

# polDNAmelt: Local Melting Within Polymeric DNA—An Improved Method and Its Applications

Sankar Basu<sup>1</sup> and Dhananjay Bhattacharyya<sup>2,\*</sup>

<sup>1</sup>Crystallography and Molecular Biology Division, Saha Institute of Nuclear Physics, 1/AF Bidhannagar, Kolkata 700064, India

<sup>2</sup>Computer Science Division, Saha Institute of Nuclear Physics, 1/AF Bidhannagar, Kolkata 700064, India

Local pseudo-unimolecular partial melting in polymeric cellular DNA is very much different from bimolecular comprehensive melting of short oligonucleotides in solutions. Several softwares estimate the melting temperature pretty accurately for the latter case and are used routinely for primer designing in PCR. However, the local partial melting in polymeric DNA involves the formation of a short-lived interior bubble constrained by double helical DNA at both the ends and thus (i) the melting-annealing transition becomes concentration independent and (ii) the fraying effect due to dangling ends are ruled out. Based on these features, the current study appropriately modifies the thermodynamic equations for calculation of local melting within polymeric DNA and incorporates them in a web-server (<http://www.saha.ac.in/biop/www/db/local/nsdnamelt.html>). Applications of the method in validating annotated promoter sequences and origin of replication have also been surveyed.

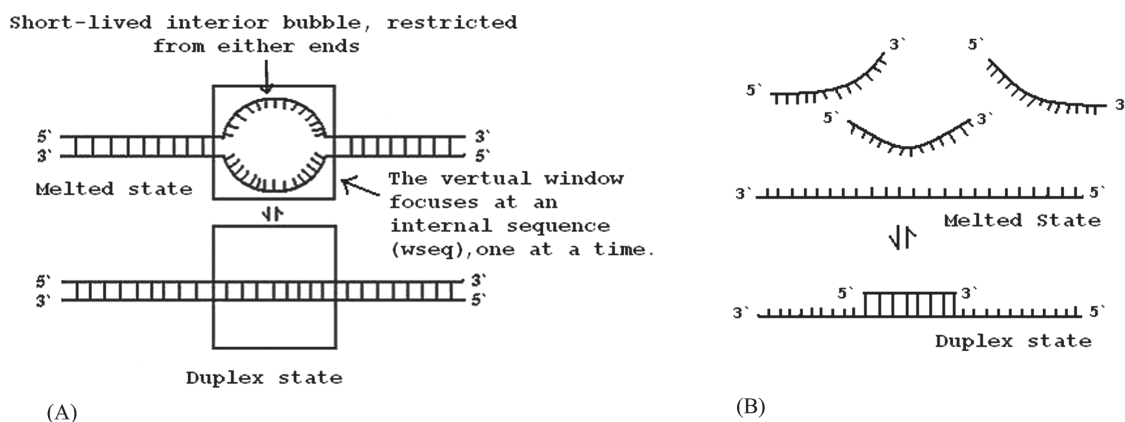
**Keywords:**  $T_m$ , Local Pseudo-Unimolecular Partial Melting, Short-Lived Interior Bubble, Concentration Independent Dna Melting, Fraying Effect.

We report the availability of a web-server, *polDNAmelt* (<http://www.saha.ac.in/biop/www/db/local/nsdnamelt.html>) to analyze local melting within polymeric DNA. Theoretical localization of melting nucleation sites distributed over chromosomal double stranded DNA (ds-DNA) is crucial to study since only the converted single stranded DNA is active for most of the biochemical processes (e.g., replication, transcription, recombination, etc.). Local melting within polymeric DNA with both the ends being held in double helical conformation is significantly different from melting of short oligonucleotides in solution. In either case, the main forces stabilizing a ds-DNA are base pairing and stacking interactions.<sup>1</sup> While there are two types of Watson-Crick base pairs (bp) present in ds-DNA with different hydrogen bonding strength, there are ten unique stacking arrangements between any two consecutive base pairs leading to ten different stacking energy increments. Thus, different base pairing and stacking interaction energies across diverge sequences (of say, identical length) require different amount of kinetic energies to transfer to unstructured melted out conformations. These evolutionarily conserved melting nucleation sites inhomogeneously distributed throughout the genome, map to local minima of stability with respect to their immediate neighboring

sequences. Although one can determine  $T_m$  considering only Watson-Crick base pairing energy (i.e., GC content) in a sequence, inclusion of heterogeneous stacking considerably improves agreements between theory and experiment for long polymeric DNA.<sup>2</sup> Experimental characterizations of duplex stability by (calorimetric and/or spectroscopic) DNA-melting studies have worked out several libraries of stacking-based melting thermodynamic parameters.<sup>2–11</sup> Quite a few softwares and servers<sup>12–16</sup> are available for prediction of melting temperature ( $T_m$ ) of short oligonucleotides using these libraries but their use to determine sequence dependent local melting can be erroneous. The present study attempts to modify appropriate thermodynamic equations for calculation of local melting within long polymeric DNA and incorporates them in *polDNAmelt*.

In a ds-DNA (devoid of mismatches), stacking occurs between every two consecutive base pairs (Nearest Neighbors: NN) once the Watson-Crick hydrogen bonding donor–acceptor pairs are stereo-specifically positioned. In the available libraries of stacking-based NN-step parameters ( $\Delta H^0$ ,  $\Delta S^0$ ,  $\Delta G^0$ ), the contribution of hydrogen bonding is implicitly included. *PolDNAmelt* utilizes the library derived by Santalucia<sup>3</sup> and adds these individual step parameters (NN propagation energies) for each consecutive NN-step along a polymeric DNA. This library is ‘a single

\*Author to whom correspondence should be addressed.



**Fig. 1.** Pictorial representation of the model: 'local pseudo-unimolecular melting in natural nucleic acid polymers.' (A) Concentration independent two-state local melting within polymeric DNA. (B) Concentration dependent template-primer melting-annealing transition during PCR and other similar molecular biology experiments.

set of unified parameters' applicable for both oligomeric and polymeric melting and had fairly good agreements with several previous experimental sets.<sup>2, 4, 5, 8, 10–12</sup>

In this study, internal sequences, focused by a virtual window in turn, are essentially non-fragmented patches of a contiguous stretch of ds-DNA (Fig. 1(A)). Precisely, all window sequences (including those at 5' and 3' ends) are internal (irrespective of its size) within a polymer i.e., the (whole) input sequence is treated as either circular or flanked by ds-DNA at either end. Having open ended termini at either end, linear single strands are produced from comprehensive melting of oligomeric duplexes (Fig. 1(B)) whereas local melting within a nucleic acid polymer yields a 'short-lived interior bubble' since the melting domain is constrained at both ends.<sup>17, 18</sup> It follows that, when two linear complementary single strands anneal to form an oligomeric duplex, they need to overcome an initiation penalty barrier due to unstacked dangling ends whereas in case of annealing of the locally melted interior bubble within a polymer, there is no such entropic barrier to overcome. Thus, initiation penalties were disregarded all throughout, in *polDNAmelt*. Again, the net free energy change ( $\Delta G_{37}^0$ ) of annealing for melted out self-complementary sequences into ds-duplex is corrected with a sequence independent entropic (C2) symmetry-penalty term ( $\Delta G_{37}^0(\text{sym}) = +0.43 \text{ kcal} \cdot \text{mol}^{-1}$ ).<sup>3</sup>

Tertiary interactions are a weaker second order effect in case of polymeric DNA and to an excellent approximation they could be neglected.<sup>19</sup> That is to say, a partially melted polymeric DNA should prefer a pseudo-unimolecular ds-duplex which is thermodynamically more stable than any alternative unimolecular secondary structures with unstacked and unpaired bases at or near each internal loop. Moreover, since the melting domain is clamped from either ends, any possible alternative secondary structure (e.g., hairpins, cruciforms etc.) will only be kinetically trapped at or near a local minima which will eventually dissolve while annealing upon thermodynamic equilibration. In this view, dynamic programming algorithms for secondary

structure prediction in DNA (UNAFold<sup>20</sup>, OMP<sup>21</sup> etc.) also neglect kinetically trapped (local minima) structures and assume that structures are populated according to an equilibrium Boltzmann distribution.<sup>19</sup> This essentially leads to a two-state approximation of the melting-annealing transition. In case of oligomers, there would always be competition among strands to anneal to the same complementary sequence (Fig. 1(B)) giving rise to a concentration dependent two-state  $T_m$ .<sup>22</sup>

$$T_m = \frac{\Delta H^0}{\Delta S^0 + R \cdot \ln(C_T/x)} \quad (1)$$

where  $C_T$  is the total strand concentration,  $\Delta H^0$  and  $\Delta S^0$  are the transition enthalpy and transition entropy respectively and  $x = 4$  for non-self complementary and  $x = 1$  for self complementary associations.

However, in case of partial melting within a polymer, an internal sequence can either remain in its locally melted (LM) or duplex (D) conformation at equilibrium (Fig. 1(A)). The concentration dependence is therefore extraneous. Consequently, following principles of two-state pseudo-unimolecular equilibrium,  $T_m$  should be concentration independent.



At  $T = T_m$ , i.e., at half denaturation,

$$[LM] = [D]$$

therefore,  $K_{\text{eq}} = 1$

$$\Rightarrow \Delta H^0 - T\Delta S^0 = -R \cdot T \cdot \ln(1) = 0$$

$$\text{Thus, } T_m = \frac{\Delta H^0}{\Delta S^0} \quad (2)$$

Annealing accompanies compaction resulting in substantial decrease in the molecular volume of a ds-DNA. Thus ds-DNA has a greater negative-charge density than ss-DNA, and hence, upon denaturation, releases

thermodynamically bound counterions. The equilibrium thermodynamic parameters ( $\Delta H^0$ ,  $\Delta S^0$ ,  $\Delta G^0$ ) derived by Santalucia,<sup>3</sup> however, apply only to nucleic acids dissolved in 1 (M) NaCl at pH 7.0.<sup>4</sup> Therefore, in attempt to imitate an apparent *in-vivo* environment, where the monovalent counterion ( $\text{Mon}^+$ ) concentration is much less ( $\sim 0.05 \rightarrow 0.165$  (M)) and also divalent  $\text{Mg}^{+2}$  is present at different concentrations among cells ( $\sim 0.005 \rightarrow 0.03$  (M)), these parameters were salt corrected according to the latest literature.<sup>23–25</sup>

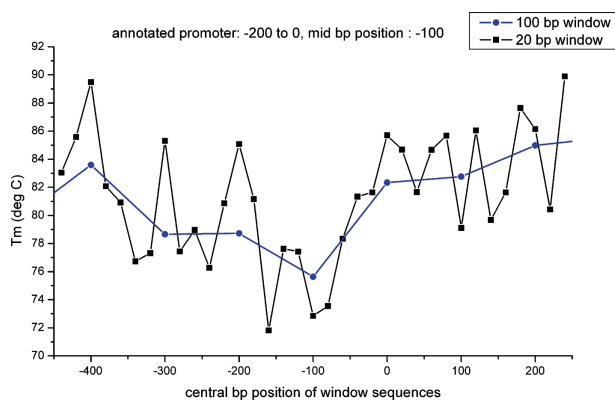
To test whether the method could actually discriminate between oligomeric and local polymeric DNA melting, a 20 bp small genomic sequence from the *E.coli* sakai strain was chosen which resides at the centre of the experimentally validated 232 bp long origin of replication O157<sup>26</sup> and  $T_m$  was calculated twice,

- (i) when the sequence was implanted at the center of the 232 bp long origin, undergoing local melting and
- (ii) when taken in isolation (oligomer).

Different DNA concentrations starting from cellular levels to regular experimental values were fed.

Genomic DNA is highly concentrated *in-vivo* (of the order of 1 (M)) which is about  $10^3$  to  $10^{12}$  fold higher in comparison to what is normally used in *in-vitro* experimental assays with oligomeric DNA (e.g., PCR). Also, oligomeric melting should have fraying effect (due to dangling ends) unlike local polymeric melting. Thus, initiation penalties were taken into account for the former.  $T_m$  of the local unimolecular melting event in the polymer, (calculated by Eq. (3)) was only slightly higher than that of the concentration dependent melting of the oligomer (Eq. (2)) at highly concentrated near cellular level ( $C_T \sim 1\text{--}4$  (M)) where the concentration term ( $\ln(C_T/4)$ ) contributed only marginally to the denominator compared to the entropic term,  $\Delta S^0$  (in cal/mol/K). On the other hand, at gradually decreased concentrations ( $10^{-3}$  to  $10^{-12}$  (M)), the concentration term being more and more pronounced resulted in markedly decreased  $T_m$  values (Table I).

The method could also be used to predict melting origins in annotated sequences. Both plasmids and genomic DNA were tested. The window size was taken to be 20 bp since

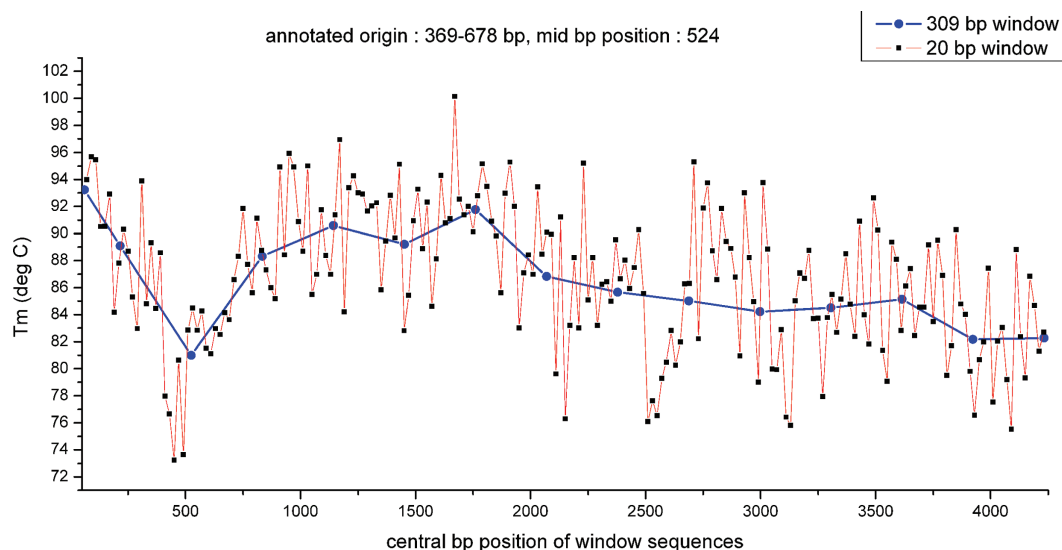


**Fig. 2.**  $T_m$  plot for PLPR0006. The annotated promoter (bp position  $-200$  to  $0$ , central bp position:  $-100$ ) was detected at a single local minima (central bp position:  $-100$ ) at the larger 100 bp scan window (plotted in blue) and the same region was detected to span across 3 local minima at the smaller 20 bp scan window (plotted in black).

such a length is preferred as the binding motif of melting associated DNA binding proteins.<sup>27</sup> Notably, the ‘two-state approximation’ should work better for smaller sequences. Both, origins of replication, and promoter elements were tested for the presence of local minima of stability. Promoter (and their flanking sequences) were collected from a plant promoter database.<sup>28</sup> For example, the candidate promoter, tagged ‘PLPR0006... AC: AF190297... OS’ of *Arabidopsis thaliana*, gene TT1 picked up from bp position  $-200$  to  $+51$  (transcription start site:  $+1$ ) was appended with its 1000 bp upstream and downstream sequences, which were mapped by local alignment of the query sequence against the non-redundant whole *Arabidopsis* genome using NCBI-BLAST.<sup>29</sup> Calculations were performed for both the default 20 bp and also the larger 100 bp window lengths. As expected, the promoter element (represented by the central bp position:  $-100$ ) mapped to a single local minima for the larger (100 bp) window length and the same region (bp position:  $-200$  to  $0$ ) was found to traverse across two local minima (centered at bp position  $-100$  and at  $-160$ ) at the smaller (20 bp) window length (Fig. 2). Thus, by gradually shortening the scan window,

**Table I.** Table of comparison between the concentration dependent melting of a 20 bp long test oligomer (sense strand sequence: 5’-AGGATCATTAAGTGTGAATT-3’) and its concentration independent melting (first row in the table, in bold) when considered to be an internal sequence, embedded at the centre of the following 232 bp long, experimentally characterized origin of replication O157 from the *E.coli* sakai strain.

Type of melting event	$C_T$ (M)	$\ln(C_T/4)$	$\Delta H^0 \times 1000$ (cal/mol)	$\Delta S^0$ (cal/mol/K)	Salt corrected $\Delta S^0$ at 0.165 (M) [ $\text{Mon}^+$ ]	$\Delta S^0 + \ln(C_T/4)$	$T_m$ ( $^{\circ}\text{C}$ ) = $\Delta H^0 / (\Delta S^0 + \ln(C_T/4) - 273.15)$
<b>Local polymeric</b>	<b>Not applicable</b>	<b>Not applicable</b>	<b>-150.9</b>	<b>-426.6</b>	<b>-440.524</b>	<b>Not applicable</b>	<b>69.39</b>
Oligomeric	1.0	-1.39	-146.3	-418.2	-432.34	-433.73	64.15
Oligomeric	2.0	-0.693	-146.3	-418.2	-432.34	-433.03	64.7
Oligomeric	3.0	-0.29	-146.3	-418.2	-432.34	-432.63	65.01
Oligomeric	4.0	0	-146.3	-418.2	-432.34	-432.34	65.24
Oligomeric	$10^{-3}$	-8.295	-146.3	-418.2	-432.34	-440.64	58.87
Oligomeric	$10^{-6}$	-15.21	-146.3	-418.2	-432.34	-447.55	53.74
Oligomeric	$10^{-9}$	-22.11	-146.3	-418.2	-432.34	-454.45	48.71
Oligomeric	$10^{-12}$	-29.02	-146.3	-418.2	-432.34	-463.36	42.58



**Fig. 3.**  $T_m$  plot for pGPS5. The annotated origin of replication was detected (bp position 369–678, central bp position: 524) to map to the global minima at the suggested 309 bp scan window (plotted in blue) which was found to be split into two highly unstable patches (central bp position: 450, 490) at the smaller 20 bp scan window (plotted in black).

it is possible to pinpoint to the actual low melting patches which may serve for the plausible melting nucleation site(s) within a fairly long annotated melting origin.

Commercially available plasmid sequences (Cloning Vectors) with well-characterized origins of replication were also considered for sample calculations.<sup>30</sup> Origins were found to map to one or more local minima of stability in most cases, flanked by relatively stable stretches at either end. They were also detected at the ‘global minima’ however, somewhat less frequently. Such a case was in the plasmid pGPS5, the 309 bp long annotated origin (bp position 369–678) mapped to the global minima of stability (central bp position: 524) against the suggested, large (309 bp) scan window (Fig. 3) and when further sub-scanned with the smaller default 20 bp window, two highly unstable local patches (central bp position: 450, 490) were detected at near-upstream to the center of the annotated origin separated by a rather stable sequence in between (centered at bp position: 470).

*PolDNAmelt* was developed in Perl (version 5.8.8) using its Common Gateway Interface (cgi). The user interface has been set to an array consisting of the compulsory input argument, 5' → 3' sense strand sequence and several other optional arguments namely the choice of thermodynamic parameters (Santalucia<sup>3</sup> or Bresler<sup>5</sup>), window size (in bp), overlap size (in bp), solution temperature (in °C) and counterion concentrations of [Na<sup>+</sup>], [K<sup>+</sup>], [NH<sub>4</sub><sup>+</sup>] and [Mg<sup>+2</sup>] (in mols · lit<sup>-1</sup>). Defaults have been set for all optional arguments (choice of thermodynamic parameters: Santalucia,<sup>3</sup> window size: 20 bp, overlap size: zero, solution temperature: 37 °C, [Na<sup>+</sup>]: 0.015 (M), [K<sup>+</sup>]: 0.15 (M), [NH<sub>4</sub><sup>+</sup>]: 0 adding up to a total monovalent counterion concentration [Mon<sup>+</sup>] of 0.165 (M) and [Mg<sup>+2</sup>]: 0.01 (M)) in order to produce a suggested output if the sequence

alone is submitted. It is to be noted that [Mon<sup>+</sup>] should be between 0.05 and 1.1 (M) (since in this range  $\Delta H^0$  does not require a salt correction<sup>3</sup>) or else the value is forcefully readjusted to the nearest cutoff. The ‘Overlap’ feature is provided in order to average out the discrete  $T_m$  data and also to take care of every possible consecutive stretch individually as potential local minima.

## References and Notes

1. P. Yakovchuk, E. Protozanova, and D. Frank-Kamenetskii, Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.* 34, 564 (2006).
2. A. V. Vologodskii, B. R. Amirikyan, Y. L. Lyubchenko, and M. D. Frank-Kamenetskii, Allowance for heterogeneous stacking in the DNA helix-coil transition theory. *J. Biomol. Struct. Dyn.* 2, 131 (1984).
3. J. Santalucia, Jr, A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci., USA* 95, 1460 (1998).
4. H. T. Allawi and J. Santalucia, Jr, Thermodynamics and NMR of internal G-T mismatches in DNA. *Biochemistry* 36, 10581 (1997).
5. K. J. Breslauer, R. Frank, H. Blocker, and L. A. Marky, Predicting DNA duplex stability from the base sequence. *Proc. Natl. Acad. Sci., USA* 83, 3746 (1986).
6. S. G. Delcourt and R. D. Blake, Stacking energies in DNA. *J. Biol. Chem* 266, 15160 (1991).
7. M. J. Doktycz, R. F. Goldstein, T. M. Paner, F. J. Gallo, and A. S. Benight, Studies of DNA dumbbells I: Melting curves of 17 DNA dumbbells with different duplex stem sequences linked by T4 end-loops evaluation of the nearest-neighbor stacking interactions in DNA. *Biopolymers* 32, 849 (1992).
8. O. Gotoh and Y. Tagashira, Locations of frequently opening regions on natural DNAs and their relation to functional loci. *Biopolymers* 20, 1033 (1981).
9. H. Kuhn and D. Frank-Kamenetskii, Labeling of unique sequences in double-stranded DNA at sites of vicinal nicks generated by nicking endonucleases. *Nucleic Acids Res.* 36, e40 (2008).

10. J. Santalucia, Jr, H. T. Allawi, and P. A. Seneviratne, Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry* 35, 3555 (1996).
11. N. Sugimoto, S. Nakano, M. Yoneyama, and K. Honda, Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. *Nucleic Acids Res.* 24, 4501 (1996).
12. R. D. Blake, J. W. Bizzaro, J. D. Blake, G. R. Day, S. G. Delcourt, J. Knowles, K. A. Marx, and J. Santalucia, Jr, Statistical mechanical simulation of polymeric DNA melting with MELTSIM. *Bioinformatics* 15, 370 (1999).
13. N. R. Markham and M. Zuker, DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res.* 33, W577 (2005).
14. Primer3: <http://frodo.wi.mit.edu/primer3>.
15. CLC Bio: <http://www.clcbio.com/index.php?id=493>.
16. PREMIER Biosoft: <http://www.premierbiosoft.com/crm/jsp/com/pbi/crm/clientside/ProductList.jsp>.
17. R. Owczarzy, P. M. Vallone, R. F. Goldstein, and A. S. Benight, Studies of DNA dumbbells VII: Evaluation of the next nearest-neighbor sequence-dependent interactions in duplex DNA. *Biopolymers* 52, 29 (1999).
18. N. Sugimoto, R. Kierzek, and D. H. Turner, Sequence dependence for the energetics of dangling ends and terminal mismatches in ribonucleic acid. *Biochemistry* 26, 4554 (1987).
19. J. SantaLucia, Jr and D. Hicks, The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct* 33, 415 (2004).
20. UNA Fold: <http://mfold.rna.albany.edu/>.
21. OMP: <http://www.dnasoftware.com/VisualOMP/tabid/108/Default.aspx>.
22. L. A. Marky and K. J. Breslauer, Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers* 26, 1601 (1987).
23. R. Owczarzy, B. G. Moreira, Y. You, M. A. Behlke, and J. A. Walder, Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations. *Biochemistry* 47, 5336 (2008).
24. J. G. Wetmur and J. Fresco, DNA Probes: Applications of the principles of nucleic acid hybridization. *Crit. Rev. BioChem. Mol. Bio.* 26, 227 (1991).
25. J. Marmur and P. Doty, Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. *J. Mol. Biol.* 5, 109 (1962).
26. NCBI: <http://www.ncbi.nlm.nih.gov/sites/entrez?Db=genome&Cmd=ShowDetailView&TermToSearch=169>.
27. S. Jones, P. van Heyningen, H. M. Berman, and J. M. Thornton, Protein-DNA interactions: A structural analysis. *J. Mol. Biol.* 287, 877 (1999).
28. Bioinformatics Web Server: <http://mendel.cs.rhul.ac.uk/mendel.php?topic=plantprom>.
29. BLAST: <http://blast.ncbi.nlm.nih.gov/Blast.cgi>.
30. New England Biolabs: [http://www.neb.com/nebecomm/tech-reference/restriction\\_enzymes/dna\\_sequences\\_maps.asp](http://www.neb.com/nebecomm/tech-reference/restriction_enzymes/dna_sequences_maps.asp).

Received: 10 August 2013. Accepted: 3 September 2013.