Sequence analysis

Enhancements and modifications of primer design program Primer3

Triinu Koressaar and Maido Remm*

Department of Bioinformatics, Institute of Molecular and Cell Biology, University of Tartu, Riia str. 23, Tartu 51010, Estonia

ABSTRACT

Summary: The determination of annealing temperature is a critical step in PCR design. This parameter is typically derived from the melting temperature of the PCR primers, so for successful PCR work it is important to determine the melting temperature of primer accurately. We introduced several enhancements in the widely used primer design program Primer3. The improvements include a formula for calculating melting temperature and a salt correction formula. Also, the new version can take into account the effects of divalent cations, which are included in most PCR buffers. Another modification enables using lowercase masked template sequences for primer design.

Availability: Features described in this article have been implemented into the development code of Primer3 and will be available in future versions (version 1.1 and newer) of Primer3. Also, a modified version is compiled under the name of mPrimer3 which is distributed independently. The web-based version of mPrimer3 is available at http://bioinfo.ebc.ee/mprimer3/ and the binary code is freely downloadable from the URL http://bioinfo.ebc.ee/download/.

Contact: maido.remm@ut.ee

1 INTRODUCTION

Primer3 (Rozen *et al.*, 2000) has been widely used for PCR primer design for more than a decade. The original version uses a table of thermodynamic parameters published by the Breslauer group (Breslauer *et al.*, 1986) and a formula for melting temperature (Tm) calculations published by Rychlik and co-workers (Rychlik *et al.*, 1990). The outdated table of thermodynamic parameters and method for calculating Tm are considered the main drawback of Primer3 (Chavali *et al.*, 2005). Improved sets of thermodynamic parameters and improved formulae are now available (Owczarzy *et al.*, 1998; Panjkovich *et al.*, 2005; SantaLucia, 1998; SantaLucia *et al.*, 2004). An important influence on primer melting temperature is the concentration of mono- and divalent cations in the solution. As nearest-neighbor parameters are measured in a specific salt concentration (e.g. 1 M NaCl), the melting temperature

should be corrected for the actual conditions of the PCR buffer. The original Primer3 release uses a salt correction formula published by Schildkraut and Lifson (Schildkraut and Lifson, 1965), which does not consider the effect of divalent cations on melting temperature calculations. Many other widely used general primer design programs like OLIGO6 (http://www.oligo.net) and PRIDE (Haas *et al.*, 1998) similarly use outdated formulas for Tm calculation. Many other primer design software exist which already use up-to-date table of thermodynamic parameters for melting temperature calculation, e.g. PrimerPremier (http://www.premierbiosoft.com/), FastPCR (http://www.biocenter.helsinki.fi/bi/Programs/fastpcr.htm) or ORFprimer (http://www.proteinstrukturfabrik.de/ORFprimer). However, from the set of programs with up-to-date thermodynamic parameters only few of them (FastPCR, ORFprimer) allow performing automatic command-line calculations for large datasets similarly to Primer3. Another problem with the current version Primer3 appears if one needs to design PCR primers to masked sequences. Primer3 (release 1.0) is optimized for use with the masking program RepeatMasker (Smith A.F.A, http://www.repeatmasker.org/), which masks the sequence with a series of 'N' characters in repeat regions if used with default parameters. By default, Primer3 excludes any primer candidates that contain 'N' characters, which is a reasonable way of avoiding most common repeats. However, other masking methods may mask only short stretches of the template DNA sequence. For example, DUST (Morgulis *et al.*, 2006), Windowmasker (Morgulis *et al.*, 2006) and GenomeMasker (Andreson *et al.*, 2006a) mask much shorter regions and occasionally only one nucleotide. Furthermore, some of these short masked features are deleterious only if they happen to overlap the 3′ end of the primer and can be tolerated when they overlap the 5′ end. If the target sequence is masked with 'N' characters, the design of primers comprising 'N' characters is typically forbidden. As far as we know, none of the existing primer design programs supports the usage of soft-masked (lowercase-masked) target sequences. Therefore, we have implemented a new feature into mPrimer3 that allows primers to be designed from lowercase masked sequences. Lowercase masking preserves the DNA sequence and allows primers to be designed that partly overlap the masked region.
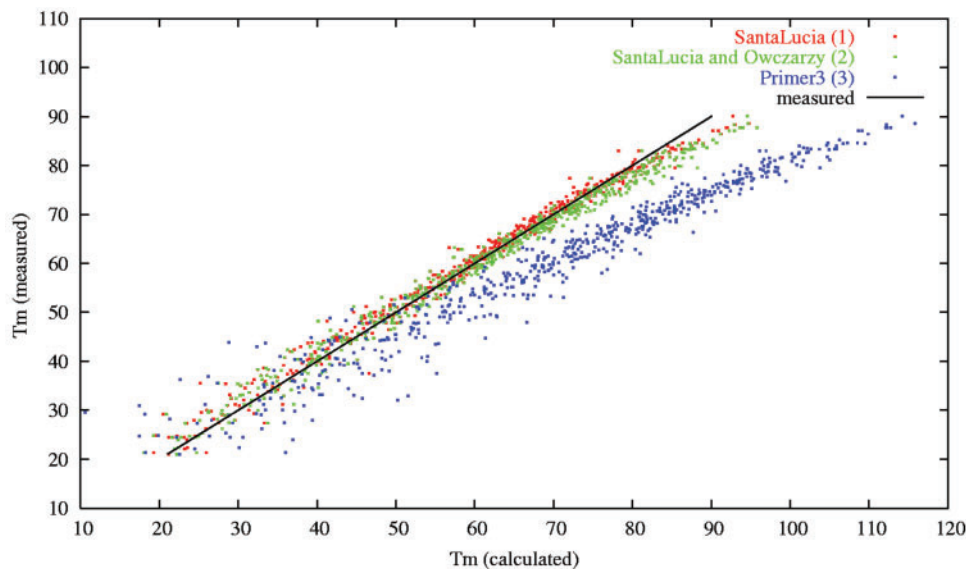
*To whom correspondence should be addressed.

Fig. 1. The correlation between predicted and measured melting temperatures for three different T_m calculation methods. Methods (1) and (2) were implemented into the modified version of Primer3. Method (3) corresponds to the original version of Primer3. For primers of typical length (15–30 nucleotides) the average differences between the experimental and predicted T_m were 1.37, 1.78 and 11.70°C for methods (1), (2) and (3) respectively. The experimental melting temperature data used in this analysis were retrieved from the literature (Owczarzy et al., 2004) and include 590 different measurements with 146 different oligonucleotides.

2   MODIFICATIONS OF THE ALGORITHM

The values of the nearest-neighbor thermodynamic parameters and the method for calculating melting temperature in the modified Primer3 are derived from existing publications of unified parameters (SantaLucia, 1998; tables 1, 2). The value of ΔS° is calculated by a method similar to that used for ΔG°(SantaLucia, 1998; equation 1):

$$S^o(total) = \sum n_i\ S^o(i) + S^o(initw/termG \cdot C) + S^o(initw/termA \cdot T) + S^o(sym) \quad (1)$$

where $\Delta S^o(i)$ are the standard entropy changes for the 10 possible Watson–Crick nearest-neighbors; n_i is the number of occurrences of each nearest neighbor pair i; $\Delta S^o(initw/termG{\cdot}C)$ and $\Delta S^o(initw/termA{\cdot}T)$ are two initiation parameters—respectively, initiation with terminal $G{\cdot}C$ and with terminal A/T; and $\Delta S^o(sym)$ is the penalty for self-complementary duplexes. Compared to the original version of the Primer3, this formula takes into account the initiation parameter for duplex formation with terminal A/T and G/C nucleotides and a symmetry correction parameter for duplex self-complementarity. The formula for calculating ΔH° is also analogous to equation 1.

Two different salt correction formulae for melting temperature calculations are implemented in modified version of Primer3. The first is the sequence-independent equation suggested by SantaLucia (SantaLucia, 1998; equation 8):

$$S^o(oligomer, [Na^+]) = S^o(1\ MNacl) + 0.368N \ln[Na^+] \quad (2)$$

where N is the total number of phosphates in the duplex and $[Na^+]$ is the total concentration of monovalent cations. ΔH° is assumed to be independent of salt concentration

over the range used in PCR work. The second is the sequence-dependent equation suggested by Owczarzy and coworkers (Owczarzy et al., 2004; equation 22):

$$\frac{1}{T_m(2)} = \frac{1}{T_m(1)} + [4.29 \times f(GC) - 3.95] \times 10^{-5} \ln \frac{[Na^+]_2}{[Na^+]_1}$$
$$+ 9.40 \times 10^{-6}(\ln^2[Na^+]_2 - \ln^2[Na^+]_1) \quad (3)$$

where $f(GC)$ is the fraction of GC base pairs in the duplex and $[Na^+]$ is the total concentration of monovalent cations. In addition, we implemented the possibility of taking account the divalent cation concentration according to the formula described by von Ahsen et al. (von Ahsen et al., 2001). Users can define the desired melting temperature calculation method by executing modified Primer3 with appropriate command line arguments or input tags in initiation file. We compared experimentally measured melting temperatures with T_m values predicted by original Primer3 and Tm values predicted by two different formulae in modified Primer3, the results are shown in Figure 1.

As explained in the introduction, there is sometimes a need to design primers overlapping the masked regions. We have added a new feature to Primer3 allowing primers overlapping lowercase-masked regions to be designed. A novel feature of the modified Primer3 is that primers with a lowercase nucleotide at the 3′ end can be rejected. This behavior relies on the assumption that masked features (e.g. repeats) can partly overlap the primer, but they cannot overlap its 3′ end. Lowercase letters in other positions are accepted, assuming that the masked features do not influence primer performance if they do not overlap the 3′ end.

ACKNOWLEDGEMENTS

REFERENCES

Ahsen von,N. *et al.* (2001) Oligonucleotide melting temperatures under PCR conditions: nearest-neighbor corrections for Mg(2+), deoxynucleotide triphosphate, and dimethyl sulfoxide concentrations with comparison to alternative empirical formulas. *Clin. Chem.*, **47**, 1956–1961.

Andreson,R. *et al.* (2006) GENOMEMASKER package for designing unique genomic PCR primers. *BMC Bioinformatics*, **7**, 172.

Breslauer,KJ. *et al.* (1986) Predicting DNA duplex stability from the base sequence. *Proc. Natl Acad. Sci.*, **83**, 4746–4750.

Chavali,S. *et al*. (2005) Oligonucleotide properties determination and primer designing: a critical examination of predictions. *Bioinformatics*, **21**, 3918–3925.

Haas,S. *et al.* (1998) Primer design for large scale sequencing. *Nucleic Acids Res.*, **26**, 3006–3012.

Morgulis,A. *et al.* (2006) Window Masker: window-based masker for sequenced genomes. *Bioinformatics*, **22**, 134–141.

Owczarzy,R. *et al.* (1998) Predicting Sequence-Dependent Melting Stability of Short Duplex DNA oligomers. *Biopolymers*, **44**, 217–239.

Owczarzy,R. *et al.* (2004) Effects of Sodium Ions on DNA Duplex Oligomers: Improved Predictions of Melting Temperatures. *Biochemistry*, **43**, 3537–3554.

Panjkovich,A. and Melo,F. (2005) Comparision of different melting temperature calculation methods for short DNA sequences. *Bioinformatics*, **21**, 711–722.

Rozen,S. and Skaletsky,H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.*, **132**, 365–386.

Rychlik,W. (1990) Optimization of the annealing temperature for DNA amplification *in vitro*. *Nucleic Acids Res.*, **118**, 6409–6412.

SantaLucia,JR. (1998) . A unified view of polymer, dumbbell and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl Acad. Sci.*, **95**, 1460–1465.

SantaLucia,JR. and Hicks,D. (2004) The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.*, **33**, 415–440.

Schildkraut,C. and Lifson,S. (1965) Dependence of the melting temperature of DNA on salt concentration. *Biopolymers*, **3**, 195–208.