

Chapter 2

A Novel DNA Model

2.1 A Feasibility Study

When this project was started in 2007, very few models were available in the literature. In particular, no thermodynamic simulations of systems involving branched duplexes had been performed. It was therefore necessary to establish whether the aim of simulating nanotechnology with a coarse-grained model was a feasible one.

At the time, the model which had been studied most rigorously and used to simulate the largest systems (involving many particles undergoing DNA-mediated aggregation) was that of Starr and Sciortino [1]. This model represents DNA as an essentially linear molecule which has the potential to form ladder-like duplexes with its complementary strand. I adapted this model and used it to investigate the formation of ‘Holliday Junctions’, branched four-armed junctions involving four strands of DNA [2].

The junctions considered were based on those used by Malo et al. [3] to construct a two-dimensional DNA crystal. These junctions, as shown in Fig. 2.1, had two 13-bp long arms and two 7-bp long arms. Due to the extra bonding in the longer arms, these were predicted to be stable at higher temperature, and indeed UV absorbance did suggest that the junction forming process proceeded in two stages as the system was cooled.

Rigorous model thermodynamics were obtained for the entire assembly process, demonstrating the possibility of using coarse-grained models to simulate DNA nanotechnology. The model reproduced the greater stability of 13-bp duplexes, and also suggested the possibility of hierarchical assembly at constant temperature (having formed the longer arms, the two shorter arms of the junction could form cooperatively at temperatures above their individual melting points). Another pleasing result was that displacement of strands from partially bound duplexes was observed, an important process in DNA nanotechnology. As is evident from Fig. 2.1b, however, there were significant issues with this model. Due to the ladder-like nature of the model and absence of an explicit stacking interaction, the geometrical and mechanical properties of duplex DNA were far from realistic. For example, the

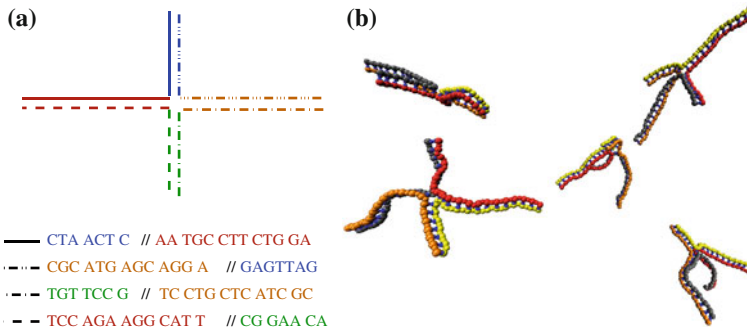


Fig. 2.1 **a** Strand sequences and schematic Holliday Junction structure from the system simulated in Ref. [2]. **b** Five identical Holliday junctions as represented by the model of Ref. [2]

duplex arms in Fig. 2.1b are unrealistically flexible perpendicular to the plane of bonding.

2.2 The Philosophy of the Model

Having established that modelling DNA nanotechnology was feasible, it was important to decide which aspects of DNA to emphasize in the new model. For the purposes of simulating much of DNA nanotechnology, I have aimed to embed the thermodynamics of transitions involving ssDNA and dsDNA (in the most common B-form) into a 3-dimensional, dynamical, coarse-grained representation that provides a reasonable description of the structural and mechanical features of the molecule. This ambition naturally coincides with a top-down approach. I have not been primarily concerned with the chemical details of interactions, but rather their net effect with regard to the properties of DNA.

Thermodynamically, the most important transitions to represent are the stacking of single strands, the formation of single-stranded hairpins and the hybridization of two separate strands to form duplexes. In terms of structure, it is vital that a model captures the ability of single strands to be both helically ordered and disordered. The helicity of dsDNA is also crucial, as it has a potentially large role in the kinetics of assembly, in particular leading to frustration of bonding when strands are topologically constrained [4]. A reasonable representation of the mechanical properties of DNA is also necessary. Single strands should be flexible, and duplexes comparatively stiff to represent their roles in nanotechnology. Further, quantities like the torsional and extensional moduli of dsDNA are important if the model is to be used to study systems involving DNA under stress, such as minicircles [5].

I have attempted to capture these properties by using only physically motivated interactions. Pairwise potentials representing excluded volume, backbone connectivity, hydrogen-bonding, stacking, coaxial stacking and cross-stacking have been

included (with no terms that possess explicitly length or loop size dependence [6, 7]). Analogously to the work of Morriss-Andrews et al. [8] and Ding et al. [6], the structure of dsDNA in the model is largely enforced by orientational dependencies of the hydrogen-bonding and stacking potentials.

An additional consideration in model design is the need for computational efficiency (if assembly transitions of complex structures are to be simulated). In the model, all interactions are pairwise (i.e., only involve two nucleotides, which are taken as rigid bodies). This pairwise character allows me to make efficient use of cluster-move Monte Carlo (MC) algorithms [9], which facilitate relaxation on all length-scales in a bound structure, and allow a much larger typical step size than possible in explicitly dynamical simulations (for more information, refer to Chap. 3). Designing the interactions to be truncated at short distances also improves simulation efficiency.

2.3 Degrees of Freedom in the Model

The model consists of rigid nucleotides with three interaction sites, illustrated in Fig. 2.2. The three interaction sites lie in a line, with the base stacking and hydrogen-bonding/base excluded volume sites separated from the backbone excluded volume site by 6.3 and 6.8 Å respectively. The orientation of bases is specified by a normal vector, which gives the notional plane of the base: the relative angle of base planes is used to modulate interactions (rather than through the use of off-axis sites).

It must be emphasized that the (often fairly complex) details of the interactions should not be over-interpreted—they are a means to an end. They are an attempt to mathematically quantify the tendency of nucleotides to interact favourably when in the geometry of duplex DNA, through the positions and orientations of the rigid model nucleotides. The widths and well depths of these potentials have been optimized to fit the properties of DNA, under the condition that the formation of unphysical structures was limited.

For the rest of this work, the model will be described in terms of reduced units, where one unit of length corresponds to $l = 8.518 \text{ Å}$ (this value was chosen to give a rise per bp of approximately 3.4 Å) and one unit of energy is equal to $E = 4.142 \times 10^{-20} \text{ J}$ (or equivalently, kT at $T = 300 \text{ K}$ corresponds to $0.1E$).

2.4 The Potential

2.4.1 Functional Forms

The potential used to model DNA consists of a sum of terms designed to represent physical interactions, such as excluded volume, base-stacking and hydrogen bonding. These terms are constructed from the functions given below:

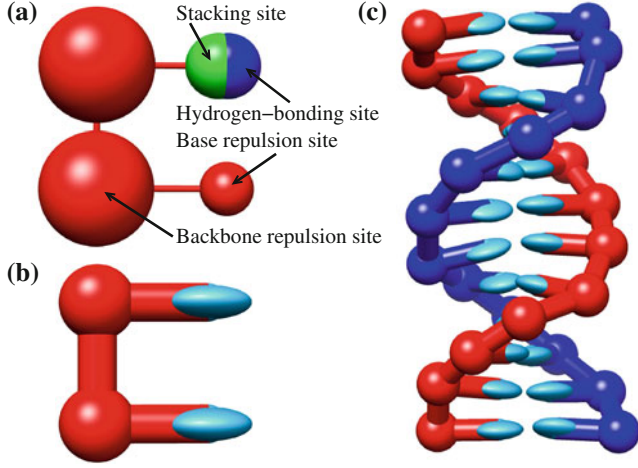


Fig. 2.2 **a** Model interaction sites. For clarity, the stacking/hydrogen-bonding sites are shown on one nucleotide and the base excluded volume on the other. The sizes of the spheres correspond to interaction ranges: two repulsive sites interact with a Lennard–Jones σ (see main text) equal to the sum of the radii shown (note that the truncation and smoothing procedure extends the repulsion slightly beyond this distance). The distance at which hydrogen-bonding and stacking interactions are at their most negative is given by the diameter of the spheres. Visualization was found to be clearer with nucleotides depicted as in **(b)**, with the subfigures **(a)** and **(b)** representing identical nucleotides on the same scale. The ellipsoidal bases allow a representation of the planarity inherent in the model, with the shortest axis corresponding to the base normal. **c** A 12 bp duplex as represented by the model

- FENE spring (used to connect backbones):

$$V_{\text{FENE}}(r, \epsilon, r^0, \Delta) = -\frac{\epsilon}{2} \ln \left(1 - \frac{(r - r^0)^2}{\Delta^2} \right). \quad (2.1)$$

- Morse potential (used for stacking and H-bonding):

$$V_{\text{Morse}}(r, \epsilon, r^0, a) = \epsilon (1 - \exp(-(r - r^0)a))^2. \quad (2.2)$$

- Harmonic potential (used for cross-stacking and coaxial stacking):

$$V_{\text{harm}}(r, k, r^0) = \frac{k}{2} (r - r^0)^2. \quad (2.3)$$

- Lennard–Jones potential (used for soft repulsion):

$$V_{\text{LJ}}(r, \epsilon, \sigma) = 4\epsilon \left(\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right). \quad (2.4)$$

- Quadratic terms (used for modulation):

$$V_{\text{mod}}(\theta, a, \theta^0) = 1 - a(\theta - \theta^0)^2. \quad (2.5)$$

- Quadratic smoothing terms for truncation:

$$V_{\text{smooth}}(x, b, x^c) = b(x^c - x)^2. \quad (2.6)$$

These functional forms are combined to give the following truncated, smooth and differentiable functions:

- The radial part of the stacking and hydrogen-bonding potentials:

$$f_1(r) = \begin{cases} V_{\text{Morse}}(r, \epsilon, r^0, a) - V_{\text{Morse}}(r^c, \epsilon, r^0, a) & \text{if } r^{\text{low}} < r < r^{\text{high}}, \\ \epsilon V_{\text{smooth}}(r, b^{\text{low}}, r^{c,\text{low}}) & \text{if } r^{c,\text{low}} < r < r^{\text{low}}, \\ \epsilon V_{\text{smooth}}(r, b^{\text{high}}, r^{c,\text{high}}) & \text{if } r^{\text{high}} < r < r^{c,\text{high}}, \\ 0 & \text{otherwise.} \end{cases} \quad (2.7)$$

- The radial part of the cross-stacking and coaxial stacking potentials:

$$f_2(r) = \begin{cases} V_{\text{harm}}(r, k, r^0) - V_{\text{harm}}(r^c, k, r^0) & \text{if } r^{\text{low}} < r < r^{\text{high}}, \\ k V_{\text{smooth}}(r, b^{\text{low}}, r^{c,\text{low}}) & \text{if } r^{c,\text{low}} < r < r^{\text{low}}, \\ k V_{\text{smooth}}(r, b^{\text{high}}, r^{c,\text{high}}) & \text{if } r^{\text{high}} < r < r^{c,\text{high}}, \\ 0 & \text{otherwise.} \end{cases} \quad (2.8)$$

- The radial part of the excluded volume potential:

$$f_3(r) = \begin{cases} V_{\text{LJ}}(r, \epsilon, \sigma) & \text{if } r < r^*, \\ \epsilon V_{\text{smooth}}(r, b, r^c) & \text{if } r^* < r < r^c, \\ 0 & \text{otherwise.} \end{cases} \quad (2.9)$$

- The angular modulation factor used in stacking, hydrogen-bonding, cross-stacking and coaxial stacking:

$$f_4(\theta) = \begin{cases} V_{\text{mod}}(\theta, a, \theta^0) & \text{if } \theta^0 - \Delta\theta^* < \theta < \theta^0 + \Delta\theta^*, \\ V_{\text{smooth}}(\theta, b, \theta^0 - \Delta\theta^c) & \text{if } \theta^0 - \Delta\theta^c < \theta < \theta^0 - \Delta\theta^*, \\ V_{\text{smooth}}(\theta, b, \theta^0 + \Delta\theta^c) & \text{if } \theta^0 + \Delta\theta^* < \theta < \theta^0 + \Delta\theta^c, \\ 0 & \text{otherwise.} \end{cases} \quad (2.10)$$

- Another modulating term which is used to impose right-handedness (effectively a one-sided modulation):

$$f_5(\phi) = \begin{cases} 1 & \text{if } x > 0, \\ V_{\text{mod}}(x, a, 0) & \text{if } x^* < x < 0, \\ V_{\text{smooth}}(x, b, x^c) & \text{if } x^c < x < x^*, \\ 0 & \text{otherwise.} \end{cases} \quad (2.11)$$

2.4.2 Interactions

The functional forms of Sect. 2.4.1 are combined to produce a potential for model DNA:

$$V = \sum_{\text{nn}} (V_{\text{backbone}} + V_{\text{stack}} + V'_{\text{exc}}) + \sum_{\text{other pairs}} (V_{\text{HB}} + V_{\text{cross_stack}} + V_{\text{coaxial_stack}} + V_{\text{exc}}). \quad (2.12)$$

Here the sum over nn runs over consecutive bases within strands.

For illustrations of the degrees of freedom to which the potential is applied, refer to Fig. 2.3.

2.4.2.1 Backbone Springs

Consecutive backbone sites on the same strand are connected by finitely-extensible nonlinear elastic (FENE) springs, which maintain backbone connectivity and specify the equilibrium separation of bases along the backbone ($\delta r_{\text{backbone}}^0$). As bases are linked by several covalent bonds in physical DNA, the extensibility of the backbone represents the possibility of rotating these covalent bonds with respect to each other, as well as the extension of the bonds themselves. Here, δr_X is the separation of interaction sites X on two nucleotides.

$$V_{\text{backbone}} = V_{\text{FENE}}(\delta r_{\text{backbone}}, \epsilon_{\text{backbone}}, \delta r_{\text{backbone}}^0, \Delta_{\text{backbone}}). \quad (2.13)$$

2.4.2.2 Excluded Volume

In order to prevent the collapse of model DNA into a dense, strongly-interacting cluster it is necessary to include excluded volume terms (which also prevent the backbones of two strands crossing during dynamical simulations). Excluded volume interactions occur between all combinations of backbone and base excluded volume sites on any two nucleotides (see Fig. 2.2). The only exception is for nearest neighbours, for which the backbone excluded volume sites do not interact (as their separation is controlled by the backbone FENE spring).

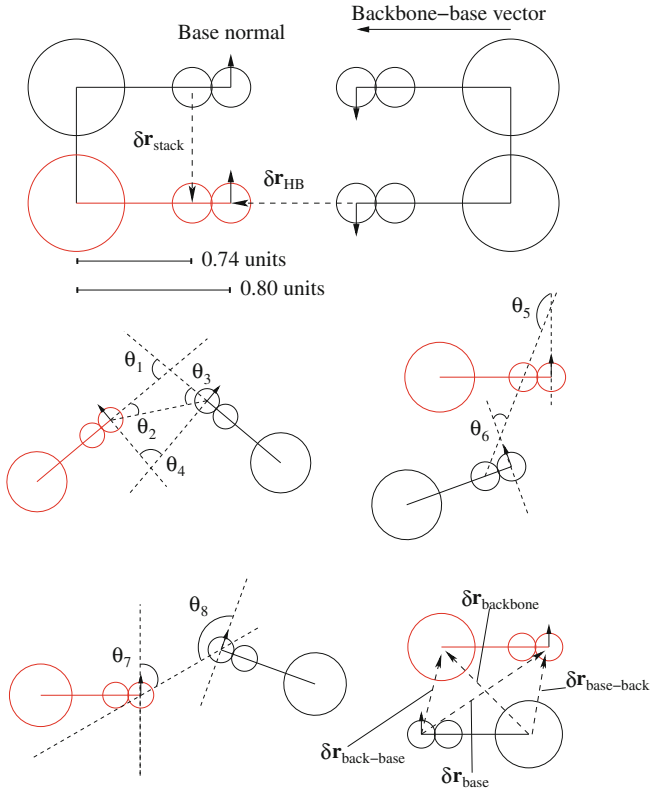


Fig. 2.3 Illustration of the variables used to parameterize the potential. Not shown in this diagram are the chirality inducing terms $\cos(\phi_1)$, $\cos(\phi_2)$, $\cos(\phi_3)$ and $\cos(\phi_4)$, which are discussed in detail in Sects. 2.4.2.3 and 2.4.2.6. It is convenient to define each pairwise interaction as if one nucleotide (coloured red in this picture) is being influenced by the other (coloured black): this allows each angle to be well-defined when calculating forces and torques. When calculating the energy, of course, the final result does not depend on the labelling of nucleotides. I define θ_2 , θ_5 and θ_7 as being measured with respect to the orientation of the red nucleotide, and θ_3 , θ_6 and θ_8 with respect to the orientation of the other nucleotide

$$\begin{aligned}
 V_{\text{exc}} = & f_3(\delta r_{\text{backbone}}, \epsilon_{\text{exc}}, \sigma_{\text{backbone}}, \delta r_{\text{backbone}}^*) + f_3(\delta r_{\text{base}}, \epsilon_{\text{exc}}, \sigma_{\text{base}}, \delta r_{\text{base}}^*) \\
 & + f_3(\delta r_{\text{back-base}}, \epsilon_{\text{exc}}, \sigma_{\text{back-base}}, \delta r_{\text{back-base}}^*) \\
 & + f_3(\delta r_{\text{base-back}}, \epsilon_{\text{exc}}, \sigma_{\text{back-base}}, \delta r_{\text{back-base}}^*).
 \end{aligned} \tag{2.14}$$

$$\begin{aligned}
 V'_{\text{exc}} = & f_3(\delta r_{\text{base}}, \epsilon_{\text{exc}}, \sigma_{\text{base}}, \delta r_{\text{base}}^*) \\
 & + f_3(\delta r_{\text{back-base}}, \epsilon_{\text{exc}}, \sigma_{\text{back-base}}, \delta r_{\text{back-base}}^*) \\
 & + f_3(\delta r_{\text{base-back}}, \epsilon_{\text{exc}}, \sigma_{\text{back-base}}, \delta r_{\text{back-base}}^*).
 \end{aligned} \tag{2.15}$$

2.4.2.3 Stacking

Bases have a tendency to form coplanar stacks, which drives the formation of helices in single- and double-stranded DNA due to the difference in length between the separation of bases along the backbone, and the optimal distance of stacking. The model reproduces this tendency with a stacking interaction between nearest neighbours on the same strand. The potential contains a radial term on the distance between stacking sites (r_{stack}), modulated by angular terms that favour the alignment of base normals with each other (θ_4) and with the separation between stacking sites ($\theta_{5'}$, $\theta_{6'}$). Right-handed helices are imposed through additional modulating factors which reduce the interaction to zero for increasing amounts of left-handed twist (measured by the quantities $\cos(\phi_1)$ and $\cos(\phi_2)$).

$$\begin{aligned}
 V_{\text{stack}} = & f_1(\delta r_{\text{stack}}, \epsilon_{\text{stack}}, a_{\text{stack}}, \delta r_{\text{stack}}^0, \delta r_{\text{stack}}^{c,\text{low}}, \delta r_{\text{stack}}^{c,\text{high}}, \delta r_{\text{stack}}^{\text{low}}, \delta r_{\text{stack}}^{\text{high}}) \\
 & \times f_4(\theta_4, a_{\text{stack},4}, \theta_{\text{stack},4}^0, \Delta\theta_{\text{stack},4}^*) \\
 & \times f_4(\theta_{5'}, a_{\text{stack},5}, \theta_{\text{stack},5}^0, \Delta\theta_{\text{stack},5}^*) f_4(\theta_{6'}, a_{\text{stack},6}, \theta_{\text{stack},6}^0, \Delta\theta_{\text{stack},6}^*) \\
 & \times f_5(-\cos(\phi_1), a_{\text{stack},1}, -\cos(\phi_1)_{\text{stack}}^*) \\
 & \times f_5(-\cos(\phi_2), a_{\text{stack},2}, -\cos(\phi_2)_{\text{stack}}^*). \tag{2.16}
 \end{aligned}$$

When calculating an interaction, it is convenient to arbitrarily label one of the nucleotides as being influenced by the other, as the angles can then all be defined with respect to one or other of the nucleotides (see Fig. 2.3). This is helpful when calculating the torques on a nucleotide, as in this situation one must consider the forces and torques on one of the nucleotides due to the other (and vice versa). If the labelling is swapped, there is of course no difference in the overall energy, and in most cases the interaction is symmetric under exchange of nucleotides so the calculation is also unaffected.

The stacking term is unusual, in that it is not symmetric under the exchange of the two nucleotides in question. $\theta_{5'}$ and $\theta_{6'}$ are calculated differently depending on which nucleotide is chosen to be influenced by the other, allowing the interaction to distinguish between the 3' and 5' directions.

- If the labelled nucleotide is in the 5' direction, $\theta_{5'} = \pi - \theta_5$ and $\theta_{6'} = \theta_6$.
- If the labelled nucleotide is in the 3' direction, $\theta_{5'} = \theta_5$ and $\theta_{6'} = \pi - \theta_6$.

By defining $\theta_{5'}$ and $\theta_{6'}$ in this way, it is possible to require that stacking can only occur when both normals point in the 5' direction. The base normals then define an axis about which the handedness of twist can be measured, and allow anti-parallel base-pairing to be enforced.

The chirality of DNA is introduced into the model via the terms $\cos(\phi_1)$ and $\cos(\phi_2)$ in the stacking interaction, which also distinguish between the 3' and 5' directions. Assume one of the nucleotides has been chosen to be labelled—the vector $\delta\hat{\mathbf{r}}_{\text{backbone}}$ is then the normalized vector **to** this nucleotide **from** the other one.

Let $\mathbf{y}$ and $\tilde{\mathbf{y}}$ by the unit vectors defined by the cross product of the normal and backbone-base vectors of the labelled and unlabelled nucleotides respectively.

- Labelled nucleotide in the 5' direction: $\cos(\phi_1) = \mathbf{y} \cdot \delta \hat{\mathbf{r}}_{\text{backbone}}$, $\cos(\phi_2) = \tilde{\mathbf{y}} \cdot \delta \hat{\mathbf{r}}_{\text{backbone}}$.
- Labelled nucleotide in the 3' direction, $\cos(\phi_1) = -\mathbf{y} \cdot \delta \hat{\mathbf{r}}_{\text{backbone}}$, $\cos(\phi_2) = -\tilde{\mathbf{y}} \cdot \delta \hat{\mathbf{r}}_{\text{backbone}}$.

When the stack forms in a right-handed fashion, $\cos(\phi_1)$ and $\cos(\phi_2)$ will be negative (this result relies on the fact that stacking normals must point in the 3' to 5' direction).

2.4.2.4 Hydrogen Bonding

DNA bases can undergo hydrogen bonding with each other, most commonly along the ‘Watson–Crick’ faces (the edge of the base furthest from the sugar group), with the planes of the bases approximately antiparallel. When taken in conjunction with stacking, the result is the famous B-DNA double helix. The model reproduces base-pairing through the V_{HB} term, which incorporates a radial term dependent on the separation of hydrogen-bonding sites, δr_{HB} . This interaction is modulated by terms that encourage the co-linear alignment of all four backbone and hydrogen-bonding sites (quantified by the angles θ_1 , θ_2 , and θ_3). A further factor is included to encourage the planes of bases to be antiparallel (measured by θ_4). Finally, terms are included that penalize pairs in which the separation of bonding sites is far from orthogonal with the base normals (these angles are θ_7 and θ_8). This final term tends to have only a small role for correctly formed base pairs, but is important in minimizing hydrogen bonding between bases that are not opposite each other in the helix.

$$\begin{aligned}
 V_{\text{HB}} = & f_1(\delta r_{\text{HB}}, \epsilon_{\text{NB}}, a_{\text{HB}}, \delta r_{\text{HB}}^0, \delta r_{\text{HB}}^{c,\text{low}}, \delta r_{\text{HB}}^{c,\text{high}}, \delta r_{\text{HB}}^{\text{low}}, \delta r_{\text{HB}}^{\text{high}}) \\
 & \times f_4(\theta_1, a_{\text{HB},1}, \theta_{\text{HB},1}^0, \Delta\theta_{\text{HB},1}^*) f_4(\theta_2, a_{\text{HB},2}, \theta_{\text{HB},2}^0, \Delta\theta_{\text{HB},2}^*) \\
 & \times f_4(\theta_3, a_{\text{HB},3}, \theta_{\text{HB},3}^0, \Delta\theta_{\text{HB},3}^*) f_4(\theta_4, a_{\text{HB},4}, \theta_{\text{HB},4}^0, \Delta\theta_{\text{HB},4}^*) \\
 & \times f_4(\theta_7, a_{\text{HB},7}, \theta_{\text{HB},7}^0, \Delta\theta_{\text{HB},7}^*) f_4(\theta_8, a_{\text{HB},8}, \theta_{\text{HB},8}^0, \Delta\theta_{\text{HB},8}^*). \quad (2.17)
 \end{aligned}$$

2.4.2.5 Cross-Stacking

$V_{\text{cross_stack}}$ represents cross-stacking interactions between a base in a base pair and nearest-neighbour bases on the opposite strand, providing additional stabilization of the duplex [10, 11]. Here it is incorporated through a potential which is a function of the distance between hydrogen-bonding sites r_{HB} , modulated by the alignment of base normals and backbone-base vectors with the separation vector (the same angles that appear in the definition of the hydrogen bonding potential) in such a way that its minimum is approximately consistent with the structure of model duplexes.

$$\begin{aligned}
V_{\text{cross_stack}} = & f_2(\delta r_{\text{HB}}, k_{\text{cross}}, \delta r_{\text{cross}}^0, \delta r_{\text{cross}}^{c,\text{low}}, \delta r_{\text{cross}}^{c,\text{high}}, \delta r_{\text{cross}}^{\text{low}}, \delta r_{\text{cross}}^{\text{high}}) \\
& \times f_4(\theta_1, a_{\text{cross},1}, \theta_{\text{cross},1}^0, \Delta\theta_{\text{cross},1}^*) \\
& \times f_4(\theta_2, a_{\text{cross},2}, \theta_{\text{cross},2}^0, \Delta\theta_{\text{cross},2}^*) f_4(\theta_3, a_{\text{cross},3}, \theta_{\text{cross},3}^0, \Delta\theta_{\text{cross},3}^*) \\
& \times \left(f_4(\theta_4, a_{\text{cross},4}, \theta_{\text{cross},4}^0, \Delta\theta_{\text{cross},4}^*) \right. \\
& \left. + f_4(\pi - \theta_4, a_{\text{cross},4}, \theta_{\text{cross},4}^0, \Delta\theta_{\text{cross},4}^*) \right) \\
& \times \left(f_4(\theta_7, a_{\text{cross},7}, \theta_{\text{cross},7}^0, \Delta\theta_{\text{cross},7}^*) \right. \\
& \left. + f_4(\pi - \theta_7, a_{\text{cross},7}, \theta_{\text{cross},7}^0, \Delta\theta_{\text{cross},7}^*) \right) \\
& \times \left(f_4(\theta_8, a_{\text{cross},8}, \theta_{\text{cross},8}^0, \Delta\theta_{\text{cross},8}^*) \right. \\
& \left. + f_4(\pi - \theta_8, a_{\text{cross},8}, \theta_{\text{cross},8}^0, \Delta\theta_{\text{cross},8}^*) \right). \tag{2.18}
\end{aligned}$$

2.4.2.6 Coaxial Stacking

The final term, $V_{\text{coax_stack}}$, is introduced to capture the tendency of stacking across nicked backbones to stabilize the binding of oligomers to duplexes with overhanging single-stranded tails [12–17] (coaxial stacking between blunt helix ends is also known to cause origami structures to associate [18]). The interaction is designed to be very similar in form to conventional stacking, with minor differences. Firstly, the radial component of the potential is taken to be of a more truncated, quadratic form—this was to prevent interactions between two bases which were stacked above and below their mutual neighbour. Secondly, it is impossible to define a 3′–5′ axis with two non-neighbouring bases: hence it was necessary to make the potential symmetric with respect to the alignment of base normals with their separation (θ_5 and θ_6). Similarly, without a 3′ to 5′ axis, the modulating terms which impose right-handedness also have to be designed differently—the new quantities $\cos(\phi_3)$ and $\cos(\phi_4)$ are defined below. Finally, in the case of nearest-neighbour stacking, the restriction of the backbone link between nucleotides means that the modulations described in Sect. 2.4.2.3 are enough to describe the geometry of stacking. For non-neighbour nucleotides, however, it was found that configurations in which the nucleotides were poorly stacked also gave significant interactions. To overcome this, an additional modulation in the angle between the two nucleotides’ backbone–base vectors (θ_1) was also included.

$$\begin{aligned}
V_{\text{coax_stack}} = & f_2(\delta r_{\text{stack}}, k, \delta r_{\text{coax}}^0, \delta r_{\text{coax}}^{c,\text{low}}, \delta r_{\text{coax}}^{c,\text{high}}, \delta r_{\text{coax}}^{\text{low}}, \delta r_{\text{coax}}^{\text{high}}) \\
& \times f_4(\theta_4, a_{\text{coax},4}, \theta_{\text{coax},4}^0, \Delta\theta_{\text{coax},4}^*) \\
& \times \left(f_4(\theta_1, a_{\text{coax},1}, \theta_{\text{coax},1}^0, \Delta\theta_{\text{coax},1}^*) \right. \\
& \left. + f_4(2\pi - \theta_1, a_{\text{coax},1}, \theta_{\text{coax},1}^0, \Delta\theta_{\text{coax},1}^*) \right) \\
& \times \left(f_4(\theta_5, a_{\text{coax},5}, \theta_{\text{coax},5}^0, \Delta\theta_{\text{coax},5}^*) \right. \\
& \left. + f_4(\pi - \theta_5, a_{\text{coax},5}, \theta_{\text{coax},5}^0, \Delta\theta_{\text{coax},5}^*) \right) \\
& \times \left(f_4(\theta_6, a_{\text{coax},6}, \theta_{\text{coax},6}^0, \Delta\theta_{\text{coax},6}^*) \right. \\
& \left. + f_4(\pi - \theta_6, a_{\text{coax},6}, \theta_{\text{coax},6}^0, \Delta\theta_{\text{coax},6}^*) \right) \\
& \times f_5(\cos(\phi_3), a_{\text{coax},3'}, \cos(\phi_3)_{\text{coax}}^*) \\
& \times f_5(\cos(\phi_4), a_{\text{coax},4'}, \cos(\phi_4)_{\text{coax}}^*). \tag{2.19}
\end{aligned}$$

Chirality is also imposed for the coaxial stacking term. Here there is no natural 3' to 5' direction, and so the vector between stacking sites rather than base normals must be used. I define the vector $\delta\hat{\mathbf{r}}_{\text{stack}}$ to be the normalized vector **to** the labelled nucleotide's stacking site **from** the other nucleotide, with a similar definition for $\delta\hat{\mathbf{r}}_{\text{backbone}}$. Taking $\mathbf{b}$ and $\tilde{\mathbf{b}}$ to be the backbone-base vectors for the labelled and unlabelled nucleotides respectively:

- $\cos(\phi_3) = \delta\hat{\mathbf{r}}_{\text{stack}} \cdot (\delta\hat{\mathbf{r}}_{\text{backbone}} \times \mathbf{b})$.
- $\cos(\phi_4) = \delta\hat{\mathbf{r}}_{\text{stack}} \cdot (\delta\hat{\mathbf{r}}_{\text{backbone}} \times \tilde{\mathbf{b}})$.

In this case, $\cos(\phi_3)$ and $\cos(\phi_4)$ are both positive for a right-handed stack.¹

2.4.3 Parameterization

The previous section is summarized in Tables 2.1 and 2.2, where the values of parameters of the model are also given. Note that the hydrogen bonding interaction is taken as non-zero only for complementary base pairs A–T and G–C, but beyond this there is no sequence dependence in the model. Also note that the factors b and x^c used in the quadratic smoothing parts of each interaction are not explicitly given. For each function, these quantities are specified by demanding differentiability and

¹ In fact, after the model was implemented it was found that $\cos(\phi_3) = \cos(\phi_4)$, and so the modulation could be simply calculated as $\cos^2(\phi_3)$.

Table 2.1 Parameter values in the model

Interaction		Parameters		
V_{backbone}	$V_{\text{FENE}}(\delta r_{\text{backbone}})$	$\epsilon_{\text{backbone}} = 2$	$\Delta_{\text{backbone}} = 0.25$	$\delta r_{\text{backbone}}^0 = 0.7525$
V_{HB}	$f_1(\delta r_{\text{HB}})$	$\epsilon_{\text{HB}} = 1.077$	$a_{\text{HB}} = 8$	$\delta r_{\text{HB}}^0 = 0.4$
		$\delta r_{\text{HB}}^c = 0.75$	$\delta r_{\text{HB}}^{\text{low}} = 0.34$	$\delta r_{\text{HB}}^{\text{high}} = 0.70$
	$f_4(\theta_1)$	$a_{\text{HB},1} = 1.50$	$\theta_{\text{HB},1}^0 = 0$	$\Delta\theta_{\text{HB},1}^* = 0.70$
	$f_4(\theta_2)$	$a_{\text{HB},2} = 1.50$	$\theta_{\text{HB},2}^0 = 0$	$\Delta\theta_{\text{HB},2}^* = 0.70$
	$f_4(\theta_3)$	$a_{\text{HB},3} = 1.50$	$\theta_{\text{HB},3}^0 = 0$	$\Delta\theta_{\text{HB},3}^* = 0.70$
	$f_4(\theta_4)$	$a_{\text{HB},4} = 0.46$	$\theta_{\text{HB},4}^0 = \pi$	$\Delta\theta_{\text{HB},4}^* = 0.70$
	$f_4(\theta_7)$	$a_{\text{HB},7} = 4.00$	$\theta_{\text{HB},7}^0 = \pi/2$	$\Delta\theta_{\text{HB},7}^* = 0.45$
	$f_4(\theta_8)$	$a_{\text{HB},8} = 4.00$	$\theta_{\text{HB},8}^0 = \pi/2$	$\Delta\theta_{\text{HB},8}^* = 0.45$
V_{stack}	$f_1(\delta r_{\text{stack}})$	$\epsilon_{\text{stack}} = 1.3448$ $+2.6568 kT$	$a_{\text{stack}} = 6$	$\delta r_{\text{stack}}^0 = 0.4$
		$\delta r_{\text{stack}}^c = 0.9$	$\delta r_{\text{stack}}^{\text{low}} = 0.32$	$\delta r_{\text{stack}}^{\text{high}} = 0.75$
	$f_4(\theta_4)$	$a_{\text{stack},4} = 1.30$	$\theta_{\text{stack},4}^0 = 0$	$\Delta\theta_{\text{stack},4}^* = 0.8$
	$f_4(\theta_{5'})$	$a_{\text{stack},5} = 0.90$	$\theta_{\text{stack},5}^0 = 0$	$\Delta\theta_{\text{stack},5}^* = 0.95$
	$f_4(\theta_{6'})$	$a_{\text{stack},6} = 0.90$	$\theta_{\text{stack},6}^0 = 0$	$\Delta\theta_{\text{stack},6}^* = 0.95$
	$f_5(-\cos(\phi_1))$	$a_{\text{stack},1} = 2.00$	$\cos(\phi_1)_{\text{stack}}^* = -0.65$	
	$f_5(-\cos(\phi_2))$	$a_{\text{stack},2} = 2.00$	$\cos(\phi_2)_{\text{stack}}^* = -0.65$	
V_{exc}	$f_3(\delta r_{\text{backbone}})$	$\epsilon_{\text{exc}} = 2.00$	$\sigma_{\text{backbone}} = 0.70$	$\delta r_{\text{backbone}}^* = 0.675$
	$+f_3(\delta r_{\text{base}})$	$\epsilon_{\text{exc}} = 2.00$	$\sigma_{\text{base}} = 0.33$	$\delta r_{\text{base}}^* = 0.32$
	$+f_3(\delta r_{\text{back-base}})$	$\epsilon_{\text{exc}} = 2.00$	$\sigma_{\text{back-base}} = 0.515$	$\delta r_{\text{back-base}}^* = 0.50$
	$+f_3(\delta r_{\text{base-back}})$	$\epsilon_{\text{exc}} = 2.00$	$\sigma_{\text{back-base}} = 0.515$	$\delta r_{\text{back-base}}^* = 0.50$

In this table, all energies and lengths are in terms of the simulation units E and l . When more than one function is listed for an interaction, the total interaction is a product of all the terms. Given the parameters of the main part of the interaction (for example, ϵ , r_0 , a and r_c for the V_{Morse} part of $f_1(r)$), the parameters of the smoothed cutoff regions are uniquely determined by ensuring continuity and differentiability at the boundaries (r^{low} and r^{high} for $f_1(r)$)

continuity at the point where the form of the potential changes. For example, consider an angular modulation $f_4(\theta)$. To be continuous at $\theta^0 \pm \Delta\theta^*$, we require:

$$1 - a(\Delta\theta^*)^2 = b(\theta^c - (\theta^0 \pm \Delta\theta^*))^2, \quad (2.20)$$

and

$$2a(\Delta\theta^*) = 2b(\theta^c - (\theta^0 \pm \Delta\theta^*)). \quad (2.21)$$

Given a , θ^0 and $\Delta\theta^*$, b and θ^c can be extracted by solving Eqs. 2.20 and 2.21 simultaneously. An equivalent procedure can be performed for any of the product functions which constitute a given interaction.

There are many free parameters in this model. This is, however, somewhat deceptive. The parameters relating to smoothing the truncation of potentials are generally of minor importance (at least for the thermodynamics of the system), as the effect of a potential well is largely determined by its width and depth rather than the details

Table 2.2 Further model parameters

Interaction	Parameters			
$V_{\text{cross_stack}}$	$f_2(\delta r_{\text{HB}})$	$k = 47.5$	$r_{\text{cross}}^0 = 0.575$	$\delta r_{\text{cross}}^c = 0.675$
		$\delta r_{\text{cross}}^{\text{low}} = 0.495$	$\delta r_{\text{cross}}^{\text{high}} = 0.655$	
	$f_4(\theta_1)$	$a_{\text{cross},1} = 2.25$	$\theta_{\text{cross},1}^0 = \pi - 2.35$	$\Delta\theta_{\text{cross},1}^* = 0.58$
	$f_4(\theta_2)$	$a_{\text{cross},2} = 1.70$	$\theta_{\text{cross},2}^0 = 1.00$	$\Delta\theta_{\text{cross},2}^* = 0.68$
	$f_4(\theta_3)$	$a_{\text{cross},3} = 1.70$	$\theta_{\text{cross},3}^0 = 1.00$	$\Delta\theta_{\text{cross},3}^* = 0.68$
	$f_4(\theta_4) + f_4(\pi - \theta_4)$	$a_{\text{cross},4} = 1.50$	$\theta_{\text{cross},4}^0 = 0$	$\Delta\theta_{\text{cross},4}^* = 0.65$
	$f_4(\theta_7) + f_4(\pi - \theta_7)$	$a_{\text{cross},7} = 1.70$	$\theta_{\text{cross},7}^0 = 0.875$	$\Delta\theta_{\text{cross},7}^* = 0.68$
	$f_4(\theta_8) + f_4(\pi - \theta_8)$	$a_{\text{cross},8} = 1.70$	$\theta_{\text{cross},8}^0 = 0.875$	$\Delta\theta_{\text{cross},8}^* = 0.68$
$V_{\text{coax_stack}}$	$f_2(\delta r_{\text{coax}})$	$k_{\text{coax}} = 46$	$\delta r_{\text{coax}}^0 = 0.4$	$\delta r_{\text{coax}}^c = 0.6$
		$\delta r_{\text{coax}}^{\text{low}} = 0.22$	$\delta r_{\text{coax}}^{\text{high}} = 0.58$	
	$f_4(\theta_1) + f_4(2\pi - \theta_1)$	$a_{\text{coax},1} = 2.00$	$\theta_{\text{coax},1}^0 = \pi - 0.60$	$\Delta\theta_{\text{coax},1}^* = 0.65$
	$f_4(\theta_4)$	$a_{\text{coax},4} = 1.30$	$\theta_{\text{coax},4}^0 = 0$	$\Delta\theta_{\text{coax},4}^* = 0.8$
	$f_4(\theta_5) + f_4(\pi - \theta_5)$	$a_{\text{coax},5} = 0.90$	$\theta_{\text{coax},5}^0 = 0$	$\Delta\theta_{\text{coax},5}^* = 0.95$
	$f_4(\theta_6) + f_4(\pi - \theta_6)$	$a_{\text{coax},6} = 0.90$	$\theta_{\text{coax},6}^0 = 0$	$\Delta\theta_{\text{coax},6}^* = 0.95$
	$f_5(\cos(\phi_3))$	$a_{\text{coax},3'} = 2.00$	$\cos(\phi_3)_{\text{coax}}^* = -0.65$	
	$f_5(\cos(\phi_4))$	$a_{\text{coax},4'} = 2.00$	$\cos(\phi_4)_{\text{coax}}^* = -0.65$	

of its edge. Furthermore, a number of parameters are essentially dictated by DNA geometry, and of those that are unconstrained, many are necessarily equal to others by symmetry. The parameters are also constrained by the requirement that the model behaves ‘like DNA’. In other words, if stacking and pairwise bonding are to drive the formation of double helices, rather than just cause collapse, the potentials need to have fairly narrow widths. Despite these caveats, there are still a large number of parameters to fit, especially via a top-down approach. This fitting is made more complex as it involves a compromise between the representation of various aspects of DNA. In particular, a given parameter may influence a wide range of properties and it is difficult to design a simple metric to compare the reproduction of thermodynamic and mechanical DNA behavior. In this case, lengths and potential minima were initially chosen to give model duplexes approximate B-DNA geometry. The stacking interaction strength and stiffness were then altered by hand to be consistent with the experimental thermodynamics reported for 14-base oligomers by Holbrook et al. [19]. Hydrogen-bonding and cross-stacking potentials were then added, and adjusted to give duplex and hairpin formation thermodynamics consistent with the SantaLucia parameterization of the nearest-neighbour model [12], which can be viewed as an accurate empirical fit to experimental data. For comparison with Ref. [12], I considered an ‘average base pair step’—details are provided in Chap. 6—as the model contains limited sequence dependence. Mechanical properties, such as persistence lengths, were then compared to experiment and interaction stiffnesses adjusted

by hand to provide improved agreement. This process was then iterated until the current set of parameters was found.²

In general, the interaction energy in a coarse-grained model should be interpreted as a free energy, as it incorporates a number of implicit degrees of freedom [23], and thus it is plausible that interaction strengths could be temperature dependent. To reduce free parameters, I have avoided this temperature dependence except with regard to the stacking interaction. It was found that it was difficult to generate a stacking transition with an entropy as small as required (for details, refer to Chap. 6) whilst maintaining an appropriate stiffness for dsDNA. I therefore took the stacking strength parameter to be linearly dependent on temperature: over the range 270–370 K, the stacking strength increases by $\sim 6\%$, in effect reducing the entropy cost of the transition.

There are several possible causes of this underestimation of the transition width. A contribution may be that in order to replicate the flexibility of single strands, the conformation of bases is unrestricted except by excluded volume, certainly a significant simplification. For physical DNA, a range of different conformations are accessible (allowing the large flexibility of unstacked single strands), but the available fraction of configuration space is restricted by specific steric clashes, which would be exceedingly difficult to reproduce in a bead-spring model such as mine. This overestimate of available configurations is the compromise necessary to allow hairpins and nanostructures to form. An alternative cause may be the lack of stacking heterogeneity. In reality, each pair of bases will stack with a different strength, resulting in a different stacking probability from other neighbouring bases. When looking at the average properties of a mixed sequence, however, an observer would see a combination of a number of transitions which would appear broader than the individual transitions themselves. This broadening effect is absent in my average base model. Finally, the stacking interaction itself is thought to rely partially on hydrophobic effects [24, 25], and hence would be temperature dependent in any model without explicit water.

2.4.4 Neglected Features of DNA

The model currently neglects some features of DNA. Although it incorporates sequence specificity (only A–T and G–C hydrogen bonds are possible), there is no other sequence dependence of interactions. I have made the simplifying assumption that noncomplementary base pairs have zero attraction, and also neglected the possibility of alternative base-pair geometries (such as Hoogsteen [24]). Due to this simplification, this version of the model cannot probe many sequence-dependent effects, such as the preferred sites of bubble formation (internal melting of base pairs) in DNA.

² as the coaxial stacking term has an almost negligible effect on the properties discussed above, it was separately fitted to thermodynamic data on coaxial stacking [12, 14–17, 20–22].

There are no explicit electrostatic interactions in the model, which may be expected to be important as bare ssDNA has a charge of $-e$ per base associated with the phosphate groups. For this reason, I fit to experimental data (where possible) at $[\text{Na}^+] = 500 \text{ mM}$, where electrostatic properties are strongly screened. Indeed, at these ionic concentrations, the Debye screening length is approximately $4.3 \text{ \AA}$, smaller than the excluded volume diameter for backbone–backbone interactions in our model ($\sim 6 \text{ \AA}$). At the shortest distances allowed by the steric interactions, charges would have an energy of $\sim 2kT$ in a Debye–Hückel approximation. Other authors have attempted to explicitly include a Debye–Hückel term [7], but also included a salt-dependent, medium-range attraction between strands in monovalent salt to facilitate hybridization, the physical origin of which is unclear. Due to this simplification, the model is unable to probe low salt regimes, and any result which relies upon DNA backbones coming into close proximity should be treated with caution.

The model also simplifies the geometry of DNA. Although the pitch, rise and diameter of helices are approximately consistent with experimental data, the grooves in between the backbones are of equal size. For physical DNA, by contrast, the major groove is significantly larger than the minor groove [24]. This is important for protein binding, and is also likely to have consequences for the strain inherent in certain origami designs [26].

Finally, it should be noted again that the description of the behaviour of the backbone of single-strands is very simplistic. In particular, bases are fully able to rotate and bend about the backbone bond with no restriction except for steric clashes. Whilst this seems to give a reasonable description of large-scale single-stranded mechanical properties (as discussed in Chap. 5), one should be wary of any predictions that rely heavily on one or two nucleotides adopting a particular configuration.

The simplifications in the model were made partly to reduce the number of possible parameters. For example, sequence dependence would give 16 combinations of stacking pairs, each pair requiring several parameters to describe their interaction. It was also felt that, as an initial step in modelling, it was important to obtain a good physical representation of the underlying properties of DNA assembly (such as the generic dependence of melting temperature on length), before incorporating sequence specific or low salt effects. Furthermore, some generic effects may be obscured by sequence-specific terms (for instance, free-energy profiles such as those in Chap. 6 would have sequence-dependent fluctuations overlying the general trend).

2.4.5 Additional Parameters Required for Dynamical Simulations

The parameters presented so far specify the thermodynamic properties of the model. To use explicitly dynamical algorithms, one must introduce masses, moments of inertia and drag coefficients. For simplicity, I assume that each nucleotide is a uniform density sphere centred on the backbone–base axis, $0.24l$ from the backbone site (this assumption is discussed in Sect. 3.2). Taking the unit of mass in simulations to be $M = 100 \text{ AMU}$, an average mass of a nucleotide is around $3.1575M$, and the moment

of inertia is 0.43512 in simulation units. Using Langevin dynamics to simulate the model is discussed further in Chap. 3 and Appendix B, where the damping coefficients are defined and given values. Note that this definition of M , when combined with the reduced length and energy scales L and E , defines the reduced unit of time in simulations $T = (E/M/L^2)^{-0.5} = 1.706$ ps.

References

1. F. W. Starr and F. Sciortino. Model for assembly and gelation of four-armed DNA dendrimers. *J. Phys.: Condens. Matter*, 18:L347–L353, 2006.
2. T. E. Ouldridge et al. The self-assembly of DNA Holliday junctions studied with a minimal model. *J. Chem. Phys.*, 130:065101, 2009.
3. J. Malo, et al. Engineering a 2D protein-DNA crystal. *Angew. Chem. Int. Ed.*, 44:3057–3061, 2005.
4. J. Bois et al. Topological constraints in nucleic acid hybridization kinetics. *Nucl. Acids Res.*, 33(13):4090–4095, 2005.
5. S. A. Harris, C. A. Laughton, and T. B. Liverpool. Mapping the phase diagram of the writhe of DNA nanocircles using atomistic molecular dynamics simulations. *Nucl. Acids Res.*, 36(1):21–29, 2008.
6. F. Ding et al. Ab initio RNA folding by discrete molecular dynamics: From structure prediction to folding mechanisms. *RNA*, 14:1164–1173, 2008.
7. E. J. Sambriski, V. Ortiz, and J. J. de Pablo. Sequence effects in the melting and renaturation of short DNA oligonucleotides: structure and mechanistic pathways. *J. Phys.: Condens. Matter*, 21(034105), 2009.
8. A. Morriss-Andrews, J. Rottler, and S. S. Plotkin. A systematically coarse-grained model for DNA and its predictions for persistence length, stacking, twist and chirality. *J. Chem. Phys.*, 132:035105, 2010.
9. S. Whitlam et al. The role of collective motion in examples of coarsening and self-assembly. *Soft Matter*, 5(6):1521–1262, 2009.
10. M. Swart et al. π - π stacking tackled with density functional theory. *J. Mol. Model.*, 13(12):1245–1257, 2007.
11. J. Sponer et al. Nature of base stacking: Reference quantum-chemical stacking energies in ten unique B-DNA base-pair steps. *Chem. Eur. J.*, 12(10):2854–2865, 2006.
12. J. SantaLucia, Jr. and D. Hicks. The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.*, 33:415–40, 2004.
13. N. Peyret. *Prediction of nucleic acid hybridization: parameters and algorithms*. PhD thesis, Wayne State University, 2000.
14. D. V. Pyshnyi and E. M. Ivanova. Thermodynamic parameters of coaxial stacking on stacking hybridization of oligodeoxyribonucleotides. *Russ. Chem. B+*, 51:1145–1155, 2002.
15. D. V. Pyshnyi and E. M. Ivanova. The influence of nearest neighbors on the efficiency of coaxial stacking at contiguous stacking hybridization of oligodeoxyribonucleotides. *Nucleos. Nucleot. Nucl.*, 23(6–7):1057–1064, 2004.
16. M. J. Lane et al. The thermodynamic advantage of DNA oligonucleotide ‘stacking hybridization’ reactions: Energetics of a DNA nick. *Nucl. Acids Res.*, 25(3):611–616, 1997.
17. V. A. Vasiliskov, D. V. Prokopenko, and A. D. Mirzabekov. Parallel multiplex thermodynamic analysis of coaxial base stacking in DNA duplexes by oligodeoxyribonucleotide microchips. *Nucl. Acids Res.*, 29(11):2303–2313, 2001.
18. S. Woo and P. W. K. Rothmund. Programmable molecular recognition based on the geometry of DNA nanostructures. *Nature Chem.*, 3:620–627, 2011.

19. J. A. Holbrook et al. Enthalpy and heat capacity changes for formation of an oligomeric DNA duplex: Interpretation in terms of coupled processes of formation and association of single-stranded helices. *Biochemistry*, 38(26):8409–8422, 1999.
20. N. Peyret et al. Nearest-neighbour thermodynamics and NMR of DNA sequences with internal AA, CC, GG and TT mismatches. *Biochemistry*, 38:3468–3477, 1999.
21. E. Protozanova, P. Yakovchuk, and M. D. Frank-Kamenetskii. Stacked-unstacked equilibrium at the nick site of DNA. *J. Mol. Biol.*, 342(3):775–785, 2004.
22. P. Yakovchuk, E. Protozanova, and M. D. Frank-Kamenetskii. Base-stacking and base-pairing contributions into thermal stability of the dna double helix. *Nucl. Acids Res.*, 34(2):564–574, 2006.
23. R. Everaers, S. Kumar, and C. Simm. Unified description of poly- and oligonucleotide DNA melting: Nearest-neighbor, poland-sheraga, and lattice models. *Phys. Rev. E*, 75:041918, 2007.
24. W. Saenger. *Principles of Nucleic Acid Structure*. Springer-Verlag, New York, 1984.
25. K. M. Guckan et al. Factors contributing to aromatic stacking in water: evaluation in the context of DNA. *J. Am. Chem. Soc.*, 122(10):2213–2222, 2000.
26. W. B. Sherman and N. C. Seeman. Design of minimally strained nucleic acid nanotubes. *Biophys. J.*, 90(12): 4546–4557, 2006.

Coarse-Grained Modelling of DNA and DNA
Self-Assembly

Ouldrige, Th.E.

2012, XII, 176 p., Hardcover

ISBN: 978-3-642-30516-0