

# Mesoscopic model parametrization of hydrogen bonds and stacking interactions of RNA from melting temperatures

Gerald Weber\*

Departamento de Física, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte-MG, Brazil

Received August 22, 2012; Revised September 21, 2012; Accepted September 24, 2012

## ABSTRACT

Information about molecular interactions in DNA can be obtained from experimental melting temperature data by using mesoscopic statistical physics models. Here, we extend the technique to RNA and show that the new parameters correctly reproduce known properties such as the stronger hydrogen bonds of AU base pairs. We also were able to calculate a complete set of elastic constants for all 10 irreducible combinations of nearest neighbours (NNs). We believe that this is particularly useful as experimentally derived information about RNA elasticity is relatively scarce. The melting temperature prediction using the present model improves over those from traditional NN model, providing thus an alternative way to calculate these temperatures for RNA. Additionally, we calculated the site-dependent base pair oscillation to explain why RNA shows larger oscillation amplitudes despite having stronger AU hydrogen bonds.

## INTRODUCTION

The biological importance of the RNA molecule grows almost on a daily basis and with it the pressing need for computationally efficient models which would be capable of reproducing the intricate physical properties of this molecule. Unlike DNA, RNA has a complex secondary structure and easily forms loops and bulges. Yet even the canonical Watson–Crick base pairing for RNA has received little attention from mesoscopic models such as the Peyrard–Bishop (PB) models.

Our understanding of the flexibility of RNA also lags behind that of DNA. Early studies indicated a large persistence length of RNA when compared with DNA (1), that is that RNA is more rigid than DNA. This is still largely the view which was supported by single molecule (2,3) and atomic force microscopy (AFM) measurements

(4). On a base pair level, that is, on a microscopic scale, most of our knowledge comes from molecular dynamic simulations (5–8). Here, the picture of RNA versus DNA flexibility is less clear. For instance, Noy *et al.* (6) have found from their simulations that RNA is not always more rigid than DNA. These apparent discrepancies between macroscopic and microscopic flexibilities are not unexpected. Even for DNA, it has been difficult to reconcile microscopic flexibility with mesoscopic persistence length measurements (9,10).

The PB model was proposed as a simplified way for the study of DNA denaturation in 1989 (11) and since then has found numerous applications. For instance, it was used recently to study multifractal denaturation process of DNA (12), its melting dynamics (13) and for promoter interaction (14,15). The model can also be used under a variety of theoretical frameworks such as path-integral formalism (16) or wavelet analysis (17), just to give a few examples. One attractive feature of the model is the ease by which one may modify the Hamiltonian to reflect certain properties of DNA such as the addition of solvation barriers (18,19), external forces (20) and other modified potentials (21).

In essence, the PB model proposes a simplified Hamiltonian which takes into account the hydrogen bond and the stacking interaction. These two interactions are modelled in such a way that they use only one variable, for which this model is commonly described as being a 1D model although it is geometrically a 2D description. This single variable is what makes the Hamiltonian tractable by established methods such as the transfer integral method (11).

Although the PB model has seen much use for DNA, there has been, to our knowledge, little research focusing on its application to RNA. This, despite the overwhelming biological importance of RNA and its challenging physics. We understand that one of the main reasons for this situation is the lack of realistic parameters which would capture the essential physics of the RNA molecule.

Here, we use published melting temperatures for RNA sequences to obtain the model parameters for the PB

\*To whom correspondence should be addressed. Tel: +55 31 3409 5633; Fax: +55 31 3409 5600; Email: gweberbh@gmail.com

model. Unlike our previous work with DNA (22), obtaining the model parameters for RNA was much more difficult due to the varying sources of experimental data. The experimental results used here were from Xia *et al.* (23) which include data from several other sources, that is, a large number of these temperatures were obtained under varying experimental conditions. Therefore, we had to take extra steps to ensure that the resulting parameters were of acceptable quality.

One advantage of using the PB model for calculating its parameters is that we can understand qualitatively these parameters in terms of hydrogen bonds and flexibility (22,24). Additionally, we use the model to calculate average base pair openings. We show that despite of a stronger hydrogen bond of the RNA r(AU) base pairs, when compared with DNA d(AT), we obtain larger base pair openings which are entirely consistent with previous findings.

## MATERIALS AND METHODS

### Sequences used

Experimental measurements are from Xia *et al.* (23), they report an experimental uncertainty of 1.3°C for 109 RNA duplexes at 1.0 M NaCl ranging from 4 to 14 bp in length (see Supplementary Tables SI–SIII for sequence information). Average prediction from nearest-neighbour (NN) model is 1.6°C for two-state sequences. For the non-two state sequences, an average prediction of 2.4°C is obtained [see Table 1 of (23)]. In our optimizations, we use all sequences, including those reported as non-two state [Table 1 of (23)]. One should note that Xia *et al.* extend the work by Freier *et al.* (25) and several other sources. Therefore, these melting temperatures were sourced from different groups and were performed under a variety of different experimental conditions.

We also used some sequences reported in the molecular dynamics study of Pan and MacKerell (26) to analyse the base opening of RNA and DNA.

### The mesoscopic model

We used the model proposed by Peyrard and Bishop (11) with harmonic stacking interaction. This simpler model is being favoured in this work over the more elaborate anharmonic model proposed in (27) as we found that the anharmonicity does not improve the melting temperature prediction (22).

The main components of this model are the hydrogen bond represented by a Morse potential, here shown for the example case of a AU base pair,

$$V(\text{AU}) = D_{\text{AU}}(e^{-y_{\text{AU}}/\lambda_{\text{AU}}} - 1)^2 \quad (1)$$

and the stacking interaction

$$w(\text{ApC}) = \frac{k_{\text{ApC}}}{2}(y_{\text{AU}}^2 - 2y_{\text{AU}}y_{\text{CG}} \cos \theta + y_{\text{CG}}^2), \quad (2)$$

here exemplified for a ApC dimer, where  $k$  is the elastic constant and  $\theta$  is a small angle (0.01 rad) introduced to

avoid numerical problems in the partition function integral (28).

For further details on the model implementation, please see (22,29). For the integration of Equation (14) of (29), we use 400 points over the interval  $[-0.1 \text{ nm}, 20.0 \text{ nm}]$  and a cut-off of  $P = 10$  of Equation (22) of (29). The calculation of the thermal index  $\tau$  is performed at 370 K, please note that this temperature is unrelated to the temperatures obtained from the regression method.

### Optimization method

The model parameters were obtained by an optimization method described in (22). For each sequence  $i$ , we calculate an adimensional melting index  $\tau_i$  which results from the PB model and its calculation is described in detail in (22). We use this index  $\tau_i$  to calculate new melting temperatures  $T'_i$  and compare them with the experimental temperatures  $T_i$ . The model parameters are then varied until we minimize the squared differences

$$\chi^2 = \sum_{i=1}^N (T'_i - T_i)^2. \quad (3)$$

We also refer in this work to an average melting temperature deviation

$$\langle \Delta T \rangle = \frac{1}{N} \sum_{i=1}^N |T'_i - T_i|. \quad (4)$$

### Length-dependent regression

As in our previous work on DNA (22), we use two equations,

$$T'_i(N) = a_0(N) + a_1(N)\tau_i, \quad (5)$$

where we split the sequences into separate regression groups of length  $N$ , that is, for each group of length  $N$ , we obtain two regression coefficients  $a_0(N)$  and  $a_1(N)$ . For these, we calculate another linear regression

$$a_k(N) = b_{0,k} + b_{1,k}N^{1/2}, \quad k = 0,1 \quad (6)$$

since we have found that the coefficients  $a_{0,1}$  are essentially linear with  $N^{1/2}$  (30).

### Single equation regression

When the model parameters are very far from their optimized values, the length-dependent regression given in Equation (6) is poorly fitted to a straight line. This causes the coefficients  $b_{0,k}$  and  $b_{1,k}$  to oscillate widely and prevents the parameter minimization from reaching any optimized values. To circumvent this problem, we used a single linear regression for the prediction without length dependence for the initial rounds of minimization

$$T'_i = f_0 + f_1\tau_i. \quad (7)$$

The coefficients  $f_{0,1}$  are linear regression coefficients calculated for all available sequences obtained from the experimental melting temperatures  $T_i$  and  $\tau_i$ .

### Error estimate by varying the melting temperatures

To obtain an error estimate for our calculations, we repeat the minimization procedure over a modified set of melting temperatures. This is achieved by adding a small random amount  $\delta T_i$  (positive or negative) to the reported melting temperature  $T_i$ . The random  $\delta T_i$  follows a Gaussian distribution such that the resulting standard deviation of the modified set matches the reported experimental uncertainty. This procedure is repeated a large number of times, always with a new set of modified temperatures.

### Error estimate by varying initial parameters

For the initial round of minimizations, we found that varying the melting temperatures as described in the previous paragraphs prevents the minimization from reaching optimized values. In this case, we opted to vary the initial parameters instead and keeping the melting temperatures unchanged.

### Minimization procedure

Due to the large experimental uncertainty of the experimental temperatures as well as complete lack of knowledge about approximate parameters for RNA, we performed the optimization in three separate steps.

- (i) Initial minimization. As we had no previous knowledge about approximate model parameters for RNA, we had to initiate the minimization procedure with some extra precautions. Also, we decided not to use the parameters optimized for DNA to avoid biasing our results. Instead, we opted for using the approximate parameters from (29) in exactly the same way as we did for DNA (22). Therefore, we used the fixed values  $\lambda_{\text{AU}} = 3.3333 \times 10^{-2}$  nm,  $\lambda_{\text{CG}} = 1.25 \times 10^{-2}$  nm and a uniform  $k = 2.5$  eV/nm<sup>2</sup>. For each round of minimization, the initial values for the Morse potential were changed randomly and uniformly within  $\pm 20\%$  of the following values  $D_{\text{AU}} = 30$  meV,  $D_{\text{CG}} = 80$  meV. Only the Morse potentials were allowed to vary during the minimization, the remaining parameters were kept as fixed values and we used the regression Equation (7). For the error estimate, we varied only the initial Morse potentials which were repeated 1000 times.
- (ii) First all-parameter minimization. We used the resulting parameters of the previous minimization as initial values for this minimization where now all parameters, except  $\theta$ , are allowed to vary. As in minimization (i), we used the regression Equation (7) and repeated the minimization 1000 times with modified melting temperatures.
- (iii) Final all-parameter minimization. Once we reach an acceptable minimization of  $\chi^2$  in step (ii), we performed the final round of minimization using as input the variables from the previous round, calculation (ii). In this step, the more sensitive length-dependent regression model was used, Equations (5) and (6). We repeated this procedure 200 times with randomly modified data sets to obtain an error estimate.

## RESULTS AND DISCUSSION

### Parameter optimization

Considering the large experimental uncertainties of the melting temperatures (23), we took a few extra precautions for the assessment of the model parameters. First, we performed the optimization by varying only the Morse potentials  $D_{\text{AU}}$  and  $D_{\text{CG}}$  and keeping  $k$  constant for all NNs. These results are shown as calculation (i) in Table 1 and already hint at a stronger r(AU) base pair when compared with d(AT), however with a  $\langle \Delta T \rangle$  twice as large as the uncertainty reported by Xia *et al.* (23).

We use these new Morse potentials of calculation (i) as input for the second round of optimization which now considers all parameters, calculation (ii) in Table 1. This optimization step is performed with a single regression, Equation (7) described in 'Materials and Methods' section. The Morse potential of the r(AU) is increased when compared with calculation (i). Unsurprisingly, given the single linear regression of Equation (7) (see 'Materials and Methods' section), the calculated parameter uncertainty is very large.

The third and final round of parameter optimization, calculation (iii), uses as starting point all the parameters obtained in calculation (ii), but now using the length-dependent regression Equations (5) and (6). Here, we obtain a much smaller uncertainty and an average deviation of melting temperatures  $\langle \Delta T \rangle$  which is comparable to the one obtained from the NN model of 1.6°C. Unlike the NN model (23), however, we do not exclude from our optimization–regression the non-two-state sequences. The average temperature deviation if we consider only the non-two-state sequences of (23) is 2.0°C, which is lower than the 2.4°C reported by Xia *et al.* (23). The complete regression coefficients are listed in Supplementary Tables SIV and SV.

### Hydrogen bonds and base pair oscillations

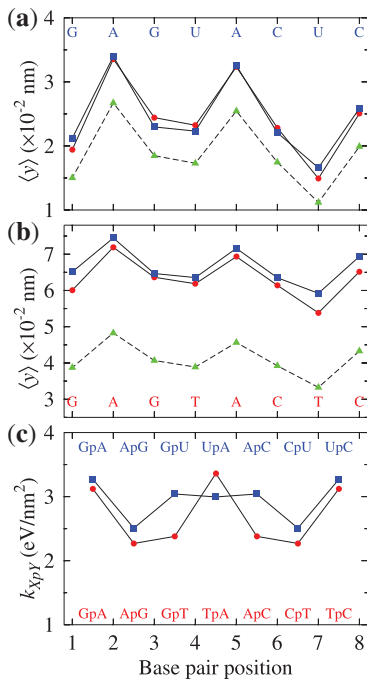
From Table 1, final minimization (iii), we clearly see that the hydrogen bond of 39(3) meV for r(AU) base pairs is stronger than for corresponding d(AT) base pairs of 33(2) meV. The AU hydrogen bond was the object of several studies and determining its strength, which was found to be stronger than AT in DNA, was of considerable experimental and theoretical difficulty (31–33). Also, it is difficult to reconcile the stronger AU hydrogen bond with larger base pair oscillations reported in molecular dynamic simulations, for example the simulations by Pan and MacKerell (26) for RNA, DNA and uracil incorporated in DNA. Larger base pair opening were also observed experimentally in nuclear magnetic resonance (NMR) measurements by Snoussi and Leroy (34) and Várnai *et al.* (35).

To verify the amplitudes of the base pair oscillations, we calculated the average base pair opening  $\langle y \rangle$  following the work by Zhang *et al.* (28). Figures 1 and 2 show these amplitudes for RNA and DNA calculated at the same temperatures. The DNA sequences were calculated with the parameters of (22). Please note that these temperatures are unrelated to the melting temperatures calculated from

**Table 1.** Morse parameters  $D$  and  $\lambda$ 

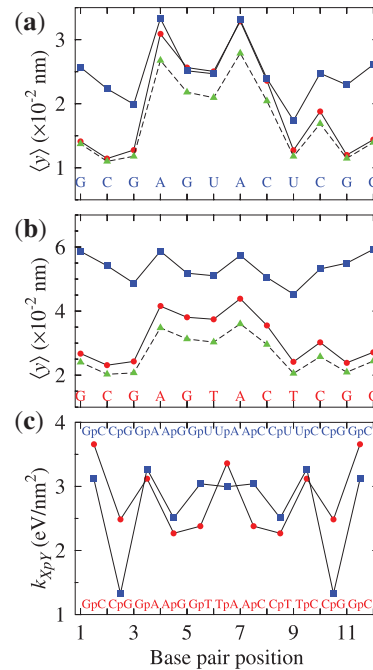
Calculation	$D_{AU}$ (meV)	$D_{CG}$ (meV)	$\lambda_{AU}$ ( $10^{-2}$ nm)	$\lambda_{CG}$ ( $10^{-2}$ nm)	$\langle \Delta T \rangle$	
i	Corresponding parameters for DNA at 1.020 M NaCl (22)	33(2)	70(2)	3.3(8)	0.96(2)	3.2
	Parameters calculated with fixed $k$ and $\lambda$ , varying the initial $D$ by 20%	35(5)	67(5)	—	—	
ii	First round of minimization using Equation (7)	38(10)	71(10)	3(1)	1.1(4)	1.9
iii	Final parameters after second round of minimization using Equations (5) and (6)	39(3)	67(4)	3(1)	0.84(3)	1.7

Shown are the parameters of the three rounds of minimizations (i)–(iii) and their DNA counterparts.



**Figure 1.** Average opening, sequences r(GAGUACUC) (blue boxes, sequence I of (26)), d(GAGTACTC) (red bullets) and d(GAGUACUC) (green triangles, sequence III of (26)). Shown are the average  $\langle y \rangle$  calculated at (a) 180 K and (b) 200 K, and (c) the NN elastic constant  $k$ .

the linear regression in the ‘Materials and Methods’ section and that for very short sequences one has to set the temperature at unrealistically low values. Nevertheless, these base openings provide a qualitative picture for a comparison of the base pair oscillations between DNA and RNA. The amplitudes for RNA were found to be generally larger, especially in Figure 2. We conclude from these results that despite the stronger r(AU) hydrogen bond, the elastic constants are such that we obtain generally larger base pair oscillation  $\langle y \rangle$ . In other words, the larger oscillations are entirely due to the stacking interaction. Also, we do not observe important qualitative variations of the base pair openings in the central parts of the sequence (Figures 1a and 2a) despite important variations in NN flexibility in Figures 1b and 2b. This is consistent with comparative studies by Steinert *et al.* (36) who concluded from NMR measurements that base pair openings are similar in DNA and



**Figure 2.** Average opening, sequences r(GCGAGUACUCGC) (blue boxes, sequence II of (26)), d(GCGAGTACTCGC) (red bullets) and d(GCGAGUACUCGC) (green triangles, sequence IV of (26)). Shown are the average  $\langle y \rangle$  calculated at (a) 200 K and (b) 220 K, and (c) the NN elastic constant  $k$ .

RNA. One should note that comparison of the base pair openings to results from molecular dynamics is not straightforward since the PB model which is used in this work is geometrically 2D, lacking especially the twisting angle. The closest parameter from molecular dynamics which relates to the openings would be the stretching of the base pairs combined to the rise of the NNs (see (37) for the definition of slide and rise).

The Morse potentials for the r(CG) hydrogen bonds are similar to the d(CG) base pairs within the calculated uncertainties (Table 1). The  $D_{CG}$  for RNA were allowed to vary freely during all minimization rounds, and the fact that it resulted in values similar to the DNA Morse potential provides some confidence that the model is consistent for all these datasets.

The Morse potential width parameters  $\lambda$  show little change with either minimization calculations (ii) and (iii) (Table 1). This is similar to what we observed in our

previous results for DNA (22,38), including the large uncertainty for r(AU) which is similar to the one found for d(AT).

With the results for the hydrogen bonds of r(AU), we are now in a position to explore for example the effects of uracil misincorporation in DNA. This is an interesting possibility which underlines the largely unexplored capabilities of the PB model. We simply replace the Morse potential of d(AT) base pairs with the stronger d(AU) potential in DNA. This mimics the occurrence of uracil in DNA which results from uracil misincorporation and is an important epigenetic factor (39). Figures 1 and 2 show a much smaller amplitude of  $\langle y \rangle$  for sequences with deoxyuracil (green triangles), when compared with canonical DNA. This shows that uracil misincorporation appears to have important effect on the DNA base pair oscillation amplitude. Such a finding may be used in bio-informatics applications to search for mutagenic hotspots or to understand the uracil repair mechanism (39).

### Elastic constants

Table 2 shows the calculated elastic constants for RNA compared with their DNA counterparts. Out of 10, only 3 elastic constants are similar in RNA and DNA. For the remaining seven, there is no clear pattern. For instance, the elastic constant of r(CpG) is almost half as large as d(CpG), while r(ApU) is twice as large as d(ApT). This supports the current view that there is very little similarity between the flexibility of RNA and DNA in general and that, depending on the actual sequence, RNA may even be less rigid than DNA (6). This finding is summarized by the calculated rank correlation (40) between the elastic constants of RNA and DNA which is only 0.37.

Figures 1c and 2c illustrate the differences in elastic constants along two different DNA and RNA sequences. For the sequence of Figure 1c, five out of seven RNA NNs do have elastic constants in the same range as DNA. In contrast, for the sequence of Figure 2c, only 5 out of 11 NNs have elastic constants in the same range as for DNA. The two highly flexible positions at the end correspond to CpG NNs, which do have a much smaller elastic constant than that for DNA (Table 2). These localized differences in elastic constants shown in Figures 1c and 2c do not result in equally localized differences of the average base pair opening  $\langle y \rangle$  (corresponding Figures 1a and b and 2a and b). The differences do exist but their effect is seen over several base pair position. For instance, the very soft r(CpG) positions in Figure 2c result in very large base

pair opening towards the extremes of the RNA molecule as shown in Figure 2a and b.

### Applicability and limitations

The applicability of results presented in this work covers their use in the framework of the PB model and for the prediction of melting temperatures. For predicting melting temperatures, the regression parameters were calculated for short RNA sequences in the range of 4–14 bp in length. We expect that predictions would be reliable within this range. For longer sequences however, which may melt in more than one step, the prediction quality is likely to suffer.

Unlike the melting temperature prediction, the PB model itself can in principle be applied to sequences of any length. It is well established that the model captures the dynamics of multistep melting quite well, especially if the more sophisticated variants of the model are used such as the anharmonic model (27). The parameters presented in this work may be used in conjunction with those more elaborate models. For instance, if we use the present PB parameters together with the following anharmonic parameters  $\rho = 2.0$  and  $\alpha = 0.35$  (30) and recalculate the regression, we obtain an average melting temperature deviation  $\langle \Delta T \rangle = 1.8^\circ\text{C}$  which is slightly larger than for the PB model.

The results presented in this work were derived from the experimental data with large experimental uncertainties and from diverse experimental sources (23). Therefore, the larger uncertainties reported here for RNA are to be expected when compared with more recent experimental data for DNA (41). Nevertheless, the convergence of the CG hydrogen bonds for RNA and DNA is an indication that the number of sequences reported by Xia *et al.* (23) is sufficient for an acceptable estimate of the RNA parameters.

### CONCLUSIONS

We obtained the parameters of canonical base pairs for RNA within the framework of the PB model from experimental melting temperature calculations. The prediction of melting temperatures is better than for the NN model with the additional benefit that non-two-state sequences need not to be taken out of the regression analysis. The resulting Morse potentials indicate a stronger r(AU) hydrogen bond which is consistent with recent NMR measurements. The calculated flexibility parameters of

**Table 2.** Elastic constants  $k$  for RNA and their corresponding counterparts in DNA

XpY	$\langle k \rangle$ (RNA)		$\langle k \rangle$ (DNA)	XpY	$\langle k \rangle$ (RNA)		$\langle k \rangle$ (DNA)
CpG	1.3(1)		2.5(2)	ApA = UpU	1.9(1)		2.5(2)
ApU	2.2(2)		1.2(4)	CpA = UpG	2.43(9)		3.1(2)
ApG = CpU	2.5(1)	↔	2.3(3)	CpC = GpG	2.8(2)		2.1(1)
UpA	3.0(3)	↔	3.4(6)	ApC = GpU	3.0(1)		2.4(2)
GpC	3.1(2)		3.7(3)	GpA = UpC	3.3(1)	↔	3.1(2)

Results are for the last minimization round, calculation (iii), and are given in eV/nm<sup>2</sup>. The arrows highlight the elastic constants of DNA and RNA which are similar within the calculated uncertainty. The results for DNA are from (22) for a salt concentration of 1.020 M NaCl.

RNA were found to have little in common with their DNA counterparts as was previously found from molecular dynamics simulations. We expect that this newly obtained set of parameters opens many new possibilities for applying the PB model to RNA on the same level as happened for DNA.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables I–V.

## FUNDING

Fundação de Amparo a Pesquisa do Estado de Minas Gerais (Fapemig); Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq); National Institute of Science and Technology for Complex Systems. Funding for open access charge: Coordenação de Aperfeiçoamento de Nível Superior (Capes).

*Conflict of interest statement.* None declared.

## REFERENCES

- Hagerman,P.J. (1997) Flexibility of RNA. *Annu. Rev. Biophys. Biomol. Struct.*, **26**, 139–156.
- Liphardt,J., Onoa,B., Smith,S.B., Tinoco,I. Jr and Bustamante,C. (2001) Reversible unfolding of single RNA molecules by mechanical force. *Science*, **292**, 733–737.
- Abels,J.A., Moreno-Herrero,F., Van der Heijden,T., Dekker,C. and Dekker,N.H. (2005) Single-molecule measurements of the persistence length of double-stranded RNA. *Biophys. J.*, **88**, 2737–2744.
- Bonin,M., Zhu,R., Klaue,Y., Oberstrass,J., Oesterschulze,E. and Nellen,W. (2002) Analysis of RNA flexibility by scanning force spectroscopy. *Nucleic Acids Res.*, **30**, e81–e81.
- Zacharias,M. and Sklenar,H. (2000) Conformational deformability of RNA: a harmonic mode analysis. *Biophys. J.*, **78**, 2528–2542.
- Noy,A., Pérez,A., Lankas,F., Javier Luque,F. and Orozco,M. (2004) Relative flexibility of DNA and RNA: a molecular dynamics study. *J. Mol. Biol.*, **343**, 627–638.
- Orozco,M., Noy,A. and Pérez,A. (2008) Recent advances in the study of nucleic acid flexibility by molecular dynamics. *Curr. Opin. Struct. Biol.*, **18**, 185–193.
- Faustino,I., Pérez,A. and Orozco,M. (2010) Toward a consensus view of duplex RNA flexibility. *Biophys. J.*, **99**, 1876–1885.
- Gross,P., Laurens,N., Oddershede,L.B., Bockelmann,U., Peterman,E.J.G. and Wuite,G.J.L. (2011) Quantifying how DNA stretches, melts and changes twist under tension. *Nat. Phys.*, **7**, 731–736.
- Van Eijck,L., Merzel,F., Rols,S., Ollivier,J., Forsyth,V.T. and Johnson,M.R. (2011) Direct determination of the base-pair force constant of DNA from the acoustic phonon dispersion of the double helix. *Phys. Rev. Lett.*, **107**, 88102.
- Peyrard,M. and Bishop,A.R. (1989) Statistical mechanics of a nonlinear model for DNA denaturation. *Phys. Rev. Lett.*, **62**, 2755–2757.
- Behnia,S., Akhshani,A., Panahi,M., Mobaraki,A. and Ghaderian,M. (2012) Multifractal properties of denaturation process based on Peyrard–Bishop model. *Phys. Lett. A*, **376**, 2538–2547.
- van Erp,T.S. and Peyrard,M. (2012) The dynamics of the DNA denaturation transition. *EPL (Europhys. Lett.)*, **98**, 48004.
- Tapia-Rojo,R., Prada-Gracia,D., Mazo,J.J. and Falo,F. (2012) Mesoscopic model for free-energy-landscape analysis of DNA sequences. *Phys. Rev. E*, **86**, 021908.
- Alexandrov,B.S., Fukuyo,Y., Lange,M., Horikoshi,N., Gelev,V., Rasmussen,K.Ø., Bishop,A.R. and Usheva,A. (2012) DNA breathing dynamics distinguish binding from nonbinding consensus sites for transcription factor YY1 in cells. *Nucleic Acids Res.*, **40**, 10116–10123.
- Zoli,M. (2010) Denaturation patterns in heterogeneous DNA. *Phys. Rev. E*, **81**, 051910.
- Machado,R.F. and Weber,G. (2009) Wavelet coefficients as a guide to DNA phase transitions. *Europhys. Lett.*, **87**, 38005.
- Weber,G. (2006) Sharp DNA denaturation due to solvent interaction. *Europhys. Lett.*, **73**, 806–811.
- Tapia-Rojo,R., Mazo,J.J. and Falo,F. (2010) Thermal and mechanical properties of a DNA model with solvation barrier. *Phys. Rev. E*, **82**, 031916.
- Sulaiman,A., Zen,F.P., Alatas,H. and Handoko,L.T. (2012) The thermal denaturation of the Peyrard–Bishop model with an external potential. *Phys. Scr.*, **86**, 015802.
- Ribeiro,N.F. and Drigo Filho,E. (2011) Thermodynamics of a Peyrard–Bishop one-dimensional lattice with on-site hump potential. *Braz. J. Phys.*, **41**, 195–200.
- Weber,G., Essex,J.W. and Neylon,C. (2009) Probing the microscopic flexibility of DNA from melting temperatures. *Nat. Phys.*, **5**, 769–773.
- Xia,T., SantaLucia,J. Jr, Burkard,M.E., Kierzek,R., Schroeder,S.J., Jiao,X., Cox,C. and Turner,D.H. (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson–Crick base pairs. *Biochemistry*, **37**, 14719–14735.
- Fagundes-Lima,D. and Weber,G. (2012) Comment on SCS: signal, context, and structure features for genome-wide human promoter recognition. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **9**, 940–941.
- Freier,S.M., Kierzek,R., Jaeger,J.A., Sugimoto,N., Caruthers,M.H., Neilson,T. and Turner,D.H. (1986) Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl Acad. Sci. USA*, **83**, 9373.
- Pan,Y. and MacKerell,A.D. Jr (2003) Altered structural fluctuations in duplex RNA versus DNA: a conformational switch involving base pair opening. *Nucleic Acids Res.*, **31**, 7131–7140.
- Dauvois,T., Peyrard,M. and Bishop,A.R. (1993) Entropy-driven DNA denaturation. *Phys. Rev. E*, **47**, R44–R47.
- Zhang,Y.-L., Zheng,W.-M., Liu,J.-X. and Chen,Y.Z. (1997) Theory of DNA melting based on the Peyrard–Bishop model. *Phys. Rev. E*, **56**, 7100–7115.
- Weber,G., Haslam,N., Essex,J.W. and Neylon,C. (2009) Thermal equivalence of DNA duplexes for probe design. *J. Phys. Condens. Matter*, **21**, 034106.
- Weber,G., Haslam,N., Whiteford,N., Prügel-Bennett,A., Essex,J.W. and Neylon,C. (2006) Thermal equivalence of DNA duplexes without melting temperature calculation. *Nat. Phys.*, **2**, 55–59.
- Vakonakis,I. and LiWang,A.C. (2004) N1...N3 hydrogen bonds of A:U base pairs of RNA are stronger than those of A:T base pairs of DNA. *J. Am. Chem. Soc.*, **126**, 5688–5689.
- Swart,M., Guerra,C.F. and Bickelhaupt,F.M. (2004) Hydrogen bonds of RNA are stronger than those of DNA, but NMR monitors only presence of methyl substituent in uracil/thymine. *J. Am. Chem. Soc.*, **126**, 16718–16719.
- Kim,Y.I., Manalo,M.N., Peréz,L.M. and LiWang,A. (2006) Computational and empirical trans-hydrogen bond deuterium isotope shifts suggest that N1–N3 A:U hydrogen bonds of RNA are shorter than those of A:T hydrogen bonds of DNA. *J. Biomol. NMR*, **34**, 229–236.
- Snoussi,K. and Leroy,J.L. (2001) Imino proton exchange and base-pair kinetics in RNA duplexes. *Biochemistry*, **40**, 8898–8904.
- Várnai,P., Canalia,M. and Leroy,J.L. (2004) Opening mechanism of G. T/U pairs in DNA and RNA duplexes: a combined study of imino proton exchange and molecular dynamics simulation. *J. Am. Chem. Soc.*, **126**, 14659–14667.
- Steinert,H.S., Rinnenthal,J. and Schwalbe,H. (2012) Individual basepair stability of DNA and RNA studied by NMR-detected solvent exchange. *Biophys. J.*, **102**, 2564–2574.

37. 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. *et al.* (2001) A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.*, **313**, 229–237.
38. Weber,G. (2011) Finite enthalpy model parameters from DNA melting temperatures. *Europhys. Lett.*, **96**, 68001.
39. Olinski,R., Jurgowiak,M. and Zaremba,T. (2010) Uracil in DNA—its biological significance. *Mutat. Res./Rev. Mutat. Res.*, **705**, 239–245.
40. Press,W.H., Teukolsky,S.A., Vetterling,W.T. and Flannery,B.P. (1988) *Numerical Recipes in C*. Cambridge University Press, Cambridge.
41. Owczarzy,R., You,Y., Moreira,B.G., Manthey,J.A., Huang,L., Behlke,M.A. and Walder,J.A. (2004) Effects of sodium ions on DNA duplex oligomers: improved predictions of melting temperatures. *Biochemistry*, **43**, 3537–3554.