BIOINFORMATICS APPLICATIONS NOTE Vol. 29 no. 10 2013, pages 1345–1347
doi:10.1093/bioinformatics/btt133

Structural bioinformatics

Advance Access publication March 16, 2013

TfReg: calculating DNA and RNA melting temperatures and opening profiles with mesoscopic models

Gerald Weber

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

Associate Editor: Anna Tramontana

ABSTRACT

Summary: The mesoscopic statistical physics models, known generically as Peyrard–Bishop (PB) models, have found many applications for the study of oligonucleotide properties. Unfortunately, PB models have not reached a wider non-specialized audience for the lack of freely available software implementations. Here we present an extensible C++ implementation of four variants of the PB model, which allows the user to calculate melting temperatures from tested model parameters. Even for a non-specialist, it should be straightforward to change these parameters to reflect different experimental environments or different types of oligonucleotides. For users with some proficiency in C++ programming, it should be feasible to extend the code to other PB models owing to the generic programming implementation adopted for TfReg. Pre-calculated parameters are included that allow the immediate calculation of melting temperatures and thermal equivalence indexes for DNA and RNA.

Availability: C++ source code and compiled binaries for several Linux distributions are available from https://sites.google.com/site/geraldweberufmg/tfreg and from OpenSuse build service at http://build.opensuse.org.

Contact: gweberbh@gmail.com

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on November 19, 2012; revised on February 6, 2013; accepted on March 9, 2013

1 INTRODUCTION

The Peyrard–Bishop (PB) model was introduced in 1989 (Peyrard and Bishop, 1989); it is a mesoscopic statistical method with two terms describing separately the hydrogen bond between the base pairs and the stacking interaction of the DNA molecule. These two terms describe the DNA physics well enough to make it a popular choice for the study of numerous physical properties (Peyrard, 2006). Additionally, the method is simple enough to make it computationally feasible to calculate a large number of different sequences in short time, a common requirement for many bioinformatics applications. The model has been successful in numerous applications and is one of the leading approaches for modelling oligonucleotides at the mesoscale; for a discussion on this see the review by Cuesta-Lopez *et al*. (2005). Being a computationally efficient method, it is unfortunate that as yet no software implementation was made available to a wider bioinformatics audience.

Here we are trying to bridge the gap by providing an implementation of the PB model that is accessible to a non-specialist, but which we hope may also be valuable to PB model specialists. The original code on which this application note is based was used to calculate the results of several of our publications on DNA (Weber, 2011; Weber *et al*., 2006, 2009a, b) and more recently on RNA (Weber, 2013). Therefore, this application note has the added purpose of providing specialists in this type of model with the opportunity to verify our published results.

Currently, software applications for melting temperature analysis fall into two broad categories: free energy models such as the one pioneered by Breslauer *et al*. (1986) and those based on Poland–Scheraga models such as uMELT by Dwight *et al*. (2011) and by Tøstesen (2008). Both approaches are nearest-neighbour (NN) methods, that is, they model the interaction between consecutive base pairs. PB models on the other hand also consider the hydrogen bonds of individual base pairs through a Morse potential (Peyrard and Bishop, 1989). Our parametrization studies revealed that these Morse potentials are remarkably consistent even when applied to completely different melting temperature datasets. For instance, the hydrogen Morse potential of CG base pairs was found to be nearly identical for both DNA and RNA (Weber, 2013). Therefore, the PB models offer the unique opportunity to study what would happen to melting temperatures and opening profiles under different strengths of hydrogen bonds.

2 METHODS

TfReg implements the transfer matrix method for calculating the partition function of the PB Hamiltonian (Weber *et al*., 2009b) for heterogeneous sequences following the method developed by Zhang *et al*. (1997). The user selects the base pair on which the expansion should take place and the application calculates all relevant matrices. The transfer matrix can be calculated for both periodic and open boundary conditions. For the calculation of melting temperatures, it implements the melting index calculation and the linear regression method described in Weber *et al*. (2006, 2009b).

Model variants: Currently, TfReg provides the option of four different types of Hamiltonians, the original PB model (Peyrard and Bishop, 1989), the anharmonic model (Dauxois *et al*., 1993), the solvent potential (Weber, 2006) and the finite enthalpy model (Joyeux and Buyukdagli, 2005). The software implementation makes extensive use of C++ generic programming, which simplifies the extension of the software to other 1D Hamiltonians. The programme can be run on a standard Linux command line and has numerous options to control the calculation, which are documented in the user manual together with some practical examples. Example shell scripts are included that reproduce some of our published results.

Melting temperatures of DNA and RNA: Using the regression method described in detail in Weber *et al*. (2009b), TfReg is capable of providing

melting temperatures of any DNA or RNA sequence. For DNA, these can be presently calculated at the salt concentrations given in Owczarzy *et al*. (2004), that is, 69, 119, 220, 621 and 1020 mM. For RNA, the application currently takes into account canonical Watson–Crick base pairs (CG and AU) and can be calculated at 1000 mM (Weber, 2013). The experimental melting temperatures used for the regression are from Xia *et al*. (1998).

Accuracy of the melting temperature calculations For DNA melting temperatures, the accuracy is around 0.8°C, which is better than for NN-model calculations, see Supplementary Table S1. For a full discussion of the accuracy when calibrated to the experimental results by Owczarzy *et al*. (2004), see Weber *et al*. (2009a). For RNA, the accuracy is also somewhat better than for NN-model calculations, especially for non-two-state sequences, as presented in Weber (2013). Further discussions on the expected prediction accuracy and its relation to the experimental data can be found in the Supplementary Section S1.

Average opening $\langle y \rangle$*:* The application is capable of calculating the average helix opening $\langle y \rangle$ (Peyrard and Bishop, 1989), that is, the average distance between base pairs, which can be used to gain a better understanding of localized openings. This capability goes beyond traditional nearest-neighbour models (Breslauer *et al*., 1986), which are limited to providing a single Gibbs free energy value for a DNA or an RNA sequence. Methods based on the Poland–Scheraga model (Dwight *et al*., 2011; Tøstesen, 2008) provide an opening probability profile that can be interpreted in a similar way to average distances. However, average distances allow for a more quantitative analysis, such as, for example, the conclusion that RNA has larger average tranversal fluctuations than DNA, despite having stronger AU hydrogen bonds (Weber, 2013).

Model matrices: The application provides all calculated model matrices [Eqs. (56,60,64) of Zhang *et al*. (1997)] for a given set of parameters, as well as the eigenvalues and eigenvectors of the transfer matrix. This may be useful for testing other programmes that implement the PB model or as input for extensions or simulations using this model.

Processing times: For the calculation of average opening profiles such as shown in Figure 1, the typical processing time was <1 s per temperature step on an i7 2.7 GHz processor. If the parameters and temperatures are kept constant, the calculation time can be considerably shortened by storing all matrices on file, as described in the previous paragraph. The processing time is linear with the size of the sequence, as shown in Supplementary Figure S3. For the calculation of melting temperatures, typical processing times are 10^{-3}s per sequence, as benchmarked by the calculation of 300 000 sequences from Panjkovich and Melo (2005).

Sequences used: For the calculation of the opening profile shown in Figure 1 and Supplementary Figure S1, we used

GTGCCCATTTAGGGTATATATGGCCGAGTGAGCGAG
CAGGATCTCCATTTTGACCGCAAATTTGAACG

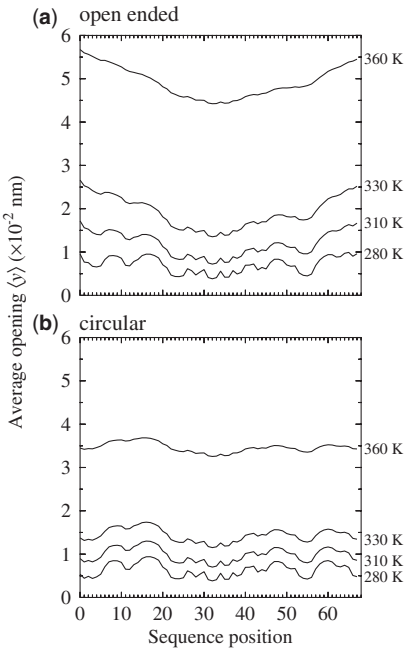

Fig. 1. Average opening $\langle y \rangle$ of an (**a**) open-ended and (**b**) circular DNA sequence as a function of sequence position, for several temperatures. Calculations were performed for optimized parameters for the harmonic PB model (Weber *et al*., 2009a) at a salt concentration of 69 mM

3 DISCUSSION

In essence, the main parameters of the PB model are the Morse potential D, which represents the hydrogen bond, and the elastic constant k, which represents the stacking interaction. The potential D depends only on the base pair and, as a rule of thumb, is roughly proportional to the strength of the hydrogen bonds. If the user wishes, for instance, to experiment with a stronger hydrogen bond, all that is needed is to set a larger value for D. In this respect, the PB model offers a much more intuitive framework for setting its parameters than the nearest neighbour models based on Gibbs free energy (Breslauer *et al*., 1986). For instance, in Figure 1a, we show an example for a DNA sequence where we calculate all opening profiles for a range of temperatures. These opening profiles were calculated for open boundary conditions and therefore clearly show the end fraying of the DNA sequence. In Figure 1b, we show the same sequence of Figure 1a but now calculated with periodic boundary conditions, which would be equivalent to a circular DNA. The characteristic end-fraying is now absent and generally this structure appears much more stable at higher temperatures than for open-ended DNA. Opening profiles calculated for a range of temperatures can be combined into opening maps, such as exemplified in Supplementary Figure S1.

In terms of the accuracy of predicting melting temperatures, TfReg is at least on par with nearest-neighbour free energy models. Supplementary Table S1 shows the melting temperature average error when compared for the DNA dataset by Owczarzy *et al*. (2004) compared with the methods implemented by DNAmate by Panjkovich and Melo (2005) and the phenomenological model by Khandelwal and Bhyravabhotla (2010). One should be aware, however, that the accuracy of predictions is largely determined by the experimental data used to calibrate a given method. In Supplementary Section S1, we discuss these issues in more detail. One aspect of interest is that TfReg has no fundamental restriction on sequence length, as shown in Supplementary Figure S2, where we present the average accuracy as a function of sequence length.

As the accuracy of any prediction method depends so strongly on the dataset used to generate the regression coefficients, it is worth mentioning that TfReg can generate new regression sets from user-provided melting temperatures. This means that users may use TfReg to predict new melting temperatures based on the calibration of their own data, instead of relying on pre-established regression sets.

Some bioinformatics applications of interest for TfReg would be the prospective studies of melting for situations where little melting experimental data are available for certain types of oligonucleotides, given that the parameters are intuitive. The opening profiles could also be used in addition to flexibility for studies of promoter sequence recognition (Fagundes-Lima and Weber, 2012).

4 CONCLUSION

We provide a complete C++ implementation of the PB model in the framework of the transfer matrix integral method. In addition to the calculation of the matrices that are essential to the method, the application TfReg also calculates melting temperatures for DNA and RNA, providing thus an alternative method to the nearest-neighbour Gibbs free energy calculations (Breslauer *et al.*, 1986). The application also provides the option of calculating the average helix opening profile. A comprehensive user manual with full explanation of the main aspects of the PB model is provided in the software package and also included as Supplementary Material.

Funding: This work was supported by Fundação de Amparo a Pesquisa do Estado de Minas Gerais (Fapemig); Conselho Nacional de Desenvolvimento Cientfico e Tecnolgico (CNPq); Universidade Federal de Minas Gerais; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior and National Institute of Science and Technology for Complex Systems.

Conflict of Interest: none declared.

REFERENCES

Breslauer,K.J. *et al.* (1986) Predicting DNA duplex stability from the base sequence. *Proc. Natl Acad. Sci. USA*, **83**, 3746–3750.

Cuesta-Lopez,S. *et al.* (2005) Can we model DNA at the mesoscale? *J. Biol. Phys.*, **31**, 273–301.

Dauxois,T. *et al.* (1993) Entropy-driven DNA denaturation. *Phys. Rev. E Stat. Nonlin. Soft. Matter Phys.*, **47**, R44–R47.

Dwight,Z. *et al.* (2011) uMELT: prediction of high-resolution melting curves and dynamic melting profiles of pcr products in a rich web application. *Bioinformatics*, **27**, 1019–1020.

Fagundes-Lima,D. and Weber,G. (2012) Comment on SCS: signal, context, and structure features for genome-wide human promoter recognition. *Trans. Comp. Biol. Bioinform.*, **9**, 940–941.

Joyeux,M. and Buyukdagli,S. (2005) Dynamical model based on finite stacking enthalpies for homogeneous and inhomogeneous DNA thermal denaturation. *Phys. Rev. E Stat. Nonlin. Soft. Matter Phys.*, **72**, 051902.

Khandelwal,G. and Bhyravabhotla,J. (2010) A phenomenological model for predicting melting temperatures of DNA sequences. *PLoS One*, **5**, e12433.

Owczarzy,R. *et al.* (2004) Effects of sodium ions on DNA duplex oligomers: improved predictions of melting temperatures. *Biochemistry*, **43**, 3537–3554.

Panjkovich,A. and Melo,F. (2005) Comparison of different melting temperature calculation methods for short DNA sequences. *Bioinformatics*, **21**, 711–722.

Peyrard,M. (2006) Melting the double helix. *Nat. Phys.*, **2**, 13–14.

Peyrard,M. and Bishop,A.R. (1989) Statistical mechanics of a nonlinear model for DNA denaturation. *Phys. Rev. Lett.*, **62**, 2755–2757.

Tøstesen,E. (2008) A stitch in time: efficient computation of genomic DNA melting bubbles. *Algorithms Mol. Biol.*, **3**.

Weber,G. (2006) Sharp DNA denaturation due to solvent interaction. *Europhys. Lett.*, **73**, 806–811.

Weber,G. (2011) Finite enthalpy model parameters from DNA melting temperatures. *Europhys. Lett.*, **96**, 68001.

Weber,G. (2013) Mesoscopic model parametrization of hydrogen bonds and stacking interactions of RNA from melting temperatures. *Nucleic Acids Res.*, **40**, e30.

Weber,G. *et al.* (2006) Thermal equivalence of DNA duplexes without melting temperature calculation. *Nat. Phys.*, **2**, 55–59.

Weber,G. *et al.* (2009a) Probing the microscopic flexibility of DNA from melting temperatures. *Nat. Phys.*, **5**, 769–773.

Weber,G. *et al.* (2009b) Thermal equivalence of DNA duplexes for probe design. *J. Phys. Condens. Matter*, **21**, 034106.

Xia,T. *et al.* (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry*, **37**, 14719–14735.

Zhang,Y.L. *et al.* (1997) Theory of DNA melting based on the Peyrard-Bishop model. *Phys. Rev. E Stat. Nonlin. Soft. Matter Phys.*, **56**, 7100–7115.