BIOINFORMATICS ORIGINAL PAPER Vol. 21 no. 6 2005, pages 711–722
doi:10.1093/bioinformatics/bti066

Structural bioinformatics

Comparison of different melting temperature calculation methods for short DNA sequences

Alejandro Panjkovich and Francisco Melo*

Departamento de Genética Molecular y Microbiología, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Alameda 340, Santiago, Chile

Received on June 23, 2004; revised on September 10, 2004; accepted on October 5, 2004
Advance Access publication October 22, 2004

ABSTRACT

Motivation: The overall performance of several molecular biology techniques involving DNA/DNA hybridization depends on the accurate prediction of the experimental value of a critical parameter: the melting temperature T_m. Till date, many computer software programs based on different methods and/or parameterizations are available for the theoretical estimation of the experimental T_m value of any given short oligonucleotide sequence. However, in most cases, large and significant differences in the estimations of T_m were obtained while using different methods. Thus, it is difficult to decide which T_m value is the accurate one. In addition, it seems that most people who use these methods are unaware about the limitations, which are well described in the literature but not stated properly or restricted the inputs of most of the web servers and standalone software programs that implement them.

Results: A quantitative comparison on the similarities and differences among some of the published DNA/DNA T_m calculation methods is reported. The comparison was carried out for a large set of short oligonucleotide sequences ranging from 16 to 30 nt long, which span the whole range of CG-content. The results showed that significant differences were observed in all the methods, which in some cases depend on the oligonucleotide length and CG-content in a non-trivial manner. Based on these results, the regions of consensus and disagreement for the methods in the oligonucleotide feature space were reported. Owing to the lack of sufficient experimental data, a fair and complete assessment of accuracy for the different methods is not yet possible. Inspite of this limitation, a consensus T_m with minimal error probability was calculated by averaging the values obtained from two or more methods that exhibit similar behavior to each particular combination of oligonucleotide length and CG-content class. Using a total of 348 DNA sequences in the size range between 16mer and 30mer, for which the experimental T_m data are available, we demonstrated that the consensus T_m is a robust and accurate measure. It is expected that the results of this work would be constituted as a useful set of guidelines to be followed for the successful experimental implementation of various molecular biology techniques, such as quantitative PCR, multiplex PCR and the design of optimal DNA microarrays.

Availability: A binary software distribution to calculate the consensus T_m described in this work for thousands of oligonucleotides simultaneously for the LINUX operating system is freely available upon request to the authors or from our website http://protein.bio.puc.cl/melting-temperatures.html

Contact: fmelo@bio.puc.cl

Supplementary information: The large set of oligonucleotides, the detailed results of the comparative and accuracy benchmarks, and hundreds of comparative graphs generated during this work are available at our website http://protein.bio.puc.cl/melting-temperatures.html.

INTRODUCTION

The experimental performance and the outcome of several molecular biology techniques depend on the accurate prediction of the DNA melting temperature (T_m). This is particularly critical in the case of those techniques that involve several oligonucleotides simultaneously such as DNA microarrays of short sequences with fixed length or 'Affymetrix chips' (Pease *et al*., 1994), quantitative polymerase chain reaction (PCR) (Buck *et al*., 1991) and multiplex PCR, where large errors in the T_m estimation can lead to the amplification of non-specific products or to an inappropriate hybridization performance in general (Steger, 1994).

A variety of methods are available to calculate the T_m of short oligonucleotide sequences. The first approach described by Marmur and Doty (1962) uses a rather simple formula where the T_m depends only on the relative content of cytosine and guanine. This formula was later improved by adding a correction factor which is also responsible for salt concentration, thus adjusting the T_m value for different experimental conditions (Wetmur, 1991). An indepth analysis of DNA oligonucleotides and their corresponding experimental T_ms has led to the conclusion that not only the relative amounts of cytosine and guanine concentrations determine the thermal denaturation of DNA, but also the sequential arrangement of different nucleotides in DNA sequences were found to play a major role in the experimental value of T_ms. Hence the 'Nearest Neighbor' (NN) model was adopted for the calculation of sequence related T_ms (Borer *et al*., 1974). The NN model postulates that the free energy for duplex formation depends mostly on two factors: first, the initiation-free energy given by an unfavorable entropy as a consequence of a loss of translational freedom after the first DNA/DNA pair is formed; and second, the sum of the complementary pairwise terms between the oligonucleotide sequences (propagation terms), which are based on dinucleotide entities. In addition to these two terms, an entropic penalty is also considered, which takes into account the maintenance of symmetry in self-complementary sequences. The calculation of T_m values by the NN method requires a set of experimental thermodynamic

*To whom correspondence should be addressed.

parameters as an input. Till date, several tables with DNA/DNA thermodynamic parameters have been published (Gotoh and Tagashira, 1981; Vologodskii *et al*., 1984; Breslauer *et al*., 1986; Delcourt and Blake, 1991; Doktycz *et al*., 1992; SantaLucia *et al*., 1996; Sugimoto *et al*., 1996; Allawi and SantaLucia, 1997). Although a detailed comparison among these tables is difficult because several independent variables are involved in their derivation, attempts to clarify the similarities and differences among the various NN parameters have been carried out (Doktycz *et al*., 1992; SantaLucia, 1998). One major conclusion of the SantaLucia's (1998) work has been that a remarkable consensus exists among most of the different thermodynamic parameters, leading to the proposition and derivation of a single and unified set of parameters. In this study, however, we will demonstrate that when different sets of NN parameters are used to predict the T_m for a large and diverse number of DNA sequences of practical application value large differences are observed in many cases, which suggest that the current consensus parameters that have been proposed must be biased toward a reduced subset of possible DNA sequences.

When all the experimental conditions are fixed, the T_m value calculated for any particular DNA sequence using the NN model depends on the thermodynamic table that is used. Till date, a single large-scale benchmark comparing the predictive accuracy of 11 different thermodynamic tables has been published previously (Owczarzy *et al*., 1998). In that work, a high-predictive accuracy with a small error was found only for three independent parameter sets: the set of Doktycz *et al*. (1992); the set of Allawi and SantaLucia (1997) and the set of Sugimoto *et al*. (1996). Although this large-scale accuracy assessment constitutes a significant contribution, it must be mentioned that the experimental data that are currently available to perform such benchmarks have several strong biases. First, there is a bias toward oligonucleotide length, because some short-length oligonucleotides (~10mers long) are clearly overrepresented in the datasets. Unfortunately, most of the practical applications in molecular biology use oligonucleotides greater than this length. Second, most of the oligonucleotides in the set fall in the CG-content range between 40 and 60%, with a few representations at the most extreme cases. Third, and most importantly, a large fraction of the oligonucleotides for which experimentally determined T_ms are available have been already used to obtain or derive the same thermodynamic parameters with the help of which performance and accuracy are going to be assessed. Finally, not all of these oligonucleotides available are known to melt in a two-state manner, which is a primary condition in order to apply the NN model to calculate the T_m. These existing biases can be explained on the basis that the larger an oligonucleotide is, greater the chances that it will not melt in a two-state manner, and hence it will be of marginal interest for the derivation or validation of thermodynamic parameters to be used in the NN model.

Therefore, there are still several limitations that need to be considered when using these methods to estimate the experimental T_m of any given oligonucleotide for a practical application. Surprisingly, the majority of the current web servers and software programs that are available to calculate the T_m of oligonucleotides do not inform about these limitations and do not limit the input provided by the user, thus any valid DNA/RNA sequence is allowed. Out of the 17 web servers that we have found as freely available on the Internet (for a list of these web servers see the Supplementary material), only 4 of them provide some sort of warning about the limitations of the T_m calculation methods. However, at the time of performing the calculations, none of these web servers limits the input to be entered by the user. Thus, we believe that it is important to be aware of the current limitations that these methods have and also to be aware of the magnitude of errors that could arise when using them for practical applications. Unfortunately, most people use these softwares with no apparent conscience about the possible costs and risks, given the magnitude of errors that could arise from these calculations. It must be advised that depending on the method that is used, for a given specific oligonucleotide sequence, the absolute differences in the calculated T_m values could be large. Then, in practical terms, the question of which T_m value should be considered arises. The major aim of this work is to compare the T_m values calculated by the different methods in a large set of oligonucleotides that are representative of real-world applications in experimental biology. Thus, on the one hand, all the oligonucleotides used in this study fall in the size range between 16mer and 30mer, which is the widely used size range for the design of PCR primers and microarrays of short-length oligonucleotides. Although it can be argued that in this size range the two-state transitions are not guaranteed and therefore the NN model should not be used, it must be stated that it is not the major aim of this work to assess or to develop a more accurate method, but to highlight the magnitude of the existing differences and variations among methods when they are applied to a typical DNA sequence that is used in the laboratory. On the other hand, the set of oligonucleotides used in this study covers the complete space of CG-content, in contrast to the oligonucleotides that have been used to derive the existing thermodynamic parameters. Therefore, the present study constitutes an extensive and representative set of benchmark sequences to assess how similar the different predictions are.

The present study is divided into two major sections: the first section containing a comparative assessment of T_ms calculated using different methods and a second section containing an accuracy assessment of different methods for experimental T_m prediction. It must be noted that, in both cases, only DNA sequences with practical application values have been used. In the first section, we compare the three methods: the basic, salt adjusted and thermodynamic. The first objective was to evaluate if simple models such as the basic and salt adjusted methods could give similar T_m predictions compared to the ones obtained by the more complex thermodynamic method. The second objective was to compare in detail the similarities and differences of the T_m predictions obtained by the thermodynamic method while using three different parameter tables. Out of the different thermodynamic tables that have been published, three tables were derived from the analysis of optical melting curves for a variety of short synthetic DNA duplexes and are the most commonly used by the scientific community as expressed in the implementations of several standalone software programs and web browsers on the Internet (for a list of web servers along with their parameterizations, see Supplementary material). These three NN tables were used in this comparative study and include the original set of values reported by Breslauer *et al*. (1986), and the putatively improved sets of values reported by SantaLucia *et al*. (1996) and Sugimoto *et al*. (1996). It could be argued that Breslauer set is not as accurate as those of SantaLucia and Sugimoto sets based on the results of the previous accuracy benchmarks mentioned above. However, till date, not a single comparison has been performed among these thermodynamic sets using a large, representative and unbiased set of DNA sequences of practical application in molecular biology. Moreover, we will demonstrate here that all these thermodynamic sets, when

compared in a pairwise fashion, exhibit some differences and share a similar behavior at different regions of the oligonucleotide feature space (represented by length and CG-content). In the second section of this study, the third and last objective of this work is addressed, which consists of the accuracy assessment of these methods in predicting the experimental T_m of several DNA sequences. This accuracy benchmark set contained all DNA sequences in the length range between 16mer and 30mer for which the experimental T_m, salt concentration and oligonucleotide concentration values were available. Also, a consensus T_m value proposed in this study was also assessed, thereby giving the lowest average error. The consensus T_m value calculation is based on the consensus T_m map among different methods observed from the comparative assessment.

To achieve the comparative assessment, we have generated a large set of oligonucleotide sequences in a computer, calculated the T_m value for each of them by using different methods, and assessed the observed similarities, differences and correlations among predictions. These comparisons were performed independently for each combination of oligonucleotide length and CG-content, for a given range of oligonucleotide length that is of practical value. The regions of consensus and disagreement are highlighted. The results show that significant differences could be obtained while using different methods and that a consensus T_m value with a minimal error probability should be defined. Owing to the lack of experimental data available till date, which is not extensive or representative enough, we address this point by averaging the T_m values of these methods that consistently exhibited a similar behaviour for each particular combination of oligonucleotide length and CG-content. To support our model, an accuracy benchmark using all DNA sequences that fall within the size range covered in this study for which experimental T_m data are available was performed, demonstrating that the consensus T_m proposed here constitutes a robust and accurate measure. Finally, a list of guidelines to calculate the T_m of short DNA sequences with a minimal error probability is provided at the end of this paper.

METHODS

Set of artificial oligonucleotide sequences used in the comparative benchmark

A total of 300 000 DNA oligonucleotide sequences were randomly generated in a computer. The length of these random DNA sequences was restricted to be in the range of 16–30 nt. The random sequences were also generated and selected in such a way that they span a homogeneous distribution in the defined size range and additionally in the possible CG-content range. For this purpose, 10 uniform CG-content classes were defined (fixed intervals of magnitude equal to 10, ranging between 0 and 100% of CG-content). Therefore, each particular combination of sequence length and CG-content class was populated with the same total number of 2000 oligonucleotide sequences (i.e. 15 length classes × 10 CG-content classes × 2000 sequences = 300 000 total sequences).

Melting temperature calculations

In this study, three different methods were used to calculate and compare the T_ms of short DNA oligonucleotides: basic, salt adjusted and NN thermodynamic calculations. The basic T_m calculations were performed according to the following equation (Marmur and Doty, 1962):

$$T_m = 64.9 + 41.0 \times \left(\frac{yG + zC - 16.4}{wA + xT + yG + zC} \right),$$

where x, y, w and z are the number of the bases of T, G, A and C, respectively. This equation assumes that the annealing occurs under standard conditions in a buffered solution of 50 mM Na^+ and 50 nM of oligonucleotide concentration, with a pH close to 7.0, but the T_m of DNA is unaffected within a significant range of pH around 7.0 due to the lack of titratable groups close to this pH in the Watson–Crick paired DNA. The salt adjusted T_m calculations were performed using the following equation (Howley *et al*., 1979):

$$T_m = 100.5 + 41.0 \times \left(\frac{yG + zC - 16.4}{wA + xT + yG + zC} \right) - \left(\frac{820.0}{wA + xT + yG + zC} \right) + 16.6 \log([Na^+])$$

where x, y, w and z are the number of the bases of T, G, A and C, respectively. In the above equation, the second term adjusts for the GC-content and the third term adjusts for the length of the sequence. The equation also assumes that the annealing occurs under standard conditions of pH close to 7.0 and 50 nM of oligonucleotide concentration.

The melting temperatures are calculated using the nearest-neighbor model and thermodynamic data as previously described by SantaLucia *et al*. (1996). The equation used is as follows:

$$T_m = \frac{\sum (\Delta H_d) + \Delta H_i}{\sum (\Delta S_d) + \Delta S_i + \Delta S_{self} + R \times \ln \frac{C_T}{b}} + 16.6 \log[Na^+],$$

where sums of enthalpy (ΔH_d) and entropy (ΔS_d) are calculated over all internal nearest-neighbor doublets, ΔS_{self} is the entropic penalty for self-complementary sequences, and ΔH_i and ΔS_i are the sums of initiation enthalpies and entropies, respectively (Table 1). R is the gas constant (fixed at 1.987 cal/K · mol), C_T is the total strand concentration in molar units and T_m is the melting temperature given in Kelvin units. Constant b adopts the value of 4 for non-self-complementary sequences or equal to 1 for duplexes of self-complementary strands or for duplexes when one of the strands is in significant excess. The thermodynamic calculations assume that the annealing occurs in a buffered solution at pH near 7.0 and that a two-state transition occurs.

For the NN T_m calculations, three different thermodynamic tables were used in this work (summarized in Table 1): the first one is the original table published by Breslauer *et al*. (1986), the second table was published by SantaLucia *et al*. (1996) and the last table included in this study was published by Sugimoto *et al*. (1996). Therefore, five different T_m values were calculated and compared for each DNA oligonucleotide: one basic term (denominated 'bas'), one salt adjusted term (denominated 'sal') and three NN thermodynamic sets [denominated by date of publication: 'Th1' for Breslauer *et al*. (1986), 'Th2' for SantaLucia *et al*. (1996) and 'Th3' for Sugimoto *et al*. (1996)]. To perform a comparison of all the methods, the T_m calculations were carried out by fixing the oligonucleotide and sodium concentrations at 50 nM and 50 mM, respectively. In the case of the accuracy benchmark reported at the end of this work, the T_m calculations were carried out using the specific experimental oligonucleotide and salt concentrations reported for each particular case. The detailed experimental and calculated data used in this accuracy benchmark is available as Supplementary material.

Salt correction in melting temperature calculations

As described in the previous section, a unique salt correction term was used in this comparative benchmark for the thermodynamic and salt adjusted methods. This salt correction factor was reported by Schildkraut and Lifson (1965) for large DNA polymers and is still the most frequently used correction term in the current implementations available for T_m calculations. Little is known about the dependence of T_ms for short DNA sequences on salt concentration, particularly for bivalent ions. A recent and complete study has addressed the effects of sodium ions on experimental T_ms for short DNA oligomers (Owczarzy *et al*., 2004). That work contains an exhaustive and complete review about salt correction formulas that have been published in the literature, and dispatches a new and more accurate salt correction factor. However, it must be stated that the present comparative study is not influenced by the above salt correction factor, because all the calculated relationships of T_ms among these methods are relative and not absolute. Thus, the differences

Table 1. Thermodynamic parameters for DNA helix initiation and propagation in 1 M NaCl

Author	Breslauer *et al*. (1986)			SantaLucia *et al*. (1996)			Sugimoto *et al*. (1996)		
Abbreviation	TH1			TH2			TH3		
Propagation sequence	ΔH (kcal/mol)	ΔS (cal/°K mol)	ΔG (kcal/mol)	ΔH (kcal/mol)	ΔS (cal/°K mol)	ΔG (kcal/mol)	ΔH (kcal/mol)	ΔS (cal/°K mol)	ΔG (kcal/mol)
AA/TT	−9.1	−24.0	−1.9	−8.4	−23.6	−1.02	−8.0	−21.9	−1.2
AT/TA	−8.6	−23.9	−1.5	−6.5	−18.8	−0.73	−5.6	−15.2	−0.9
TA/AT	−6.0	−16.9	−0.9	−6.3	−18.5	−0.60	−6.6	−18.4	−0.9
CA/GT	−5.8	−12.9	−1.9	−7.4	−19.3	−1.38	−8.2	−21.0	−1.7
GT/CA	−6.5	−17.3	−1.3	−8.6	−23.0	−1.43	−9.4	−25.5	−1.5
CT/GA	−7.8	−20.8	−1.6	−6.1	−16.1	−1.16	−6.6	−16.4	−1.5
GA/CT	−5.6	−13.5	−1.6	−7.7	−20.3	−1.46	−8.8	−23.5	−1.5
CG/GC	−11.9	−27.8	−3.6	−10.1	−25.5	−2.09	−11.8	−29.0	−2.8
GC/CG	−11.1	−26.7	−3.1	−11.1	−28.4	−2.28	−10.5	−26.4	−2.3
GG/CC	−11.0	−26.6	−3.1	−6.7	−15.6	−1.77	−10.9	−28.4	−2.1
Any G·C pair?	0.0	−16.77	+5.0	0.0	−5.9	+1.82	+0.6	−9.0	+3.4
Only A·T pairs?	0.0	−20.13	+6.0	0.0	−9.0	+2.8	+0.6	−9.0	+3.4
Symmetry correction	0.0	−1.34	+0.4	0.0	−1.4	+0.4	0.0	−1.4	+0.4
5′-terminal-T·A-3 bp	0.0	0.0	0.0	+0.4	0.0	+0.4	0.0	0.0	0.0

or correlations among T_ms calculated by any two methods are not sensitive to this factor, as it is a constant value on both sides of the equation that is subtracted and therefore eliminated.

In the case of accuracy benchmark reported at the end of this work, the salt correction factor of course becomes relevant, because the calculated T_m is then compared to the experimentally observed temperature. Thus, for the accuracy benchmark we have used not only the correction factor described above, but also the new and improved correction factor recently published by Owczarzy *et al*. (2004). In addition, we also used the correction factor published by SantaLucia *et al*. (1996), which is slightly smaller than the one described above. The best-performing salt corrections were used to assess each method.

Comparative measures

Several measures of similarity between the T_m values reported for any two methods were used in this study. All comparisons were done within each grid point (i.e. for each fixed combination of oligonucleotide length and percentage of CG-content class), thus involving a total of 2000 oligonucleotide sequences. The calculated measures include the maximal observed absolute difference (MaxAD) of the 2000 pairwise comparisons between any two methods, the minimal observed absolute difference (MinAD), the average absolute difference (AveAD), the standard deviation of absolute differences (DevAD), the maximal observed difference (MaxD), the minimal observed difference (MinD), the average observed difference (AveD), the correlation coefficient (CC) and the percentage of cases, where the absolute difference between T_m values was equal to or less than 10 (Per10C), 5 (Per5C) and 3 (Per3C)°C. The correlation coefficient was only calculated between thermodynamic methods because the other two methods have a variance equal to zero for oligonucleotides of fixed length and CG-content (with the salt concentration being fixed, they depend only on these two variables, which were of course identical at each grid point where the analysis was carried out). Because of space constraints, only some of the comparative measures are reported in this paper. The complete comparative data is available graphically as Supplementary material at our web site http://protein.bio.puc.cl/melting-temperatures.html.

Accuracy benchmark dataset

The experimental data used here were extracted from different sources and consisted of a total of 108 unique oligonucleotide sequences in various salt concentrations, accounting for a total of 348 data points. All these sequences have a length ranging between 16mer and 30mer. A total of 37 unique sequences were extracted from Owczarzy *et al*. (1998); a total of 11 unique sequences were extracted from the NTDB database (Chiu *et al*., 2003); and finally, a total of 60 unique sequences were extracted from Owczarzy *et al*. (2004). The experimental T_m for each one of these unique 60 sequences was measured at five different salt concentrations, thus constituting a total of 300 different experimental points (Owczarzy *et al*., 2004). Therefore, a total of 348 data points were used in the accuracy benchmark reported at the end of this work. The detailed dataset information is available as Supplementary material at our website http://protein.bio.puc.cl/melting-temperatures.html.

Electronic availability of software, raw and calculated data

The software used in this work was fully written by us in ANSI C language and the executable LINUX version is freely available upon request or from our website. The complete set of oligonucleotides used in this work, along with the experimental and calculated data, and the complete set of the generated comparative graphs in full color are available at our website http://protein.bio.puc.cl/melting-temperatures.html.

RESULTS

In this study, thousands of short DNA sequences were generated in a computer, and the T_m for each of them was calculated using five different approaches. Several comparative measures were used to assess the differences and similarities of the calculated T_ms. The approaches included the basic T_m calculation (denominated 'bas'), the salt adjusted T_m calculation (denominated 'sal') and the NN thermodynamic method based on three of the most commonly used parameter sets. The thermodynamic parameter sets included the Breslauer table (Breslauer *et al*., 1986) denominated here as Th1, the SantaLucia table (SantaLucia *et al*., 1996) called Th2 and the Sugimoto table (Sugimoto *et al*., 1996) denominated as Th3. These parameters are listed in Table 1 and the mathematical expressions used to calculate the T_ms are described in detail in the Methods section.

The length of the DNA sequences was limited between 16 and 30 nt, which is the most commonly used length range for PCR primer design and *in situ* synthesized oligonucleotide microarrays. For each length, 10 different CG-content classes ranging between 0 and 100 were defined, thus covering the complete CG-content range. Finally, a total of 2000 DNA sequences were randomly generated for each particular combination of length and CG-content class. For each sequence, the T_ms were calculated using the methods described above and several comparisons were carried out (see Methods section).

The first comparison involved the maximal observed absolute difference of T_ms among all the methods at each combination of sequence length and percentage of CG-content. The results are shown in Figure 1A. It is clear that large differences are observed, ranging between 16 and 24°C. The sequences in the middle range of CG-content present the smallest differences, irrespective of their length. When the same comparison is carried out, but only for the thermodynamic methods (Fig. 1B), the absolute differences remain in a similar range, but with the maximal values slightly lower at 20°C. However, the difference dependencies are totally different. In this case, sequences with a larger fraction of C and G nucleotides exhibit larger differences in the T_m estimations, almost irrespective of sequence length. When the average absolute differences were plotted, the previous trends were perfectly conserved, but the variation range decreased as expected (Fig. 1C and D). The average differences range between 5 and 9°C for all the methods and between 4 and 12°C for the thermodynamic methods. Average differences were smaller for sequences in the CG-content range of 40–60%, but still above 4–5°C. The average differences slightly increased with sequence length, irrespective of the CG-content.

Another comparison was a similarity measure that consisted in the percentage of cases or oligonucleotide sequences where all the methods shared T_m estimations with a maximal difference of a fixed figure in Celsius degrees. When all the methods are compared simultaneously, only a small fraction of the plotted area exhibit significant similarities within 10°C. This area was restricted to short sequences of 16mer with a CG-content between 40 and 80% (Fig. 1E). When the thermodynamic methods are compared using this measure, similar results are observed, but the similarity area increases if the same figure value is considered (Fig. 1F). These results demonstrate that large and significant differences exist among different methods in the T_m estimations. High similarities are only observed for short oligonucleotide sequences with a medium and restricted CG-content.

To assess the possibility that a particular method could be concealing some additional existing similarities among the methods, a pairwise comparison of the thermodynamic methods was carried out. The results of these comparisons are shown in Figure 2. The comparison of the Breslauer set (Breslauer *et al*., 1986) or Th1 and the SantaLucia set (SantaLucia *et al*., 1996) or Th2 (Fig. 2A) demonstrates that the overall similarities observed earlier among all the thermodynamic methods (Fig. 1F) were a simple consequence of the existing similarities among these two methods, which present the largest differences in the all-against-all pairwise comparisons among the three thermodynamic methods (Fig. 2). However, it is quite surprising that despite the low similarity observed, the T_m estimations of these two methods were found to be highly correlated in a large fraction of the oligonucleotide feature space (Fig. 2B). In fact, these two thermodynamic methods are the most dissimilar ones, but indeed the most highly correlated for a large fraction of the oligonucleotide sequences tested in this study. This can be explained partially based on the observed differences of T_m predictions offered by these two methods, where Breslauer T_m values were consistently higher than SantaLucia estimations, in the whole range of sequence length and CG-content (data not shown). The comparison of Breslauer (Breslauer *et al*., 1986) or Th1 and Sugimoto (Sugimoto *et al*., 1996) or Th3 calculations exhibit high similarity in the CG-content range of 10–40%, irrespective of the sequence length (Fig. 2C). These two approaches presented a good correlation (Fig. 2D), but it is puzzling that the highest correlation values are not expressed at the same points where the highest similarities are observed (Fig. 2C and D). Finally, the comparison of SantaLucia (SantaLucia *et al*., 1996) or Th2 and Sugimoto (Sugimoto *et al*., 1996) or Th3 showed similarities only for medium to rich CG sequences, irrespective of the length (Fig. 2E). However, the same previously described effect was observed, the correlation is high only for the low-similarity regions (Fig. 2F).

To illustrate the magnitude of the observed disparity between similarities and correlation while comparing two thermodynamic methods, two particular grid points of the comparison between SantaLucia (SantaLucia *et al*., 1996) and Sugimoto (Sugimoto *et al*., 1996) calculations were chosen. It should be noted that these two grid points are not the most extreme points in the graph. On the one hand, it is clearly demonstrated that the correlation is high for a grid point where the similarity is low (Fig. 3A). On the other hand, it is demonstrated that the correlation is low for a grid point where the similarity is high (Fig. 3B).

In practical terms, irrespective of the correlation and similarities among different methods, what is desired is to have the possible most accurate T_m estimation. This simple study has clearly shown that the methods do not correlate nor exhibit a clear common pattern of similarity. Thus, it is quite difficult to address the question of which method or approach one should follow in order to be successful at obtaining an accurate T_m prediction for a wide and diverse range of oligonucleotide sequences. The best solution available, given the biased selection of oligonucleotide sequences that has been used to obtain and parameterize the current existing thermodynamic tables, is to minimize the possible error that can arise when estimating the T_m of a given DNA oligonucleotide. Owing to the lack of sufficient representative experimental data, this can only be achieved by following the average T_m of such methods that exhibited a similar behavior at a particular combination of sequence length and CG-content. The observed consensus among thermodynamic sets that has been obtained in this study is illustrated in Figure 4A. It can be seen that consensus among all the NN sets is only observed for short sequences in the middle CG-content range (Fig. 4A, white color). This is not surprising because almost all the sequences that have been used to obtain the thermodynamic parameters used in this study lie exactly in this area (Breslauer *et al*., 1986; SantaLucia *et al*., 1996; Sugimoto *et al*., 1996). However, two additional regions where only two NN sets overlap are also highlighted. These showed the similarity regions of Breslauer (Breslauer *et al*., 1986) and Sugimoto (Sugimoto *et al*., 1996) on the one hand (Fig. 4A, light gray color), and that of SantaLucia (SantaLucia *et al*., 1996) and Sugimoto (Sugimoto *et al*., 1996) on the other hand (Fig. 4A, dark gray color). In addition to these zones, some dangerous regions where none of these methods exhibits a similar behaviour are also illustrated (Fig. 4A, black color). These regions mostly occur at extreme CG-content values, almost irrespective of the sequence length.

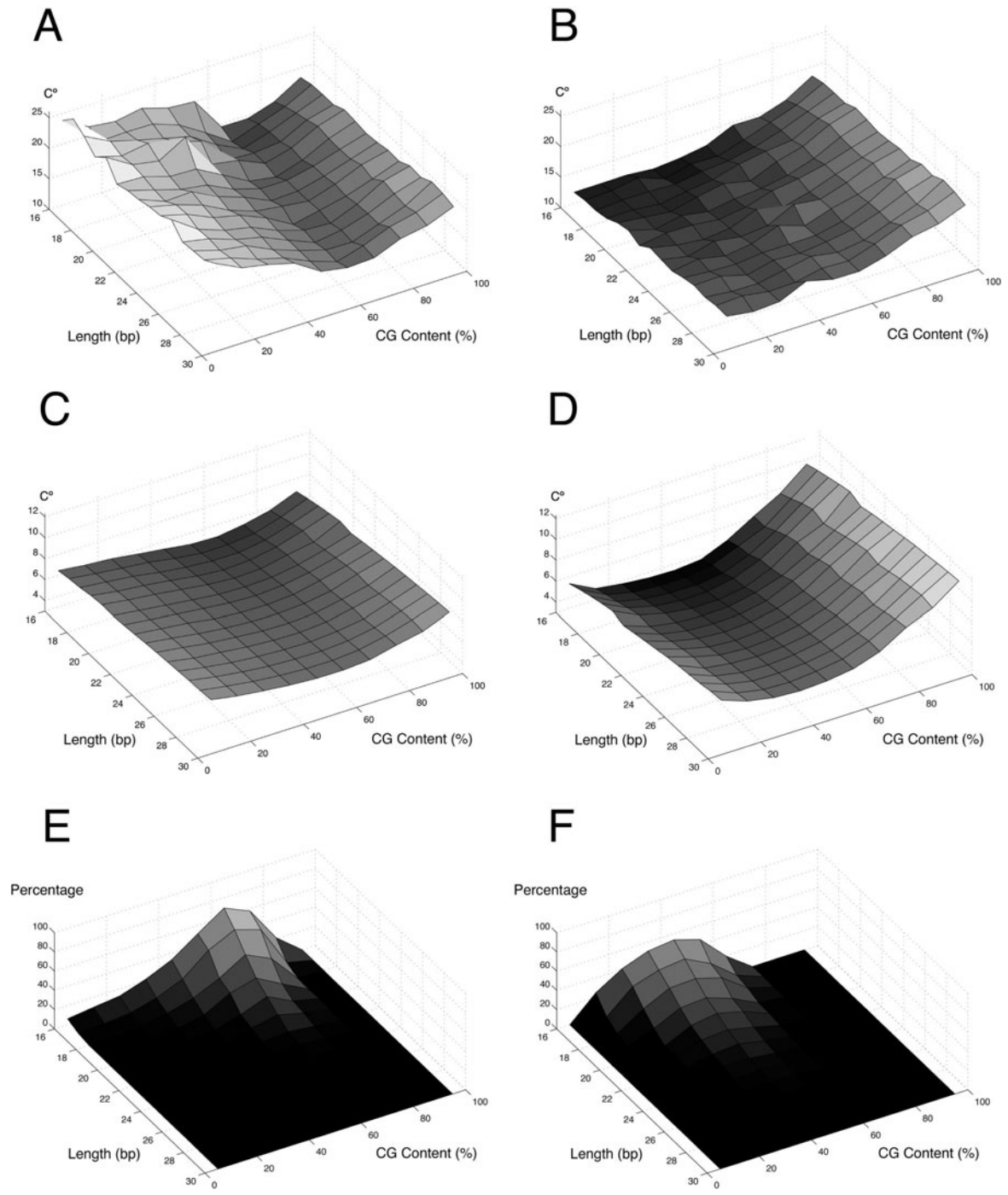


Fig. 1. Simultaneous comparison of T_ms calculated using all the methods. For each class of oligonucleotide length and percentage of CG-content, 2000 DNA sequences were randomly generated in a computer. The T_m of each sequence was calculated using different methods: basic, salt adjusted and the NN thermodynamic method based on three published tables (Table 1). Subsequently, the similarities and differences of the calculated values were evaluated. (**A**, **C** and **E**) The simultaneous comparison of all the methods is shown. (**B**, **D** and **F**) The simultaneous comparison of the three thermodynamic methods is shown. The maximum observed absolute differences are shown in A and B; the average absolute differences in C and D; and finally, the percentage of cases where the absolute difference is $\leq$10 or 5°C is shown in E and F. In the case of the simultaneous comparison of all the thermodynamic methods, a threshold of 5°C was used to define similarity, because in our judgement this figure represents a reasonable error estimation of T_m values, as it has been suggested previously. In the case of the simultaneous comparison of all methods, a threshold value of 10°C was used because no similarity was observed below that value (i.e. a flat graph in the XY plane was generated).

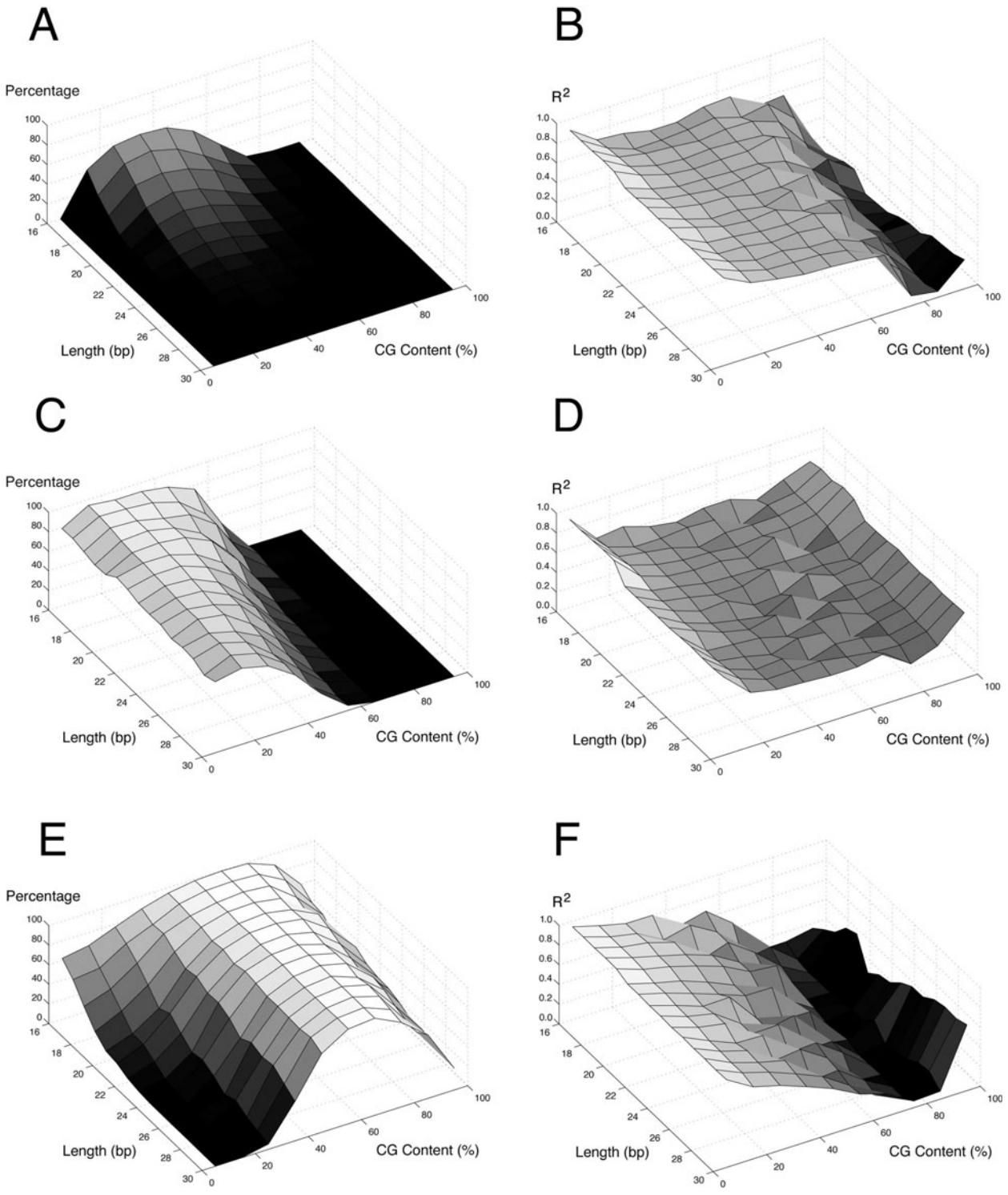


Fig. 2. Correlation and percentage of similarities among thermodynamic parameter sets. The pairwise similarities and correlation of the T_m values calculated by the thermodynamic sets using the parameters described in Table 1 were assessed. The procedure adopted was the same as described in the legend of Figure 1. (**A**, **C** and **E**) The percentage of oligonucleotide sequences where the absolute T_m difference is $\leq 5°C$ is shown as a function of sequence length and CG-content. (**B**, **D** and **F**) The correlation coefficient of the calculated T_ms is shown as a function of sequence length and percentage of CG-content. The corresponding pairwise comparisons are as follows: (A and B) Th1 versus Th2; (C and D) Th1 versus Th3 and (E and F) Th2 versus Th3. Th1 stands for Breslauer *et al*. (1986), Th2 for SantaLucia *et al*. (1996) and Th3 for Sugimoto *et al*. (1996).

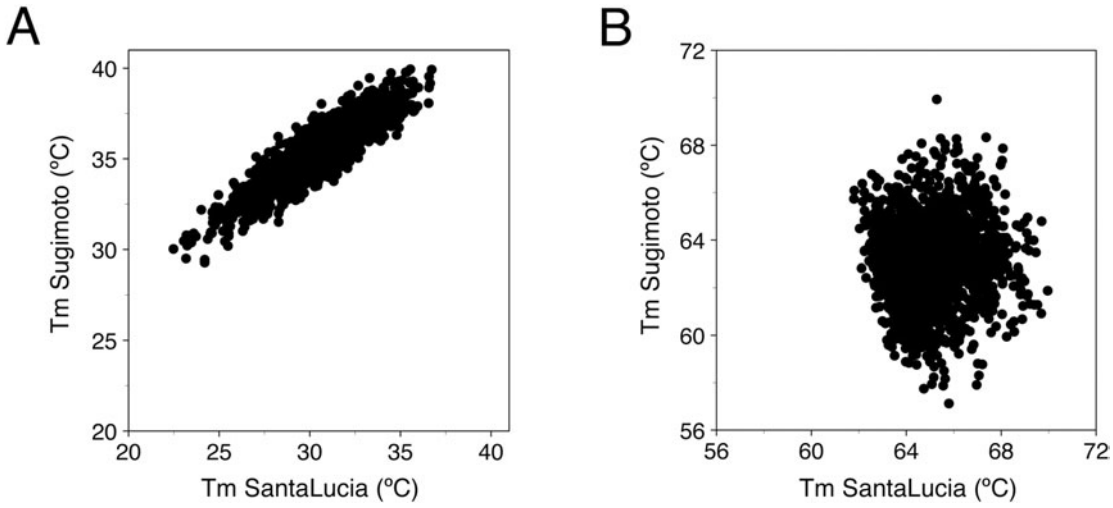


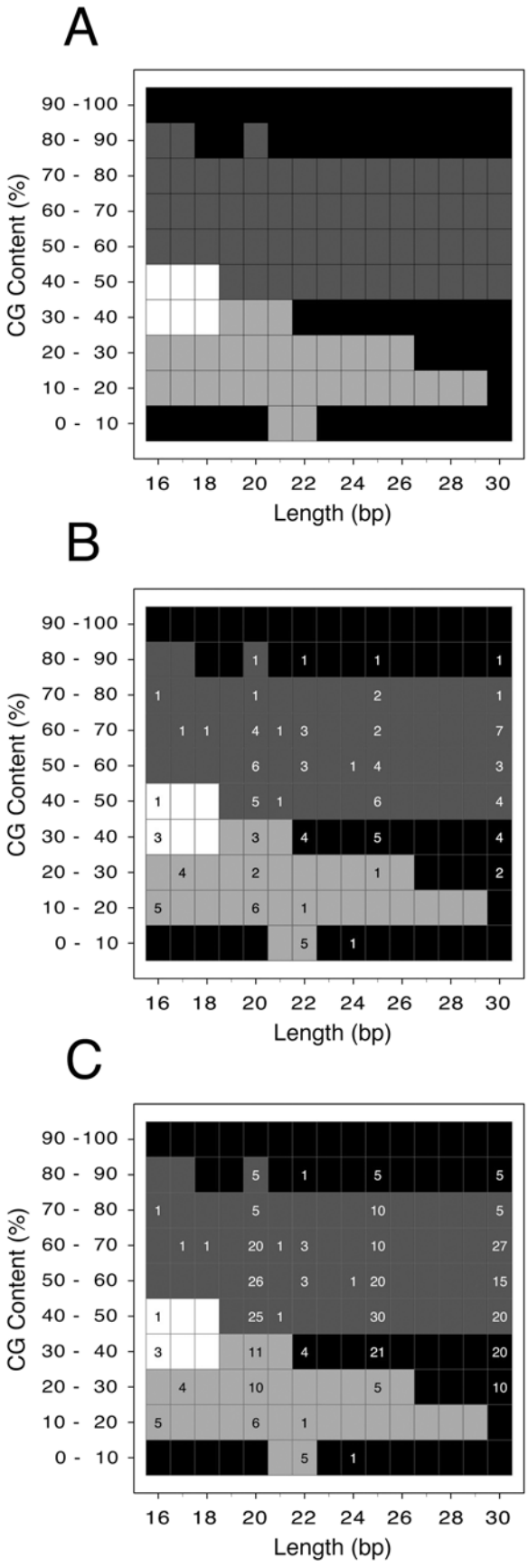
Fig. 3. Correlation of T_m estimations. The correlation of the T_m estimations by two thermodynamic parameter sets are shown at two specific grid points of Figure 2F. The NN sets compared were that of SantaLucia (SantaLucia *et al*., 1996) and Sugimoto (Sugimoto *et al*., 1996) as shown in Figure 2E and F. (**A**) Scatter plots of T_m values estimated by SantaLucia and Sugimoto for 2000 sequences of length equal to 18mer and CG-content ranging between 20 and 30%. The observed correlation coefficient value is 0.821994. (**B**) Scatter plots of T_m values estimated by SantaLucia and Sugimoto for 2000 sequences of length equal to 23mer and CG-content ranging between 70 and 80%. The observed correlation coefficient value is 0.006366.

In summary, four different regions or zones in oligonucleotide feature space were obtained: (1) Zone 3, where Th1, Th2 and Th3 exhibit similar T_m values (Fig. 4A, white color); (2) Zone 1, where Th1 and Th3 exhibit similar T_m values (Fig. 4A, light gray color); (3) Zone 2, where Th2 and Th3 exhibit similar T_m values (Fig. 4A, dark gray color); and (4) Zone 0, where none of the NN parameter sets share similar T_m values (Fig. 4A, black color). Based on these comparative results, a consensus T_m is then defined as the average of the T_m values calculated using those NN sets that exhibited a similar behavior at a given length and percentage of CG-content. Thus, depending on the grid point where a given oligonucleotide maps, the T_m values that are considered and averaged. Hence, the consensus T_m value of oligonucleotides mapping in Zone 3 would be the average of the T_m values calculated by the Th1, Th2 and Th3 sets. In the case of a DNA sequence falling in Zone 1, the consensus T_m value would be the average of the T_m values obtained by Th1 and Th3 sets. Finally, the consensus T_m value of a sequence falling in Zone 2 would be the average of the T_m values obtained by Th2 and Th3 sets. When a sequence maps into Zone 0 (black regions of Fig. 4A), it is not clear which parameter sets should be used. Although an average of all T_m values could be used for these regions where no similarities are observed, we suggest that oligonucleotides falling in the black regions should be avoided, because it is not clear which NN set could be considered.

A final objective of this work was to assess the accuracy of different methods at predicting the experimental T_m values for DNA sequences with a length between 16mer and 30mer. For that purpose, we recollected all the unique DNA sequences in that size range for which the experimental T_ms were available. A total of 108 unique oligonucleotide sequences that fulfill these requirements have been considered, for which frequencies of occurrence were mapped on oligonucleotide feature space and are displayed in Figure 4B. It must be mentioned that 60 of these 108 sequences had a total of five experimental T_m values determined at different salt concentrations, thus giving a total of 348 experimental data points to support the accuracy benchmark. The frequency of occurrence for all these sequences mapped onto oligonucleotide feature space is shown in Figure 4C.

The accuracy benchmark set was performed by using the salt correction factor that yielded the best results for each method, except for the consensus method, which used the T_m from each method calculated with the salt correction factor reported in the Methods section. Although this is unfair for the consensus method, the reason for doing so was to obtain values consequent with the comparative benchmark that generated the consensus map, which used this unique and fixed salt correction factor for all the methods (salt adjusted, Th1, Th2 and Th3). Three different salt correction factors were calculated and applied to the T_m predictions of each individual T_m calculation method. First, the salt correction described in the Methods section (Schildkraut and Lifson, 1965), where the salt adjusted method and the NN model with the Th1 and Th3 parameter sets showed the best performance. Second, the salt correction factor suggested by SantaLucia *et al*. (1996), where the NN model with the Th2 parameter set showed the best performance. None of the methods performed well when the new salt correction factor recently suggested by Owczarzy *et al*. (2004) was used. In fact, when this correction factor was considered, the accuracy of different predictions was significantly reduced in all the cases.

The results of the benchmark are summarized in Table 2. The best-performing method in terms of giving the closest value to the experimental T_m most of the times is Th2 (SantaLucia *et al*., 1996). In 40% of the cases, this method gives the most accurate experimental T_m prediction. However, when a more useful accuracy measure for the experimental biologist is used to assess the performance of these methods, the overall picture changes. When the percentage of cases where the predictions containing a maximal error of 3 or 5°C is used, the best performance is achieved by the consensus method proposed in this study, closely followed by the NN model using the thermodynamic parameters of Th3 (Sugimoto *et al*., 1996) and Th2



(SantaLucia *et al*., 1996). The same trend is observed when the average error is used as the accuracy measure. It must be noted that when the benchmark is carried out considering only those sequences that map into a consensus area of the oligonucleotide feature space (i.e. excluding those sequences that fall into a black region of Fig. 4A), the observed trend is not only confirmed, but also reinforced (Table 2, bottom values within each cell).

Statistical significance tests for different methods' average accuracies were performed (Table 3). No significant average accuracy differences were observed among Th2, Th3 and consensus methods when all the sequences from the benchmark set are used. These three methods are however more accurate than the basic, salt-adjusted and Th1, with statistical significance. When only those 281 sequences that fall in the consensus regions of the consensus map are used (Fig. 4C), the situation is different. In this case, Th3 and consensus methods are more accurate than the others, with statistical significance. These results demonstrated that the T_ms estimated by the Th3 NN set and the consensus method are the most accurate. Thus, the consensus T_m is a robust measure that exhibits a minimum average error and the largest fraction of cases predicted within an acceptable experimental error.

DISCUSSION

In this study, we have compared the similarities and correlations of the T_m values calculated using different methods. For this purpose, we have used a large and representative benchmark set of short oligonucleotide sequences. We did not address completely the problem of judging which method gives the closest value to the experimental T_m. We have only performed an accuracy benchmark using all the relevant sequences for which experimental T_m data are available to support the idea of using a consensus T_m with a minimal error probability. But we still believe that a benchmark based on the currently available sequences is not sufficient to validate or discard the performance of any existing method for T_m estimation. Although this last issue has been addressed by other authors previously and an effort to reconcile the existing differences has been made (SantaLucia, 1998; Rouzina and Bloomfield, 1999; Owczarzy *et al*., 1998), we believe that this kind of assessment is biased because most of the experimental data available correspond to very short DNA sequences and

Fig. 4. Consensus of T_m values among thermodynamic parameter sets. The consensus among two or three parameter sets is defined when at least 80% of the sequences exhibit an absolute difference between the calculated T_m values $<5°C$. All possible pairwise comparisons were carried out, as well as simultaneous comparison of the three thermodynamic sets. Th1 stands for Breslauer *et al*. (1986); Th2 stands for SantaLucia *et al*. (1996) and Th3 stands for Sugimoto *et al*. (1996). Th1 and Th2 did not show similar behavior in the whole range of sequence length and percentage of CG-content. (**A**) The observed consensus among the methods is as follows: Simultaneously, Th1 and Th3, Th2 and Th3, exhibit similar values (white color); only Th1 and Th3 exhibit similar values (light gray color); only Th2 and Th3 exhibit similar values (dark gray color) and finally, no consensus is observed among any of the methods (black color). (**B**) At each grid point, the total number of unique oligonucleotide sequences for which existing experimental T_m data available are also displayed (Legend to Table 2). (**C**) At each grid point, the total number of different cases (i.e. a unique combination of oligonucleotide sequence, salt and oligonucleotide concentration) for which the existing experimental T_m data available are also displayed.

Table 2. Accuracy benchmark of methods

Accuracy measure	BAS	SAL	BRE	SAN	SUG	CON
BEST (%)	0.6	5.2	3.5	40.8	26.2	23.9
	(0.0)	(6.4)	(2.1)	(38.1)	(27.1)	(26.3)
ERROR WITHIN 5°C (%)	11.2	31.0	26.2	83.3	83.6	83.9
	(9.6)	(35.2)	(24.6)	(84.0)	(84.3)	(86.1)
ERROR WITHIN 3°C (%)	3.7	14.9	14.4	60.9	60.1	61.5
	(2.9)	(17.4)	(12.5)	(57.3)	(62.6)	(64.1)
AVERAGE ERROR (°C)	12.3	7.1	8.5	2.9	2.9	2.8
	(12.6)	(6.7)	(8.6)	(3.0)	(2.7)	(2.6)

A total of 348 DNA sequences 16–30mers long with experimental T_m, salt and oligonucleotide concentrations available were used in this benchmark: 37 sequences were obtained from the work of Owczarzy *et al*. (1998); 11 sequences were obtained from the NTDB database (Chiu *et al*., 2003); and the remaining 300 sequences were obtained from Owczarzy *et al*. (2004). The complete table containing all the experimental values and the theoretical predictions made by using the various methods is available as Supplementary material. The T_ms were predicted with the basic method (BAS), the salt adjusted method (SAL), and the NN model with the thermodynamic parameters of Breslauer *et al*. (1986) (BRE), SantaLucia *et al*. (1996) (SAN) and Sugimoto *et al*. (1996) (SUG). The T_m was also predicted using the consensus method (CON) proposed in this study, which is based on the results obtained and shown in Figure 4. The consensus T_m corresponds to the average T_m of those methods that exhibit similar results at a given grid point of the oligonucleotide feature space. In those cases where no similarities are observed among methods (black regions of Fig. 4A), the average of all melting temperature values was used (top values within each cell). The results of the benchmark using the 281 sequences that are mapped in Zones 1, 2 and 3 (excluding the 67 sequences from the black regions in Fig. 4C) are shown in this table within parentheses. Four different accuracy measures are reported here. First, the percentage of cases where the method gives the closest prediction to the experimental T_m (BEST); second, the percentage of cases where the method gives a prediction within 5 and 3°C from the experimental T_m (ERROR WITHIN); and finally, the average of the absolute differences between the prediction method and the experimental T_m for all the cases considered (AVERAGE ERROR).

have also been used to optimize and parameterize the existing methods. In addition, it is also true that the currently available data are not representative of the oligonucleotide sequence space either. Most of the experimental data include oligonucleotide sequences <16mer and with a CG-content in the range of 40–60%. The aim of this study has not been 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. In practical terms, a large 'error' in the estimation of the T_m could easily arise, irrespective of which method is used. Thus, we believe that this comparative analysis will provide some guidelines to be followed in order to avoid or minimize large and frequent errors in the estimation of the experimental T_m of short oligonucleotide sequences. Also, we encourage people working in the experimental determination of T_ms to cover more extensively the practical oligonucleotide feature space when deriving new and improved thermodynamic tables.

It has been suggested that NN thermodynamics apply to duplexes ranging from 4 to 20 bp, because beyond 20 bp the transitions are unlikely to be two-state (SantaLucia *et al*., 1996). Thus, it could be argued that most of the oligonucleotides used in this study would not be following a two-state transition. However, in the same work, the authors suggested that the NN model can also provide reasonable approximations for oligonucleotide sequences that do not have two-state transitions. Using the NN model, they obtained good predictions of the experimental T_ms (5°C) for several oligonucleotides that were not following a two-state transition. Although it is still not clear if the NN model could be a valid approximation for larger sequences, where long range interactions and salt dependence could have a complex effect, most of the computational implementations currently available on the Internet or standalone software use this method for sequences that fall in the length range covered in this study. Thus, we believe that it is important to be aware of the potential errors and/or existing variations in the T_m predictions that will be generated by using those softwares without care.

Our results showed that complex relationships exist among methods. For instance, as it was shown in Figure 2, the similarity and correlation of T_m values among the methods are usually not aligned or in phase as may be expected. It was observed quite often that when two methods give similar predictions, they do not correlate. In addition, the pairwise comparison of average T_m differences between SantaLucia (SantaLucia *et al*., 1996) and Sugimoto (Sugimoto *et al*., 1996) revealed that SantaLucia predicts higher T_m values for sequences in the middle to high-CG-content range, irrespective of the sequence length (Supplementary material). In the low-CG-content range, the opposite result was observed. On the other hand, Breslauer (Breslauer *et al*., 1986) T_m predictions were higher than SantaLucia and Sugimoto T_m estimations, in the whole range of sequence length and CG-content, but with a different magnitude and behavior (Supplementary material). When the standard deviations of the T_m predictions were calculated for each grid point, it was surprising to note that Breslauer presented a large figure (>4°C) in the middle range of the CG-content and for short length sequences (Supplementary material). Perhaps this is the reason why Breslauer T_m predictions do not agree very well with the predictions based on other parameter sets (SantaLucia *et al*., 1996). SantaLucia and Sugimoto T_m predictions showed low-standard deviations at each grid point (the maximal values were ~2°C). However, when the minimum absolute difference T_m value was plotted at each grid point for all pairwise comparisons among thermodynamic methods, large differences were observed in all cases (Supplementary material). As an example of these observations, Sugimoto and Breslauer showed large differences in the high-CG-content range (>12°C). Breslauer and SantaLucia showed a similar trend, but in this case the differences were >4°C. SantaLucia and Sugimoto showed the opposite trend, with the largest differences occurring in the low-CG-content range. In this case, the minimum differences were >3°C.

As a complement of the comparative assessment, we finally performed an accuracy benchmark to evaluate the ability of different methods to predict the experimental T_m of oligonucleotide sequences with a practical application value, under varying conditions of salt and oligonucleotide concentrations. It must be mentioned that this benchmark was not fair in one major respect: all the methods were compared under such conditions which did not apply to some of them, such as the basic method under varying conditions of salt and oligonucleotide concentration, and the salt adjusted method under varying conditions of oligonucleotide concentration. This is one of the reasons why these methods performed very poorly in this benchmark. In this respect, it is important to note that in our comparative assessment, the basic method and the NN model using the Th3 parameters (Sugimoto *et al*., 1996) exhibit very similar T_m values in the complete oligonucleotide feature space (data not shown, see Supplementary material). This means that under certain experimental conditions of salt and oligonucleotide concentrations, a very simple method that

Table 3. Statistical significance of differences in average T_m prediction errors among the methods

	BAS	SAL	BRE	SAN	SUG	CON
BAS	—	9.1E – 47	2.8E – 23	7.3E – 83	1.3E – 112	1.6E – 107
SAL	2.9E – 46	—	2.0E – 20	3.0E – 53	5.2E – 67	1.1E – 56
BRE	6.7E – 26	1.0E – 10	—	3.2E – 73	1.2E – 75	7.6E – 71
SAN	2.7E – 99	7.9E – 68	3.1E – 85	—	0.050	0.009
SUG	5.1E – 129	2.1E – 92	3.0E – 81	0.479	—	0.1702
CON	5.0E – 122	8.3E – 75	4.0E – 83	0.477	0.432	—

Paired one-tail distribution Student's *t*-tests were carried out to assess the statistical significance of average error or accuracy differences between any two methods, as described by Press *et al*. (1997). The null hypothesis stated that there is no difference in the average error of melting temperatures predicted by any two methods (the observed average differences could arise only by chance). The upper right triangle of the table (dark gray shaded cells) contains the obtained *P*-values for the pairwise statistical tests of all the methods based on the 281 sequences used in the accuracy benchmark mapped in Zones 1, 2 and 3 of the consensus map shown in Figure 4. This set does not contain the sequences falling in the black regions (zones of no consensus among any two methods), which were excluded from the analysis. The lower left triangle of the table (light gray shaded cells) contains the obtained *P*-values for the same analysis mentioned above, but considering all 348 sequences used in the accuracy benchmark. The abbreviation of methods is as follows: BAS for the basic, SAL for the salt adjusted, BRE for Th1 NN set (Breslauer *et al*., 1986), SAN for Th2 NN set (SantaLucia *et al*., 1996), SUG for Th3 NN set (Sugimoto *et al*., 1996) and CON for the consensus method described in this study. For more details about the sequences used in the accuracy benchmark, see the legend of Table 2 or the section Methods.

does not take into account these parameters can give results similar to the more complex methods that indeed considered these variables. However, under varying conditions of salt and oligonucleotide concentration, the NN model with proper thermodynamic parameters clearly outperforms the simple methods. The NN model using the Th1 thermodynamic parameters (Breslauer *et al*., 1986) showed a very low performance in our accuracy benchmark, when compared to the other parameter sets. This was in agreement with what was observed in previous works for shorter oligonucleotide sequences (SantaLucia, 1998; Owczarzy *et al*., 1998). The accuracy of the NN model using the Th1 set was too low, that even the salt adjusted method had a better performance, although highly unsatisfactory.

The NN model using the Th2 (SantaLucia *et al*., 1996) and Th3 (Sugimoto *et al*., 1996) thermodynamic sets exhibited a good performance in this accuracy benchmark. This result suggests that both sets could be successfully used to predict the experimental T_m of oligonucleotide sequences in the range of 16–30mers. However, the best result was achieved by the consensus method described in this study, which is based on the three thermodynamic sets and in the consensus map generated in the comparative assessment carried out in this study. The consensus T_m value is a more robust measure, which is less sensitive to large errors that could arise while using a single parameter table. Irrespective of the method that is used, oligonucleotide sequences falling in regions where no consensus was observed are more prone to large errors in the experimental T_m estimation, as it was demonstrated in our accuracy benchmark. Thus, we recommend the use of the consensus T_m value for sequences in the range of 16–30mers, avoiding those sequences that fall in those regions where no consensus was observed (black regions of Fig. 4A).

The consensus T_m suggested in this work will minimize the error in the long run. Owing to the Lack of enough experimental data covering the complete oligonucleotide feature space under varying conditions of salt and oligonucleotide concentration, this is the safest way to proceed. The consensus T_m measure does not certainly guarantee the minimal error in all individual cases, but none of the methods can do that either. It must be mentioned that the accuracy benchmark proposed here favors those T_m values obtained with the Th2 (SantaLucia *et al*., 1996) and Th3 (Sugimoto *et al*., 1996) parameter sets, because most of the sequences fall in the regions where these two methods are highly similar, as shown by the comparative benchmark results. Out of the 348 experimental data points used in this benchmark, a total of 282 sequences fall in Zone 1 (dark gray region, Fig. 4C). On the other hand, it was also shown that the poor performance in this benchmark by the Th1 set (Breslauer *et al*., 1986) did not degrade the overall performance of the consensus method in Zone 2, although the T_m value obtained with the Th1 set is averaged in this zone with values obtained from the Th3 set. This demonstrates that the consensus T_m is a robust measure and validates the usefulness of this comparative study. Although the accuracy differences of T_m predictions obtained in this benchmark for Th3 and consensus method were not statistically significant, we suggest that in a large-scale application, the consensus T_m will turn out to be significantly more accurate than the T_m estimated by the Th3 NN set alone. Unfortunately, not enough experimental data are available yet to perform such a large-scale accuracy benchmark. Thus, we finally suggest that additional experimental data, covering a larger fraction of the oligonucleotide sequence space, are required to derive more accurate and robust thermodynamic parameters. Also, a large, heterogeneous, representative and unbiased set of sequences should be used to carry out a complete assessment of accuracy for the existing methods. The consensus method proposed in this study does not guarantee the best accuracy for any possible sequence. However, it minimizes the chances of error when using the existing methods for a diverse and large number of sequences, which is the case in the currently used practical molecular biology applications.

CONCLUSIONS

Significant differences are observed for the T_m values of short DNA oligonucleotides calculated by different T_m prediction methods. Additional experimental data covering a larger fraction of the oligonucleotide feature space are required in order to evaluate the accuracy of the current methods or to obtain a more precise estimation of the experimental T_m for any short oligonucleotide sequence. Meanwhile, the use of a consensus T_m calculation with a minimal error probability is suggested, which should be derived from the comparison of existing methods in a large benchmark set of sequences, as it was the case in this study. The guidelines to follow in order

to increase the success for practical molecular biology applications, from top to bottom priority, are as follows: (1) apply safely the current methods after considering the restrictions or limitations they have (i.e. avoid sequences that form stable alternative secondary structures, because such sequences do not follow a two-state transition); (2) if possible, use oligonucleotide sequences that fall in the middle range of CG-content and are shorter than 20–22mers (i.e. where most of the current T_m prediction methods agree); (3) avoid the use of sequences that fall in those regions of oligonucleotide feature space where none of the current methods agrees (black regions of Fig. 4A); (4) for large-scale applications with sequences where a two-state transition is not known to occur, use a consensus T_m calculation method like the one suggested in this study, and then maximizes the rate of success and (5) refer to the upcoming literature, wherein new and improved methods for T_m prediction will be developed.

ACKNOWLEDGEMENTS

We gratefully acknowledge the helpful comments and suggestions made by the two anonymous reviewers of this manuscript. This work was funded by grants from Fundación Andes (#13600/4), FONDECYT (#1010959) and DIPUC (#2004/01PF).

REFERENCES

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

Borer,P.N., Dengler,B., Tinoco,I.,Jr. and Uhlenbeck,O.C. (1974) Stability of ribonucleic acid double-stranded helices. *J. Mol. Biol.*, **86**, 843–853.

Breslauer,K.J., Frank,R., Blöcker,H. and Marky,L.A. (1986) Predicting DNA duplex stability from the base sequence. *Proc. Natl Acad. Sci. USA*, **83**, 3746–3750.

Buck,K.J., Harris,R.A. and Sikela,J.M. (1991) A general method for quantitative PCR analysis of mRNA levels for members of gene families: application to GABAA receptor subunits. *Biotechniques*, **11**, 636–641.

Chiu,W.L., Sze,C.N., Ma,N.T., Chiu,L.F., Leung,C.W. and Au-Yeung,S.C. (2003) NTDB: Thermodynamic Database for Nucleic Acids, Version 2.0. *Nucleic Acids Res.*, **31**, 483–485.

Delcourt,S.G. and Blake,R.D. (1991) Stacking energies in DNA. *J. Biol. Chem.*, **266**, 15160–15169.

Doktycz,M.J., Goldstein,R.F., Paner,T.M., Gallo,F.J. and Benight,A.S. (1992) Studies of DNA dumbbells. I. Melting curves of 17 DNA dumbbells with different duplex stem sequences linked by T4 endloops: evaluation of the nearest-neighbor stacking interactions in DNA. *Biopolymers*, **32**, 849–864.

Gotoh,O. and Tagashira,Y. (1981) Stabilities of nearest-neighbor doublets in double-helical DNA determined by fitting calculated melting profiles to observed profiles. *Biopolymers*, **20**, 1033–1042.

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

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

Owczarzy,R., Vallone,P.M., Gallo,F.J., Paner,T.M., Lane,M.J. and Benight,A.S. (1998) Predicting sequence-dependent melting stability of short duplex DNA oligomers. *Biopolymers*, **44**, 217–239.

Owczarzy,R., You,Y., Moreira,B.G., Manthey,J.A., Huang,L., Behlke,M.A. and Walder,J.A. (2004) Effects of sodium ions on DNA duplex oligomers: improved predictions of melting temperatures. *Biochemistry*, **43**, 3537–3554.

Pease,A.C., Solas,D., Sullivan,E.J., Cronin,M.T., Holmes,C.P. and Fodor,S.P.A. (1994) Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc. Natl Acad. Sci. USA*, **91**, 5022–5026.

Press,W.H., Teukolsky,S.A., Vetterling,W.T. and Flannery,B.P. (1997) *Numerical Recipes in C. The Art of Scientific Computing*, 2nd edn. Cambridge University Press, NY.

Rouzina,I. and Bloomfield,V.A. (1999) Heat capacity effects on the melting of DNA. 2. Analysis of nearest-neighbor base pair effects. *Biophys. J.*, **77**, 3252–3255.

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, Allawi,H.T. and Seneviratne,P.A. (1996) Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry*, **35**, 3555–3562.

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

Steger,G. (1994) Thermal denaturation of double-stranded nucleic acids: prediction of temperatures critical for gradient gel electrophoresis and polymerase chain reaction. *Nucleic Acids Res.*, **22**, 2760–2768.

Sugimoto,N., Nakano,S., Yoneyama,M. and Honda,K. (1996) Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. *Nucleic Acids Res.*, **24**, 4501–4505.

Vologodskii,A.V., Amirikyan,B.R., Lyubchenko,Y.L. and Frank-Kamenetskii, M.D. (1984) Allowance for heterogeneous stacking in the DNA helix-coil transition theory. *J. Biomol. Struct. Dyn.*, **2**, 131–148.

Wetmur,J.G. (1991) DNA probes: applications of the principles of nucleic acid hybridization. *Crit. Rev. Biochem. Mol. Biol.*, **26**, 227–259.