Computer Program for Calculating the Melting Temperature of Degenerate Oligonucleotides Used in PCR or Hybridization

BioTechniques 22:1158-1160 (June 1997)

Haoyuan Chen and Guan Zhu[1]
Emory University, Atlanta, GA and [1]New York State Department of Health, Albany, NY, USA

ABSTRACT

Degenerate primers or probes have been widely used in molecular cloning, but the calculation of their melting temperatures could not simply be done using thermodynamic parameters because of degeneracy and the lack of a computer program. We present here a simple computer program named dPrimer for the calculation of melting temperature of degenerate oligonucleotides based on the nearest-neighbor model. The program was written in C++ computer language and implemented in Macintosh® with a Symantec® C++ compiler. The degenerate sequencing data were read into a graph data structure. All possible oligonucleotide sequences were then determined by a depth-first search algorithm. Their melting temperature (T_m) values were individually calculated, and output was given as T_m range, mean and standard deviation. These data could help one in the selection of PCR annealing and hybridization temperatures as well as in the design of degenerate oligonucleotides with a desired range of T_m.

INTRODUCTION

Degenerate primers or probes, usually designed from partially sequenced peptides or conserved regions on the basis of comparison of several proteins, have been widely used in the polymerase chain reaction (PCR), the DNA library screening or Southern blot analysis (9,10). It is necessary to estimate the melting temperature (T_m) of a primer or probe so that one can select reasonable annealing or hybridization temperatures.

The nearest-neighbor model calculates the free energy of the double helix for an oligonucleotide based on the sequence, thus providing an accurate estimation of the T_m of the oligonucleotide (1–3,6,7). Although the algorithm of the model is not complicated, a computer program may be needed in the calculation of T_m of oligonucleotides because of the number of parameters and steps involved in the calculation (6,7). A degenerate primer or probe is actually a number of mixed oligonucleotides. The number of oligonucleotides is determined by the degeneracy, which ranges from at least 2 to several thousand. Since each oligonucleotide has its own T_m, a degenerate primer or probe actually has a range of T_m values. When using the nearest-neighbor model, it is not practical for one to dissect the degeneracy and individually calculate the T_m value of all possible oligonucleotides.

A popular algorithm could easily be applied to calculate T_m of a degenerate oligonucleotide:

$$T_m\ (°C) = 2 \times AT + 4 \times GC \qquad \text{[Eq. 1]}$$

where AT and GC are the numbers of AT and GC base pairs, respectively (8). Based on previous tests with oligonucleotides whose experimental T_m values were known, this does not seem to be an accurate method for either PCR or hybridization (6,7). An empirical way to determine the optimal annealing or probing temperature is to test a series of temperatures in a set of experiments, which is a time-consuming process, especially when several pairs of degenerate primers or several probes are used in one experiment. Therefore, a

convenient computer program for the calculation of T_m values of degenerate primers or probes using the nearest-neighbor model would greatly help researchers in the selection of an annealing or hybridization temperature or in the design of degenerate oligonucleotides with a desired range of T_m values. Hence, we were motivated to design the present computer program to fulfill the need.

MATERIALS AND METHODS

There have been several reported algorithms for the determination of T_m of an oligonucleotide duplex, but the most accurate are based on the nearest-neighbor interaction model (1,3,6,7). If the oligonucleotide is not self-complementary, the T_m can be calculated as:

$$T_m\ (°C) = \Delta H/(\Delta S + R \times \ln(c/4)) - 273.15 + 16.6 \times \log [salt] \quad \text{[Eq. 2]}$$

where ΔH and ΔS are the enthalpy and entropy for the helix formation of an oligonucleotide duplex, respectively; R is the molar gas constant (1.987 cal/°C × mol); c is the total concentration of the annealing oligonucleotide; and [salt] is the potassium ion (K^+) concentration in the PCR or sodium ion (Na^+) in the hybridization of a blot. The effect of magnesium ions on the T_m of PCR primers has not been considered here because of the lack of published thermodynamic data. In a standard protocol, $K^+ = 50$ mM in PCR, while $Na^+ \approx 1.0$ M in hybridization of a blot, in which 6× standard saline citrate (SSC) or sodium chloride sodium phosphate EDTA (SSPE) buffer is usually used for both hybridization and washings (5). We also empirically assigned c = 50 nM or 0.1 nM to Equation 2 for the calculation of T_m values of degenerate oligonucleotides used in PCR or hybridization, which were in good agreement when those oligonucleotides with published thermodynamic data were tested (6,7).

The program, named dPrimer, was written in C++ computer language and implemented in Macintosh® with a Symantec® C++ compiler (Symantec, Cupertino, CA, USA). The degenerate DNA sequencing data were read into a graph data structure. All possible oligonucleotide sequences were then determined by a depth-first search algorithm (4). Individual T_m values of all oligonucleotides were calculated with Equation 2 using the nearest-neighbor thermodynamic data determined by Breslauer et al. (2). The output was given as a range of T_m values, the mean of T_m values and the standard deviation (SD).

RESULTS AND DISCUSSION

The launched dPrimer program provides two small windows for the input of either amino acids or nucleotides. A brief instruction is also given for the use of one-letter symbols. The amino acids can be converted to degenerate oligonucleotides by checking the 'A.A. to DNA' button. The oligonucleotides can also be converted to complementary and reverse sequences by checking the 'complement & reverse'

button. The default maximum degeneracy is set at 5000 but can be adjusted accordingly. The thermodynamic parameters for both PCR and hybridization have been preset in the program, and a user can select one of them by checking the corresponding button. The program will begin the calculation if the 'Process' button has been checked. All results will then be shown in the large window on the right. Although the dPrimer program is designed for degenerate oligonucleotides, the T_m of a regular oligonucleotide can also be calculated with a degeneracy of 1.

The depth-first search algorithm was applied in the program to search for all possible sequences. T_m values of all possible oligonucleotides in a degenerate primer or probe were then calculated, and output was given as a T_m range (minimum and maximum T_m) with the mean T_m values and SD. The length of the degenerate oligonucleotides could easily be adjusted in the input windows for quick calculations if a desired range of T_m is required.

For example, one of the degenerate PCR primers used in the molecular cloning of the β-tubulin gene of *Eimeria tenella* has an amino acid sequence of DNEALY (9). The dPrimer program will convert it into a degenerate primer sequence as GAY-AAY-GAR-GCN-YTN-TA and calculate the degeneracy (256), T_m range (30.81°–51.24°C), mean of T_m (40.89°C) and SD (4.01°C). If this primer is used for hybridization, the range of T_m values, the mean of T_m values and the SD would be 41.78°–61.83°, 51.75° and 4.00°C, respectively. The two sets of data reflect the difference in the stability of the DNA duplex of oligonucleotides, which is caused by the change of concentrations of salt and oligonucleotides (1).

Considering that most degenerate oligonucleotides are designed for PCR amplification, DNA library screenings or Southern blot analysis, we include only thermodynamic data for the formation of a DNA duplex in this program. Therefore, the T_m obtained using the program is only accurate when the degenerate oligonucleotides are annealing or hybridizing to DNA templates.

Although dPrimer will give users a T_m range covering all individual oligonucleotides in a degenerate primer or probe, the T_m of the oligonucleotide that actually forms the DNA duplex is not clear. Previous studies have shown that, besides the T_m of primers, other factors such as the length and T_m of PCR product, should also be considered when determining an optimal annealing temperature for PCR (7). In the hybridization of a membrane blot, the hybridization temperature may be much lower than the T_m of an oligonucleotide. The first two washings are usually done at room temperature, while the final washing is conducted for 5–15 min at the T_m of oligonucleotide (5,6). Degenerate primers have been intensively used in our laboratory where we generally begin with an annealing temperature between the mean and maximum T_m of primers for PCR. In the case of hybridization, temperature for the final washing is usually chosen between the minimum and mean T_m of the probe, which may be increased if higher stringency is required. Using this strategy, we have successfully amplified desired genomic DNA fragments of the β-tubulin gene in *Eimeria tenella* (9) and a P-ATPase gene in *Cryptosporidium parvum* (G. Zhu, unpublished) by PCR with degenerate primers. We have also cloned other types of P-ATPases in *C. parvum* by screening a genomic DNA library with a degenerate probe (G. Zhu, unpublished).

The dPrimer Version 1.0 for Macintosh computer is available upon request.

REFERENCES

1. **Breslauer, K.J.** 1995. Extracting thermodynamic data from equilibrium melting curves for oligonucleotide order-disorder transitions. Methods Enzymol. *259*:221-242.
2. **Breslauer, K.J., R. Frank, H. Blšcker and L.A. Marky.** 1986. Predicting DNA duplex stability from the base sequence. Proc. Natl. Acad. Sci. USA *83*:3746-3750.
3. **Freier, S.M., R Kierzek, J.A. Jaeger, N. Sugimoto, M.H. Caruthers, T. Neilson and D.H. Turner.** 1986. Improved free-energy parameters for predictions of RNA duplex stability. Proc. Natl. Acad. Sci. USA *83*:9373-9377.
4. **Horowitz, E., S. Sahni and D. Mehta.** 1995. Fundamentals of Data Structures in C++, p. 345-346. Computer Science Press, New York.
5. **Rychlik, W., L.L. Domier, P.R. Gardner, G.M. Hellmann and R.E. Rhoads.** 1987. Amino acid sequence of the mRNA cap-binding protein from human tissues. Proc. Natl. Acad. Sci. USA *84*:945-949.
6. **Rychlik, W. and R.E. Rhoads.** 1989. A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in vitro amplification of DNA. Nucleic Acids Res. *17*:8543-8551.
7. **Rychlik, W., W.J. Spencer and R.E. Rhoads.** 1990. Optimization of the annealing temperature for DNA amplification in vitro. Nucleic Acids Res. *18*:6409-6412.
8. **Suggs, S.V., T. Hirose, E.H. Miyake, M.J. Kawashima, K.I. Johnson and R.B. Wallace.** 1981. Use of synthetic oligodeoxyribonucleotides for the isolation of specific cloned DNA sequences, p. 683-693. *In* D.D. Brown (Ed.), ICN-UCLA Symposium: Development Biology Using Purified Genes. Vol. 23. Academic Press Inc., New York.
9. **Zhu, G. and J.S. Keithly.** 1996. The beta tubulin gene of *Eimeria tenella*. Mol. Biochem. Parasitol. *76*:315-319.
10. **Zhu, Y.Y., M.J. Fillenwarth, D. Crabb, L. Lumeng and R.C. Lin.** 1996. Identification of the 37-kD rat liver protein that forms an acetaldehyde adduct in vivo as delta(4)-3-ketosteroid 5-beta-reductase. Hepatology *23*:115-122.

Received 20 February 1997; accepted 9 April 1997.

Address correspondence to:

Guan Zhu, PhD
Wadsworth Center, DAI, Rm 3063
New York State Department of Health
PO Box 22002
Albany, NY 12201-2002, USA
Inrernet:zhug@wadsworth.org