Supporting Information for World Wide Web Edition

Richard Owczarzy, Yong You, Bernardo G. Moreira, Jeffrey A. Manthey, Lingyan Huang, Mark A. Behlke, and Joseph A. Walder (2004) **Effects of Sodium Ions on DNA Duplex Oligomers: Improved Predictions of Melting Temperatures**.

Approach to Derive Nearest-Neighbor T_m Salt Correction

The T_m salt correction given by equation [22] can be expanded to include sequence specific nearest-neighbor interactions (*73, 74*). The coefficients $B_4(g)$ and $C_4(g)$ from equation [18] for DNA duplex g can be partitioned into the sum of nearest-neighbor parameters,

$$B_4(g) = \sum_{i,j=A,T,C,G,E} (B_4)_{ij}^{n-n} \times \frac{N_{ij}(g)}{\sum_{ij=A,T,C,G,E} N_{ij}(g)} = \sum_{i,j=A,T,C,G,E} (B_4)_{ij}^{n-n} \times f_{ij}(g) \qquad [S1]$$

$$C_4(g) = \sum_{i,j=A,T,C,G,E} (C_4)_{ij}^{n-n} \times \frac{N_{ij}(g)}{\sum_{ij=A,T,C,G,E} N_{ij}(g)} = \sum_{i,j=A,T,C,G,E} (C_4)_{ij}^{n-n} \times f_{ij}(g) \qquad [S2]$$

where $(B_4)_{ij}^{n-n}$ and $(C_4)_{ij}^{n-n}$ are sequence dependent nearest-neighbor parameters. The coefficient, $N_{ij}(g)$, is the number of times the particular nearest-neighbor doublet (i, j = A, T, C, G, E) appears in the DNA duplex g. 'E' denotes the ends (*8, 73*). The sums in the equations [S1] and [S2] are calculated over the 10 internal and 2 end (initiation) nearest-neighbor parameters. The fractions of the occurrence for the particular nearest-neighbor doublet, $f_{ij}(g)$, in the sequence of

duplex *g* are calculated from $N_{ij}(g)$ coefficients. Because the duplex has $N_{bp}(g)$ base pairs, it contains $N_{bp}(g) + 1$ nearest-neighbor doublets including the end interactions, i.e. $\sum_{ij=A,T,C,G,E} N_{ij}(g) = N_{bp}(g) + 1$. Our data set consists of 92 molecules with unique base sequences ($g = 1, 92$). Thus the combination of the equations [18] and [S1] provides a system of 92 linear equations, which can be solved for 12 unknown $(B_4)_{ij}^{n-n}$ nearest-neighbor parameters. This can be performed (*53*) by minimizing the value of χ^2,

$$\chi^2 = | (\mathbf{F} \times \mathbf{B}^{\mathbf{n-n}} - \mathbf{B}) \times \boldsymbol{\sigma}_{\boldsymbol{B}}^{-1} |^2 \qquad [S3]$$

where $\boldsymbol{\sigma}_{\boldsymbol{B}}$ is the diagonal matrix whose elements are experimental errors of $B_4(g)$ and **F** is the 92 X 12 design matrix of the elements $f_{ij}(g)$. Each row of **F** comprises the elements $f_{ij}(g)$ for the particular DNA duplex *g* and each column of **F** contains the elements $f_{ij}(g)$ for the particular nearest-neighbor *ij*. $\mathbf{B}^{\mathbf{n-n}}$ is the column vector of 12 unknown $(B_4)_{ij}^{n-n}$ parameters to be solved for and **B** is the column vector of coefficients $B_4(g)$, $g = 1, 92$ determined for each DNA duplex from quadratic fits using equation [18]. A unique solution $\mathbf{B}^{\mathbf{n-n}}$ is achieved only if matrix **F** is not rank deficient, i.e., it has no singular values (*53*). That was verified to be the case for all our calculations. An analogous system of linear equations can be constructed for coefficients $C_4(g)$ and solved by minimizing χ^2,

$$\chi^2 = | (\mathbf{F} \times \mathbf{C}^{\mathbf{n-n}} - \mathbf{C}) \times \boldsymbol{\sigma}_{\boldsymbol{C}}^{-1} |^2 \qquad [S4]$$

where $\boldsymbol{\sigma}_C$ is the diagonal matrix whose elements are experimental errors of $C_4(g)$, $\mathbf{C}^{\mathbf{n\text{-}n}}$ is the column vector of 12 unknown nearest-neighbor parameters $(C_4)_{ij}^{n-n}$ to be solved for and **C** is the column vector of coefficients $C_4(g)$, g = 1, 92 determined from equation [18] for each DNA duplex. Solutions for equations [S3] and [S4] were obtained by singular value decomposition (*53*) for our UV melting data set of 92 duplex DNA oligomers. Resulting nearest-neighbor parameters $(B_4)_{ij}^{n-n}$ and $(C_4)_{ij}^{n-n}$ are summarized in Table S-1 and they were used to scale melting temperatures between buffers of different sodium ion concentrations using the nearest-neighbor salt correction function [23].

Supporting Information for World Wide Web Edition

Table S-1. The nearest-neighbor salt correction parameters and their standard deviations were derived using SVD. Extra significant figures are presented to prevent accumulation of rounding errors in calculations.

Nearest-neighbor sequence (5' to 3')	$(B_4)_{ij}^{n-n}$	$(C_4)_{ij}^{n-n}$
AT/AT	$(-1.75 \pm 2.02) \times 10^{-5}$	$(7.46 \pm 7.49) \times 10^{-6}$
TA/TA	$(-8.02 \pm 1.98) \times 10^{-5}$	$(0.82 \pm 7.33) \times 10^{-6}$
AA/TT	$(-5.07 \pm 0.49) \times 10^{-5}$	$(6.28 \pm 1.81) \times 10^{-6}$
AC/GT	$(-1.54 \pm 1.49) \times 10^{-5}$	$(1.16 \pm 0.55) \times 10^{-5}$
CA/TG	$(-0.19 \pm 1.37) \times 10^{-5}$	$(8.78 \pm 5.07) \times 10^{-6}$
TC/GA	$(-6.18 \pm 1.39) \times 10^{-5}$	$(4.71 \pm 5.16) \times 10^{-6}$
CT/AG	$(0.62 \pm 1.40) \times 10^{-5}$	$(9.88 \pm 5.17) \times 10^{-6}$
CG/CG	$(4.60 \pm 1.86) \times 10^{-5}$	$(8.04 \pm 6.90) \times 10^{-6}$
GC/GC	$(-3.98 \pm 1.83) \times 10^{-5}$	$(6.70 \pm 6.77) \times 10^{-6}$
GG/CC	$(-1.12 \pm 0.57) \times 10^{-5}$	$(7.62 \pm 2.10) \times 10^{-6}$
EA/TE and ET/AE	$(1.87 \pm 1.19) \times 10^{-5}$	$(2.88 \pm 0.44) \times 10^{-5}$
EG/CE and EC/GE	$(1.81 \pm 1.13) \times 10^{-5}$	$(2.25 \pm 0.42) \times 10^{-5}$

Supporting Information for World Wide Web Edition

Table S-2. Predicted and experimental calorimetric melting temperatures of DNA duplex oligomers in 69 mM and 1 M Na^+ buffers.

	1 M $[Na^+]^a$			69 mM $[Na^+]^b$	
	Exper. C_t^1	Exper. T_m^1	Exper. T_m^0 (°C)	Exper. T_m	Predicted T_m
DNA Sequence (5' to 3')	(μM)	(°C) at C_t^1	at C_t = 180 μM	(°C)	(°C)^c
11-mer					
CGTACACATGC	119	62.5	63.6	53.3	51.5
15-mers					
TACTAACATTAACTA	151	62.6	63.0	46.8	46.8
ATACTTACTGATTAG	91	62.0	63.4	47.8	48.0
GTACACTGTCTTATA	147	66.3	66.6	51.6	51.7
GTATGAGAGACTTTA	142	66.7	67.2	50.7	52.2
TTCTACCTATGTGAT	151	65.6	65.9	51.7	51.0
AGTAGTAATCACACC	140	68.6	69.1	54.6	54.8
ATCGTCTCGGTATAA	153	71.2	71.5	56.5	57.0
ACGACAGGTTTACCA	84	72.4	74.1	59.6	60.2
CTTTCATGTCCGCAT	128	74.0	74.8	61.1	60.8
TGGATGTGTGAACAC	139	72.9	73.4	58.4	59.5

Table S-2. (cont.)

	1 M $[Na^+]^a$			69 mM $[Na^+]^b$	
	Exper. C_t^1	Exper. T_m^1	Exper. T_m^0 (°C)	Exper. T_m	Predicted T_m
DNA Sequence (5' to 3')	(μM)	(°C) at C_t^1	at C_t = 180 μM	(°C)	(°C)^c
ACCCCGCAATACATG	146	74.3	74.7	62.3	61.7
GCAGTGGATGTGAGA	156	74.1	74.3	61.8	61.3
GGTCCTTACTTGGTG	164	71.4	71.5	58.8	58.7
CGCCTCATGCTCATC	118	76.0	76.8	64.3	64.5
AAATAGCCGGGCCGC	131	82.4	83.1	70.3	71.2
CCAGCCAGTCTCTCC	156	77.6	77.8	65.1	66.3
GACGACAAGACCGCG	149	80.0	80.3	68.9	68.7
CAGCCTCGTCGCAGC	108	82.0	83.1	71.8	72.2
CTCGCGGTCGAAGCG	151	82.9	83.2	72.1	72.2
GCGTCGGTCCGGGCT	163	85.7	85.9	75.4	75.7
20-mers					
TATGTATATTTTGTAATCAG	150	69.1	69.4	53.0	52.6
TTCAAGTTAAACATTCTATC	108	70.1	70.9	54.1	54.6

Table S-2. (cont.)

	1 M $[Na^+]^a$			69 mM $[Na^+]^b$	
	Exper. C_t^1	Exper. T_m^1	Exper. T_m^0 (°C)	Exper. T_m	Predicted T_m
DNA Sequence (5' to 3')	(μM)	(°C) at C_t^1	at C_t = 180 μM	(°C)	(°C)^c
TGATTCTACCTATGTGATTT	149	73.2	73.5	57.6	57.5
GAGATTGTTTCCCTTTCAAA	126	74.1	74.6	57.1	59.2
ATGCAATGCTACATATTCGC	170	78.0	78.1	64.0	63.1
CCACTATACCATCTATGTAC	147	73.0	73.2	59.6	58.6
CCATCATTGTGTCTACCTCA	167	78.1	78.2	64.2	63.7
CGGGACCAACTAAAGGAAAT	153	77.5	77.8	62.5	63.4
TAGTGGCGATTAGATTCTGC	155	80.2	80.4	65.6	65.8
AGCTGCAGTGGATGTGAGAA	165	82.2	82.3	68.8	68.2
TACTTCCAGTGCTCAGCGTA	142	83.3	83.6	69.7	69.4
CAGTGAGACAGCAATGGTCG	154	81.9	82.1	68.9	68.7
CGAGCTTATCCCTATCCCTC	143	79.3	79.6	65.3	66.4
CGTACTAGCGTTGGTCATGG	141	80.7	81.0	68.2	67.7
AAGGCGAGTCAGGCTCAGTG	161	86.5	86.6	73.5	73.6

Table S-2. (cont.)

	1 M $[Na^+]^a$			69 mM $[Na^+]^b$	
	Exper. C_t^1	Exper. T_m^1	Exper. T_m^0 (°C)	Exper. T_m	Predicted T_m
DNA Sequence (5' to 3')	(μM)	(°C) at C_t^1	at C_t = 180 μM	(°C)	(°C)^c
ACCGACGACGCTGATCCGAT	162	87.2	87.3	74.7	74.3
AGCAGTCCGCCACACCCTGA	157	87.8	88.0	75.9	75.6
CAGCCTCGTTCGCACAGCCC	163	88.2	88.3	76.2	76.6
GTGGTGGGCCGTGCGCTCTG	160	90.9	91.1	78.8	79.9
GTCCACGCCCGGTGCGACGG	174	91.0	91.0	80.0	80.6
25-mers					
GATATAGCAAAATTCTAAGTTAATA	141	74.1	74.3	56.4	57.0
ATAACTTTACGTGTGTGACCTATTA	153	80.3	80.5	64.4	64.2
GTTCTATACTCTTGAAGTTGATTAC	154	76.1	76.2	59.7	60.3
CCCTGCACTTTAACTGAATTGTTTA	146	80.2	80.5	65.1	64.7
TAACCATACTGAATACCTTTTGACG	123	78.7	79.1	63.8	63.4
TCCACACGGTAGTAAAATTAGGCTT	158	81.4	81.5	66.9	66.2
TTCCAAAAGGAGTTATGAGTTGCGA	142	81.5	81.8	66.6	66.4

Table S-2. (cont.)

	1 M $[Na^+]^a$			69 mM $[Na^+]^b$	
	Exper. C_t^1	Exper. T_m^1	Exper. T_m^0 (°C)	Exper. T_m	Predicted T_m
DNA Sequence (5' to 3')	(μM)	(°C) at C_t^1	at C_t = 180 μM	(°C)	(°C)[c]
AATATCTCTCATGCGCCAAGCTACA	156	84.8	84.9	69.9	69.8
TAGTATATCGCAGCATCATACAGGC	165	82.6	82.6	68.6	67.7
TGGATTCTACTCAACCTTAGTCTGG	144	80.9	81.1	66.2	66.3
CGGAATCCATGTTACTTCGGCTATC	124	82.9	83.3	68.4	68.9
CTGGTCTGGATCTGAGAACTTCAGG	153	83.9	84.1	69.4	70.1
ACAGCGAATGGACCTACGTGGCCTT	155	88.9	89.1	75.9	75.3
AGCAAGTCGAGCAGGGCCTACGTTT	163	89.4	89.5	76.2	75.8
GCGAGCGACAGGTTACTTGGCTGAT	110	87.4	88.0	74.6	74.3
AAAGGTGTCGCGGAGAGTCGTGCTG	158	90.6	90.8	77.6	77.4
ATGGGTGGGAGCCTCGGTAGCAGCC	162	90.7	90.8	78.0	78.6
CAGTGGGCTCCTGGGCGTGCTGGTC	158	92.2	92.3	80.0	80.6
GCCAACTCCGTCGCCGTTCGTGCGC	154	92.2	92.4	81.0	80.7
ACGGGTCCCCGCACCGCACCGCCAG	159	95.8	96.0	84.6	85.2

Table S-2. (cont.)

	1 M $[Na^+]^a$			69 mM $[Na^+]^b$	
DNA Sequence (5' to 3')	Exper. C_t^1 (μM)	Exper. T_m^1 (°C) at C_t^1	Exper. T_m^0 (°C) at C_t = 180 μM	Exper. T_m (°C)	Predicted T_m (°C)[c]
30-mers					
TTATGTATTAAGTTATATAGTAGTAGTAGT	158	74.1	74.2	57.4	57.0
ATTGATATCCTTTTCTATTCATCTTTCATT	152	78.0	78.2	60.4	60.9
AAAGTACATCAACATAGAGAATTGCATTTC	156	80.9	81.0	64.6	64.4
CTTAAGATATGAGAACTTCAACTAATGTGT	163	79.6	79.7	63.5	63.2
CTCAACTTGCGGTAAATAAATCGCTTAATC	161	83.0	83.1	67.5	67.2
TATTGAGAACAAGTGTCCGATTAGCAGAAA	160	84.2	84.3	68.1	68.3
GTCATACGACTGAGTGCAACATTGTTCAAA	156	84.7	84.9	69.8	69.2
AACCTGCAACATGGAGTTTTTGTCTCATGC	145	85.8	85.9	71.5	70.7
CCGTGCGGTGTGTACGTTTTATTCATCATA	149	85.4	85.6	71.2	70.3
GTTCACGTCCGAAAGCTCGAAAAAGGATAC	150	86.4	86.6	71.1	71.7
AGTCTGGTCTGGATCTGAGAACTTCAGGCT	141	87.4	87.6	73.1	73.1
TCGGAGAAATCACTGAGCTGCCTGAGAAGA	120	87.4	87.8	72.9	73.3

Table S-2. (cont.)

	1 M $[Na^+]^a$			69 mM $[Na^+]^b$	
DNA Sequence (5' to 3')	Exper. C_t^1 (μM)	Exper. T_m^1 (°C) at C_t^1	Exper. T_m^0 (°C) at C_t = 180 μM	Exper. T_m (°C)	Predicted T_m (°C)[c]
CTTCAACGGATCAGGTAGGACTGTGGTGGG	160	87.7	87.8	74.5	74.2
ACGCCCACAGGATTAGGCTGGCCCACATTG	158	91.5	91.6	78.3	78.2
GTTATTCCGCAGTCCGATGGCAGCAGGCTC	159	92.0	92.1	78.1	78.7
TCAGTAGGCGTGACGCAGAGCTGGCGATGG	149	92.6	92.8	80.0	79.8
CGCGCCACGTGTGATCTACAGCCGTTCGGC	147	92.0	92.2	80.4	79.7
GACCTGACGTGGACCGCTCCTGGGCGTGGT	152	93.7	93.9	81.8	81.8
GCCCCTCCACTGGCCGACGGCAGCAGGCTC	166	95.4	95.5	83.8	84.3
CGCCGCTGCCGACTGGAGGAGCGCGGGACG	156	96.5	96.6	85.4	85.8

[a] T_m^0 was estimated from T_m^1 by equation [2]. [b] In 69 mM buffer, DSC experiments were conducted at the same total single strand concentration, C_t = (180 ± 10) μM. [c] Melting temperatures in 69 mM Na^+ buffer were predicted from experimentally measured T_m^0 in 1 M Na^+ using Owczarzy et al. T_m salt correction [22].

Supporting Information for World Wide Web Edition

Table S-3. Analysis of T_m predictions using new T_m salt correction function [22] and data set of 92 DNA duplex oligomers from Table 2. Subsets of DNA duplex oligomers of various lengths were examined. Accuracy of T_m predictions was calculated for each subset using equations [3] and [4].

Number of Base Pairs in DNA Duplexes	$\lvert\langle T_m \rangle\rvert_{AVE}$ (°C)	χ^2_r
10,11	0.7	8.1
15	0.5	4.9
20	0.6	6.2
25	0.4	2.4
30	0.4	2.5

Supporting Information for World Wide Web Edition

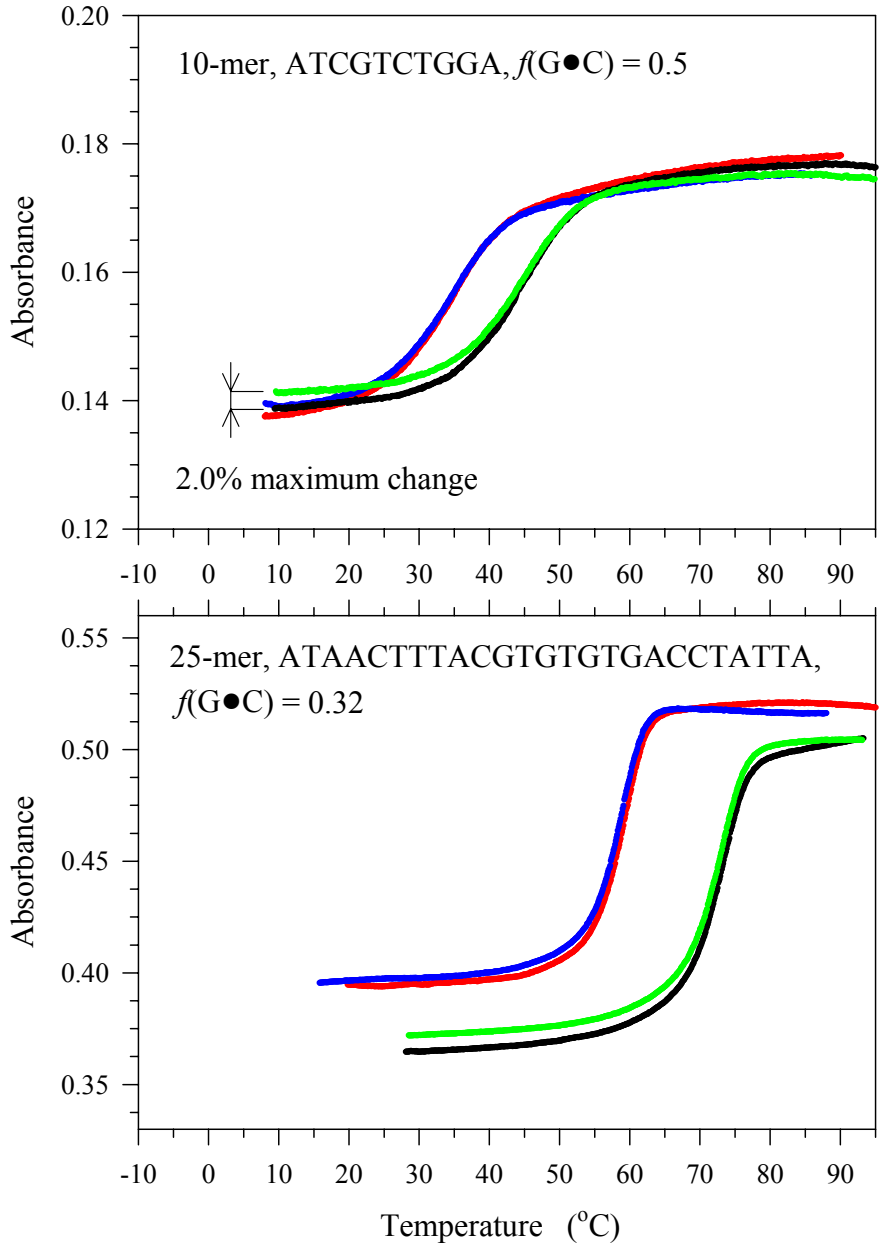

Figure S-1. Typical melting curves for two DNA duplex oligomers in 69 mM Na^+ buffer (●, ●) and 1M Na^+ buffer (●, ●). Both heating (denaturation) curves (●, ●) and cooling (renaturation) curves (●, ●) overlap within experimental errors. The results demonstrate reproducibility of melting curves and negligible evaporation during experiments.