

SUPPLEMENTARY MATERIALS

Specificity of the Mbp1-DNA Interactions: a spectroscopic and modeling study.

Anna Chernatynskaya^{1,2}, Lynn Deleeuw^{1,2}, John O. Trent^{1,2}, Tom Brown³ and Andrew N. Lane^{1,2}

DNA structure calculations by molecular dynamics (MD) and energy minimization methods.

Restrained Molecular Dynamics Simulations

To evaluate the role of the force field in the structure determination, we carried out calculations using two different forcefields, namely the AMBER forcefield in the Discover package (Accelrys, San Diego) and the Charm forcefield in the XPLOR-NIH package.

Restrained molecular dynamics were performed using the DISCOVER module of Insight II. The simulations were made using an AMBER-derived force field and the dielectric constant $\epsilon=4r_{ij}$ to simulate the effects of electrostatics⁵⁷. There were no limits on the non-bond interactions. Structure calculations were performed in three-step procedure using a protocol modified from⁵⁸. First, an unrestrained energy minimization was used (100 steps steepest descent followed by of 2000 steps conjugate gradient minimization). Molecular dynamics were then carried out for 10 ps at 3500 K with low weight of the non-bonded and Coulombic parameters. NOE restraints and hydrogen bond restraints were used with zero lower bounds for hydrogen bonding interactions, which allows atoms to move through each other in order to satisfy constraints, followed by 25.5 ps of MD with normal lower bounds for the hydrogen bonds using normal force constant of the non-bonded and electrostatic interactions. The distance restraint force constant was $50 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ during all calculations. The temperature was gradually cooled to 300 K using 100 K temperature step. The second step of the protocol used in calculating structures consists of initial steepest descent minimization for 100 steps and 200 steps of conjugate gradient minimization using NOE and hydrogen bond restraints. Next, conjugate gradient minimization (10000 steps) followed by MD at 3500 K by applying NOE, hydrogen bond and torsion restraints including sugar angles δ, ν_1 and glycosidic angle χ with force constants equal. During the second stage of refinement non-bonded interactions and Coulombic terms had low force constants. The torsion restraint force constant was gradually increased from 50 to a final value of $500 \text{ kcal mol}^{-1} \text{ rad}^{-2}$ over 10 cycles of MD runs, corresponding to a total of 25 ps. This was followed by minimization (10000 steps) and MD with the presence of all experimental restraints (NOE, hydrogen bonds, all torsion angles). Temperature was gradually cooled to 300 K in a total of 31.5 ps (50 K temperature step) followed by equilibration for 20 ps at 300 K. The final step of the protocol included refinement of structures at 1000 K for 8 ps using normal force constant for non-bonded and Coulombic parameters. Again, the temperature was gradually cooled to 300 K in a total of 28 ps with 50 K temperature step followed by final conjugate gradient minimization for 10000 iterations. In the event of any constraint

violation, other rounds of dynamics were performed by varying the time step size and time of MD. This procedure was repeated several times, until well-converged structures were obtained with close to zero violations. The total of 136 structures were obtained, from which the 12 lowest energy structures were chosen.

When A-form and B-form DNA were used as the starting structures for MD runs, an annealing protocol was similar to protocol for refinement starting from random structure. Comparing protocols there were some differences such as MD calculations have been run at lower temperatures, 1500 K for A-form DNA and 1000 K for B-form DNA and convergence of structures was obtained at shorter MD time. The distance and torsion angle force constant restraints for both forms of DNA were equal to 50 kcal mol⁻¹ Å⁻² and 50 kcal mol⁻¹ rad⁻² correspondently.

Restrained Molecular Dynamics Simulations using XPLOR-NIH.

Structures that satisfy all of the measured constraints to within experimental error were calculated by restrained molecular dynamics that includes the residual dipolar couplings. New internal variable module within XPLOR-NIH was used to perform torsion angle molecular dynamics. Lennard-Jones and electrostatic energy terms in different combinations were included at different stages of refinement. Switched van der Waals function in combination with a switched electrostatic function and a $1/r_{ij}$ dielectric function were selected for non-bond energy terms. A $1/r_{ij}$ dependent dielectric constant introduces an approximate solvent screening term by setting the constant $\epsilon_{r_{ij}}$. The switched distances are 9.5 Å and 10.5 Å.

Structure calculations in XPLOR-NIH were performed in a similar way, in three-step procedure as in DISCOVER plus refinement step with dipolar couplings. In the first step, structures were minimized for 1000 steps with only covalent energy terms (bonds, angles and improper torsions) followed by refined during 20 ps at high temperature, 3500 K, with low nonbond repel term for van der Waals interactions and zero lower bounds for hydrogen bonds. The force constant for distance restraints was 20 kcal mol⁻¹ Å⁻² and force constant for phosphate backbone torsion angle restrains was very low 5 kcal mol⁻¹ rad⁻². Another round of simulating annealing was done (20 ps) with addition of distance restraints followed by next 20 ps refinement with normal lower bounds for hydrogen bonding interactions. At this stage the soft-square function and biharmonic function were used for distance and hydrogen bonding restraints correspondently. Next, the van der Waals force constant was increased during 100 rounds of 3 ps MD followed by gradually cooling temperature to 300 K with 25 K temperature step. During cooling cycles soft-square function of distance restraints was replaced by square-well function and force

constants of distance and torsion angle restraints were increased to $50 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ and $200 \text{ kcal mol}^{-1} \text{ rad}^{-2}$ correspondently. Structures were minimized using conjugate gradient method (Powell minimization) for 500 steps with the non-bond repel term replaced by Lennard-Jones potential with electrostatics. In the second step, the globally folded structures from the first part were undergo refinement with all torsion angles by slowly increasing torsion angle restraints force constant from 10 to $200 \text{ kcal mol}^{-1} \text{ rad}^{-2}$ during 60 ps at 3500 K, followed by cooling temperature to 300 K and final minimization for 1000 steps using Lennard-Jones and electrostatic potentials.

Residual dipolar couplings

To incorporate dipolar couplings into structure calculations the axial and rhombic component of the alignment tensor need to be specified. The grid search method was used to determine optimal values D_a and R, where structures were calculated for each grid point. The best 5 structures from refinement stage with NOE distance and torsion angle restraints were used as initial structures. The D_a parameter was varied from -30 to -4 with 1Hz step while R was varied from 0 to 0.6 with 0.05 step. The grid search procedure consists of 1000 steps of rigid body minimization for each combination of D_a and R. Next, initially determined values of D_a and R were used for refinement with dipolar couplings. After the first cycle of refinement with dipolar couplings, again the 5 best structures were chosen to repeat the grid search. There are five rounds of refinement followed by grid search was made. Then D_a was varied from -23 to -17 (0.2 step size) and R was varied from 0 to 0.25 (0.01 step size). Another two rounds were made. Final grid search of parameters was from -25 to 19 for D_a with 0.2 step size and from 0 to 0.25 for R with 0.01 step size followed by refinement with dipolar couplings.

All structures during refinement with dipolar couplings were calculated using the values of $D_a = -22.4$ and $R = 0.17$. The structures generated with NOEs and torsion angles were starting structures for dipolar coupling refinement. The harmonic potential function was used as the dipolar coupling energy term. The force constant for dipolar coupling restraints was $0.2 \text{ kcal mol}^{-1} \text{ Hz}^{-2}$ for class 1 dipolar couplings and $0.04 \text{ kcal mol}^{-1} \text{ Hz}^{-2}$ for class 2. These constants were chosen so that the dipolar rms equal to the error in the measured residual dipolar couplings. The refinement procedure includes initial conjugate gradient minimization of structures (50 steps) with only covalent energy terms turned on, followed by 10 ps MD at 400 K. Next, during cooling step, temperature was gradually decreased to 300 K while the dipolar coupling force constant was gradually increased from $0.01 \text{ kcal mol}^{-1} \text{ Hz}^{-2}$ to their final values. The second round of refinement consists of 10 ps MD at 400 K, followed by reduction of temperature to 300 K (10 K temperature step) while dipolar coupling constants were kept at their final

values. The cooling step was followed by long dynamics at 300 K (25 ps). Then, structures were minimized for 1000 steps using conjugate gradient minimization, with Lennard-Jones and electrostatic potentials. During these calculations, the remaining (distance and torsion angle) restraints were maintained with the force constants set to their final values.

The final step, molecular dynamics with van der Waals and electrostatic forces was applied to the structures generated with NOEs, torsion angle and dipolar couplings. This last step includes 25 ps MD at 300 K, followed by conjugate gradient minimization for 1000 steps.

The final structures from XPLOR-NIH were accepted based on following criteria: a large, negative potential energy, good stereochemistry with no NOE violations and torsion angles exceeding 0.2 Å and 3° correspondently, and with no dipolar coupling violations in excess of 5 Hz. The criteria for structures from DISCOVER module (INSIGHTII) was similar, except no torsion angle violations be greater than 2° and a low (<10 kcal mol⁻¹) residual restraint energy. Structures were analyzed using INSIGHTII. The rms differences between coordinates of each final structure were performed in XPLOR-NIH.

Modeling of the protein-DNA complex

The structure calculations were done using Amber 8⁵⁹. The first phase of the modeling started with the three structures from a set of 19 structures in the PDB under accession number 1L3G, one structure of Mbp1 from MD in explicit solvent and lowest d(CTTACGCGTCATTG).d(CAATGACGCGTAAG) energy structure from XPLOR-NIH. The atomic coordinates of protein and DNA were incorporated in the GRAMM program (version 1.03) to obtain the orientation and position of protein relative to the DNA. The exhaustive six-dimensional search was performed through the translations and rotations of protein relative to the DNA with a step size of 12°. The resulting 400 models were examined and chosen for the next phase based on following criteria: helix B of protein is near or in the major groove of DNA, close to the binding site 5'- (ACGCGTCA)-3' and the wing region of the protein is close to minor groove of the DNA, with overall low energy of the system compare to others. In the next phase, solvation parameters were added to the four resulting protein-DNA systems using LEaP program of Amber. These solvated systems undergo two-step approach: equilibration and free molecular dynamics simulations in explicit solvent. Equilibration period includes minimization of solvated system with solute fixed using conjugate gradient minimization of 5000 steps, followed by 10 ps MD at temperature of 100 K. Next, two minimization were performed for 5000 steps each with and without position restraints. The position restraints were fixed using 100

kcal/mol \AA^{-2} force constant. The same MD was performed for 10 ps at temperature of 100 K using position restraints. Last part of equilibration period includes several steps of MD at 300 K for 10 ps each with solute fixed using 100, 50, 10 and 1 kcal/mol \AA^{-2} force constant. The production MD run was performed at 300 K for 5 ns for each system.

At this point the models did not satisfy experimental data and they were used as initial models for the next stage of modeling. The coordinate files of dry system were used to run MD using distance restraints in the implicit solvent (using generalized Born model) in Sander program of Amber 8. Leap was used to create solvation parameters of implicit solvent. First, an initial steepest descent minimization of 250 steps followed by conjugate gradient minimization of 750 steps was performed on the three systems. No restraint forces were applied during this time. During MD simulations protein and DNA were fixed, using position restraints with force constant equal to 500 kcal/mol \AA^{-2} , except wing (residues 58-79) and C-terminal region (residues 102-124 residues). At the same time, C-terminal of the protein was enforced with artificial distance restraints to be in the closer contact with DNA. The force constant for the intermolecular and intramolecular distance restraints was equal to 50 and 100 kcal/mol \AA^{-2} correspondently. Intramolecular restraints include hydrogen bond restraints for each base pair in duplex, using standard distance values from Saenger⁶⁰. Other intramolecular distance restraints are NOE restraints between core (residues 5, 6, 7, 8) and C-terminal region (residues 102, 113 and 114) of the protein with the range 1.8-4 \AA . Intermolecular distance were applied between Lys116, Lys 122 and His 117 and closest 2-4 phosphate of DNA. Distance restraints were applied for 100 ps MD simulations which were performed in several steps: during initial 3 ps temperature was 300 K and the force constant of inter- and intramolecular distance restraints were increased from 0 to the corresponding final value and it was used during all calculations. During final 2 ps temperature was decreased to 0 K.

The next phase included random Monte Carlo calculations on C-terminal region of the protein to determine possible orientations of it relative to the DNA. The dry coordinate files of the three models were transferred to the MacroModel, Interactive Molecular Modeling System version 6 (Department of Chemistry, Columbia University, 1997). The protein and DNA were fixed, while random rotation of bonds $\text{C}\alpha\text{-N}$, $\text{CO-C}\alpha$ and $\text{C}\alpha\text{-C}\beta$ of C-terminal residues from Lys116 to Asp124 was performed. For each model of protein-DNA complex 150 structures was obtained. The lowest energy conformations were minimized in Sander with steepest descent minimization for 1000 steps followed by conjugate gradient minimization for 5000 steps in vacuum with 9 \AA cutoff for non-bond

forces. The models were examined for the best conformation of protein relative to the DNA using the same criteria as for structures of the complex after initial docking using GRAMM program (see above). Additional analysis was done for examination of interaction between protein C-terminal of the protein and phosphodiester bond of DNA.

The last phase of modeling was carried out using MD simulations for one of the models of protein-DNA complex in the explicit solvent. For this purpose the model was solvated again in the LEaP program. MD simulations in the explicit solvent were performed in the Sander using the same approach, containing equilibration period and free MD run. This procedure was described previously in great detail. In addition to position restraints, hydrogen bond restraints for the terminal base pairs and the same NOE restraints between the core (residues 5-8) and the tail (residues 102, 113, 114) of the protein were applied. The production run of MD simulations was not enforced by any distance restraints. The average structure of the Mbp1-DNA complex module was calculated for the last 10 ps using the Carnal utility in Amber 8.

Table S1. The Average Backbone Torsion Angles α - ζ , and Glycosyl Angles χ for the Insight II Structure of d(CTTACGCGTCATTG)•d(CAATGACGCGTAAG).

residue	α	β	γ	δ	ϵ	Z	X
C1	0	0	145	151	-178	-114	-116
T2	-70	148	66	126	-179	-97	-117
T3	-68	173	60	123	-177	-91	-121
A4	-71	179	54	121	-181	-96	-110
C5	-70	173	59	121	-175	-85	-122
G6	-78	172	98	155	-170	-173	-122
C7	-89	143	84	129	-177	-92	-115
G8	-75	141	54	126	-182	-96	-112
T9	-67	176	60	120	-177	-98	-127
C10	-71	176	59	116	-179	-97	-125
A11	-71	145	55	132	-179	-103	-112
T12	-69	170	64	108	-169	-93	-130
T13	-72	172	60	133	-142	-90	-108
G14	-72	171	67	112	0	0	-133

Table S2. Structural Refinement Statistics

Xplor-NIH structure	Final dnaa	Final lj dnab	Insight II structure	
RMSD from restraints				
NOE distances (Å)(346)	0.024	0.024	NOE distances (Å)	0.039
torsion angles (deg)(188)	0.65	0.65	torsion angles (deg)	0.26
RMSD from dipolar couplings(Hz)				
D _{CH} ribose(9) ^c	0.92	0.92	D _{CH} ribose (0)	
D _{CH} ribose(22) ^d	3.15	3.15	D _{CH} ribose (0)	
D _{CH} base(11) ^c	0.76	0.76	D _{CH} base (0)	
D _{CH} base(14) ^d	1.96	1.96	D _{CH} base (0)	
Number of violations				
NOE (>0.2Å)	0	0	NOE (>0.2Å)	0.2
torsion angles (>3deg)	1.17	1.67	torsion angles (>2deg)	0
dipolar couplings (>5Hz)	4.08	3.92		
Deviation from idealized covalent geometry				
Bonds (Å)	0.0027	0.0049	bonds (Å)	0.00028
Angles (deg)	0.72	0.72		
impropers(deg)	0.33	0.34		
Energies (kcal mol⁻¹)				
total	-277.18	-314.77	total	-170.85
Bonds	6.75	7.25	Bonds	9.63
Angles	135.61	136.86	angles theta	112.34
Impropers	11.72	12.43	angles phi	263.21
Electrostatic	-101.12	-104.97	Coulomb	-218.42
VdW	-363.10	-395.66	nonbond	-318.70
NOE	10.56	10.35	restrain force	5.04
Torsions	4.81	4.88	hydrogen bond	-19.13
dipolar couplings	17.57	14.24		

^a The final structures obtained from rMD using only Repel function of Lennard-Jones potential.

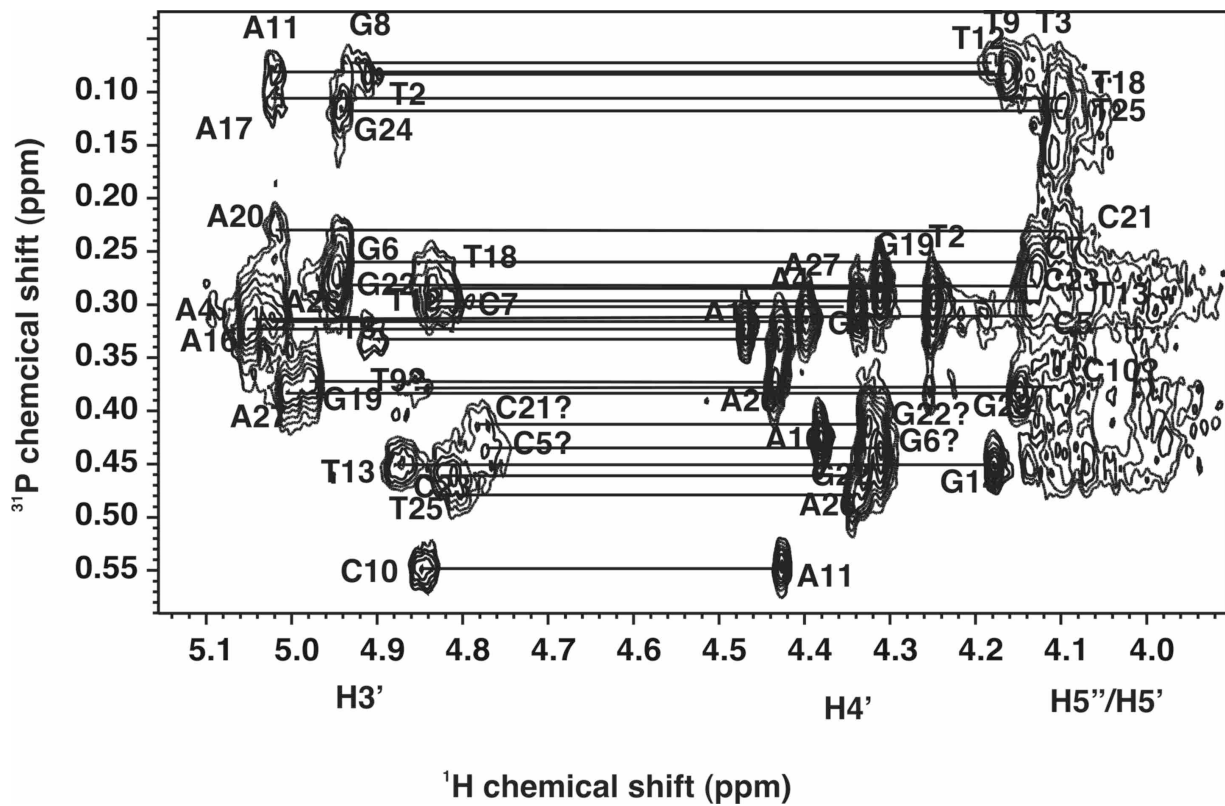
^b The final structures obtained from rMD using Lennard-Jones and electrostatic potential.

^c Class 1 of dipolar coupling with estimated errors less than 2 Hz.

^d Class 2 of dipolar couplings with estimated errors less than 4 Hz.

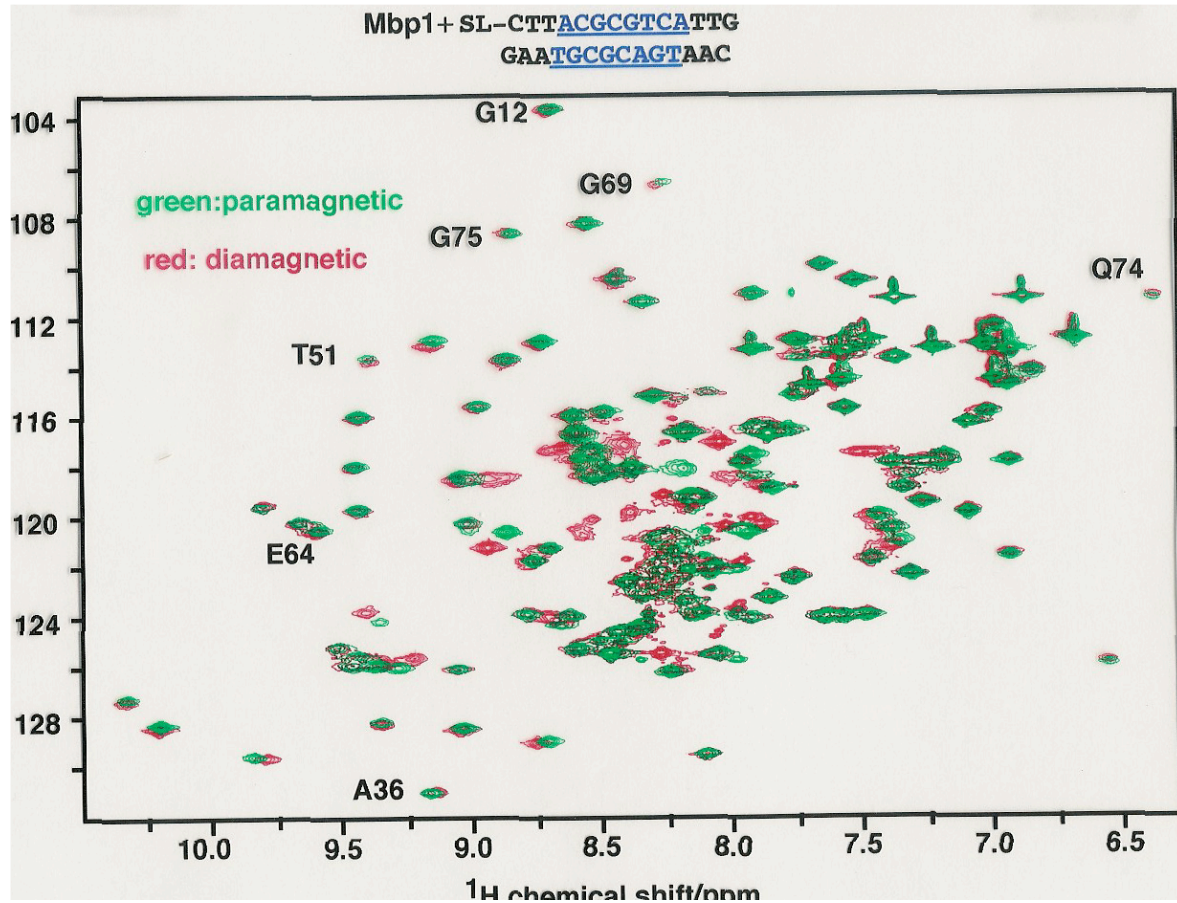
^e The number in parentheses corresponds to the number of restraints.

Figure S1.



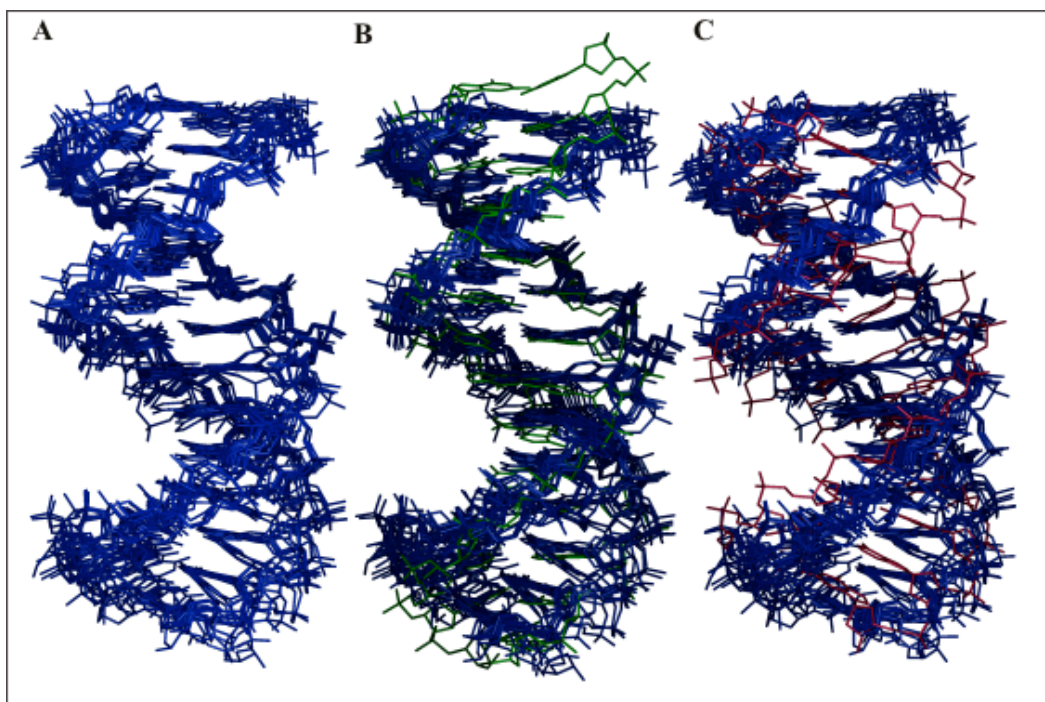
Legend to Figure S1. ^{31}P - ^1H HSQC spectrum of the DNA 14-mer. The spectrum was recorded at 18.8 T, 30 °C. Most of the H3'-P cross-peaks were assigned, and also the purine H4'-P cross peaks assignments as shown.

Figure S2



Legend to Figure S2. Paramagnetic perturbation with spin-labeled DNA. ^{15}N labeled Mbp1 was mixed with DNA labeled 5'- TEMPO in the upper (d(CTTACGCGTCATTG)) strand as described in the text. Green: paramagnetic state; Red: diamagnetic state after reduction with sodium dithionite.

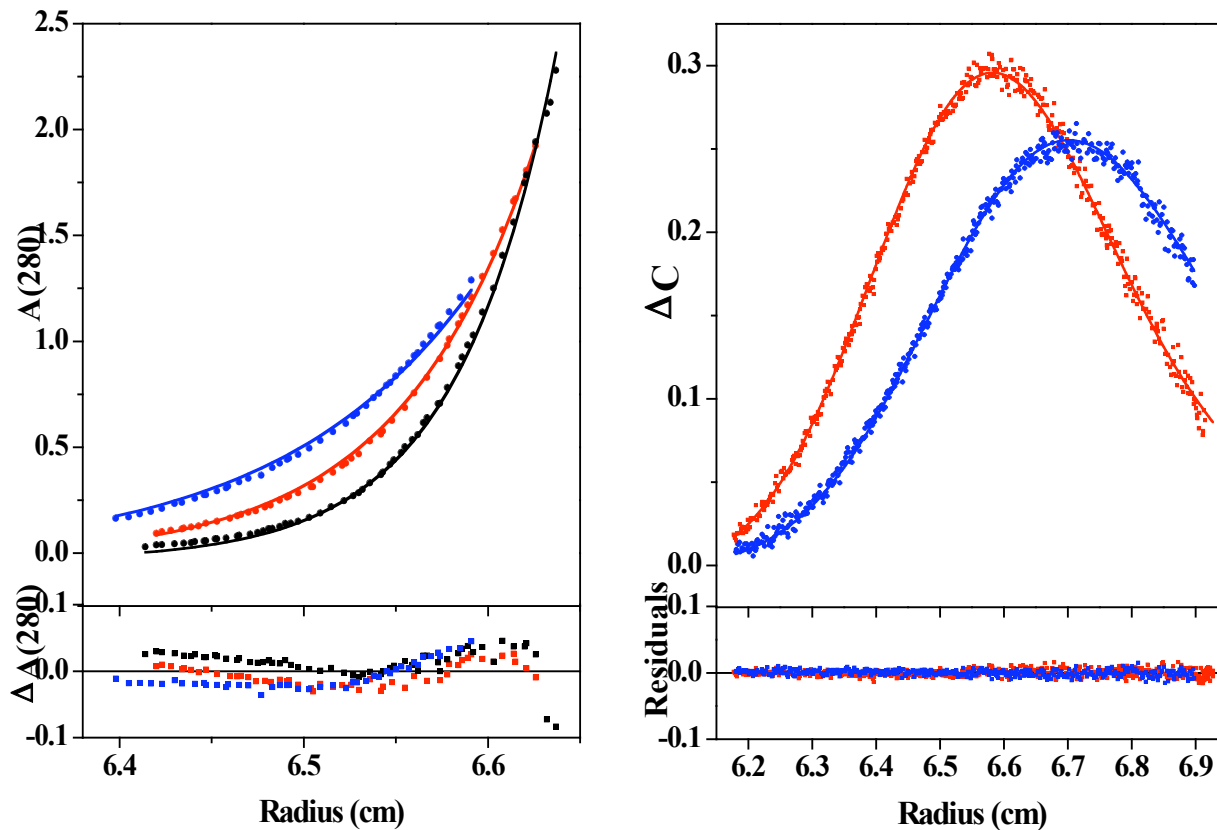
Figure S3 Insight structures



Legend to Figure S3.

Overlay of the ten best structures from rMD Insight/Discover. A view of the 10 overlaid structures. B,C overlays compared with A-DNA (B) or B-DNA (C). Only non-hydrogen atoms were used for the superposition.

Figure S4. Analytical ultracentrifugation



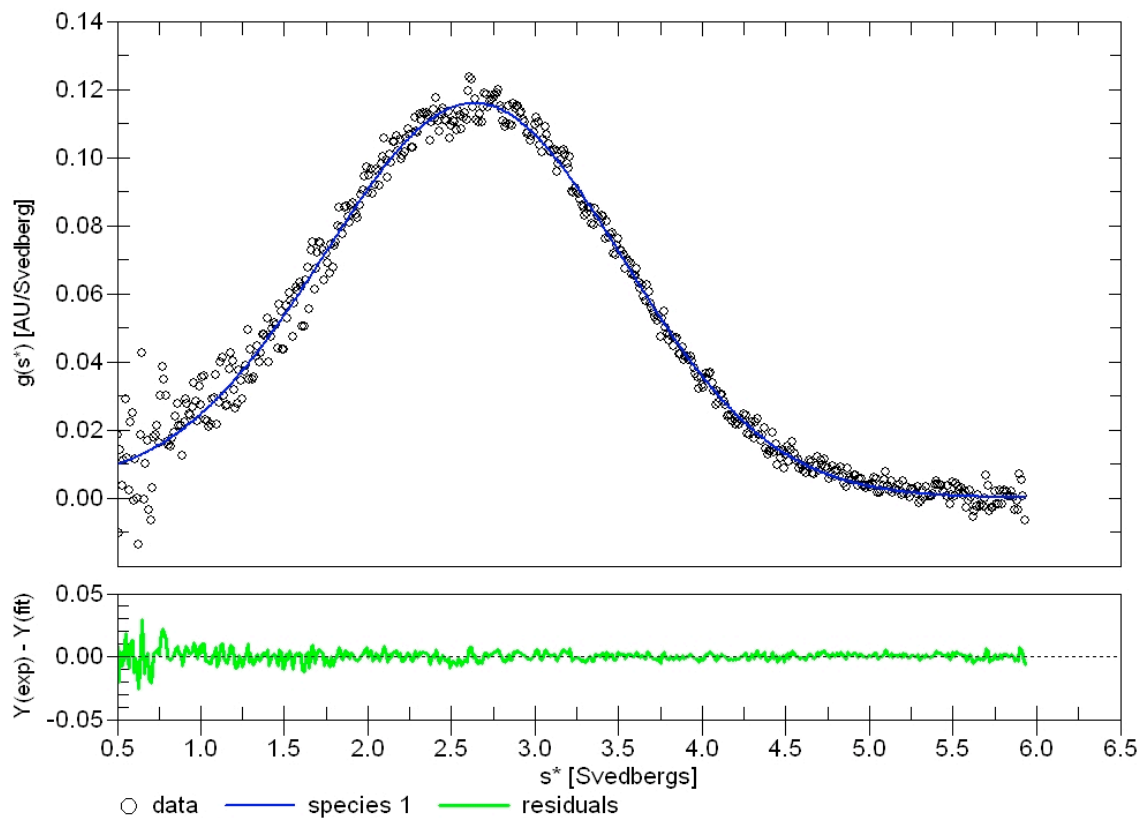
Legend to Figure S4.

Analytical centrifugation of Mbp1 at 20°C on a Beckman Optima XLA centrifuge.

Left equilibrium centrifugation at three protein concentrations (0.25, 0.5 and 0.9 mg/ml) – radial absorbance profile at 280 nm with associated single (global) exponential. Residuals are shown below,

Right DCDT+ analysis of the Mbp1 at two concentrations and associated fit to a single sedimenting species. The residuals are shown below.

Figure S5



Legend to Figure S5. Analytical centrifugation of Mbp1-DNA complex at 20°C on a Beckman Optima XLA centrifuge.

DCDT+ profile of a mixture of 5'-fluorescein-labeled DNA with an equimolar concentration of protein monitored at 495 nm. The rotor speed was 40 000 rpm. The associated fit and residuals (bottom panel) indicate a single sedimenting species under these conditions.