

A critical examination of predictions for oligo properties

Reidar Andreson

24.10.2005

$$T_m$$

- T_m – the **melting temperature** T_m refers to the temperature where 50% of the DNA is in a duplex form. In other words, 50% of the DNA has been denatured into single strands.
- Usually the optimal annealing temperature (T_a) is 5°C lower than the melting temperature of primer-template DNA duplex.

T_m

- T_m calculation types:
 - Basic** – % G+C content
(Marmur and Doty, 1962; Wallace *et al.*, 1979)
 - Salt Adjusted (SA)** – usually Na^+
(Howley *et al.*, 1979)
 - Nearest Neighbor (NN)** – closest nucleotides
thermodynamics + salt and oligo concentrations
(Breslauer *et al.*, 1986; SantaLucia *et al.*, 1998 etc.)
- Several programs to determine the properties of oligonucleotides

Secondary structures

- Self-dimers, hairpins, cross dimers of oligos
(Gamper *et al.*, 1987)
- Target DNA secondary structures
(Fedorova *et al.*, 1992)
- Prediction relies on energy minimization algorithms
- Many methods give poor predictions
(Dong *et al.*, 2001)

Aim of the study

- Identification of the best oligonucleotide properties calculator that predicts T_m with least deviation
- Evaluate secondary structure prediction methods
- Testing the efficiency of primer designing software

Experimental data

- 108 different oligos for thermal studies
 - 18 self-designed (6 with rich AT content [GC<45%], 6 with equal proportions [GC 45-55%], 6 with rich GC content [GC>55%])
 - 30 (Owczarzy *et al.*, 1997)
 - 60 (Owczarzy *et al.*, 2004)
- 25 freely available programs for T_m prediction
- Root mean square deviation (rmsd) were computed to determine the program providing least deviation between experiment and prediction

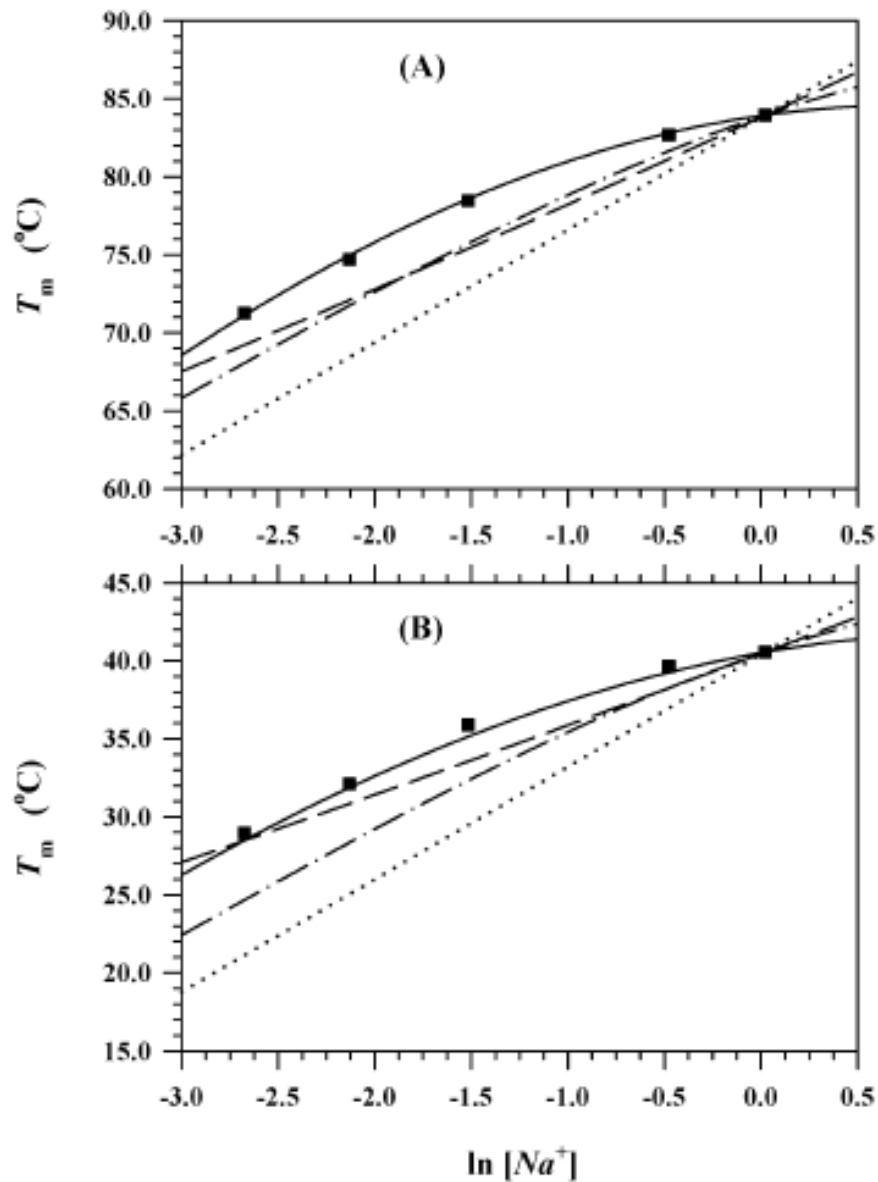


FIGURE 7: Comparison of some commonly used salt corrections and our new T_m salt correction 22. Experimentally measured (■) and predicted melting temperatures for two DNA duplex oligomers from our set, (A) 5'-ACGCCCACAGGATTAGGCTGGCCCA-CATTG-3' and (B) 5'-CCA ACTTCTT-3', are shown. Salt-corrected melting temperatures from 1 M Na^+ buffer to lower Na^+ concentrations are calculated by Owczarzy et al. T_m salt correction (—), Schildkraut-Lifson equation (···), Wetmur salt correction (-·-), and SantaLucia unified parameters salt correction (- - -).

Experimental data

- 18 different oligos for secondary structure studies
 - 9 sense and 9 antisense
- 6 programs was used for structure prediction
- Self-dimer and hairpins were checked

Experimental data

- 9 regions (1 kb each) from human genome for primer design studies
 - 3 AT-rich (from AMPK, NKX6.1, F3 genes)
 - 3 equal AT-GC (from PKLR, NKX6.1, ISL1 genes)
 - 3 GC-rich (from VEGF, IPF1, INS genes)
- RepeatMasker scanned
- 11 programs was used for primer design
- GC-content, salt conc., strand conc., length of the sequence
- One best candidate for each tool selected and PCR experiments were attempted

T_m results (1)

- 56 different modules were tested
- Default parameters:
 - Basic -> no salt and oligo concentrations
 - SA -> $[\text{Na}^+]$ 50 mM
 - NN -> $[\text{Na}^+]$ 50 mM, [oligo] 250 pM
- Experimental parameters were used where possible
- 25 modules (rmds ≤ 5) were selected
- All Basic modules were filtered out

T_m results (2)

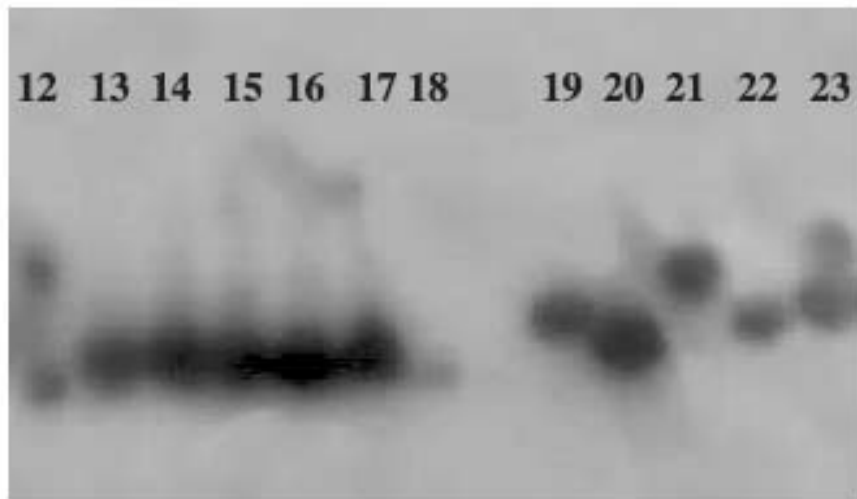
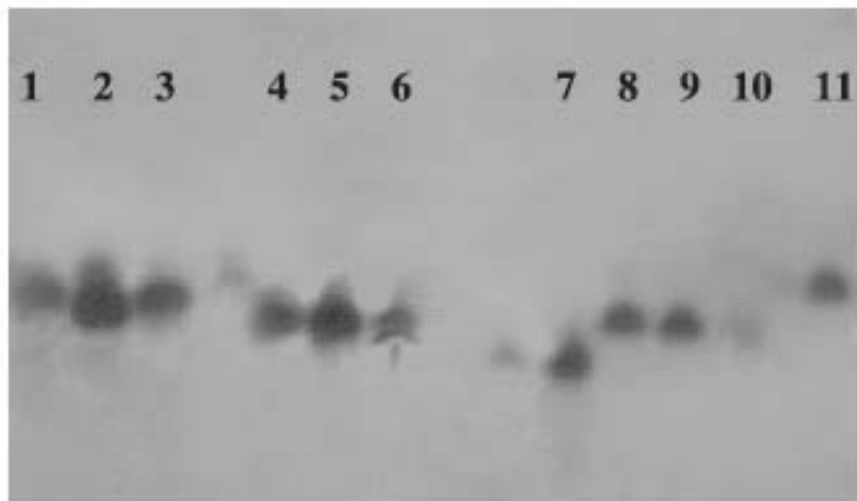
- 17 of 25 modules were tested with different salt concentrations
- 60 (Owczarzy *et al.*, 2004) oligos
- SA modules could not perform better than NN modules
- Best predictors:
 - MELTING (Le Novere, 2001) with Allawi (Santa Lucia, 1998) module
 - Oligo calculator (McLab)
(http://tool.mclab.com/toolbox/oligo_calculator.jsp)
 - HYTHER (<http://ozone2.chem.wayne.edu/>)
 - T_m calculator for oligos (Biomath Calculator; Promega)

T_m results (3)

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 (RNAure)	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.



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

Structure results

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
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	CCTTCAACCCAGCCTTGGCC	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	GATTGAGGCCCATCCTTGTCACCTT	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.

Primer design results (1)

- Optimal annealing temperature prediction:

$$T_a^{\text{OPT}} = 0.34(T_{m_{\text{primer}}}) + 0.46(T_{m_{\text{product}}})$$

modified (Rychlik *et al.*, 1990) formula

- Used best T_m predictors (MELTING and Oligo calculator (McLab))
- T_a^{OPT} calculated for all amplicons
- Experiments with those optimised T_a^{OPT} values
- +/- <- successful PCR or not

Primer design results (2)

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.

Consensus T_m estimation method

- Comparison of differences among of the published DNA/DNA T_m calculation methods
- Oligos ranging 16 – 30 bp
- Whole range GC-content
- 348 experimentally validated oligos

Consensus T_m estimation method

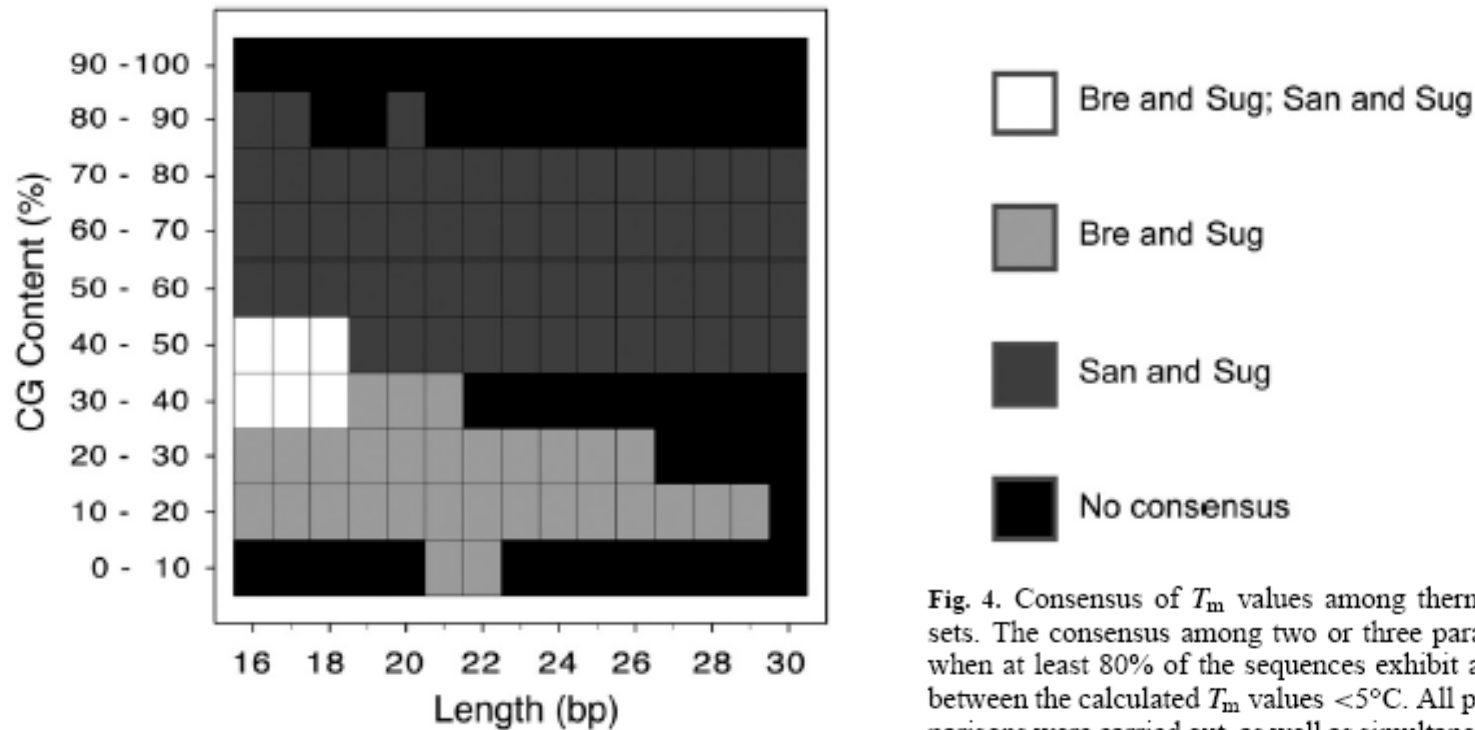


Fig. 4. Consensus of T_m values among thermodynamic parameter sets. The consensus among two or three parameter sets is defined when at least 80% of the sequences exhibit an absolute difference between the calculated T_m values $<5^\circ\text{C}$. All possible pairwise comparisons were carried out, as 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).

Table 2. Accuracy benchmark of methods

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

Consensus T_m estimation method

- Avoid secondary structures (because such sequences are not going to follow a two-state transition such sequences are not going to follow a two-state transition)
- Avoid using sequences that fall in those regions of oligonucleotide feature space where none of the current methods agrees whole range GC-content
- If possible, use oligonucleotide sequences that fall in the middle range of CG-content and of a length 16–22mer