebert, D.H., Merusi, C., Nowak, J., Selfridge, J., Guy, J., Kastan, N.R., Robinson, N.D., de Lima Alves, F., Rappsilber, J., Greenberg, M.E. and Bird, A. (2013) Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nat Neurosci*, **16**, 898-902.
7. Bedogni, F., Rossi, R.L., Galli, F., Gigli, C.C., Gandaglia, A., Kilstrup-Nielsen, C. and Landsberger, N. (2014) Rett syndrome and the urge of novel approaches to study MeCP2 functions and mechanisms of action. *Neurosci Biobehav R*, **46**, 187-201.
8. Guy, J., Cheval, H., Selfridge, J., Bird A. (2011) The role of MeCP2 in the brain. *Ann Rev Cell Dev Biol*, **27**, 631-652.
9. Georgel, P.T., Horowitz-Scherer, R.A., Adkins, N., Woodcock, C.L., Wade, P.A. and Hansen, J.C. (2003) Chromatin compaction by human MeCP2. Assembly of novel secondary chromatin structures in the absence of DNA methylation. *The Journal of biological chemistry*, **278**, 32181-32188.
10. Skene, P.J., Illingworth, R.S., Webb, S., Kerr, A.R., James, K.D., Turner, D.J., Andrews, R. and Bird, A.P. (2010) Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. *Molecular cell*, **37**, 457-468.
11. Cheng, T.L.a.Q., Z. (2014) MeCP2: multifaceted roles in gene regulation and neural development. *Neurosci Bull*, **30**, 601-609.
12. Adams, V.H.M., S.J.; Wade, P.A.; Woodcock, C.L. and Hansen, J.C. (2007) Intrinsic disorder and autonomous domain function in the multi functional nuclear protein, MeCP2. *The Journal of biological chemistry*, **282**, 15057–15064.
13. Bellini, E., Pavesi, G., Barbiero, I., Bergo, A., Chandola, C., Nawaz, M.S., Rusconi, L., Stefanelli, G., Strollo, M., Valente, M.M. *et al*. (2014) MeCP2 post-translational modifications: a mechanism to control its involvement in synaptic plasticity and homeostasis? *Front Cell Neurosci*, **8**.
14. Zhou, Z., Hong, E.J., Cohen, S., Zhao, W.N., Ho, H.Y., Schmidt, L., Chen, W.G., Lin, Y., Savner, E., Griffith, E.C., Hu, L., Steen, J.A., Weitz, C.J. and Greenberg, M.E. (2006) Brain-specific phosphorylation of MeCP2 regulates activity-dependent Bdnf transcription, dendritic growth, and spine maturation. *Neuron*, **52**, 255-269.
15. Stefanelli, G., Gandaglia, A., Costa, M., Cheema, M.S., Di Marino, D., Barbiero, I., Kilstrup-Nielsen, C., Ausio, J. and Landsberger, N. (2016) Brain phosphorylation of MeCP2 at serine 164 is developmentally regulated and globally alters its chromatin association. *Sci Rep-Uk*, **6**.
16. Ghosh, R.P., Nikitina, T., Horowitz-Scherer, R.A., Gierasch, L.M., Uversky, V.N., Hite, K., Hansen, J.C., Woodcock, C.L. (2010) Unique physical properties and interactions of the domains of methylated DNA binding protein 2. *Biochemistry*, **49**, 4395-4410.
17. Wakefield, R.I.D., Smith, B.O., Nan, X.S., Free, A., Soteriou, A., Uhrin, D., Bird, A.P. and Barlow, P.N. (1999) The solution structure of the domain from MeCP2 that binds to methylated DNA. *Journal of molecular biology*, **291**, 1055-1065.
18. Ho, K.L., Mcnae, L.W., Schmiedeberg, L., Klose, R.J., Bird, A.P. and Walkinshaw, M.D. (2008) MeCP2 binding to DNA depends upon hydration at methyl-CpG. *Molecular cell*, **29**, 525-531.
19. Chia, J.Y., Tan, W.S., Ng, C.L., Hu, N.J., Foo, H.L. and Ho, K.L. (2016) A/T Run Geometry of B-form DNA Is Independent of Bound Methyl-CpG Binding Domain, Cytosine Methylation and Flanking Sequence. *Sci Rep-Uk*, **6**.
20. Scarsdale, J.N., Webb, H.D., Ginder, G.D. and Williams, D.C. (2011) Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. *Nucleic acids research*, **39**, 6741-6752.
21. Wan, W.Y. and Milner-White, E.J. (1999) A natural grouping of motifs with an aspartate or asparagine residue forming two hydrogen bonds to residues ahead in sequence: Their occurrence at alpha-helical N termini and in other situations. *Journal of molecular biology*, **286**, 1633-1649.
22. Wan, W.Y. and Milner-White, E.J. (1999) A recurring two-hydrogen-bond motif incorporating a serine or

threonine residue is found both at alpha-helical N termini and in other situations. *Journal of molecular biology*, **286**, 1651-1662.
23. Yusufzai, T.M. and Wolffe, A.P. (2000) Functional consequences of Rett syndrome mutations on human MeCP2. *Nucleic acids research*, **28**, 4172-4179.
24. Agarwal, N., Becker, A., Jost, K.L., Haase, S., Thakur, B.K., Brero, A., Hardt, T., Kudo, S., Leonhardt, H. and Cardoso, M.C. (2011) MeCP2 Rett mutations affect large scale chromatin organization. *Human molecular genetics*, **20**, 4187-4195.
25. Kudo, S., Nomura, Y., Segawa, M., Fujita, N., Nakao, M., Schanen, C. and Tamura, M. (2003) Heterogeneity in residual function of MeCP2 carrying missense mutations in the methyl CpG binding domain. *J Med Genet*, **40**, 487-493.
26. Goffin, D., Allen, M., Zhang, L., Amorim, M., Wang, I.T.J., Reyes, A.R.S., Mercado-Berton, A., Ong, C., Cohen, S., Hu, L.D. *et al*. (2012) Rett syndrome mutation MeCP2 T158A disrupts DNA binding, protein stability and ERP responses. *Nat Neurosci*, **15**, 274-283.
27. Hansen, J.C., Wexler, B.B., Rogers, D.J., Hite, K.C., Panchenko ,T., Ajith, S. and Black, B.E. (2011) DNA binding restricts the intrinsic conformational flexibility of methyl CpG binding protein 2 (MeCP2). *The Journal of biological chemistry*, **286**, 18938-18948.
28. Kucukkal, T.G., Yang, Y., Uvarov, O., Cao, W.G. and Alexov, E. (2015) Impact of Rett Syndrome Mutations on MeCP2 MBD Stability. *Biochemistry*, **54**, 6357-6368.
29. Yang, Y., Kucukkal, T.G., Li, J., Alexov, E. and Cao, W.G. (2016) Binding Analysis of Methyl-CpG Binding Domain of MeCP2 and Rett Syndrome Mutations. *Acs Chem Biol*, **11**, 2706-2715.
30. Bergo, A., Strollo, M., Gai, M., Barbiero, I., Stefanelli, G., Sertic, S., Gigli, C.C., Di Cunto, F., Kilstrup-Nielsen, C. and Landsberger, N. (2015) Methyl-CpG Binding Protein 2 (MeCP2) Localizes at the Centrosome and Is Required for Proper Mitotic Spindle Organization. *Journal of Biological Chemistry*, **290**, 3223-3237.
31. Dephoure, N., Zhou, C., Villen, J., Beausoleil, S.A., Bakalarski, C.E., Elledge, S.J. and Gygi, S.P. (2008) A quantitative atlas of mitotic phosphorylation. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 10762-10767.
32. Inui, K., Akagi, M., Ono, J., Tsukamoto, H., Shimono, K., Mano, T., Imai, K., Yamada, M., Muramatsu, T., Sakai, N. *et al*. (2001) Mutational analysis of MECP2 in Japanese patients with atypical Rett syndrome. *Brain Dev-Jpn*, **23**, 212-215.
33. Case, D.A.B., V.; Berryman, J.T.; Betz, R.M.; Cai, Q.; Cerutti, D.S.; Cheatham III, T.E.; Darden, T.A.; Duke, R.E.; Gohlke, H.; Goetz, A.W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T.S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K.M.; Paesani, F.; Roe, D.R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C.L.; Smith, W.; Swails, J.; Walker, R.C.; Wang, J.; Wolf, R.M.; Wu X. and Kollman P.A. (2014) Amber14 University of California San Francisco.
34. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impery, R.W., Klein,M.L. . (1983) Comparison of simple potential functions for simulating liquid water. . *J. Chem. Phys.*, **79**, 926-935.
35. Cheatham, T.E., Miller, J.L., Fox, T., Darden, T.A. and Kollman, P.A. (1995) Molecular-Dynamics Simulations on Solvated Biomolecular Systems - the Particle Mesh Ewald Method Leads to Stable Trajectories of DNA, Rna, and Proteins. *Journal of the American Chemical Society*, **117**, 4193-4194.
36. Ceriotti, M., Bussi, G. and Parrinello, M. (2009) Langevin Equation with Colored Noise for Constant-Temperature Molecular Dynamics Simulations. *Physical review letters*, **102**.
37. Feller, S.E., Zhang, Y.H., Pastor, R.W. and Brooks, B.R. (1995) Constant-Pressure Molecular-Dynamics Simulation - the Langevin Piston Method. *Journal of Chemical Physics*, **103**, 4613-4621.
38. Ryckaert, J.P., Ciccotti, G. Berendsen, H.J.C. . (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. . *J. Comput. Phys.*, **23**, 327-341.
39. P.A., M.S.a.K. (1992) Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of computational chemistry*, **13**, 952–962.
40. Hess, B., Kutzner, C., van der Spoel, D. Lindahl, E. . (2008) GROMACS 4: algorithm for highly Efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, **4**, 435-447.
41. van Zundert, G.C.P.R., J.P.G.L.M.; Trellet,M.; Schmitz, C.; Kastritis, P.L.; Karaca, E.; Melquiond, A.S.J.; van Dijk, M.; de Vries S.J. and Bonvin. A.M.J.J. (2016) The HADDOCK2.2 webserver: User-friendly integrative modeling of biomolecular complexes. *Journal of molecular biology*, **428**, 720-725.
42. Dominguez, C., Boelens, R. and Bonvin, A.M.J.J. (2003) HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *Journal of the American Chemical Society*, **125**, 1731-1737.
43. Ho, S.Y., Li, X.G., Wakefield, A., Barbarese, E., D'Arrigo, J.S. and Simon, R.H. (1997) The affinity of lipid-coated microbubbles for maturing brain injury sites. *Brain research bulletin*, **43**, 543-549.
44. DePristo, M.A., de Bakker, P.I.W. and Blundell, T.L. (2004) Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography. . *Structure*, **12**, 831-838.

45. Ghosh, R.P., Horowitz-Scherer, R.A., Nikitina, T., Gierasch, L.M. and Woodcock, C.L. (2008) Rett syndrome-causing mutations in human MeCP2 result in diverse structural changes that impact folding and DNA interactions. *Journal of Biological Chemistry*, **283**, 20523-20534.

Figure1

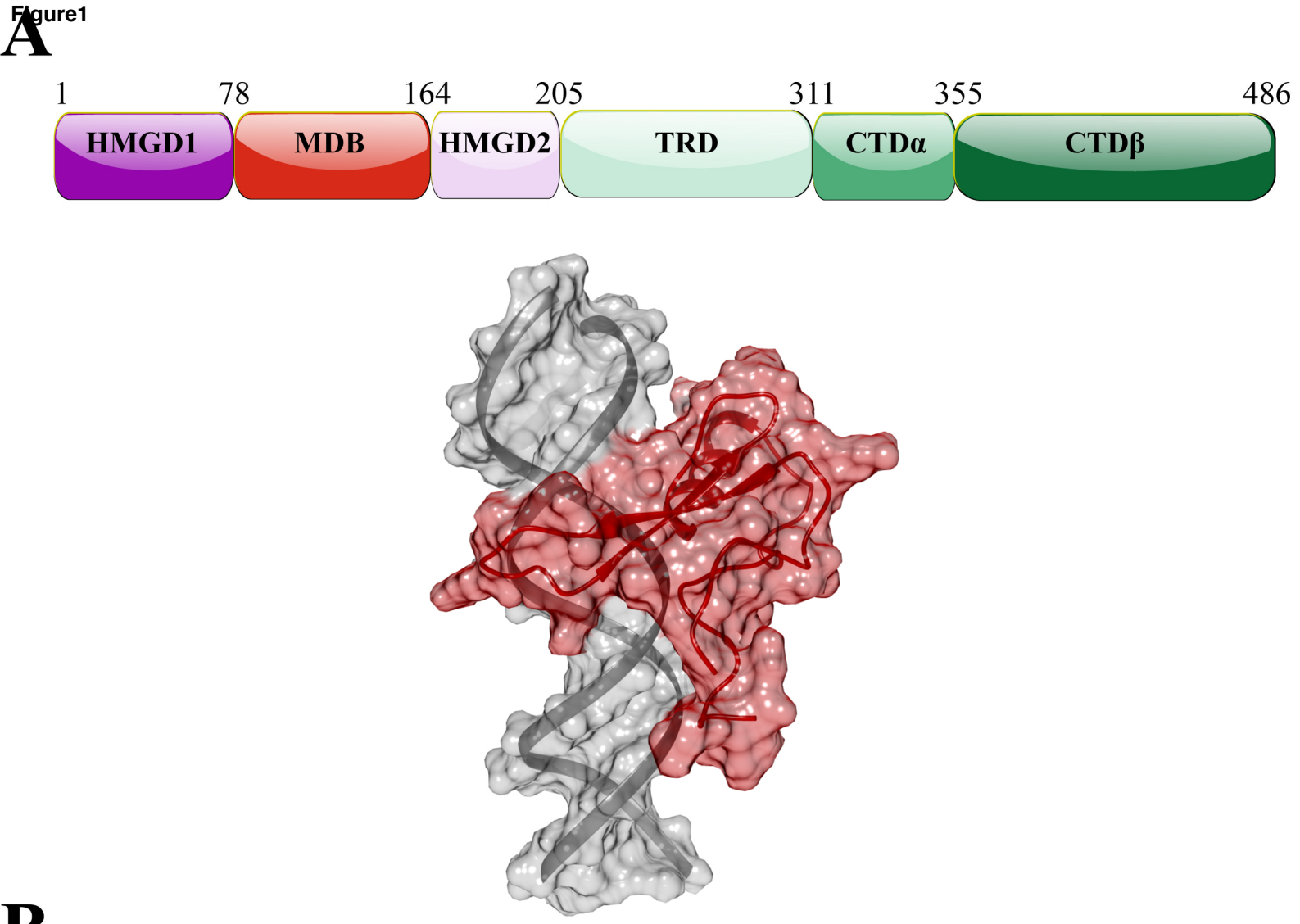
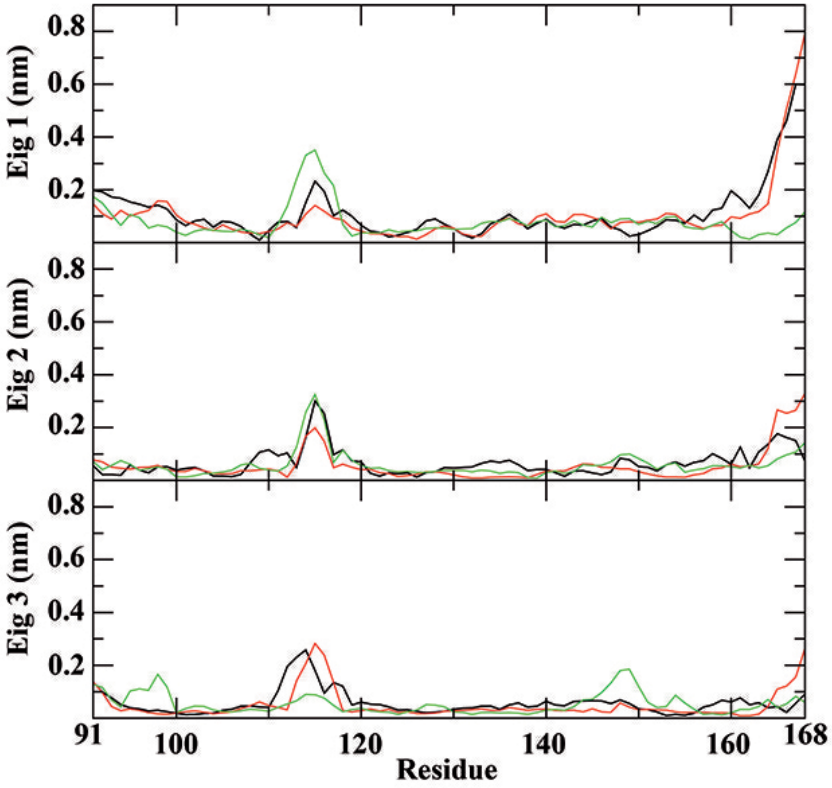

A
1
78
164
205
311
355
486
HMGD1
MDB
HMGD2
TRD
CTDα
CTDβ
B
K112
K109
R133
S113
R115
R111
S116
D121
S134
Y120
V159
K112
K109
R133
S113
R115
R111
S116
Y120
T160
R167
R168

Figure2


Eig 1 (nm)
Eig 2 (nm)
Eig 3 (nm)
0.8
0.6
0.4
0.2
91
100
120
140
160
168
Residue

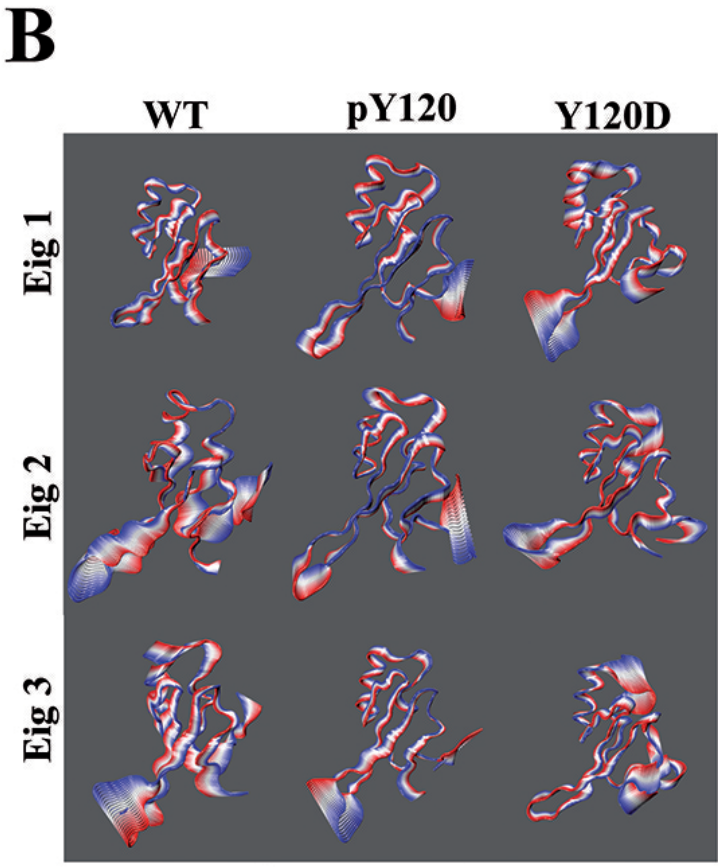
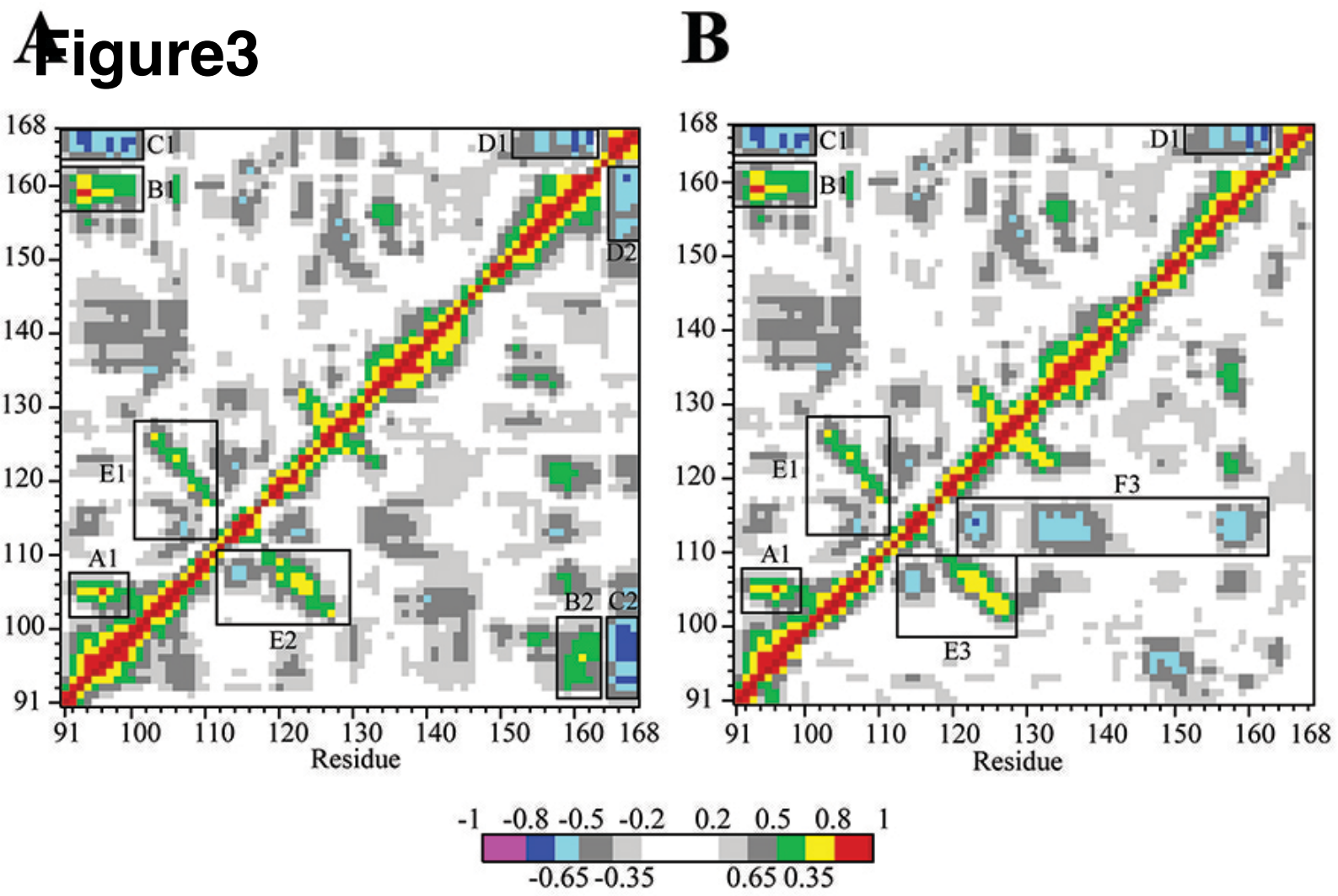

B
WT
pY120
Y120D
Eig 1
Eig 2
Eig 3

Figure3


A
B
C1
B1
D1
D2
E1
A1
E2
B2
C2
F3
E3
168
160
150
140
130
120
110
100
91
Residue
-1 -0.8 -0.5 -0.2 0.2 0.5 0.8 1
-0.65 -0.35 0.65 0.35

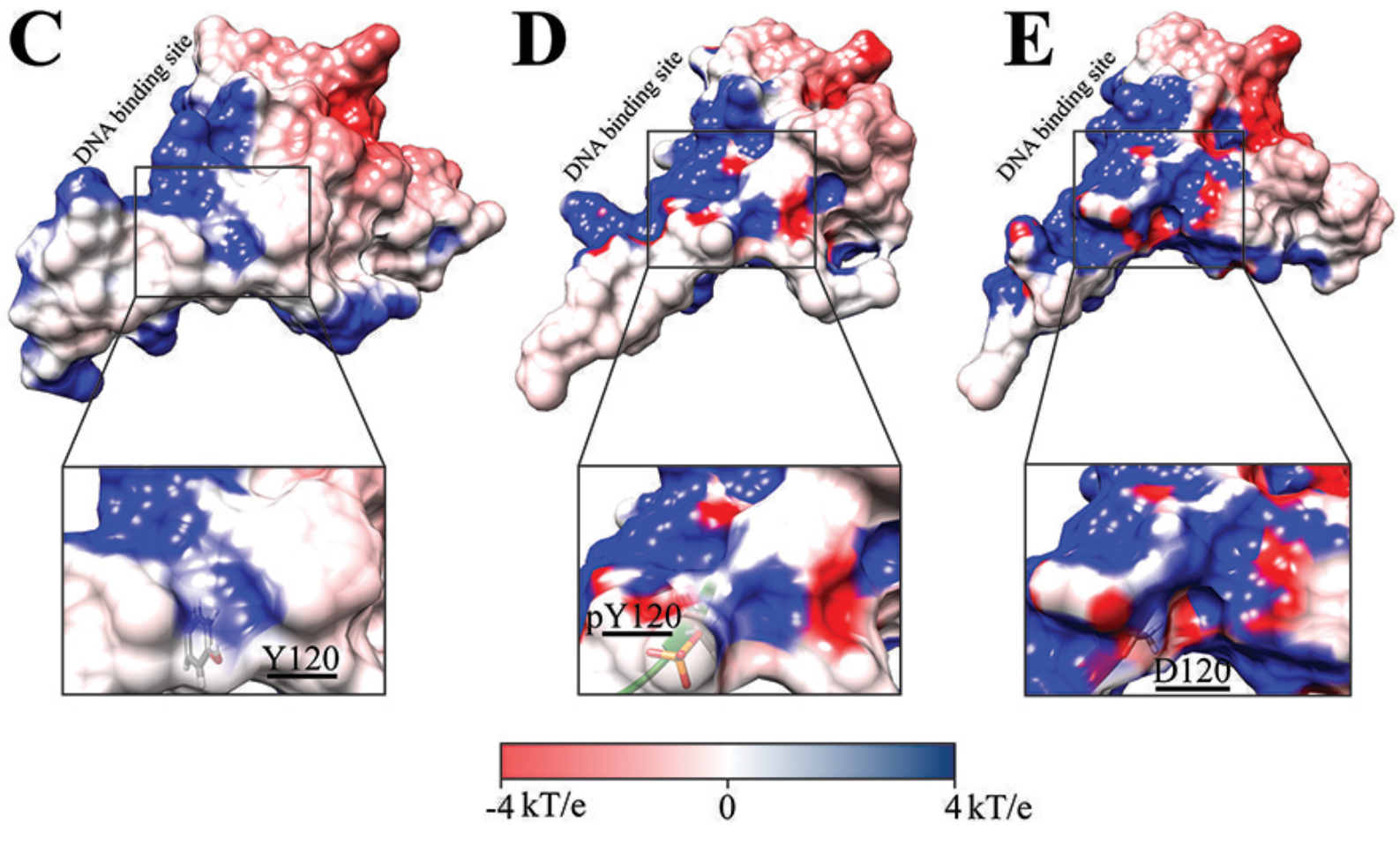

C
D
E
DNA binding site
Y120
pY120
D120
-4 kT/e
0
4 kT/e

Figure4

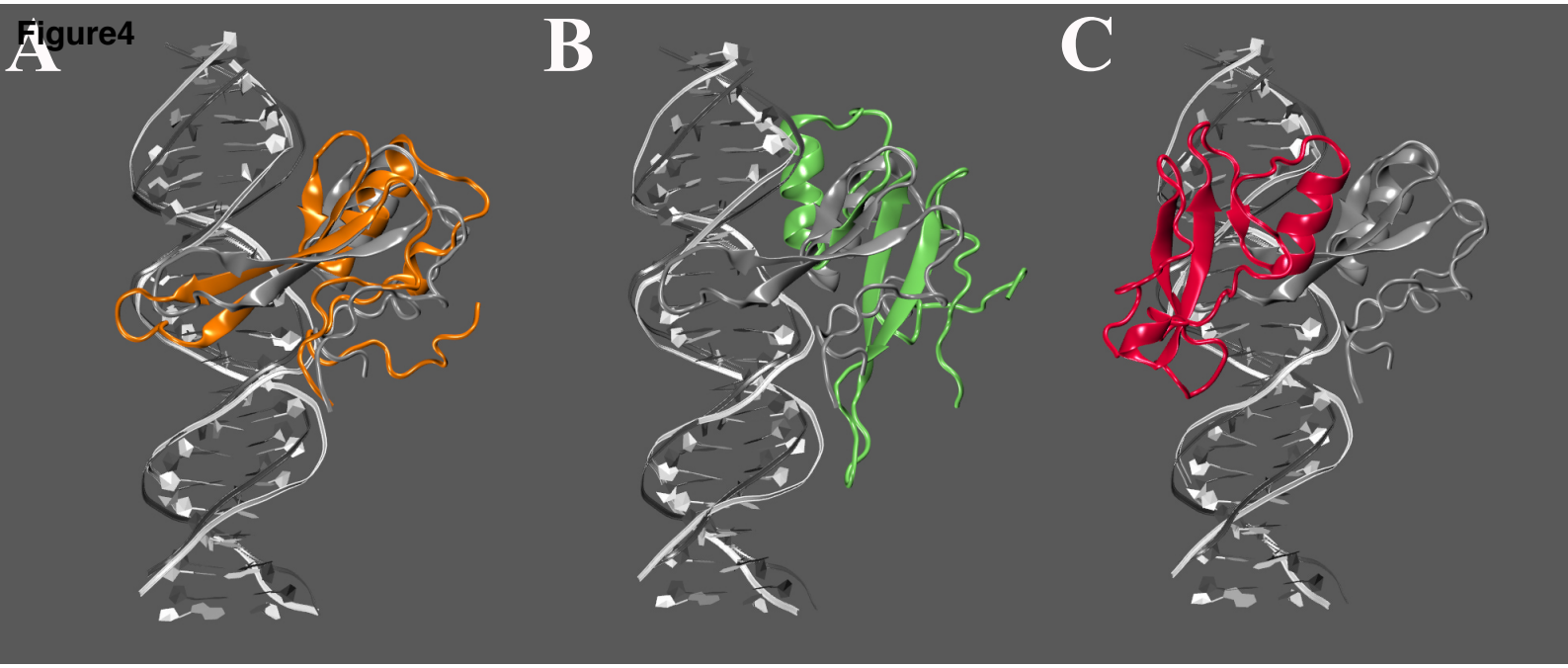
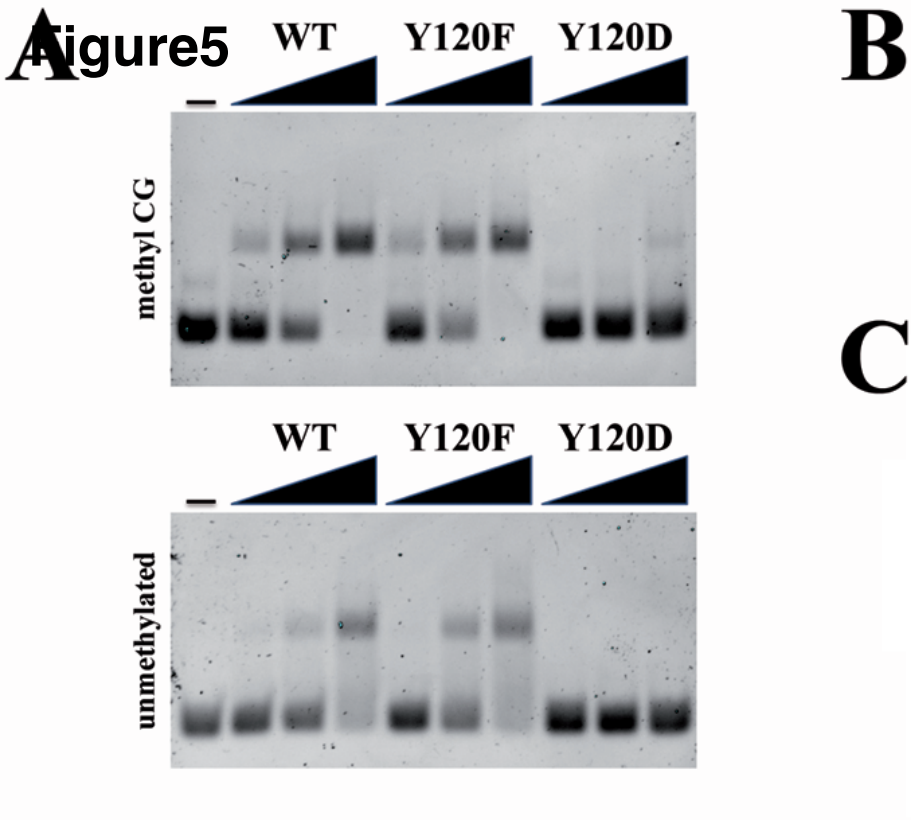


Figure5


A
WT
Y120F
Y120D
methyl CG
WT
Y120F
Y120D
unmethylated
B
WT
Y120F
Y120D
C
*
K_d(μM)
0
1
2
3
4
WT
Y120F
Y120D
WT
Y120F
Y120D
methyl CG
unmethylated

Protein-DNA Hydrogen Bonds	
X-Ray	**MD**
Lys109-Ade7(+)	81.35
Arg111-Gua9(+)	96.99
Lys112-mCyt8(+)	91.25
Ser113-Cyt8(+)	65.14
Arg115-Gua9(+)	95.53
Arg115-Gua10(+)	95.59
Ser116-Gua9(+)	94.14
Asp121-mCyt8(+)	**-**
Arg133-Gua14(-)	55.50
Ser134-mCyt13(-)	**-**
Val159-Thy11(-)	**-**
Thr160-Thy11(-)	**99.11**
Arg167-Thy16(+)	**85.64**
Arg168-Thy16(+)	**92.43**

Supplementary Table T1. Hydrogen Bonds present in the MeCP2 MBD-DNA X-Ray structure and in the MD trajectory in the last 230 ns of simulation. The interactions present only in the X-ray or in the trajectory are highlighted in bold.

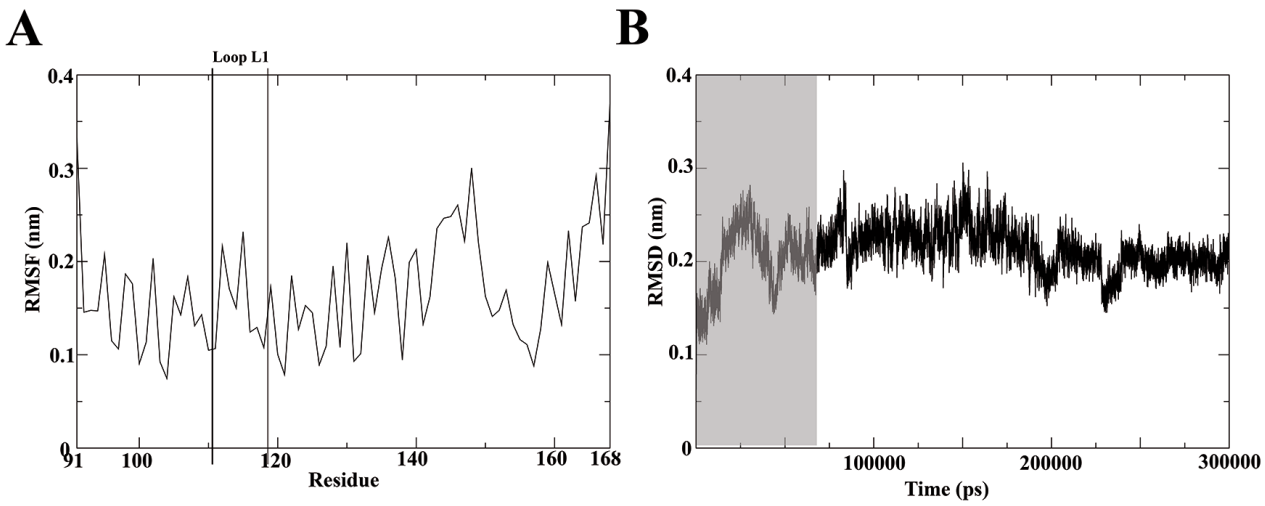


Figure S1. (A) Per-residue RMSF of the MeCP2 MBD domain in complex with DNA. The loop L1, important for the interaction with the nucleic acid, is highlighted. **(B)** RMSD of the Cα atoms. The grey box underlines the equilibration time after which the protein reaches a plateau.

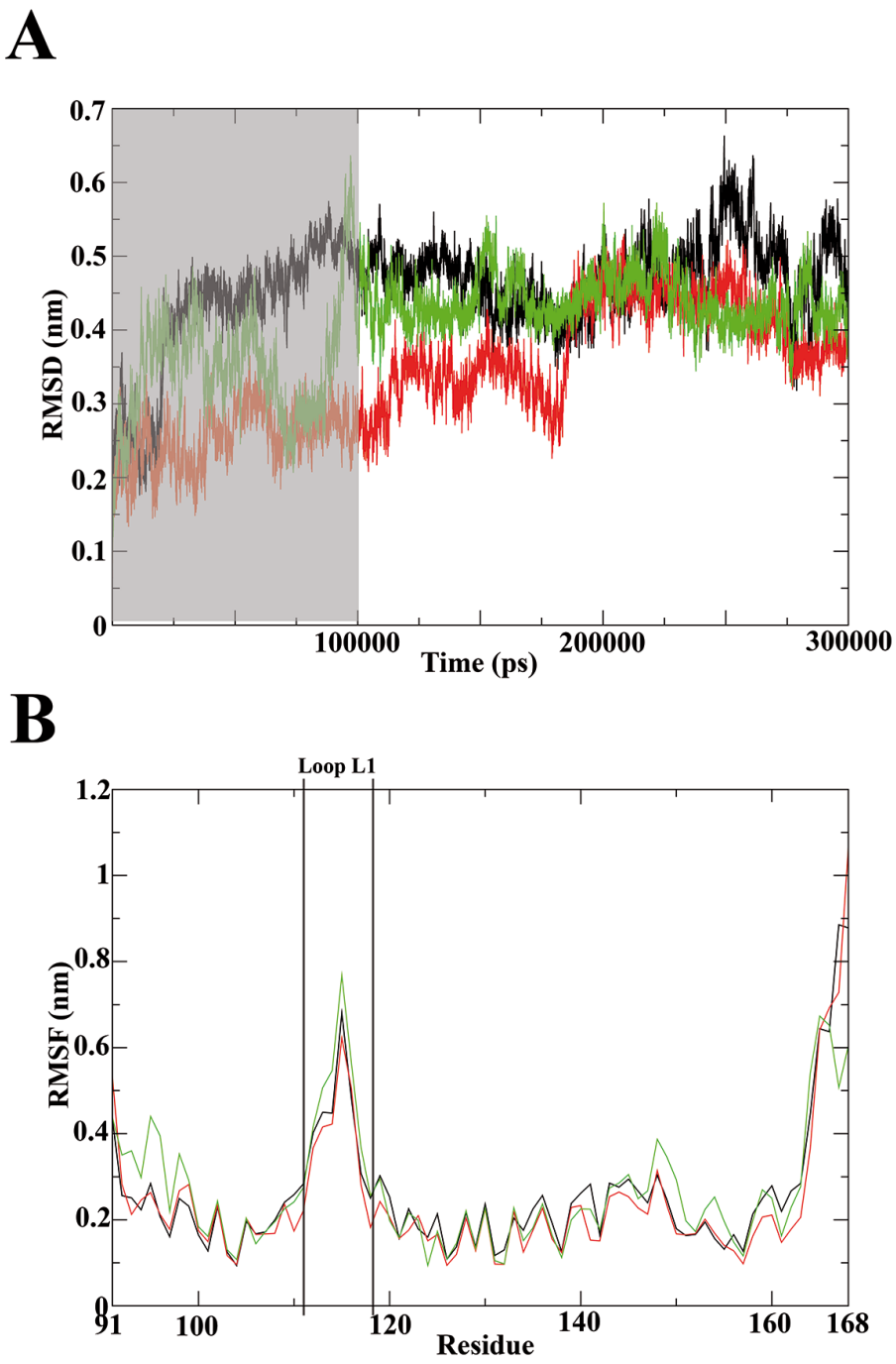


Figure S2. (A) RMSD of the Cα atoms for the WT (black line), pY120 (red line) and Y120D (green line) domains. The grey box underlines the equilibration time avoided to carry out further analyses. **(B)** Per-residue RMSF of the WT (black line), pY120 (red line) and Y120D (green line) domains. The loop L1, important for the interaction with the nucleic acid, is highlighted.

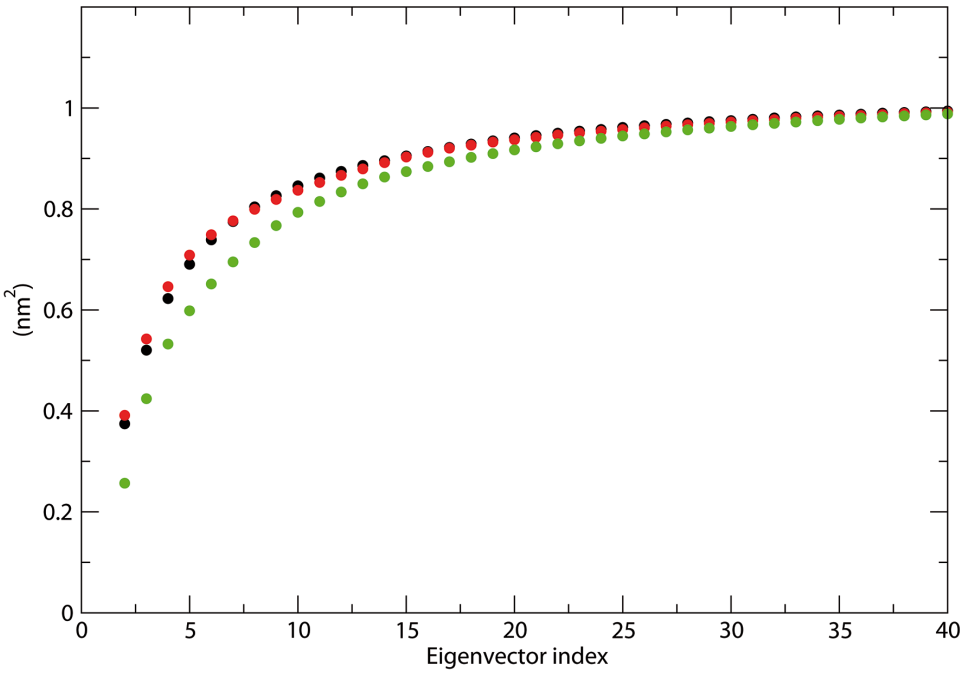


Figure S3. Cumulative percentage of the description of the motion by the first 40 eigenvectors in the three systems: WT=black dots, pY120=red dots, Y120D=green dots.

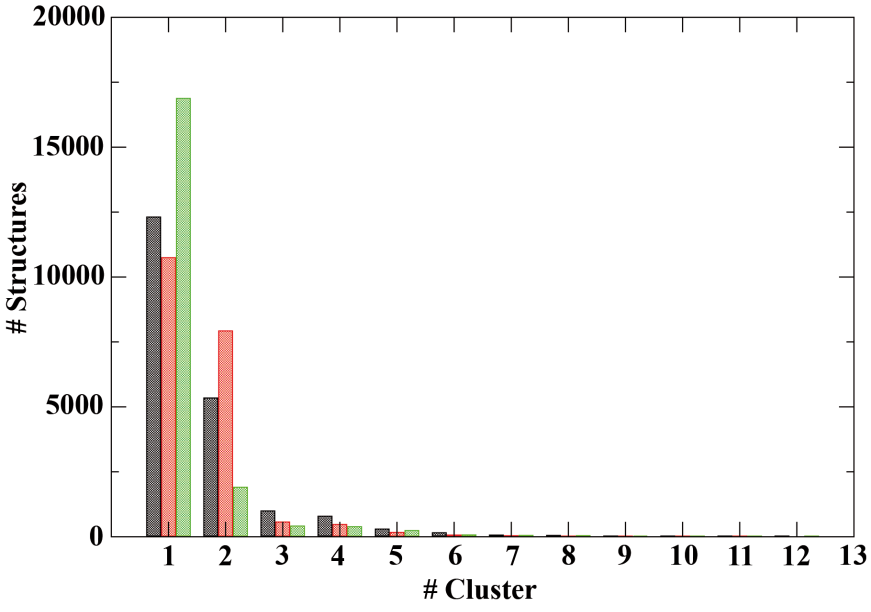


Figure S4. Numbers of structures represented in each cluster. WT=black bar, pY120=red bar, Y120D=green bar.

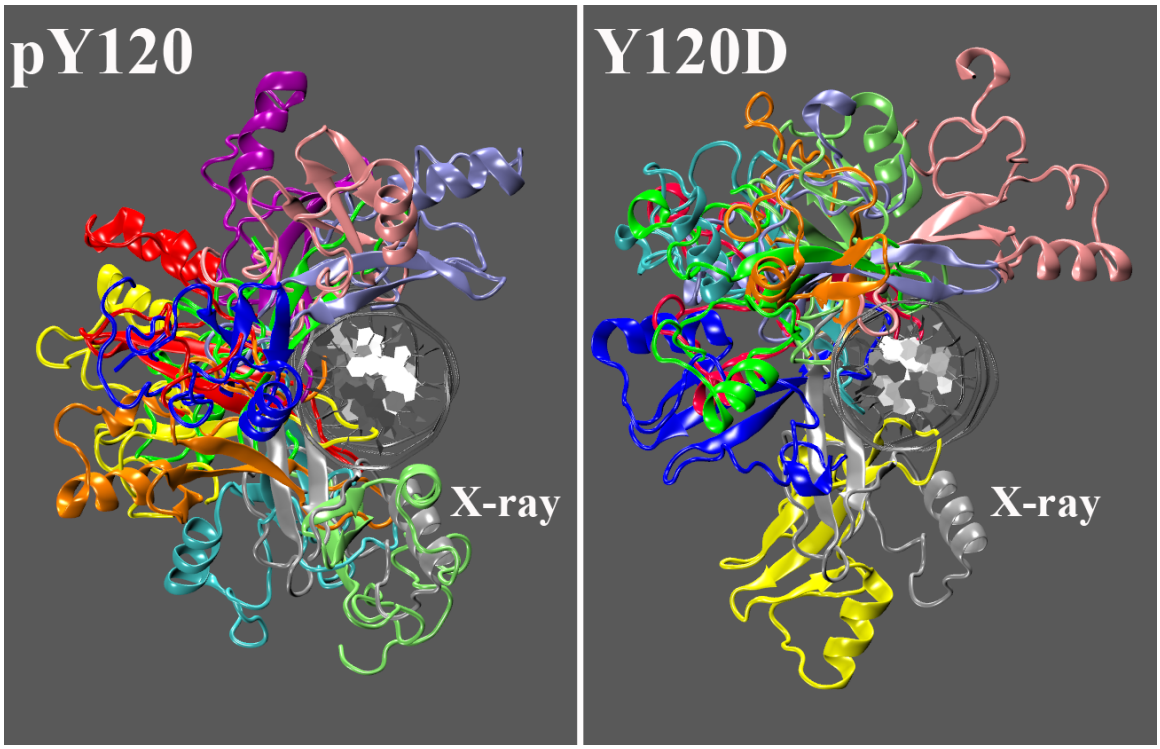


Figure S5. Upper view of the superposition on the DNA atoms between the MeCP2 MBD-DNA X-ray structure and all the best pose of each family of docked structures obtained for pY120 **(A)** and Y120D **(B).** The domain from the X-ray complex is depicted in grey in both panels. For the sake of clearness only the DNA from the X-ray is shown.