Supporting Information for

PHYSICAL MODEL FOR RECOGNITION TUNNELING

Predrag Krstic[1,*], Brian Ashcroft[2] and Stuart Lindsay[2,3,4]

[1]Institute for Advanced Computational Science, Stony Brook University, Stony Brook, NY

11794-5250, USA

[2]Biodesign Institute, [3]Department of Physics, [4]Department of Chemistry and Biochemistry

Arizona State University, PO Box 875001, Tempe, Arizona 85287, USA,

*E-mail: predrag.krstic@stonybrook.edu

1. Orientational and radial dependences of hydrogen bonding in the oxDNA model

The oxDNA model is described in detail in the published literature[25-30] (see the References in the paper) as well as at the oxDNA website https://dna.physics.ox.ac.uk/index.php/Main_Page). We here show some illustrations of the model of importance for our application. It should be noted that oxDNA has high level of coarse graining (each nucleotide is built of two "atoms", one representing sugar-phosphorous back-bone, and one representing a particular base). The interactions included in the model are shown in figure S1. The orientational dependence of the interactions between the nucleotides captures the planarity of bases, of importance for formation of helical duplexes. The coaxial stacking term captures stacking interactions between bases that are not immediate neighbors along the backbone of a strand. Hydrogen-bonding interactions, V_{HB}, in the model are possible only between complementary, Crick-Watson (A-T and C-G) base pairs, with the additional condition of approximate anti-alignnment of the phosphodiester linkage, which leads to the formation of double helical structures. This is only sequence-dependence that is included in oxDNA and is limited to thermodynamic properties. Bases and backbones also have excluded volume interactions, as well as backbones interactions, V_{BB}. The model neglects several features of the DNA structure and interactions due to the high level of coarse-graining. For example, oxDNA is only appropriate for the study of systems at high salt concentration. The model does not have any explicit electrostatic interactions, but is fitted to reproduce presence of the strong electrostatic screening in the DNA behavior at salt concentration $[Na^+]$=0.5M, some short-range electrostatic properties being incorporated into a short-ranged repulsive excluded volume potential between bases. In addition all four nucleotides have the same structure, with no size dependence. Although the interactions do not provide realistic chemical interactions, the oxDNA model still provides a physically realistic description of effects of the interactions[25]. A deficiency of the model is that only Watson-Crick pairs are allowed. We introduce a base Z in Section 4 of the paper, which interacts with all four DNA bases. Still, oxDNA has the ability to model correctly the double DNA helix, and nontrivial DNA nanodevices, like a two-footed DNA walker[30], and DNA origami[25].

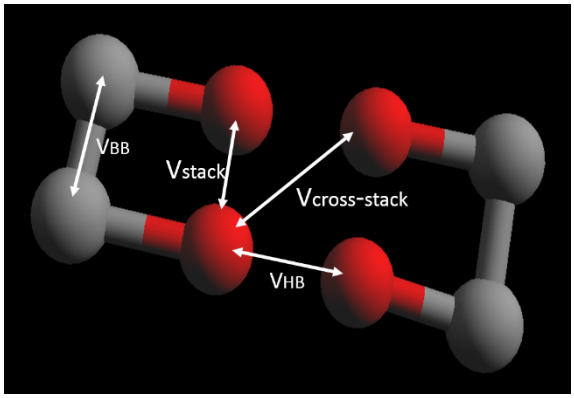

Figure S1. Interactions in the coarse-grained oxDNA model

Figure S2 shows the radial dependence of the $V_{HB}(R)$ for randomly chosen sets of angles $\bar{\theta}$ for the Watson-Crick pairs A-T and G-C. This set of angles defines the mutual alignment of the two antiparallel bases. The range of coupling extends to more than 6 Angstroms, while the absolute minimum of the coupling is close to 3.4 Angstroms and is about 0.32 eV for G-C and about 0.23 eV for A-T hydrogen bonding (not achieved in the figure). We note that these values include two-H bonds for A-T and three for G-C.

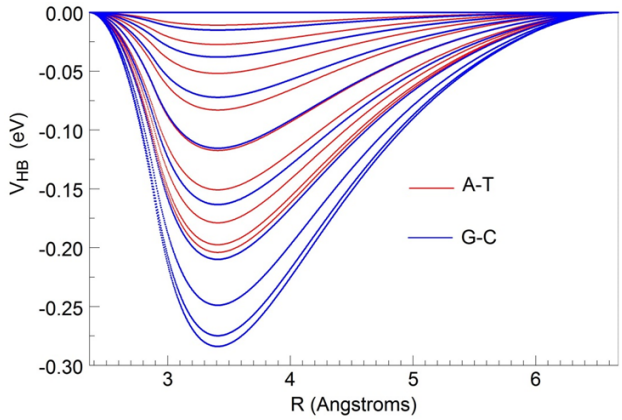

Figure S2. The hydrogen bonding potential energy V_{HB} as function of the distance between the interacting bases, for random sets of the orientation angles.

The V_{HB} curves, shown in figure S2, were extracted from the analytic forms embedded in the oxDNA model.

2. Coarse-grained dynamics for interacting nucleotides

We tested the oxDNA model using Watson-Crick (W-C) base pairing only, before introducing our "Z" base reader. We set a short ssDNA segment of a length of few tens of bases, applying at its ends a stretching force of 24.3 pN, whose purpose is to prevent the folding of the DNA segment (Fig S3). In the computing box (7-12 nm) we also set some number of dimers containing a base which W-C pairs with the DNA segment. The dynamics was run (at 300K) for

a long time (step 15 fs, 10^9 to 10^{10} steps at a time). The state of the system is dumped each 200 steps, including positions and velocities of all particles, energies etc.

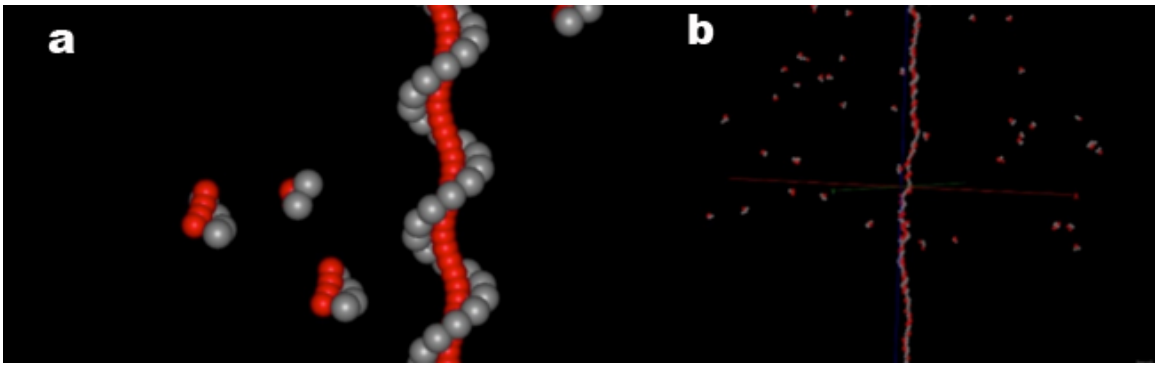

Figure S3. Configuration of the DNA segment in a bath of the DNA dimers. (a) Close look to the segment and a few dimers, and (b) view to the whole bath.

The gray-spheres in Fig. S3 represent the sugar-phosphate back bone, while the red spheres are the bases. We note that the DNA segment is composed of either single or of mixed bases.

As a test of the approach to the dynamics, we perform the calculations in which the DNA segment in figure S3 is either poly-A or poly-G containing 30 nucleotides probed with 15 dimers TT or CC, respectively. Typical traces of the H-bonding energy, E_{HB}, the bonding length, R_{HB}, the resulting bond conductance G, for TT-dimer bonding a poly-A, and their time-averaged values are shown in figure S4a. Notice that the bonding probability P_{HB}, defined as ratio of the time when at least one Z-dimer is bonded to the DNA, and the total time (shown here for the interval of 0.6μs) is P_{HB}=0.11. Multiplying the average value of the conductance (which doesn't includes no-bonding cases) with this probability, yields the average conductance of 1.19 nS. This gives ~600 pA current if bias is 0.5V, surprisingly not too much bigger than the experimental values of the tunneling current. The full set of over 200,000 points of E_{HB} vs. R_{HB} for this example is shown in figure S4b. The boundaries in figure S4b follow the shape of the potential well of the strongest hydrogen bonding (aligned directions of the bonding to the mid hydrogen). The distribution of the bond lengths creates a Gaussian-like distribution (figure S4c). We stress that the probability for the bonding of CC to the poly-G DNA is 0.92, caused by the stronger hydrogen bonding of C-G than of T-A in our model. This is further illustrated at figure S5.

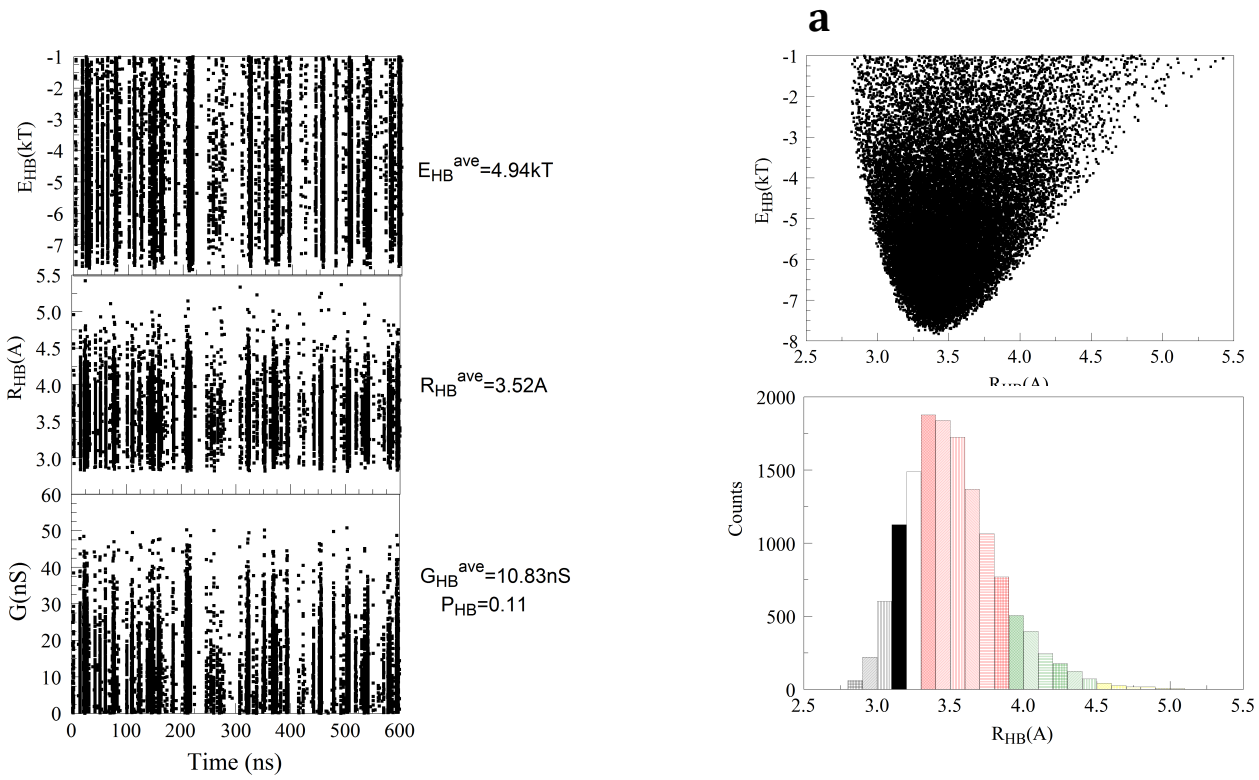

Figure S4. A sample of 200,000 time points, with step of 3 ps, showing for probing ssDNA of A-type with the Z-dimers (a) distribution of hydrogen bond energies, bond lengths and hydrogen bond conductance's, (b) the hydrogen bond energy vs. bonding distance, and (c) the Maxwellian-like distribution of the bonding lengths.

The comparison of the calculated distributions of the H-bond lengths, as well as of the total hydrogen bond energies as functions of time for two independent homopolymers A and G probed by the TT and CC dimers are shown in figure S5. As mentioned above, the larger H-bond energies (in average) of the G-C bonds reflect larger strength of the G-C H-bond. The bonds appear in clusters over extended interval of time, and this clustering, as a rule, lasts much longer for G-C than for A-T bonds. For illustration, figure S5b shows typical short-lasting clusters of AT bonds, on the phone of the much longer-lasting GC bond, in a short interval of 5 ns.

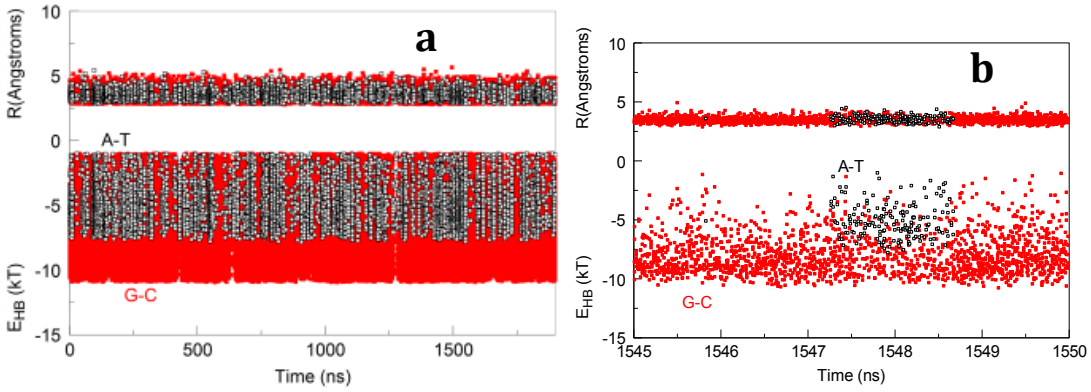

Figure S5. a) Distributions of the H-bond distances and energies of the A-T and G-C bonds; b) Snapshot of a) during only 5 ns.

The fluctuating conductance through the AT and GC bonds is shown in figure S6. This is obtained for a single heteropolymer containing string of 20 A nucleotides, followed by a string of 20 G nucleotides, probed by 10 TT dimers and 10 CC dimers. As can be seen in figure S6b, besides double GC multiplicity (2GC) and double AT multiplicity (2AT) reflecting bond of one CC dimer to the G polymer and one TT dimer to A polymer, there is a significant number of higher-order bonds from a simultaneous bonding of two CC dimers to G (4GC), bonding three CC polymers (6GC) to G, bonding one CC dimer to G and one TT dimer to A. Interestingly, pairing of just one monomer of a dimer is a relatively probable event (GC, 3GC, AT,…). Of course, non-pairing at all is the most probable event in this numeric experiment (~65%). The conductance of all multiple bonds is included in figure S6a. Although the peak values of the conductance are significantly larger for the AT bonds, as consequence of choice of smaller exponential factor β in Eq. 2 in the text of the paper, the significantly larger occurrence of the GC bonds in figure S6b, indicates a comparable or larger tunneling current through the GC hydrogen bonds, in comparison to the AT.

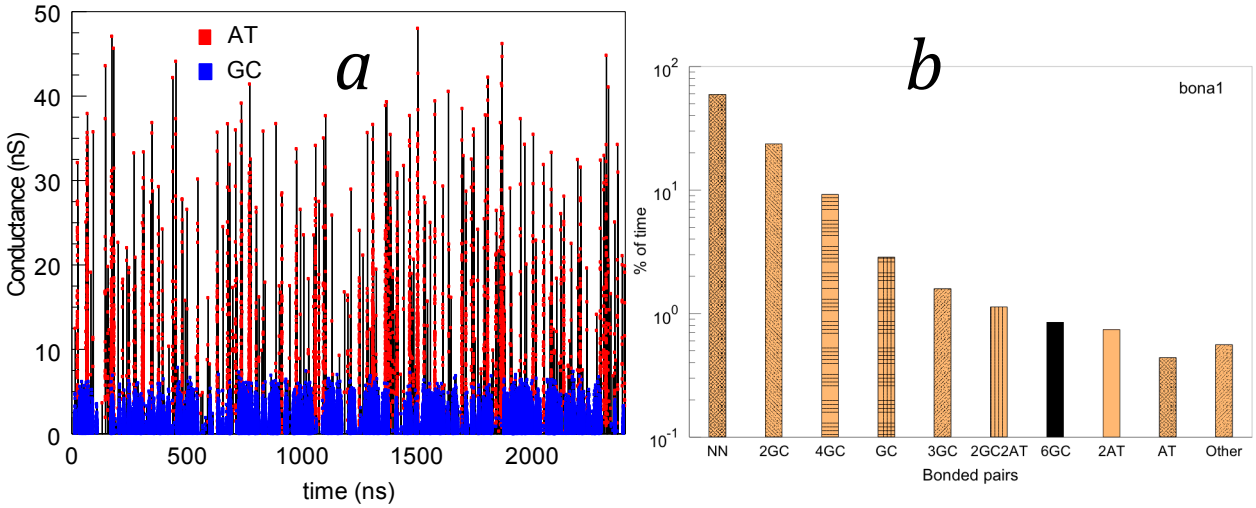

Figure S6. (a) Conductance signal through A-T (red symbols) and G-C bonds (blue). The GC signals are significantly smaller in the peak values due to the larger factor β, in Eq. 2 of the paper. (b) Multiplicity of the bonds in % of time. NN is the time when no dimer is bonded to polymer by a hydrogen bond.

3. Reader-Base Interactions with ZZ-Dimers

When the probe dimer is made of two Z monomers, both Z in ZZ will likely bond, increasing the bonding time and bonding frequency, as can be seen when figure S7 is compared with single Z bonds, as shown in figure 5 of the paper.

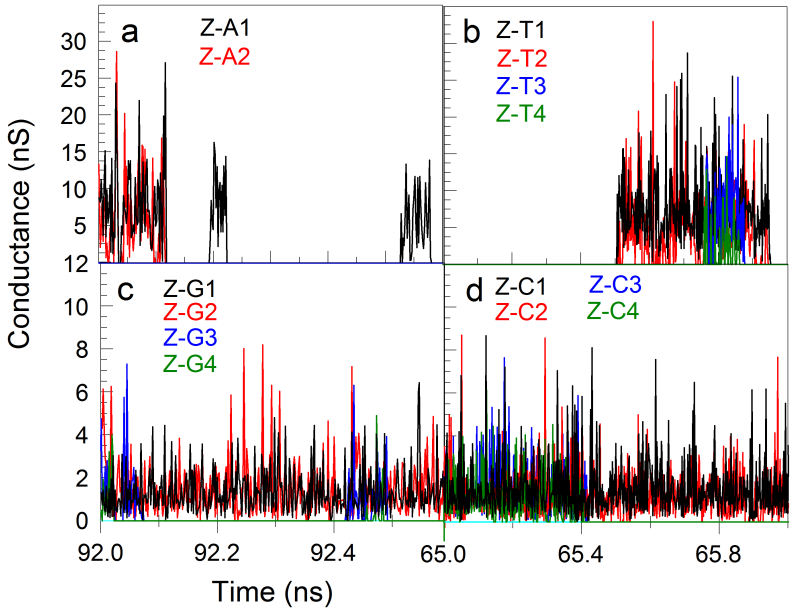

Figure S7: Conductance vs. time for Z-A (a) Z-T (b), Z-G (c) and Z-C (d) complexes. The multiplicity of each complex is indicated by the color code. To accentuate multiplicity the ZZ-dimers were used for the probes. The same beta β=3.0 is used for all complexes.

The signal from heterogeneous ssDNA of 70 bases A,T,C, and G, GCCGTTCGCACGGCGCGAAGGAGCGGCTGCCAGTTCCAAGTGCGGACGCGGCTGCC GCAACGGAGCTCGT probed by dimers ZZ over a period of 130 ns is shown in figure S8. The frequency of bonds to different bases is a visible characteristic of the signal spectrum.

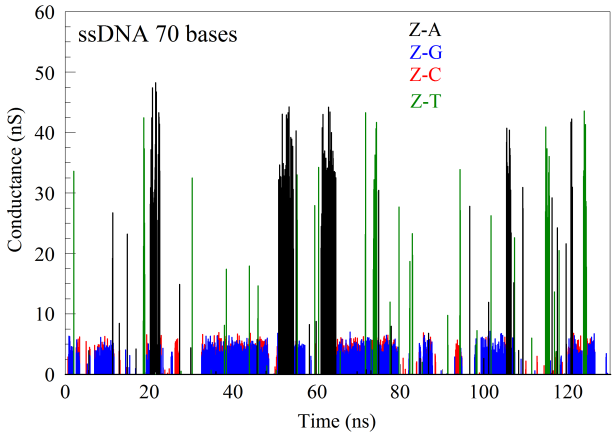

Figure S8. Hydrogen-bond conductance between ZZ-probes and various bases in a 70-tople heterogeneous DNA which contains all 4 bases, with base-dependent β, as discussed below Eq. 2 of the paper. The RT Signal form G and C are much smaller because of the larger β. The frequency of the peaks for various bond-types is also a visible characteristic.

4. Example of a “Single Peak” in the simulated data

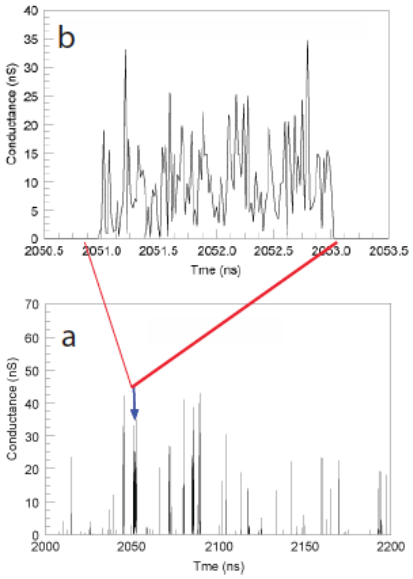

Figure S9 (a) Current-time trace showing RT signal peaks. (b) Expanded plot of the peak pointed to by the arrow showing the fine structure.