

Sequence-Dependent Twist-Stretch Coupling in DNA

Timothée Lionnet* and Filip Lankaš†

*Laboratoire de Physique Statistique, Centre National de la Recherche Scientifique, UMR 8550, Ecole Normale Supérieure, 75005 Paris, France; and †Institut de Mathématiques B, Ecole Polytechnique Fédérale de Lausanne, Station 8, CH-1015 Lausanne, Switzerland

ABSTRACT Recent single-molecule micromanipulation experiments on DNA subject to small distortion revealed positive coupling between DNA stretching and twisting—for instance, DNA elongates when overtwisted. Here we propose a method to calculate the twist-stretch coupling constant specific to a DNA fragment of a given sequence. The method employs a sequence-dependent dinucleotide force field and is based on constrained minimization of the fragment's deformation energy. Using a force field inferred from atomistic molecular dynamics simulations, we obtain the twist-stretch coupling for random sequence to be 0.30 nm/turn, close to experimental values. An exhaustive calculation for all oligomers of nine basepairs yields values between 0.14 and 0.45 nm/turn, positively correlated with the contents of pyrimidine-purine steps in the sequence. Our method is simple to use and allows one to explore the hypothesis that some sequences may be optimized for twist-stretch coupling.

Received for publication 15 October 2006 and in final form 21 November 2006.

Address reprint requests and inquiries to Filip Lankaš, E-mail: filip.lankas@epfl.ch.

Mechanical properties of DNA play an important role in a range of biological processes such as DNA binding to proteins or DNA packaging in viruses. The helical structure of DNA implies continuum elastic models with extra terms in the quadratic elastic energy function, absent in the case of a homogeneous isotropic rod. In particular, a coupling term involving both twisting and extension is required (1,2). Previous analysis (1–3) of experimental single-molecule data has suggested that DNA shortens upon overtwisting, in agreement with a simple picture of DNA as a stack of plates connected by rigid backbones. However, two independent micromanipulation studies have recently revealed that kilobase-long DNA molecules under small distortion, gently pulled to reduce thermal fluctuations and suppress stress-induced buckling, actually elongate when overtwisted (4,5) as well as overwind when stretched (5). Within the framework of linear elasticity, this behavior is described by a single proportionality constant indicating the length change upon the imposed rotation, which has been found to be 0.4 ± 0.2 nm/turn by one study (4) and 0.5 ± 0.1 nm/turn by the other (5). This coupling constant does not depend on the length of the molecule. For shorter molecules, it might be more practical to express it in the units of angstroms per degree ($1 \text{ Å/deg} = 0.1 \times 360 \text{ nm/turn} = 36 \text{ nm/turn}$).

These values describe the average property of the long DNA molecules investigated in the experiments. However, it has been proposed (5,6) that proteins may exploit the positive coupling between elongation and overtwisting on a much shorter length scale to recognize binding sites with variable sequence lengths; in case of basepair deletion (or insertion), proteins may recognize the site by simultaneously stretching and overtwisting (or compressing and undertwisting) the DNA. Certain sequences may be optimized for coupled twisting and stretching (5), and the question arises concerning the sequence-dependence of the twist-stretch (TS) coupling. This problem is addressed in this study.

We consider a DNA fragment as a chain of dinucleotide (or basepair (bp)) steps. There are 10 unique dinucleotide sequences. The geometry of each step is described in our model by the six basepair step parameters, namely tilt, roll, twist, shift, slide, and rise (7), defining the translation and rotation of the second basepair in the step with respect to the first one. In the absence of external forces, the parameters adopt their minimum energy values. We assume that the energy associated with the deformation of a basepair step away from its minimum energy geometry has the form of a six-dimensional harmonic potential:

$$E = (1/2) \mathbf{w}^T \mathbf{K}^{(a)} \mathbf{w}, \quad (1)$$

where the index a ($a = 1, \dots, 10$) labels the unique step sequence, E is the deformation energy, $\mathbf{K}^{(a)}$ the sequence-specific stiffness matrix, and \mathbf{w} the six-component column vector containing the deviations of the step parameters from their minimum energy values (T indicates the matrix transpose). We further assume that the total deformation energy of a DNA fragment is a sum of the energies of the individual steps.

Let us look first at the twist-stretch coupling in the individual dinucleotide steps. Suppose that twist ω in excess to the minimum energy twist is imposed on the step. The corresponding deviations of the other parameters can be found by minimizing the energy (1) under the constraint of the imposed twist.

If i is the index of the twist component in \mathbf{w} (in the parameter ordering mentioned above, $i = 3$), we have

$$\begin{aligned} \partial E / \partial w_j &= 0, \quad j \neq i \\ w_i &= \omega, \end{aligned} \quad (2)$$

which yields (we are omitting the superscript a)

$$\sum_{j \neq i} K_{kj} w_j + K_{ki} w_i = 0, \quad k \neq i \text{ (no sum over } i)$$

$$w_i = \omega. \quad (3)$$

The solution to Eq. 3 can be written as $\omega \tilde{\mathbf{w}}$, where $\tilde{\mathbf{w}}$ is the solution to Eq. 3 with $\omega = 1$. Thus, the change of step parameters is proportional to the imposed excess twist, the proportionality coefficients being the components of $\tilde{\mathbf{w}}$. In particular, \tilde{w}_6 is the coupling coefficient between twist and the parameter rise, which measures the step extension along the helical axis. Thus, \tilde{w}_6 (we will subsequently denote it by d) is the twist-elongation coupling at the level of individual dinucleotide steps.

To calculate its sequence-dependent values, we take the matrices $\mathbf{K}^{(a)}$ from a dinucleotide force field based on atomistic molecular dynamics simulations (8). These stiffness data have been recently found to perform particularly well in predicting binding affinity in a protein-DNA complex with an important indirect readout component (9). The results are shown in Table 1. Note that values for different steps may differ by an order of magnitude.

What is the energy cost of imposing excess twist ω on a basepair step? From Eq. 1 we find

$$E = (1/2)(\omega \tilde{\mathbf{w}}^T) \mathbf{K} (\omega \tilde{\mathbf{w}}) = (1/2)(\tilde{\mathbf{w}}^T \mathbf{K} \tilde{\mathbf{w}}) \omega^2 \quad (4)$$

so that the energy is a quadratic function of the excess twist, the effective stiffness constant being $\alpha = \tilde{\mathbf{w}}^T \mathbf{K} \tilde{\mathbf{w}}$. Its sequence-dependent values are summarized in Table 1. Note the difference from the twist stiffness reported earlier (8) where it referred to the diagonal element of \mathbf{K} corresponding to twist, i.e., stiffness with respect to twist while the other parameters retain their minimum energy values. Here, in contrast, α refers to the stiffness associated with twisting while the other parameters are free to relax.

Let us now calculate the twist-stretch coupling constant for an arbitrary DNA molecule. We assume that: a), thermal fluctuations are suppressed, b), all the imposed torsion is distributed into twists of the dinucleotide steps, and c), the total length of the molecule is equal to the sum of rises of the dinucleotide steps. These assumptions are plausible for molecules under moderate tension and fragments of small curvature bound to proteins.

Consider first just two steps in a row, with effective twist stiffness constants α_1 and α_2 . The imposed total excess twist ω will be distributed in the step twists ω_1 and ω_2 , $\omega = \omega_1 + \omega_2$, in such a way that the energy, $E = (1/2)\alpha_1\omega_1^2 + (1/2)\alpha_2\omega_2^2$, be minimal. Performing the minimization, we find $E = (1/2)\alpha_{\text{tot}}\omega^2$ where the total twist stiffness α_{tot} is given by the relation $1/\alpha_{\text{tot}} = 1/\alpha_1 + 1/\alpha_2$. By induction we obtain the

following result: the total twist stiffness α_{tot} for a chain of n steps with twist stiffnesses $\alpha_1, \dots, \alpha_n$ is given by the relation

$$1/\alpha_{\text{tot}} = (1/\alpha_1) + \dots + (1/\alpha_n). \quad (5)$$

Now consider n steps with the total imposed excess twist ω , distributed among the steps: $\omega = \omega_1 + \dots + \omega_n$. To find the excess twists $\omega_1, \dots, \omega_n$ of the individual steps, we have to minimize the deformation energy of the chain, $E = (1/2)\alpha_1\omega_1^2 + \dots + (1/2)\alpha_n\omega_n^2$, with the constraint $\omega = \omega_1 + \dots + \omega_n$. The method of Lagrange multipliers implies unconstrained minimization of the function $\Phi = E - \lambda g$, where $g = \omega_1 + \dots + \omega_n - \omega$. We obtain the following system of $n+1$ linear equations:

$$\partial\Phi/\partial\omega_j = \alpha_j\omega_j - \lambda = 0, \quad j = 1, \dots, n$$

$$\omega_1 + \dots + \omega_n = \omega. \quad (6)$$

It is obvious by mere inspection that the solution to this system is $\omega_j = (\alpha_{\text{tot}}/\alpha_j)\omega$, $\lambda = \omega\alpha_{\text{tot}}$, where α_{tot} is given by Eq. 5. Thus, the excess twist in a particular step is proportional to the ratio of the total twist stiffness of the whole molecule to the twist stiffness of that step. Consequently, steps of the same type get the same amount of excess twist.

Finally, denoting by d_j the twist-rise coupling of the individual steps, we find that the total change in rise is $r = r_1 + \dots + r_n = d_1\omega_1 + \dots + d_n\omega_n$, or

$$r = (d_1\alpha_{\text{tot}}/\alpha_1 + \dots + d_n\alpha_{\text{tot}}/\alpha_n)\omega. \quad (7)$$

But according to assumption c), the total change in rise is equal to the elongation of the molecule. Thus, the expression in the parentheses is the desired TS coupling constant for the whole molecule. Note that equal step types contribute equally to the elongation. Assuming that there are $n^{(a)}$ steps of type a in the molecule ($a = 1, \dots, 10$), we can rewrite Eq. 7 as

$$r = d_{\text{tot}}\omega$$

$$d_{\text{tot}} = d^{(1)}n^{(1)}\alpha_{\text{tot}}/\alpha^{(1)} + \dots + d^{(10)}n^{(10)}\alpha_{\text{tot}}/\alpha^{(10)}$$

$$\alpha_{\text{tot}} = 1/[(n^{(1)}/\alpha^{(1)}) + \dots + (n^{(10)}/\alpha^{(10)})], \quad (8)$$

where $\alpha^{(a)}$ and $d^{(a)}$ are the values of α and d for step of type a taken from Table 1. Equations 8 thus enable one to easily calculate the TS coupling d_{tot} of a DNA molecule provided the number of steps of the ten unique types, $n^{(1)}, \dots, n^{(10)}$, is known.

Alternatively we may express $n^{(a)}$ in terms of the fraction $x^{(a)}$ of the step type a in the sequence with a total of n steps: $n^{(a)} = nx^{(a)}$, so that Eq. 8 now becomes

TABLE 1 Twist-rise coupling d ($\text{\AA} \cdot \text{deg}^{-1}$) and effective twist stiffness α ($\text{kcal} \cdot \text{mol}^{-1} \cdot \text{deg}^{-2}$) as a function of the dinucleotide sequence

| | CG | CA | TA | AG | GG | AA | GA | AT | AC | GC |
|----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| d | 0.0190 | 0.0172 | 0.0078 | 0.0101 | 0.0048 | 0.0042 | 0.0037 | 0.0042 | 0.0020 | 0.0012 |
| α | 0.0173 | 0.0178 | 0.0268 | 0.0307 | 0.0312 | 0.0251 | 0.0218 | 0.0404 | 0.0332 | 0.0261 |

$$r = d_{\text{tot}}\omega$$

$$d_{\text{tot}} = d^{(1)}x^{(1)}\tilde{\alpha}_{\text{tot}}/\alpha^{(1)} + \dots + d^{(10)}x^{(10)}\tilde{\alpha}_{\text{tot}}/\alpha^{(10)}$$

$$\tilde{\alpha}_{\text{tot}} = 1/[(x^{(1)}/\alpha^{(1)}) + \dots + (x^{(10)}/\alpha^{(10)})]. \quad (9)$$

The TS coupling of a sequence, d_{tot} , is thus computed as a certain kind of weighted average of the TS coupling coefficients of the individual basepair steps, $d^{(a)}$. Note that molecules of different length but identical composition would have the same TS coupling constant.

Let us calculate, as an example, the TS coupling for the tetramer ATAT. We have $n^{(\text{AT})} = 2$, $n^{(\text{TA})} = 1$ and Eqs. 8 yield

$$\alpha_{\text{tot}} = 1/(2/\alpha^{(\text{AT})} + 1/\alpha^{(\text{TA})}) = 0.0115 \text{ kcal} \cdot \text{mol}^{-1} \text{ deg}^{-2},$$

$$d_{\text{tot}} = 6.82 \times 10^3 \text{ } \text{\AA}/\text{deg}, \text{ or } 0.25 \text{ nm/turn}.$$

For a long alternating polymer $(\text{AT})_m$, $m \gg 1$, we have $n^{(\text{AT})} = m$, $n^{(\text{TA})} = m - 1 \approx m$ and using Eqs. 9 with $x^{(\text{AT})} = x^{(\text{TA})} = 0.5$ (this is of course equivalent to the use of Eqs. 8 with $n^{(\text{AT})} = n^{(\text{TA})} = 1$) yields 0.23 nm/turn. Similarly we obtain 0.23 nm/turn also for $(\text{GA})_m$, and 0.43 nm/turn for both $(\text{CG})_m$ and $(\text{AC})_m$.

We now calculate the TS coupling for a random sequence. By random we mean that at any position and independently on the rest of the sequence, each of the four nucleotides, A, T, C, G, is present with equal probability, namely 1/4. This implies that any dinucleotide of the form IJ, where I and J can be any of the four, is present with probability 1/16. For dinucleotides that are not self-complementary, one of the 10 unique steps is represented by two different sequences. Thus, the unique steps that are not self-complementary must be taken with a weight of 2 in the calculation, and the self-complementary ones with a weight of 1. Using Eqs. 8 with $n^{(\text{AT})} = n^{(\text{TA})} = n^{(\text{GC})} = n^{(\text{CG})} = 1$ and other $n^{(a)}$ equal to 2, we obtain 0.30 nm/turn as the value of TS coupling for a random sequence.

One may be tempted to calculate the TS coupling in a different way, for instance computing first the arithmetic mean of the stiffness matrices, thus obtaining a sort of “average step”, then using Eq. 3. This procedure, however, seems to lack any rigorous justification and indeed yields, for a random sequence, a value of 0.21 nm/turn, further away from the experimental results. Similarly, a mere arithmetic mean of $d^{(a)}$ over a random sequence would yield 0.26 nm/turn.

Equations 8 or 9 allow one to investigate a large pool of sequences. As an example, we performed an exhaustive calculation for all nine bp oligomers (131,072 unique sequences) and got the distribution of TS coupling constants shown in Fig. 1. The values span the range from 0.14 to 0.45 nm/turn and are positively correlated with the contents of pyrimidine-purine steps in the sequence (correlation coefficient 0.76). The distribution is essentially bimodal (the origin

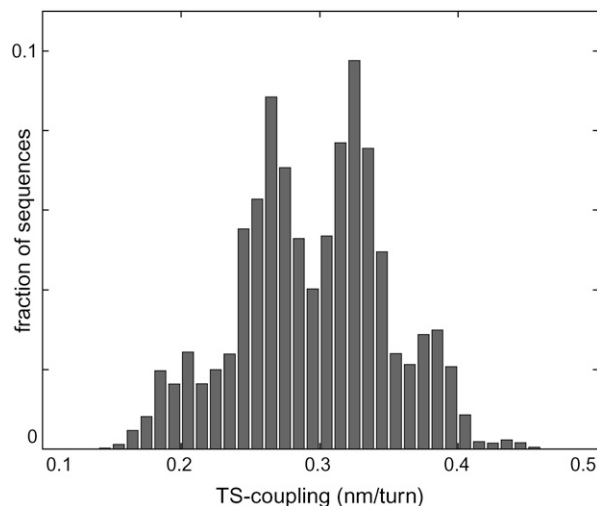


FIGURE 1 Distribution of TS coupling constants in the ensemble of all nine bp sequences. The two highest maxima are at 0.27 and 0.33 nm/turn, respectively.

of this shape remains to be determined). Similar calculations may serve to explore the hypothesis that some sequences may be optimized for TS coupling (5). Such a task is, however, outside the scope of this study.

ACKNOWLEDGMENTS

We thank David Bensimon, Vincent Croquette, and Richard Lavery for helpful discussions.

This work was partially supported by the Swiss National Science Foundation (F.L.) and Centre National de la Recherche Scientifique, Université Pierre et Marie Curie, Paris 6 (T.L.).

REFERENCES and FOOTNOTES

1. Kamien, R. D., T. C. Lubensky, P. Nelson, and C. S. O'Hern. 1997. Direct determination of DNA twist-stretch coupling. *Europhys. Lett.* 38:237–242.
2. Marko, J. F. 1997. Stretching must twist DNA. *Europhys. Lett.* 38:183–188.
3. Moroz, J. D., and P. Nelson. 1998. Entropic elasticity of twist-storing polymers. *Macromolecules.* 31:6333–6347.
4. Lionnet, T., S. Joubaud, R. Lavery, D. Bensimon, and V. Croquette. 2006. Wringing out DNA. *Phys. Rev. Lett.* 96:178102.
5. Gore, J., Z. Bryant, M. Nollmann, M. U. Le, N. R. Cozzarelli, and C. Bustamante. 2006. DNA overwinds when stretched. *Nature.* 442:836–839.
6. Kosikov, K. M., A. A. Gorin, V. B. Zhurkin, and W. K. Olson. 1999. DNA stretching and compression: large-scale simulations of double helical structures. *J. Mol. Biol.* 289:1301–1326.
7. Dickerson, R. E., M. Bansal, C. R. Calladine, S. Diekmann, W. N. Hunter, O. Kennard, R. Lavery, H. C. M. Nelson, W. K. Olson, W. Saenger, Z. Shakked, H. Sklenar, D. M. Soumpasis, C.-S. Tung, E. von Kitzing, A. H.-J. Wang, and V. B. Zhurkin. 1989. Definitions and nomenclature of nucleic acid structure parameters. *J. Mol. Biol.* 205:787–791.
8. Lankas, F., J. Spöner, J. Langowski, and T. E. Cheatham III. 2003. DNA basepair step deformability inferred from molecular dynamics simulations. *Biophys. J.* 85:2872–2883.
9. Becker, N., L. Wolff, and R. Everaers. 2006. Indirect readout: detection of optimized subsequences and calculation of relative binding affinities using different DNA elastic potentials. *Nucleic Acids Res.* 34:5638–5649.