Supporting Information

for the manuscript

B-DNA to Zip-DNA: Simulating DNA Transition to a Novel Structure with Enhanced Charge Transport Characteristics

Alexander Balaeff [1], Stephen L. Craig [2], and David N. Beratan [1,2,3]
Departments of [1] Chemistry, [2] Biochemistry, and [3] Physics, Duke University,
Box 90349, Durham, NC 27708-0349

Table of Contents

Preparation of the Initial Structure for the Steered Molecular Dynamics (SMD) Simulations

The modeled 16 bp-long DNA segment has the sequence GGATATCCGCTTAAGC (strand A; the complementary strand B has the sequence GCTTAAGCGGATATCC). An all-atom structure of that segment was built using the online tool *model.it* [1], available at *http://hydra.icgeb.trieste.it/~kristian/dna/model_it.html*. Among possible DNA models provided by *model.it*, the dinucleotide "NMR" model was selected, since the resulting structure was the closest (RMSD-wise) to the DNA structures resulting from all the other models. The "NMR" model is based on the average helicoidal parameters of DNA dinucleotides observed in NMR structures of DNA.

The "NMR" DNA model was solvated in a box of TIP3 water molecules, using VMD [2]; ~20Å padding was applied in every dimension. The resulting system size was 65Å x 65Å x 95Å (the DNA was aligned with the Z axis of the box). 30 Na^+ counterions were added to the box using *meadionize* VMD plugin (*http://www.chem.duke.edu/~ilya/Software/Meadionize/docs/meadionize.html*) based on the finite-difference Poisson-Boltzmann solver MEAD [3] (Macroscopic Electrostatics with Atomic Detail). The final system included 37,788 atoms.

The system was minimized for 400 steepest-descent steps and then equilibrated for 1ns. Thereafter, 2 independent sampling simulations were performed using different sets of initial atom velocities (randomly assigned according to the Maxwell distribution for 298.15K). Simulation 1 ran for 5.5ns; Simulation 2 ran for 3.65ns. The first 0.5ns of Simulation 1 and 0.4ns of Simulation 2 were used for velocities equilibration.

The simulations were performed using NAMD 2.61b [4] and employed the CHARMM 27 force field [5,6], periodic boundary conditions, an NPT ensemble (T=298.15K, P=1atm), full electrostatics computed with the particle mesh Ewald (PME) method, and a multiple timestepping integration scheme (nonbonded interactions within the 16 Å cutoff distance computed every 2 steps, and long-distance component of the electrostatic interactions computed every 4 steps).

Other simulation parameters:

- 1 fs time step,
- pairlist regenerated every 10 steps (twice per integrator cycle),
- PME grid size 64x64x96,
- 14 Å cutoff for short-range electrostatic and van der Waals interactions, with pairlists maintained out to 16 Å; between 12 Å and 14 Å, the interactions are smoothly reduced to zero using SWITCH algorithm,
- a dielectric constant of 1,
- all bonds rigid (constrained to their equilibrium values using SETTLE algorithm),
- Langevin dynamics is used for temperature control, with all the heavy atoms in the system coupled to the 298.15 K thermal bath. The coupling constant was 5 ps^-

^{1} during the equilibration stage, 1 ps^{-1} during the second (velocity) equilibration stage, and 0.2 ps^{-1} during the sampling runs,
- Nose-Hoover Langevin barostat used for pressure control; the oscillation period of the barostat piston was 100 fs, and the damping time for the piston was 500 fs,
- TIP3P water model.

5500 snapshots, saved every 1ps during Simulation 1, and 3650 snapshots, saved every 1ps during Simulation 2, constituted our structural ensemble for the simulated DNA segment. A root-mean-square deviation (RMSD) of the DNA structure was calculated for every snapshot against every other snapshot; the resulting RMSD map is shown in Fig.S1.

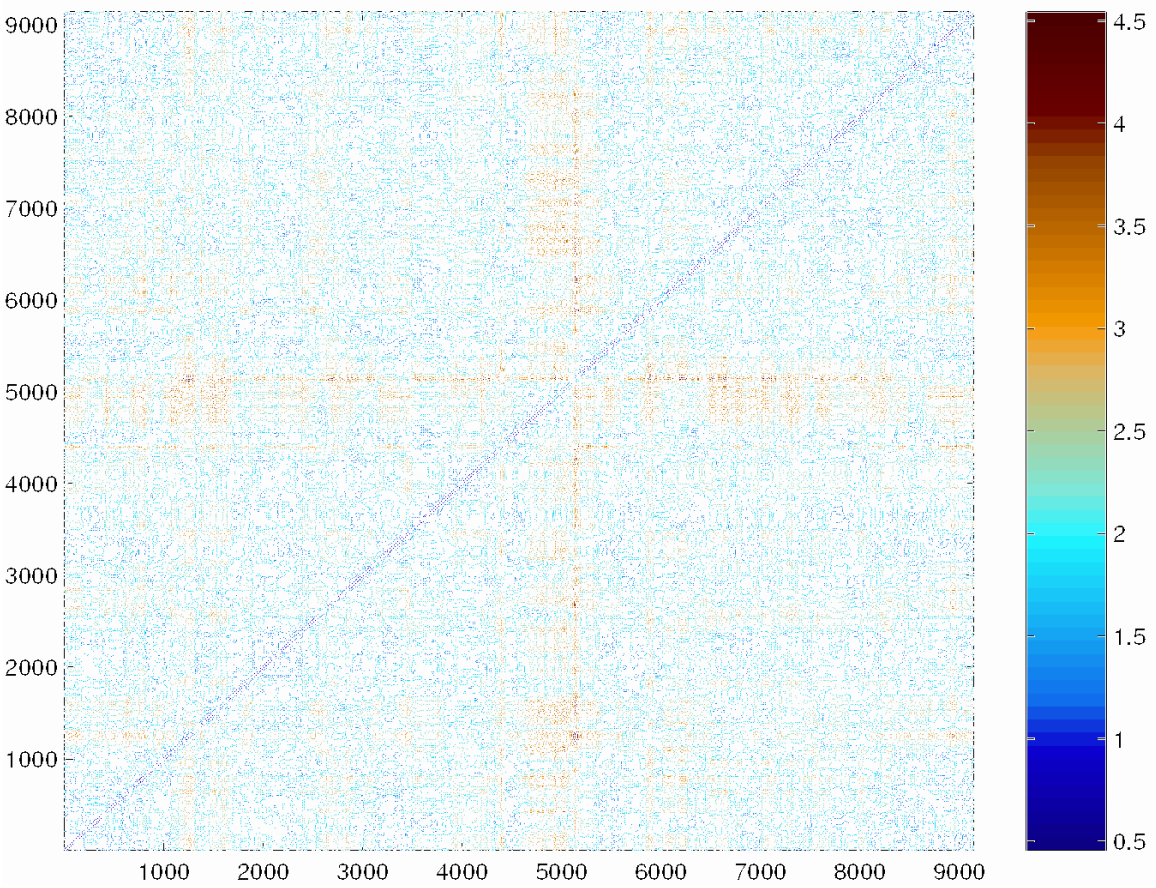

Figure S1. A composite RMSD map for the 9.1 ns sampling run. Each point (i,j) of the map corresponds to the RMSD between DNA structures in snapshots i and j. Only non-hydrogen atoms of residues 2-15 of each DNA chain are used to compute the RMSD. 5500 snapshots of Simulation 1 corresponds to i = 3651 – 9150. 3650 snapshots of Simulation 2 correspond to i = 3650 – 1 (in reverse order). The map is colored according to the RMSD value; the color scale (in Å) is shown on the right.

The map shows that most of the DNA structures are similar (RMSD within 2.5 Å), except for one island between time steps ~1 ns and ~1.5 ns of Simulation 1. Excluding that

island and the velocity equilibration stages (first 0.5 ns of Simulation 1 and first 0.4 ns of Simulation 2 – see above), we computed the ensemble-average RMSD for every snapshot. A snapshot with the third lowest ensemble-average RMSD (1.54572 Å), corresponding to time step 3.36 ns of Simulation 1, was selected as the initial point for the subsequent SMD simulations. For comparison, the two lowest ensemble-average RMSDs were 1.51941 Å and 1.54171 Å.

Then the DNA structure from the selected "best" snapshot was re-solvated in the SMD solvent bath. The "core" part including: (i) the DNA segment, (ii) a 17 Å-thick shell of water and Na^+ ions around the DNA, and (iii) additional 6 Å-thick shells of water around each selected Na^+ ion, was cut out of the selected snapshot. The core system was re-solvated in the SMD water box, with the dimensions of 120Å x 65 Å x 65 Å. The DNA was aligned with the X axis of the box. More Na^+ ions were added to the system to bring the total number to 30. The combined water shell was minimized for 400 steepest-descent steps and then equilibrated for 0.5 ns around the harmonically restrained DNA. The simulation parameters were the same as above, except:

- PME grid size: 128x64x64,
- Langevin coupling constant for temperature control: 5 ps^{-1},
- DNA non-hydrogen atoms harmonically restrained to their initial positions with the force constant of 25 kcal/(mol Å).

The SMD simulations

The final structure from the 0.5 ns equilibration run for water and ions around the constrained DNA snapshot, as described above, was used to initiate the Steered Molecular Dynamics (SMD)[7,8] simulations (listed in Table 1 of the main manuscript). The initial velocities were randomized at the beginning of each simulation. The simulation parameters were the same as those during the water equilibration, except:

- no harmonical restraints on DNA, except for those on two "fixed" atoms (see below)
- Langevin coupling constant for temperature control: 0.2 ps^{-1}.

The SMD forces were applied to two atoms on one end of the DNA segment, namely, C_3' of G_{15} of strand A and C_4' of C_2 of strand B. The chosen atoms are located on the ribose rings of G_{15A} and C_{2B}. The pulling force was not applied to the terminal residues C_{16A} and G_{1B} due to their expectedly disordered state during the simulation. The applied SMD forces were equal to half of the desired total pulling force and were directed along the X axis of the water box.

On the other end of the DNA segment, two atoms were harmonically restrained to their initial positions: C_4' of G_2 of strand A and C_3' of C_{15} of strand B. The restraining force constant was 10 kcal/(mol Å). Similarly, the restraints were not applied to the terminal residues G_{1A} and C_{16B} due the expected disorder of those.

The C_4'/ C_3' atoms to which the restraints and the pulling force were applied are located in the DNA backbone between the last and the next to last residue on each DNA strand. Therefore, the conformation of both residues has been affected by the pulling force and the restrains. For example, the restrained residues G_{A2} and C_{B15} were not allowed to approach each other to form a zip-conformation. In view of that, two terminal bases rather than only one terminal base on each DNA strand were excluded from the RMSD analysis of zip-DNA structure.

The SMD simulations are summarized in Tables 1,2 and Figs. 2-4 of the main manuscript. Additional data are shown below.

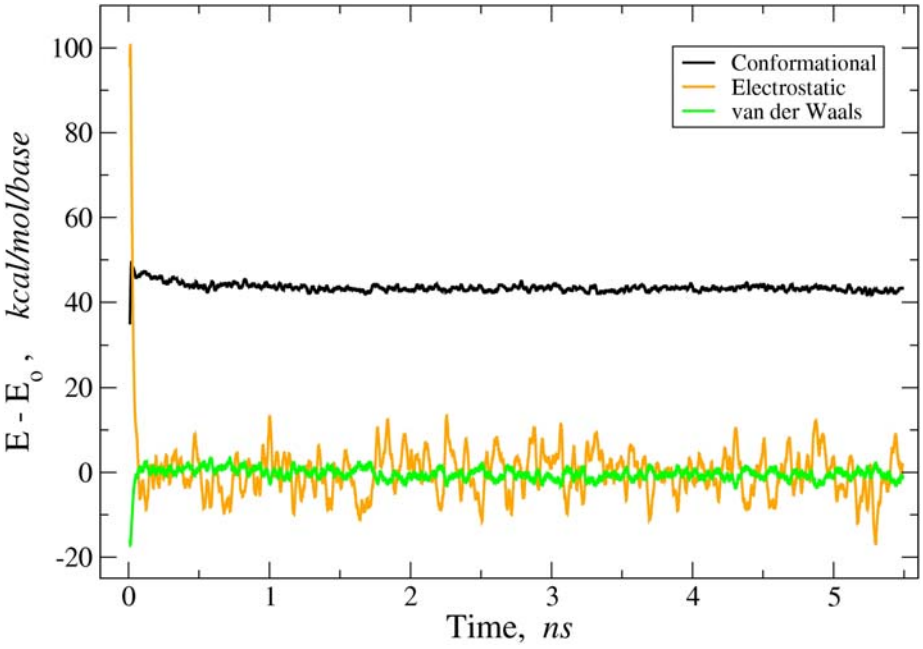

Figure S2. Energy plot for SMD run SMD-H5 (F = 11.2 nN), similar to the plot for SMD-H1 in Fig. 4C of the main manuscript. Note that the large pulling force of 11.2 nN significantly increases the conformational energy of the DNA segment compared to both the small-force runs and the 5.6 nN runs.

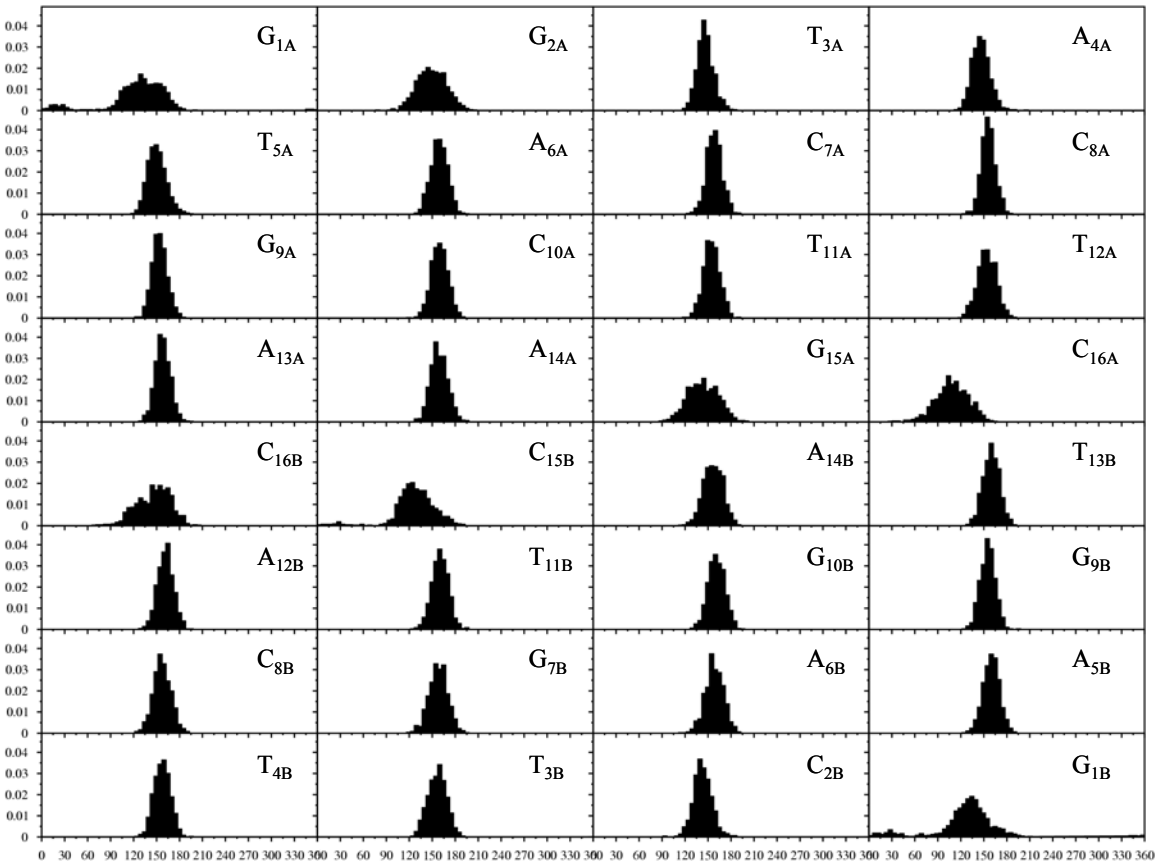

Figure S3. Sugar pucker angle histograms for the last 500 ps of SMD-H1 simulation (cf. Table 1). Note that the pucker angles in this, extended zip-DNA conformation fluctuate in the 130°-170° range characteristic of the C_2'-*endo* conformation of B-DNA. Similar pucker angle distributions are found in the other SMD simulations. The pucker angles were computed with 3DNA [9].

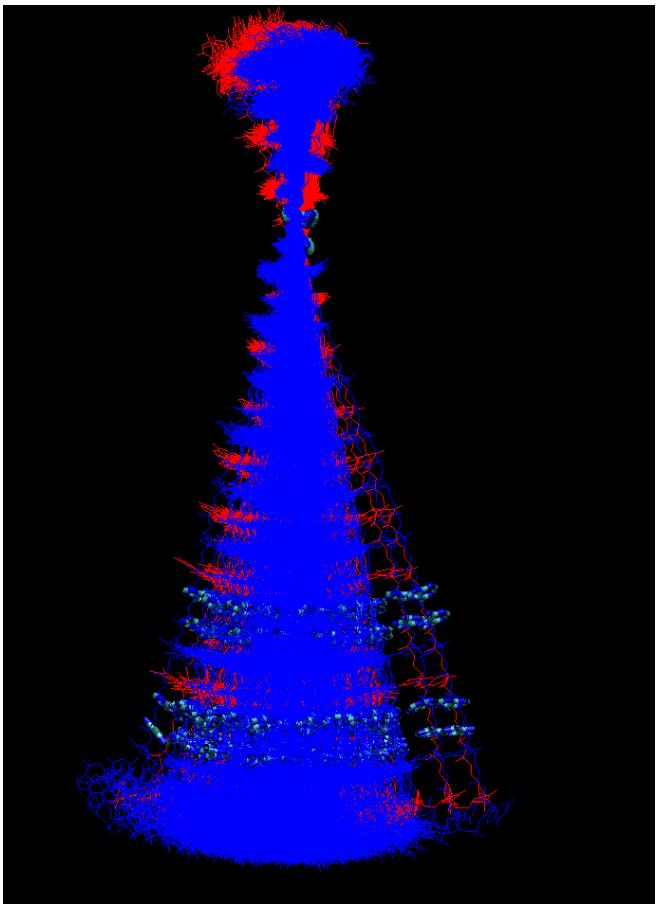

Figure S4. Precession motion of the rigid zip-DNA. The last 500 snapshots of the SMD1 simulation are aligned with respect to the structure of the $A_{14B}A_{4A}$ base step. While the DNA segment is very rigid (RMSD of 0.93 ± 0.11 Å), it exhibits a significant rigid body-like motion. The fixed end of the DNA is on the top of the Figure; the pulling force points down. The DNA structure is shown as lines; red and blue colors indicate the two DNA chains. 2 more AA base steps are shown in bonds representation to emphasize the pendulum-like motion of the zip-DNA segment.

Solvent Partition for the Electrostatic Calculations

In addition to the energy of the full system, the electrostatic energies of several subsystems were computed: (i) the energy of interaction between the DNA strands; (ii) the energies of interaction of each DNA strand with itself; (iii) the energy of DNA-solvent interaction (including both water and Na^+ ions); and (iv) the energy of solvent-solvent interactions (see the Methods section and Table 3 of the main manuscript).

The subsystems are electrostatically charged: -16 elementary charge units per DNA strand and +32 units for the solvent. Therefore, the Ewald series for those subsystems would diverge and one could not compute the subsystem energies using the Particle-Mesh Ewald scheme (the infinite cutoff interactions in the periodic system). A different electrostatic energy function had to be chosen.

We chose to compute the by-component electrostatic interactions with an infinite cutoff but for a single unit cell only. However, the DNA coverage by the solvent in a rectangular unit cell, typically saved during an MD simulation, is uneven (Fig. S5). An electrostatic calculation for such a cell could lead to computational artifacts. Therefore, we rearranged the solvent in the unit cell prior to the electrostatic calculations in order to achieve the most even coverage of the DNA by the solvent.

That was accomplished by using the copy of e water molecule or a Na^+ ion from such a unit cell that places that water/ion at the shortest possible distance from the DNA. That is equivalent to shifting the waters and ions in every dimension by a single unit cell vector until finding the closest position to the DNA. Such rearrangement of the water molecules results in a Voronoi partition of the simulation space with respect to the simulated DNA.

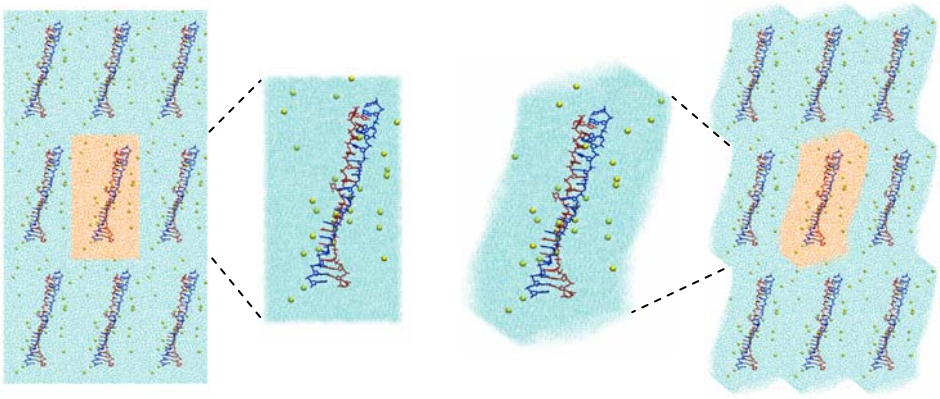

Figure S5. A system snapshot from the end of the SMD7 simulation (40 ns) is used to illustrate the different partitions of the simulations space. On the left, the solvent molecules (water and Na^+ ions) are wrapped into a rectangular unit cell which is replicated in every dimension by the periodic boundary conditions during the SMD simulation. On the right, the solvent molecules are partitioned into Voronoi cells around the DNA molecule. The DNA is shown in red and blue chemical bond representation, the Na^+ ions are shown as yellow spheres, and the water O atoms are show as cyan dots. In the grid representations, the central unit cells are shown in orange.

Nucleobase Coupling in zip-DNA

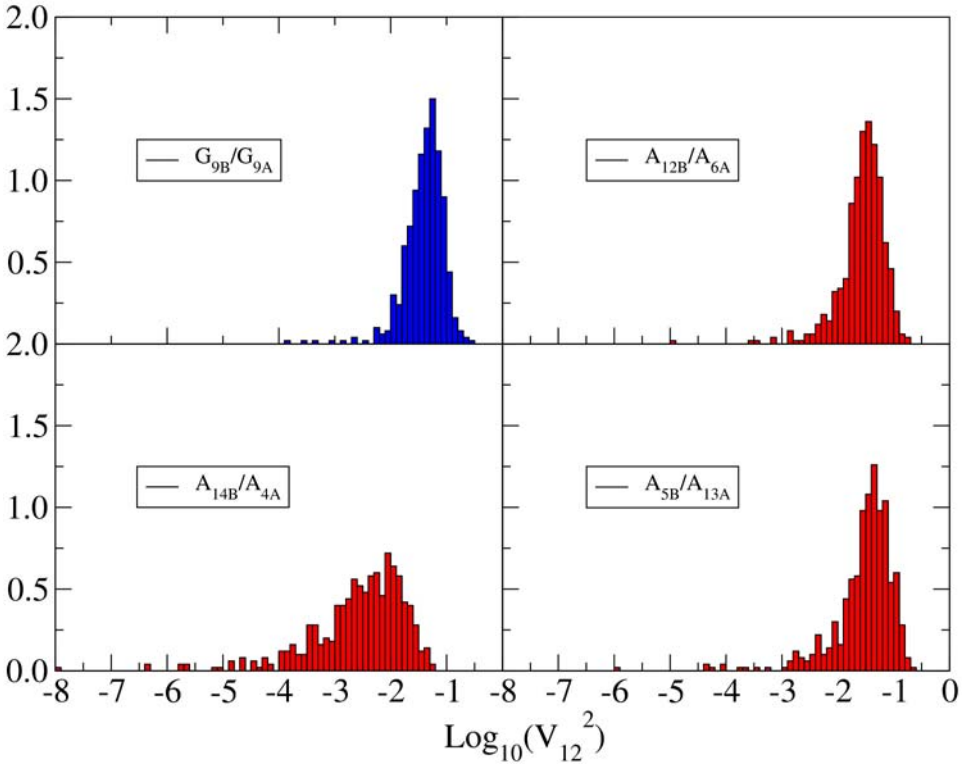

Figure S6. Electronic coupling between HOMO orbitals of the nearest-neighbor purine bases in zip-DNA, computed as described in Methods. Note that the coupling distribution for the bases A_{14B} / A_{4A} is different from those for the bases A_{12B} / A_{6A} and A_{5B} / A_{13A}, apparently due to a higher degree of thermal disorder between the former base step compared to the two latter base steps. Therefore, the base step A_{14B} / A_{4A} was excluded from the structural ensemble for A-A coupling computations in zip-DNA.

Critical Tilt Estimate

We want to estimate the critical tilt at which S-DNA base steps switch into zip-DNA conformation. For this oversimplified estimate, let us consider a completely unwound S-DNA base step that has developed tilt α (Fig. S7). Let us assume that the distance $h \approx 3.4$ Å between the aromatic bases has not changed between B- and zip-DNA. Simple geometrical calculations yield the intrastrand overlap o between the bases A_1 and B_2 (measured from the midpoint between A_1 and B_1 to the midpoint between A_2 and B_2):

$$o = h\,\mathrm{tg}(\alpha)$$

Let us assume that the critical overlap, at which the zip-contact between bases A_1 and B_2 is effectively created, is equal to $o_{cr} \approx 5$ Å (roughly the size of a pyrimidine base). Then,

$$\alpha_{cr} = \mathrm{arctg}(o/h) \approx 55.8^{o}$$

The corresponding length L_{cr} of an N base pair-long DNA is

$$L_{cr} = Nh/\cos(\alpha_{cr}) = L_o/\cos(\alpha_{cr}) \approx 1.78 L_o$$

This calculation is, of course, oversimplified both mathematically and physically because it reduces the 3-dimensional structure of DNA to a 2-dimensional projection, does not consider the purine/pyrimidine type of A_i and B_i bases, neglects the DNA thermal fluctuations, *etc.*

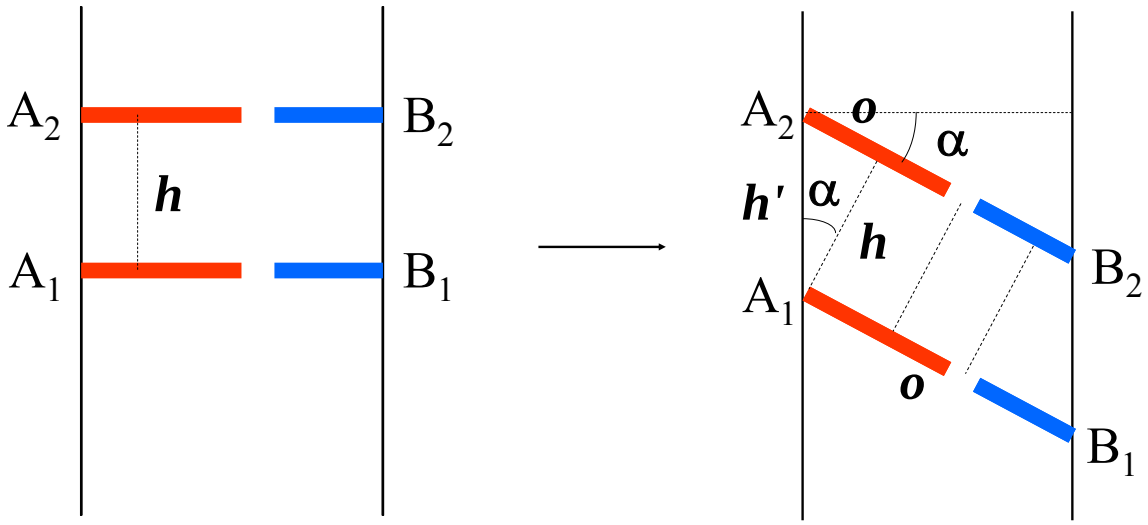

Figure S7. A simplified calculation of the intrastrand base overlap.

Bibliography

(1) Munteanu, M. G.; Vlahovicek, K.; Parthasarathy, S.; Simon, I.; Pongor, S. *Trends in Biochemical Sciences* **1998**, *23*, 341-347.

(2) Humphrey, W.; Dalke, A.; Schulten, K. *Journal of Molecular Graphics* **1996**, *14*, 33-38.

(3) Bashford, D. In *ISCOPE 97*; Ishikawa, Y., Oldehoeft, R. R., Reynders, J. V. W., Tholburn, M., Eds.; Springer: 1997; Vol. 1343, p 233-240.

(4) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *Journal of Computational Chemistry* **2005**, *26*, 1781-1802.

(5) MacKerell, A. D.; Banavali, N. K. *Journal of Computational Chemistry* **2000**, *21*, 105-120.

(6) Foloppe, N.; MacKerell, A. D. *Journal of Computational Chemistry* **2000**, *21*, 86-104.

(7) Isralewitz, B.; Gao, M.; Schulten, K. *Current Opinion in Structural Biology* **2001**, *11*, 224-230.

(8) Sotomayor, M.; Schulten, K. *Science* **2007**, *316*, 1144-1148.

(9) Lu, X.-J.; Olson, W. K. *Nucl. Acids Res.* **2003**, *31*, 5108-5121.