Energy Landscape and Pathways for Transitions Between Watson-Crick and Hoogsteen Base Pairing in DNA

Debayan Chakraborty*,†,‡ and David J. Wales*,†

E-mail: dc550@cam.ac.uk; dw34@cam.ac.uk

Supporting Information

Preparation of Initial Structures

The initial coordinates for the Hoogsteen conformation corresponding to the ($ATTAAT_2$) duplex were taken from a previously published high resolution crystal structure (PDB ID: 4U9M). The initial structure of the Watson-Crick duplex was constructed using the nucleic acid builder (NAB) module available within the AMBER code.[1] The duplexes were modeled using a properly symmetrized version[2] of the AMBER99bsc0 force field,[3] employing the latest $\chi OL4$ corrections.[4] Symmetrization is essential to ensure that accessible permutational isomers have the same energy. For exploration of the energy landscape using the discrete path sampling method, the solvent effects were considered implicitly using a generalized Born model.[5,6] An effective salt concentration of 0.1 M was maintained using the Debye-Hückel approximation.[7]

Molecular Dynamics Simulations

The initial structures corresponding to the Watson-Crick and Hoogsteen duplexes were solvated in a truncated octahedral box of TIP3P water molecules, with a solvent buffer of at least 10 Å on each side. The net charge in the simulation box was neutralized by adding an appropriate number of Na^+ counterions. The ion parameters appropriate for the TIP3P water model, as proposed by Joung and Cheatham,[8] were used. The particle mesh Ewald summation technique[9] was used to compute the electrostatic energy, using the standard settings in the AMBER code. A 10 Å cutoff for the non-bonded interactions was employed. A integration time-step of 2 fs was used in conjunction with SHAKE constraints[10]

for all the bonds involving hydrogen atoms. The MD simulations were carried out using the GPU-enabled version of the AMBER12 code.[11]

The solvated systems were first minimized using sequential applications of the steepest-descent and conjugate gradient algorithms to remove steric clashes. After minimization, the temperature of the system was gradually increased from 0 to 300 K with 25 ps of NVT dynamics. During this time, positional restraints of 50 kcal mol^{-1} were applied on the DNA molecule. The restraints were systematically relaxed through five subsequent cycles of NVT simulations, each of length 25 ps. During this equilibration phase, temperature control was maintained via a Langevin thermostat,[12] employing a collision frequency of 0.2 ps^{-1}. After heating, and progressive relaxation, the density of the system was equilibrated using 2 ns of NPT simulation, at a constant pressure of 1 bar, and temperature of 300 K. Following equilibration, production runs of duration 200 ns were carried out in the NPT ensemble.

The various structural parameters of the DNA molecule were analyzed using the *cpptraj* module available within the AmberTools13 distribution.[11]

Discrete Path Sampling

The energy landscape for the DNA duplex was explored using the discrete path sampling (DPS) technique.[13] DPS exploits geometry optimization to provide a coarse-grained description of the underlying landscape in terms of stationary points (minima and transition states). The connectivity between stationary points on the potential energy surface (PES) is described in terms of discrete paths. A discrete path between two endpoints of interest consist of a sequence of minima connected by intervening transition states. The endpoints are denoted as reactant, and product, respectively. A geometric criterion is employed to identify minima and transition states. For each stationary point, the normal-mode frequencies are obtained from the eigenvalues of the mass-weighted Hessian matrix. A stationary point for which all the nonzero normal mode frequencies are positive is a local minimum. In contrast, a transition state has a single imaginary frequency.[14,15] Displacements parallel and antiparallel to the corresponding eigenvector are used to initiate the steepest-descent paths that lead to the adjoining minima.

The OPTIM code[16] interfaced with the AMBER9 package[1] was used for all the geometry optimizations, transition state searches, and normal-mode analysis. A modified version of the LBFGS algorithm, described by Liu and Nocedal, was employed for the local minimizations.[17] To identify candidate transition state structures between pairs of local minima, we used the doubly-nudged elastic band (DNEB) method.[18–20] The transition state candidates were further refined using the hybrid eigenvector-following technique.[21] Geometry optimizations were deemed to have converged when the root-mean-square-gradient fell below 10^{-6} kcal mol^{-1} Å^{-1}. After each cycle of connection-making attempts, a large number of intervening minima and transition states may be located, especially if the endpoints are far apart in configuration space. As a result, the number of possible connections between local minima that might be tried to generate a fully connected discrete path grows with each cycle. To avoid a combinatorial problem, we employed the missing connection algorithm[22] to construct a priority list of connection attempts based on an appropriate edge-weight metric.

After an initial discrete path was found between between the endpoints of interest, the stationary point databases were further expanded using various refinement schemes.

The refinement of the databases were carried out using the PATHSAMPLE code,[23] which distributes parallel OPTIM jobs across compute nodes to connect different pairs of minima. The SHORTCUT BARRIER scheme[24] is efficient in locating pathways characterized by lower energy barriers. In this procedure, connection attempts are prioritized between pairs of minima on either side of, and an equal number of steps away from, the largest potential energy barriers. Another scheme, SHORTCUT, is used to locate shorter pathways.[24] In this procedure, pairs of local minima that are closest together in configuration space, but are separated by a minimum number of steps on the discrete path, have a higher priority in the connection-making attempts. During the database refinement, we exploited some recently introduced interpolation techniques based on natural internal coordinates,[25] and quasi-continuous schemes,[26] which exploit the connectivity of the covalently bonded network. In our experience, these methods are effective in circumventing common problems associated with linear interpolations, such as unphysical chain crossings, and steric clashes in the intervening images.

Extensive sampling of specific discrete paths may often introduce artificial frustration (kinetic traps) into the databases. This frustration is caused by undersampling of certain regions of the landscape, and needs to removed to make the network a faithful representation of the global kinetics. To remove the artificial traps we used the UNTRAP scheme,[24] available within the PATHSAMPLE code. In this procedure, candidate minima for connection attempts are chosen based on the ratio of the potential energy barrier to the potential energy difference from the product region.

The stationary point databases (kinetic transition network) were refined using sequential applications of the SHORTCUT BARRIER, SHORTCUT, and UNTRAP schemes until the phenomenological rate constants corresponding to the WC$\longleftrightarrow$ HG switch converged to within an order of magnitude, with respect to the addition of new stationary points.

Calculation of Free Energies

The database of stationary points obtained from DPS simulations was used to estimate the free energies using the superposition approach.[27] Here, the total energy density of states, $\Omega(E)$, and the canonical partition function, $Z(T)$ are written as a sum of contributions from the catchment basin of each local minimum.[28,29]

$$\Omega(E) = \sum_i \Omega_i(E), \tag{1}$$

and

$$Z(T) = \sum_i Z_i(T). \tag{2}$$

$\Omega_i(E)$ and $Z_i(T)$ are respectively the density of states (DOS) and the partition function (PF) for the basin of attraction of minimum i. The basin of attraction is defined as the region in configuration space from which a steepest-descent minimization leads to minimum i, so each point in the configuration space can only belong to the basin of a single minimum, unless it lies on a boundary.[15]

The equilibrium occupation probability of minimum i is

$$p_i^{eq}(T) = \frac{Z_i(T)}{Z(T)}. \tag{3}$$

Since the sums in equations (1) and (2) are over all the geometrically distinct minima on the landscape, the DOS and the PF must also contain a multiplying prefactor to take into account identical contributions from permutation-inversion isomers of each minimum. For a system containing N_A atoms type A, N_B atoms of type B, N_C atoms of type C, etc., the prefactor is given by $n_i = 2N_A!N_B!N_C!.../o_a$, where o_a is the order of the point group for the minimum i.[15]

The density of states of local minima can be estimated analytically using certain approximations. The simplest approach is to assume that the potential well around each local minimum is harmonic in nature, and the Taylor series expansion of the energy in the neighborhood of the minimum can be truncated to second order.[15] A vibrational analysis (which involves diagonalization of the mass weighted Hessian matrix) at the configuration corresponding to the local minimum yields the normal mode angular frequencies ω_j, where $1 \leq j \leq \kappa$. Here $\kappa = 3N - 6$, the number of vibrational degrees of freedom of the system.

The microcanonical density of states for a system with κ degrees of freedom can then be expressed as:[30]

$$\Omega(E) = \sum_i \Omega_i(E) = \sum_i \frac{n_i \left(E - V_i\right)^{\kappa-1}}{\Gamma\left(\kappa\right) \prod_{\alpha=1}^{\kappa} h\nu_\alpha(i)}, \tag{4}$$

where $\nu_\alpha(i) = \omega_\alpha(i)/2\pi$ is the vibrational frequency of mode α for minimum i, V_i is the potential energy of minimum i, $\Gamma(\kappa) = (\kappa - 1)!$, n_i is the multiplicative prefactor, and h is Planck's constant.

The canonical partition function can be obtained from the total energy density of states by a Laplace transform.[15]

$$Z(T) = \int \Omega(E) e^{-\beta E} dE = \sum_i \frac{n_i e^{-\beta V_i}}{(\beta h \overline{\nu}_i)}, \tag{5}$$

where $\beta = 1/k_B T$ and $\overline{\nu}_i = \left[\prod_{\alpha=1}^{\kappa} h\nu_\alpha(i)\right]^{1/\kappa}$, the geometric mean of the vibrational frequencies of minimum i. The free energy of each minimum is expressed in terms of its associated partition function as:[15]

$$F_i(T) = -k_B T \ln Z_i(T). \tag{6}$$

A similar expression is also used for the free energy of a transition state j:

$$F_j^{\dagger}(T) = -k_B T \ln Z_j^{\dagger}(T) \tag{7}$$

$Z_j^{\dagger}(T)$ is defined in the same way as the corresponding partition functions for the minima, but the normal mode corresponding to the negative Hessian eigenvalue is omitted from the expressions.

The heat capacity C_v can be expressed in terms of the partition function, $Z(T)$, using standard thermodynamic relations:[30]

$$C_v = \left(\frac{\partial U(T)}{\partial T}\right)_{N,V} \tag{8}$$

where $U = -\partial \ln Z(T)/\partial\beta$ is the internal energy. Using equation (5), C_v corresponding to the superposition partition function is[15]

$$C_v = \kappa k_B - \frac{z_1(T)^2}{k_B T^2 z_0(T)^2} + \frac{z_2(T)}{k_B T^2 z_0(T)} \tag{9}$$

where

$$z_r(T) = \sum_i n_i \left(V_i\right)^r \left(\frac{1}{\beta h \overline{\nu}_i}\right)^\kappa e^{-\beta V_i} \tag{10}$$

Calculation of Rate Constants

The unimolecular rate constant $k_i^\dagger(T)$ for minimum i crossing the transition state $\dagger$ at temperature T, can be estimated using harmonic transition state theory (TST):[31–33]

$$k_i^\dagger(T) = \frac{k_b T Z^\dagger(T)}{h Z_i(T)} e^{-\beta \Delta V}, \tag{11}$$

where $Z^\dagger(T)$ is the partition function for the transition state; $Z_i(T)$ is the partition function of the minimum; ΔV is the potential energy difference between the transition state and the minimum i. The total rate constant $k_{ji}(T)$ for an elementary transition from minimum i to j is obtained by summing the $k_i^\dagger(T)$ values for all transition states that connect the two minima. As TST does not account for recrossing events at the dividing surface, the computed rate constants are upper bound estimates.

Analyzing Global Dynamics from Kinetic Transition Networks

Once the unimolecular rate constants for elementary min-TS-min transitions are known, the phenomenological rate constants k_{AB} and k_{BA} between the reactant (A) and product (B) states can be computed. The two states can be connected via multiple discrete paths, and as discussed, the objective in DPS is to systematically refine the stationary point databases, and locate those pathways that are kinetically relevant and contribute to the rate constant significantly. The phenomenological rate constant is expressed as a weighted sum of contributions from all the discrete paths.

Rate constants between the two sets of minima (A and B) will only be meaningful if the minima within each set are in local equilibrium. The condition for local equilibrium in terms of occupation probabilities is follows:

$$\frac{p_a(t)}{p_A(t)} = \frac{p_a^{eq}}{p_A^{eq}} \qquad \text{and} \qquad \frac{p_b(t)}{p_B(t)} = \frac{p_b^{eq}}{p_B^{eq}}, \tag{12}$$

where a and b denote minima within the states A and B, respectively. The above equations imply that the occupation probability of a minimum within a particular set at time t does not change relative to the occupation probability of the whole set.

If the dynamics are Markovian,[34] the time evolution of the occupation probability for a particular minimum a can be written in the form a linear master equation:[13,35]

$$\frac{dp_a(t)}{dt} = \sum_{b \neq a} \left[k_{ab} p_b(t) - k_{ba} p_a(t) \right], \tag{13}$$

where k_{ab} is the unimolecular rate constant for transitions to minimum a, starting from minimum b, and $p_b(t)$ is occupation probability of minimum b at time t. The sum is over all the geometrically distinct minima, but excluding the permutation-inversion isomers.[15]

If there is a direct connection between the A and B states (i.e. without any other intervening minima), the master equation can be written as:

$$\frac{dp_A(t)}{dt} = -\frac{dp_B(t)}{dt} = k_{AB} p_B(t) - k_{BA} p_A(t), \tag{14}$$

where the phenomenological rate constants k_{AB} and k_{BA} are expressed as weighted sums of the unimolecular rate constants for all the min-TS-min transitions that lie on the boundary of A and B.[13,35]

However, for complex conformational changes, it is unlikely that there will be a direct connection between the A and B states. Instead, A and B are connected via a set of intervening minima $i_1, i_2, i_3, ...i_n$, which can be considered as members of the set I. Within the steady-state approximation, it is assumed that the rate of change of the occupation probabilities of the minima in the set I are low:[13,35]

$$\frac{dp_i(t)}{dt} = \sum_{j \neq i} k_{ij} p_j(t) - p_i(t) \sum_{j \neq i} k_{ji} \approx 0. \tag{15}$$

Substituting the expressions for the occupational probabilities of the minima in the set I in the master equation, the rate constant expressions, within the steady-state approximation, can be written as:[13,35]

$$k_{AB}^{SS} = \frac{1}{p_B^{eq}} \sum_{a \leftarrow b} \frac{k_{ai_1} k_{i_1 i_2} k_{i_2 i_3} ... k_{i_n b} p_b^{eq}}{\sum_{j_1} k_{j_1 i_1} \sum_{j_2} k_{j_2 i_2} \sum_{j_3} k_{j_3 i_3} ... \sum_{j_n} k_{j_n i_n}} \tag{16}$$

$$k_{BA}^{SS} = \frac{1}{p_A^{eq}} \sum_{b \leftarrow a} \frac{k_{bi_1} k_{i_1 i_2} k_{i_2 i_3} ... k_{i_n a} p_a^{eq}}{\sum_{j_1} k_{j_1 i_1} \sum_{j_2} k_{j_2 i_2} \sum_{j_3} k_{j_3 i_3} ... \sum_{j_n} k_{j_n i_n}} \tag{17}$$

The individual sums in the denominators of equations (16) and (17) include the unimolecular rate constants corresponding to all direct transitions from minimum j_k to geometrically distinct minima i_k.

If all the possible transitions out of a certain minimum α are considered as independent Poisson processes,[36] then the mean waiting time in that minimum can be expressed as:[13,35]

$$\tau_\alpha = \frac{1}{\sum_\delta k_{\delta\alpha}}. \tag{18}$$

The transition probability from a minimum α to another directly connected minimum γ can

then be written as:[13,35]

$$P_{\gamma\alpha} = \frac{k_{\gamma\alpha}}{\sum_\delta k_{\delta\alpha}} = k_{\gamma\alpha}\tau_\alpha. \tag{19}$$

Using the expressions in equations (18) and (19), the steady-state rate constants can be rewritten as:

$$k_{AB}^{SS} = \frac{1}{p_B^{eq}} \sum_{a\leftarrow b} P_{ai_1} P_{i_1 i_2} P_{i_2 i_3} ... P_{i_n b} p_b^{eq} \tau_b^{-1}, \tag{20}$$

$$k_{BA}^{SS} = \frac{1}{p_A^{eq}} \sum_{b\leftarrow a} P_{bi_1} P_{i_1 i_2} P_{i_2 i_3} ... P_{i_n a} p_a^{eq} \tau_a^{-1}. \tag{21}$$

The discrete path between A and B that makes the largest contribution to the steady-state rate constant k^{SS} is termed the 'fastest path', and can be extracted from the network using Dijkstra's shortest path algorithm[22] or recursive enumeration analysis.[37,38] For this analysis, an appropriate edge-weight is also required, and it is taken to be the product of the transition probabilities in equations (20) and (21).[13,15] These products therefore represent the statistical weight associated with each discrete path.

The steady-state approximation for the intervening set I can be relaxed, and the non-steady-state rate constants can be expressed in a similar form as equations (20) and (21):[39]

$$k_{AB}^{NSS} = \frac{1}{p_B^{eq}} \sum_{a\leftarrow b} P_{ai_1} P_{i_1 i_2} P_{i_2 i_3} ... P_{i_n b} p_b^{eq} t_b^{-1}, \tag{22}$$

$$k_{BA}^{NSS} = \frac{1}{p_A^{eq}} \sum_{b\leftarrow a} P_{bi_1} P_{i_1 i_2} P_{i_2 i_3} ... P_{i_n a} p_a^{eq} t_a^{-1}. \tag{23}$$

where t_a (t_b) is the mean waiting time for a transition out of minimum a (b) to any minimum in the B or A regions.

Several approaches have been suggested to compute the overall rate constants. For example, the mean waiting times t_a and t_b can be estimated by averaging over multiple kinetic Monte Carlo (KMC) runs.[40] The rate constants in this case can be expressed in terms of the average over the mean first passage times (MFPT) calculated for multiple KMC trajectories starting at either minimum a or b, and ending at a minimum in the B or A region, respectively (T_{Ba} and T_{Ab}):[39]

$$k_{AB} = \frac{1}{p_B^{eq}} \sum_{a\leftarrow b} \frac{p_b^{eq}}{T_{Ab}}. \tag{24}$$

$$k_{BA} = \frac{1}{p_A^{eq}} \sum_{b\leftarrow a} \frac{p_a^{eq}}{T_{Ba}}. \tag{25}$$

However, the KMC method is computationally intensive and scales poorly with the size of the network, as well as temperature. On the other hand, techniques based on the solution of the master equation[41] by diagonalization of the transition matrix encounter numerical instabilities for large stationary point databases, or if slow relaxation time scales are present within the kinetic transition network. In contrast, the new graph transformation (NGT)

method[39] provides a robust formalism for estimating phenomenological rate constants. NGT scales better with temperature and is robust in terms of numerical precision.[42] The NGT procedure removes all minima in the intervening region I progressively, and the transition probabilities as well as the waiting times are renormalized to conserve the mean first passage time (MFPT). The rate constants are then written as:

$$k_{AB} = \frac{1}{p_B^{eq}} \sum_{a \leftarrow b} \frac{P_{Ab}^{'} p_b^{eq}}{\tau_b^{'}} \tag{26}$$

$$k_{BA} = \frac{1}{p_A^{eq}} \sum_{b \leftarrow a} \frac{P_{Ba}^{'} p_a^{eq}}{\tau_a^{'}} \tag{27}$$

In the above equations, $P_{Ab}^{'}$ ($P_{Ba}^{'}$) and $\tau_b^{'}$ ($\tau_a^{'}$) represent the renormalized transition probability and waiting times, respectively. It can be shown that the MFPT for the transition from minimum a to a minimum in the B region is in fact $\tau_a^{'}/P_{Ba}^{'}$,[39] which can be identified as T_{Ba} in equation (25). Therefore, the NGT and KMC formulations are formally equivalent.

In this work, the rate constants describing the global dynamics associated with the HG$\longleftrightarrow$WC transformation were computed using the NGT procedure, in conjunction with a self-consistent regrouping scheme,.[43] For regrouped stationary point databases, the equilibrium occupation probability, and the free energy of the group J is written as:[43]

$$p_J^{eq}(T) = \sum_{j \in J} p_j^{eq}(T), \tag{28}$$

$$F_J = -k_b T \ln \sum_{j \in J} Z_j(T). \tag{29}$$

where minimum j is a member of group J. The free energy of the group of transition states linking J and K is:[43]

$$F_{KJ}^{\dagger} = -k_b T \ln \sum_{k \leftarrow j} Z_{kj}^{\dagger}(T) \equiv -k_b(T) \ln Z_{KJ}^{\dagger}(T), \tag{30}$$

To analyze global dynamics corresponding to regrouped databases, the rate constants corresponding to transitions between different free energy groups are required, which can then be used in the appropriate expressions for k^{SS}, k^{NSS} and k. The inter-group rate constant from J to K is written as:[43]

$$\begin{aligned} k_{KJ} &= \sum_{k \leftarrow j} \frac{p_j^{eq}(T)}{p_J^{eq}(T)} k_{kj}(T) = \sum_{k \leftarrow j} \frac{Z_j(T)}{Z_J(T)} \frac{k_b T}{h} \frac{Z_{kj}^{\dagger}(T)}{Z_j(T)}, \\ &= \frac{k_b T}{h} \frac{Z_{KJ}^{\dagger}(T)}{Z_J(T)} = \frac{k_b T}{h} e^{-\left[F_{KJ}^{\dagger}(T) - F_J(T)\right]/k_b T}. \end{aligned} \tag{31}$$

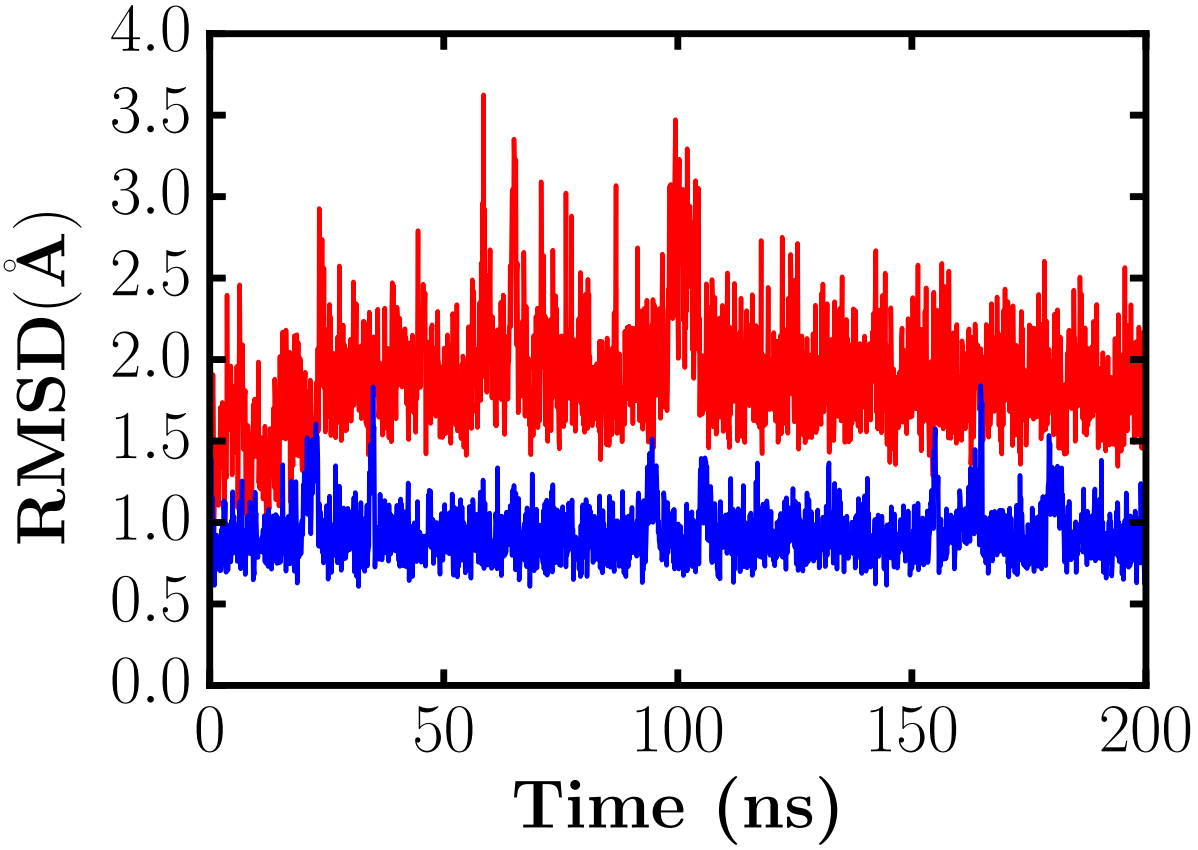

Figure S1. The evolution of the root-mean-square deviation (RMSD) with respect to the initial conformation for the WC (red), and the HG (blue) duplex. The average RMSD along the trajectory is higher for the WC duplex, indicating that it is more flexible.

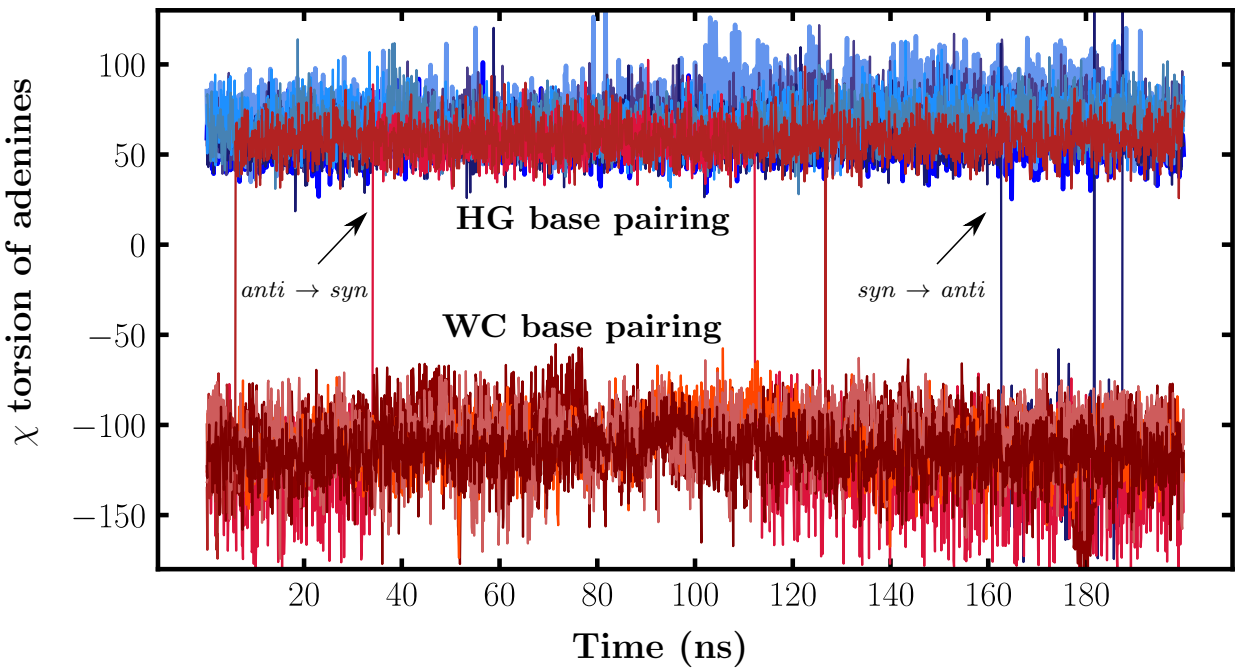

Figure S2. The evolution of the χ torsions for the adenine bases along the parmbsc1 trajectories corresponding to the HG (blue lines), WC (red lines) duplexes.

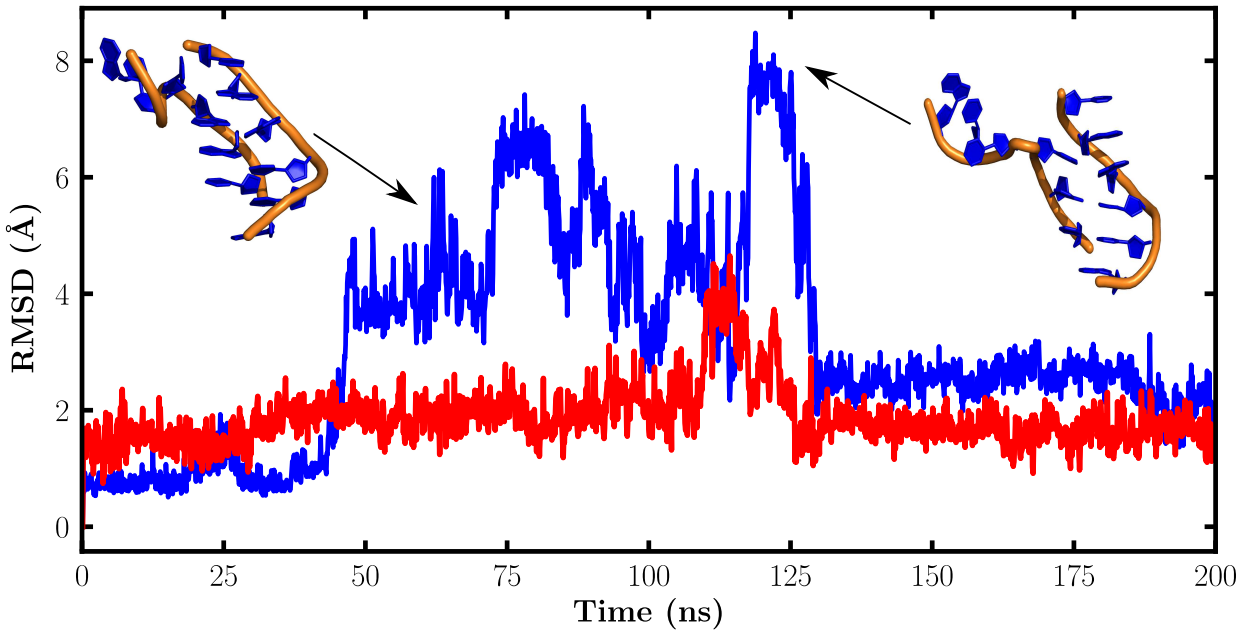

Figure S3. The evolution of the root-mean-square-deviation (RMSD) with respect to the initial conformation for the WC (red), and the HG(blue) duplex, with the parmbsc1 force field. Significant strand slippage is observed for the HG duplex. Two representative snapshots are shown.

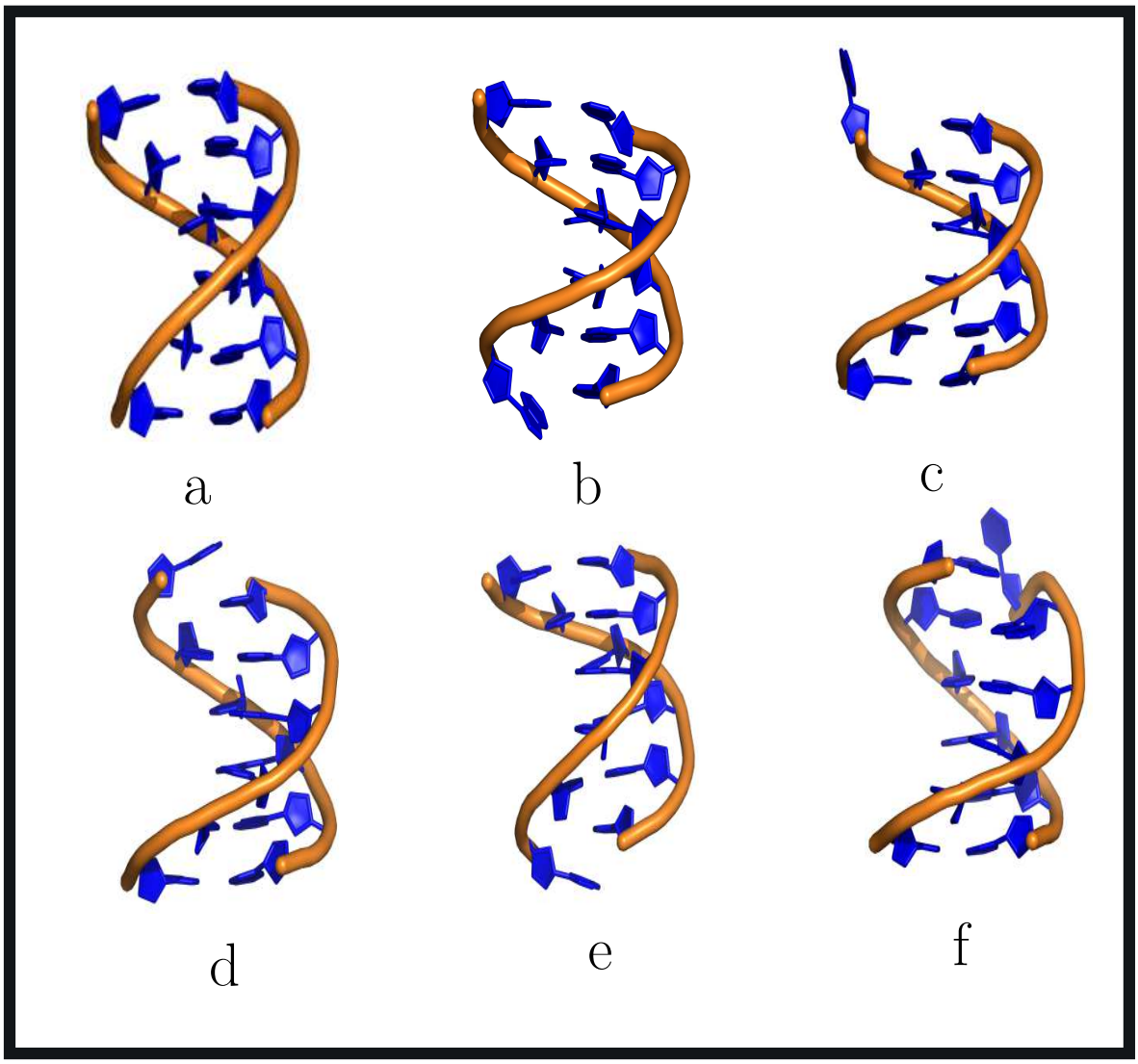

Figure S4. Constituent members of the HG ensemble (a) full HG duplex. (b) A6 frayed out of the helix. (c) A1 frayed out of the helix. (d) A1 stacked on top of T1. (e) A6 stacked on top of T6. (f) T6 frayed out of the helix.

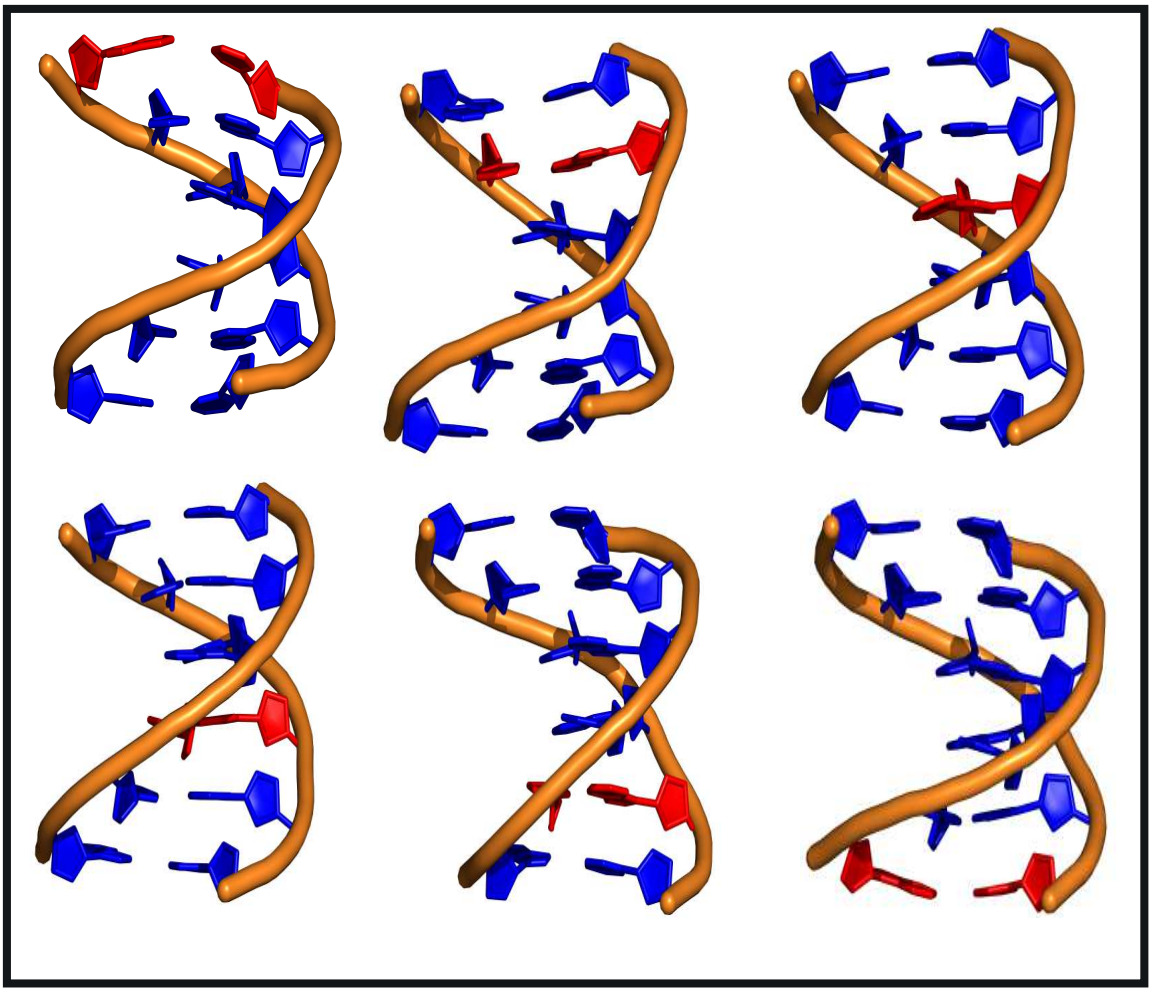

Figure S5. Duplexes of the type **5HG+1WC** with appreciable occupation probabilities at 300 K.

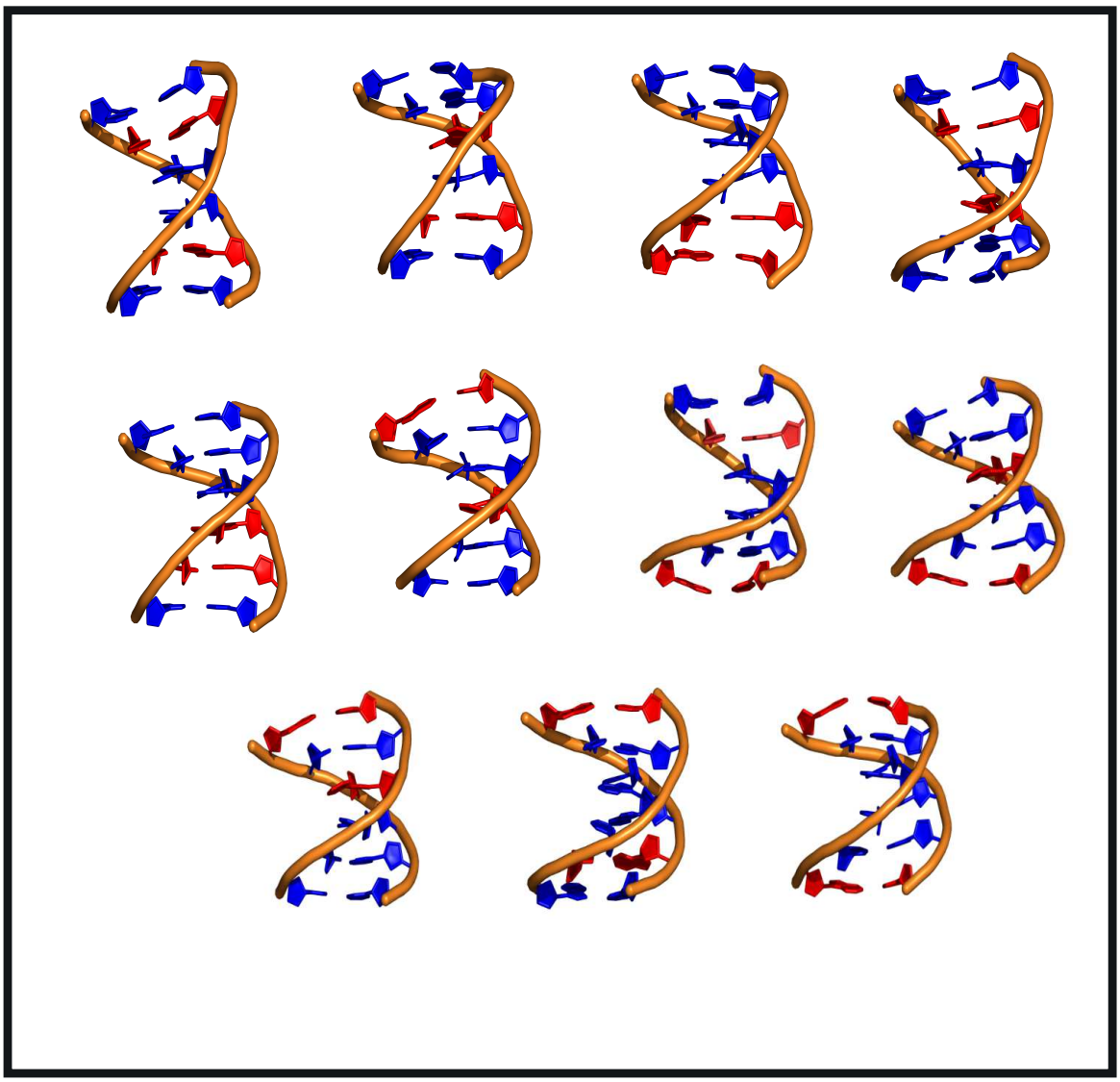

Figure S6. Duplexes of the type **4HG+2WC** with appreciable occupation probabilities at 300 K.

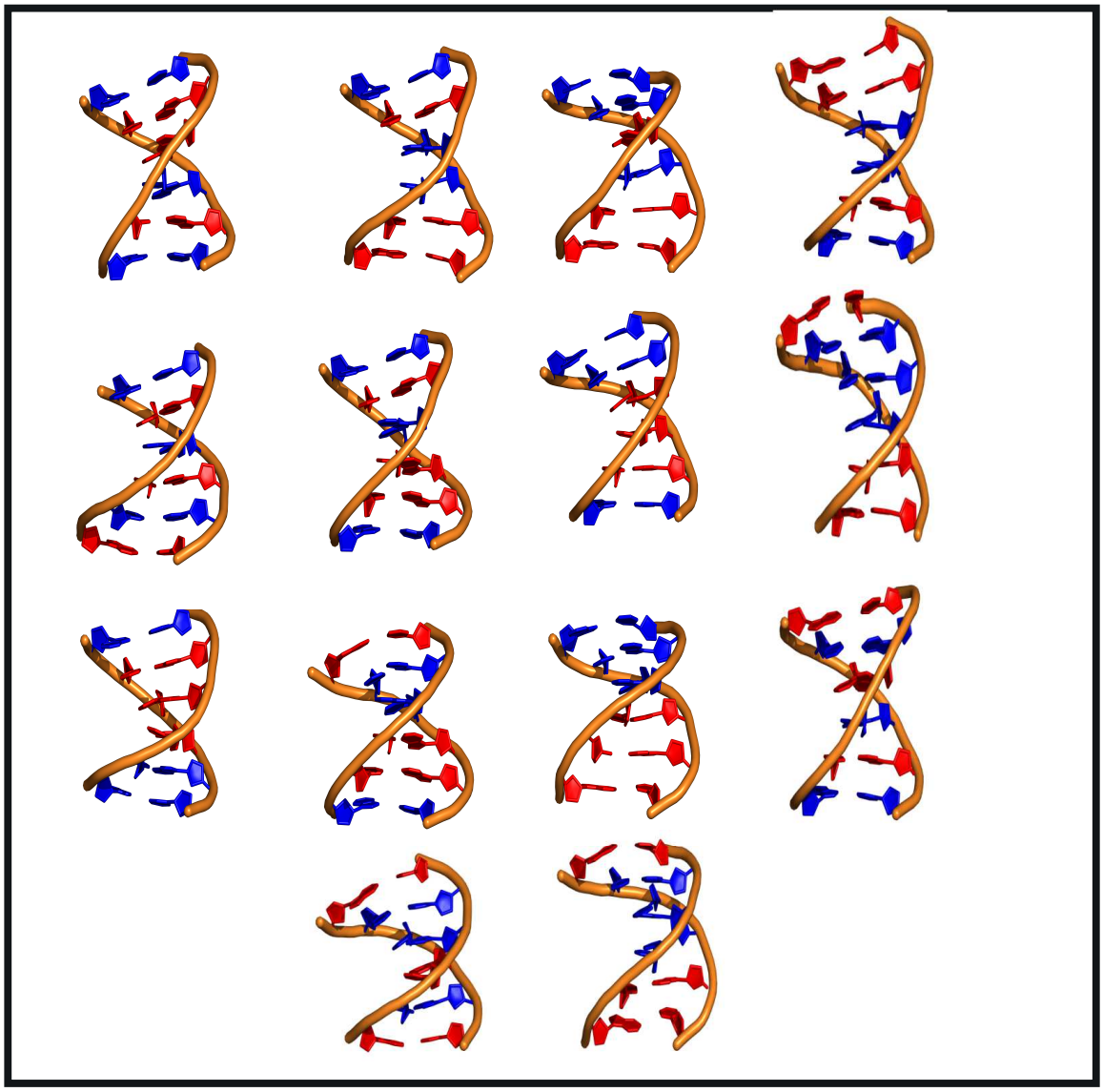

Figure S7. Duplexes of the type **3HG+3WC** with appreciable occupation probabilities at 300 K.

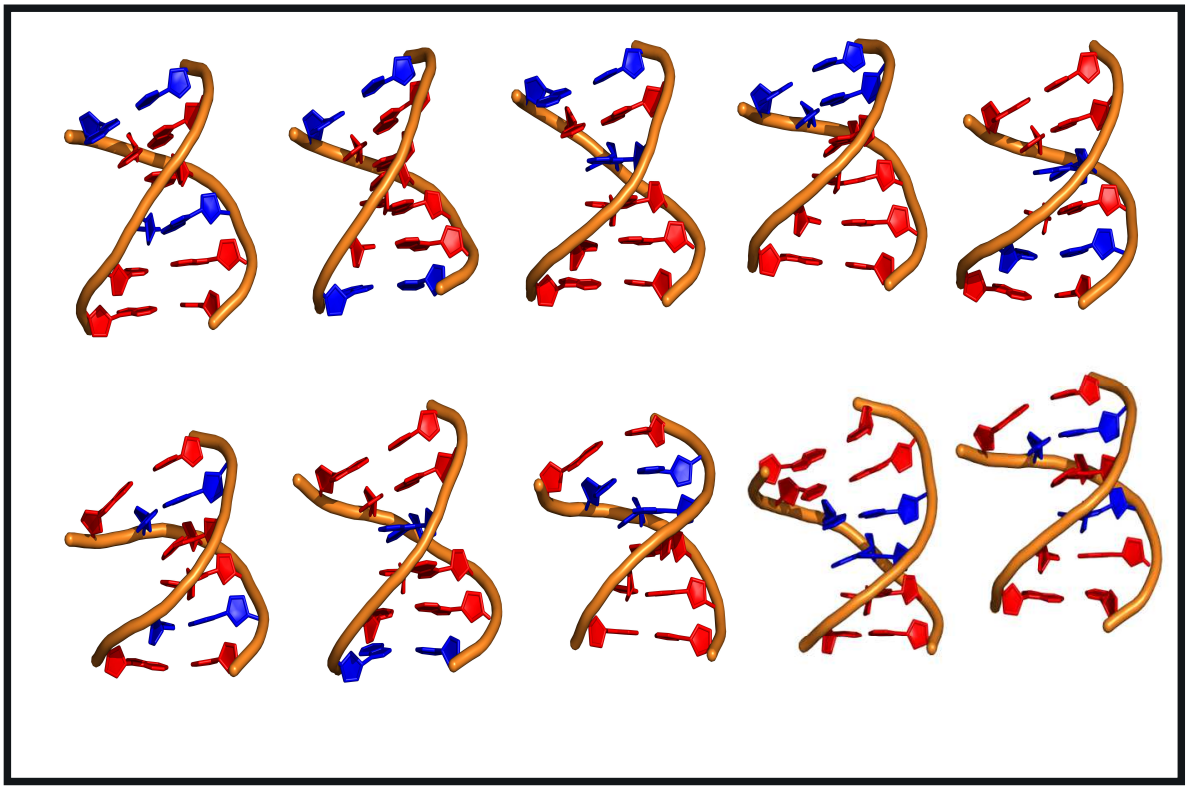

Figure S8. Duplexes of the type **2HG+4WC** with appreciable occupation probabilities at 300 K.

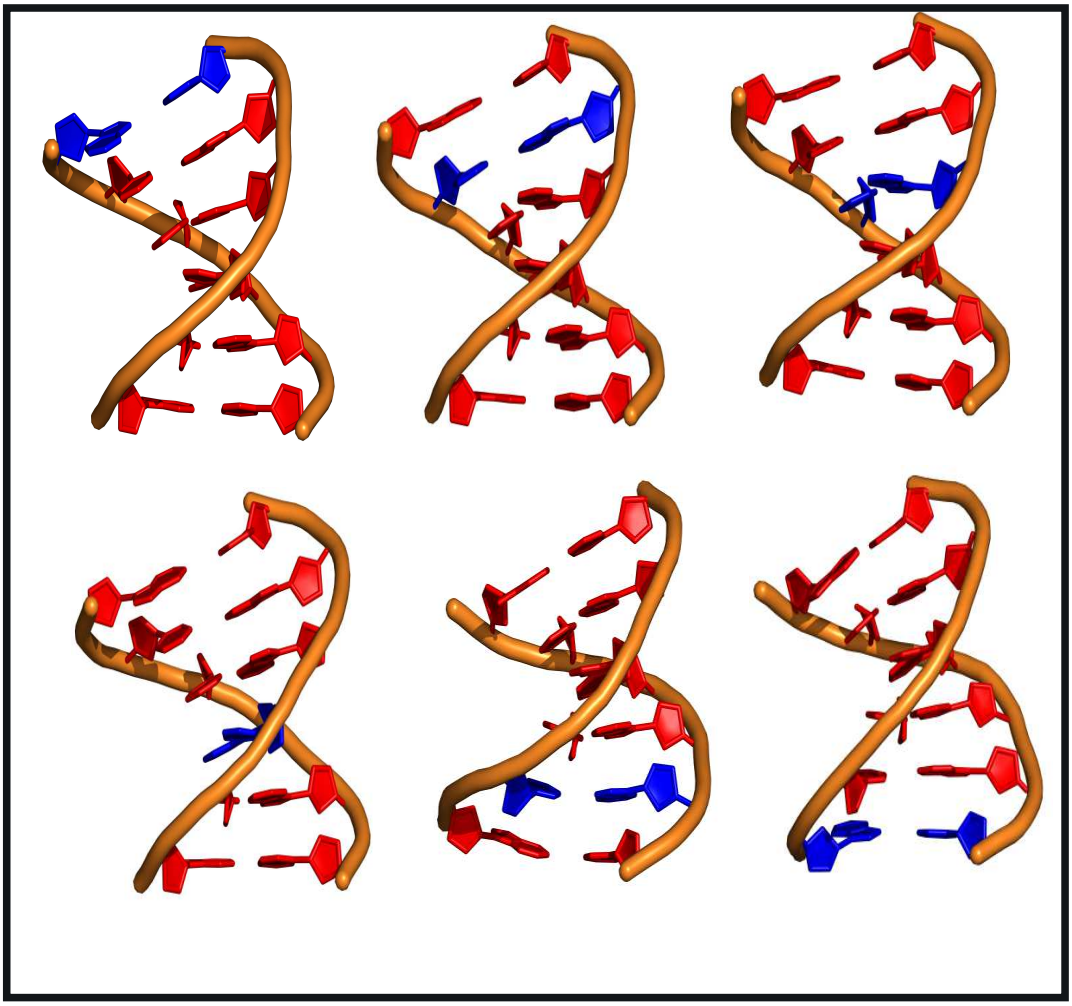

Figure S9. Duplexes of the type **1HG+5WC** with appreciable occupation probabilities at 300 K.

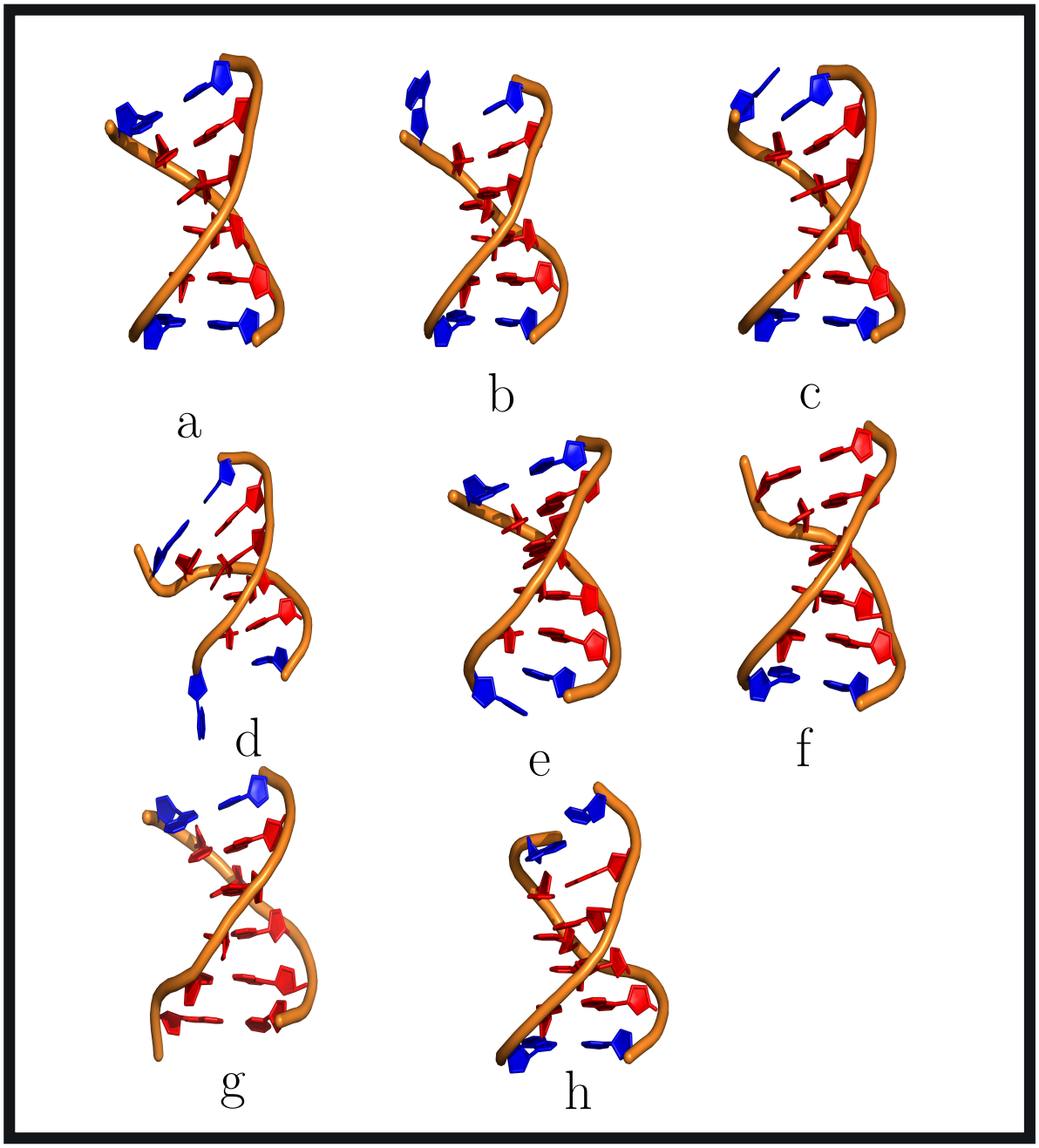

Figure S10. Constituent members of the lowest energy WC ensemble (a) WC duplex with HG pairing at both terminals. (b) WC duplex with HG pairing at both terminals, and A1 frayed out of the helix. (c) WC duplex with HG pairing at both terminals, and A1 stacked on top of T1. (d) WC duplex with HG pairing at both terminals, and A6 frayed out of the helix. (e) WC duplex with HG pairing at both terminals, and A6 stacked on top of T6. (f) WC duplex with HG pairing between A6 and T6. (g) WC duplex with HG pairing between A1 and T1. (h) WC duplex with HG pairing at both terminals, and T1 frayed out of the helix.

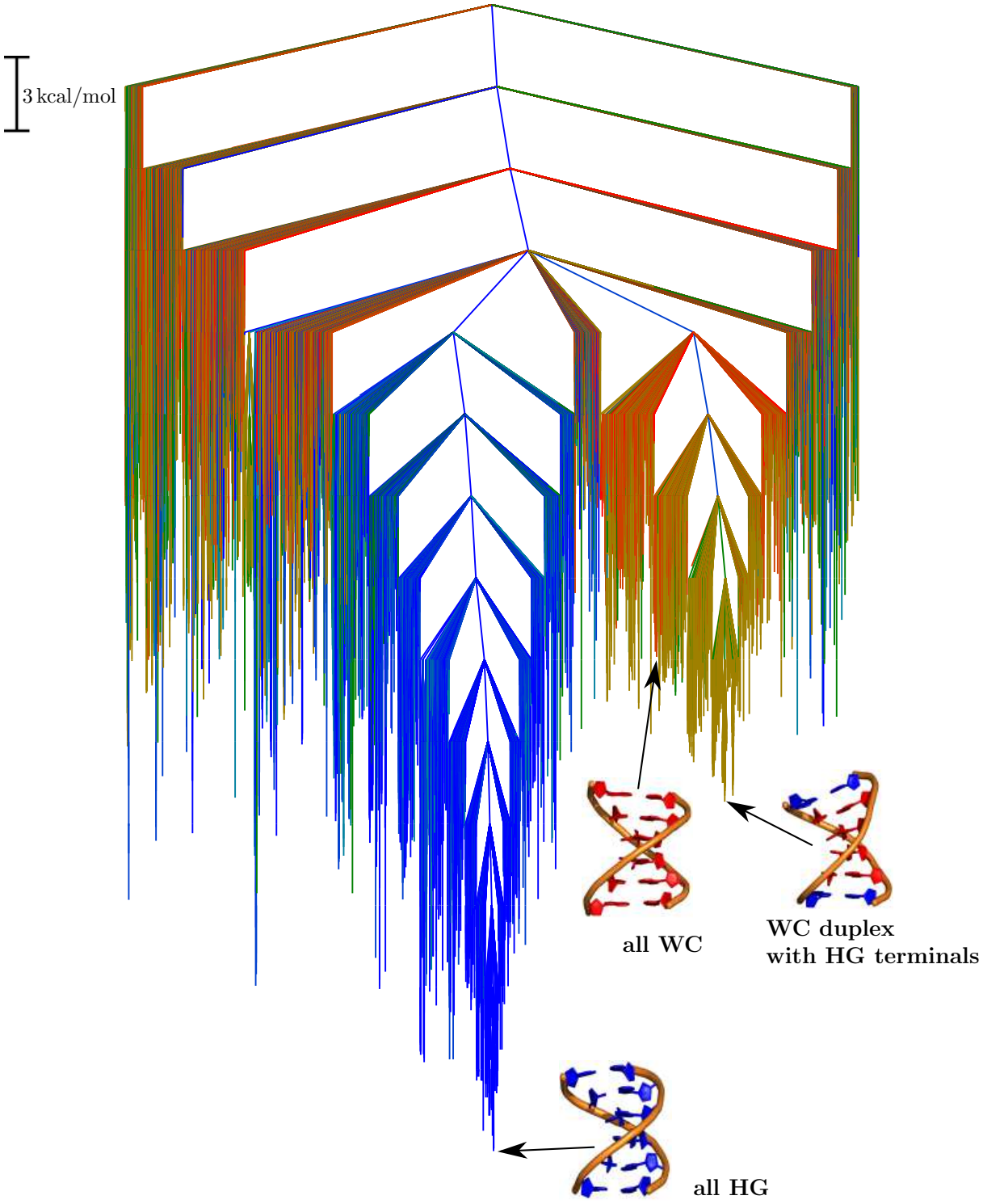

Figure S11. Potential energy disconnectivity graph. The color coding is the same as in Figure 3. Snapshots corresponding to the all HG, all WC, and WC duplex with HG terminals are shown.

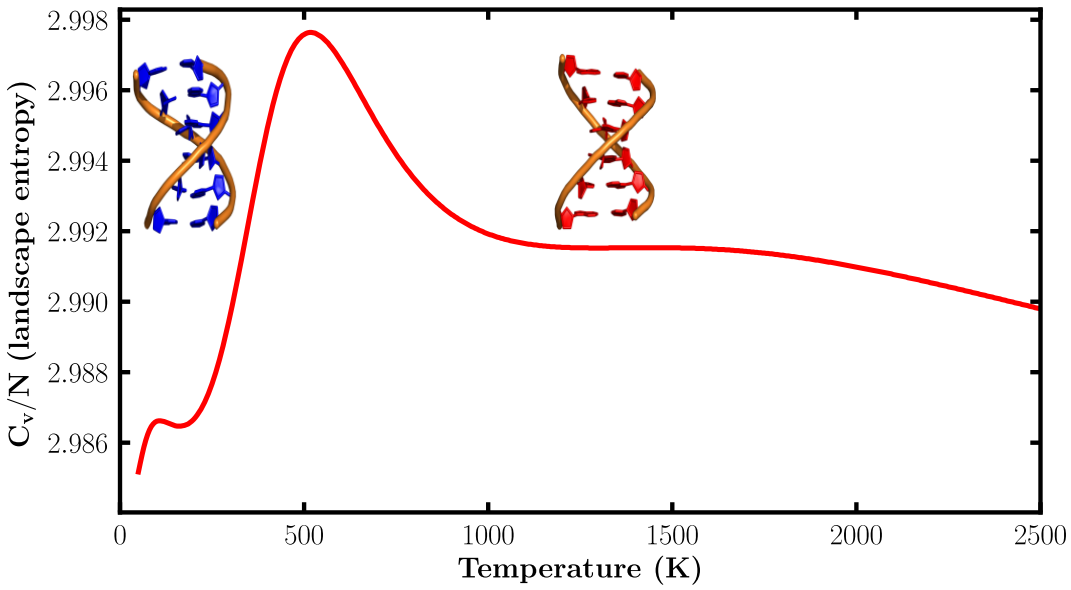

Figure S12. The normalized heat capacity profile excluding the vibrational degrees of freedom, illustrating the effect of landscape entropy. The WC duplex (red) has a higher landscape entropy due to more numerous high-lying potential energy minima, compared to the HG duplex (blue).

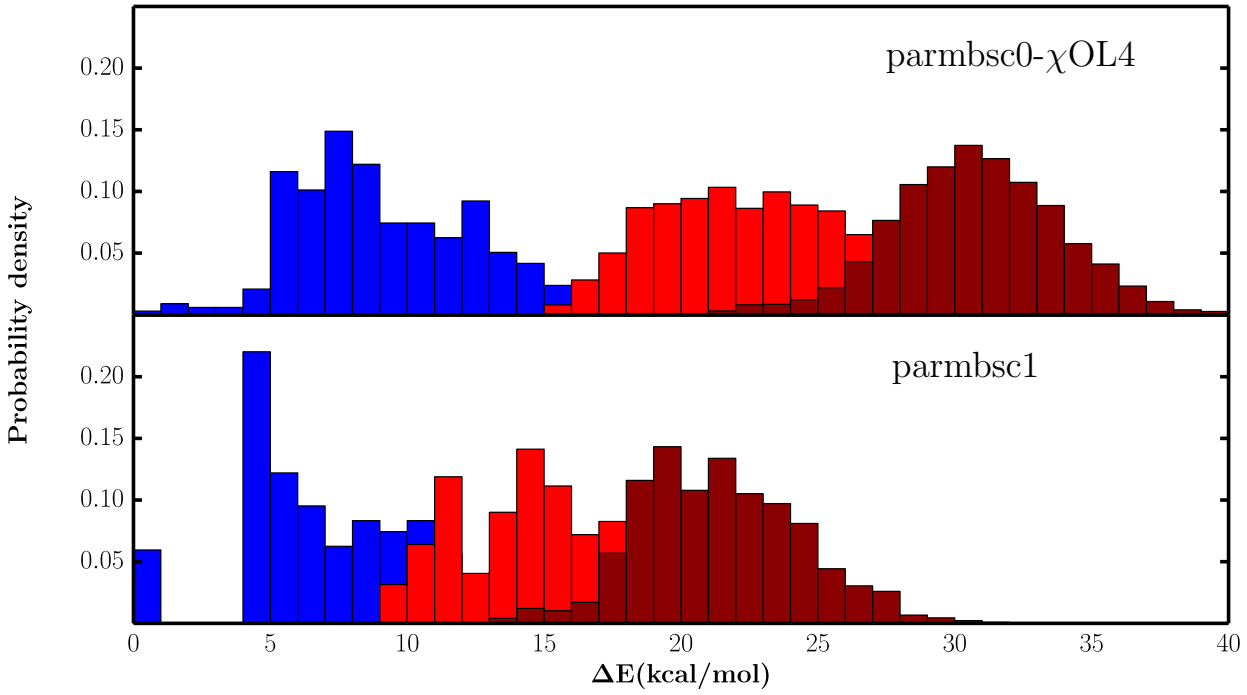

Figure S13. Distribution of the relative potential energies with respect to the global minimum computed using the χOL4, and the parmbsc1 force fields. The blue bars correspond to the ensemble comprising HG duplexes, the red bars correspond to the lowest energy WC duplexes having terminal HG base pairs, and the dark red bars correspond to the full WC conformations.

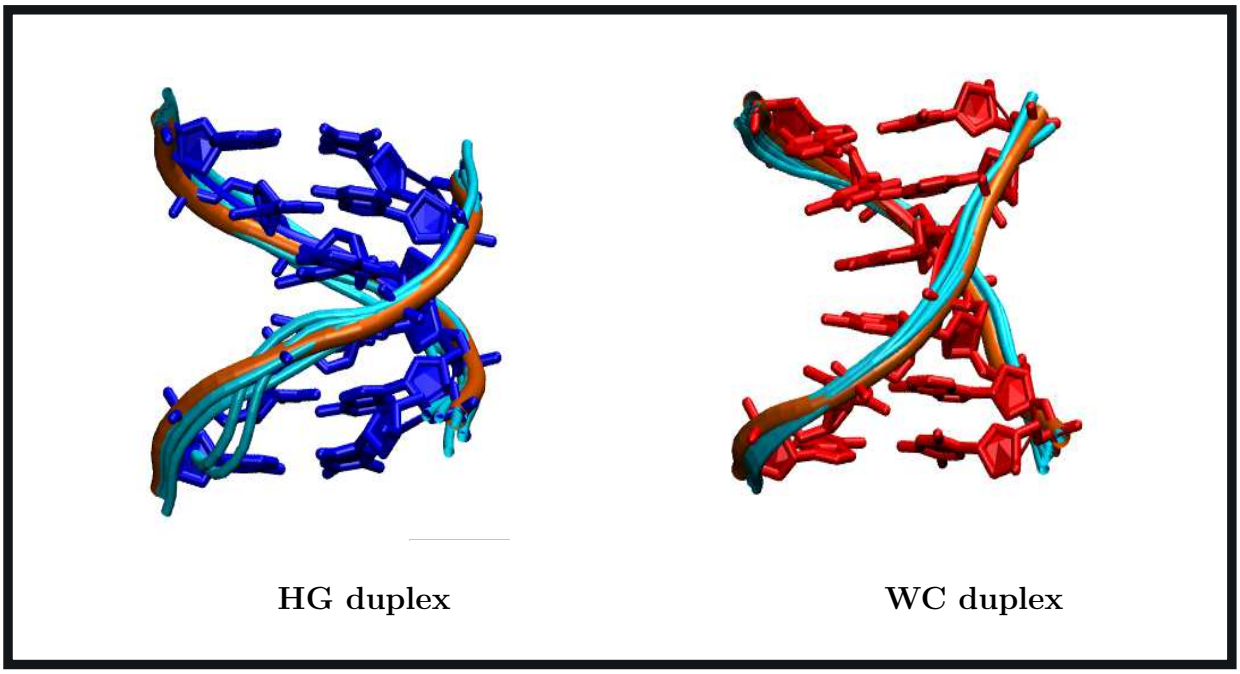

Figure S14. Several distinct potential energy minima located with the χOL4 parametrization (represented as cyan tubes) collapse to a single minimum on the parmbsc1 surface. Left: Minima from the HG ensemble, which collapse to the potential energy global minimum on the parmbsc1 surface. Right: WC duplex minima, which collapse to a single minimum on the parmbsc1 surface.

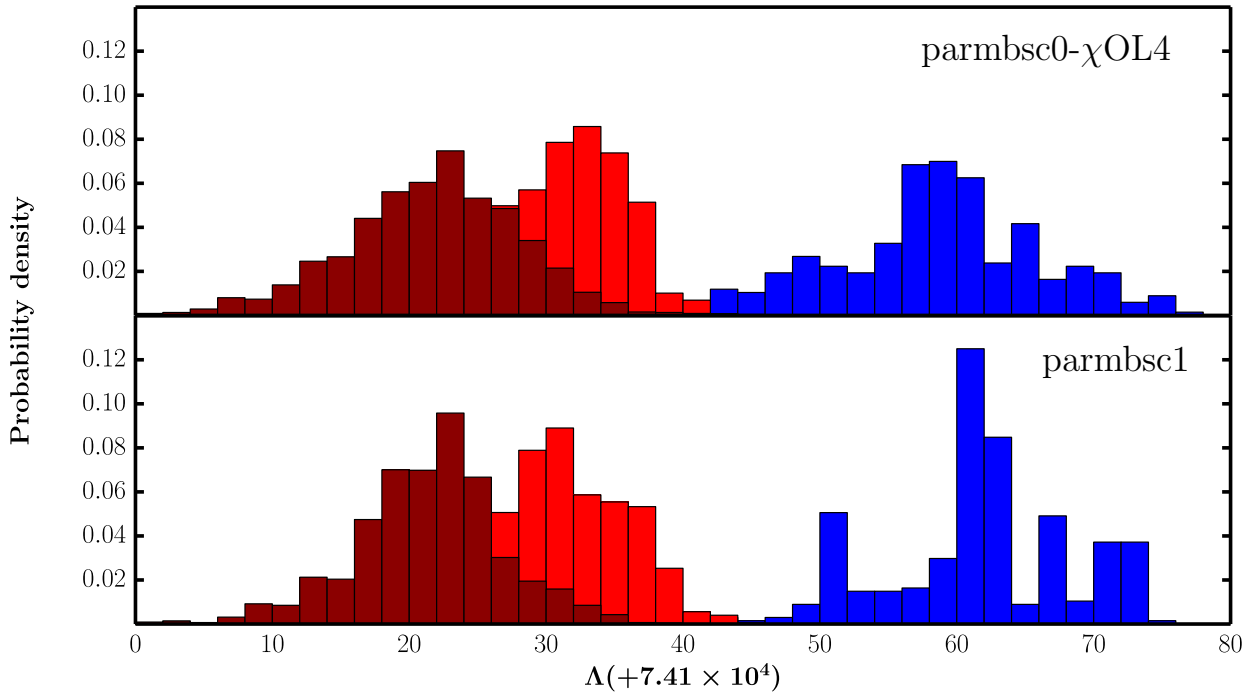

Figure S15. Distribution of the Λ values for minima comprising the free energy groups. The color coding is the same as Figure S13.

Reorganization of minima within the different funnels takes place after reoptimization with the parmbsc1 force field. This effect can be quantified by calculating the relative potential energies of minima, with respect to a fixed reference structure. For convenience, the free energy global minimum identified with the χOL4 parametrization is selected as the reference structure. As shown in Figure S16, the relative ordering of the different funnels is largely preserved. The dashed line provides a guide to the eye, and points which lie on this line correspond to minima whose relative positions (in terms of potential energy) with respect to the other minima in the database do not change upon reoptimization. Points that lie below the dashed line correspond to minima, which shift to higher ranks (in terms of potential energy) upon reoptimization. Points that lie above the dashed line correspond to minima, which are destabilized. In the graph, there are several points that have the same $\Delta E^{parmbsc1}$, but different $\Delta E^{parmbsc0\chi OL4}$. These points correspond to structures, which collapse to the same minimum on the parmbsc1 surface.

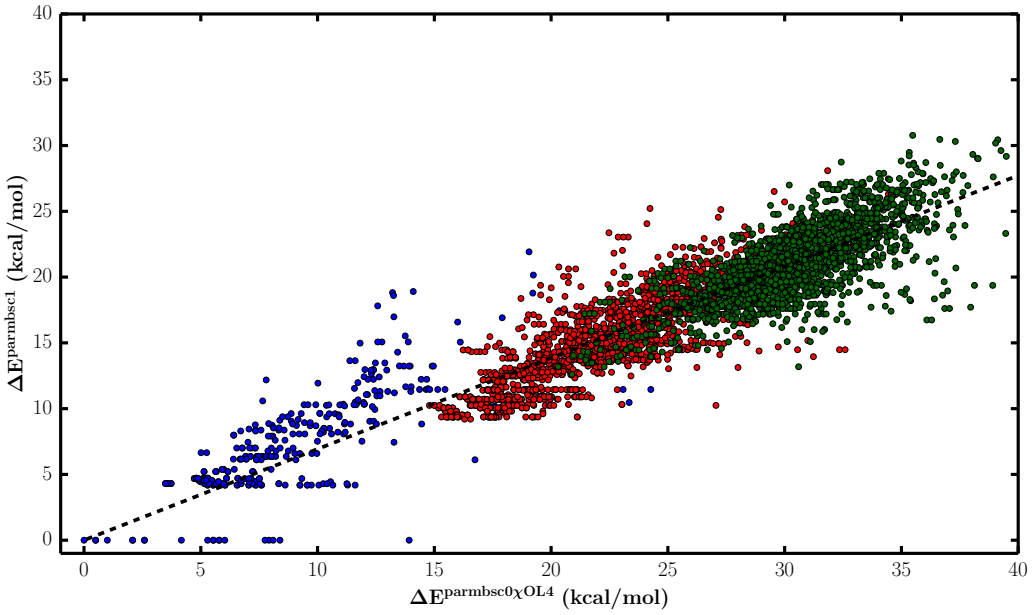

Figure S16. The blue circles correspond to minima from the HG duplex ensemble. The red circles correspond to minima constituting the lowest energy WC duplex ensemble. The green circles correspond to minima from the full WC duplex ensemble. The dashed line denotes a linear fit, and provides a guide to the eye. The larger number of high-lying potential energy minima in the WC ensemble, compared to the HG ensemble, illustrates the effect of landscape entropy (discussed in the main text, and Figure S12).

References

(1) Case, D. A.; Darden, T. A.; Cheatham, T.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I. AMBER 9. http://ambermd.org/, 2006.

(2) Malolepsza, E.; Strodel, B.; Khalili, M.; Trygubenko, S.; Fejer, S. N.; Wales, D. J. Symmetrization of the AMBER and CHARMM force fields. *J. Comput. Chem.* **2010**, *31*, 1402–1409.

(3) Pérez, A.; Marchán, I.; Svozil, D.; Šponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophys. J.* **2007**, *92*, 3817–3829.

(4) Krepl, M.; Zgarbova, M.; Stadlbauer, P.; Otyepka, M.; Banáš, P.; Koca, J.; Cheatham, T. E.; Jurecka, P.; Šponer, J. Reference simulations of noncanonical nucleic acids with different χ variants of the AMBER force field: quadruplex DNA, quadruplex RNA and Z-DNA. *J. Chem. Theory Comput.* **2012**, *8*, 2506–2520.

(5) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **2004**, *55*, 383–394.

(6) Onufriev, A.; Bashford, D.; Case, D. A. Modification of the Generalized Born Model Suitable for Macromolecules. *J. Phys. Chem. B.* **2000**, *104*, 3712–3720.

(7) Srinivasan, J.; Trevathan, M. W.; Beroza, P.; Case, D. A. Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects. *Theor. Chem. Acc.* **1999**, *101*, 426–434.

(8) Joung, S.; Cheatham, T. E. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **2008**, *112*, 9020–9041.

(9) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(10) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(11) Case, D. A.; Darden, T. A.; Cheatham, T.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I. AMBER 12. http://ambermd.org/, 2012.

(12) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin dynamics of peptides: The frictional dependence of isomerization rates of N-actylananyl-N′-methylamide. *Biopolymers* **1992**, *32*, 523–535.

(13) Wales, D. J. Discrete Path Sampling. *Mol. Phys.* **2002**, *100*, 3285–3305.

(14) Murrell, J. N.; Laidler, K. J. Symmetries of activated complexes. *Trans. Faraday Soc.* **1968**, *64*, 371–377.

(15) Wales, D. J. *Energy Landscapes*; Cambridge University Press, U.K., 2003.

(16) Wales, D. J. OPTIM: A program for optimising geometries and calculating pathways. http://www-wales.ch.cam.ac.uk/software.html.

(17) Liu, D.; Nocedal, J. On the Limited Memory Method for Large Scale Optimization. *Math. Program.* **1989**, *45*, 503–528.

(18) Henkelman, G.; Uberuaga, B. P.; Jònsson, H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* **2000**, *113*, 9901–9904.

(19) Henkelman, G.; Jònsson, H. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *J. Chem. Phys.* **1999**, *111*, 7010–7022.

(20) Trygubenko, S. A.; Wales, D. J. A Doubly Nudged Elastic Band Method for Finding Transition States. *J. Chem. Phys.* **2004**, *120*, 2082–2094.

(21) Munro, L. J.; Wales, D. J. Defect Migration in Crystalline Silicon. *Phys. Rev. B.* **1999**, *59*, 3969–3980.

(22) Dijkstra, E. W. A Note on Two Problems in Connexion with Graphs. *Numer. Math.* **1959**, *1*, 269–271.

(23) Wales, D. J. PATHSAMPLE: A program for generating connected stationary point databases and extracting global kinetics. http://www-wales.ch.cam.ac.uk/software.html.

(24) Strodel, B.; Whittleston, C. W.; Wales, D. J. Thermodynamics and Kinetics of Aggregation for the GNNQQNY Peptide. *J. Am. Chem. Soc.* **2007**, *129*, 16005–16014.

(25) Bauer, M. S.; Strodel, B.; Fejer, S. N.; Koslover, E.; Wales, D. J. Interpolation schemes for peptide rearrangements. *J. Chem. Phys.* **2010**, *132*, 054101.

(26) Wales, D. J.; Carr, J. M. A Quasi-Continuous Interpolation Scheme for Pathways Between Distant Configurations. *J. Chem. Theory Comput.* **2012**, *8*, 5020–5034.

(27) Strodel, B.; Wales, D. J. Free energy surfaces from an extended harmonic superposition approach and kinetics for alanine dipeptide. *Chem. Phys. Lett.* **2008**, *466*, 105–115.

(28) Hoare, M. R.; McInnes, J. J. Statistical mechanics and morphology of very small atomic clusters. *Faraday Discuss. Chem. Soc.* **1976**, *61*, 12–24.

(29) Hoare, M. R. *Advances in Chemical Physics*; John Wiley and Sons, USA, 1979; Vol. 40; pp 49–129.

(30) McQuarrie, D. A. *Statistical Mechanics*; University Science Books, 2000.

(31) Laidler, J. K.; Christine King, M. Development of transition-state theory.*J. Phys. Chem.* **1983**, *87*, 2657–2664.

(32) Eyring, H. J. The Activated Complex in Chemical Reactions.*J. Chem. Phys.* **1935**, *3*, 107–115.

(33) Evans, M. G.; Polyani, M. Some applications of the transition state method to the calculation of reaction velocities, especially in solution. *Trans. Faraday Soc.* **1935**, *31*, 875.

(34) van Kampen, N. G. *Stochastic processes in physics and chemistry*; Elsevier, Amsterdam, 1981.

(35) Wales, D. J. Some Further Applications of Discrete Path Sampling to Cluster Isomerization. *Mol. Phys.* **2004**, *102*, 891–908.

(36) Bulmer, M. G. *Principles of Statistics*; Dover, New York, 1979.

(37) Carr, J. M.; Wales, D. J. In *Latest Advances in Atomic CLuster Collisions: Structure and Dynamics from the Nuclear to the Biological Scale*; Connerade, J. P., Solov'yov, A., Eds.; Imperial College Press, London, 2008; pp 321–330.

(38) Jimènez, J.; Marzal, A. *Algorithm Engineering: 3rd International Workshop, WAE'99, London, UK, July 1999*; Springer, Berlin, 1999; pp 15–29.

(39) Wales, D. J. Calculating Rate Constants and Committor Probabilities for Transition Networks by Graph Transformation. *J. Chem. Phys.* **2009**, *130*, 204111.

(40) Fitchthom, K. A.; Weinberg, W. Theoretical foundations of dynamical Monte Carlo simulations. *J. Chem. Phys.* **1990**, *95*, 1090–1096.

(41) Fain, B. Theory of Rate Constants: Master Equation Approach. *J. Stat. Phys.* **1981**, *25*, 475–489.

(42) Stevenson, J. D.; Wales, D. J. Analysing kinetic transition networks for rare events. *J. Chem. Phys.* **2014**, *141*, 041104.

(43) Carr, J. M.; Wales, D. J. Folding Pathways and Rates for the Three-Stranded beta-sheet Peptide Beta3s Using Discrete Path Sampling. *J. Phys. Chem. B* **2008**, *112*, 8760–8769.