Supporting Information: Rigid Base Biasing in Molecular Dynamics enables enhanced sampling of DNA conformations

Aderik Voorspoels,*,† Jocelyne Vreede,*,‡ and Enrico Carlon*,†

†*Soft Matter and Biophysics, Department of Physics and Astronomy, KU Leuven, Celestijnenlaan 200D, 3000 Leuven, Belgium*
‡*Van 't Hoff Institute for Molecular Sciences, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, the Netherlands*

E-mail: aderik.voorspoels@kuleuven.be; j.vreede@uva.nl; enrico.carlon@kuleuven.be

S1 Mathematical derivations in the algorithm

The purpose of the RBB-NA algorithm is to bias the rotational and translational rigid base coordinates $\vec{X}$ of a DNA (or RNA) strand, given the representation of this strand as a series of bases $\mathcal{B}_b$. Here b ranges from 0 to $2N$ with N being the length of the strand in question in base pairs. Notably any base will during the simulation be represented by a collection of atom positions $\left\{\vec{A}_i\right\}$. It is from these atom positions that the calculation of the collective variables must start, and it is on these atoms that in the end forces can be applied using Eq. (2). For sake of completeness and to introduce the required notations and concepts this section begins with a description of the curves algorithm developed by Lavery et al.[S1] which is used to calculate the collective variables. This algorithm has been re-implemented in RBB-NA to fit with the PLUMED code base.

S1.1 Curves in brief

The standard choice in the Curves+ software is to first fit a set of ideal base coordinates $\left\{\vec{A}_i^*\right\}$ to the measured ones by finding the appropriate rotation in a procedure described by McLachlan.[S2] This is usually done to reduce the influence of fluctuations within a base on the collective variables, thus reducing the noise. For the purposes of this work this step is useful as it incorporates all heavy atoms in a given base into the calculations of the collective variables. As such it is this step that facilitates both treating the entire base as a rigid body as well as preventing forces from interfering with fluctuations in the bases themselves.

After the optional fitting the curves algorithm starts by defining reference frames $\mathcal{B}_b$ for each base consisting of a triad of orthonormal vectors $\mathbf{B}_b$ and a reference point $\vec{r}_b$ where this frame is to be attached.[S1] Here $\mathbf{B}_b$ can be seen as a rotation matrix

$$\mathbf{B}_b = [\hat{e}_1, \hat{e}_2, \hat{e}_3]\,. \tag{S1}$$

where the vectors $\hat{e}_1, \hat{e}_2, \hat{e}_3$ collectively form the base triad with $\hat{e}_1$ pointing into the major groove, $\hat{e}_2$ connecting the backbones and $\hat{e}_3$ being normal to the base plane.

To keep track of the different frames and many variables needed in the calculations that follow we include here a diagram (S1). On this diagram the reference frames involved in the calculation of rigid base coordinates in this par-

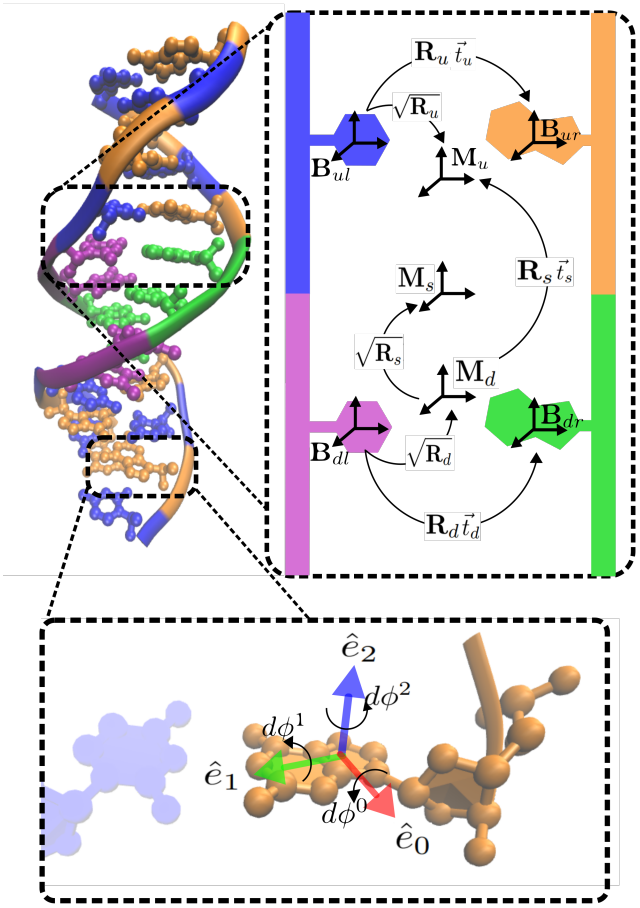

Figure S1: A strand of DNA. Attached to the right is a diagram showing the relation between different rotations and reference triad. Attached below is an illustration of the unit vectors forming a base triad and the action small rotations have on them

ticular step are labelled according to their position in relation to the step. We denote all frames in the upward direction from the step with a subscript 'u', The frames in the downward direction are denoted similarly with subscript 'd'. For frames on the left strand we add 'l' and for those on the right 'r'.

When the frames are constructed one defines the transformation from one base to the opposing one, say in the diagram, (S1) from $\mathcal{B}_{ul}$ to $\mathcal{B}_{ur}$, by rotation matrix $\mathbf{R}_u$ and translation $\vec{t}_u$ which are defined as

$$\mathbf{R}_u = \mathbf{B}_{ur}\mathbf{B}_{ul}^T, \quad \vec{t}_u = \vec{r}_{ur} - \vec{r}_{ul}. \tag{S2}$$

In addition to these transformations a frame in which to express them is needed. This is given by the midframe $\mathcal{M}_u$ of this particular base pair for which the triad can be written as

$$\mathbf{M}_u = \sqrt{\mathbf{R}_u}\mathbf{B}_{ul} = \sqrt{\mathbf{R}_u^T}\mathbf{B}_{ur}. \tag{S3}$$

Here the square-root $\sqrt{\mathbf{R}}$ is a short notation for $\exp\frac{1}{2}\log\mathbf{R}$ which yields a rotation about the same axis by half the angle.

In practice the midframe is not found by explicit calculation of the square root but simply by rotating the three vectors in $\mathbf{B}_{ul}$ around the axis of rotation of $\mathbf{R}_u$ by half its rotation angle. For any rotation matrix the angle of rotation $\Theta(\mathbf{R})$ can be found by

$$\Theta(\mathbf{R}) = \arccos\frac{1}{2}\left(\mathrm{Tr}(\mathbf{R}) - 1\right), \tag{S4}$$

and subsequently the axis of rotation $\widehat{K}(\mathbf{R})$ is given as

$$\widehat{K}(\mathbf{R}) = \frac{1}{2\sin\Theta(\mathbf{R})}\begin{bmatrix} R_{23} - R_{32} \\ R_{31} - R_{13} \\ R_{12} - R_{21} \end{bmatrix}. \tag{S5}$$

This conversion of rotation matrices to axis angle representation also allows us to define the rotational collective variables of a base pair as

$$\vec{\omega}_u^T = \Theta(\mathbf{R}_u)(\widehat{K}(\mathbf{R}_u)^T \cdot \mathbf{M}_u). \tag{S6}$$

Similarly the translational collective variables of the basepair are defined by

$$\vec{d}_u^T = \vec{t}_u^T \cdot \mathbf{M}_u. \tag{S7}$$

After the calculation of the basepair collective variables the algorithm largely repeats itself. The transformation from one pair to the next, say from $\mathcal{M}_d$ to $\mathcal{M}_u$ on the diagram, is now given by a rotation $\mathbf{R}_s$ and a translation $\vec{t}_s$ which are defined by

$$\mathbf{R}_s = \mathbf{M}_u\mathbf{M}_d^T, \quad \vec{t}_s = \frac{1}{2}(\vec{r}_{ur} + \vec{r}_{ul}) - \frac{1}{2}(\vec{r}_{dr} + \vec{r}_{dl}). \tag{S8}$$

Like before a midframe is then defined by half rotation:

$$\mathbf{M}_s = \sqrt{\mathbf{R}_s}\mathbf{M}_d = \sqrt{\mathbf{R}_s^T}\mathbf{M}_u, \tag{S9}$$

and finally the transformations are expressed in

this midframe

$$\vec{\Omega}_u^T = \Theta(\mathbf{R}_s)(\widehat{K}(\mathbf{R}_s)^T \cdot \mathbf{M}_s) \quad \text{(S10)}$$
$$\vec{D}_s^T = \vec{t}_s^{\,T} \cdot \mathbf{M}_s. \quad \text{(S11)}$$

With these four sets of collective variables we can turn to the calculation of forces, which as noted in the main text, requires the calculation of gradients with respect to the cartesian coordinates of atom positions.

S1.2 Calculus of rotations

The calculated rigid base-pair coordinates fall into two different categories: translations and rotations. For the two sets of translations derivation and thus the application of forces is simple, the two sets of rotations require more work. Here the construction of gradients of those two sets of rotations $\vec{\Omega}$ and $\vec{\omega}$ to small rotations of the underlying base frames $\mathrm{d}\vec{\phi}_b$ is discussed.

To start, the derivative of the base reference frame $\mathbf{B}_b$ to rotations $\mathrm{d}\vec{\phi}^b$ about its own axis can be found as follows. Subsequently imposing the infinitesimal rotations $\mathrm{d}\phi_b^0, \mathrm{d}\phi_b^1, \mathrm{d}\phi_b^2$ on $\mathbf{B}_b$ yields

$$\mathbf{B}_b + \mathrm{d}\mathbf{B}_b = \mathbf{B}_b \circ \begin{pmatrix} 1 & -\mathrm{d}\phi^3 & \mathrm{d}\phi^2 \\ \mathrm{d}\phi^3 & 1 & -\mathrm{d}\phi^1 \\ -\mathrm{d}\phi^2 & \mathrm{d}\phi^1 & 1 \end{pmatrix}$$

Which implies the derivative of $\mathbf{B}^b$ is given by

$$\frac{\mathrm{d}\mathbf{B}_b}{\mathrm{d}\vec{\phi}_b} = \mathbf{B}_b\vec{\mathbf{S}}. \quad \text{(S12)}$$

Where $\vec{\mathbf{S}}$ is a vector of skew matrices defined by

$$\vec{\mathbf{S}} = \left[\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\right]^T.$$

From this result the derivatives of the other relevant rotation matrices used above can be calculated. To start the derivatives of the $\mathbf{R}^u$ are given by:

$$\frac{\partial \mathbf{R}_u}{\partial \vec{\phi}_{ul}} = \mathbf{B}_{ur}\vec{\mathbf{S}}^T\mathbf{B}_{ul}^T, \quad \frac{\partial \mathbf{R}_u}{\partial \vec{\phi}_{ur}} = \mathbf{B}_{ur}\vec{\mathbf{S}}\mathbf{B}_{ul}^T. \quad \text{(S13)}$$

For the mid-pair frame $\mathbf{M}_u$ we find:

$$\frac{\partial \mathbf{M}_u}{\partial \vec{\phi}_{ur}} = \frac{\partial \sqrt{\mathbf{R}_u}}{\partial \vec{\phi}_{ur}}\mathbf{B}_{ul} = \frac{1}{2}\sqrt{\mathbf{R}_u}\mathbf{R}_u^T\frac{\partial \mathbf{R}_u}{\partial \vec{\phi}_{ur}}\mathbf{B}_{ul} = \frac{1}{2}\mathbf{M}_u\vec{\mathbf{S}}, \quad \text{(S14)}$$

and likewise:

$$\frac{\partial \mathbf{M}_u}{\partial \vec{\phi}_{ul}} = \frac{1}{2}\mathbf{M}_u\vec{\mathbf{S}}. \quad \text{(S15)}$$

Where we used the fact that $\mathbf{R}_u$, $\sqrt{\mathbf{R}_u}$, $\mathbf{R}_u^T$, and $\sqrt{\mathbf{R}_u^T}$ all commute and we used both representations of the midframe in Eq. (S3). Similar results can be obtained for the lower basepair in a given base-pair step. As such the derivatives of the rotation matrices $\mathbf{R}_s$ and $\mathbf{M}_s$ relevant for the step parameters can also be written down. Doing this gives for the rotation:

$$\frac{\partial \mathbf{R}_s}{\partial \vec{\phi}_{ul}} = \frac{\partial \mathbf{R}_s}{\partial \vec{\phi}_{ur}} = \frac{1}{2}\mathbf{M}_u\vec{\mathbf{S}}\mathbf{M}_d^T, \quad \text{(S16)}$$
$$\frac{\partial \mathbf{R}_s}{\partial \vec{\phi}_{dl}} = \frac{\partial \mathbf{R}_s}{\partial \vec{\phi}_{dr}} = \frac{1}{2}\mathbf{M}_u\vec{\mathbf{S}}^T\mathbf{M}_d^T. \quad \text{(S17)}$$

where it should be noted derivatives to both base in the same pair have become equal. Likewise for the midframe of the step we find:

$$\frac{\partial \mathbf{M}_s}{\partial \vec{\phi}_{ul}} = \frac{\partial \mathbf{M}_s}{\partial \vec{\phi}_{ur}} = \frac{\partial \mathbf{M}_s}{\partial \vec{\phi}_{dl}} = \frac{\partial \mathbf{M}_s}{\partial \vec{\phi}_{dr}} = \frac{1}{4}\mathbf{M}_s\vec{\mathbf{S}}. \quad \text{(S18)}$$

Now we have calculated the derivatives of all rotation matrices and $\mathbf{R}$ and mid frames $\mathbf{M}$ to small rotations of the bases. Subsequently we need to compute the derivatives of axis angle coordinates to the corresponding rotation matrix $\mathbf{R}$. To do this one can simply refer to the formulas to construct the axis angle representation from a matrix in (S4) and (S5), and derive them:

$$\frac{\partial \Theta(\mathbf{R})}{\partial \mathbf{R}} = \frac{-1}{2\sin\Theta(\mathbf{R})}\mathbf{I}_{3x3}, \quad \text{(S19)}$$
$$\frac{\partial \hat{K}(\mathbf{R})}{\partial \mathbf{R}} = \frac{\vec{\mathbf{S}} - \cot\Theta(\mathbf{R})\hat{K}(\mathbf{R}) \otimes \mathbf{I}_{3x3}}{2\sin\Theta(\mathbf{R})} \quad \text{(S20)}$$

With this final bit of information the derivative of any one of our angular coordinates to a

small rotation of a base frame can be written. To do this one needs only the relevant formula (S10) and the equations derived in this section. For example deriving some step rotation α to a rotation of base b about axis k would read:

$$\frac{\partial \Omega_s^\alpha}{\partial \phi_b^k} = \frac{\partial \Theta_s}{\partial \phi_b^k}\left[\hat{K}_s \cdot \mathbf{M}_s[\alpha]\right] + \Theta_s\left[\frac{\partial \hat{K}_s}{\partial \phi_b^k}\cdot \mathbf{M}_s[\alpha] + \hat{K}_s \cdot \frac{\partial \mathbf{M}_s[\alpha]}{\partial \phi_b^k}\right]. \quad \text{(S21)}$$

Here we dropped the explicit indication that Θ_s and $\hat{K}_s$ are functions of $\mathbf{R}_s$. Their derivatives are calculated using the chain rule:

$$\frac{\partial \Theta_s}{\partial \phi_b^k} = \frac{\partial \Theta_s}{\partial \mathbf{R}_s}\frac{\partial \mathbf{R}_s}{\partial \phi_b^k}, \quad \frac{\partial \hat{K}_s}{\partial \phi_b^k} = \frac{\partial \hat{K}_s}{\partial \mathbf{R}_s}\frac{\partial \mathbf{R}_s}{\partial \phi_b^k}. \quad \text{(S22)}$$

Subsequently these derivatives can be combined into the gradient of a rotational rigid base coordinate to the orientation of an underlying base by:

$$\vec{\nabla}_b^r \Omega_s^\alpha = \sum_k \frac{\partial \Omega_s^\alpha}{\partial \phi_b^k}\mathbf{B}_b[k]. \quad \text{(S23)}$$

Note that replacing Ω by ω and s by p in the previous three equation would yield the equivalent formulas for a pair rotation.

S1.3 Rigid body dynamics

Having calculated the gradients of the desired collective variables, we can recall Eq. (2). While the precise gradients to atomic positions $\vec{A}_i$ are not known, and thus we can not write out equations for the atomic forces, the rotational gradients to small rotations of the bases are known. Considering the bases as rigid we can then write out an equation for the torque $\vec{\tau}_b$ applied by some potential $V(X^\alpha)$ on a base $\mathcal{B}_b$

$$\vec{\tau}_b = -\frac{\partial V(X^\alpha)}{\partial X^\alpha}\vec{\nabla}_b^r X^\alpha. \quad \text{(S24)}$$

Here the torque is given in coordinates with respect to the lab frame (the reference frame in which the atomic coordinates are provided by the MD engine). In order to have this torque act during an MD simulation it needs to be converted to a force for every atom i. The needed force can be found as follows

$$\vec{F}_i = m_i\vec{a}_i = m_i[\vec{r}_i \times \vec{\alpha}_b] = m_i\frac{\vec{r}_i \times \vec{\tau}_b}{I_b}. \quad \text{(S25)}$$

Here $\vec{r}_i$ is the vector pointing from the center of mass of the base $\vec{r}_b^{com}$ to the position $\vec{A}_i$, of atom i $\vec{r}_i^T = [\vec{A}_i - \vec{r}_b^{com}]^T$. Additionally I_b has been used to represent the moment of inertia of base with respect to the axis of the torque. Note that as the axis of the torque changes every time the constraint forces need to be calculated, so does I_b. Additionally I_b will change due to fluctuations within the base.

Aside from the rotations $\vec{\omega}$ and $\vec{\Omega}$ one might also want to impose some bias on the translations $\vec{d}$ and $\vec{D}$. These are essentially the displacement between two rigid bodies expressed in the relevant midframe $\mathbf{M}$, and as such the gradients of these to atom position $\vec{A}_i$ are

$$\vec{\nabla}_i d^\alpha = \pm\frac{m_i}{m_{base}}\mathbf{M}_{pair}[\alpha], \quad \text{(S26)}$$

$$\vec{\nabla}_i D^\alpha = \pm\frac{m_i}{m_{pair}}\mathbf{M}_{step}[\alpha]. \quad \text{(S27)}$$

Where the plus is used for the base above the step (noted ul and ur in the diagram of Fig. S1) and for the bases on the righthand strand (noted dr and ur in the diagram of Fig. S1) when considering $\vec{D}$ and $\vec{S}$ respectively.

Finally equations (S24) and (S25) can be combined with equations (S26), (S27) and (2) to give the final force on an atom i in base b (and thus in pair $p = b/2$ and steps $s = b/2$ and $s = b/2-1$) due to a potential on the collective

variables $V(\vec{\omega}_p, \vec{\Omega}_s, \vec{d}_p, \vec{D}_s)$

$$\begin{aligned}\vec{F}_i = &-\sum_\alpha \frac{\partial V}{\partial \Omega^\alpha_{\frac{b}{2}}} \frac{m_i}{I_b}\left[\vec{r}_i \times \vec{\nabla}^r_b \Omega^\alpha_{\frac{b}{2}}\right] \\ &-\sum_\alpha \frac{\partial V}{\partial D^\alpha_{\frac{b}{2}}} \vec{\nabla}_i D^\alpha_{\frac{b}{2}} \\ &-\sum_\alpha \frac{\partial V}{\partial \omega^\alpha_{\frac{b}{2}}} \frac{m_i}{I_b}\left[\vec{r}_i \times \vec{\nabla}^r_b \omega^\alpha_{\frac{b}{2}}\right] \\ &-\sum_\alpha \frac{\partial V}{\partial d^\alpha_{\frac{b}{2}}} \vec{\nabla}_i d^\alpha_{\frac{b}{2}} \\ &-\sum_\alpha \frac{\partial V}{\partial \Omega^\alpha_{\frac{b}{2}-1}} \frac{m_i}{I_b}\left[\vec{r}_i \times \vec{\nabla}^r_b \Omega^\alpha_{\frac{b}{2}-1}\right] \\ &-\sum_\alpha \frac{\partial V}{\partial D^\alpha_{\frac{b}{2}-1}} \vec{\nabla}_i D^\alpha_{\frac{b}{2}-1}. \end{aligned} \tag{S28}$$

For completeness we note that with respect to (2) we have replaced the exact positional gradients with versions that apply rigid body dynamics. For translations they are replaced with gradients weighed to ensure all atoms in a base move together. For rotations we have substituted $\frac{m_i}{I_b}\left[\vec{r}_i \times \vec{\nabla}^r_b\right]$ for $\vec{\nabla}_i$. This last substitution is equivalent to projecting $\vec{\nabla}_i$ onto the plane normal to $\vec{r}$ and again reweighing which ensures that constraint forces do not distort the internal structure of the bases.

S2 Biasing Twist and Roll

To explain the response of the DNA under the effect of a bias in the twist and in the roll, we consider the following free energy model

$$\begin{aligned} f(\Omega_2, \Omega_3) \;=\; & \frac{1}{2}\left\{A_2\Omega_2^2 + C\Omega_3^2 + 2G\Omega_2\Omega_3 + \right. \\ & \left. K_r(\Omega_2 - \bar{\Omega}_2)^2 + K_t(\Omega_3 - \bar{\Omega}_3)^2\right\} \end{aligned} \tag{S29}$$

with Ω_2 and Ω_3 excess roll and twist and A_2, C and G the roll stiffness, the twist stiffness and the twist-roll coupling, respectively. In addition to this free energy, perturbations similar to (5) and (6) have been included. These add a quadratic potential of strengths K_r and K_t around an imposed roll $\bar{\Omega}_2$ and twist $\bar{\Omega}_3$ respectively. Equation (S29) ignores higher order anharmonic terms which become important in the high deformation regime. The stiffnesses A_2, C and G are sequence-dependent.[S3] Here we consider the base pair step considered in Fig. 3, which is the central AT step of a DD sequence. The average excess roll and twist, $\langle\Omega_2\rangle$ and $\langle\Omega_3\rangle$ are obtained from the minimization of Eq. (S29).

We consider first the two cases of a pure twist and pure roll bias. A twist bias corresponds to $K_r = 0$ and $K_t \neq 0$ in (S29). The minimum is obtained by setting the partial derivatives of the free energy f with respect to Ω_2 and Ω_3 to zero. In particular the condition $\partial f/\partial\Omega_2 = 0$ gives

$$\langle\Omega_3\rangle = -\frac{A_2}{G}\langle\Omega_2\rangle \tag{S30}$$

In the case of a roll bias, corresponding to $K_r \neq 0$ and $K_t = 0$ in (S29), the condition $\partial f/\partial\Omega_3 = 0$ gives

$$\langle\Omega_3\rangle = -\frac{G}{C}\langle\Omega_2\rangle \tag{S31}$$

Equations (S30) and (S31) give a linear dependence of $\langle\Omega_3\rangle$ vs. $\langle\Omega_2\rangle$. From prior work we know that $G > 0$,[S4] which implies that $\langle\Omega_3\rangle$ vs. $\langle\Omega_2\rangle$ lines must have negative slopes. This is agreement with the simulation data of Fig. 3(b). Note that the two slopes are different, which is in agreement with (S30) and (S31). On average G is smaller than the two other couplings A_2 and C, which suggests a large slope for a twist bias (S30) and a small slope for a roll bias (S31), which is indeed observed in the simulation data of Fig. 3(c).

From the equilibrium fluctuations of the central dinucleotide step of the DD dodecamer and using the equipartition theorem[S3] we obtain the following estimates of the elastic constants $A_2 = 43.5$ nm, $C = 80$ nm and $G = 10$ nm. We recall that elastic couplings in DNA are both sequence-dependent[S3,S5] and length-scale dependent,[S6,S7] being smaller at short length scales as compared to the asymtptotic values. Fits of the data of Fig. 3(c) in the vicinity of the origin, using Eqs. (S30) and (S31), give the following estimates $A_2/G = 3.8(1)$ and $G/C = 0.24(5)$. The former is rather consistent with the vales of A_2 and G from fluc-

tuations data. The latter value overestimates the expected G/C, which may be due to the quadratic approximation of free energies. We consider next biasing simultaneously the twist and roll with $K_t = K_r = K$ and $\bar{\Omega}_2 = \bar{\Omega}_3$, as done in simulations. Minimizing (S29) under these constraints we find the relation

$$(K + C - G)\langle\Omega_3\rangle = (K + A_2 - G)\langle\Omega_2\rangle \quad \text{(S32)}$$

As $K = 1000$ KJ/mol in simulations, one has $K \gg C, G, A_2$ (in length units as the other elastic constant we get $K = 138$ nm), hence (S32)

$$\langle\Omega_3\rangle \approx \langle\Omega_2\rangle \quad \text{(S33)}$$

consistent with a slope close to one, as observed for the green line of Fig. 3(c). In this limit the stiffness of the biasing potential is much larger than any local bending and torsional stiffness of the DNA molecule.

S3 Biasing twist and base pair opening: non-canonical DNA conformations

A constraint that strongly perturbs the DNA structure can lead to base pair opening, disrupting the canonical DNA form. An example, which we discuss in some detail here, is the base-pair opening induced by a strong over twist, shown in the configuration of Fig. 3(b). One may ask if the biased simulations of the RBB-NA algorithm remain meaningful for non-canonical DNA. This algorithm biases rigid base and rigid base-pair coordinates, which were designed to describe conformations of DNA in which complementary bases remain bound to each other.[S8] When a base pair breaks, the twist Ω_3, as mathematically calculated following Curves+,[S1] may not represent a physically meaningful twist (the relative rotation of two consecutive base pairs around the helical axis). However, Ω_3 remains a well-defined quantity and umbrella sampling simulations estimate the free energy in a correct manner.

RBB-NA can be used to stabilize canonical DNA conformations. To illustrate this, we repeated the simulations with a bias on the twist via Eq. (6) together with a bias on the intra-base coordinate "opening", which measures the angle between complementary base pairs in the plane orthogonal to the helical axis. This was done by adding quartic upper and lower walls to the opening in the two base pairs of the central AT-step of the DD-dodecamer. We used the following potential[1]

$$\begin{aligned} V_{\text{Upper}}(\omega_3 > 0.2 \text{ rad}) &= \frac{K}{2}(\omega_3 - 0.2)^4 \\ V_{\text{Upper}}(\omega_3 \leq 0.2 \text{ rad}) &= 0 \\ V_{\text{Lower}}(\omega_3 < -0.25 \text{ rad}) &= \frac{K}{2}(\omega_3 + 0.25)^4 \\ V_{\text{Lower}}(\omega_3 \geq -0.25 \text{ rad}) &= 0 \end{aligned} \quad \text{(S34)}$$

Here ω_3 is the opening angle and $K = 1000$ kJ/mol. The above potential does not bias ω_3 in $[-0.25, 0.2]$ (an interval chosen based on the typical fluctuations), but strongly suppresses values of ω_3 outside this interval. We performed seven umbrella sampling simulations overtwisting the DNA using (6) (increasing $\bar{\Omega}_3$ by 0.1 radians) and simultaneously enforcing the ω_3 constraint (S34), ensuring the DNA remains closed. The simulations with only a twist bias potential are indicated as TWIST and the simulations with a twist bias potential and the ω_3 constraint are labelled TWIST+ω_3.

Figure S2(a) shows the results of these simulations as a orange line, which can be compared with the pure twist bias simulations (blue line) already discussed in the main text. Both biasing schemes produce overlapping (quadratic) free energies for twist in the range $30° \leq \Omega_3 \leq 45°$. The additional opening constraint produces a free energy which follows the quadratic behavior for a wider range of twist angles (yellow line). The deviations from the parabolic free energy arise from non-canonical DNA con-

[1] To ensure that these constraints only act when the opening is outside the $[-0.25, 0.2]$ interval, "UPPER_WALLS" and "LOWER_WALLS" keywords in PLUMED were used.

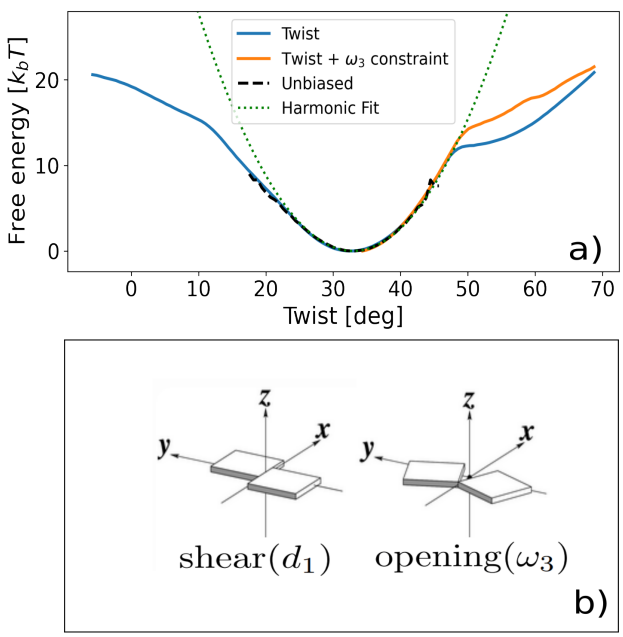

Figure S2: (a) Comparison of umbrella sampling free energy obtained from a bias in twist via Eq. (6), as in the main text (blue line), with that obtained from an additional constraint on the opening parameter given by Eq. (S34) (orange line). (b) Sketch of the intra-base parameters shear (d_1) and opening (ω_3).

formations, which however are different in the two cases. In the pure twist biasing simulation, as discussed in the main text, non-canonical DNA involves base-flipping. The constraint on opening does not allow flipping, but produces a strong shearing of the two bases (Fig. S2(b) illustrates the intra basepair variables opening and shear). This is due to geometric correlations between twist and shear already pointed out in prior literature.[S8] Snapshots of the configurations generated by the simultaneous application of the constraints on twist and opening is shown in Fig. S3.

In conclusion, the examples discussed show how RBB-NA could be useful in studying non-canonical structures, their conformations and free energies, as well. There is a large flexibility on the type of constraint(s) that can be used. As a matter of fact any of the 12 rigid base parameters, in various combinations, could be biased, which may result in different non-canonical structures.

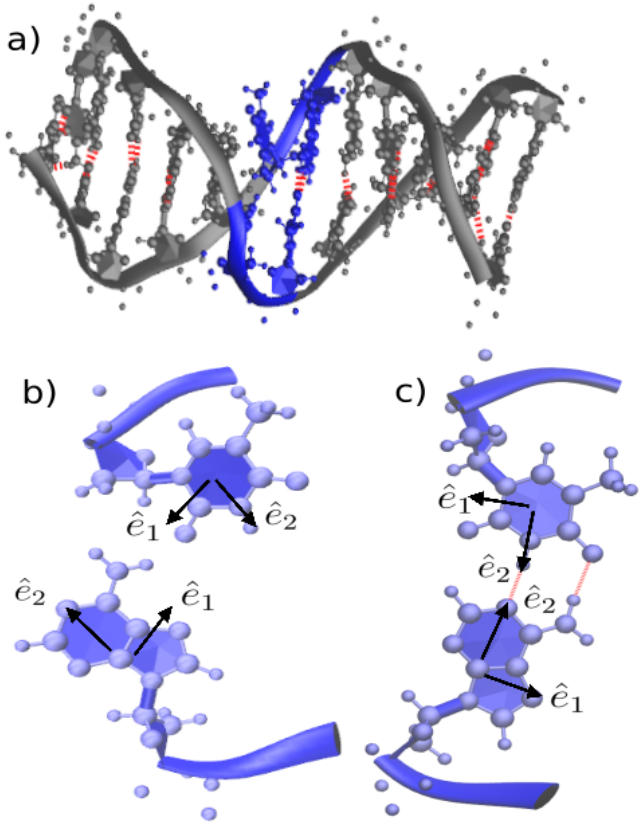

Figure S3: (a) A snapshot of a simulation in which an over-twist (34.4°) and an opening constraint, via (6) and (S34) respectively, are applied simultaneously. The central AT base pair step of the DD sequence, in which the constraints are applied is shown in blue. The red dashed lines indicate hydrogen bonds between complementary bases. (b,c) Detailed view of the structure of the two base pairs forming the central base pair steps of the snapshot (a). Note that the constraint in the opening is applied to both base pairs. Of these two, one assumes a non-canonical conformation (b), while the other shows a canonical Watson-Crick base pairing with hydrogen bonding (c). The simultaneous rotation of both bases in (b) leads to no change in the opening angle, but to a sheared conformation.

References

(S1) Lavery, R.; Moakher, M.; Maddocks, J.; Petkeviciute, D.; Zakrzewska, D. Confor-

mational analysis of nucleic acids revisited: Curves+. *Nucl. Acids Res.* **2009**, *37*, 5917–5929.

(S2) McLachlan, A. D. Gene Duplications in the Structural Evolution of Chymotrypsin. *J. Mol. Biol.* **1979**, *128*, 49–79.

(S3) Lankaš, F.; Šponer, J.; Langowski, J.; Cheatham, T. E. DNA basepair step deformability inferred from molecular dynamics simulations. *Biophys J.* **2003**, *85*, 2872–2883.

(S4) Skoruppa, E.; Nomidis, S.; Marko, J. F.; Carlon, E. Bend-Induced Twist Waves and the Structure of Nucleosomal DNA. *Phys. Rev. Lett.* **2018**, *121*, 088101.

(S5) Lankaš, F.; Lavery, R.; Maddocks, J. H. Kinking occurs during molecular dynamics simulations of small DNA minicircles. *Structure* **2006**, *14*, 1527–1534.

(S6) Noy, A.; Golestanian, R. Length Scale Dependence of DNA Mechanical Properties. *Phys. Rev. Lett.* **2012**, *109*, 228101.

(S7) Skoruppa, E.; Voorspoels, A.; Vreede, J.; Carlon, E. Length-scale-dependent elasticity in DNA from coarse-grained and all-atom models. *Phys. Rev. E* **2021**, *103*, 042408.

(S8) Olson, W. K.; Bansal, M.; Burley, S. K.; Dickerson, R. E.; Gerstein, M.; Harvey, S. C.; Heinemann, U.; Lu, X. J.; Neidle, S.; Shakked, Z.; Sklenar, H.; Suzuki, M.; Tung, C. S.; Westhof, E.; Wolberger, C.; Berman, H. M. A standard reference frame for the description of nucleic acid base-pair geometry. *Journal of Molecular Biology* **2001**, *313*, 229–237.