MELTDNA: TOOL FOR PREDICTION OF DNA DUPLEX HYBRIDIZATION & MELTING THERMODYNAMICS

Abhishek Tiwari and Vipin Wadhwa

Department of Biotechnology, Vellore Institute of Technology, Vellore, India.
e-mail: abhishek_twr@yahoo.com e-mail: abhishek_twr@gmail.com

Receiving Date: 29-04-05 Accepted On: 27-07-05

Abstract: Prediction of DNA duplex stability and thermodynamics is invaluable for many molecular biology applications involving sequence dependent hybridization reactions. Sequence dependent stability of duplex DNA plays a major role in fundamental processes of the living cell, such as replication, transcription, and recombination. Many techniques in molecular biology depend on the oligonucleotide melting temperature (Tm), and several formulas have been developed to estimate Tm. We have developed a Perl based tool to predict DNA duplex hybridization & melting thermodynamics.

This tool consider all factors those determine thermodynamic values like-Base composition,NN parameters, salt dependency (Na+, Mg++), internal mismatch, dangling ends ,loops etc. For improve NN model based calculations we introduced fusion matrices. MeltDNA provides better prediction accuracy for DNA duplex hybridization especially its ability to handle the structural features in DNA duplex like loops, loop size etc. It has got wide applicability in primer selection in PCR, probe selection in Microarrays, DNA secondary structure stability prediction, DNA computing etc.

Keywords: Hybridization, Melting, Prediction, Oligonucleotides, Nearest-neighbour (NN), Mismatch, Dangling-ends, Hairpins, Bulges, Loops, Salt dependency.

INTRODUCTION

DNA amplification and detection techniques often depend on oligonucleotide melting temperature (Tm).The Tm of a DNA duplex, defined as the temperature where one-half of the nucleotides are paired and one-half are unpaired , corresponds to the midpoint of the spectroscopic hyperchromic absorbance shift during DNA melting. The Tm indicates the transition from double helical to random coil formation. For Tm (Melting Point) prediction there are many formulas some based on nucleotide composition (A+T % & G+C %) while others on Nearest Neighbour Model. Nearest-neighbor (N-N) models provide the highest accuracy for Tm prediction for short oligonucleotide sequences.

GC % Based Methods

The Wallace-Ikatura rule is often used as a rule of thumb when primer Tm is to be estimated at the bench (1, 2). However, the formula was originally applied to the hybridization of probes in 1 mol/L NaCl (l) and is an estimate of the denaturation temperature (Td):

Ta (ç^{a}C) = 2(A+T) +4(G+C)

Another equation for the effective priming temperature (Tp) was suggested:

Tp (ç^{a}C) = 22+1.46Ln

where Ln=2(G+C) + (A+T).The Wallace-Ikatura rule (Eq. 1) overestimates the Tm of long duplexes and gives reasonable results only in the range of 14-20 bp. The equation for the effective priming temperature by equation 2 is similar to the Wallace-Ikatura rule. Only oligonucleotides with Ln <39 produce better results.Marmur and Doty originally established a formula to correlate GC content (%GC-in percentage and not in fractions like 50% and not 0.5) to the Tm of long duplexes at a given ionic strength. Chester and

Marshak (12) added a term to account for DNA strand length (n in base pairs) to estimate primer Tm:

Tm (c^{a}C) =69.3 + 0.41(%GC) 2-650/n

The Marmur-Schildkraut-Doty equation also accounts for ionic strength with a term for the Na^{+} concentration (1, 2, 14-16).

Tm (°C) = 81.5 °C + 16.6(log [Na^{+}]) +0.41(%GC) - b/n

Values between 500 and 750 have been used for b a value that may increase with the ionic strength and b=500 are very commonly used. Another modification is that of Wetmur (1):

Tm=81.5+16.6 log ([Na^{+}] / (1.0+0.7[Na^{+}])) +0.41(%GC)-500/n

Eqs. 4 and 5 assume that the stabilizing effects of cations are the same on all base pairs. However, Owen et al. (4) observed that the slopes of Tm vs. log [Na^{+}] decrease with increasing GC content, leading to the following final formula for the estimation of Tm in polymer DNA (5-7):

Tm (°C) = 87.16 +0.345(%GC) +log [Na^{+}] × [20.17 - 0.066(%GC)]

Eq. 6 has been suggested as the best predictor of polymer DNA Tm (6, 7).

Effect of Monovalent cation (Na+, K+), Bivalent cation (Mg++), dNTP on Tm

A general model commonly used is Tm (observed) = Tm (predicted) - a × DMSO (%), with [Na+]equivalent= [monovalent cations] + b × ([Mg^{++}]-[dNTP])c. Here parameters a, b, and c: commonly with a = 0.75, b = 120, and c = 0.5.The nonlinear effect of [Mg^{++}] on [Na^{+}]equivalent are best approximated by the square-root function, which is in agreement with a different publications (/).Note that under only PCR conditions Mg^{++} and dNTP are present. Also Mg^{++} has larger stabilizing effect on duplex DNA with respect to monovalent Na^{+}. Deoxynucleotide triphosphates (dNTPs) are also essential and chelate some of the available Mg^{++}.

Effect of DMSO on Tm

In addition, dimethyl sulfoxide (DMSO) is commonly used as a cosolvent to facilitate amplification from difficult templates. Addition of DMSO decreases the Tm (9-13), which must be taken into account when primer Tm is calculated (12). Each percentage of DMSO (by volume) decreased the Tm by 0.75 °C(13) for example 5% DMSO will cause decrease in Tm by 0.75×5 °C .GC content had no obvious influence on the DMSO factor. The effect of DMSO on thermal stability of DNA has been investigated before. A factor of 0.75 °C decrease in Tm per 1% DMSO is similar to previous findings of 0.6 °C per 1% DMSO (9, 12), 0.675 °C per 1% DMSO (11), and 0.5 °C per 1% DMSO (10).

Actually prior studies were performed on polymer DNA, suggesting that DMSO may have a slightly greater effect on oligomer DNA. Pattern of decrease in Tm due to DMSO is applicable for both GC % based methods as well as NN model.

Nearest-Neighbor (NN) Model

Measurements on a large number of oligomers revealed that thermodynamic values for Helix-Coil transition or DNA duplex melting($f'G°$,$f'H°$,$f'S°$) did not depend simply on base pair identity (i.e. AT (A-U), or G-C). Rather the energetics depended on the dinucleotide step. This is the clearest manifestation of the importance of nearest-neighbor base stacking in helix stability. In the NN model, sequence dependent stability is considered in terms of n-n doublets. In duplex DNA there are 10 such unique internal nearest-neighbor doublets (INN).

These are AA/TT, AT/TA, TA/AT, CA/GT, GT/CA, CT/GA, GA/CT, CG/GC, GC/CG,GG/CC.Note that the slash indicates the sequences are given in antiparallel orientation. (e.g., CA/GT means 5'-CA-3' is Watson-Crick base paired with 3'- GT-5').And since the strands are antiparallel, MN/M'N' is identical to N'M'/NM. Thus, there are only 10 distinct MN entries (out of the 16 dinucleotide sequences). As a specific example, 5' -CA- 3' is identical to 5' -TG- 3' so CA/GT and TG/AC are identical stacking doublets. Both correspond to the same helical stack so only CA/GT is listed. In our software implementation we used Unified NN parameter set (Table 1) finally revised by J. SantaLucia, Jr (17, 18).

Calculation of Tm, $f'G°$, $f'H°$ and $f'S°$

1. Principal of Additivity: The $f'G°$, $f'H°$ and $f'S°$ contributions of each dinucleotide in a sequence can be summed to obtain the overall values for a given helix.

$$\Delta H^{\circ} = \Delta H^{\circ}_{\text{AT terminal}} + \Delta H^{\circ}_{\text{init}} + \Delta H^{\circ}_{\text{symmetry}} + \sum_{\text{kc\{stacks\}}} \Delta H^{\circ}$$

$$\Delta S^{\circ} = \Delta S^{\circ}_{\text{AT terminal}} + \Delta S^{\circ}_{\text{init}} + \Delta S^{\circ}_{\text{symmetry}} + \sum_{\text{kc\{stacks\}}} \Delta S^{\circ}$$

$$\Delta G^{\circ}_{37} = \Delta G^{\circ}_{\text{AT terminal}} + \Delta G^{\circ}_{\text{init}} + \Delta G^{\circ}_{\text{symmetry}} + \sum_{\text{kc\{stacks\}}} \Delta G^{\circ}$$

Temperature Dependency: Within experimental error and for simple duplexes, neither $f'H°$ nor $f'S°$ has any temperature dependence, i.e. $f'Cp = 0$(.Cp,H =d.H/d Tm & .Cp,S =d.S/d lnTm -both zero). This is an excellent approximation for nucleic acids. Note that predictions of .G¢ª are most accurate at temperatures near 50¢ªC, even though the .G¢ª is traditionally given at 37¢ªC because that is the temperature of the human body, and get worse as the temperature deviates from 50¢ªC.We have fallowed the same approach but it is possible for one to do calculations using considering $f'Cp$ as non zero for $f'H°$ and $f'S°$ (19).

Tm Calculation: Theoretical melting temperature is typically calculated assuming that the coil-helix transition is two-state, which is a justifiable assumption for small oligonucleotides. SantaLucia,et al. (20) suggest that the two-state model is capable of providing a reasonable approximation of melting temperature for duplexes with non-two-state transitions, but the applicability of the assumption obviously decreases as the size of the duplex under consideration increases. For a two-state model of

single-strand + single-strand = double-strand

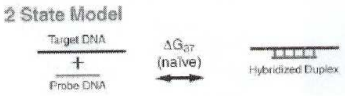


Fig.: 1. Two-State Model of Coil Helix

Transition transition between two distinct oligonucleotides in equimolar concentration, the equilibrium constant is given by

$$K = \frac{2f}{(1-f)^2[C_T]}$$

But the equilibrium constant can also be expressed in thermodynamic terms as

$$K = \exp(-\frac{\Delta G°}{RT}) = \exp(-\frac{\Delta H° - T\Delta S°}{RT})$$

where R is Boltzmann's constant and T is the temperature in Kelvin. Since the melting temperature, Tm, is defined as the temperature at which half of the strands are in the double-stranded state, it follows that f = 1=2 when T = Tm. Setting f = 1=2 and setting the two expressions for K in above two equations equal to one another, we get an equation relating the melting temperature, Tm, to the total molar strand concentration and the enthalpy and entropy of the forward state transition:

$$T_m = \frac{\Delta H°}{\Delta S° + R\ln([C_T]/4)}$$

To get the temperature in degree Celsius

$$T_m = \frac{\Delta H°}{\Delta S° + R\ln([C_T]/4)} - 273.15$$

For self-complementary molecule [CT]/4 in above equation is replaced by CT. For non-selfcomplementary molecule replace [CT]/4 by ((CA -CB)/2) if the strands are at different concentrations, (where CA and CB are the concentrations of the more concentrated and less concentrated strands respectively), otherwise [CT]/4.

Effect of Monovalent cation (Na^+, K^+), Bivalent cation (Mg^{++}), dNTP: The entropy ($f'S°$) & free energy change($f'G°$) is salt dependent, and hence $f'S°$ & $f'G°$ must be corrected if the ionic environment is different from 1 mol/L NaCl, the salt concentration at which most NN model thermodynamic values have been derived. However, Mg^{++} is present in PCR as an important cofactor for Taq DNA polymerase and strongly nfluences $f'S°$ & $f'G°$.Deoxynucleotide triphosphates (dNTPs) are also essential and chelate some of the available Mg^{++}. In addition, dimethyl sulfoxide (DMSO) is commonly used as a cosolvent and presence of DMSO decreases the Tm as described before. Thermodynamic value after salt correction are calculated by fallowing formulas

$$\Delta G°_{37}[Na+] = \Delta G°_{37}[1\ M\ NaCl] - 0.114 \times N/2 \times \ln[Na+]_{eq}$$

$$\Delta S°[Na+] = \Delta S°[1\ M\ NaCl] + 0.368 \times N/2 \times \ln[Na+]_{eq}$$

where .G¢ª37¢ª[Na^+] and .S¢ª[Na^+] are salt corrected values from NN model and N is the total number of phosphates in the duplex (For duplex without terminal phosphate N/2=oligonucleotide length-1, e.g., for an 8-bp duplex without terminal phosphates, N/2=7), and [NaCl] is the total concentration of monovalent cations from all sources (the same equation works for sodium, potassium, and ammonium; J. SantaLucia, unpublished experiments). The salt correction for DNA is assumed to be independent of sequence but to be dependent on oligonucleotide length. The $f'H°$ is assumed to be independent of [NaC], which is valid for nucleic acids for total sodium concentrations above 0.05 M and below 1.1 M. Above equations performs better only if oligonucleotide sequences are short. For polymers the coefficient in front of the natural logarithm changes to 0.175, presumably owing to counterion condensation effects. A salt dependence function that accounts for all lengths

has not yet been derived. In addition, the equation applies only to duplexes that melt in a two-state fashion, which often is not the case for longer duplexes where single strand folding can compete with duplex formation and where slow dissociation kinetics can inhibit equilibration. Also [Na^+]aq is given by [Na^+]eq= [monovalent cations] + b × ([Mg^{++}]-[dNTP])c. Here parameters a, b, and c: commonly with a = 0.75, b = 120, and c = 0.5 as described before.

Effect of Internal Singal Mismatch: DNA mismatches occur as a result of errors during replication, due to heteroduplex formation during homologous recombination and mutagenic chemicals and ionizing radiation or spontaneous deamination. Mismatches also occur in the secondary structures of single-stranded DNA viruses. In addition to stable canonical Watson-Crick base pairs (G·C and A·T) there are eight possible mispairs of varying stability and structure, namely A·A, A·C, C·C, C·T, G·G, G·A, G·T and T·T. In order to understand the origins of various mismatch occurrences and to help in our interpretation of mismatch recognition and repair mechanisms, thermodynamics of these mismatches are experimentally determined by Allawi HT, SantaLucia J Jr. We used the values determined by Allawi HT, SantaLucia J Jr. .So when $f'G°$, $f'H°$, $f'S°$ & Tm are calculated using NN method a correction is introduced for internal single mismatch. Values for internal multiple mismatch still to be determined experimentally.

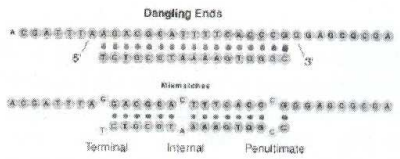


Fig.:2. Dangling ends & Mismatches in DNA duplex

Also there is possibility of terminal mismatches. It is observed that all terminal mismatches are stabilizing, whereas internal mismatches may be either stabilizing or destabilizing; presumably, the destabilizing internal mismatches are due to unfavorable helical constraints that prevent the formation of the optimal stacking and H-bond geometry.

Effect of Dangling Ends: Oligonucleotide probe hybridization to a long target DNA involves not only base pairing, but also two dangling end contributions. Accounting for dangling end effects is crucial to accurately calculate probe/target binding strength and dangling-end parameters should be incorporated during thermodynamic calculations. The average 5'-dangling end contributes -0.45 kcal mol-1, while the average 3'-dangling end contributes -0.29 kcal mol-1.

Effect of Loops: Secondary structure elements (like hairpins, bulges, internal loops) etc have very vast effect on nucleic acid thermodynamics. All loop .H¢^{a} parameters are assumed to equal zero. The loop .S¢^{a} increment may be calculated from: .S¢^{a} = .G¢^{a}37¢^{a} ×1000/310.15 where entropy is in eu. A Jacobson-Stockmayer entropy extrapolation is then used to fill in the gaps and provide parameters for closure of long loops (26) according to equation

$$\Delta G^{\circ}_{37}(\text{total}) = \Delta G^{\circ}_{37}(\text{Hairpin of 3}) + \Delta G^{\circ}_{37}(\text{triloop bonus}) + \text{closing AT penalty},$$

where .G¢^{a}37¢^{a} (loop-n) is the free energy increment of a loop of length n, .G¢^{a}37¢^{a} (loop-x) is the freeenergy increment of the longest loop of length x for which there are experimental data, and R is the gas constant. The increments for loop lengths not shown in table 6 may be calculated by above equation.

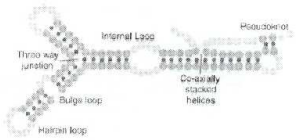


Fig:3. Different type of loop element

Hairpin Loops: Hairpins with lengths of 3 and 4 are treated differently than longer hairpin loops because certain sequences are particularly stable. Importantly, these stable triloop and tetraloop sequences have a significant probability of occurring by random chance, in probes, primers, and targets, and they can significantly inhibit hybridization in various assays. Hairpin loops with lengths shorter than 3 are sterically prohibited. For hairpins of length 3

$$\Delta G^{\circ}_{37}(\text{total}) = \Delta G^{\circ}_{37}(\text{Hairpin of 4}) + \Delta G^{\circ}_{37}(\text{tetraloop bonus}) + \Delta G^{\circ}_{37}(\text{terminal mismatch}).$$

where .G¢^{a}37¢^{a} (total) is total increment due to hairpin loop, .G¢^{a}37¢^{a}(Hairpin of 3) is +3.5 kcal mol-1 (Table 6) and the closing AT penalty is +0.5 kcal mol-1 and is applied only to hairpin sequences that are closed by AT. Similarly for hairpins of length 4

$$\Delta G^{\circ}_{37}(\text{total}) = \Delta G^{\circ}_{37}(\text{Hairpin of N}) + \Delta G^{\circ}_{37}(\text{terminal mismatch}),$$

where .G¢^{a}37¢^{a} (total) is total increment due to

hairpin loop, .G¢ª37¢ª (Hairpin of 4) is +3.5 kcal mol-1 (Table 4) and .G¢ª37¢ª (terminal mismatch) is the increment for terminal mismatches. Values for .G¢ª37¢ª (triloop bonus) and .G¢ª37¢ª(tetraloop bonus) are available along with this article. For hairpin loops with lengths longer than 4,

$$\Delta G^{\circ}_{37}(\text{Loop total}) = \Delta G^{\circ}_{37}(\text{Internal Loop of N}) + \Delta G^{\circ}_{37}(\text{asymmetry}) + \Delta G^{\circ}_{37}(\text{left terminal mismatch}) + \Delta G^{\circ}_{37}(\text{right terminal mismatch}),$$

where .G¢ª37¢ª (total) is total increment due to hairpin loop, .G¢ª37¢ª(Hairpin of N) is given in Table 6. To compute the stability of a complete hairpin + stem, one simply adds the salt-corrected base pair NN contributions to the loop energy correction that is

.G¢ª37¢ª(net) =Salt corrected .G¢ª37¢ª from NN model + .G¢ª37¢ª total correction for hairpin loops

Similar to .G¢ª37¢ª, .H¢ª and .S¢ª can also calculated. The thermodynamic contributions of loop nucleotides of a hairpin are assumed to be salt concentration independent. The two-state Tm for hairpins is calculated from Tm=.H¢ª/.S¢ª-273.15. Internal Loops: Internal loop stability is calculated according to fallowing equation

$$\Delta G^{\circ}_{37}(\text{Loop total}) = \Delta G^{\circ}_{37}(\text{Internal Loop of N}) + \Delta G^{\circ}_{37}(\text{asymmetry}) + \Delta G^{\circ}_{37}(\text{left terminal mismatch}) + \Delta G^{\circ}_{37}(\text{right terminal mismatch}),$$

where .G¢ª37¢ª(asymmetry) =|length A -length B| × 0.3 kcal mol-1 and A and B are the lengths of both sides of the internal loop. Internal loops of two are calculated using the mismatch nearest neighbor parameters.

Bulges: Bulges of length 1 are calculated assuming that they are "flipped out," and thus the intervening base pair stack is added (the same approximation is used in RNA).

$$\Delta G^{\circ}_{37}(\text{Loop total}) = \Delta G^{\circ}_{37}(\text{Bulge Loop of 1}) + \Delta G^{\circ}_{37}(\text{intervening NN}) + \text{closing AT penalty},$$

For bulge loops longer than 1, above equation is applied, but the intervening NN term is not added. The large destabilizing contribution of bulges means that they are relatively rare in DNA secondary structures of random sequences.

Coaxial Stacking Parameters: Coaxial stacking parameters are important for accurately predicting the stability of multibranched loops. Coaxial stacking occurs when two oligonucleotides hybridize at adjacent locations on a template or when a probe DNA binds next to a unimolecular hairpin of a template.Alternatively,coaxial stacking may be thought of as occurring as the result of a strand "nick". Consider the two structures below:

GGTCGCTCTG	GGTCG–CTCTG
CCAGCGAGAC	CCAGC/GAGAC
ΔG°_{37}(Total) = −12.19	ΔG°_{37}(Total) = −11.34

The structure on the left is a normal 10-bp bimolecular duplex, and the structure on the right is a trimolecular coaxially stacked complex of two 5-mers bound to one 10-mer. The "nick" site is indicated by the slash, "/"; the dash "-" indicates the covalently continuous strand. The method for computing the total stability of the coaxially stacked complex is shown in fallowing equation

$$\Delta G^{\circ}_{37}(\text{w/nick}) = \Delta G^{\circ}_{37}(\text{without nick}) - \Delta G^{\circ}_{37}(\text{GC/CG}) + \Delta G^{\circ}_{37}(\text{G-C + C/G coaxial}) + \text{extra initiation},$$

where -3.35 kcal mol-1 is the measured coaxial stacking contribution. The extra initiation penalty is required because another bimolecular event must take place to form the coaxially stacked complex.

Thermodynamic analysis of DNA duplexes on microchips

By using fallowing relationships several approaches for calculating thermodynamic parameters of DNA duplexes in solution can be used to estimate the same parameters for DNA duplexes on microchips, and vice versa.

$$\Delta G^{\circ}\ (\text{microchip}) = a \times \Delta G^{\circ}\ (\text{solution}) + g,$$

where a = 1.1 ± 0.2 and g = -3.2 ± 0.4 kcal/mol. Or, assuming that a = 1:

$$\Delta G^{\circ}\ (\text{microchip}) = a \times \Delta G^{\circ}\ (\text{solution}) + g,$$

The best linear fit for plots of Tm for the microchips versus solution is:

$$\Delta G^{\circ}\ (\text{microchip}) = \Delta G^{\circ}\ (\text{solution}) - (3.2 \pm 0.4)\ \text{kcal mol}^{-1},$$

Correction parameters between microchip and solution plots for .S¢ª and .H¢ª may also be treated as additive constants:

ΔH° (microchip) $= \Delta H^\circ$ (solution) – (24±4) kcal mol^{-1},
ΔS° (microchip) $= \Delta S^\circ$ (solution) – (70±12) cal mol^{-1}.K,

Prediction of Extinction (absorption) coefficient of DNA

Extinction coefficient at 260 nm, 25 degrees of Celsius, and neutral pH for the single-strand DNA is predicted by the nearest-neighbor method. Values used for extinction coefficient determination are given in table 7. For example, the extinction coefficient of oligomer 5'-ATGCTTC-3' is

$$\epsilon_{ATGCTTC} = 2(\epsilon_{dApdT}+\epsilon_{dTpdG}+\epsilon_{dGpdC}+\epsilon_{dCpdT}+\epsilon_{dTpdT}+\epsilon_{dTpdC}) - \epsilon_{pdT}-\epsilon_{pdG}-\epsilon_{pdC}-\epsilon_{pdT}-\epsilon_{pdT}$$

For linear DNA molecule

$$\epsilon_{ATGCTTC} = \epsilon_{dApdT}+\epsilon_{dTpdG}+\epsilon_{dGpdC}+\epsilon_{dCpdT}+\epsilon_{dTpdT}+\epsilon_{dTpdC}+\epsilon_{dCpdA}$$

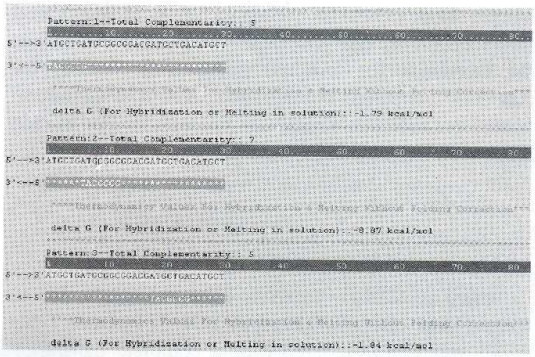
For circular DNA molecule

ANNEALING TEMPERATURE

In PCR technique, DNA is amplified in vitro by a series of polymerization cycles consisting of three temperature dependent steps: DNA danaturation, primer-template annealing, and DNA synthesis by a thermostable DNA polymerase. The purity and yield of reaction products depend on several parameters, one of which is the annealing temperature (Ta).Optimal Ta is found to be function of the melting temperature of less stable primer-template pair and of the product. So Ta is Ta=0.3×TmPrimer+0.7×TmProduct-14.9 in which TmPrimer is the calculated Tm of less stable primer template pair by NN model and TmProduct is the Tm of the PCR product and calculated using equation (4) and not by NN mode(as NN model is not applicable to long DNA molecule)

METHODS

Figure 5.Output from MeltDNA for given two sequences 5'-atgctgatgcggcggacgatgctgacatg ct-3' and 3'-tacgccg-5' maximum complementarity is 7 so cutoff binding ability is 5 and hence 3 binding patterns are possible. For all possible patterns thermodynamics calculations is done. It exhibits that the binding pattern 2 is most stable as lower .Gc^a.



Fusion Matrices

Calculation of thermodynamic values (Tm, .Gc^a, .Hc^a, .Sc^a) using NN model is complex process as one has to consider NN doublets, internal mismatch, dangling ends etc and then use corresponding parameters. So overcome this problem we developed a fusion matrix of 25×25 in which all possible doublet from four nucleotide characters (A, T, G, C) and an * character for dangling ends are taken.

Three fusion matrices were calculated for $f'G^\circ$, $f'H^\circ$ and $f'S^\circ$ doublet parameters for 1 M NaCl salt concentration. In fusion matrices all data for NN model-normal values (Unified NN parameters), dangling end values, and internal single mismatch values were fused to form a single value matrix for each $f'G^\circ$, $f'H^\circ$ and $f'S^\circ$. Just now fusion matrices are not complete as values for internal double mismatch and terminal mismatch are not available. For computational simplicity we have considered unavailable values as zero. In future as fusion matrixes become more complete and improved they will get more new applications especially in DNA computing.

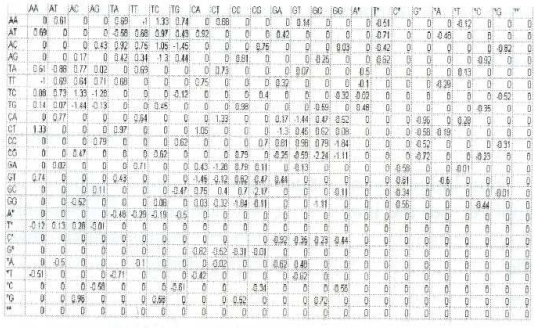


Fig.: 6. Fusion Matrix For .Gc^{a}37c^a .Similarly fusion matrices for .Hc^a and .Sc^a are given.

Possible Binding Patterns

MeltDNA first of all calculates all possible binding patterns between two DNA sequences (fully or partially complementary) those have cutoff binding ability in terms of total complementarity. We have taken cutoff binding ability as 60 % of maximum complementarity between two sequences. But user can use lower or higher cutoff limits. Further for each binding pattern thermodynamic values are calculated (for example see figure 4).

Calculations

When two strands (e.g., upper-ATGTCCGTTGCAT and lower-GGCAACG) are entered they are first complimentary aligned using end character * as shown below

5'-ATGTCCGTTGCAT-3'

3'-****GGCAACG**-5'

Then possible theoretical doublets in above example are AT/**, TG/**, GT/**, TC/*G, CC/GG,..., CA/G*, AT/** .Then from theoretical doublets using principal of additivity and fusion matrices .Gc^a,. Hc^a, .Sc^a are calculated. Salt correction for .Gc^a & .Sc^a is carried out. Default calculation is done using 1 M NaCl salt concentration and .0002 M as strand concentration considering absence of Mg^{++}, dNTPs and DMSO at temperature 37c^a C. For loop elements a separate calculation for $f'G°$, $f'H°$ and $f'S°$ is done according to loop size and loop type (internal, bulge, hairpin etc).Other structural factors also considered as closing AT penalty for loops, terminal mismatches, tri and tetra hairpin loop sequences ,coaxial stacking, intervening NN etc. So net values are given by

ΔG°_{37} (net) =Salt corrected ΔG°_{37} from NN model + ΔG°_{37} total correction for loops

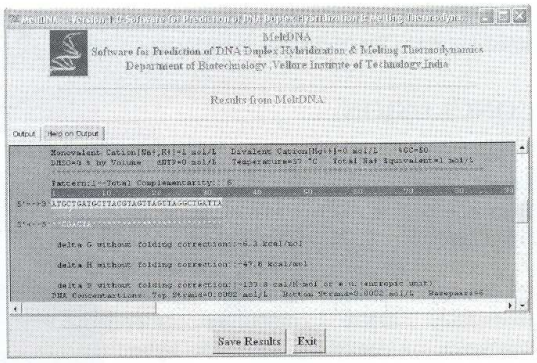
ΔH°(net) =ΔH° from NN model + ΔH° total correction for loops

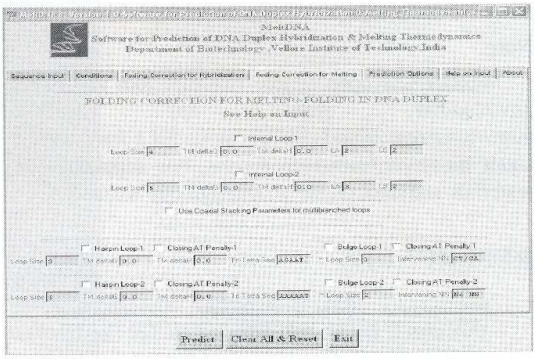
ΔS° (net) =Salt corrected ΔS° from NN model + ΔS° total correction for loops

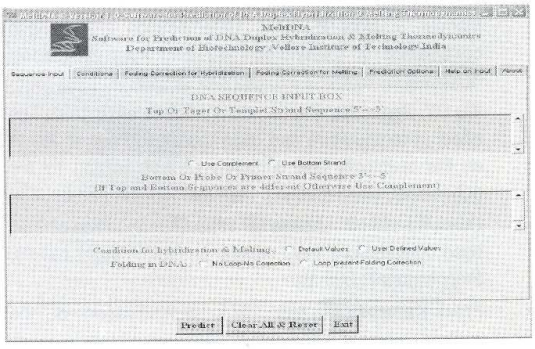
where, Total correction for loops = .hairpin loop + .bulge loop + .internal loop + .coaxial stacking After calculating net values Tm is calculated using $f'H°$ and $f'S°$ .Further Tm is corrected for DMSO. MeltDNA also calculates annealing temperature (for primer-template annealing) and extinction coefficient for DNA molecules at 260 nm.

RESULTS & DISCUSSION

MeltDNA gives good prediction accuracy for thermodynamic values. For a dataset of 100 sequences ranging in length from 4 to 24 bp (the experimentally determined values for these sequences with experimental condition were extracted from Nucleic Acid Thermodynamic Database and other published sources) the average deviation between experimental and predicted is 1.63c^a C (corresponding to a standard deviation of 2.4c^aC).Similar type accuracy is found for $f'G°$, $f'H°$ and $f'S°$ prediction. This level of prediction accuracy is sufficient for most applications of nucleic acids, and performance will improve more if more improved NN model parameters will be available.







FURTHER IMPROVEMENTS

In future a lot of improvements are possible in MeltDNA. One of them is to introduce RNA-RNA hybridization and DNA-RNA hybridization calculations in MeltDNA. For this purpose we are calculating 36×36 fusion matrices in which all possible doublet from five nucleotide characters (A, T, G, C,U) and an * character for dangling ends are taken. Also folding algorithms will be added so that it can identify type of loops & loop size. Currently for secondary structure thermodynamics prediction user has to enter information about loop elements and their size. A web based interface for MeltDNA is under development

CONCLUSION

MeltDNA is powerful and efficient tool for DNA duplex hybridization & melting thermodynamics prediction both in solution as well as on microchip with broad applicability. MeltDNA has got wide applicability in primer selection in PCR, probe selection in Microarrays, DNA secondary structure stability prediction, DNA computing (thermodynamic Simulation of deoxyoligonucleotide hybridization for DNA computation) etc.

Note

MeltDNA is open source freeware tool and easily downloaded from our website http://bioserver.tripod.com or ask us for a free copy by email bioserver.team@gmail.com. For supplementary materials and data please go through our website.

REFERENCES

- Wetmur JG. 1991; DNA probes: applications of the principles of nucleic acid hybridization. *Crit Rev Biochem Mol Biol* **26:** 227-59.
- Marmur J, Doty P. 1962; Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. *Mol Biol* **5:** 109-18.
- Wu DY, Ugozzoli L, Pal BK, Qian J, Wallace RB 1991. The effect of temperature and oligo -nucleotide primer length on the specificity and efficiency of amplification by the polymerase chain reaction. *DNA Cell Biol* **10:** 233-8.
- Owen RJ, Hill LR, Lapage SP 1969. Determination of DNA base compositions from melting profiles in dilute buffers. *Biopolymers* **7:** 503-16.
- Frank-Kamenetskii MD 1971. Simplification of the empirical relationship between melting temperature of DNA, its GC content and concentration of sodium ions in solution. *Biopolymers* **10:** 2623-4.
- Blake RD 1996. Denaturation of DNA. In: Meyers RA, ed. Encyclopedia of molecular biology and molecular medicine, Vol. 2. Weinheim, Germany: *VCH Verlagsgesellschaft,* 1-19.
- Blake RD, Delcourt SG 1998. Thermal stability of DNA. *Nucleic Acids Res* **26:** 3323 -32.
- Musielski H, Mann W, Laue R, Michel S 1981. Influence of dimethylsulfoxide on transcript -ion by bacteriophage T3-induced RNA polymerase. *Z Allg Mikrobiol* **21:** 447-56.
- Cullen BR, Bick MD 1976. Thermal denaturation of DNA from bromodeoxyuridine substituted cells. *Nucleic Acids Res*; **3:** 49-62.
- Escara JF, Hutton JR 1980. Thermal stability and renaturation of DNA in dimethyl sulfoxide solutions: acceleration of the renaturation rate. *Biopolymers*; **19:** 1315-27.
- Chester N, Marshak DR 1993. Dimethyl sulfoxide-mediated primer Tm reduction: a method for analyzing the role of renaturation temperature in the polymerase chain reaction. *Anal Biochem*; **209:** 284-90.
- Nicolas von Ahsen, Carl T. Wittwer and Ekkehard Sch.tz 2001. Oligonucleotide Melting Temperatures under PCR Conditions: Nearest-Neighbor Corrections for Mg++, dNTP, and Dimethyl Sulfoxide Concentrations with Comparison to Alternative Empirical Formulas. Clinical Chemistry -*Molecular Diagnostics and Genetics*; 1956-1961.
- Schildkraut C, Lifson S 1965. Dependence of the melting temperature of DNA on salt concentration. *Biopolymers*; **3:** 195-208.
- Wahl GM, Berger SL, Kimmel AR 1987. Molecular hybridization of immobilized nucleic acids: theoretical concepts and practical considerations. *Methods Enzymol*; **152:** 399-407.
- Britten RJ, Graham DE, Neufeld BR 1974. Analysis of repeating DNA sequences by reassociation. *Methods Enzymol*; **29:** 363-418.
- SantaLucia J Jr. 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA* **95:** 1460-65
- Peng Wu1, Shu-ichi Nakano and Naoki Sugimoto 2002. Temperature dependence of thermodynamic properties for DNA/DNA and RNA/DNA duplex formation. *Eur. J. Biochem.* **269:** 2821-2830.
- SantaLucia, J., Allawi, H. & Seneviratne, P. A. (1996) *Biochemistry* **35:** 3555-3562.

- Allawi HT, SantaLucia J Jr. 1997. Thermodynamics and NMR of internal G-T mismatches in DNA. *Biochemistry* **36:** 10581-94
- AllawiHT, SantaLucia J Jr. 1998. Nearest neighbor thermodynamics of internal A-C mismatches in DNA: sequence dependence and pH effects. *Biochemistry* **37:** 9435-44
- AllawiHT, SantaLucia J Jr. 1998. Nearestneighbor thermodynamics parameters for internal G-A mismatches in DNA. *Biochemistry* **37:** 2170-79
- Allawi HT, SantaLucia J Jr. 1998. Thermodynamics of internal C-T mismatches in DNA. *Nucleic Acids Res.* **26:** 2694-701
- Peyret N, Seneviratne PA, Allawi HT, SantaLucia J Jr. 1999. Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A-A, C-C, G-G, and *T-T mismatches. Biochemistry* **38:** 3468-77
- Jacobson H, Stockmayer WH. 1950. Intramolecular reaction in polycondensations.I. The theory of linear systems. *J.Chem. Phys.* **18:** 1600-6
- Alexander V. Fotin, Aleksei L. Drobyshev, Dmitri Y. Proudnikov, Alexander N. Perov and Andrei D. Mirzabekov 1998 . Parallel thermodynamic analysis of duplexes on oligodeoxyribonucleotide microchips. *Nucleic Acids Research.* 1515-1521
- Cantor C.R., Warshaw M.M., Shapiro, H 1970. *Biopolymers,* **9:** 1059-1077.
- Handbook of Biochemistry and Molecular Biology 1975, Volume 1: Nucleic Acids, Fasman, G.D. editor, page 589, 3rd edition, *CRC Press.*
- Rychlik,W., W.J.Spencer and R.E.Rhoads. 1990. Optimization of the annealing temperature for DNA amplification in vitro [published erratum appears in Nucleic Acids Res 1991 Feb 11; 19(3):698]. *Nucleic Acids Res* **18:** 6409-6412.
- John SantaLucia, Jr. and Donald Hicks 2004. The Thermodynamics of DNA Structural Motifs. *Annu. Rev. Biophys. Biomol. Struct.* **33:** 415-40.