A Letters Journal Exploring the Frontiers of Physics

December 2011

EPL, **96** (2011) 68001
doi: 10.1209/0295-5075/96/68001

www.epljournal.org

Finite-enthalpy model parameters from DNA melting temperatures

G. Weber[(a)]

Department of Physics, Federal University of Minas Gerais - 31270-901 Belo Horizonte-MG, Brazil

received 3 October 2011; accepted in final form 31 October 2011
published online 15 November 2011

PACS 87.14.gk – Biomolecules: types: DNA
PACS 87.15.Zg – Biomolecules: structure and physical properties: Phase transitions
PACS 87.15.A- – Biomolecules: structure and physical properties: Theory, modeling, and computer simulation

Abstract – Peyrard-Bishop (PB) models are used for the study of denaturation in DNA. Unfortunately, there is little connection of these models to linear nearest-neighbour models which are extensively used for the calculation of melting temperatures in biochemistry. Here we use the Joyeux-Buyukdagli (JB) model, a variant of the PB model which incorporates stacking enthalpies, and carry out a fitting procedure to experimental melting temperatures where we let the enthalpies vary freely. We start out with a single value for the enthalpy for all combinations of base pairs and after the fitting we obtain a new set of enthalpies which correlate very strongly with the measured enthalpies. This result provides the needed support for the use of experimental enthalpies in the JB/PB model.

One of the earliest approaches to predict experimental melting temperatures are first-neighbour or nearest-neighbour (NN) models [1–4]. In this approach, the double-stranded DNA sequence is broken up into units of sequential neighbouring base pairs. To each of these nearest neighbours one assigns a value for entropy and enthalpy representing the stability of the molecule. This results in a simple linear model and the entropy and enthalpy are easily obtained from melting temperatures with the use of standard numerical techniques. The resulting parameters can then be used to predict melting temperatures for unknown DNA sequences and its good accuracy has made it a very popular method. The NN models, however, do not provide much insight into what is happening to the DNA molecule. For instance, they are of limited use when it comes to understand the local base pair opening during denaturation.

Fortunately, the denaturation of DNA can be modelled by a large number of theoretical techniques, ranging from Ising-type Hamiltonians [5] to atomistic molecular dynamics [6]. One important model, the Peyrard-Bishop (PB) model [7], finds a compromise between computational efficiency and physics by employing a Hamiltonian which accounts for the basic ingredients of the stability of the DNA molecule: the hydrogen bonds and the stacking interaction. It can be used to calculate the average openings of the double strand from an equilibrium partition function which can be correlated to melting temperatures of DNA [8]. PB models have found their use in numerous applications, for instance they were recently applied to the study of thermal transport in DNA [9], genomic melting [10], pre-melting dynamics of DNA [11], to the analysis of localisation in DNA [12], and to study the order of the denaturation transition [13]. Please note that this list highlights just a few recent applications and is by no means exhaustive. Given the broad application of the model, there is an active interest in pursuing modifications to the Hamiltonian to cover a range of important properties of the DNA molecule. For instance, it is possible to add solvation barriers [14,15] and other modified potentials [16]. Using this model the thermodynamic properties of oligonucleotides can be studied with a variety of theoretical techniques such as wavelet analysis [17], Langevin [18] or Fokker-Plank formalism [19], again just to cite but a few examples.

Unfortunately, the PB model suffers from a chronic lack of realistic parameters. One of the earliest works towards realistic parameters, and still one of the most frequently used, was obtained by Campa and Giansanti [20]. More recently, stacking terms were obtained from Monte Carlo

[(a)]E-mail: gweberbh@gmail.com

simulations by Alexandrov *et al.* [21] and from direct fitting of experimental melting temperatures [22]. One way to overcome this problem would be to use well-established free-energy parameters from NN models such as the unified set from SantaLucia [3]. Unfortunately, using these parameters is not straightforward for the PB model, except for a variant of the PB model proposed by Joyeux and Buyukdagli (JB) [23–26] which uses a finite-enthalpy stacking term. In principle, one may use enthalpy parameters from NN models for the JB model and in fact this has been attempted with promising results [23,27]. Clearly, further independent validation of this procedure would be valuable.

In this work we calculate a new set of finite stacking enthalpies for the JB model from experimental melting temperatures with the use of the fitting technique which we developed recently [22]. We will show that this new independent set of enthalpies correlates very strongly to the enthalpies from the unified SantaLucia set of parameters [3] which is essentially the gold standard in terms of NN model calculations.

The JB model [23–26] incorporates a finite-enthalpy stacking term ΔH_{XpY} in the Hamiltonian,

$$W(\mathrm{XpY}) = \frac{\Delta H_{\mathrm{XpY}}}{C}\left[1-\exp\left(-b(y_n-y_{n-1})^2\right)\right] + K_b(y_n-y_{n-1})^2 \quad (1)$$

where XpY are the nearest-neighbour base pairs, X and Y from $5'$ to $3'$. For instance, an AT base pair (X=A) followed by CG (Y=C) will be represented as ApC in this work. C is a constant and K_b was described as a backbone stiffness constant which is several orders of magnitude smaller than the elastic constant usually employed in PB models [7,28]. The Morse potential is given in its usual form [7]

$$V(\mathrm{XY}) = D_{\mathrm{XY}}(e^{-y_n/\lambda_{\mathrm{XY}}}-1)^2. \quad (2)$$

where D_{XY} and λ_{XY} are the potential depth and width of the XY base pair.

Recently, we developed a fitting procedure to adjust PB model parameters from melting temperatures [22]. With this method we were able to obtain a much better fit to experimental melting temperatures and gained an insight into the physics of DNA, especially in regard to the elastic constants. This method uses the equivalent melting index [8] which speeds up the numerical calculation and makes the fitting procedure feasible. The method is computationally efficient enough to allow for a robust statistics estimation of the error of these parameters. The basic approach is to calculate the predicted melting temperatures for a large set of sequences for which melting temperatures were measured. Initially, we start with a tentative set of parameters which are optimised using standard numerical methods to obtain progressively better fittings to the experimental data, that is, we seek to reduce the squared difference between the measured temperatures T_i and predicted temperatures T_i',

$$\chi^2 = \sum_{i=1}^{N}\left(T_i'-T_i\right)^2. \quad (3)$$

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

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

for a more intuitive discussion of the accuracy of the melting temperature predictions.

To avoid common problems associated to the optimisation such as local minima and to estimate the relative uncertainty of the optimised parameters, the optimisation was repeated several times with a modified set of melting temperatures. We modify the original experimental temperatures by small random amounts ΔT_i such that the standard deviation from the original data matches the reported experimental error. The optimisation is then carried out again and new sets of parameters are obtained which provide the closest fit to the modified data. This procedure is repeated several times, creating new modified sets of experimental data and optimising them all over again. Eventually, after a large enough number of optimisations we obtain average parameters and their respective standard deviations.

In this work we used published melting temperature data from ref. [29] which allow us to explore the salt dependence of melting temperatures. The melting temperatures were modified by small random amounts, as described previously, and we obtained a standard deviation of 0.3 °C which is similar to the reported experimental uncertainty [29].

The initial parameters used for the JB model [30] were: a uniform enthalpy of $\Delta H = 0.409\,\mathrm{eV}$ for all NN pairs. The initial Morse parameters were $D_{\mathrm{AT}} = 41\,\mathrm{meV}$, $D_{\mathrm{CG}} = 54\,\mathrm{meV}$, $\lambda_{\mathrm{AT}} = \lambda_{\mathrm{CG}} = 1.7\times10^{-2}\,\mathrm{nm}$. The specific initial parameters for the JB model were $b = 80\,\mathrm{nm}^{-2}$, $K_b = 40\,\mathrm{meV}\cdot\mathrm{nm}^{-2}$ and $C = 4$. When using these initial parameters we obtained melting temperatures with an average temperature deviation $\langle\Delta T\rangle = 5.6\,^{\circ}\mathrm{C}$ and $\chi^2 = 18000\,^{\circ}\mathrm{C}^2$, when compared to the melting temperatures of ref. [29]. All these parameters were allowed to vary during the optimisation, except for the parameter C which was kept constant.

The optimisation was carried out in two steps. First, we optimised the parameters while keeping a log dependence for the regression parameters, see eq. (37) of ref. [31], in this way all five sets are optimised simultaneously. For this step we performed 100 optimisations with randomly modified melting temperatures and we obtained a new average melting temperature deviation $\langle\Delta T\rangle = 1.0\,^{\circ}\mathrm{C}$. The resulting Morse potentials were $\langle D_{\mathrm{AT}}\rangle = 28(3)\,\mathrm{meV}$ and $\langle D_{\mathrm{CG}}\rangle = 65(3)\,\mathrm{meV}$ which deviate significantly from

Table 1: Average finite enthalpies $\langle \Delta H_{\mathrm{XpY}} \rangle$ for 100 rounds of optimisations with log-dependent regression parameters. Also shown are nearest-neighbour equivalent representations. The standard deviations are shown in parenthesis as uncertainties to the last significant digit.

XpY	$\langle \Delta H \rangle$ (eV)	XpY	$\langle \Delta H \rangle$ (eV)
ApT	0.35(4)	TpA	0.38(3)
CpC=GpG	0.39(1)	ApA=TpT	0.39(2)
ApG=CpT	0.39(2)	ApC=GpT	0.42(2)
CpG	0.43(2)	GpA=TpC	0.43(1)
CpA=TpG	0.45(2)	GpC	0.51(3)

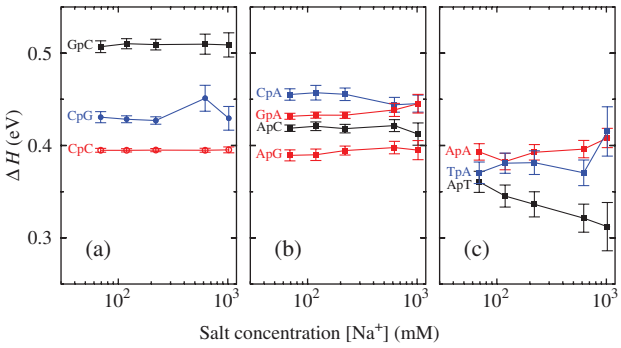

Fig. 1: (Colour on-line) Average finite enthalpy $\langle \Delta H_{\mathrm{XpY}} \rangle$ as function of salt concentration [Na $^+$]. The enthalpies are grouped according to the strength of the hydrogen bond: (a) strong, (b) intermediate and (c) weak. Error bars are standard deviations from 250 runs.

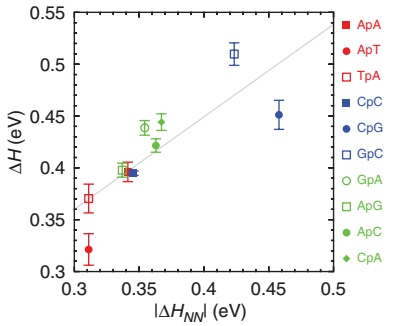

Fig. 2: (Colour on-line) Correlation between the calculated finite enthalpies $\langle \Delta_{\mathrm{XpY}} \rangle$ and the enthalpies obtained from nearest-neighbour models [3]. The calculated finite enthalpies shown are for 620 mM salt concentration which displays the best ranking coefficient $\rho_{\Delta H(621)|3} = 0.92$ and the lowest p-value 0.0014.

their initial values. The potential widths $\langle \lambda_{\mathrm{CG}} \rangle = 1.5(2) \times 10^{-2}$ nm and $\langle \lambda_{\mathrm{AT}} \rangle = 1.8(3) \times 10^{-2}$ nm, on the other hand, show only a minor change. The model-specific constants $\langle b \rangle = 79(8)\,\mathrm{nm}^{-2}$ and $\langle K_b \rangle = 40(3)\,\mathrm{meV \cdot nm}^{-2}$ essentially vary around their initial values. The finite enthalpies, shown in table 1, all deviate from the initial value.

We use the Spearman's rank correlation coefficient $\rho_{\mathrm{Set\,1|Set\,2}}$ [32] to evaluate the correlation between our calculated finite enthalpies and those derived from NN models. We will represent the sets either by the variable symbol or by a reference citation number, for instance $\rho_{\Delta H|3}$ is the correlation between the enthalpies ΔH and the corresponding data from ref. [3]. This ranking correlation shows if the values of the two sets do increase or decrease in the same order. When compared to the enthalpies obtained from the NN model by SantaLucia [3] we obtain a strong rank correlation of $\rho_{\Delta H|3} = 0.89$. For comparison, the rank coefficient for the enthalpies obtained by Huguet *et al.* [33] from single-molecule measurements, when compared to those of SantaLucia [3] is only $\rho_{33|3} = 0.73$.

The next step of optimisation is to use the average parameters from table 1 as new initial parameters, but this time we optimise independently for each salt concentration, that is, we do not use eq. (37) of ref. [31]. Since each set contains less data we increased the number of randomised sets to 250. In fig. 1 we show the resulting average enthalpies $\langle \Delta H_{\mathrm{XpY}} \rangle$ of the second optimisation step as a function of the salt concentration. For most enthalpies (fig. 1(a) and (b)) we observe no pronounced variation with salt concentration. This bears some relation to enthalpies from NN models which are generally assumed not to vary with salt concentration [3]. However, for enthalpies related to weakly bonded nearest-neighbours we observe an important dependence with salt concentration, in a very similar way as was found for elastic constants in the canonical PB model [22]. These similarities were expected since the elastic constants k were replaced by finite enthalpies ΔH in the Hamiltonian of eq. (1) and the exponential dependence with base pair distances y essentially accounts for most of the differences seen in fig. 1.

The correlation of the fitted enthalpies with those from nearest-neighbour models are exemplified in fig. 2 where we show the enthalpies for a salt concentration of 621 mM and the enthalpies reported by SantaLucia [3]. The Spearman's rank correlation coefficient is $\rho_{\Delta H(621)|3} = 0.92$, which is generally agreed to be a very high-ranking correlation. Still, given the small number of parameters this rank coefficient could be misleading. Therefore, we also evaluated the p-value which is the probability of obtaining this rank coefficient by chance. We calculated the p-value by randomly reassigning the enthalpies to different nearest neighbours and counting how many times we obtained the same or a larger rank correlation coefficient. For 621 mM we obtain a p-value of 0.0014 which indicates a very low probability of obtaining the given rank correlation by chance. The poorest rank coefficient found is 0.73 for a salt concentration of 1020 mM which is mainly due to the change in position of TpA and ApA seen in fig. 1(c). The complete set of ranking coefficients are shown in table 2. The evaluation of the linear regression, however, given by its goodness of fit R^2 [32] of 0.66, is somewhat poorer. Nevertheless, this still larger than the $R^2_{33|3} = 0.43$ for the enthalpies of Huguet *et al.*

Table 2: Statistical coefficients as a function of salt concentration for finite stacking enthalpies and nearest-neighbour enthalpies from ref. [3]. Shown are the rank correlation coefficient $\rho_{\Delta H|3}$ between calculated enthalpies and NN enthalpies from ref. [3], their respective goodness of fit $R^2_{\Delta H|3}$ and p-values $p_{\Delta H|3}$; the rank correlation coefficient $\rho_{\Delta H|34}$ between calculated enthalpies and NN enthalpies from ref. [34], and $\rho_{k|3}$ between calculated elastic constants k from ref. [22] and NN enthalpies from ref. [3]. Also shown are the optimisation merit values $\langle \Delta T \rangle$ and χ^2.

$[Na^+]$ (mM)	$\rho_{\Delta H\|3}$	$R^2_{\Delta H\|3}$	$p_{\Delta H\|3}$	$\rho_{\Delta H\|34}$	$\rho_{k\|3}$	$\langle \Delta T \rangle$ (°C)	χ^2 (°C^2)
all		— initial parameters —				5.6	18000
all	0.89	0.54	0.015	0.68	—	1.0	830
69	0.90	0.56	0.014	0.64	0.76	0.83	120
119	0.89	0.52	0.017	0.68	0.75	0.83	120
220	0.89	0.51	0.017	0.68	0.59	0.74	100
621	0.92	0.66	0.0014	0.72	0.76	0.83	110
1020	0.73	0.40	0.023	0.39	0.36	0.80	120

Since the results for finite enthalpies of fig. 1 are qualitatively similar to those of the elastic constants (see fig. 2 of ref. [22]) one may ask whether the elastic constants k may also correlate with the NN enthalpies. We calculated the rank correlation $\rho_{k|3}$ of the elastic constants k presented in ref. [22] and obtained values ranging from 0.36 to 0.76, see table 2. Therefore, the enthalpies show a significantly better correlation than the elastic constants when compared to the results from SantaLucia [3].

Joyeux and Buyukdagli [23,27] used a set of experimental NN enthalpies from Blake *et al.* [34,35] which were measured for tandem sequences inserted in recombinant plasmid at low salt concentration. However, the rank correlation $\rho_{\Delta H|34}$ shown in table 2 are not as good as for the unified SantaLucia set [3].

The linear regression calculated for 621 mM salt concentration, shown as a straight line in fig. 2, is given by

$$\Delta H = 0.0924 + 0.892|\Delta H_{\rm NN}|. \qquad (5)$$

We may ask: how worse does the fit to the experimental melting temperatures become if we use enthalpy parameters calculated with eq. (5) instead of the actual ones shown in fig. 2? We performed this calculation and obtained a $\langle \Delta T \rangle$ of 0.93 °C, roughly 10% up compared to the optimised parameters shown in table 2. Therefore, the use of eq. (5) is already a quite acceptable approximation.

The Morse potential parameters, shown in fig. 3 are qualitatively very similar to those for the harmonic model [22]. The Morse potential depth D for the JB model is somewhat smaller for CG base pairs than for the harmonic Morse potential [22]. The potential widths λ for both types of base pairs are roughly of the same magnitude in contrast to the harmonic model where they are markedly different [22], yet this is likely due to the initial parameters used for both models. For the JB model we used the same potential width λ as initial parameter for AT and CG base pairs, while for the canonical PB model [22] we used different initial values. The parameters introduced in the finite-enthalpy model, b and k_b, remain largely constant with varying salt concentration as shown in fig. 4.

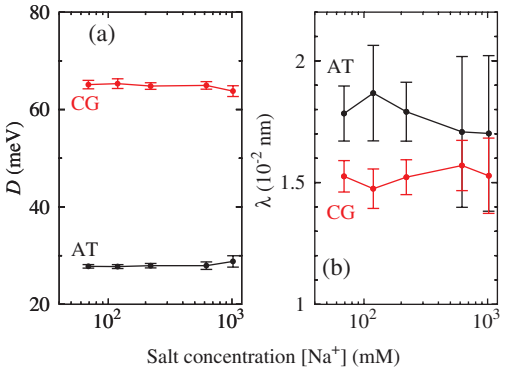

Fig. 3: (Colour on-line) Average Morse potential parameters (a) $\langle D \rangle$ and (b) $\langle \lambda \rangle$ as a function of salt concentration.

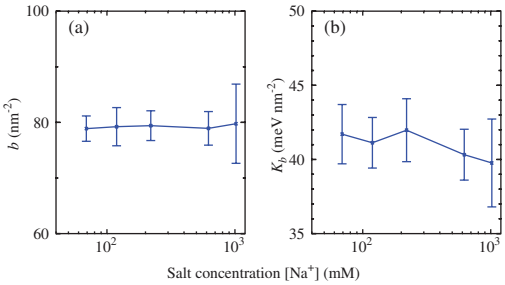

Fig. 4: (Colour on-line) Average finite-enthalpy parameters $\langle b \rangle$ and $\langle K_b \rangle$ as functions of salt concentration.

A number of new experimental techniques to study local melting or bubble opening were developed recently. For instance, Cuesta-Lopez *et al.* [36] used guanine probes to spatially resolve bubbles in DNA, and Reisner *et al.* [37] demonstrated how local melting can be used as a barcoding method in nanofluidic channels. The NN model, while being successful in predicting average melting temperatures, cannot provide the necessary details of the melting mechanism which are needed for the interpretation of these experimental results (see ref. [38] for an interesting discussion on this). We believe that validating the use of NN enthalpies in the JB model provides an increased level of confidence for using PB-type models to study local DNA

melting. More importantly perhaps, it makes the application of PB-type models to non-DNA oligonucleotides a lot more straightforward. One attractive possibility is the study of local melting in RNA. Given the validation presented in this letter one should be able to use, with reasonable confidence, the experimental NN enthalpies for RNA from Xia *et al.* [39]. Such studies could even be of clinical importance as they may be helpful in understanding the intricate mechanism of microRNA target hybridisation [40–42].

In conclusion, we have provided an independent confirmation that the finite stacking enthalpies for the Joyeux-Buyukdagli Hamiltonian are consistent with the experimental enthalpies from nearest-neighbour regression models. This provides a much needed validation of the use of experimental enthalpies in the Peyrard-Bishop model and opens up the opportunity to use this model for more complicated situations such as nucleotide mismatches or RNA.

* * *

We acknowledge financial support by Fapemig, CNPq and National Institute of Science and Technology for Complex Systems.

REFERENCES

[1] Gray D. M. and I. Tinoco J., *Biopolymers*, **9** (1970) 223.
[2] Breslauer K. J., Frank R., Blocker H. and Marky L. A., *Proc. Natl. Acad. Sci. U.S.A.*, **83** (1986) 3746.
[3] SantaLucia J. jr., *Proc. Natl. Acad. Sci. U.S.A.*, **95** (1998) 1460, http://www.pnas.org/cgi/content/abstract/95/4/1460.
[4] Licinio P. and Guerra J. C. O., *Biophys. J.*, **92** (2007) 2000.
[5] Poland D. and Scheraga H. A., *J. Chem. Phys.*, **45** (1966) 1464.
[6] Mitchell J., Laughton C. and Harris S., *Nucl. Acids Res.*, **39** (2011) 3928.
[7] Peyrard M. and Bishop A. R., *Phys. Rev. Lett.*, **62** (1989) 2755.
[8] Weber G., Haslam N., Whiteford N., Prügel-Bennett A., Essex J. W. and Neylon C., *Nat. Phys.*, **2** (2006) 55.
[9] Velizhanin K. A., Chien C. C., Dubi Y. and Zwolak M., *Phys. Rev. E*, **83** (2011) 050906.
[10] Theodorakopoulos N., *Phys. Rev. E*, **82** (2010) 21905.
[11] Alexandrov B., Voulgarakis N. K., Rasmussen K. Ø., Usheva A. and Bishop A. R., *J. Phys.: Condens. Matter*, **21** (2009) 034107.
[12] Tabi C. B., Mohamadou A. and Kofané T. C., *J. Phys.: Condens. Matter*, **20** (2008) 415104.
[13] Romero-Enrique J. M., de los Santos F. and Munoz M. A., *EPL*, **89** (2010) 40011.
[14] Weber G., *Europhys. Lett.*, **73** (2006) 806.
[15] Tapia-Rojo R., Mazo J. J. and Falo F., *Phys. Rev. E*, **82** (2010) 031916.
[16] Ribeiro N. F. and Drigo Filho E., *Braz. J. Phys.*, **41** (2011) 195.
[17] Machado R. F. and Weber G., *EPL*, **87** (2009) 38005.
[18] Das T. and Chakraborty S., *EPL*, **83** (2008) 48003.
[19] Srivastava S. and Singh Y., *EPL*, **85** (2009) 38001.
[20] Campa A. and Giansanti A., *Phys. Rev. E*, **58** (1998) 3585.
[21] Alexandrov B., Gelev V., Monisova Y., Alexandrov L., Bishop A., Rasmussen K. and Usheva A., *Nucl. Acids Res.*, **37** (2009) 2405.
[22] Weber G., Essex J. W. and Neylon C., *Nat. Phys.*, **5** (2009) 769.
[23] Joyeux M. and Buyukdagli S., *Phys. Rev. E*, **72** (2005) 051902.
[24] Buyukdagli S. and Joyeux M., *Phys. Rev. E*, **73** (2006) 51910.
[25] Buyukdagli S. and Joyeux M., *Phys. Rev. E*, **76** (2007) 021917.
[26] Buyukdagli S. and Joyeux M., *Phys. Rev. E*, **77** (2008) 031903.
[27] Buyukdagli S., Sanrey M. and Joyeux M., *Chem. Phys. Lett.*, **419** (2006) 434.
[28] Dauxois T., Peyrard M. and Bishop A. R., *Phys. Rev. E*, **47** (1993) R44.
[29] Owczarzy R., You Y., Moreira B. G., Manthey J. A., Huang L., Behlke M. A. and Walder J. A., *Biochemistry*, **43** (2004) 3537.
[30] Joyeux M. and Florescu A.-M., *J. Phys.: Condens. Matter*, **21** (2009) 034101.
[31] Weber G., Haslam N., Essex J. W. and Neylon C., *J. Phys.: Condens. Matter*, **21** (2009) 034106.
[32] Press W. H., Teukolsky S. A., Vetterling W. T. and Flannery B. P., *Numerical Recipes in C* (Cambridge University Press, Cambridge) 1988.
[33] Huguet J. M., Bizarro C. V., Forns N., Smith S. B., Bustamante C. and Ritort F., *Proc. Natl. Acad. Sci. U.S.A.*, **107** (2010) 15431.
[34] Blake R. and Delcourt S., *Nucl. Acids Res.*, **26** (1998) 3323, http://nar.oupjournals.org/cgi/content/abstract/26/14/3323.
[35] Blake R. D., Bizzaro J. W., Blake J. D., Day G. R., Delcourt S. G., Knowles J., Marx K. A. and SantaLucia J. jr., *Bioinformatics*, **15** (1999) 370.
[36] Cuesta-Lopez S., Angelov D. and Peyrard M., *EPL*, **87** (2009) 48009.
[37] Reisner W., Larsen N., Silahtaroglu A., Kristensen A., Tommerup N., Tegenfeldt J. and Flyvbjerg H., *Proc. Natl. Acad. Sci. U.S.A.*, **107** (2010) 13294.
[38] Gonzalez R., Zeng Y., Ivanov V. and Zocchi G., *J. Phys.: Condens. Matter*, **21** (2009) 034102.
[39] Xia T., SantaLucia J. jr., Burkard M. E., Kierzek R., Schroeder S. J., Jiao X., Cox C. and Turner D. H., *Biochemistry*, **37** (1998) 14719.
[40] Long D., Lee R., Williams P., Chan C., Ambros V. and Ding Y., *Nat. Struct. Mol. Biol.*, **14** (2007) 287.
[41] Ui-Tei K., Naito Y., Nishi K., Juni A. and Saigo K., *Nucl. Acids Res.*, **36** (2008) 7100.
[42] Schmidt T., Mewes H. and Stümpflen V., *PLoS ONE*, **4** (2009) e6473.