Spreadsheet Software for Thermodynamic Melting Point Prediction of Oligonucleotide Hybridization with and without Mismatches

BioTechniques 27:1218-1224 (December 1999)

Ekkehard Schütz and Nicolas von Ahsen
Georg-August-University, Göttingen, Germany

ABSTRACT

The use of thermodynamic parameters for the calculation of oligonucleotide duplex stability provides the best estimates of oligonucleotide melting temperatures (T_m). Such estimates can be used for evidence-based design of molecular biological experiments in which oligonucleotide melting behavior is a critical issue, such as temperature or denaturing gradient gel electrophoreses, Southern blotting or hybridization probe assays on the LightCycler™. We have developed a user friendly program for T_m calculation of matched and mismatched probes using the spreadsheet software Microsoft Excel®. The most recently published values for entropy and enthalpy of Watson-Crick pairs are used, and salt and oligonucleotide concentrations are considered. The 5′ and 3′ end stability is calculated for the estimation of primer specificity. In addition, the influence of all possible mutations under a given probe can be calculated automatically. The experimental evaluation of predicted T_m with the LightCycler, based on 14 hybridization probes for different gene loci, showed an excellent fit between measured results and values predicted with the thermodynamic model in 14 matched, 25 single mismatched and 8 two-point mismatched assays (r=0.98; Sy.x= 0.90; y=1.01× –0.38). This program is extremely useful for the design of oligonucleotide probes because the use of probes that do not discriminate with a reasonable T_m difference between wild-type and mutation can be avoided in advance.

INTRODUCTION

The chemical- or temperature-driven melting behavior of short oligonucleotides that are hybridized to PCR-amplified DNA is the basic principle of several mutation detection techniques. These techniques include temperature or denaturing gradient gel electrophoreses (TGGE/DGGE) (17), DNA blotting (7) and the hybridization probe assay on the Light Cycler™ (Roche Molecular Biochemicals) (8). They are only possible because the stability of dsDNA oligonucleotides is grossly influenced by mismatches under a specifically designed synthetic probe (18). During the last decade, several publications that discuss the influence of mismatches on dsDNA stability, based on thermodynamic principles, have offered insight into the underlying physico-chemical mechanisms (1–4,10). The stability of a single base pair binding is influenced by the surrounding base pairs (nearest neighbor). The thermodynamic parameters for every matched or mismatched base pair are experimentally defined (1–4,10). Such definition allows the prediction of oligonucleotide duplex stability with or without mismatches. For longer sequences (>50 bp), alternative methods based on length and GC content give reliable estimates of the melting temperature (T_m) (18), but only for duplex stability without mismatches. However, mismatches have only a minor influence on the T_m of longer dsDNAs.

For the enthalpy and entropy of matched pairs, unified values for the prediction of DNA duplex stability are available that provide a reliable calculation of the T_m (12). The T_m is the temperature at which 50% of the oligonucleotides in solution are in a duplex formation. Because entropy depends on salt and oligonucleotide concentration, the T_m is also influenced by these parameters. Therefore, they must be considered by the use of experimentally derived correction equations (12).

We have used these parameters to develop a user friendly program for the T_m calculation using the Excel® spreadsheet software (Microsoft, Redmond, WA, USA). It offers a sequence that can be pasted from any source into the input table. The mismatch must be defined together with sodium buffer equivalents and DNA concentration to calculate the T_m of both matched and mismatched oligonucleotide probes. A function to calculate the influence of any possible single base pair mismatch under the probe is also included. Nearly every laboratory will be able to meet the program's hardware and software needs. The program is fully compatible with the Windows environment. Copy and paste into every program running under Windows is possible (e.g., into e-mail oligonu-

cleotide custom synthesis order programs). This program reliably calculates estimates of hybridization probes' T_m when compared with experimental data from genotyping assays used in our laboratory and is of particular interest for the design of such hybridization-probe mutation detection systems.

MATERIALS AND METHODS

Every standard PC equipped with Microsoft Windows (NT4.0, WIN95 or WIN98) or a Macintosh® computer (Apple Computer, Cupertino, CA, USA) with Microsoft Excel Ver. 5.0 and later can run the program. For the calculation of enthalpy (ΔH^0) and entropy (ΔS^0), we used the unified parameters derived from six independent laboratories (12). Parameters for the stability of mismatched duplexes were taken from publications of the same group (1–4,10). The Appendix gives the equations that are used for the calculations in this program. Oligonucleotides (PCR primers and fluorescein probes) were custom synthesized (MWG-Biotech, Germany). LC-Red640-N-hydroxysuccimide ester (Roche Molecular Biochemicals, Mannheim, Germany) was linked with the respective oligonucleotide by using an aminolinker purchased from Glen Research (Sterling, VA, USA) and was purified by HPLC. Experimental T_m values were determined with the LightCycler using standard techniques in hybridization probe format (16). The actual T_m was calculated by the LightCycler software (Ver. 3.0) as the maximum of the first derivative of the fluorescence as opposed to the temperature function. Experiments were run in triplicates and the CV was <1% for experiments repeated on different days.

Program Overview

The Excel spreadsheet consists of three parts: an input table, an output table for all possible mismatches under the probe and the program itself. After loading, the program adds a new menu item to the standard menu that incorporates four program functions (Figure 1): *(i)* calculation of the T_m of self-complementary oligonucleotides; *(ii)* calculation of the T_m of non-self-complementary oligonucleotides; *(iii)* generation of a spreadsheet with the T_m of any possible one-point mutation; *(iv)* calculation of buffer strength in sodium equivalents.

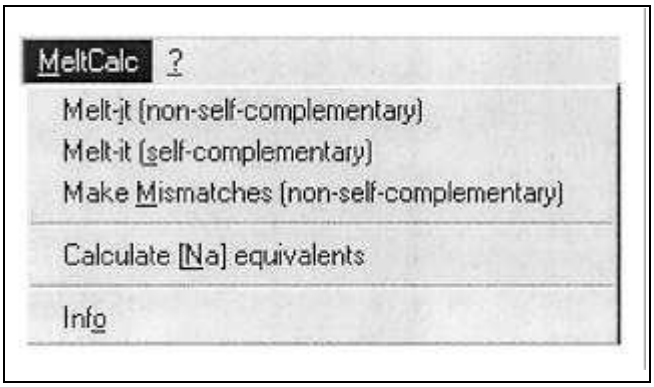

Figure 1. Menu for the calculation of oligonucleotide thermodynamics.

The calculation of oligonucleotide T_m takes <1 s for the program items 1 and 2. The time for calculation of all possible single-point mismatch T_m under the probe (program 3) depends on the length of the probe because 3× (x-2) sequences are considered (where x is the length of the probe). For example, it takes approximately 5 s to generate the output spreadsheet for a 30-mer probe on a PC equipped with an AMD K6-350 that runs on Windows NT 4.0 and Microsoft Excel 7.0.

Users can type the sequence (only nucleotides A,T,G and C are allowed) into the input table or paste it from other sources, such as the GenBank® database (**HTTP://www2.ncbi.nlm.nih.gov/cgi-bin/genbank**). The sequence must be free of unnecessary characters, such as numbers and spaces. Two sequences are always required as input. The length of the input sequence is restricted to <256 bp because the model is not useful for longer DNA sequences. The oligonucleotide concentration and the concentration of sodium equivalents must be entered in the respective cells because these parameters also influence the T_m. For user convenience, a sodium equivalent calculator computes the monovalent ion concentration on user input of Mg^{++} and K^+ concentration (6) of the used buffers; this number is then pasted into the respective cells of the input table. Until the influence of Mg^{++} on DNA duplex stability is fully investigated, we recommend this formula. However, Mg^{++} concentrations higher than approximately 8 mM will not be properly accounted for. When dimethyl sulfoxide (DMSO) is included in the PCRs, its influence on T_m is considered when a value >0% is given in the respective cell.

After these steps are completed, the user can start the calculation by choosing the desired function from the "MeltCalc" menu. When starting the T_m calculation "Melt-it" (self-complementary or non-self-complementary), the given sequences are separated into single nearest-neighbor pairs, and the thermodynamic values of the resulting pairs are taken from the internal table. If the second sequence contains a mismatch, then the respective mismatch parameters are used. Total thermodynamic values for the sequences are then calculated according to the equations given in the Appendix. The program writes back T_m values for the matched and mismatched oligonucleotide and the difference in T_m between the given wild-type and mutation (mismatch). In addition, GC%, the free energy (ΔG^0) of the 5′ and 3′ pentamer end of the oligonucleotide are also calculated. The pentamer stability is useful for the estimation of primer performance because it has been shown that primers with high priming efficiency are somewhat more stable at their 3′ end (11). The system calculates enthalpy and entropy values under standard conditions (1 mol/L NaCl, 100 mol/L oligonucleotide). For the entropy, the values after correction of the given salt concentration are calculated and shown. Because it might be favorable to probe the antisense strand for a mutation detection caused by a higher T_m difference between matched and unmatched sequence, the same outputs just described are given for this strand together with the respective sequence. Figure 2 shows the calculation of duplex thermodynamics of an oligonucleotide with a mismatch.

Additionally, the program will calculate the influence of

every possible mutation that can occur under a given probe. This is helpful to estimate the detection specificity of a mutation. When the user selects this menu item, the system calculates the T_m for every possible single point mismatch between the perfect matched and the mismatched sequence. Results are sorted by the T_m difference with the perfect match and are written into a separate spreadsheet together with the respective sequence. The position of the mismatch in the sequence is indicated by an uppercase letter together with the base that was changed. The complementary reversed sequence is also displayed in an additional row (Figure 3).

Because no thermodynamic parameters for dangling ends, double and terminal mismatches are available as yet, these are not calculated by the program. The same restrictions apply for double mismatches as direct neighbors, with the exception of those pairs (e.g., GG/**TT** or GT/**TG**) for which respective values are known (1). Such calculations are only possible with the HyTher program of SantaLucia's group, available at **HTTP://jsl1.chem.wayne.edu** that uses as yet unpublished thermodynamic data. For all other situations, the presented spreadsheet application calculates results equivalent to the HyTher program.

RESULTS

We have generated eight different single base pair mutations by site-directed mutagenesis under a wild-type and a mutation complementary probe that we used for the detection of the apolipoprotein-B3500 mutation (14). These mutations were used with two different probes for experimental validation of the calculated T_m values together with 12 other hybridization-probe sets that are currently used in our laboratory for genotyping of various gene loci (15,16 and unpublished

Oligoconc. [µM]	0.1	DMSO [%]	5	Unified parameters					
Na eq. [mM]	300								
Length [bp]	19	Tm	ΔTm	ΔG-5(5'-end)	ΔG-5(3'-end)	GC[%]	ΔH	ΔS1	ΔSmod
Sequence (perfect match)	cggcgattgtcgcaccagt	64.5	1.7	-8.6	-6.0	63.2	-155.0	-412.2	-420.2
Sequence (mismatch)	cggcgcttgtcgcaccagt	62.8		-8.6	-6.0	68.4	-140.9	-372.9	-380.9
Reverse sequence (pm)	actggtgcgacaatcgccg	64.5	5.5	-6.0	-8.6	63.2	-155.0	-412.2	-420.2
Reverse sequence (mm)	actggtgcgacaagcgccg	59.1		-6.0	-8.6	68.4	-138.5	-370.4	-378.4

Figure 2. Input spreadsheet after calculation of thermodynamic values (example). Cells for user input are marked with dark letters on white background. Note the difference in ΔT_m when the detection probe is sited on the antisense strand.

Mismatch	Sequence	Reverse Sequence	ΔG_{37}	TMmm	ΔTm
(perfect match)	cggcgattgtcgcaccagt	actggtgcgacaatcgccg	-27.09	64.5	
c6	cggcgCttgtcgcaccagt	actggtgcgacaaGcgccg	-25.14	62.8	1.7
c10	cggcgattgCcgcaccagt	actggtgcgGcaatcgccg	-24.86	62.2	2.3
c14	cggcgattgtcgcCccagt	actggGgcgacaatcgccg	-24.69	62.0	2.5
c3	cgCcgattgtcgcaccagt	actggtgcgacaatcgGcg	-25.23	61.2	3.3
c8	cggcgatCgtcgcaccagt	actggtgcgacGatcgccg	-24.77	61.0	3.5
c7	cggcgaCtgtcgcaccagt	actggtgcgacaGtcgccg	-24.43	60.7	3.8
c18	cggcgattgtcgcaccaCt	aGtggtgcgacaatcgccg	-24.63	60.6	3.9
t18	cggcgattgtcgcaccaTt	aAtggtgcgacaatcgccg	-24.21	60.5	4.0
g6	cggcgGttgtcgcaccagt	actggtgcgacaaCcgccg	-23.33	60.4	4.1
t3	cgTcgattgtcgcaccagt	actggtgcgacaatcgAcg	-23.78	60.0	4.5
t6	cggcgTttgtcgcaccagt	actggtgcgacaaAcgccg	-24.13	59.9	4.6
c17	cggcgattgtcgcaccCgt	acGggtgcgacaatcgccg	-24.22	59.9	4.7
c9	cggcgattCtcgcaccagt	actggtgcgaGaatcgccg	-23.89	59.6	4.9
g7	cggcgaGtgtcgcaccagt	actggtgcgacaCtcgccg	-23.84	59.4	5.1
a8	cggcgatAgtcgcaccagt	actggtgcgacTatcgccg	-24.08	59.4	5.1
t9	cggcgattTtcgcaccagt	actggtgcgaAaatcgccg	-23.32	59.1	5.4
a3	cgAcgattgtcgcaccagt	actggtgcgacaatcgTcg	-23.52	59.1	5.5
a2	cAgcgattgtcgcaccagt	actggtgcgacaatcgcTg	-23.87	59.0	5.5

Figure 3. Output after calculation of all possible mismatches under a given probe (example; truncated).

data). A total of 14 matched and 33 unmatched hybridization experiments were performed in triplicate. The annealing oligonucleotide concentration is needed for T_m calculation according to [Equation 3]. We have used the final concentration of the hybridization probe in each assay (0.1 and 0.2 µmol/L) for parameterization of the model. If DMSO was present in the PCRs the lowering influence on T_m, which is 0.6°C/% DMSO (9), was considered. The results in Figure 4 show excellent agreement of calculated and measured T_m, for both wild-type experiments and for those with mismatches (n=47; r=0.98, Sy/x=0.90). The T_m of the perfectly matched and the mismatched hybridization was predicted with an accuracy of ±2.4°C (±2 SD). Figure 4 shows a Bland-Altman difference plot (5) of T_m values predicted with the thermodynamic model vs. those values derived from LightCycler experiments plotted against the mean of both.

DISCUSSION

An easy-to-use spreadsheet program is presented that calculates the T_m of oligonucleotide duplexes on the basis of thermodynamic parameters for nearest-neighbor pairs. Recently published values for enthalpy and entropy allow a precise prediction of duplex stability and the T_m with an error of approximately 2°C (1–4,10,13). Our data show that this holds true for hybridization-probe experiments after enzymatic amplification of genomic DNA with an estimated error of 2.4°C, with appropriate parameterization of the thermodynamic model for matched, one-point and two-point mismatched oligonucleotide sequences.

The application of this program for the design of hybridization assays on the LightCycler is of special interest when selecting a probe/anchor pair that gives the best discrimination (difference in T_m between wild-type and mutation). In our experience, a minimal difference of 1.5°C is necessary, and at least 3°C is advisable for the safe detection of heterozygosity. Furthermore, it estimates the discriminating power of such assays for the detection of a mutation of interest and other mutations that may be present in the region covered by the probe by calculating the T_m of all possible mutations.

A more detailed discussion about hybridization-probe assays on the LightCycler and their discrimination power was published recently (14). Because dye labeled probes (ca. $3.80 per assay) are expensive, added to the cost for technician time in setting up a suitable and reliable genotyping assay, the use of this thermodynamic approach before probe design and synthesis may save time and money. The program can be used in almost every scientific laboratory that works with T_m-based molecular biological assays.

From a theoretical point of view, such evidence-based design of probes is also applicable to DNA blotting, microarray experiments or for the estimation of suitable stringency conditions (sodium concentration) that give reliable and reproducible results for respective experiments. How valid the model is for other hybridization assays can not be concluded from our data. However, other assay systems (e.g., TaqMan) should also profit from a probe design that generates a maximum difference between perfectly matched and mismatched sequences. In addition, the "make all mismatches" feature

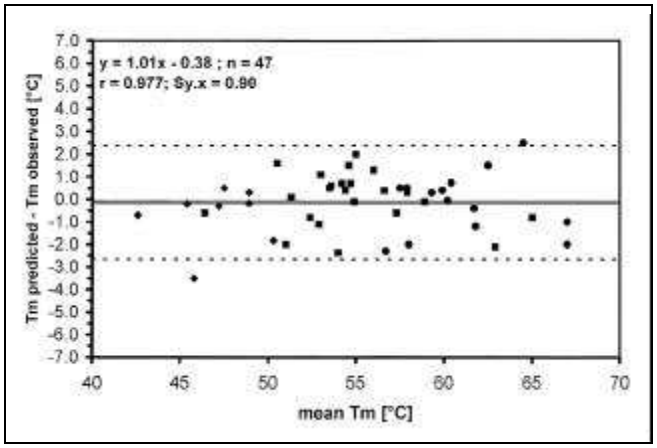

Figure 4. Difference plot for predicted and observed T_m derived from 14 different LightCycler hybridization-probe assays vs. the mean T_m of both. Probe lengths were 17–29 bp, GC content was between 24% and 63%. Melting temperatures from 14 matched (●), 25 different single mismatches (■) and 8 two-point mismatches (◆) evaluated in 14 hybridization-probe assays are shown (r=0.977). The solid line (—) represents the mean difference of predicted and observed T_m (-0.13°C), and (- - - -) shows ±2 SD (2.4°C). Regression estimates are derived from principle component analysis of calculated vs. measured T_m values. Correlation coefficient represents Pearson's r.

may be helpful to calculate how specific a mutation can be detected by a certain T_m shift. We encourage scientists working with hybridization assays to use this program and to report their experience to the scientific community.

The program is available at **HTTP://server1.medikc.med.uni-goettingen.de/meltcalc.htm**, which can also be accessed through a link from the **Software Library** on the *BioTechniques* Web site (**www.biotechniques.com**), or by e-mail upon request from the authors.

Formulas and Equations

The enthalpy (ΔH^0) and entropy (ΔS^0) and free energy ΔG^0 of an oligonucleotide duplex is calculated as the sum of the entire nearest neighbors plus helix initiation and symmetry terms:

$$\Delta H^0_{total} = \Sigma i \Delta H^0\ (i) + \Delta H^0\ (5'init) + \Delta H^0\ (3'init) + \Delta H^0\ (symmetry) \quad \text{[Eq. 1]}$$

For ΔS^0 and ΔG^0, the same formula applies. The ΔH^0, ΔS^0 and ΔG^0 values for nearest neighbors for matched and mismatched base pairs and those for initiation and symmetry correction were taken from published data (1–4,10,12).

Apart from entropy and enthalpy of the nucleotides in the sequence, the concentration of monovalent ions and oligonucleotides must be taken into account as the data basically provided by Allawi and SantaLucia are derived at 1 mol/L NaCl. Because enthalpy is independent of salt concentration, only entropy values for a given [Na] were calculated according to the formula:

$$\Delta S^0\ [Na] = \Delta S^0\ [1M\ Na] + 0.368 \times N \times \ln\ [Na] \quad \text{[Eq. 2]}$$

where N is the total number of phosphates in the duplex divided by two (12) which usually equals the length minus one.

The T_m of self-complementary oligonucleotides at a given concentration CT in a solution with the sodium equivalence concentration [Na] is calculated as:

$$T_m = \Delta H^0/(\Delta S^0\ [Na] + R \times \ln\ CT) \quad \text{[Eq. 3]}$$

where R is the gas constant (1.987 cal/K×mol); for non-self-complementary sequences, CT is the concentration of oligonucleotides divided by 4. If strands are in different concentrations, CT is defined as Ca-Cb/2, where Ca and Cb are the concentrations of either oligonucleotide and Ca > Cb.

Calculation of monovalent cations (6):

$$[Na\ equivalents] = [MgCl_2] \times 100 + [KCl]$$

AKNOWLEDGMENT

We gratefully thank Sandra Hartung and Reiner Andag for their skillful technical assistance and Prof. Dr. V.W. Armstrong for the critical revision of the manuscript.

REFERENCES

1. **Allawi, H.T. and J. SantaLucia, Jr.** 1997. Thermodynamics and NMR of internal GT mismatches in DNA. Biochemistry *36*:10581-10594.
2. **Allawi, H.T. and J. SantaLucia, Jr.** 1998. Nearest-neighbor thermodynamics of internal AC mismatches in DNA: sequence dependence and pH effects. Biochemistry *37*:9435-9444.
3. **Allawi, H.T. and J. SantaLucia, Jr.** 1998. Nearest-neighbor thermodynamic parameters for internal GA mismatches in DNA. Biochemistry *37*:2170-2179.
4. **Allawi, H.T. and J. SantaLucia, Jr.** 1998. Thermodynamics of internal CT mismatches in DNA. Nucleic Acids Res. *26*:2694-2701.
5. **Bland, J.M. and D.G. Altman.** 1986. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet *1*:307-310.
6. **Chester, N. and D.R. Marshak.** 1993. Dimethyl sulfoxide-mediated primer T_m reduction: a method for analyzing the role of renaturation temperature in the polymerase chain reaction. Anal. Biochem. *209*:284-290.
7. **Conner, B.J., A.A. Reyes, C. Morin, K. Itakura, R.L. Teplitz and R.B. Wallace.** 1983. Detection of sickle cell beta S-globin allele by hybridization with synthetic oligonucleotides. Proc. Natl. Acad. Sci. USA *80*:278-282.
8. **Lay, M.J. and C.T. Wittwer.** 1997. Real-time fluorescence genotyping of factor V Leiden during rapid-cycle PCR. Clin. Chem. *43*:2262-2267.
9. **Musielski, H., W. Mann, R. Laue and S. Michel.** 1981. Influence of dimethylsulfoxide on transcription by bacteriophage T3- induced RNA polymerase. Z. Allg. Mikrobiol. *21*:447-456.
10. **Peyret, N., P.A. Seneviratne, H.T. Allawi and J. SantaLucia, Jr.** 1999. Nearest-neighbor thermodynamics and NMR of DNA sequences with internal AA, CC, GG, and TT mismatches. Biochemistry *38*:3468-3477.
11. **Rychlik, W.** 1995. Priming efficiency in PCR. BioTechniques *18*:84-90.
12. **SantaLucia, J., Jr.** 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc. Natl. Acad. Sci. USA *95*:1460-1465.
13. **SantaLucia, J., Jr., H.T. Allawi and P.A. Seneviratne.** 1996. Improved nearest-neighbor parameters for predicting DNA duplex stability. Biochemistry *35*:3555-3562.
14. **von Ahsen, N., M. Oellerich, V.W. Armstrong and E. Schütz.** Application of a thermodynamic nearest-neighbor model to estimate nucleic acid stability and optimize probe design: prediction of melting points of different mutations of apolipoprotein B 3500 and factor V Leiden with a hybridization probe genotyping assay on the LightCycler. Clin. Chem. (In press).
15. **von Ahsen, N., M. Oellerich and E. Schütz.** Using two reporter dyes without interference in a single tube rapid cycle PCR: alpha1-antitrypsin genotyping by multiplex real time fluorescence PCR with the LightCycler. Clin. Chem. (In press).
16. **von Ahsen, N., E. Schütz, V.W. Armstrong and M. Oellerich.** 1999. Rapid detection of prothrombotic mutations of prothrombin (G20210A), factor V (G1691A) and methylenetetrahydrofolate reductase (C677T) by real time fluorescence PCR with the LightCycler. Clin. Chem. *45*:694-696.
17. **Wartell, R.M., S. Hosseini, S. Powell and J. Zhu.** 1998. Detecting single base substitutions, mismatches and bulges in DNA by temperature gradient gel electrophoresis and related methods. J. Chromatogr. A *806*:169-185.
18. **Wetmur, J.G.** 1991. DNA probes: applications of the principles of nucleic acid hybridization. Crit. Rev. Biochem. Mol. Biol. *26*:227-259.

Received 28 July 1999; accepted 12 October 1999.

Address correspondence to:

Dr. Ekkehard Schütz
Dept. Clinical Chemistry
University Hospital
Robert-Koch-Str. 40
37075 Göttingen,
Germany
Internet: eschuetz@med.uni-goettingen.de