

Statistical mechanical simulation of polymeric DNA melting with MELTSIM

R. D. Blake^{1,*}, J. W. Bizzaro², J. D. Blake³, G. R. Day¹,
S. G. Delcourt¹, J. Knowles¹, K. A. Marx² and J. SantaLucia, Jr⁴

¹Department of Biochemistry, Microbiology and Molecular Biology, University of Maine, Orono, ME 04469-5735, USA, ²Department of Chemistry, University of Massachusetts, Lowell, MA 01854, USA, ³Department of Cellular and Molecular Pharmacology, Box 0450, University of California, 513 Parnassus Avenue, San Francisco, CA 94143-0450, USA and ⁴Department of Chemistry, Wayne State University, Detroit, MI 48202, USA

Received on October 19, 1998; revised on December 6, 1998; accepted on February 9, 1999

Abstract

Motivation: MELTSIM is a windows-based statistical mechanical program for simulating melting curves of DNAs of known sequence and genomic dimensions under different conditions of ionic strength with great accuracy. The program is useful for mapping variations of base compositions of sequences, conducting studies of denaturation, establishing appropriate conditions for hybridization and renaturation, determinations of sequence complexity, and sequence divergence.

Results: Good agreement is achieved between experimental and calculated melting curves of plasmid, bacterial, yeast and human DNAs. Denaturation maps that accompany the calculated curves indicate non-coding regions have a significantly lower (G+C) composition than coding regions in all species examined. Curves of partially sequenced human DNA suggest the current database may be heavily biased with coding regions, and excluding large (A+T)-rich elements.

Availability: MELTSIM 1.0 is available at: <http://www.uml.edu/Dept/Chem/UMLBIC/Apps/MELTSIM/MELTSIM-1.0-Win/meltsim.zip>. Melting curve plots in this paper were made with GNU PLOT 3.5, available at: http://www.cs.dartmouth.edu/gnuplot_info.html

Contact: blake@maine.maine.edu; jbizzaro@alexandria.uml.edu

Introduction

The familiar helical conformation of DNA is the consequence of Watson–Crick hydrogen bonds on opposing strands, and of more substantial interactions between closely stacked base pairs. Strengths of these interactions are well

below those of covalent bonds, so the helix can be denatured by relatively mild alterations of conditions supporting the native structure. Since covalent bonds are unaffected, denaturation is reversible; convenient for the investigation of different aspects of equilibrium thermodynamics. Although thermally induced denaturation differs in mechanism from the natural process of replication, studies of equilibrium melting provide valuable insight into the thermodynamics of this process and DNA.

The most sensitive and convenient means of following equilibrium denaturation is by UV electronic spectroscopy, where the perturbation of stacked base pairs during the helix–coil transition is accompanied by large changes in absorption. Figure 1 shows a melting curve of the linear form of the 4662 bp plasmid DNA pN/MCS-13, obtained by the difference-approximation method at 270 nm, where the differential extinction coefficients for the dissociation of A·T and G·C pairs, $\Delta\epsilon_{A\cdot T}$ and $\Delta\epsilon_{G\cdot C}$, are equal (Blake and Hydorn, 1985). Such curves can be compared directly with calculated curves, $d\theta_h/dT$, where θ_h is the calculated fraction of base pairs remaining in the helical state.

Denaturation occurs in steps, as regions of sequence partition into discrete thermodynamic ‘domains’ that melt cooperatively as intermediate subtransitions. The stability of domains depends on their sequence, but as seen in Figure 2, on their position as well: internal loops (Case I), ends (Case II), enlargements of pre-existing coils (Case III), expansion of loops through to the ends of the helix (Case IV), intervening domains, resulting in the coalescence of neighboring loops (Case V), and finally the separation of the strands. In modeling these transitions, it is assumed that melting is everywhere at equilibrium, and fully reversible. Except under extraordinarily high DNA concentrations, conditions that undermine the unique sensitivity of the spectroscopic approach, the separation of strands is always a non-equilibrium process and is therefore not considered in the model.

*To whom correspondence should be addressed

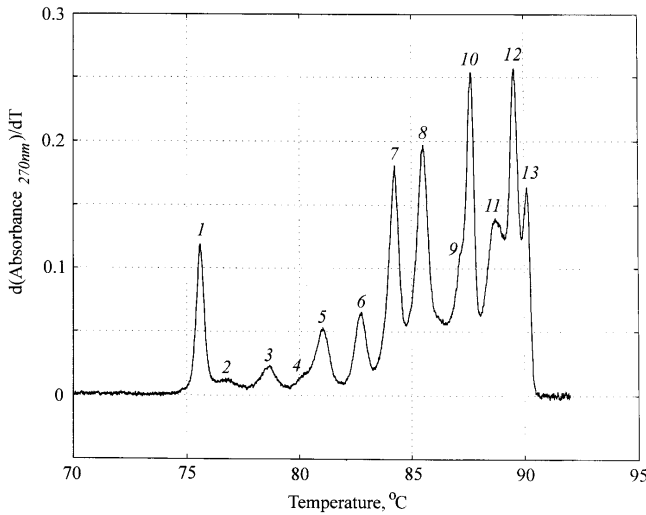


Fig. 1. Experimental derivative melting curve of 20 µg of pN/MCS-13 DNA in 0.0745 M-Na⁺, followed at 270 nm. The 4662 bp plasmid DNA was linearized by *EcoRV* at position 185 in preparation for melting. Peaks were numbered only for identification purposes, and except for nos 1–5, do not denote individual subtransitions. Peak no. 1 represents the subtransition for the 245 bp repetitive insert: [AAGTTGAACAAAAA]₁₇AAGTTGA, located 805 bp from the *EcoRV* locus, so that the domain melts as a closed loop.

Denaturation is a highly cooperative process, where the conformational states of base pairs exhibit a strong dependence on the states of their neighbors (Poland and Scheraga, 1970). Accounting for the stability and cooperativity of DNA, and the various intermediate subtransitions seen in Figure 1, and modeled in Figure 2, has approached quantitative levels, so that it has become possible to calculate accurate melting curves just from knowledge of DNA sequences; which gets us to the purpose of this report. With current para-

eters in a statistical mechanical program, we show that it is possible to calculate melting curves and denaturation maps of DNA sequences of even genomic dimensions with great accuracy.

Thermodynamic considerations

As seen in Figure 2, the foremost thermodynamic effect on melting temperature, T_M , is the statistical weight for the dissociation of base pairs. This quantity, denoted by s for helix–coil transitions at equilibrium, includes both enthalpic and entropic factors,

$$s = \exp(-\Delta G/RT) = \exp(-(\Delta H - T\Delta S)/RT) \quad (1)$$

The enthalpic loss arises predominately from unstacking pair i from j at the ends of helical domains (Case II). Recognition of nearest-neighbor effects is taken into account with s_{ij} ; where ΔH_{ij} and ΔS_{ij} represent values for the ten unique stacked pairs; given in Table 1, for the standard buffer conditions. Given thermodynamic additivity, the equilibrium constant for a domain of length N is represented by the product of individual stability constants, s_{ij}^N , and its variation with temperature by the van't Hoff expression

$$\frac{\partial \ln(s_{ij}^N)}{\partial T} = \frac{N \cdot \Delta \bar{H}}{RT^2} \quad (2)$$

where

$$\Delta \bar{H} = \sum_1^N \Delta H_{ij}/N \quad (3)$$

Also, the average statistical weight for a DNA of length L , melting in k domains,

$$\langle s \rangle = \prod_1^k s_{ij}^N / L \quad (4)$$

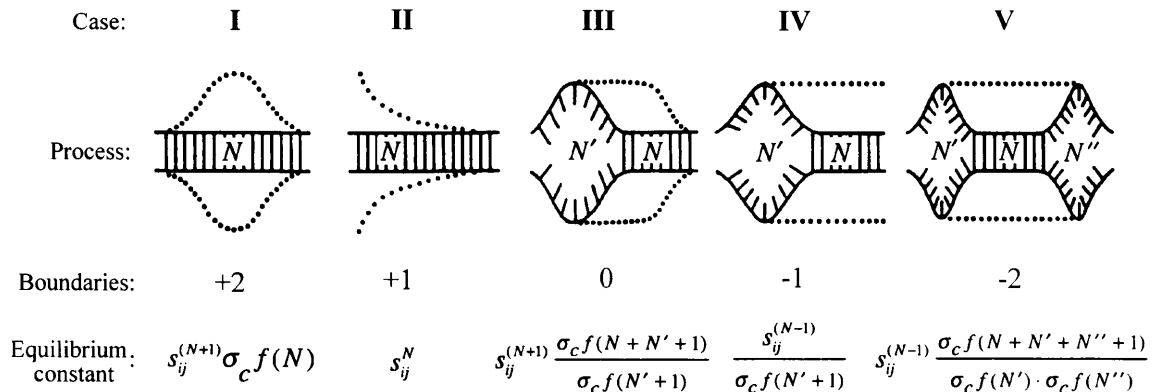


Fig. 2. Schematic representation of five of the six dissociation processes of polymeric DNAs, with changes in numbers of boundaries and equilibrium constants.

Table 1. Nearest stacked and paired neighbor energies

s'	i	$3'$	$\frac{dT_{ij}}{d\log[Na^+]}$	$T_{ij}^{1.0M-Na^+}$		$T_{ij}^{0.0745M-Na^+}$		ΔH_{ij}^a	ΔS_{ij}^b
				$^{\circ}C$	$^{\circ}K$	$^{\circ}C$	$^{\circ}K$		
1	A	T	21.00	81.85	355.01	58.23	8.00	22.53	
2	T	A	20.11	86.72	359.88	64.10	8.31	24.64	
3	A	T	19.78	89.08	362.24	66.77	8.45	24.86	
4	G	C	17.76	99.49	372.65	79.51	9.13	24.50	
5	A	T	17.10	103.18	376.34	83.94	9.36	24.87	
6	G	C	16.87	104.43	377.59	85.45	9.44	25.00	
7	C	G	16.21	107.96	381.12	89.72	9.67	25.37	
8	G	C	14.18	118.49	391.65	102.50	10.34	27.52	
9	C	G	13.20	124.54	397.70	109.69	10.72	26.95	
10	C	G	13.20	124.61	397.77	109.76	10.72	26.95	

a) kcal·mol⁻¹.
b) cal·mol⁻¹·deg⁻¹.

A second, smaller, and predominately entropic effect on melting temperature arises from the formation of loops (Poland and Scheraga, 1970), requiring sufficient thermal energy to interrupt helices sustained by cooperative interactions; to release internal paired and stacked residues from interacting with their neighbor-pairs. Consequently, the stability constant s_{ij}^N , is weighted by a cooperativity parameter, σ_c , and by a loop function for the greater probability of loop closure:

$$f(N) = (N + D)^{-\alpha} \quad (5)$$

where N is the number of residues in the loop domain, D an empirical stiffness parameter (Fixman and Freiere, 1977), and α an adjustable loop-closure exponent that allows for spatial constraints on the chains. A value of $D = 1$ is assumed, while the value for α derived from Monte Carlo and direct enumeration studies of the probability of ring closure for random walks with excluded volume, is taken to be 1.7 ± 0.2 (Fisher, 1966; Poland and Scheraga, 1970). The overall weighting function for Case I melting of interior loops is therefore given by,

$$S_{ij}^{N+1} \cdot \sigma_c \cdot (N + 1)^{-1.7} \quad (6)$$

while those for other processes involving loops, represented by cases III–V, are given in Figure 2.

The common, most dominant weighting function in all cases is the stability factor. For the standard $[Na^+]$ conditions used in obtaining the experimental curve in Figure 1 (0.0745 M- Na^+), values of s_{ij} are determined from Eqn (1) and ΔH_{ij} and ΔS_{ij} from columns seven and eight in Table 1. The nearest-neighbor parameters in Table 1 were evaluated from high-resolution melting experiments of 42 plasmids containing well-behaved synthetic tandemly repeated inserts (Blake and Delcourt, 1998). Domain T_M were arranged into a set of linear algebraic expressions:

$$T_M = \sum f_{ij} \cdot T_{ij} \quad (7)$$

where f_{ij} represents the fractional neighbor pair frequencies in the domain. T_{ij} were determined by Crout's method for LU matrix decomposition.

However, it is not possible to uniquely determine all 10 nearest neighbor parameters from polymer studies (Goldstein and Benight, 1992; Vologodskii *et al.*, 1984); instead, only eight invariants can be determined. Thus, the method of singular value decomposition was applied to provide the solution with the minimum sequence dependence of the 10 nearest-neighbors consistent with the eight polymer invariants (SantaLucia, 1998). This procedure has been shown to provide polymer parameters that are in good agreement with rank-determinant oligonucleotide parameters (SantaLucia, 1998; Allawi and SantaLucia, 1997). The nearest-neighbor enthalpy changes, ΔH_{ij} , were determined by fitting shapes of experimental curves to Eqn (2).

Values of ΔS_{ij} for other $[Na^+]$ can be derived from columns 3 and 5, by assuming the salt effect is completely entropic, and,

$$\Delta G_{ij} = \Delta H_{ij} - T_M \cdot \Delta S_{ij} = 0 \quad (8)$$

The salt dependence presented in Table 1 shows a systematic dependence on $\%(G + C)$, consistent with previous observations (Frank-Kamenetskii, 1971).

Literature values for the cooperativity parameter, σ_c , have generally varied between 1×10^{-5} and 5×10^{-5} (Oliver *et al.*, 1977; Amirikyan *et al.*, 1981; Blake, 1987). A recent value determined from a large collection of well-behaved model systems, 1.26×10^{-5} , seems to be of universal applicability (Blake and Delcourt, 1998).

The model generally adopted in theoretical treatments of melting is the Ising one-dimensional lattice (Hill, 1959; Wartell and Montroll, 1972; Wartell and Benight, 1985), to which loop entropy has been appended. Exact algorithms for the behavior of this model have computer times proportional to N^2 (Wartell and Montroll, 1972; Poland, 1974). Poland's method used here, employs exact recursion relations for probabilities as transient variables for reverse recursion formulae. His forward formula then serves in the calculation of unconditional probabilities for the fraction of paired residues $\theta_h(T)$ and $d\theta_h(T)/dT$. We include the approximation of the

loop function with an exponential series as proposed by Fixman and Freiere (1977); which reduces computer time to $\sim 6 \cdot N$.

Experiment and theory

The pN/MCS13 DNA that produced the curve in Figure 1 was constructed (Blake and Delcourt, 1998) from pBR322, with a 245 bp repetitive sequence: [AAGTTGAACAAAAT]₁₇AAGTTGA, installed between two large, (G+C)-rich domains that serve as energetic barriers to further melting. The subtransition for the repetitive domain is no. 1, while those for the neighboring barrier domains are no. 10 and no. 11 in Figure 1.

pN/MCS13 was linearized at the unique *EcoRV* locus (position 185) for both experimental and calculated melting (Figure 3a). The *EcoRV* locus is 805bp away from the repetitive sequence domain, which therefore melts as a closed loop (case I in Figure 2). The level of agreement between theory and experiment can be seen in magnified form in Figure 3b, where both experimental and calculated subtransitions for the repetitive sequence were isolated from closely transitions by subtracting curves for the plasmid without the repetitive insert from corresponding curves with the insert. The experimental curve, with T_m^{obs} (no. 1) = 75.59°C, is represented by the continuous heavy line, and the statistical mechanical curve, with T_M^{calc} (no. 1) = 75.66°C, by the thin line, punctuated by cross-shaped (+) symbols. The agreement between observed and calculated T_M , amplitudes, breadths and areas is good; characteristic of all tests of the model.

The calculated denaturation map in Figure 3a indicates a concerted dissociation of the region between 824 and 1069, corresponding precisely with the location of the repetitive sequence; indicating this domain dissociates in an abrupt, essentially two-state fashion. This is substantiated by the calculated two-state van't Hoff curve (Blake and Delcourt, 1998), denoted by the diamond-shaped symbols in Figure 3b; with a T_M of 75.59°C, a mean enthalpy of 8.87 kcal/mol-bp, and a size of 245 bp. Other domains dissociating as internal loops at positions 1400, 2000, 3300 and 4300, do so by more gradual accretion of residues at helix boundaries with increasing temperature.

Experiment and theory agree just as well for DNAs one thousand times longer; where numbers of subtransitions are $\sim 50\times$ greater, resulting in essentially featureless, smooth melting curves. The example of the melting of the 4 638 858 bp *E. coli* DNA in the standard buffer (0.0745 M-Na⁺) is shown in Figure 4, where the experimental curve is represented by the continuous line and calculated by the diamond-shape points. There is similar good agreement between experimental and calculated curves of *Saccharomyces cerevisiae* DNA of 12 067 277 bp (Bizzaro *et al.*, 1998). The denaturation map from calculated curves of the 16 yeast

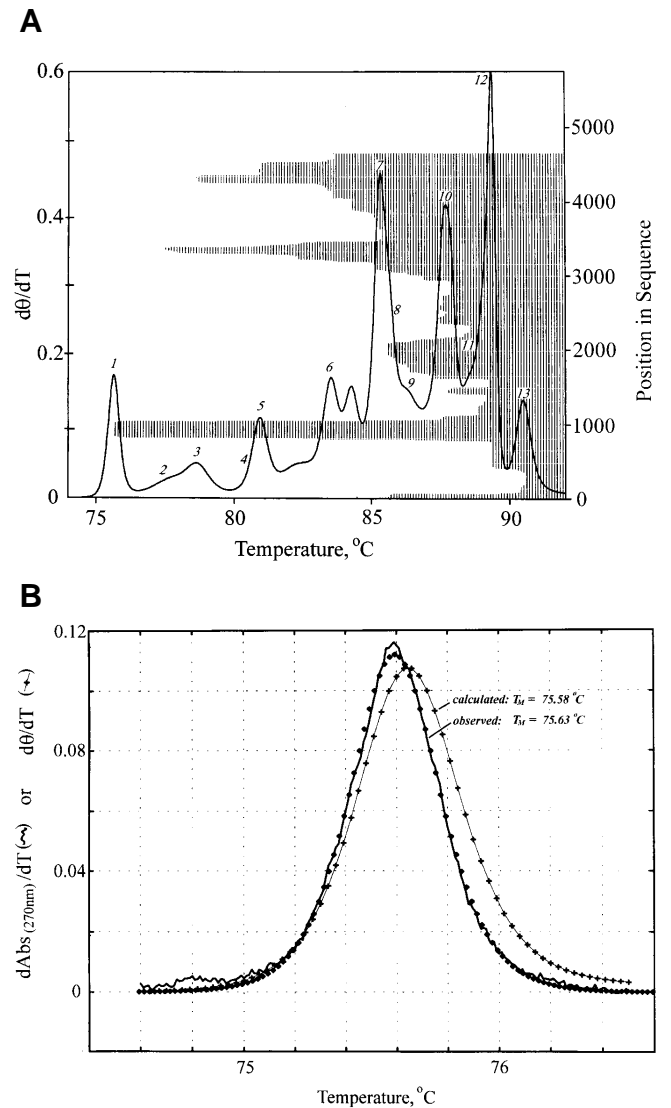


Fig. 3. (A) Calculated derivative melting curve of the pN/MCS-13 DNA sequence, under the same conditions used to obtain the experimental curve in Figure 1. For convenience in the correlation of subtransitions with features of the denaturation map, the third dimension: position in sequence, is here plotted on the right hand vertical axis with the scale for the two-dimensional melting curve on the left. (B) Enlargement of the melting curve of the 245 bp repetitive sequence in pN/MCS-13, labeled no. 1 in Figures 1 and 3A. The heavy, somewhat irregular trace represents the experimental curve. The trace punctuated by cross-shaped symbols (+) denotes the calculated curve. The curve denoted by filled diamonds represents a melting curve calculated on the assumption of two-state equilibrium melting of the repetitive sequence domain.

chromosomal DNAs indicate that flanking regions are significantly more (A+T)-rich than coding regions, as observed in other species (Suyama and Wada, 1983; Wada and Suyama,

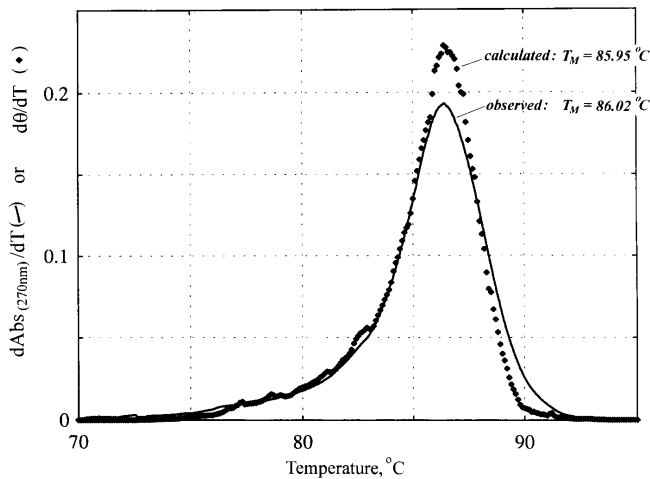


Fig. 4. Experimental (solid trace) and calculated (diamond points) melting curves of *E. coli* DNA.

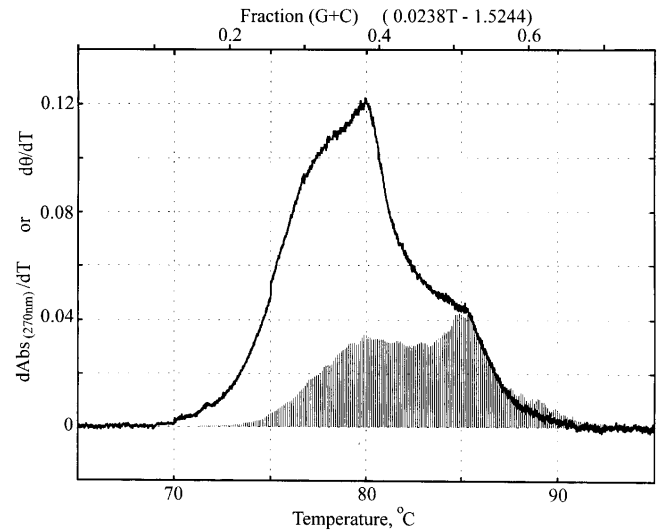


Fig. 6. Experimental (solid trace) and calculated (impulses) melting curves of human DNA. The former was obtained on total nuclear DNA, while the latter was calculated for only 5.7×10^6 bp from the GenBank database.

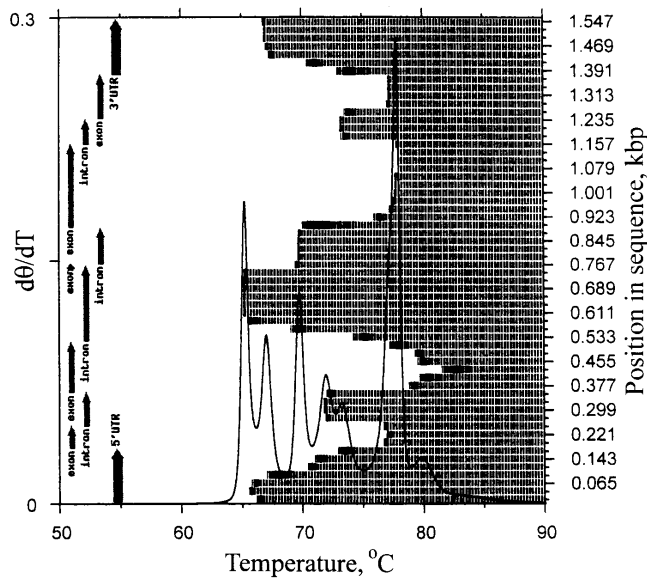


Fig. 5. Calculated melting curve of a 1600 bp segment of a *D. discoideum* multi exon-intron gene, together with corresponding denaturation and gene maps, showing the correlation of low temperature melting-low (G+C) content regions and high temperature melting-high (G+C) content regions with non-coding and coding elements, respectively.

ma, 1984; Blake and Earley, 1986). Differences in stability of coding and non-coding regions are particularly evident in Figure 5, showing the melting curve of a *Dictyostelium discoideum* multi exon-intron gene, together with denaturation and gene maps; for which a correlation of stabilities with functional elements was found (Marx *et al.*, 1998).

The story is different for the larger (6×10^9 bp) human DNA, where the sequence is only partially known. The experimental melting curve is broad, with a distinct peak at 80 °C and a shoulder at 85 °C, and differs from the calculated curve in an interesting way. The latter was produced from a large sequence of 5.7×10^6 bp by concatenation of three dozen large non-redundant human DNA sequences from the GenBank database; obtained from at least eight different chromosomes. The calculated curve, denoted by the shaded area in Figure 6, was reduced by 62% in area so that it would fit under the experimental curve. There is good correspondence of the high-temperature region from 85–90 °C, and of the two T_M of the bimodal calculated curve with the peak and shoulder of the experimental curve. However, assuming our partial collection of human sequences is representative of all the sequences in the database, it would appear that sequences thus far determined have excluded large (A+T)-rich elements that are present in the human genome. From the linear relationship of temperature and (G+C) content (Marmur and Doty, 1962; Blake, 1996), the peak at 80 °C corresponds to a (G+C) content of 38%, and from the denaturation map arises mainly from intergenic and flanking regions; whereas the peak (shoulder) at 85 °C corresponds to a (G+C) content of 50%, and arises almost exclusively from coding regions.

Computer programs

MELTSIM 1.0, compiled for the Windows OS (95, 98, or NT), requires 32 MB RAM, and is capable of handling sequences up to 2×10^6 base pairs at one time. The input file

consists of a title line, followed by lines of A, C, G and T sequences in upper or lowercase, with or without line breaks. An associated program, COMPILE, converts GenBank-formatted files into MELTSIM input files with a title line. Both two-dimensional and three-dimensional output files are produced, together with a log file of details of the run. Melting curves of a 4662 bp sequence took 1.5 min on a PC with 166 MHz Pentium chip, when the temperature increment was 0.03°C, and the range was 20°C. The *E. coli* sequence (4 638 858 bp) took 170 min, with a temperature increment of 0.1°C/25°C.

References

- Allawi, H.T. and SantaLucia, J., Jr (1997) Thermodynamics and NMR of internal G·T mismatches in DNA. *Biochemistry*, **36**, 10 581–10 594.
- Amirikyan, B.R., Vologodskii, A.V. and Lyubchenko, Y.L. (1981) Determination of DNA cooperativity factor. *Nucl. Acids Res.*, **9**, 5469–5482.
- Bizzaro, J.W., Marx, K.A. and Blake, R.D. (1998) Comparison of experimental with theoretical melting of the yeast genome and individual yeast chromosome denaturation mapping using the program MELTSIM. *Proceedings, Materials Research Society, Pittsburgh, PA*, **489**, 73–78.
- Blake, R.D. (1987) Cooperative lengths of DNA during melting. *Biopolymers*, **26**, 1063–1074.
- Blake, R.D. (1996) In Meyers, R.A. (ed.), *Encyclopedia of Molecular Biology and Molecular Medicine*, Vol. 2. VCH, New York, pp. 1–19.
- Blake, R.D. and Delcourt, S.G. (1998) Thermal stability of DNA. *Nucleic Acids Res.*, **26**, 3323–3332.
- Blake, R.D. and Earley, S. (1986) Distribution and evolution of sequence characteristics in the *E. coli* genome. *J. Biomol. Struct. Dyn.*, **4**, 291–307.
- Blake, R.D. and Hydorn, T.G. (1985) Spectral analysis for base composition of DNA undergoing melting. *J. Biochem. Biophys. Methods*, **11**, 307–316.
- Fisher, M.E. (1966) Effect of excluded volume on phase transitions in biopolymers. *J. Chem. Phys.*, **45**, 1464–1469.
- Fixman, M. and Freiere, J.J. (1977) Theory of DNA melting curves. *Biopolymers*, **16**, 2693–2704.
- Frank-Kamenetskii, M.D. (1971) Simplification of the empirical relationship between melting temperature of DNA, its GC content and concentration of sodium ions in solution. *Biopolymers*, **10**, 2623–2624.
- Goldstein, R.F. and Benight, A.S. (1992) How many numbers are required to specify sequence-dependent properties of polynucleotides? *Biopolymers*, **32**, 1679–1693.
- Hill, T.L. (1959) Generalization of the one-dimensional Ising model to helix transitions in nucleic acids and proteins. *J. Chem. Phys.*, **30**, 383–387.
- Marmur, J. and Doty, P. (1962) Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. *J. Mol. Biol.*, **5**, 109–118.
- Marx, K.A., Assil, I.Q., Bizzaro, J.W. and Blake, R.D. (1998) Comparison of experimental to MELTSIM calculated DNA melting of the (A+T) rich dictyostelium discoideum genome: denaturation maps distinguish exons from introns. *J. Biomol. Struct. Dyn.*, **16**, 329–340.
- Oliver, A.L., Wartell, R.M. and Ratliff, R.L. (1977) Helix coil transitions of d(A)_n-d(T)_n, d(A-T)_n-d(A-T)_n, and d(A-A-T)_n-d(A-T-T)_n; evaluation of parameters governing DNA stability. *Biopolymers*, **16**, 1115–1137.
- Poland, D. (1974) Recursion relation generation of probability profiles for specific-sequence macromolecules with long-range correlations. *Biopolymers*, **13**, 1859–1871.
- Poland, D. and Scheraga, H.A. (1970) *Theory of Helix-Coil Transitions in Biopolymers*. Academic Press, New York.
- SantaLucia, J., Jr (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl Acad. Sci. USA*, **95**, 1460–1465.
- Suyama, A. and Wada, A. (1983) Correlation between thermal stability maps and genetic maps of double-stranded DNAs. *J. Theoret. Biol.*, **105**, 133–145.
- Vologodskii, A.V., Amirikyan, B.R., Lyubchenko, Y.L. and Frank-Kamenetskii, M.D. (1984) Allowance for heterogeneous stacking in the DNA helix-coil transition theory. *J. Biomol. Struct. Dyn.*, **2**, 131–148.
- Wada, A. and Suyama, A. (1984) Stability distribution in the phage lambda-DNA double helix: a correlation between physical and genetic structure. *J. Biomol. Struct. Dyn.*, **2**, 573–591.
- Wartell, R.M. and Benight, A.S. (1985) Thermal denaturation of DNA molecules: a comparison of theory with experiment. *Physics Rep.*, **126**, 67–107.
- Wartell, R.M. and Montroll, E.W. (1970) Equilibrium denaturation of natural and of periodic synthetic DNA molecules. *Adv. Chem. Phys.*, **22**, 129–203.