BIOINFORMATICS ORIGINAL PAPER Vol. 21 no. 20 2005, pages 3918–3925
doi:10.1093/bioinformatics/bti633

Data and text mining

Oligonucleotide properties determination and primer designing: a critical examination of predictions

Sreenivas Chavali[1], Anubha Mahajan[1], Rubina Tabassum[1], Souvik Maiti[2] and Dwaipayan Bharadwaj[1,2,*]

[1]Functional Genomics Unit, and [2]Proteomics and Structural Biology Unit, Institute of Genomics and Integrative Biology, CSIR, Mall Road, Delhi 110 007, India

Received on April 20, 2005; revised on August 14, 2005; accepted on August 15, 2005
Advance Access publication August 16, 2005

ABSTRACT

Motivation: Precise prediction of melting temperature (T_m), secondary structures and design of oligonucleotides determine the efficiency and success of experimentation in molecular biology. Availability of a plethora of software and the users unawareness about their limitations compromises the accuracy and reliability of the predictions.

Results: Comparative analysis of 56 modules was done for T_m prediction using a large set of oligonucleotide sequences spanning the whole range of GC-content and length. Allawi module of the calculator 'MELTING', Nearest Neighbor (NN) of oligo calculator (McLab), NN of T_m Calculation for Oligos (Biomath Calculator, Promega) and HYTHER provided the most precise T_m predictions. A model has also been proposed to calculate the optimum annealing temperature integrating the already reported formulations. Secondary structure predictions of oligonucleotides reveal a large number of structures in contrast to the experimental observations. Of the 11 primer designing tools evaluated, Primer 3 and WebPrimer performed the best for the AT-rich templates, Exon Primer for AT = GC templates, and Primer Design Assistant, Primer3 and Primer Quest for GC-rich templates. This study provides optimal choice for application to the user, increasing the success of a variety of experimentations, especially those that have high-throughput and complex assay designs.

Contact: db@igib.res.in

Supplementary Information: The details of the oligonucleotides and of the different modules of T_m prediction considered for the study are provided as Supplementary Information, available at Bioinformatics online.

INTRODUCTION

Experimentation in molecular biology invariably revolves around oligonucleotides, a stretch of deoxynucleotides. Some of the techniques that employ oligonucleotides as their prime component are polymerase chain reaction (PCR), hybridization, Southern blotting, sequencing, etc. Perhaps the most universal method in use today is the PCR, with applications including amplification, cloning, mutation detection, mutagenesis and restriction fragment length polymorphism studies. The principal oligonucleotide characteristics that determine the efficiency and accuracy of these experiments are melting temperature (T_m), secondary structures and design of the primer. Choice of a non-optimal sequence or temperature can lead to amplification or detection of wrong sequences. These characteristic features of oligonucleotides gain much more importance when we go for high-throughput design and for complicated assays, such as DNA microarray and multiplex PCR, that have many interacting oligonucleotides, as the cost incurred in each reaction is enormous.

Several programs are available commercially as well as freely on the World Wide Web to determine the properties of oligonucleotides and to design primers. A large number of computer software based on different methods, considering a wide variety of parameters, are available to theoretically estimate the T_m of oligonucleotides; the number is ever increasing. The methodologies adopted in the software can be broadly classified into Basic, which considers %G+C content for T_m calculation (Marmur and Doty, 1962) (Wallace *et al*., 1979), Salt Adjusted (SA), which takes into account the salt concentration (usually concentration of Na^+) (Howley *et al*., 1979) and Nearest Neighbor (NN) modules, which calculate T_m using the known DNA NN parameters, considering the salt and oligonucleotide concentrations (Breslauer *et al*., 1986; Freier *et al*., 1986; Sugimoto *et al*., 1995, 1996; SantaLucia *et al*., 1996; Allawi and Santa Lucia, 1997; SantaLucia, 1998; Xia *et al*., 1998). T_m estimation by different software results in huge and significant differences, forcing the users to doubt the reliability of the estimations. Besides this, most users seem to be unaware of the limitations of the software they use and the method it employs, further compromising on accuracy and reliability.

Another parameter of prime importance that is considered for the qualitative assessment of an oligonucleotide is the formation of secondary structures, which includes self-dimers, hairpins and cross dimers, with other interacting oligonucleotides. This characteristic of nucleic acids has been well documented to affect oligonucleotide binding (Gamper *et al*., 1987). Secondary structure prediction for any given sequence relies on energy minimization algorithms and the probability of correct predictions is quite poor because of the limitations of the mathematical model and the uncertainties in the thermodynamic parameters used in these methods (Dong *et al*., 2001). Secondary structure of the target DNA to which the oligonucleotide binds is also a matter of concern (Fedorova *et al*., 1992) and needs due attention.

*To whom correspondence should be addressed.

Primer design is fundamentally important for all PCR-based studies and detection methods. Though a great deal of research has been undertaken over the years in designing primers, complete comprehension of the various factors that affect the amplification and yield of the product is still lacking. Hence, it is always desirous to consider all the known parameters (Robertson and Walsh-Weller, 1998) and have software that adopts this wholistic approach to provide us with the best possible primer pair for target DNA amplification. A considerable amount of failure of amplifications can be attributed to faulty design of primers, necessitating the identification of the best primer designing software among those available.

This study is divided into three sections: first, identification of the best oligonucleotide properties calculator that would predict T_m with least deviation, which could be used in calculating the optimum annealing temperature for PCR amplifications; second, evaluation of the secondary structure predictions; and third, testing the efficiency of primer designing software and identifying the best one.

METHODS

Thermal melting studies

Oligonucleotides used were either obtained from Sigma Genosys or synthesized in-house (at The Centre for Genomic Application) on the Applied Biosystems 3900 high-throughput DNA synthesizer using standard Phosphoramidite chemistry (Caruthers, 1982). Strands were deprotected, purified, desalted and vacuum dried on a speed vac concentrator and stored at −20°C. Strand concentrations were determined using extinction coefficients at 260 nm, estimated by the NN method (Cantor *et al*., 1970). Sample purity and fidelity were confirmed by the MALDI-TOF mass spectrometric analysis on Sequenom (Sequenom Inc., San Diego, CA). Oligonucleotides were designed and synthesized in such a way that six were rich in AT (GC-content <45%), six had equal proportions of AT and GC (GC-content 45–55%), while six were rich in GC (GC-content >55%).

Optical melting experiments were performed on Cary 400 BIO (Varian Inc., Palo Alto, CA) UV-visible spectrophotometer to determine the melting temperature of these oligonucleotides. Samples were rehydrated in melting buffer (10 mM sodium phosphate, pH 8.3 containing 20 mM NaCl, resulting in a total $[Na^+]$ of 50 mM. Absorbance versus temperature profile was measured at 260 nm with a heating rate of 1°C. For each sample, at least three complete transition curves were obtained. The water condensation on the cuvette exterior at a low temperature range was avoided by flushing with a constant stream of dry N_2 gas. T_m values were calculated by using the first derivatives of the melting curves. Out of the 90 other oligonucleotides considered for the study, 30 were obtained from the work of Owczarzy *et al*. (1997) and 60 from the report of Owczarzy *et al*. (2004). In the latter case T_m for each oligonucleotide has been determined at different Na^+ concentrations ($[Na^+]$), of 69 mM, 119 mM, 220 mM, 621 mM and 1020 mM.

T_m predictions

After thorough searching of the World Wide Web, a total of 25 freely available oligonucleotide properties calculators were selected. T_m prediction was also performed using DNASTAR (available with us) and T_m calc module of the GeneAmp-PCR System 9700 (Applied Biosystems, Foster City, CA). T_m was predicted for a total of 108 oligonucleotides, which had lengths ranging from 17 to 30 nt and a %GC-content ranging from 8 to 80.

Statistical analysis

All the oligonucleotides were grouped into three categories—AT-rich (GC <45%), AT≈GC (GC-content ranging from 45 to 55%) and GC-rich (GC >55%). For each oligonucleotide the difference between the experimentally determined T_m and the predicted T_m was calculated. Least square estimates were computed for predictions of each category of oligonucleotides of each calculator. Root mean square deviations (rmsd) were computed to determine the precision of the predictions.

The calculators that provide NN T_m predictions with a window to change salt and oligonucleotide concentrations were selected. Predictions were performed taking into account the salt and oligonucleotide concentrations at which melting temperatures were experimentally determined. For these calculators, average deviation in T_m prediction was calculated. To determine the significance of differences among the deviations of these calculators with respect to the calculator providing the least deviation, *t*-test was performed and *P*-value was calculated.

Calculation of optimal annealing temperature

Using our empirical dataset of 41 amplicons, we predicted the optimal annealing temperature (T_a^{opt}) of PCR. Our model was a modification of the formulation developed by Rychlik *et al*. (1990)

$$T_a^{opt} = m(T_{m_{primer}}) + n(T_{m_{product}}), \quad (1)$$

in which $T_{m_{primer}}$ is the T_m of the less stable primer–template pair, which is the most precise prediction possible from the available calculators at a monovalent cation concentration given at $[K^+_{eq}]$ and respective oligo concentrations ([oligo]).

$[K^+_{eq}]$ is the K^+ equivalents that are calculated from the approximation derived for $[Na^+_{eq}]$ by von Ahsen *et al*. (2001), since it is well known that K^+ is similar to Na^+ in regard to duplex stabilization (Nakano *et al*., 1999). This equation takes into account the effect of $[Mg^{2+}]$ and the deoxynucleotide triphosphates concentration ([dNTPs]).

$$[K^+_{eq}] = [K^+] + 120(\sqrt{[Mg^{2+}] - [dNTPs]}). \quad (2)$$

Note that the concentrations in Equation (2) are in mmol/l.

$T_{m_{product}}$ is the T_m of the product determined using NN thermodynamics at $[K^+_{eq}]$ and respective template concentrations used. Application of NN thermodynamic model for the prediction of T_m of polymers has been well established (Santa Lucia, 1998). The parameters *m* and *n* in equation (1) were optimized to minimize the prediction error by stepwise incremental iterations using sum of squares. Optimization of no parameters in Equation (2) was attempted since their precision were thoroughly discussed previously (von Ahsen *et al*., 2001).

Secondary structure studies

Electrophoresis of 18 (9 sense and 9 antisense, complementary) 5′ labeled oligonucleotides (20 pmol) was carried out as described in Yoshizawa *et al*. (1997) on 20% polyacrylamide gel containing 7 M urea in TBE (pH 8.3) at 4°C. This was then analyzed on an autoradiogram. Concentration-dependent single-strand melting studies were performed to determine whether the secondary structures observed were self-dimers or hairpins. The number of secondary structures each oligonucleotide can form was predicted using six different calculators.

Primer designing studies

Nine regions (1 kb each) were selected from the human genome. Of these templates three were AT-rich (from AMPK, NKX6.1, F3 genes), three had equal proportions of AT and GC (from PKLR, NKX6.1, ISL1 genes), and the three were GC-rich (from VEGF, IPF1, INS). Repeat Masker analysis ensured that these regions were free of repeats. Primers were designed to amplify these regions, using 10 freely available primer-designing tools and 'Primer Select' module of DNASTAR. The best primer pair designate from each tool was synthesized and PCR optimization was attempted. Initially, primers were designed, synthesized and optimized only for one template of

each category. After initial analysis, primers for two more templates of each kind were analyzed.

RESULTS

T_m predictions

Melting temperature for all the 108 oligonucleotides (Supplementary Table 1) considered for the study was predicted using 25 oligonucleotide properties calculators. As mentioned earlier certain calculators had more than one module, such as Basic, SA and NN. Each module adopts different methodology for the calculation of T_m. The modules that lacked the specifications of the methodology adopted were categorized under 'miscellaneous' group. Modules that calculated Basic T_m did not consider salt and oligonucleotide concentrations. In the modules that calculate SA T_m, the default value of $[Na^+]$ is 50 mM and for the NN T_m calculation the default values of $[Na^+]$ and [oligo] are 50 mM and 250 pM, respectively. In the modules that had a window to change parameters, the values of $[Na^+]$ and [oligo] at which the T_m was experimentally determined were provided.

In the first phase of the analysis, T_m was predicted for all the oligonucleotides using all the 56 modules, (Supplementary Table 2), providing the applicable parameters. For the 60 oligonucleotides for which the T_m has been determined at five different $[Na^+]$ (Owczarzy *et al*., 2004), in this phase the T_m determined at 69 mM alone was used. Least square estimates and rmsd values were computed for each prediction. Equations that compute T_m using %G+C-content performed badly for oligonucleotide duplexes; particularly, since these equations do not account for bimolecular initiation and the effect of strand concentration, resulting in high-rmsd values. As a result of this analysis 25 modules, which had predictions with rmsd values $\leqslant 5$, were selected. The GC-content of the oligonucleotides had no effect on the precision of predictions.

T_m is not just a property of composition of an oligonucleotide but a property of an oligonucleotide under specific conditions and at a given concentration. Hence, in the second phase of the analysis T_m for 60 oligonucleotides at varying $[Na^+]$ concentrations (Owczarzy *et al*., 2004) was calculated using 17 out of the 25 modules (Fig. 1) that had a window to change parameters. The selected oligonucleotides were grouped into AT-rich (25), AT≈GC (13) and GC-rich (22) in accordance with the described criterion. Least square and rmsd values were computed for each prediction. All the modules of NN type adopt the doublet format of DNA sequence-dependent thermodynamic stability prediction, wherein they consider the entire NN interactions (H bonding and stacking) in a single parameter (Owczarzy, *et al*., 1997).

The SA modules could not perform better than the NN modules. Even in the NN modules, those that adopted outdated thermodynamic parameters (Breslauer *et al*., 1986) gave a huge average T_m deviation of the range 7°C, which is well in accordance with earlier reports (SantaLucia and Hicks, 2004). Allawi module of the calculator MELTING (Le Novere, 2001) using default salt corrections (Santa Lucia, 1998) (http://bioweb.pasteur.fr/seqanal/interfaces/melting.html), NN of oligo calculator (McLab) (http://tool.mclab.com/toolbox/oligo_calculator.jsp) and NN of T_m Calculation for Oligos (Biomath Calculator; Promega) (http://www.promega.com/biomath/default.htm) had very less average deviation in T_m prediction across different $[Na^+]$ and GC-content (Fig. 1). Statistical analysis (Table 1) reveals that the difference between their average deviations at different $[Na^+]$ was not significant ($P>0.05$). This emphasizes that these calculators perform equally well. HYTHER (http://ozone2.chem.wayne.edu/) performed well at all applicable $[Na^+]$ above 69 mM, and Oligo analyzer 3.0 (IDT Biotools) at 1020 mM $[Na^+]$. The average deviations in T_m predictions of HYTHER and Oligo analyzer 3.0 (IDT Biotools) (http://scitools.idtdna.com/scitools/Applications/OligoAnalyzer/Default.aspx) when considered alone (Fig. 1) appear to be less at other concentrations of $[Na^+]$, but when compared with other calculators (Table 1) their predictions deviate significantly ($P<0.05$). However, given the uncertainties of up to 2°C on experimental determinations of melting temperature the statistically significant deviations observed in the predictions of HYTHER at low salt concentrations ($P = 0.022$) may be insignificant from the experimental point of view. Melting Temperature Calculator-Consensus (now called dnaMATE 1.0) (http://dna.bio.puc.cl/cardex/servers/dnaMATE/index.html) has been recently developed, but it does not appear to perform any better than the existing ones. We note that 50.92% of the oligos we have used for our analysis lie in the Zone 2, 25.93% in the Zone 1, 23.15% in the Zone 0 and none in Zone 3 of Panjkovich plots. Briefly, Zone 1 corresponds to similarity in T_m predictions between Breslauer *et al*. (1986) and Sugimoto *et al*. (1996); Zone 2 to similarity between Santalucia *et al*. (1996) and Sugimoto *et al*. (1996); Zone 3 between all the three and in Zone 0 none of the NN parameter sets shared similar T_m values (Panjkovich and Melo, 2005). This positioning of the oligos in the Panjkovich plots may provide insight in the observed differences of predictions.

Apart from T_m, free energy of hybridization and fraction of the oligomers hybridized are also of high-predictive value in oligo designing. Of the calculators that performed well, Allawi module of MELTING, NN of T_m Calculation for Oligos (Biomath Calculator; Promega) and HYTHER predict free energy.

Calculation of T_a^{opt}

Optimal annealing temperature was predicted empirically for a set of 41 amplicons. For this, the $T_{m_{primer}}$ was established from the predictions of the Allawi module of MELTING and $T_{m_{product}}$ from NN of oligo calculator (McLab), as Allawi module of MELTING cannot take large sequences as input. With $m = 0.34$ and $n = 0.46$ in Equation (1) we were able to obtain predicted values for all these amplicons with an average deviation of 1°C.

$$T_a^{opt} = 0.34(T_{m_{primer}}) + 0.46(T_{m_{product}}). \quad (3)$$

The empirical formulations developed by von Ahsen *et al*. (2001) for T_m evaluation aim at NN corrections for parameters, including Mg^{2+}, dNTPs and dimethyl sulfoxide (DMSO). The following equation has been recommended as a tradeoff between accuracy and ease of use.

$$T_m = 77.1°C + 11.7 \times \log[Na^+] + 0.41(\%GC) - 528/n \quad (4)$$

The factor that limits the extensive applicability of this equation is the T_m offset that has been empirically derived. Our attempt overcomes this limitation and allows the sequence and base composition dependent prediction of T_m of the least stable primer and predicts the T_m of the product based on NN thermodynamics considering the NN corrections for Mg^{2+} and dNTPs. The rationale

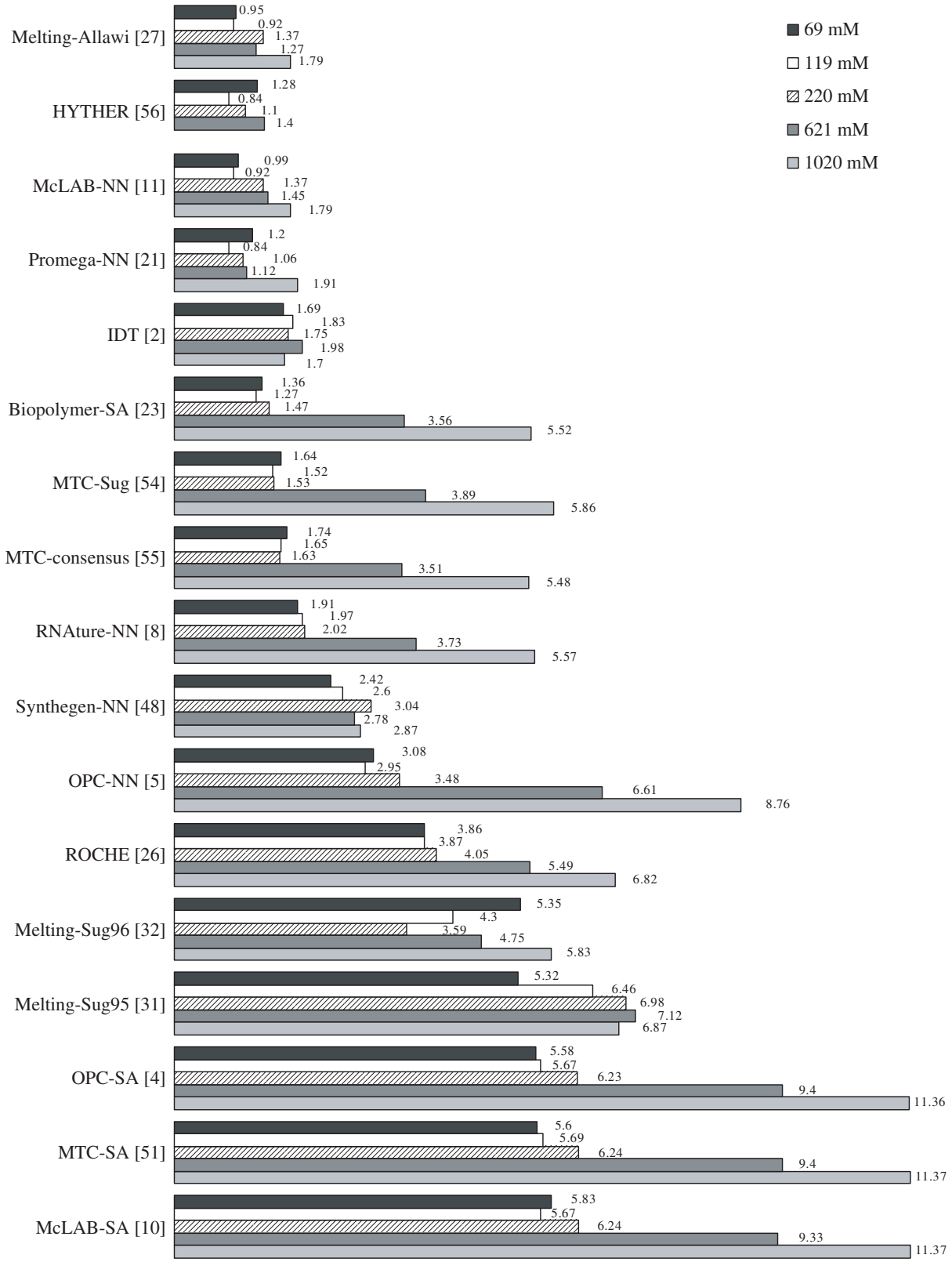


Fig. 1. Comparison of different oligonucleotide properties calculators with window to change salt and oligo concentrations. Predictions at five different [Na^+] concentrations are compared. The value given against each bar provides the average T_m deviation (°C) in predictions of each calculator at the specified [Na^+] concentration. The number indicated in the square brackets corresponds to the serial no. of the module in the Supplementary Table 2.

Table 1. Statistical analysis of average deviations in T_m predictions

Modules	Type	*P*-values for average T_m prediction at different [Na^+]				
		69 mM	119 mM	220 mM	621 mM	1020 mM
Oligo analyzer 3.0 (IDT Biotools)	NN	<0.0001	<0.0001	<0.0001	<0.0001	LD
Oligonucleotide properties calculator	SA	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Oligonucleotide properties calculator	NN	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Oligonucleotide analyzer (RNAture)	NN	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Oligo calculator (McLab)	SA	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Oligo calculator (McLab)	NN	0.795	0.475	0.055	0.241	0.667
T_m calculation for oligos (Biomath calculator; Promega)	NN	0.073	LD	LD	LD	0.295
Biopolymer calculator (Schepartz lab)	SA	0.011	0.0054	0.102	<0.0001	<0.0001
T_m calc (Roche)	Misc	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
MELTING-Allawi	NN	LD	0.472	0.0524	0.405	0.661
MELTING-Sugimoto 95	NN	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
MELTING-Sugimoto 96	NN	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Sequence analyzer (Synthegen)	NN	<0.0001	<0.0001	<0.0001	<0.0001	0.0004
Melting temperature calculator	SA	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Melting temperature calculator; Sugimoto	NN	<0.0001	<0.0001	0.0065	<0.0001	<0.0001
Melting temperature calculator; Consensus	NN	<0.0001	<0.0001	0.007	<0.0001	<0.0001
Hybridization thermodynamics (HYTHER)	NN	0.022	1.0	0.811	0.185	Option not available

Comparative analysis of the average deviations in T_m predictions of different calculators with that of the calculator that gave the least deviation. *P*-value <0.05 signifies that the difference in predictions are not merely out of chance alone. NN, nearest neighbor; SA, salt adjusted; Misc, miscellaneous (modules in which details of the methodology applied for T_m calculation is not provided); LD, least deviation.

Table 2. Secondary structure predictions of different tools

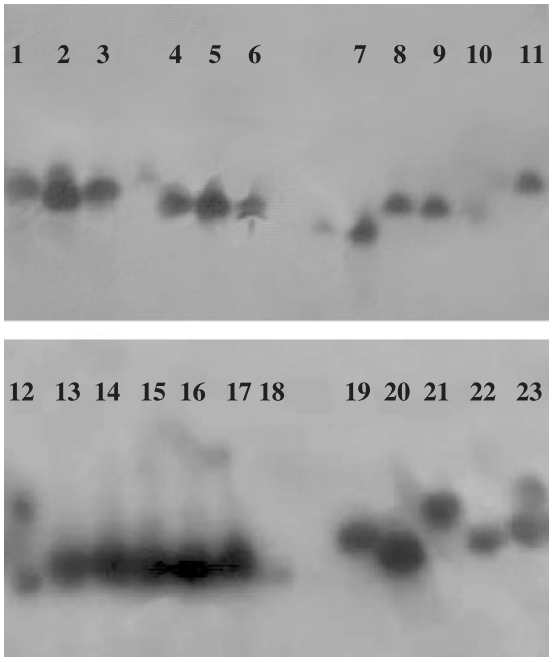
S.no.	Oligonucleotide sequence	MFOLD	Oligo Analyzer 3.0 (IDT Biotools)		Oligonucleotide Properties Calculator		Net Primer (Premier Biosoft)		Oligonucleotide Analyzer (RNAture)		Primer Select (DNA STAR)	
		HP	SD	HP	SD	HP	SD	HP	SD	HP	SD	HP
1	GATTTGGCTGTGATTAGCCC	2	8	2	0	0	1	1	1	1	6	3
2	GGGCTAATCACAGCCAAATC	1	8	1	0	1	1	1	1	1	5	3
3	CTCCGGGGGCCACACTCACGC	1	8	1	0	0	2	0	1	1	4	4
4	GCGTGAGTGTGGCCCCCGGAG	6	8	1	0	0	2	0	1	1	8	4
5	GGCCAAGGCTGGGGTTGAAGG	2	8	2	0	0	1	1	1	1	4	7
6	CCTTCAACCCCAGCCTTGGCC	2	8	2	0	0	4	2	1	1	7	7
7	AAAACAAAGACTTTCTTAAGAGAT	2	8	2	2	1	4	0	1	1	6	15
8	ATCTCTTAAGAAAGTCTTTGTTTT	5	8	2	2	0	5	2	1	1	10	15
9	CATATGTTTCATATATTAGCTAGA	1	8	1	2	2	8	1	1	1	10	10
10	TCTAGCTAATATATGAAACATATG	1	8	1	2	0	8	2	1	1	9	10
11	AAGTGACAAGGATGGGCCTCAATC	1	8	1	0	0	3	1	1	1	10	7
12	GATTGAGGCCCATCCTTGTCACTT	2	8	1	0	0	2	1	1	1	5	7
13	AGAGATGTAAAATTTTCATGATGTT	5	8	1	4	0	3	0	1	1	9	12
14	AACATCATGAAAATTTTACATCTCT	6	8	1	4	0	2	0	1	1	7	12
15	TTAGGTCAGTGGTCCCAAGTAG	3	8	1	0	0	0	0	1	1	5	4
16	CTACTTGGGACCACTGACCTAA	5	8	2	0	0	0	0	1	1	5	4
17	TGAGGCAGCCCCGTTGAG	5	8	3	0	0	0	0	1	1	3	5
18	CTCAACGGGGCTGCCTCA	4	8	2	0	0	1	0	1	1	7	5

SD, self dimers; HP, hair pins, *Note*: MFOLD predicts only hairpins.

behind using the T_m of the least stable primer has already been explained (von Ahsen *et al*., 2001).

Secondary structure studies

We selected complementary sense and antisense strands for our studies because it is well established that the predicted secondary structures of complementary single strands are not mirror images nor does their folding result in the same free energy decrease (Nielsen *et al*., 1995). For the sequences we have selected, a large number of secondary structures have been predicted by different oligonucleotide properties calculators, including the most widely used secondary structure prediction algorithm, MFOLD



Lane	1	2	3	4	5	6	7	8	9	10	11
Oligonucleotide	13	14	C2	15	16	C1	17	18	C3	1	2

Lane	12	13	14	15	16	17	18	19	20	21	22	23
Oligonucleotide	7	8	9	10	11	12	C2	3	4	5	6	C1

Fig. 2. Analysis of secondary structures in oligonucleotides. C1: T-21; C2: T-24; C3: T-18. Oligonucleotide numbers correspond to the sequences in Table 3.

(Zuker, 2003) (Table 2). Different predictions are provided for sense and antisense strands. But from Figure 2 it is evident that only oligonucleotides 7, 10, 11, 12 and 14 form secondary structure. Concentration-dependent single-strand melting studies established that these structures were not because of intermolecular interactions. Hence, these structures are considered to be hairpins. Presence of negligible secondary structures in these oligonucleotides at 4°C poses doubts regarding the mere existence of stable structures at temperatures used during molecular biology experimentation, questioning the authenticity of such predictions. However, all the programs except for oligonucleotide properties calculator provide the free energy of the structures predicted. We note that this may provide information about the stability of the predicted structures.

Primer designing studies

The details of the performance of each primer-design tool considered are tabulated (Table 3). Although biologists can modify the primer sequence, the target DNA cannot be modified. Little attention is usually paid to the analysis of the PCR template before experimental procedure. In an attempt to address this, we have selected AT-rich, GC-rich and AT = GC templates to check for the performance of different tools. In the initial analysis the primers generated by Primer Selection worked well. Web Primer, Primer 3 and Primer Design Assistant brought about amplifications in AT-rich and GC-rich templates only, and those generated by Do Primer, Exon Primer and Pride 1.2 (Haas *et al*., 1998) worked well for AT-rich and AT = GC templates. For the second phase of the analysis Do primer and Primer Selection were not available on the World Wide Web. The second phase was carried on only for the remaining nine tools. In the AT-rich category Primer 3 and Web Primer, in the AT = GC Exon primer and in the GC rich category Primer Design Assistant, Primer 3 and Primer Quest performed the best. Only one out of the three sequences in the AT = GC class could be amplified using Primer 3 generated primers. Primer 3 is user friendly as it considers a wide range of parameters and provides their complete description, but the usage of outdated Breslauer parameters (Breslauer *et al*., 1986) is its main drawback. The intricacies in designing the primers on web (Binas, 2000) and the details of primer design tools have been dealt previously (Abd-Elsalam, 2003). We emphasize that no reason is being attributed neither to failures of certain tools nor to the success of others.

DISCUSSION

In this work we have compared different oligonucleotide properties calculators for estimating T_m of oligonucleotides and have performed an accuracy benchmark using all relevant sequences for which experimental melting data are available. The aim of this study is not to disqualify any of the existing methods, but to demonstrate that significant differences in the T_m predictions of short DNA sequences are observed among them when a large number of sequences of practical application value are tested. This kind of assessment is biased to some extent because most of the experimental data available correspond to very short DNA sequences and have also been used to optimize and parameterize the existing methods, and the currently available data are not representative of the oligonucleotide sequence space either. This analysis encompasses every known parameter required for evaluating the accuracy of predictions, such as GC-content of the oligonucleotide, concentration of the salt at which it is melted, concentration of the strand and length of the sequence. This study would help the user to utilize that particular calculator which would give T_m with minimal deviation. Further, the model developed by us for calculating T_a^{opt} integrating the formulations developed by several groups can be applied in different laboratories by just finding the best fit for m and n in Equation (1) from the already existing amplicon dataset. The accuracy of this model depends on the precision of the values derived for m and n. Here, we demonstrate the use of NN thermodynamics for predicting T_m of the product.

Considerations such as avoidance of complementarity at the 3′ termini of the primers, which promotes the formation of primer–dimer artifacts, and avoidance of stable self-complementary hairpin loops that increase primer stability are essential to increase specificity of oligonucleotides. Though a large number of secondary structures have been predicted for certain sets of primers, successful amplification and good product yield have been observed in these cases. Although length-based approximations have yielded important predictions, sequence-dependent rules and parameters for hairpins, bulges, internal loops and multibranched loops are needed. Refinement of thermodynamic parameters characterizing the various structural motifs and incorporation of corrections for divalent cation (required for

Table 3. Details of Primer-design tools and their performances

Primer-design tools Used	AT-rich sequence	AT = GC sequence	GC-rich sequence	URL
DNASTAR	− + +	− + +	− + +	
Do Primer	+ na na	+ na na	− na na	http://doprimer.interactiva.de/input/frameset.html
Exon Primer	+ + −	+ + +	− + +	http://ihg.gsf.de/ihg/ExonPrimer.html
Gene Fisher	− − −	− + −	− + +	http://bibiserv.techfak.uni-bielefeld.de/genefisher/
Primer Design Assistant	+ + −	− − +	+ + +	http://dbb.nhri.org.tw/primer/
Pride 1.2	+ + −	+ + −	− − −	http://www.dkfz-heidelberg.de/tbi/services/Pride/search_primer
Primer Selection	+ na na	+ na na	+ na na	http://alces.med.umn.edu/rawprimer.html
Primer 3	+ + +	− + −	+ + +	http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi
Primer Quest	− − +	− + +	+ + +	http://biotools.idtdna.com/primerquest/
Primo Pro 3.4	− + +	− + +	− − +	http://www.changbioscience.com/primo/primo.html
Web Primer	+ + +	− + +	+ − +	http://seq.yeastgenome.org/cgi-bin/web-primer

'+' indicates that PCR optimization was not achieved using the primer designed using a specified tool. '−' indicates that PCR optimization was not achieved using the primer designed using a specified tool. The templates for which the performance of the tools under each category are as follows: AT-rich, AMPK, NKX6.1, F3; AT = GC, PKLR, NKX6.1, ISL1; GC-rich, VEGF, IPF1, INS. na, the tool was not available for analysis.

enzymatic activity) concentrations are required to further improve the quality of these predictions (SantaLucia and Hicks, 2004). We emphasize that prediction of the correct secondary structure is not the only goal but the accurate prediction of the energy required to unfold the portion of a long DNA so that an oligonucleotide can bind is also important (Dong *et al*., 2001), and a mention regarding this in the software would be of great help to the naive users. When a probe or primer has some secondary structures, there is a competition between secondary structure formation and target binding. The respective propensities of secondary structure formation and duplex formation ultimately determine the relevance of a given secondary structure. To evaluate the outcome of this competition, prediction of free energies is required. We thus caution the end users on the elimination of an otherwise good primer during primer design for different applications just on the basis of the number of the secondary structures predicted by the software they use.

Though there is a luxury of experimentally optimizing various parameters, poor predictions in primer design may not be tolerated. The expenses and time incurred for experimentation are huge and hence minimization of failures is utmost essential. Overall success rates of PCR are usually compromised due to the variation in template sequence and length. Also, regionalized GC-content of template DNA has been reported to be a strong predictor of PCR success (Benita *et al*., 2003). By using three different categories of templates for analysis, we provide an idea as to which tool provides better design depending on the template GC-content. Different tools compromise on different considerations and hence suffer a plethora of limitations. Though this is not an exhaustive exercise to determine the efficiency of each tool, our study does provide an insight into this aspect. We believe that this comparative analysis will provide some guidelines to follow in order to avoid or minimize large and frequent flaws in primer designing.

We readily acknowledge that this study aims at providing a safe application of the current methods available freely to all by considering their restrictions and limitations rather than to develop new methods or adding one more software or a server to the existing ones. We believe that our study would help in increasing the success of a variety of practical applications involving oligonucleotides.

ACKNOWLEDGEMENTS

The authors are thankful to Prof. S.K. Brahmachari, Director, IGIB for his support. The authors thank Ms Vijaya Banerjee and Mr Amitabh Sharma (Institute of Genomics and Integrative Biology) for their critical inputs. The authors acknowledge Dr K. Narayanasamy and Mr Mohd Nadeem Khan (The Centre for Genomic Application) for their help in oligo synthesis. SC, AM and RT acknowledge CSIR, Government of India, for providing their pre-doctoral fellowships. This study was supported by the CSIR task force project 'Predictive medicine using repeat and single nucleotide polymorphisms' (CMM0016) and the project 'Genetic and Proteomic studies in Diabetes Mellitus' (OLP 0030) funded by CSIR. The authors also thank Dr Alex Bateman, Executive Editor, Bioinformatics for giving them a fair chance to publish their work.

Conflict of Interest: none declared.

REFERENCES

Abd-Elsalam,K.A. (2003) Bioinformatic tools and guideline for PCR primer design. *Afr. J. Biotechnol*., **2**, 91–95.

Allawi,H.T. and Santa Lucia,J.,Jr (1997) Thermodynamics and NMR of internal G–T mismatches in DNA. *Biochemistry*, **36**, 10581–10594.

Benita,Y. *et al*. (2003) Regionalized GC-content of template DNA as a predictor of PCR success. *Nucleic Acids Res*., **31**, e99.

Binas,M. (2000) Designing PCR primers on the web. *Biotechniques*, **29**, 988–990.

Breslauer,K.J. *et al*. (1986) Predicting DNA duplex stability from the base sequence. *Proc. Natl Acad. Sci. USA*, **83**, 3746–3750.

Cantor,R.C. *et al*. (1970) Oligonucleotide interactions. III. Circular dichroism studies of the conformation of deoxyoligonucleotides and deoxyoligonucleotide analogs. *Biopolymers*, **9**, 1059–1077.

Caruthers,M.H. (1982) In Gassen,H.G. and Lang,A. (eds), *Chemical and Enzymatic Synthesis of Gene Fragments. A Laboratory Manual*. Verag-Chemie, Weinheim, Germany, pp. 71–79.

Dong,F. *et al*. (2001) Secondary structure prediction and structure-specific sequence analysis of single-stranded DNA. *Nucleic Acids Res*., **29**, 3248–3257.

Fedorova,O.S. *et al*. (1992) The influence of the target structure on the efficiency of alkylation of single-stranded DNA with reactive derivatives of antisense oligonucleotides. *FEBS Lett*., **302**, 47–50.

Freier,S.M. *et al*. (1986) Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl Acad. Sci. USA*, **83**, 9373–9377.

Gamper,H.B. *et al*. (1987) Solution hybridization of crosslinkable DNA oligonucleotides to bacteriophage M13 DNA. Effect of secondary structure on hybridization kinetics and equilibria. *J. Mol. Biol*., **197**, 349–362.

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

Howley,P.M. *et al*. (1979) A rapid method for detecting and mapping homology between heterologous DNAs. Evaluation of polyomavirus genomes. *J. Biol. Chem*., **254**, 4876–4883.

Le Novere,N. (2001) MELTING, computing the melting temperature of nucleic acid duplex. *Bioinformatics*, **17**, 1226–1227.

Marmur,J. and Dotty,P. (1962) Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. *J. Mol. Biol*., **5**, 109–118.

Nakano,S. *et al*. (1999) Nucleic acid duplex stability: influence of base composition on cation effects. *Nucleic Acids Res*., **27**, 2957–2965.

Nielsen,D.A. *et al*. (1995) SSCP primer design based on single-strand DNA structure predicted by a DNA folding program. *Nucleic Acids Res*., **23**, 2287–2291.

Owczarzy,R. *et al*. (1997) 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) Comparison of different melting temperature calculation methods for short DNA sequences. *Bioinformatics*, **21**, 711–722.

Robertson,J.M. and Walsh-Weller,J. (1998) An introduction to PCR primer design and optimization of amplification reactions. *Methods Mol. Biol*., **98**, 121–154.

Rychlik,W. *et al*. (1990) Optimization of the annealing temperature for DNA amplification *in vitro*. *Nucleic Acids Res*., **18**, 6409–6412.

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

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

SantaLucia,J.,Jr *et al*. (1996) Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry*, **35**, 3555–3562.

Sugimoto,N. *et al*. (1995) Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry*, **34**, 11211–11216.

Sugimoto,N. *et al*. (1996) Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. *Nucleic Acids Res*., **24**, 4501–4505.

vonAhsen,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.

Wallace,R.B. *et al*. (1979) Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: the effect of single base pair mismatch. *Nucleic Acids Res*., **6**, 3543–3557.

Xia,T. *et al*. (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson–Crick base pairs. *Biochemistry*, **37**, 14719–14735.

Yoshizawa,S. *et al*. (1997) GNA trinucleotide loop sequences producing extraordinarily stable DNA minihairpins. *Biochemistry*, **36**, 4761–4767.

Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*., **31**, 3406–3415.