Supplementary Materials for "Multi-scale coarse-graining for the study of assembly pathways in DNA-brick self assembly"

Pedro Fonseca,[1,*] Flavio Romano,[2] John S. Schreck,[3] Thomas E. Ouldridge,[4] Jonathan P. K. Doye,[5] and Ard A. Louis[1]

[1] *Rudolf Peierls Centre for Theoretical Physics, University of Oxford, 1 Keble Road, Oxford, OX1 3NP, United Kingdom*
[2] *Dipartimento di Scienze Molecolari e Nanosistemi, Universitá Ca' Foscari, Via Torino 155, 30172 Venezia Mestre, Italy.*
[3] *Department of Chemical Engineering, Columbia University, 500 W 120th St, New York, NY 10027, USA*
[4] *Department of Bioengineering, Imperial College London, London, SW7 2AZ, United Kingdom*
[5] *Physical and Theoretical Chemistry Laboratory, Department of Chemistry, University of Oxford, South Parks Road, Oxford, OX1 3QZ, United Kingdom*

(Dated: Tuesday 20th February, 2018)

S.I OXDNA MODEL

oxDNA [1, 2] is a top-down model developed to study self-assembly processes associated with DNA nanotechnology. oxDNA makes use of three coarse-grained units per nucleotide plus the full orientation of the base to account for angle-dependent interactions. In Fig. S1 we show a schematic representation of the potentials: pairwise interactions between the nucleotides represent base-pairing between complementary Watson-Crick bases, stacking, coaxial stacking, and chain connectivity; in addition, all nucleotides also interact through repulsive excluded volume. oxDNA potentials were parametrised to reproduce the hybridisation transitions of duplexes as predicted by Santa Lucia's two-state nearest-neighbour (NN) model [3, 4], and some fundamental structural and mechanical properties of single-stranded DNA and duplex DNA. Since its conception, oxDNA has successfully predicted functional properties of many biophysical and nanotechnological systems beyond its original parametrisation. An extensive overview of the model can be found in references [1, 2], and the source code and full list of publications can be accessed in the website dna.physics.ox.ac.uk.

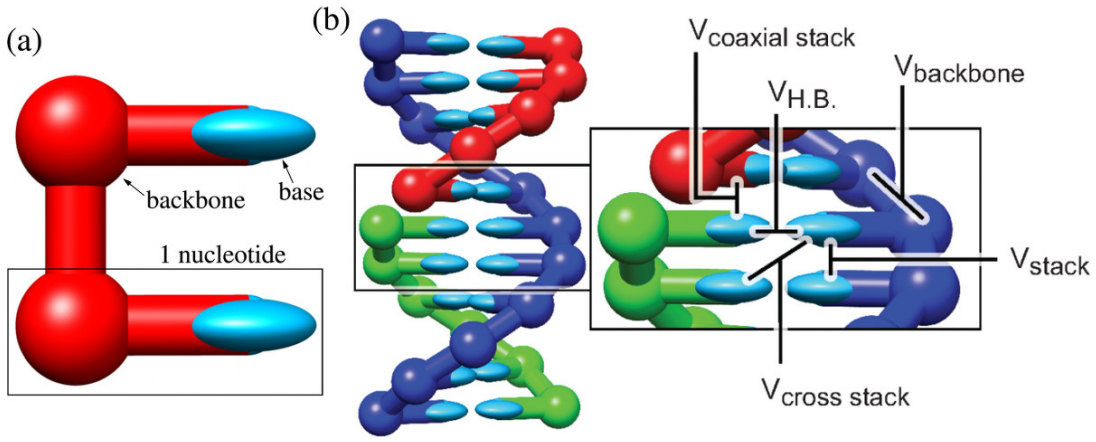


FIG. S1. Representation of the oxDNA model. The backbone sites are represented as spheres; the bases are represented by ellipsoids emphasizing the orientation dependence of the interactions. (a) Representation of rigid nucleotides. (b) Representation of three hybridized stands. The pairwise interactions of the model are depicted in the inset. All nucleotides also interact through repulsive excluded volume interactions. Reproduced with permission of J. Schreck *et al.* [5]

* pedro.fonseca@physics.ox.ac.uk

S.II SIMULATION METHODS

A Virtual Move Monte Carlo (VMMC)

Standard Monte Carlo (MC) methods, in which sequential one-particle moves are attempted, can be very inefficient to sample dilute systems with strong short-ranged interactions. This is the case for the systems studied throughout this work where moves in which a single nucleotide is displaced (the nucleotide is the basic constituent of DNA within oxDNA) suffer from low acceptance rates, resulting in strongly suppressed collective motion of strands and slow sampling of diffusion. Simulations can be made more efficient by using the VMMC algorithm proposed by Whitelam and Geissler [6], which allows for collective moves involving clusters of nucleotides.

To generate a trial move the algorithm first selects a single nucleotide, and attempts a random seed move according to standard Metropolis rules. Additional nucleotides in the neighbourhood of the selected nucleotide are then added, with probabilities that depend on the resultant energy changes, forming a co-moving cluster. Trial moves are then accepted (or rejected) with a probability that ensures correct sampling from the canonic ensemble. In the simulations performed through this work, seed moves were:

- Rotation of a nucleotide about its backbone site, with the axis chosen uniformly on the unit sphere and the angle drawn from a normal distribution with a mean of zero and a standard deviation of 0.22 radians.
- Translation of a nucleotide, where the displacement along each Cartesian axis is drawn from a normal distribution with a mean zero and a standard deviation of 1.28 Å.

B Single Histogram Reweighting

In this work, we use oxDNA extensively to calculate tile-association free energies at $T = 50°\mathrm{C}$, in the temperature range we expected the assembly transition to be (the experimental peak of the rate-of-assembly curve is at $T = 53°\mathrm{C}$). To avoid having to repeat simulations for temperatures other than $T = 50°\mathrm{C}$, we follow earlier work on oxDNA [7] and use the single histogram reweighting method, first introduced by Ferrenberg and Swensden [8], to extrapolate our results to other temperatures.

To use this method, the states of the system sampled in the simulations are grouped according to their energy E, and value of some observable of interest A, and the distribution $p(A, E; T_0)$ is produced during the run (all runs were performed at $T_0 = 50°\mathrm{C}$). From the distribution $p(A, E; T_0)$, we can derive the density of states $p(A, E; T_0) \propto \Omega(A, E)e^{-\beta_0 E}$, where $\beta_0 = 1/k_B T_0$. The density of states can then be used to estimate values of A at a different temperature T:

$$\langle A(T)\rangle = \frac{\int\int A\Omega(A,E)e^{-\beta E}dEdA}{\int\int \Omega(A,E)e^{-\beta E}dEdA}, \tag{1}$$

where $\beta = 1/k_B T$. Eq. (2) can be rewritten as:

$$\langle A(T)\rangle = \frac{\int\int Ap(A,E;T_0)e^{-(\beta_0-\beta)E}dEdA}{\int\int p(A,E;T_0)e^{-(\beta_0-\beta)E}dEdA}, \tag{2}$$

S.III MISBOND FORMATION CALCULATIONS

In this section we estimate an upper bound on the equilibrium probability of misbonding at binding sites at the interface of the assembly. We also estimate the effect of misbonds on assembly dynamics. In the following sections we will use the term "binding site type" to differentiate between binding sites that have a different set of first neighbours.

We start our analysis by observing that for all binding site types, the cumulative length of binding domains is at most 42-bases long (see S.VI, environments 27-32). For most binding site types however, the cumulative length of binding domains is less than 42 bases (see S.VI, environments 1-26), but for simplicity we assume it to be 42 bases for all binding site types. Note that by using this simplification, we overestimate the stability of misbonds and thus, also the probability of misbonding at equilibrium.

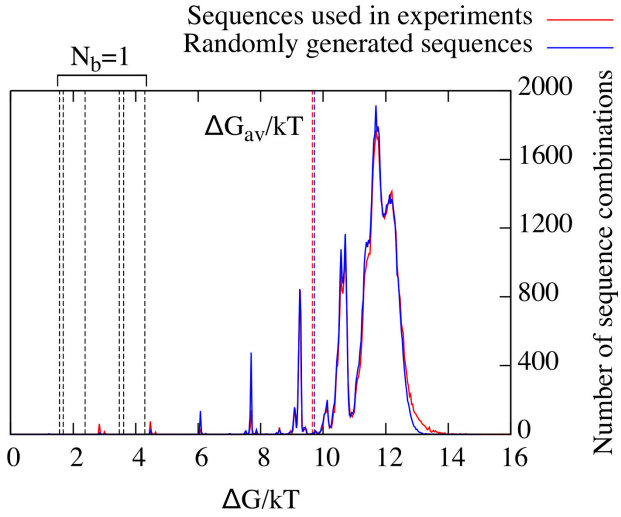


FIG. S2. Effective free energies of interaction determined for pairs of sequences (each 42-base long) that were not designed to interact, estimated using the NN model [3], at 100nM per strand species and $T = 50°$C. Red curve: calculation using sequences used in experiments [9]. Blue curve: calculation using randomly generated sequences. Dashed coloured lines: contact free-energies ($\Delta G_{\rm av}/kT$) correspondent to the average probability of bonding between strands (taken over all combinations of sequences). Also shown for comparison (dashed black lines) are the free energies of tile association determined using oxDNA for binding sites with a single domain available for binding ($N_b = 1$), also at 100 nM per strand species and $T = 50°$C.

We further simplify our analysis by assuming that, instead of being distributed through multiple neighbouring tiles, binding domains can be treated as a single continuous segment with 42 bases. Again, we note that this simplification overestimates the probability of misbonding at equilibrium - as can be seen in the free-energy profiles of tile-association (see S.VI, environments 7-32) there is a significant entropic penalty associated with initiating a second or subsequent domains, which is not accounted for if binding domains are considered to be concatenated in a single-stranded segment. We emphasise that our goal is to estimate an upper bound on the equilibrium probability of misbonding rather than obtaining a precise value.

A Calculation of misbond free energies

We generate a representative set of 42-base sequences for the sites by using the sequences of the 334 tiles themselves. The challenge then reduces to estimating the binding free energy of pairs of sequences (each 42-base long) from the ensemble used in experiments [9]. We exclude from the calculations sequences that share a complementary domain designed to hybridise in the assembled structure, which leaves a total of $334 \times 333/2 - 616 = 54\,995$ different combinations of 2 sequences. For each pair of sequences, we chose one (A) to represent a binding site in the assembly, and the other (s) to represent an incoming strand that may incorrectly form bonds with the binding site. We then use use a simple algorithm that finds complementary sections between both sequences that are at least 2 bases long. For each of the complementary sections found, and all pairs of such sections, we calculate the free energy of the base-pairing configuration using the average-base parameterisation of the NN model [4], together with internal loop parameters and asymmetry corrections [4]. To each different complementary section between A and s (or two simultaneously), we add a label η to which we call the "contact type" between A, s.

For each contact type η between a pair of sequences A, s the equilibrium probability of contact type η being formed with relation to the reference state $\emptyset$ (where no contacts are formed) is given by:

$$p_\eta(A,s)/p_\emptyset(A,s) = e^{-\Delta\beta G_\eta^\ominus(A,s)}\frac{[s]}{1\mathrm{M}} = \kappa_\eta(A,s). \tag{3}$$

Using $p_\emptyset(A,s) + \sum_\eta p_\eta(A,s) = 1$, we determine the probability of *any* contact being formed between the pair A, s, which we denote by $p(A,S) \equiv \sum_\eta p_\eta(A,s)$:

$$p(A,s) = \frac{\sum_\eta \kappa_\eta(A,s)}{1+\sum_\eta \kappa_\eta(A,s)}. \tag{4}$$

Finally, we determine an overall free energy of interaction for the pair A, s at the concentration $[s]$, $\Delta G^{\ominus}(A,s)$, by solving the equation:

$$p(A,s)/p_{\emptyset}(A,s) = \kappa(A,s) = e^{-\beta\Delta G^{\ominus}(A,s)}[s]/(1\mathrm{M}). \tag{5}$$

In Fig. S2 we show the resulting distribution of effective free energies of interaction ($\Delta G^{\ominus}(A,s)$) obtained by repeating the calculation for the different 54995 combinations of two sequences from the set used in the experiments (red curve). Also shown in the figure (blue curve) is the distribution of effective free energies of interaction determined using a similar calculation, but this time using a batch of strands whose sequences were generated at random (with the constraint of observing the complementary pattern needed for the target structure to assemble).

B Estimation of the effect of misbonds on equilibrium and dynamic properties of assembly

To estimate the effect of misbonds on assembly at equilibrium, we first consider the average probability at equilibrium that a given strand with an incorrect sequence binds to a binding site of the assembly. In this calculation, we take the average over binding probabilities for all combinations of representative binding site and tile sequences: $\bar{p} = \sum_{A,S} p(A,s)/M$, with $M = 54995$ being the total number of different pairs of sequences. Next, we calculate the interaction free energy $\Delta G^{\ominus}_{\mathrm{av}}$ corresponding to $\bar{p}$ by solving the equation:

$$\frac{\bar{p}}{(1-\bar{p})} = \kappa_{\mathrm{av}} = e^{-\Delta G^{\ominus}_{\mathrm{av}}/RT}[s]/(1\mathrm{M}). \tag{6}$$

In Fig. S2 we show (dashed coloured lines) the free energies $\Delta G^{\ominus}_{\mathrm{av}}$ determined using sequences taken from the experimental batch (red) and sequences generated randomly (blue).

1 Misbonding model

We are now in a position to incorporate this overestimate of typical mismatch binding propensity into the self-assembly model presented in the main text. We assume that for each binding site, there are 333 incorrect strands, all present at concentration $[s]$, which all bind with a bimolecular rate constant k^0_+, yielding an overall misbonding rate

$$k^{\mathrm{mm}}_+ = 333 \times k^0_+[s], \tag{7}$$

and dissociate at a rate

$$k^{\mathrm{mm}}_- = k^0_+ e^{\beta\Delta G^{\ominus}_{\mathrm{av}}} \times (1\mathrm{M}). \tag{8}$$

This is the simplest kinetic approach that captures the overall tendency of misbonds to form in equilibrium, whilst averaging over the sequence details.

2 Overall effect of misbonding on assembly at equilibrium

Binding sites in the assembly can be either unoccupied or occupied by either a correctly bonded tile or a strand with an incorrect sequence. To estimate the relative probabilities in equilibrium we use the rates in Eqs. 7, 8 and Eqs. 1 and 2 of the main text. For the relative probabilities of the misbonded and unoccupied states, we have

$$p_{\mathrm{misbond}}/p(\emptyset'') = k^{\mathrm{mm}_+}/k^{\mathrm{mm}_-}, \tag{9}$$

For the correct tile, from Eq. 2 of the main text, we have

$$p_{(\mathrm{correct})}/p(\emptyset'') = \exp\left(-\beta\Delta G^{\ominus}(\vec{x},\vec{y})\right)[s]/1\mathrm{M}, \tag{10}$$

which depends on the nature of the states $\vec{x}$ and $\vec{y}$ with empty and full sites, respectively.

In Fig. S3 we show the ratio of the equilibrium binding probabilities for correct bonding and misbonding, incorporating the variation over site type into the estimate of correct bonding.

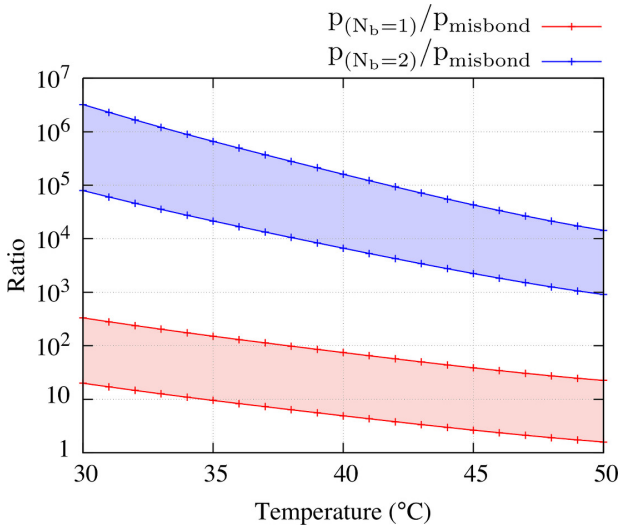


FIG. S3. Ratio between the equilibrium probability of bonding for correctly bonded tiles with with one ($p_{(N_b=1)}$) or two ($p_{(N_b=2)}$) binding domains, and the equilibrium probability of misbonding ($p_{misbond}$) as a function of temperature. The solid curves limiting the coloured areas correspond to bonding probabilities for the least strongly bonded (bottom curve) and most strongly bonded (top curve) correct tiles.

3 Effect of misbonding on assembly dynamics

To estimate the effect of misbonding on dynamics, we use a simplified variant of our assembly model, where tiles with the correct sequence can associate the assembly with rate $k_+^0[s]$ but are not allowed to leave the assembly once bound (which is the typical behaviour we observe at lower temperatures ($T \lesssim 43°C$)), and misbonding is described explicitly using the rates determined above (Eq. 7 and Eq. 8).

We perform our calculations using two sub-variants of the model. In the first we allow misbonds to occur (as described above) in addition to correct bonding; whereas in the second sub-variant only correct bonding is allowed to occur. We define t as the average time needed for the assembly to reach full completion in the first sub-variant of the model, i.e. to reach a state where 334 tiles are correctly bonded; and define t_0 as the average completion time observed using the second sub-variant of the model. In Fig. S4 we show the ratio between average completion times observed in the two sets of simulations as a function of temperature.

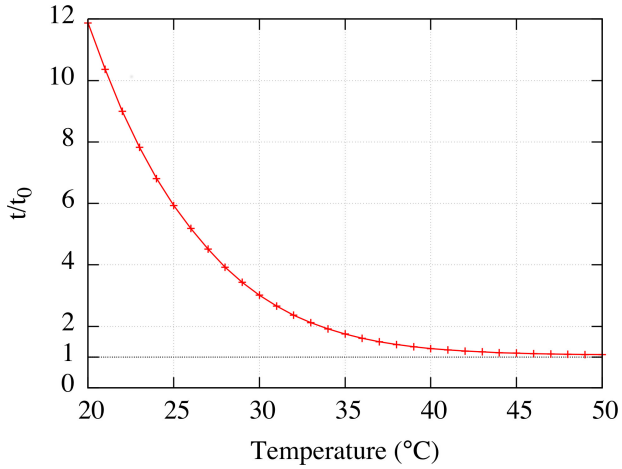


FIG. S4. Ratio between average completion time when misbonds are explicitly described in the model (t), and average completion time when misbond are not allowed to occur (t_0).

S.IV CRITICAL NUCLEUS STRUCTURE

As described in the main text, we calculated the probability of finding a critical nucleus with a certain number of bonded domains E. This is shown in Fig. S5 for the critical temperature $T = 51.64°$C where the critical nucleus size is $N^* = 50$. While the structure with the maximum $E_{\max} = 85$ has the lowest enthalpy, it is not the most common value of E because it also has a lower entropy. Instead the peak in probability is around $E = 83$.

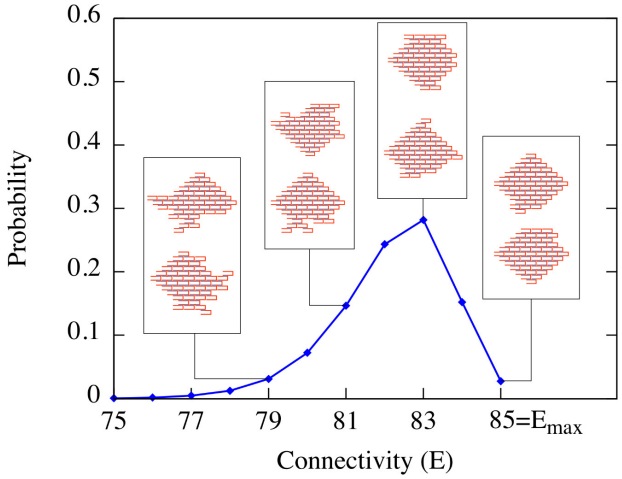


FIG. S5. Probability distribution for the number of bonded domains (E) between tiles for structures with $N = 50$ components (critical size,critical temperature - $T = 51.64°$C). Above the curve are depicted examples of assembly states, representative of states $(N^* = 50, E)$ with different connectivities.

S.V RANDOMLY GENERATED SEQUENCES - PROBABILITY DISTRIBUTION OF THE LENGTH OF THE LONGEST COMPLEMENTARY SECTION BETWEEN 2 SEQUENCES

In this section we determine the probability distribution for the length of the longest complementary section between two randomly generated sequences, each 42-base-long. For this purpose, we use an algorithm that generates pairs of 42-base-long random sequences, checks for complementary sections between the two sequences, and stores the length (in number of bases) of the longest complementary section found. In Fig. S7 we show the resulting probability distribution, obtained by using the algorithm to produce and analyse 10^7 pairs of random sequences.

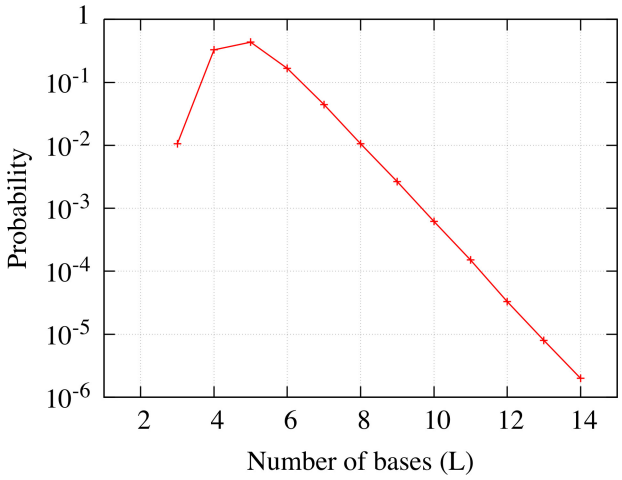


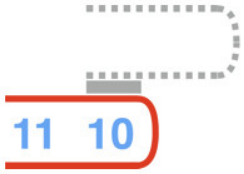
FIG. S6. Probability distribution for the length (L) of the longest complementary section between two randomly generated (42-base-long) sequences.

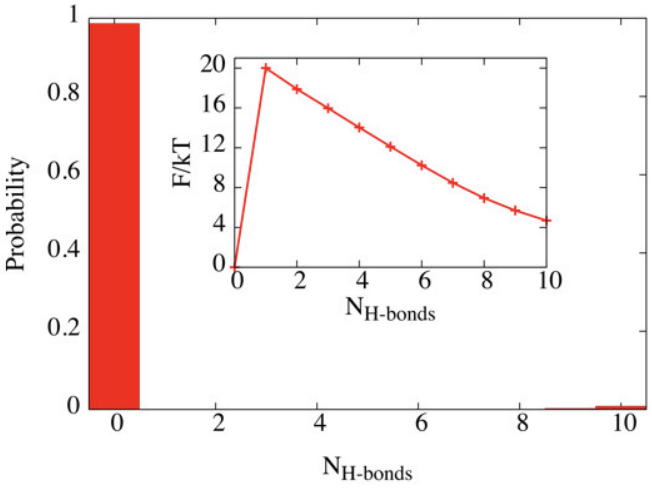
S.VI FREE-ENERGY PROFILES OF TILE-ASSOCIATION

In this section we detail the set of local environments used in the implementation of the kinetic model, and the respective free-energy profiles of tile-association calculated using oxDNA, at $T = 50°C$ and 100 nM per tile species. The numbers above each diagram (top left), are used to identify each local environment (also called environment index in the main text). In the schemes on the left, tiles coloured red represent the assembly fragment that contains the first neighbours of the binding site (which is colored gray). Rectangles coloured blue and grey represent duplex sections in the assembly fragment, and between assembly fragment and binding tile, respectively. The numbers in the schemes (coloured blue) represent the length (in nucleotides) of domains positioned above and bellow the numbers. In the plots on the right, $N_{H\text{-}bonds}$ denote the the number of base-pairs formed between tile and assembly fragment.

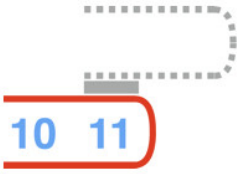
A Local environments with a single binding domain

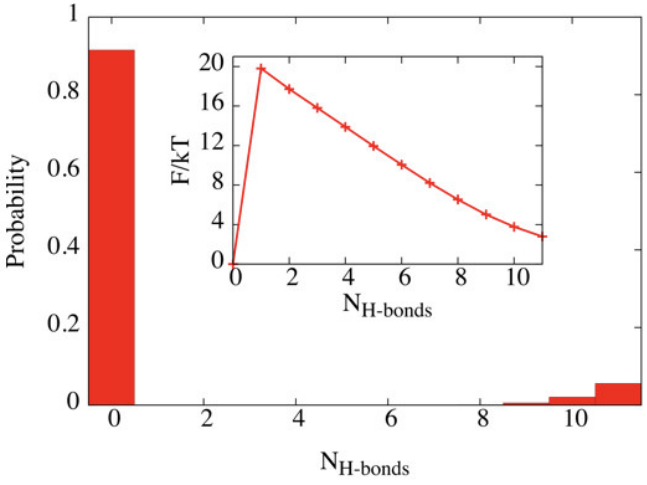
1



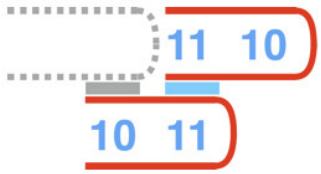


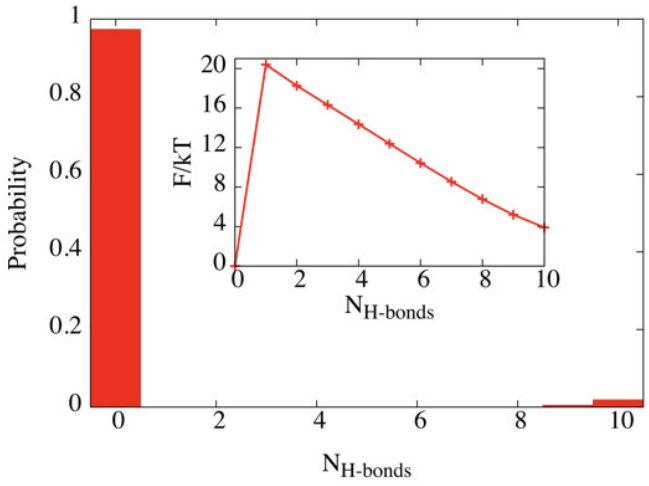
2



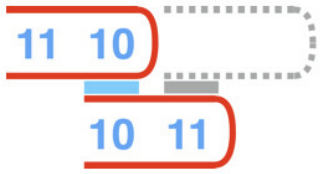


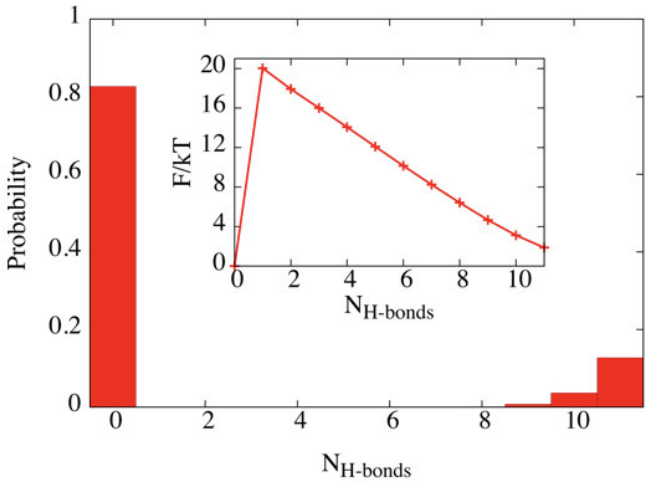
3


11 10
10 11

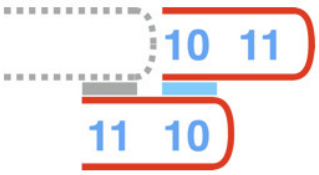

Probability
F/kT
N_H-bonds
N_H-bonds

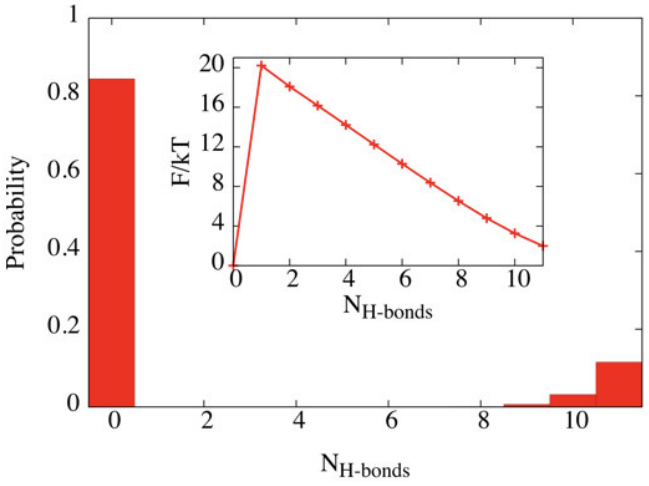
4


11 10
10 11

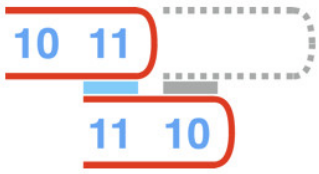

Probability
F/kT
N_H-bonds
N_H-bonds

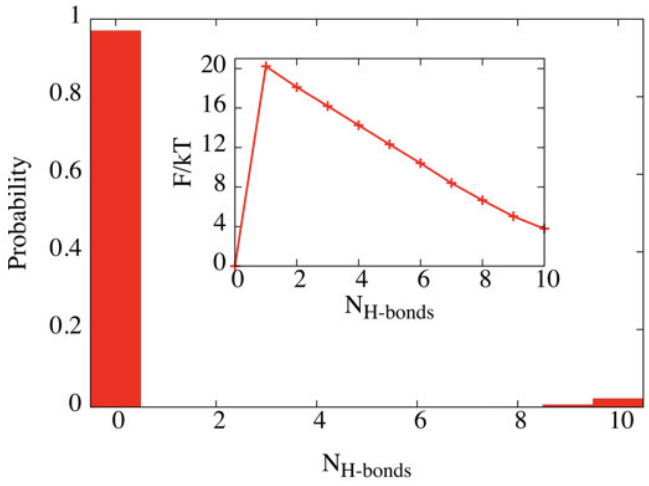
5


10 11
11 10


Probability
F/kT
N_H-bonds
N_H-bonds

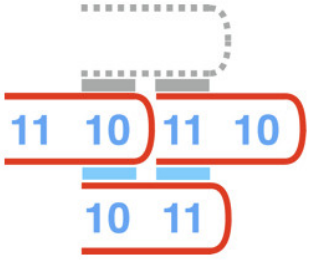
6

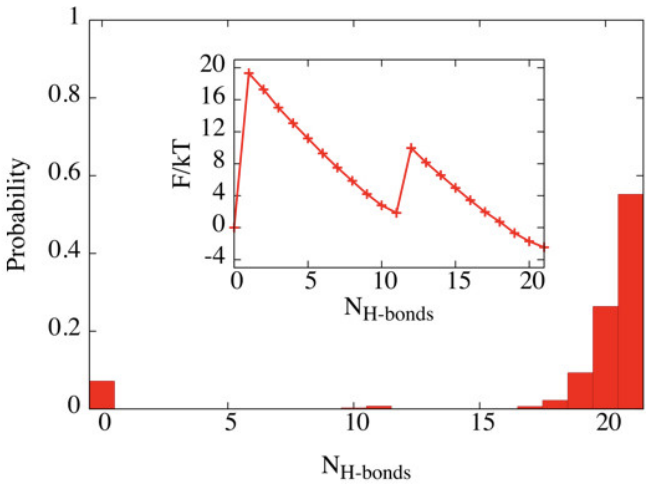




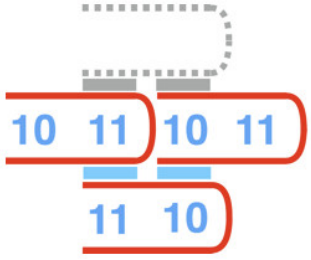
B Local environments with two binding domains

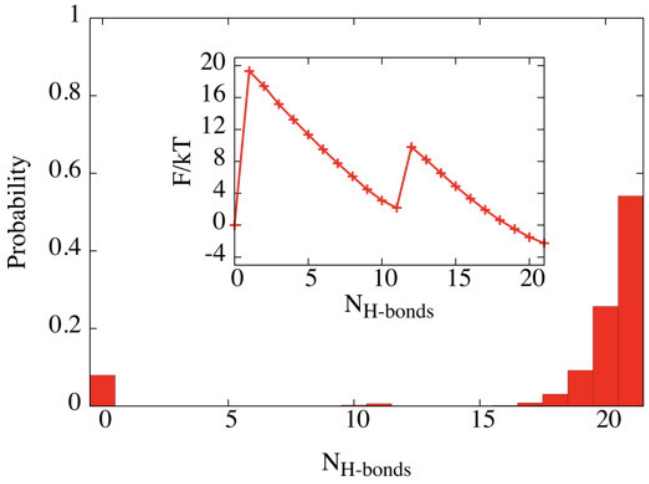
7



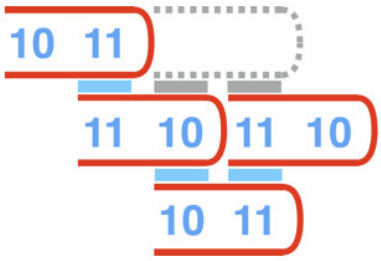


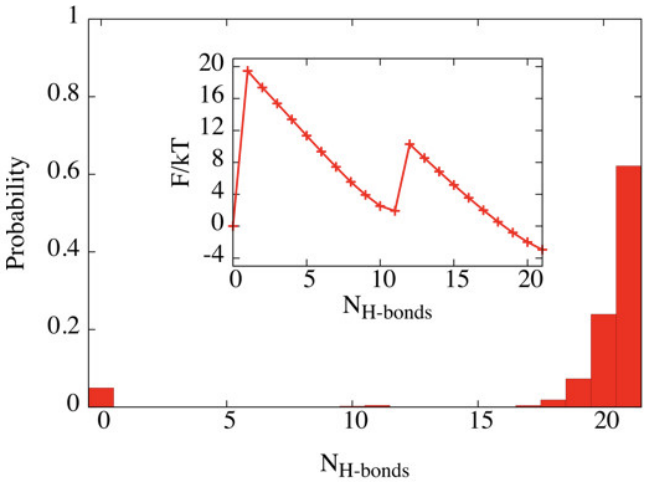
8



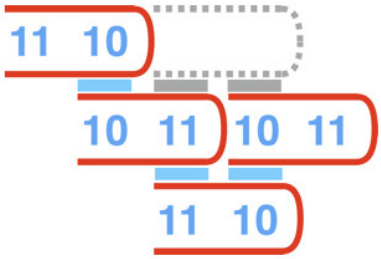


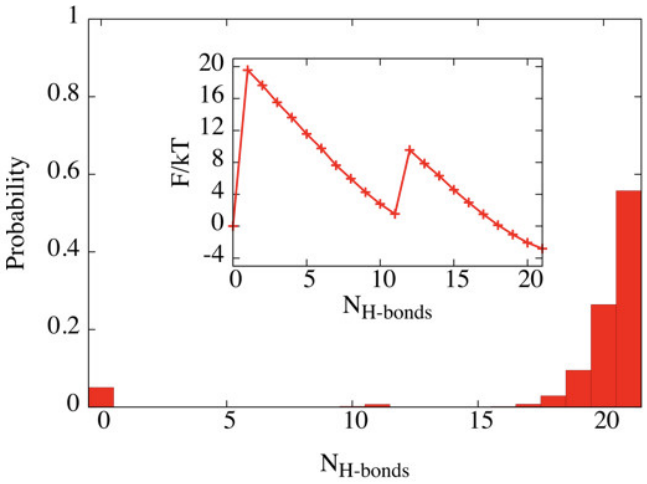
9


10 11
11 10 11 10
10 11

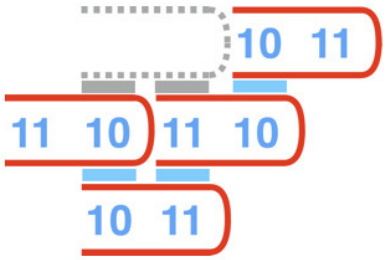

Probability
F/kT
N H-bonds
N H-bonds

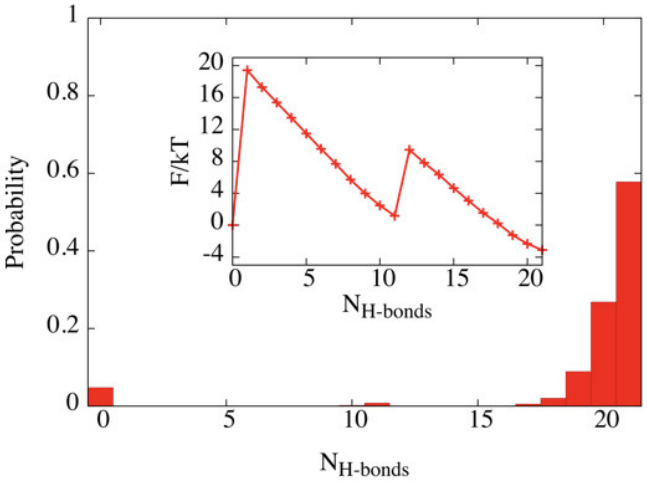
10


11 10
10 11 10 11
11 10

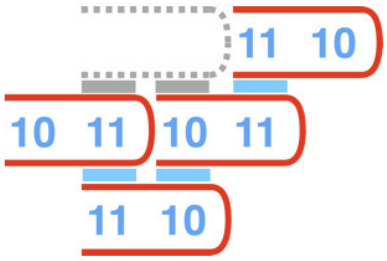

Probability
F/kT
N H-bonds
N H-bonds

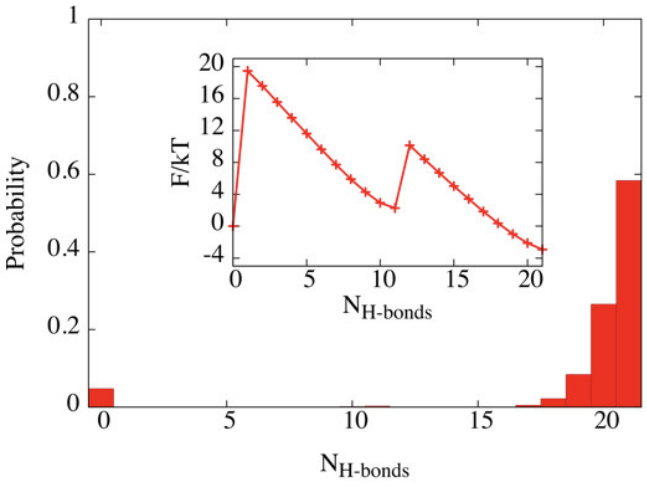
11


10 11
11 10 11 10
10 11

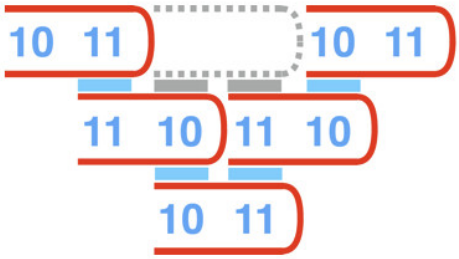

Probability
F/kT
N H-bonds
N H-bonds

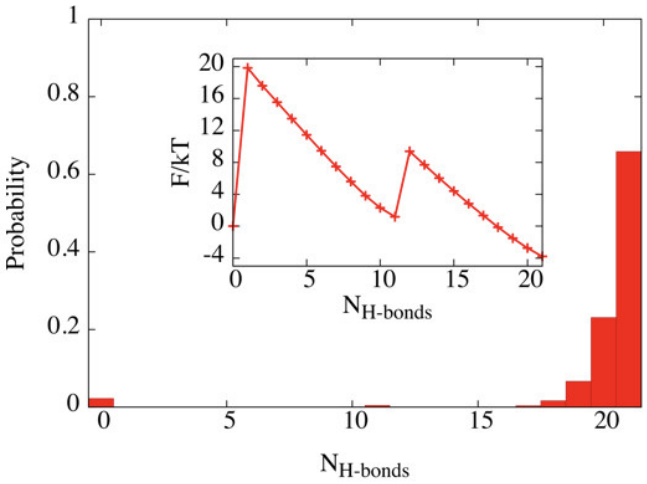
12


11 10
10 11 10 11
11 10

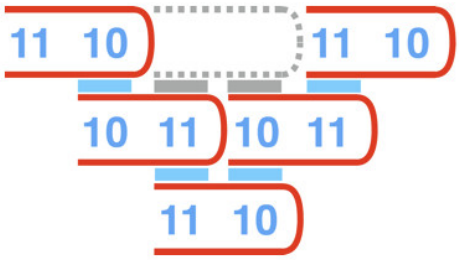

Probability
F/kT
N H-bonds
N H-bonds

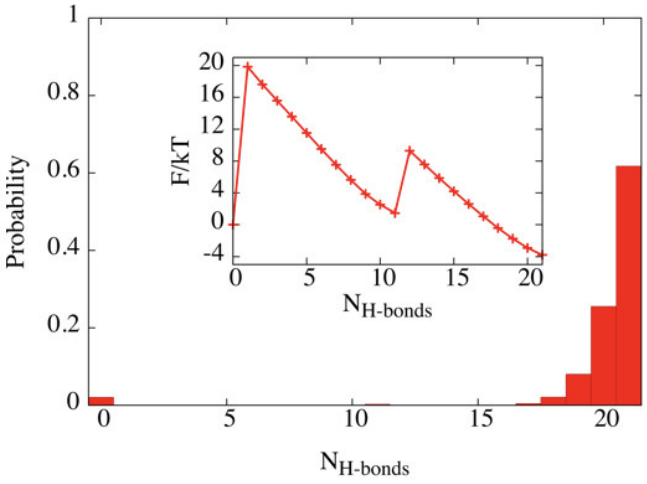
13


10 11 10 11
11 10 11 10
10 11

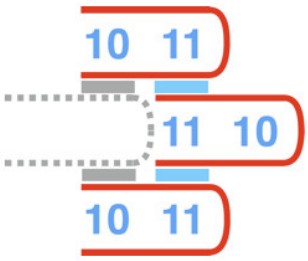

Probability
F/kT
N H-bonds
N H-bonds

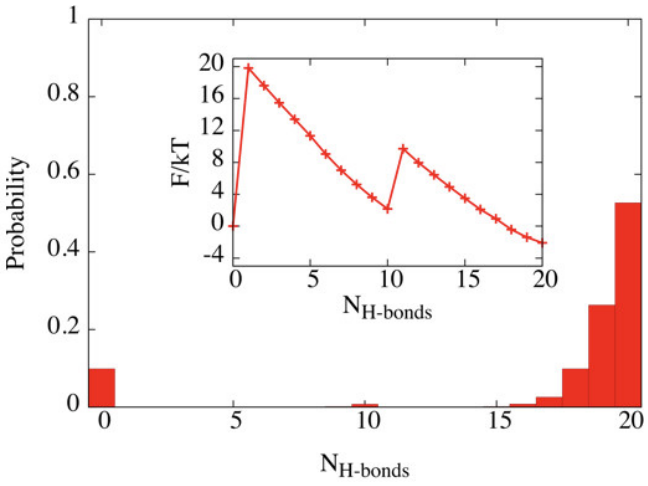
14


11 10 11 10
10 11 10 11
11 10

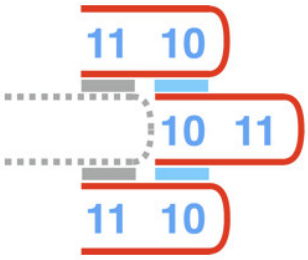

Probability
F/kT
N H-bonds
N H-bonds

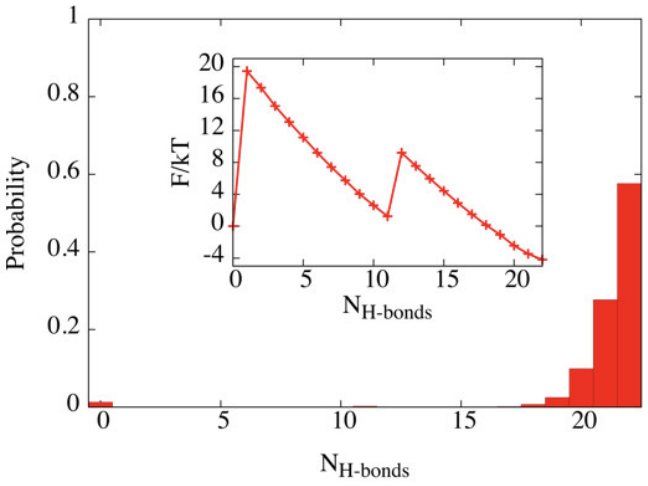
15


10 11
11 10
10 11

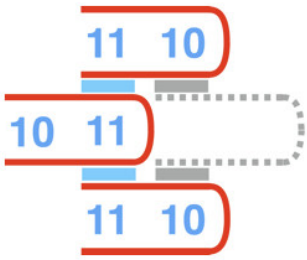

Probability
F/kT
N H-bonds
N H-bonds

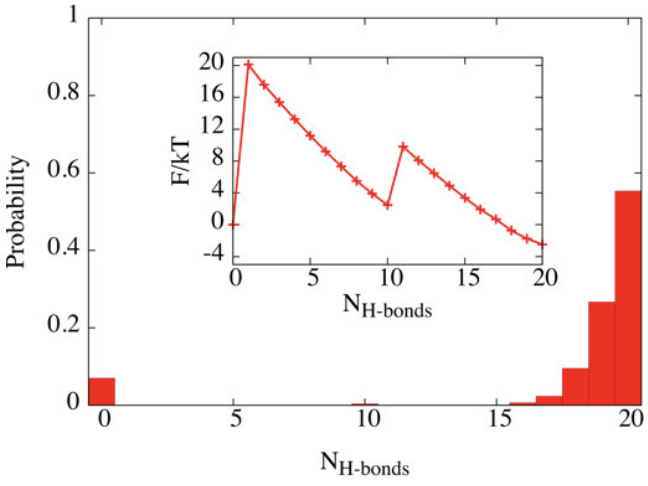
16


11 10
10 11
11 10

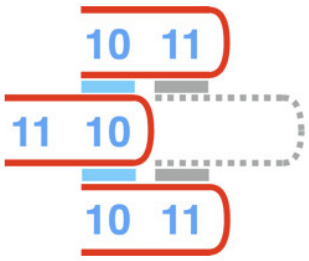

Probability
F/kT
N H-bonds
N H-bonds

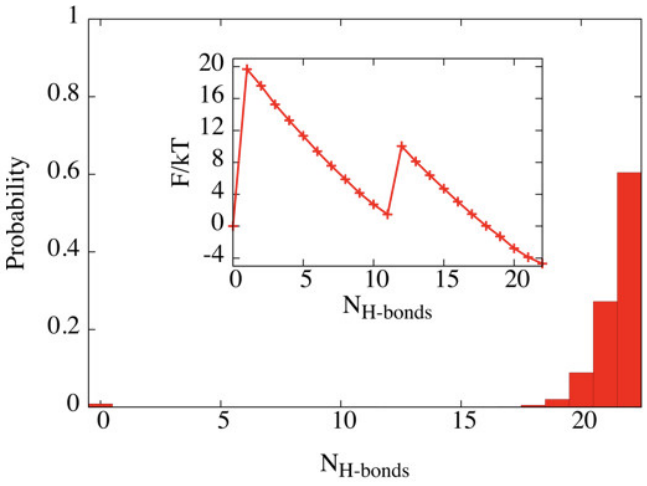
17


11 10
10 11
11 10


Probability
F/kT
N H-bonds
N H-bonds

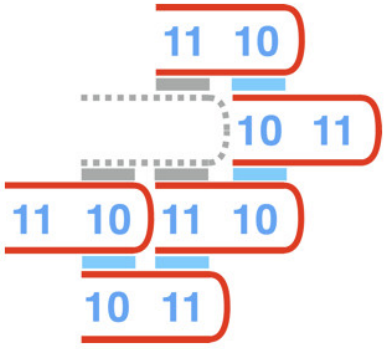
18

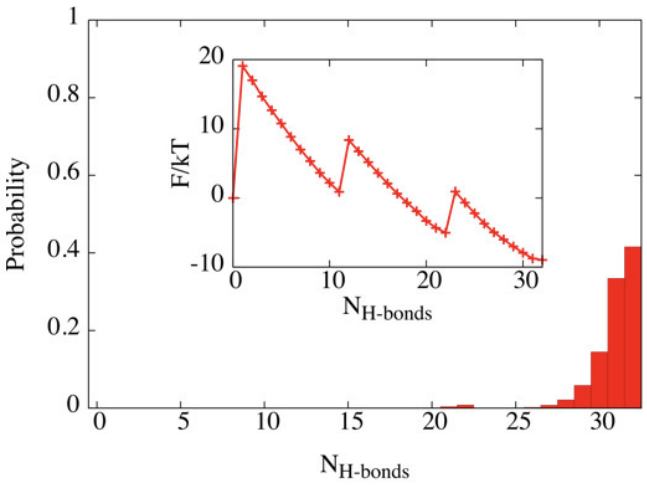




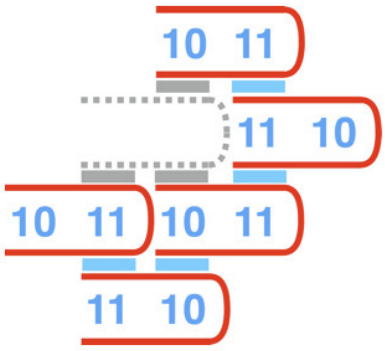
C Local environments with three binding domains

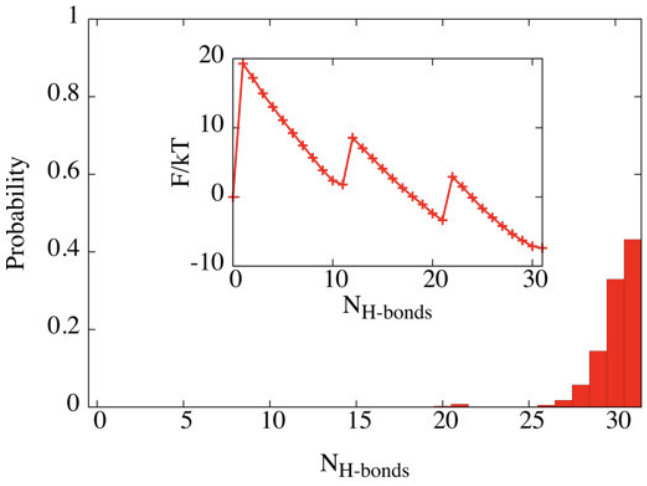
19



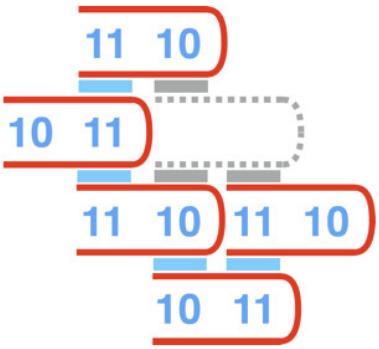


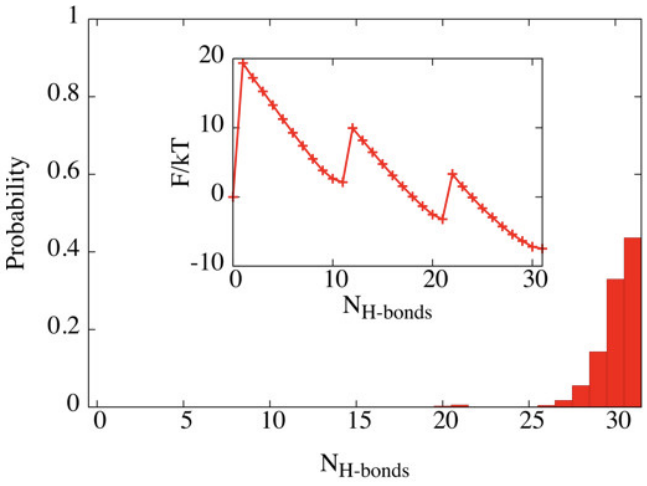
20



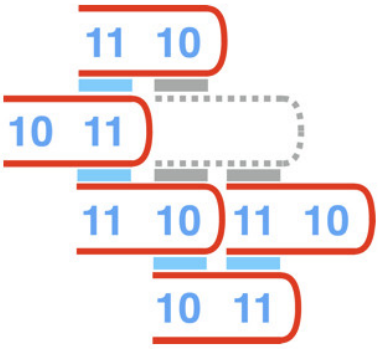


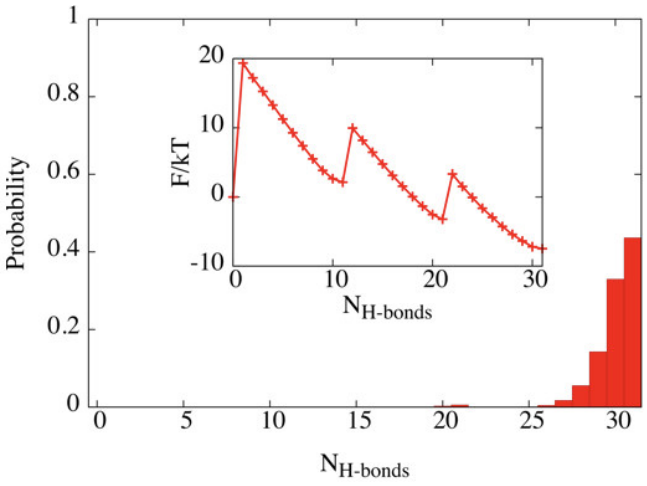
21


11 10
10 11
11 10 11 10
10 11

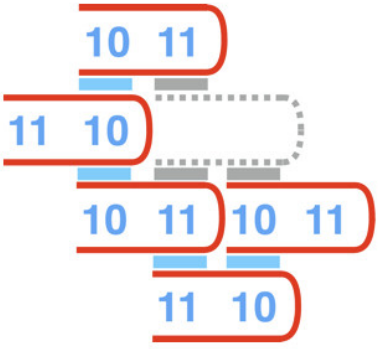

Probability
F/kT
N H-bonds
N H-bonds

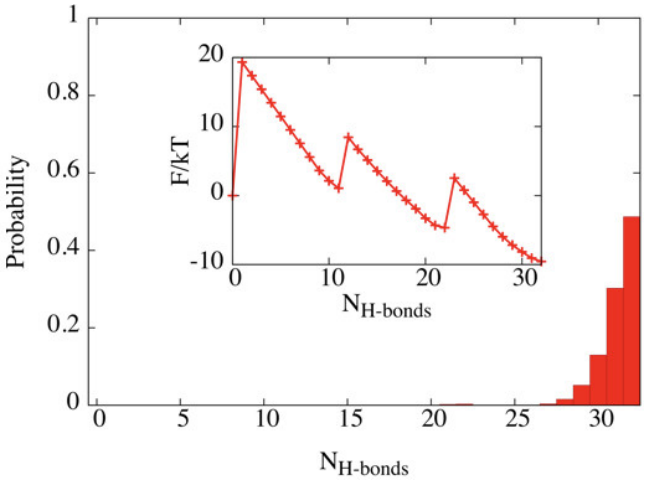
21


11 10
10 11
11 10 11 10
10 11

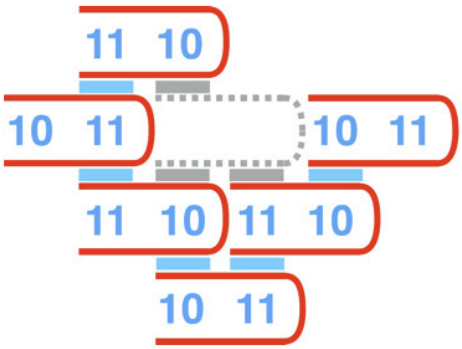

Probability
F/kT
N H-bonds
N H-bonds

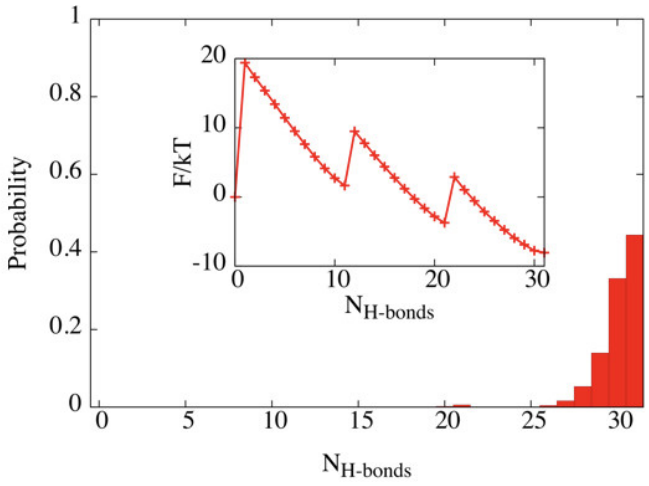
22


10 11
11 10
10 11 10 11
11 10

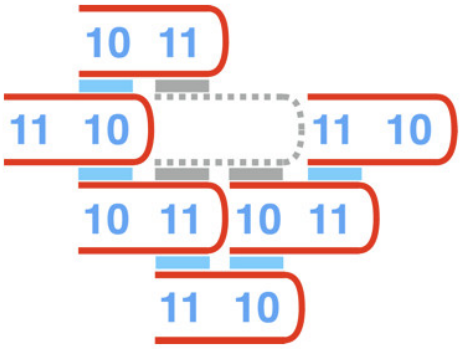

Probability
F/kT
N H-bonds
N H-bonds

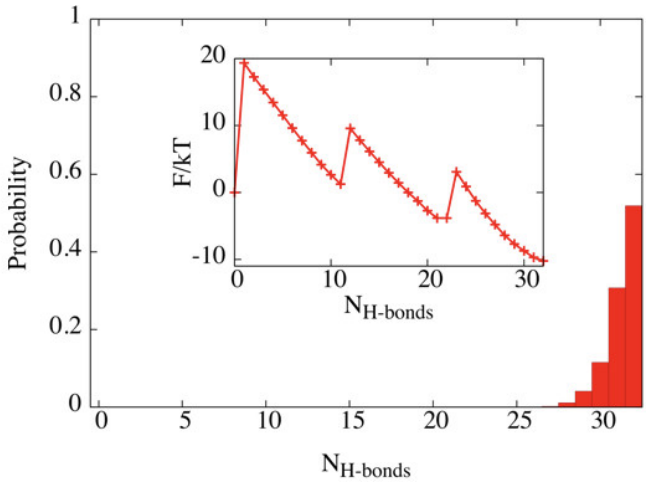
23


11 10
10 11
10 11
11 10
11 10
10 11

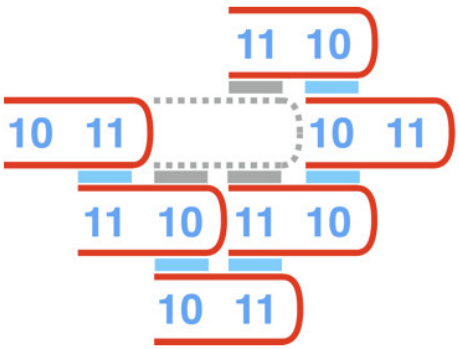

Probability
F/kT
N H-bonds
N H-bonds

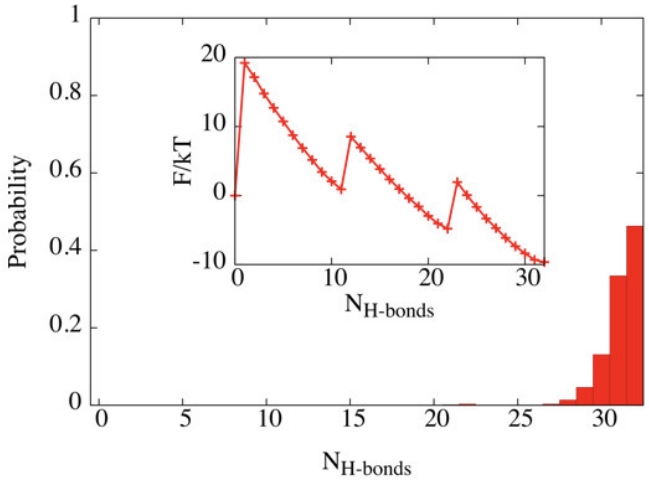
24


10 11
11 10
11 10
10 11
10 11
11 10

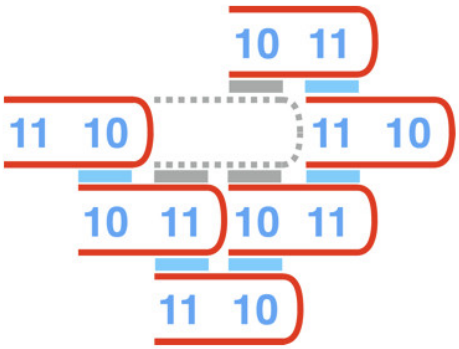

Probability
F/kT
N H-bonds
N H-bonds

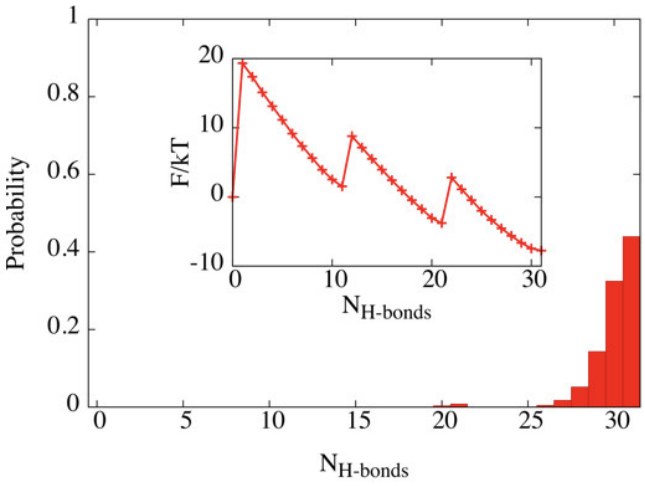
25


11 10
10 11
10 11
11 10
11 10
10 11


Probability
F/kT
N H-bonds
N H-bonds

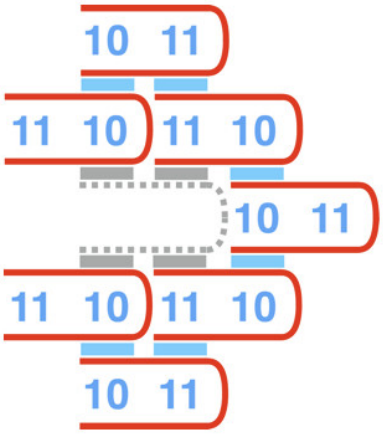
26

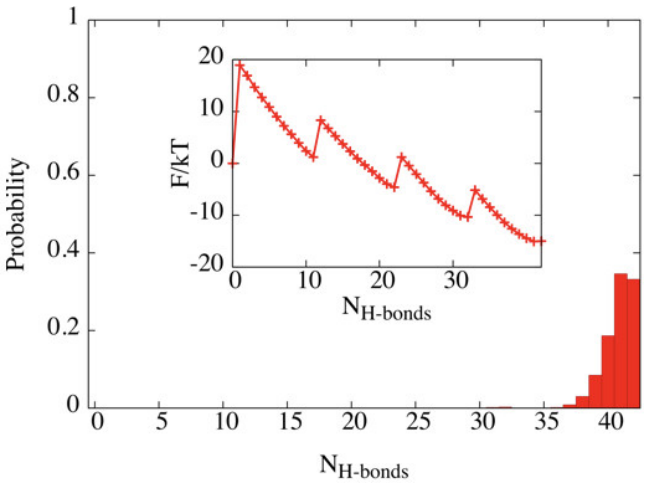




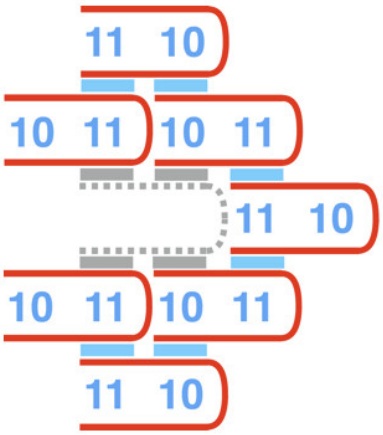
D Local environments with four binding domains

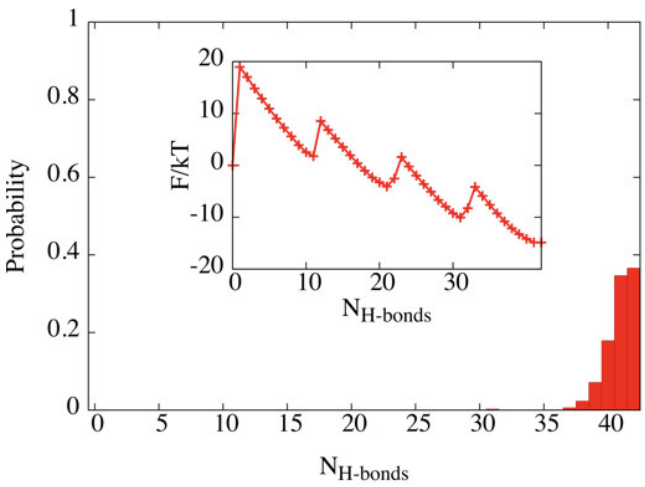
27



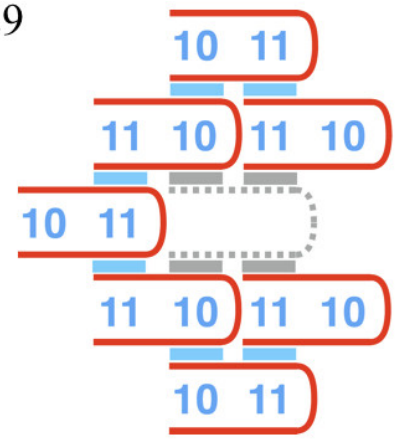


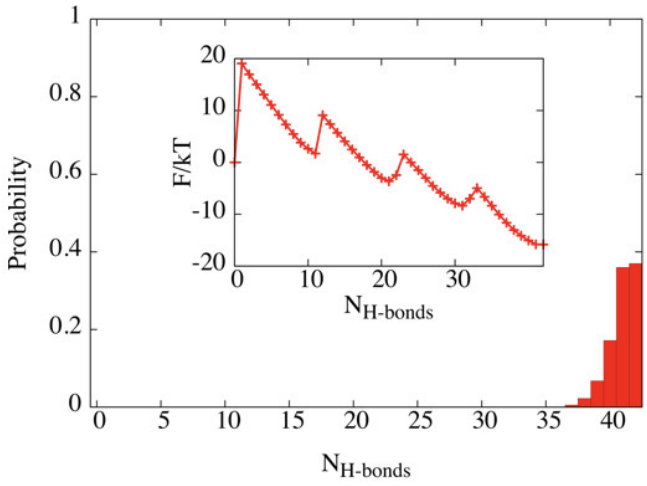
28



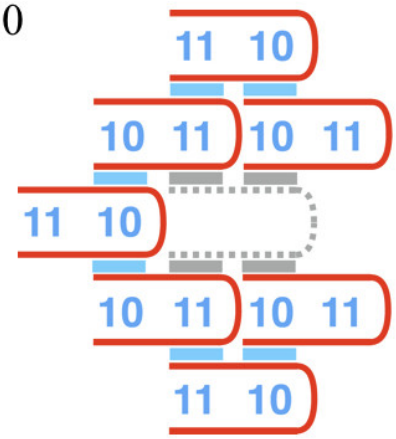


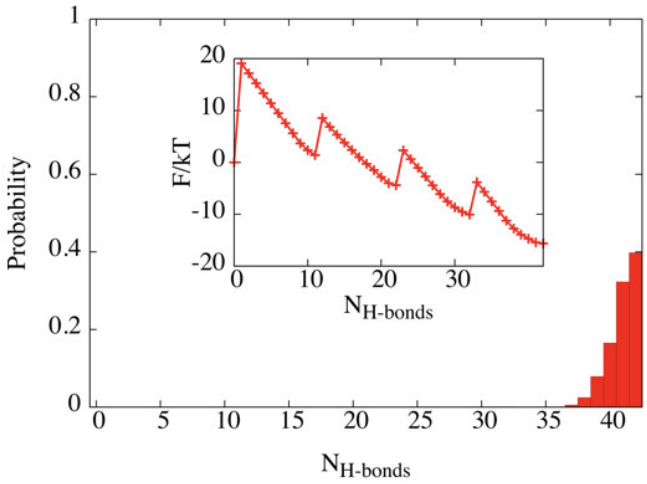
29


10 11
11 10 11 10
10 11
11 10 11 10
10 11

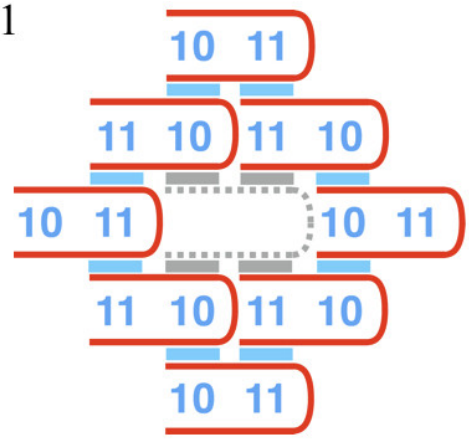

Probability
F/kT
N H-bonds
N H-bonds

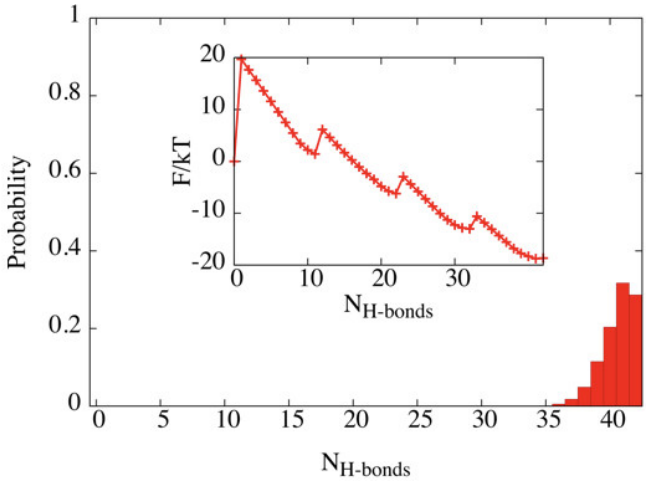
30


11 10
10 11 10 11
11 10
10 11 10 11
11 10

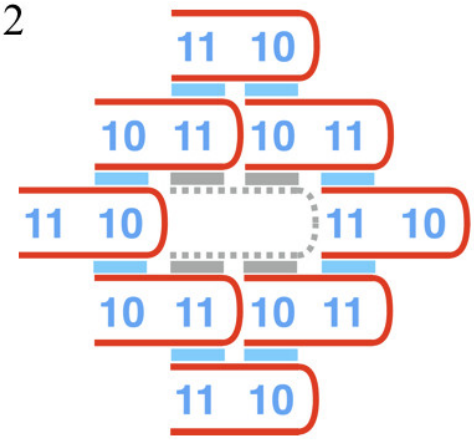

Probability
F/kT
N H-bonds
N H-bonds

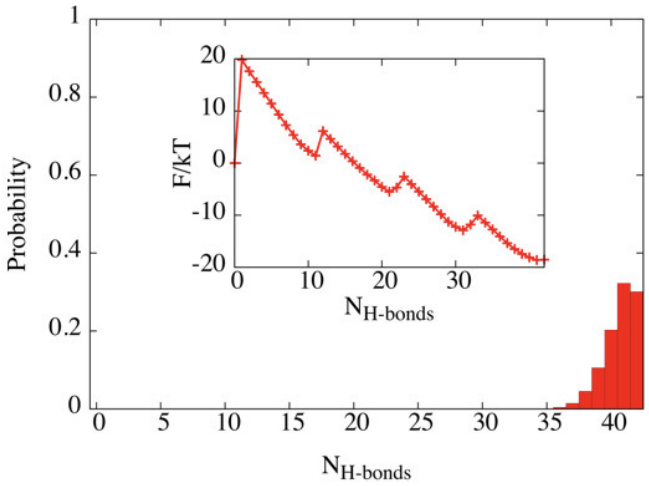
31


10 11
11 10 11 10
10 11 10 11
11 10 11 10
10 11


Probability
F/kT
N H-bonds
N H-bonds

32


11 10
10 11 10 11
11 10 11 10
10 11 10 11
11 10


Probability
F/kT
N H-bonds
N H-bonds

S.VII LONG RANGE BRIDGING - EXAMPLE

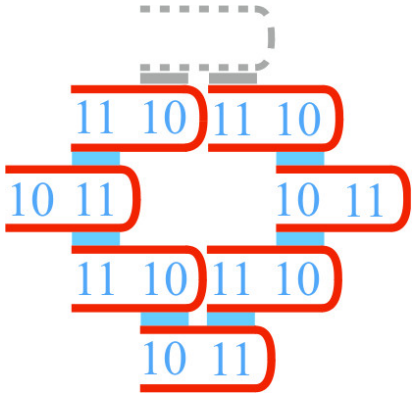


FIG. S7. Example of a "bridging tile" (coloured grey) whose association involves bringing together two initially distant sections of the assembly fragment. The binding domains in the assembly fragment that are hybridized are shown in blue. The numbers 10 and 11 indicate the length (in bases) of the binding domains positioned above.

S.VIII DEVIATIONS FROM DETAILED BALANCE I

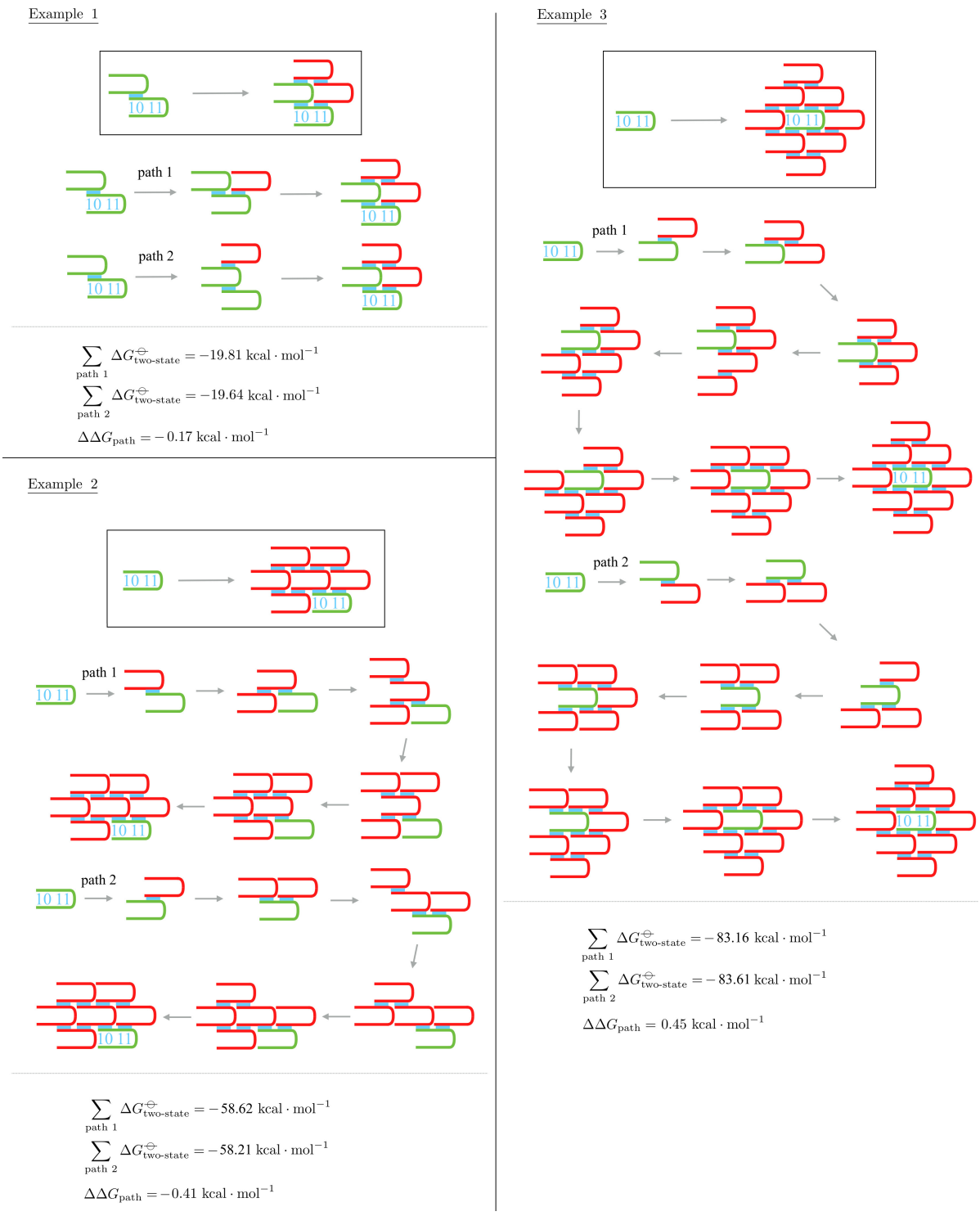


FIG. S8. Calculation of free-energy differences using different reaction paths. Calculation performed at $T = 50°\text{C}$ using tile-association free energies determined using oxDNA.

S.IX DEVIATIONS FROM DETAILED BALANCE II

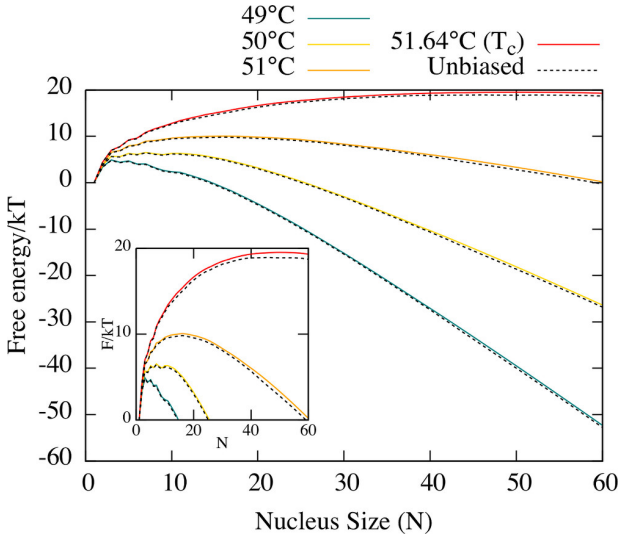


FIG. S9. Free-energy profiles as function of assembly size determined using biased sampling (coloured curves) and unbiased sampling (dashed curves). To make sampling more efficient, the assembly size coordinate was split into 6 sampling windows: $N_1 \in [1, 11]$, $N_2 \in [9, 21]$, $N_3 \in [19, 31]$, $N_4 \in [29, 41]$, $N_5 \in [39, 51]$ and $N_6 \in [49, 60]$; the resulting statistics were then combined by defining $F(N_1 = 1) = 0$, and fixing $F(N_1 = 10) = F(N_2 = 10)$, $F(N_2 = 20) = F(N_3 = 20)$, $F(N_3 = 30) = F(N_4 = 30)$, $F(N_4 = 40) = F(N_5 = 40)$ and $F(N_5 = 50) = F(N_6 = 50)$.

S.X ESTIMATION OF k_+^0 FROM TOEHOLD-MEDIATED STRAND DISPLACEMENT EXPERIMENTS

To parameterise the rate constant of tile-association (k_+^0) we consider the work of D. Zhang and E. Winfree [10], in which toehold exchange (TESD) and toehold-mediated (TMSD) strand displacement processes were characterised experimentally and compared against predictions of simple kinetic models, in which both TESD and TMSD are treated as simple bimolecular processes (with two reactants and two products each).

In the experiments carried out, a spectrofluorimetry analysis was used that provided a real-time probe for changes in the concentration of one of the final products (labelled "Y" in the original work). The experimental data was than used to parameterise the rate constants used in the kinetic models, so that a best fit was achieved between experimental and simulated curves.

Of especial interest for the present work is the analysis made on the TMSD system, which has the particularity that for long toeholds (~ 7 bases or longer) the forward rate constant in the kinetic model corresponds approximately to the rate constant associated with the hybridisation of the toehold with its complementary domain in an initially diffusing single strand, a process which we assume to be kinetically similar to the hybridisation of the first domain during tile-association.

We choose the value 6×10^6 for k_+^0, which corresponds to the rate constant estimated for the hybridisation of a G-C rich toehold in the TMSD system with a length of 10 bases (the maximum toehold length considered in Ref. [10]), with a complementary domain in a diffusing single strand.

[1] T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, J. Chem. Phys. **134**, 085101 (2011).
[2] J. P. K. Doye, T. E. Ouldridge, A. A. Louis, F. Romano, P. Sulc, C. Matek, B. E. K. Snodin, L. Rovigatti, J. S. Schreck, R. M. Harrison, and W. P. J. Smith, Phys. Chem. Chem. Phys. **15**, 20395 (2013).

[3] J. SantaLucia, Proc. Natl. Acad. Sci. U.S.A. **95**, 1460 (1998).
[4] J. SantaLucia, Jr. and D. Hicks, Annu. Rev. Biophys. Biomol. Struct. **33**, 415 (2004).
[5] J. S. Schreck, T. E. Ouldridge, F. Romano, A. A. Louis, and J. P. K. Doye, J. Chem. Phys. **142**, 165101 (2015).
[6] S. Whitelam and P. L. Geissler, J. Chem. Phys. **127**, 154101 (2007).
[7] P. Šulc, F. Romano, T. E. Ouldridge, L. Rovigatti, J. P. K. Doye, and A. A. Louis, J. Chem. Phys. **137**, 135101 (2012).
[8] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **61**, 2635 (1988).
[9] B. Wei, M. Dai, and P. Yin, Nature **485**, 623 (2012).
[10] D. Y. Zhang and E. Winfree, J. Am. Chem. Soc. **131**, 17303 (2009).