Sequence-Dependent Thermodynamic Parameters for Locked Nucleic Acid (LNA)−DNA Duplex Formation†

Patricia M. McTigue,‡,§ Raymond J. Peterson,*,# and Jason D. Kahn*,‡

Department of Chemistry and Biochemistry, University of Maryland, College Park, College Park, Maryland 20742-2021, and Celadon Laboratories, Inc., Technology Growth Center, 6525 Belcrest Road, Suite 500, Hyattsville, Maryland 20782

Received November 5, 2003; Revised Manuscript Received February 26, 2004

ABSTRACT: The design of modified nucleic acid probes, primers, and therapeutics is improved by considering their thermodynamics. Locked nucleic acid (LNA) is one of the most useful modified backbones, with incorporation of a single LNA providing a substantial increase in duplex stability. In this work, the hybridization $\Delta H°$, $\Delta S°$, and melting temperature (T_M) were measured from absorbance melting curves for 100 duplex oligonucleotides with single internal LNA nucleotides on one strand, and the results provided $\Delta\Delta H°$, $\Delta\Delta S°$, $\Delta\Delta G°_{37}$, and ΔT_M relative to reference DNA oligonucleotides. LNA pyrimidines contribute more stability than purines, especially $\mathbf{A^L}$, but there is substantial context dependence for each LNA base. Both the 5′ and 3′ neighbors must be considered in predicting the effect of an LNA incorporation, with purine neighbors providing more stability. Enthalpy−entropy compensation in $\Delta\Delta H°$ and $\Delta\Delta S°$ is observed across the set of sequences, suggesting that LNA can stabilize the duplex by either preorganization or improved stacking, but not both simultaneously. Singular value decomposition analysis provides predictive sequence-dependent rules for hybridization of singly LNA-substituted DNA oligonucleotides to their all-DNA complements. The results are provided as sets of $\Delta\Delta H°$, $\Delta\Delta S°$, and $\Delta\Delta G°_{37}$ parameters for all 32 of the possible nearest neighbors for LNA+DNA:DNA hybridization (5′ M$\mathbf{X^L}$ and 5′ $\mathbf{X^L}$N, where M, N, and X = A, C, G, or T and $\mathbf{X^L}$ represents LNA). The parameters are applicable within the standard thermodynamic prediction algorithms. They provide T_M estimates accurate to within 2 °C for LNA-containing oligonucleotides, which is significantly better accuracy than previously available.

DNA and RNA analogues and modified nucleic acids with covalently attached ligands can provide more sensitive, specific, and robust nucleic acid tests than can be obtained from unmodified DNA, and they are becoming more commonly used for probes and primers and for antisense or catalytic therapeutics. Modifications such as the minor groove binder (MGB)[1] (*1*), propynl dU and dC (*2*), and locked nucleic acid (LNA, reviewed in refs *3* and *4*) increase the stability of duplex nucleic acids, which allows shorter primers and probes that avoid secondary structure, that increase allelic discrimination, and that are of manageable length in A:T-rich genes. Universal bases such as deoxyinosine (*5*) and 5-nitroindole (*6*) allow placing a primer or probe over a known single nucleotide polymorphism (SNP) or in a consensus sequence of closely related strains. Phosphorothioates increase nuclease resistance (*7*). 8-Aza-7-deazaguanine ("Super G") reduces guanine self-association (*8*). Peptide nucleic acid (PNA) maintains stability and specificity in low ionic strength solutions, and can displace DNA−DNA hybrids (*9*).

Accurate predictions of nucleic acid hybridization thermodynamics are fundamental to primer and probe design and to the prediction of RNA and DNA secondary structure (*10*). For natural RNA (*11*), DNA (*12*, *13*), and RNA:DNA (*14*) hybrids, there is an extensive literature of predictive rules for hybridization enthalpy and entropy. In contrast to simple formulas based on GC content, these nearest-neighbor models explicitly account for the sequence dependence of nucleic acid hybridization (*15*). The models additively partition the total entropy ($\Delta H°$), enthalpy ($\Delta S°$), and free energy ($\Delta G°_{37} = \Delta H° - 310.15\ \text{K}\ \Delta S°$) of a sequence across one or more intercept variables and as many additional variables as are needed to represent each unique dinucleotide that is possible given the chemical alphabets of the sequence and its complement. The intercepts account for duplex initiation and symmetry, and the dinucleotide terms account for the marginal contribution to helix propagation due to hydrogen bonding and stacking upon additional base pair formation. For example, a recent parameter set for RNA−RNA duplex hybridization (*11*) includes helix initiation terms for $\Delta H°$ and $\Delta S°$, a correction term for each terminal AU base pair that correctly accounts for hydrogen bonds, a symmetry correction for $\Delta S°$, and $\Delta H°$ and $\Delta S°$ values for each of the 10 unique nearest neighbors. The parameter estimates are statistically robust because the number of test

† Supported by grants from the Maryland Industrial Partnerships (MIPs) Foundation.

* To whom correspondence should be addressed. E-mail: peterson@celadonlabs.com or jdkahn@umd.edu.

‡ University of Maryland.

Celadon Laboratories.

§ Present address: Accenture, One Financial Plaza, Suite 16, Hartford, CT 06103.

[1] Abbreviations: RSS, residual sum of squares; SVD, singular value decomposition; LNA, locked nucleic acid; PNA, peptide nucleic acid; T_M, melting temperature; RMSD, root-mean-square deviation; Y = pyrimidine, C or T; R = purine, A or G; SNP, single-nucleotide polymorphism.

10.1021/bi035976d CCC: $27.50
Published on Web 04/15/2004

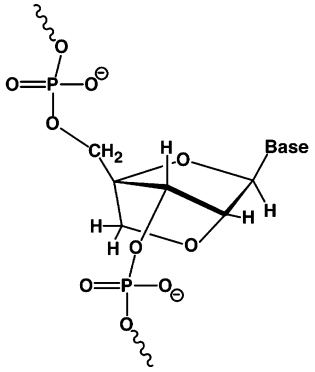

FIGURE 1: The structure of a locked nucleic acid (LNA) nucleotide. The figure illustrates that LNA is readily incorporated into a standard DNA backbone.

sequences is well in excess of the number of model parameters, and the inclusion of each parameter has been shown to improve the accuracy of predictions significantly. The database of available thermodynamic values has been expanded to include consideration of salt effects (*12*), mismatched bases (*16*), dangling ends (*17*), and special tertiary structure elements such as tetraloops (*18*).

A primary function of nucleic acid test design software is to predict the stability of candidate primers and probes using a nearest-neighbor model and the other available parameters. Stability predictions are essential to the computation of likely secondary structures. For example, in PCR primer design it is advantageous to design primer pairs with minimal self-structure, minimal potential for primer−dimer formation, and similar melting temperatures (T_M's). The applications of modified chemistries would be enhanced by the availability of detailed thermodynamic information for them. It would be ideal to be able to predict the stability of a primer or probe of any length, any sequence, and any chemical composition. Stability prediction is difficult because of nearest neighbor or longer-range effects from any nearby modifications.

Nearest-neighbor models have been presented with model parameters for deoxyinosine (*5*), isoG and isoC (*19*), 5-nitroindole (*6*), and PNA (*20*−*22*). However, we are not aware of any systematic investigation of a full set of nearest-neighbor dinucleotide parameters for a DNA or RNA analogue incorporated into a DNA backbone. We have used the LNA modification as a test case for a general framework for sequence design and data reduction that can be used to establish such a parameter set for any modification. We use standard absorbance melting curve methods and build on existing DNA parameters. Robust estimation of model parameters for chemically heterogeneous duplexes presents additional challenges: the number of nearest neighbors is potentially much larger, the oligonucleotides are more expensive and difficult to handle, and there may be long-range effects for some modifications. Our approaches to these challenges may be of utility in all applications of nucleic acid analogues.

LNA (Figure 1) is a particularly attractive backbone modification (*3*, *4*) introduced by the Wengel (*23*) and Imanishi (*24*) groups. The LNA sugar is a bicyclic ribose derivative with a bridging methylene group between C4′ and O2′, which confers rigidity and a preference for the C3′ endo sugar pucker characteristic of A-form RNA. A DNA duplex with a single LNA in each strand adopts the A form in the crystal (*25*), but NMR suggests that sugar pucker changes are more localized to the vicinity of the LNA residue (*26*). The LNA modification confers the largest increase in base pairing stability of any commercially available DNA or RNA analogue: a positive ΔT_M of up to 10 °C has been obtained upon substitution of one LNA nucleotide in DNA, and multiple incorporations give larger effects (*4*, *23*, *27*−*29*). LNA phosphoramidites are available with all four bases, A, C, G, and T. LNA can be readily incorporated into LNA/DNA "mixmers," it is soluble and stable, it is resistant to nuclease digestion, and it can form triple helical DNA (*30*). It has been incorporated into antisense agents (reviewed in ref *31*) and in microarrays (*4*, *32*). LNA's increased stability makes it an ideal candidate to allow for shorter, therefore more specific, primers and probes.

The literature on the effects of LNA has been couched mainly in terms of the ΔT_M per LNA substitution. This is straightforward to measure, but it does not allow for quantitative predictions of the effect of an LNA in a new sequence context, since the ΔT_M obtained depends on the $\Delta H°$ and $\Delta S°$ of the reference DNA. Longer oligonucleotides or multiple LNAs will give much smaller ΔT_M values per LNA than a typical short reference sequence because the fractional changes in $\Delta H°$ and $\Delta S°$ per LNA are much smaller. The thermodynamic origin of LNA's enhanced base pairing stability has not been clearly delineated, although it has been suggested that decreased entropy of duplex formation and improved stacking in the duplex both play a part.

Hybridization rules analogous to the nearest-neighbor rules for DNA would aid in the design of all LNA-containing agents by providing more accurate sequence-dependent T_M estimates. There are design tools available for LNA, but the underlying principles used are not clear or transportable to other algorithms (*32*). In this work, we have measured the $\Delta\Delta H°$ and $\Delta\Delta S°$ increments attributable to LNA substitution, for the hybridization of DNA oligonucleotides that contain a single LNA mismatch to their all-DNA complements. The resulting parameter set is presented as a set of $\Delta\Delta H°$ and $\Delta\Delta S°$ values for the 32 possible LNA−DNA nearest-neighbors. The results show that LNA pyrimidines contribute much more stability than purines (in terms of $\Delta\Delta G°_{37}$, the order is LNA-C > T > G ≫ A), but there is substantial context dependence for each LNA base, which must be considered in making accurate predictions. There is substantial enthalpy−entropy compensation among sequences, so the thermodynamic origin of stabilization is found to be either enthalpic or entropic, but not both at the same time. The nearest-neighbor parameters provide T_M estimates for LNA-substituted oligonucleotides that should be accurate to within about 2 °C, and can be readily incorporated into standard prediction algorithms.

EXPERIMENTAL PROCEDURES

Oligonucleotide Synthesis and Purification. LNA-containing oligonucleotides were synthesized, deprotected, and desalted by PrOligo, LLC. DNA oligonucleotides were from Invitrogen. Oligonucleotides were ethanol precipitated and their concentrations were determined by absorbance, using the nearest-neighbor rules for single-strand DNA absorbance. Three micrograms of each was analyzed on 15% (40:1 bis:

acrylamide) denaturing polyacrylamide gels and visualized with Stains-All (Aldrich). If there were visible impurity bands, the oligonucleotide was then purified on a preparative 15% gel or resynthesized; otherwise, it was used directly. For gel purification, gel slices visualized by UV shadowing were eluted into 50 mM NaOAc (pH 7), 1 mM EDTA, and the eluant was ethanol precipitated. Oligonucleotides were stored in 10 mM TrisHCl (pH 7.5), 1 mM EDTA. After prolonged storage (months), it was necessary to remeasure the oligonucleotide concentration before use.

Absorbance Melting Curve Procedures. Absorbance versus temperature curves were measured by standard methods, using a Varian Cary 100 UV−visible spectrophotometer equipped with a 12-position cell holder and Peltier heating/cooling unit. A reference DNA−DNA duplex and several corresponding LNA-substituted duplexes were melted in the same run. Absorbance was monitored at 260 nm, using single-beam mode and subtracting the absorbance from a matched cuvette containing buffer. The buffer was 1 M NaCl, 10 mM Na cacodylate, and 0.5 mM Na_2EDTA, pH 7.0 (total $[Na^+] = 1011$ mM). Buffer was degassed thoroughly before each use. Self-masking quartz cuvettes with 1-cm path lengths were used, except a 2-mm path length was used for 50 μM samples. Temperature was monitored by a thermocouple probe in a reference cuvette containing the sample buffer.

Complementary oligonucleotides for forming duplexes were mixed at a 1:1 concentration ratio. The total single-strand concentration (C_T) was 5 μM, except $C_T = 1$ μM was used for the LT4 and LT5 series. To check the accuracy of C_T, the high-temperature absorbance was extrapolated back to 21 °C using the slope of the ssDNA absorbance, and the result was compared to the expected value for the sum of the single-strand components. Several duplexes were melted at $C_T = 1$, 5, and 50 μM to spot-validate that $\Delta H°$ and $\Delta S°$ are independent of concentration. Melting curves for several oligonucleotides were obtained at 1011, 115, and 21 mM total $[Na^+]$.

Oligonuclotides were annealed in the cuvette: samples were heated to 85 °C at 3 °C/min, maintained at 85 °C for 5 min, and cooled to 0 °C at a rate of 3 °C/min. The analyzed data was then acquired while heating to 95 °C at 0.8 °C/min. Measurements were recorded at intervals of 1 °C. Control experiments at slower heating rates showed that equilibrium was attained at each temperature. In addition, data were acquired for several duplexes with temperature decreasing from 95 °C at 0.8 °C/min to check reversibility. The sample chamber was purged with nitrogen whenever it was below 30 °C to prevent condensation.

Melting Curve Data Analysis. The entire melting curve for each oligonucleotide was fit by the van't Hoff expression for two-state melting. Briefly, the equilibrium constant describing the hybridization of non-self-complementary oligonucleotides L and D is given by

$$K_{eq} = \frac{[L\cdot D]}{[L][D]} = \frac{\alpha C_T/2}{[(1-\alpha)C_T/2]^2} = \frac{2\alpha}{(1-\alpha)^2 C_T} \quad (1)$$

where α is the fraction of the total DNA strands found in double-stranded form. Solving the equation for α and choosing the physically realizable root, we have

$$\alpha = \frac{1 + C_T K_{eq} - \sqrt{1 + 2C_T K_{eq}}}{C_T K_{eq}} \quad (2)$$

The van't Hoff expression for K_{eq} is

$$K_{eq} = \exp(-\Delta G°/RT) = \exp(-\Delta H°/RT + \Delta S°/R) \quad (3)$$

where the enthalpy ($\Delta H°$) and entropy ($\Delta S°$) changes refer to hybridization, so both are negative. Substitution of (eq 3) into (eq 2) gives $\alpha(C_T, \Delta H°, \Delta S°, T)$, not shown.

The absorbance A_{260} at temperature T is given by

$$A_{260}(T) = \alpha[A_{260}(T_{min}) + m_{ds}(T - T_{min})] + (1-\alpha)[A_{260}(T_{max}) - m_{ss}(T_{max} - T)] \quad (4)$$

where $A_{260}(T_{min})$ and $A_{260}(T_{max})$ are the measured absorbances at temperatures T_{min} and T_{max}, and m_{ss} and m_{ds} contain the temperature dependence of the extinction coefficients of the ssDNA and dsDNA, respectively. $A_{260}(T_{min})$ and $A_{260}(T_{max})$ are taken as the average absorbances for the first five and last five temperatures, respectively, and m_{ds} and m_{ss} are the average slopes for the first 20 and last 20 temperatures, respectively, unless the melting transition overlaps these ranges. Substitution of $\alpha(C_T, \Delta H°, \Delta S°, T)$ into (eq 4) gives $A_{260}(T)$ as a function of $A_{260}(T_{min})$, $A_{260}(T_{max})$, m_{ss}, m_{ds}, C_T, $\Delta H°$, and $\Delta S°$. C_T is measured, and $A_{260}(T_{min})$, $A_{260}(T_{max})$, m_{ss}, and m_{ds} are obtained directly from the data. This leaves $\Delta H°$ and $\Delta S°$ to be fit to the experimental $A_{260}(T)$ as described below. Finally, the melting temperature (T_M) is given by

$$T_M = \frac{\Delta H°}{\Delta S° + R\ln(C_T/4)} \quad (5)$$

Typical melting curves and analysis results are shown in Figure 2. In practice, the derivative of the absorbance with respect to temperature (dA_{260}/dT) from 0 to 95 °C was fit by m_{ss}, m_{ds}, $\Delta H°$, and $\Delta S°$. Fitting to the derivative makes any deviations from apparent two-state behavior much more obvious. For the longer oligonucleotides, the LT4 and LT5 sets, some "premelting" was observed before strand separation. This non-two-state behavior is more pronounced for LNA-substituted DNA (if more than 5% of the change in absorbance was observed to be non-two-state for the DNA alone, the sequence was discarded and redesigned). In principle, non-two-state melting requires a statistical mechanical treatment; in practice, data for oligonucleotides exhibiting non-two-state melting are generally just discarded. However, we observed that in every case the high-temperature side of the derivative melting curve was well-behaved. Since our main interest is in applications, in which complete dissociation is the key event, we retained the data. Fitting the shallower dA_{260}/dT dependence of non-two-state data to the two-state equation gives artifactually low absolute values for $\Delta H°$ and $\Delta S°$. In these cases, the premelting derivative data, generally the 20 °C below the temperature at one-third of the maximum in dA_{260}/dT, were eliminated from the fit. Comparison of the experimental A_{260} vs T curves to those predicted from the derived $\Delta H°$ and $\Delta S°$ showed that the premelting accounted for no more than 15% of the total absorbance change in going from duplex to single strand. The worst case of non-two-state behavior is shown in Figure

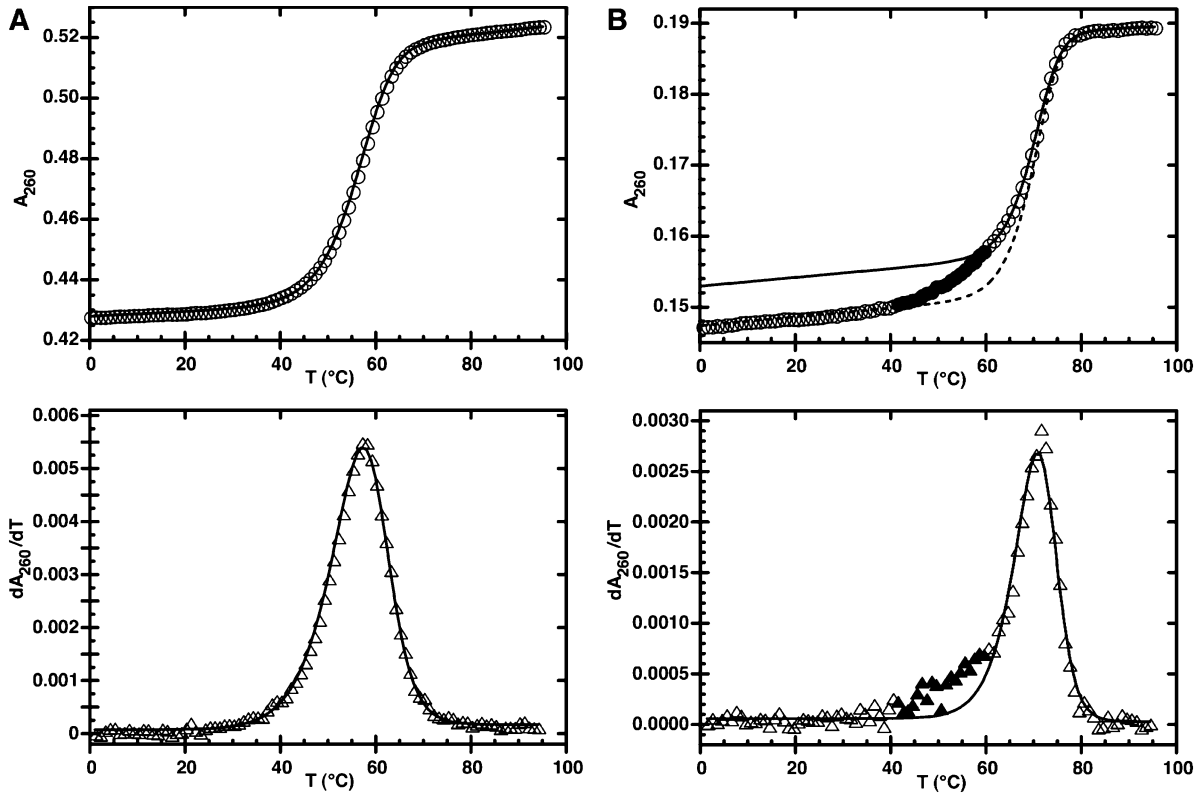

FIGURE 2: Absorbance melting curves. Absorbance (circles) versus temperature is shown in the top panel and its derivative (triangles) in the bottom panel. Solid lines are the van't Hoff fit to the derivative data. Panel A illustrates two-state melting (for oligonucleotide L5d). Panel B, for LT4d, exhibits multistate character. The filled symbols show the data omitted in calculating the van't Hoff fit. The dashed line shows the rescaled absorbance, which would have been observed if the melting were pure two-state.

2B: it is clear that most of the data are adequately described, especially near the T_M.

To obtain $\Delta H°$ and $\Delta S°$ and their uncertainties, we varied $\Delta H°$ systematically, determined the corresponding best-fit $\Delta S°$ using Microsoft Excel's Solver, and calculated the residual sum of squares (RSS) between the observed and predicted dA_{260}/dT for each combination of $\Delta H°$ and $\Delta S°$. The $\Delta H°$ and $\Delta S°$ values in Table 1 minimize the RSS, and their estimated uncertainties correspond to the values that increase the RSS by 10%, which gives a visibly though subtly poorer fit (not shown). The estimated uncertainties in $\Delta H°$ and $\Delta S°$ are about $\pm$ 2−5% of the best-fit values. This is larger than the measured deviation between repeated experiments (typically < 2%), and comparable to the differences among $\Delta H°$ values determined at different concentrations. It is, however, smaller than the standard deviations typically observed for $\Delta H°$ and $\Delta S°$ in absorbance melting experiments (*11*, *13*). In general, the RMSD in predicting $\Delta H°$ or $\Delta S°$ from nearest-neighbor parameters is about 8%, so the precision of $\Delta H°$ and $\Delta S°$ for an individual sample is not particularly relevant to predictive accuracy (*13*).

Near the optimal $\Delta H°$ and $\Delta S°$ values, the RSS is a quadratic function of the deviation in $\Delta H°$ from its optimum, which means that setting a larger error threshold (for example, 120% of the optimal RSS) would increase all of the calculated uncertainties proportionately (by a factor of $\sqrt{2}$), which propagates into the analysis below as a constant multiplier (1/2) of χ^2 values. Thus, while the uncertainties estimated from the van't Hoff analysis may be too low, the relative weighting of the experiments should be independent of the actual absolute errors. We take the estimated uncertainties derived in this way to be estimates for the standard deviations $\sigma(\Delta H°)$ and $\sigma(\Delta S°)$, which would otherwise be obtained from a $1/T_M$ versus $\ln(C_T)$ analysis or from averaging the $\Delta H°$ and $\Delta S°$ values obtained at different values of C_T. The van't Hoff fitting and the dependence of $1/T_M$ on C_T have previously been shown to give equivalent results (*13*).

Errors in $\Delta H°$ and $\Delta S°$ are highly correlated, so calculated values for the change in free energy $\Delta G°_{37} = \Delta H° - 310\ \text{K} \times \Delta S°$ and the T_M are much more precise (typically within <2% and <1 °C, respectively). We estimate the standard deviation in $\Delta G°_{37}$ as

$$\sigma(\Delta G°_{37}) = \Delta G°_{37,\text{min}} - \Delta G°_{37,\text{opt}} \tag{6}$$

where $\Delta G°_{37,\text{min}} = \Delta H°_{\text{min}} - 310\ \text{K} \times \Delta S°_{\text{min}}$ is calculated at the lower limits of the best-fit $\Delta H°$ and $\Delta S°$, and $\Delta G°_{37,\text{opt}}$ is calculated at the optimal values. The results are close to those obtained using $\sigma^2(\Delta G°_T) = (\sigma(\Delta H°) - T\sigma(\Delta S°))^2$ as in Xia et al. (*11*). Finally, following (*11*) we estimate the variance in T_M as follows:

$$\sigma^2(T_M) = \frac{\sigma^2(\Delta G°_{T_M})}{(\Delta H°)^2} T_M{}^2 \tag{7}$$

The accuracy of the experimental thermodynamics can be assessed by comparing measurements for the reference DNA: DNA duplexes to predictions from the unified thermodynamic parameters for DNA (*12*). For the 24 DNA duplexes used, the RMSDs for predicted vs observed $\Delta H°$, $\Delta S°$, and T_M are 9.5 kcal/mol (11%), 29 cal/mol K (12.5%), and 1.6

Table 1: Experimental Thermodynamic Values for LNA-Containing DNA Oligonucleotides[a]

"training data set" oligonucleotides			results from absorbance melting curves				changes due to LNA substitution			
name	sequence (5′ to 3′)	length, trinuc	$\Delta H°$ (kcal/ mol)	$\Delta S°$ (e.u.)	$\Delta G°_{37}$ (kcal/mol)	T_M (°C)	ΔT_M (°C)	$\Delta\Delta H°$ (kcal/mol)	$\Delta\Delta S°$ (e.u.)	$\Delta\Delta G°_{37}$ (kcal/mol)
D1	GTCGAACAGC	10	−80.6 ± 1.6	−221 ± 5	−12.1 ± 0.1	51.8				
L1a	GTC**G**^{L}AACAGC	C**G**^{L}A	−74.2 ± 2.8	−202 ± 9	−11.6 ± 0.1	51.2	−0.6	6.5 ± 3.2	19.4 ± 9.9	0.43 ± 0.16
L1b	GTCG**A**^{L}ACAGC	G**A**^{L}A	−76.2 ± 1.5	−206 ± 5	−12.3 ± 0.1	53.8	2.0	4.4 ± 2.2	14.9 ± 6.8	−0.24 ± 0.12
L1c	GTCGA**A**^{L}CAGC	A**A**^{L}C	−77.2 ± 1.3	−209 ± 4	−12.4 ± 0.1	53.9	2.1	3.4 ± 2.1	12.0 ± 6.4	−0.32 ± 0.11
L1d	GTCGAA**C**^{L}AGC	A**C**^{L}A	−75.5 ± 1.1	−202 ± 3	−12.8 ± 0.1	56.1	4.3	5.1 ± 2.0	18.7 ± 6.0	−0.71 ± 0.11
D2	ATCTATCCGGC	11	−80.3 ± 1.0	−219 ± 3	−12.3 ± 0.1	52.7				
L2a	ATC**T**^{L}ATCCGGC	C**T**^{L}A	−79.3 ± 2.6	−213 ± 8	−13.3 ± 0.2	57.3	4.6	1.0 ± 2.7	6.5 ± 8.3	−1.01 ± 0.18
L2b	ATCTA**T**^{L}CCGGC	TA**T**^{L}T	−76.8 ± 2.0	−207 ± 6	−12.5 ± 0.1	54.5	1.8	3.5 ± 2.2	12.2 ± 6.7	−0.23 ± 0.13
L2c	ATCTAT**C**^{L}CGGC	AT**C**^{L}C	−73.9 ± 2.4	−198 ± 7	−12.5 ± 0.2	55.4	2.7	6.4 ± 2.6	21.5 ± 7.9	−0.27 ± 0.16
L2d	ATCTATC^{L}CGGC	TC^{L}C	−74.0 ± 1.3	−198 ± 4	−12.7 ± 0.1	56.1	3.5	6.3 ± 1.6	21.9 ± 4.9	−0.44 ± 0.10
L2e	ATCTATC**C**^{L}GGC	C**C**^{L}G	−76.6 ± 1.0	−206 ± 3	−12.8 ± 0.1	55.9	3.2	3.7 ± 1.4	13.6 ± 4.3	−0.53 ± 0.08
D3	CGCTGTTACGC	11	−79.3 ± 1.2	−214 ± 4	−13.0 ± 0.1	56.1				
L3a	CGC**T**^{L}GTTACGC	C**T**^{L}G	−82.9 ± 1.4	−221 ± 4	−14.2 ± 0.1	60.5	4.3	−3.6 ± 1.8	−7.6 ± 5.5	−1.22 ± 0.13
L3b	CGCT**G**^{L}TTACGC	T**G**^{L}T	−81.0 ± 1.8	−217 ± 5	−13.6 ± 0.1	58.4	2.3	−1.7 ± 2.1	−3.4 ± 6.3	−0.63 ± 0.15
L3c	CGCTG**T**^{L}TACGC	G**T**^{L}T	−81.1 ± 1.9	−217 ± 6	−13.8 ± 0.1	59.2	3.1	−1.8 ± 2.2	−3.2 ± 6.7	−0.81 ± 0.16
L3d	CGCTGT**T**^{L}ACGC	T**T**^{L}A	−78.2 ± 2.4	−208 ± 7	−13.6 ± 0.2	59.2	3.1	1.0 ± 2.7	5.3 ± 8.0	−0.62 ± 0.19
L3e	CGCTGTTA**L**CGC	TA^{L}C	−78.7 ± 1.4	−211 ± 4	−13.3 ± 0.1	57.6	1.5	0.6 ± 1.8	2.8 ± 5.6	−0.30 ± 0.12
D4	GGACCTCGAC	10	−77.2 ± 1.2	−209 ± 4	−12.3 ± 0.1	53.6				
L4a	GGA**C**^{L}CTCGAC	A**C**^{L}C	−74.5 ± 1.4	−198 ± 4	−13.3 ± 0.1	58.7	5.1	2.7 ± 1.8	11.6 ± 5.6	−0.94 ± 0.12
L4b	GGAC**C**^{L}TCGAC	C**C**^{L}T	−74.1 ± 1.6	−197 ± 5	−13.0 ± 0.1	57.9	4.2	3.1 ± 2.0	12.4 ± 6.0	−0.74 ± 0.13
L4c	GGACC**T**^{L}CGAC	C**T**^{L}C	−74.7 ± 1.1	−200 ± 3	−12.6 ± 0.1	55.5	1.9	2.4 ± 1.7	8.8 ± 5.0	−0.29 ± 0.10
L4d	GGACCT**C**^{L}GAC	TC^{L}G	−75.0 ± 1.4	−200 ± 4	−13.1 ± 0.1	57.8	4.2	2.2 ± 1.9	9.6 ± 5.6	−0.78 ± 0.12
D5	CCATTGCTACC	11	−82.4 ± 1.1	−227 ± 3	−11.9 ± 0.1	50.9				
L5a1	CC**A**^{L}TTGCTACC	C**A**^{L}T	−75.8 ± 1.6	−207 ± 5	−11.6 ± 0.1	50.6	−0.3	6.6 ± 1.9	20.1 ± 5.9	0.35 ± 0.09
L5a	CCAT**T**^{L}TGCTACC	AT**T**^{L}T	−77.2 ± 1.9	−209 ± 6	−12.3 ± 0.1	53.7	2.8	5.2 ± 2.1	18.0 ± 6.5	−0.42 ± 0.12
L5b	CCATT**G**^{L}GCTACC	TT^{L}G	−78.6 ± 1.5	−213 ± 5	−12.6 ± 0.1	54.4	3.5	3.8 ± 1.8	14.2 ± 5.6	−0.64 ± 0.10
L5c	CCATT**G**^{L}CTACC	TG^{L}C	−77.3 ± 1.1	−210 ± 3	−12.1 ± 0.1	52.5	1.6	5.1 ± 1.5	16.9 ± 4.7	−0.15 ± 0.07
L5d	CCATTG**C**^{L}TACC	GC^{L}T	−79.4 ± 1.1	−214 ± 4	−13.0 ± 0.1	56.0	5.1	3.0 ± 1.6	13.0 ± 4.8	−1.04 ± 0.08
D6	GGTGCCAA	8	−51.6 ± 0.9	−138 ± 3	−8.65 ± 0.02	38.7				
L6a	GGT**G**^{L}CCAA	GT**G**^{L}G	−52.3 ± 1.6	−137 ± 5	−9.7 ± 0.1	44.8	6.2	−0.7 ± 1.8	1.0 ± 5.6	−1.01 ± 0.06
L6b	GGT**G**^{L}CCAA	TG^{L}C	−50.1 ± 1.4	−132 ± 4	−9.31 ± 0.04	42.9	4.2	1.5 ± 1.6	6.8 ± 5.2	−0.66 ± 0.04
L6b2	GGTG**C**^{L}CAA	GC^{L}C	−52.5 ± 2.0	−138 ± 6	−9.8 ± 0.1	45.9	7.3	−0.9 ± 2.2	0.9 ± 6.8	−1.19 ± 0.07
L6c	GGTGC**C**^{L}AA	CC^{L}A	−49.6 ± 1.5	−130 ± 5	−9.44 ± 0.04	43.8	5.1	2.0 ± 1.7	8.9 ± 5.4	−0.78 ± 0.05
D7	GGAACAAGATGC	12	−90.8 ± 1.0	−250 ± 3	−13.3 ± 0.1	54.9				
L7a	GGAAC**A**^{L}AGATGC	C**A**^{L}A	−88.5 ± 1.3	−242 ± 4	−13.4 ± 0.1	55.8	0.9	2.3 ± 1.6	7.7 ± 4.9	−0.11 ± 0.10
L7b	GGAACAA**L**GATGC	AA^{L}G	−89.3 ± 2.2	−243 ± 7	−13.9 ± 0.2	57.6	2.6	1.5 ± 2.4	6.6 ± 7.2	−0.59 ± 0.16
L7c	GGAACAA**G**^{L}ATGC	A**G**^{L}A	−89.4 ± 1.7	−242 ± 5	−14.2 ± 0.1	58.6	3.7	1.4 ± 1.9	7.3 ± 5.8	−0.87 ± 0.13
L7d	GGAACAAG**A**^{L}TGC	G**A**^{L}T	−89.8 ± 1.1	−244 ± 3	−14.0 ± 0.1	57.6	2.7	1.0 ± 1.4	5.2 ± 4.4	−0.64 ± 0.09
D8	CACGGCTC	8	−58.5 ± 1.3	−158 ± 4	−9.7 ± 0.1	44.0				
L8a1	CA**C**^{L}GGCTC	A**C**^{L}G	−66.5 ± 3.6	−179 ± 11	−11.0 ± 0.2	49.8	5.7	−8.0 ± 3.8	−21 ± 12	−1.33 ± 0.18
L8a	CAC**G**^{L}GCTC	C**G**^{L}G	−59.5 ± 1.6	−159 ± 5	−10.3 ± 0.1	47.4	3.4	−1.0 ± 2.0	−1.2 ± 6.3	−0.63 ± 0.10
L8b	CACG**G**^{L}CTC	G**G**^{L}C	−58.9 ± 2.2	−155 ± 7	−10.7 ± 0.1	49.9	5.9	−0.4 ± 2.6	2.3 ± 7.9	−1.06 ± 0.13
L8c	CACGG**C**^{L}TC	GC^{L}T	−64.4 ± 3.4	−172 ± 11	−11.2 ± 0.2	51.2	7.2	−5.9 ± 3.7	−14 ± 11	−1.53 ± 0.18
D9	GCAGGTCTGC	10	−70.5 ± 1.3	−188 ± 4	−12.1 ± 0.1	54.6				
L9a	GCA**L**GGTCTGC	CA**L**G	−71.9 ± 1.8	−190 ± 5	−12.9 ± 0.1	57.7	3.2	−1.4 ± 2.3	−2.2 ± 7.0	−0.81 ± 0.15
L9b	GCA**G**^{L}GTCTGC	A**G**^{L}G	−76.5 ± 1.4	−202 ± 4	−14.0 ± 0.1	61.5	6.9	−6.1 ± 1.9	−13.7 ± 5.9	−1.90 ± 0.14
L9c	GCAG**G**^{L}TCTGC	G**G**^{L}T	−70.7 ± 2.2	−186 ± 7	−13.0 ± 0.2	58.4	3.8	−0.2 ± 2.6	1.8 ± 7.9	−0.86 ± 0.18
D10	GTAGCGATGTA	11	−75.3 ± 1.2	−205 ± 4	−11.8 ± 0.1	51.9				
L10a	GTA**G**^{L}CGATGTA	A**G**^{L}C	−77.4 ± 1.4	−210 ± 4	−12.4 ± 0.1	53.9	2.0	−2.1 ± 1.8	−4.9 ± 5.6	−0.54 ± 0.10
L10b	GTAG**C**^{L}GATGTA	G**C**^{L}G	−72.8 ± 2.4	−194 ± 7	−12.7 ± 0.2	56.6	4.7	2.6 ± 2.7	11.2 ± 8.2	−0.87 ± 0.17
L10c	GTAGC**G**^{L}ATGTA	C**G**^{L}A	−72.3 ± 1.8	−194 ± 6	−12.0 ± 0.1	53.6	1.7	3.0 ± 2.2	10.4 ± 6.6	−0.21 ± 0.12
L10d	GTAGCGA**T**^{L}GTA	AT^{L}G	−73.7 ± 2.8	−196 ± 8	−13.0 ± 0.2	57.7	5.8	1.6 ± 3.0	9.0 ± 9.1	−1.16 ± 0.20
D11	ACGTCTTCG	9	−57.2 ± 1.6	−154 ± 5	−9.35 ± 0.04	42.4				
L11a1	AC**G**^{L}TCTTCG	C**G**^{L}T	−57.4 ± 1.6	−154 ± 5	−9.46 ± 0.04	43.0	0.6	−0.1 ± 2.3	−0.1 ± 7.1	−0.11 ± 0.06
L11a	ACG**T**^{L}CTTCG	G**T**^{L}C	−58.1 ± 2.0	−155 ± 6	−10.2 ± 0.1	47.0	4.7	−0.9 ± 2.5	−0.2 ± 7.9	−0.85 ± 0.08
L11b	ACGT**C**^{L}TTCG	TC^{L}T	−63.0 ± 1.4	−168 ± 4	−10.8 ± 0.1	49.6	7.2	−5.7 ± 2.2	−13.7 ± 6.7	−1.49 ± 0.08
L11c	ACGTC**T**^{L}TCG	CT^{L}T	−59.1 ± 2.1	−158 ± 7	−10.0 ± 0.1	45.7	3.3	−1.9 ± 2.7	−4.0 ± 8.3	−0.64 ± 0.08
L11e	ACGTCTT^{L}CG	TT^{L}C	−58.9 ± 1.8	−158 ± 6	−10.0 ± 0.1	45.5	3.1	−1.7 ± 2.4	−3.4 ± 7.6	−0.60 ± 0.07
D13	TTGGGAGTAGC	11	−72.6 ± 1.3	−197 ± 4	−11.5 ± 0.1	51.1				
L13a	TT**G**^{L}GGAGTAGC	T**G**^{L}G	−72.3 ± 2.5	−194 ± 8	−12.1 ± 0.1	53.8	2.6	0.4 ± 2.8	3.0 ± 8.5	−0.54 ± 0.16
L13b	TTG**G**^{L}GAGTAGC	G**G**^{L}G	−74.0 ± 2.7	−198 ± 8	−12.6 ± 0.2	55.9	4.8	−1.4 ± 3.0	−0.9 ± 9.0	−1.09 ± 0.18
L13c	TTGG**G**^{L}AGTAGC	G**G**^{L}A	−75.2 ± 1.4	−202 ± 4	−12.7 ± 0.1	55.9	4.8	−2.6 ± 1.9	−4.6 ± 5.9	−1.16 ± 0.11
L13d	TTGGGA**L**GTAGC	GA^{L}G	−76.9 ± 1.4	−206 ± 4	−13.0 ± 0.1	56.9	5.7	−4.3 ± 2.0	−9.1 ± 5.9	−1.47 ± 0.12
L13e	TTGGGA**G**^{L}TAGC	A**G**^{L}T	−69.4 ± 1.9	−184 ± 6	−12.1 ± 0.1	54.8	3.7	3.3 ± 2.4	12.5 ± 7.2	−0.60 ± 0.14
L13f	TTGGGAG**T**^{L}AGC	GT^{L}A	−72.4 ± 2.1	−193 ± 6	−12.6 ± 0.1	56.3	5.1	0.2 ± 2.5	4.2 ± 7.6	−1.07 ± 0.15
L13g	TTGGGAGTA^{L}GC	TA^{L}G	−71.8 ± 1.6	−192 ± 5	−12.3 ± 0.1	54.9	3.8	0.9 ± 2.1	5.3 ± 6.3	−0.76 ± 0.12
L13rv	GCTACTC**C**^{L}CAA	C**C**^{L}C	−70.9 ± 2.2	−189 ± 7	−12.2 ± 0.1	54.8	3.7	1.7 ± 2.6	7.8 ± 7.8	−0.68 ± 0.15
D14	CTAAATAGCG	10	−67.3 ± 1.7	−187 ± 6	−9.14 ± 0.03	40.6				
L14a	CTA^{L}AATAGCG	TA^{L}A	−61.8 ± 2.1	−170 ± 7	−9.17 ± 0.04	41.1	0.5	5.5 ± 2.7	17.8 ± 8.6	−0.03 ± 0.05
L14b	CTAA^{L}ATAGCG	AA^{L}A	−67.2 ± 2.0	−185 ± 6	−9.73 ± 0.05	43.4	2.8	0.1 ± 2.7	2.2 ± 8.4	−0.59 ± 0.06
L14c	CTAA**A**^{L}TAGCG	AA^{L}T	−65.4 ± 2.0	−180 ± 6	−9.56 ± 0.05	42.7	2.1	1.9 ± 2.6	7.3 ± 8.3	−0.42 ± 0.05
L14d	CTAAAT^{L}AGCG	AT^{L}A	−66.4 ± 1.4	−182 ± 4	−10.03 ± 0.04	44.9	4.3	0.8 ± 2.2	5.5 ± 7.0	−0.89 ± 0.05
D15	TGCACGCTA	9	−49.0 ± 1.4	−125 ± 4	−10.1 ± 0.1	48.2				
L15a	TG**C**^{L}ACGCTA	GC^{L}A	−48.1 ± 1.5	−121 ± 5	−10.5 ± 0.1	51.0	2.8	0.9 ± 2.0	4.1 ± 6.2	−0.36 ± 0.08
L15b	TGCA^{L}CGCTA	CA^{L}C	−53.7 ± 1.8	−139 ± 6	−10.5 ± 0.1	49.8	1.6	−4.7 ± 2.3	−13.8 ± 7.0	−0.42 ± 0.10

Table 1: Continued

training data set			results from melting curves				changes due to LNA			
name	sequence (5′ to 3′)	length, trinuc	$\Delta H°$ (kcal/ mol)	$\Delta S°$ (e.u.)	$\Delta G°_{37}$ (kcal/mol)	T_M (°C)	ΔT_M (°C)	$\Delta\Delta H°$ (kcal/mol)	$\Delta\Delta S°$ (e.u.)	$\Delta\Delta G°_{37}$ (kcal/mol)
L15c	TGCAC**G**^{L}CTA	C**G**^{L}C	−53.5 ± 1.6	−138 ± 5	−10.9 ± 0.1	52.1	3.8	−4.5 ± 2.1	−12.1 ± 6.5	−0.77 ± 0.10
D16	ATTTGACTCAG	11	−76.9 ± 1.7	−214 ± 5	−10.7 ± 0.1	46.5				
L16a	AT**T**^{L}TGACTCAG	T**T**^{L}T	−69.0 ± 2.4	−188 ± 8	−10.8 ± 0.1	48.3	1.7	7.9 ± 3.0	25.9 ± 9.2	−0.12 ± 0.12
L16b	ATTT**G**^{L}ACTCAG	T**G**^{L}A	−66.8 ± 1.8	−181 ± 6	−10.6 ± 0.1	47.6	1.0	10.1 ± 2.4	32.3 ± 7.5	0.09 ± 0.09
L16c	ATTTGA**C**^{L}CTCAG	GA**C**^{L}C	−67.1 ± 1.7	−181 ± 5	−11.0 ± 0.1	49.5	2.9	9.8 ± 2.4	32.7 ± 7.4	−0.30 ± 0.10
L16d	ATTTGAC**T**^{L}TCAG	AC**T**^{L}T	−66.0 ± 2.6	−176 ± 8	−11.3 ± 0.1	51.1	4.6	11.0 ± 3.1	37.2 ± 9.5	−0.58 ± 0.14
L16e	ATTTGACT**C**^{L}AG	T**C**^{L}A	−72.3 ± 1.4	−197 ± 4	−11.1 ± 0.1	49.3	2.8	4.6 ± 2.2	16.4 ± 6.8	−0.46 ± 0.09
test data set			**results from melting curves**				**changes due to LNA**			
name	sequence (5′ to 3′)	length, trinuc	$\Delta H°$ (kcal/ mol)	$\Delta S°$ (e.u.)	$\Delta G°_{37}$ (kcal/mol)	T_M (°C)	ΔT_M (°C)	$\Delta\Delta H°$ (kcal/mol)	$\Delta\Delta S°$ (e.u.)	$\Delta\Delta G°_{37}$ (kcal/mol)
DT1	GTGGATCTTTA	11	−66.8 ± 1.7	−182 ± 5	−10.3 ± 0.1	45.8				
LT1a	GT**G**^{L}GATCTTTA	T**G**^{L}G	−69.2 ± 1.4	−188 ± 4	−11.1 ± 0.1	49.1	3.3	−2.4 ± 2.2	−5.4 ± 6.8	−0.76 ± 0.08
LT1b	GTGGA**T**^{L}CTTTA	A**T**^{L}C	−63.7 ± 1.7	−171 ± 5	−10.8 ± 0.1	48.7	2.8	3.0 ± 2.4	11.3 ± 7.4	−0.46 ± 0.09
LT1c	GTGGATCT**T**^{L}TA	T**T**^{L}T	−63.6 ± 2.4	−171 ± 7	−10.6 ± 0.1	47.8	2.0	3.1 ± 2.9	11.0 ± 9.0	−0.29 ± 0.12
DT2	ACTGGCATCTG	11	−70.0 ± 2.4	−187 ± 7	−12.1 ± 0.1	54.0				
LT2a	AC**T**^{L}GGCATCTG	C**T**^{L}G	−77.0 ± 1.1	−207 ± 3	−13.0 ± 0.1	56.5	2.6	−7.0 ± 2.7	−19.6 ± 8.2	−0.93 ± 0.16
LT2b	ACTGG**C**^{L}ATCTG	G**C**^{L}A	−78.1 ± 1.6	−207 ± 5	−13.9 ± 0.1	59.8	5.9	−8.1 ± 2.9	−20.4 ± 8.9	−1.72 ± 0.18
LT2c	ACTGGCAT**C**^{L}TG	T**C**^{L}T	−73.4 ± 1.2	−195 ± 4	−12.9 ± 0.1	57.0	3.0	−3.4 ± 2.7	−8.2 ± 8.3	−0.82 ± 0.16
DT3	CGGTTGTGGCG	11	−88.3 ± 1.5	−238 ± 4	−14.6 ± 0.1	60.1				
LT3a	CG**G**^{L}TTGTGGCG	G**G**^{L}T	−87.6 ± 2.7	−233 ± 8	−15.5 ± 0.2	63.4	3.3	0.7 ± 3.0	4.8 ± 9.0	−0.74 ± 0.25
LT3b	CGGTT**G**^{L}TGGCG	T**G**^{L}T	−87.9 ± 3.3	−235 ± 10	−15.2 ± 0.3	62.0	1.9	0.4 ± 3.6	2.5 ± 10.8	−0.43 ± 0.29
LT3c	CGGTTGTG**G**^{L}CG	G**G**^{L}C	−87.1 ± 1.7	−232 ± 5	−15.2 ± 0.1	62.7	2.5	1.2 ± 2.2	5.7 ± 6.6	−0.53 ± 0.17
DT4	TATTAAGCGACCA CACATAA	20	−128.3 ± 3.8	−346 ± 11	−21.3 ± 0.4	67.7				
LT4a	TA**T**^{L}TAAGCGACC ACACATAA	A**T**^{L}T	−124.6 ± 5.3	−335 ± 15	−21.3 ± 0.5	68.3	0.6	3.7 ± 6.5	11.4 ± 19.0	0.13 ± 0.62
LT4b	TATTA**A**^{L}GCGACC ACACATAA	A**A**^{L}G	−129.5 ± 4.4	−348 ± 13	−21.9 ± 0.4	69.0	1.3	−1.2 ± 5.8	−2.1 ± 16.8	−0.57 ± 0.56
LT4c	TATTAAGC**G**^{L}ACC ACACATAA	C**G**^{L}A	−117.6 ± 4.1	−314 ± 12	−20.5 ± 0.4	68.1	0.4	10.7 ± 5.6	31.7 ± 16.4	0.83 ± 0.53
LT4d	TATTAAGCGAC**C**^{L} ACACATAA	C**C**^{L}A	−116.6 ± 4.6	−310 ± 13	−21.0 ± 0.5	69.9	2.3	11.6 ± 6.0	36.3 ± 17.4	0.35 ± 0.59
LT4e	TATTAAGCGACCA C**A**^{L}CATAA	CA**L**C	−115.3 ± 4.6	−307 ± 13	−20.7 ± 0.5	69.1	1.5	12.9 ± 5.9	39.4 ± 17.3	0.72 ± 0.58
LT4f	TATTAAGCGACCA CACA**T**^{L}AA	A**T**^{L}A	−118.8 ± 4.3	−317 ± 13	−20.8 ± 0.4	68.6	0.9	9.5 ± 5.7	28.7 ± 16.8	0.57 ± 0.55
DT5	ATTATGCTCCAAT CATGTCG	20	−136.8 ± 3.8	−372 ± 11	−21.9 ± 0.4	67.4				
LT5a	AT**T**^{L}ATGCTCCAA TCATGTCG	T**T**^{L}A	−137.9 ± 5.3	−373 ± 16	−22.7 ± 0.5	68.8	1.4	−1.1 ± 6.6	−1.5 ± 19.1	−0.63 ± 0.63
LT5b	ATTAT**G**^{L}CTCCAA TCATGTCG	T**G**^{L}C	−137.8 ± 5.4	−373 ± 16	−22.7 ± 0.5	68.8	1.5	−0.9 ± 6.6	−1.0 ± 19.3	−0.62 ± 0.63
LT5c	ATTATGCT**C**^{L}CAA TCATGTCG	T**C**^{L}C	−125.3 ± 7.8	−335 ± 23	−22.2 ± 0.8	69.7	2.4	11.5 ± 8.7	36.4 ± 25.3	0.22 ± 0.87
LT5d	ATTATGCTCCAA**A**^{L} TCATGTCG	A**A**^{L}T	−124.8 ± 6.0	−335 ± 18	−21.4 ± 0.6	68.4	1.0	12.1 ± 7.1	36.5 ± 20.8	0.75 ± 0.68
LT5e	ATTATGCTCCAAT C**A**^{L}TGTCG	CA**L**T	−126.5 ± 5.7	−340 ± 17	−21.6 ± 0.6	68.6	1.2	10.3 ± 6.9	31.6 ± 20.0	0.52 ± 0.66
LT5f	ATTATGCTCCAAT CATG**T**^{L}CG	G**T**^{L}C	−126.6 ± 4.5	−339 ± 13	−22.1 ± 0.5	70.2	2.8	10.2 ± 5.9	33.1 ± 17.2	−0.04 ± 0.59
DT6	CGAACGTCTAT	11	−67.0 ± 1.6	−181 ± 5	−11.0 ± 0.1	49.4				
LT6a	CG**A**^{L}ACGTCTAT	C**A**^{L}A	−69.5 ± 1.4	−187 ± 4	−11.6 ± 0.1	51.8	2.4	−2.5 ± 2.2	−6.3 ± 6.7	−0.58 ± 0.11
LT6b	CGAAC**G**^{L}TCTAT	GA**L**A	−69.2 ± 1.6	−186 ± 5	−11.6 ± 0.1	51.5	2.0	−2.2 ± 2.3	−5.5 ± 7.1	−0.50 ± 0.11
LT6c	CGAACGTC**T**^{L}AT	C**T**^{L}A	−74.9 ± 4.3	−203 ± 13	−12.4 ± 0.2	53.2	3.8	−8.0 ± 4.6	−22.0 ± 14.1	−1.14 ± 0.25
DT7	GTAGCAGGAGT	11	−83.4 ± 2.7	−230 ± 8	−12.3 ± 0.1	51.6				
LT7a	GT**A**^{L}GCAGGAGT	T**A**^{L}G	−75.9 ± 2.1	−206 ± 6	−12.3 ± 0.1	53.3	1.7	7.5 ± 3.4	24.2 ± 10.4	−0.05 ± 0.18
LT7b	GTAGC**A**^{L}GGAGT	C**A**^{L}G	−73.4 ± 1.7	−197 ± 5	−12.3 ± 0.1	54.2	2.6	10.0 ± 3.1	32.7 ± 9.6	−0.11 ± 0.16
LT7c	GTAGCAGG**A**^{L}GT	GA**L**G	−73.9 ± 1.8	−198 ± 6	−12.8 ± 0.1	56.1	4.5	9.5 ± 3.2	32.3 ± 9.9	−0.54 ± 0.18
DT8	TTGCTCGATGT	11	−71.1 ± 1.6	−191 ± 5	−11.9 ± 0.1	52.9				
LT8a	TT**G**^{L}CTCGATGT	T**G**^{L}C	−73.4 ± 1.5	−197 ± 5	−12.4 ± 0.1	54.6	1.6	−2.3 ± 2.2	−5.9 ± 6.6	−0.46 ± 0.13
LT8b	TTGCT**C**^{L}GATGT	T**C**^{L}G	−70.8 ± 2.0	−188 ± 6	−12.7 ± 0.1	56.6	3.7	0.2 ± 2.5	3.1 ± 7.7	−0.74 ± 0.16
LT8c	TTGCTCGA**T**^{L}GT	A**T**^{L}G	−74.9 ± 1.5	−200 ± 5	−13.0 ± 0.1	56.9	4.0	−3.8 ± 2.2	−9.0 ± 6.6	−1.04 ± 0.13
DT9	ACAAGCGACTC	11	−74.2 ± 1.4	−199 ± 4	−12.5 ± 0.1	54.9				
LT9a	AC**A**^{L}AGCGACTC	C**A**^{L}A	−72.9 ± 1.6	−195 ± 5	−12.6 ± 0.1	55.7	0.8	1.3 ± 2.1	4.4 ± 6.4	−0.09 ± 0.13
LT9b	ACAAG**C**^{L}GACTC	G**C**^{L}G	−74.9 ± 2.6	−198 ± 8	−13.5 ± 0.2	59.0	4.1	−0.7 ± 3.0	0.7 ± 8.9	−0.91 ± 0.21
LT9c	ACAAGCGA**C**^{L}TC	A**C**^{L}T	−67.4 ± 2.5	−175 ± 7	−13.1 ± 0.2	59.6	4.7	6.8 ± 2.8	23.7 ± 8.5	−0.54 ± 0.21

[a] Each LNA and DNA top strand shown was hybridized to the complementary DNA bottom strand. Total strand concentration, *CT* is 5 μM for all oligonucleotides except the LT4 and LT5 sets, for which it is 1 μM. The estimated uncertainties were calculated as in Materials and Methods. For the training data set, the uncertainties in T_M and ΔT_M are <1.0 °C. For the test and combined data sets, the uncertainty in T_M is <1.6 °C, and the uncertainty in ΔT_M is <2.4 °C. All changes due to LNA were calculated using nonrounded values of the thermodynamic parameters for individual oligonucleotides.

°C (3.3%), respectively. The observed absolute values of $\Delta H°$ and $\Delta S°$ averaged 7% less than predicted, perhaps because the sequences used here are longer than many of the sequences used in deriving the nearest-neighbor parameters (deviations were −14% and −10% for DT4 and DT5, the longest sequences). However, the observed average T_M was only 0.76 °C less than the prediction. Thus, overall the agreement between prediction and experiment is quite good. Systematic underestimation of $\Delta H°$ and $\Delta S°$ by 7% would lead to 7% errors in $\Delta\Delta H°$ and $\Delta\Delta S°$. The standard deviations in $\Delta\Delta H°$ and $\Delta\Delta S°$ are much larger, so any such systematic bias should make a small difference to the quality of the results. The uncertainties in $\Delta H°$ and $\Delta S°$ derived as above correlated with the deviations from predicted values, suggesting that the relative uncertainties are sensible. There were no obvious experimental causes for the largest deviations from predicted values, except that DT4 and DT5 showed slight non-two-state behavior.

The $\Delta\Delta H°$, $\Delta\Delta S°$, $\Delta\Delta G°_{37}$, and ΔT_M attributable to an LNA nucleotide are estimated from $\Delta H°$, $\Delta S°$, and $\Delta G°_{37}$. The change in $\Delta H°$ is given by the equation below, with $\Delta\Delta S°$, $\Delta\Delta G°_{37}$, and ΔT_M given by analogous expressions.

$$\Delta\Delta H° \equiv \Delta H°_{\text{LNA+DNA:DNA}} - \Delta H°_{\text{DNA:DNA}} \quad (8)$$

where LNA+DNA:DNA refers to the duplex with a single LNA incorporated in one strand and DNA:DNA refers to to the reference DNA duplex. The variance in the estimated $\Delta\Delta H°$ is given below, again with analogous expressions for $\Delta\Delta S°$, $\Delta\Delta G°_{37}$, and ΔT_M.

$$\sigma^2(\Delta\Delta H°) = \sigma^2(\Delta H°_{\text{LNA+DNA:DNA}}) + \sigma^2(\Delta H°_{\text{DNA:DNA}}) \quad (9)$$

As for $\Delta G°_{37}$ and T_M, $\Delta\Delta G°_{37}$ and ΔT_M are more precisely determined than $\Delta\Delta H°$ or $\Delta\Delta S°$ because of the correlated errors in $\Delta\Delta H°$ and $\Delta\Delta S°$. To the extent that the errors in $\Delta H°_{\text{LNA+DNA:DNA}}$ and $\Delta H°_{\text{DNA:DNA}}$ are correlated with each other, the true variance in $\Delta\Delta H°$ and the other difference parameters will be smaller.

Singular Value Decomposition. Singular value decomposition (SVD) was applied separately to the sets of $\Delta\Delta H°$, $\Delta\Delta S°$, and $\Delta\Delta G°_{37}$ values, essentially as described by Xia et al. (*11*), using Matlab (The Mathworks, Inc.). The variances estimated as above were used as weighting factors. The construction of design matrices and solution vectors is described in Results. Initial parameter sets were derived exclusively from the "training data set" and then validated against the "test data set." The final parameter sets given in Table 4 are derived from repeating the SVD analysis using the combined data set of 100 oligonucleotides.

RESULTS

This paper presents a general approach to deriving predictive thermodynamic rules for modified nucleic acids. It is possible to predict the T_M of unmodified DNA oligonucleotides accurately using nearest-neighbor parameters for $\Delta H°$ and $\Delta S°$, suggesting that the additional results needed here are the incremental effects of a modification on $\Delta H°$ and $\Delta S°$. Sequence-dependent enthalpy and entropy increments for LNA were obtained as follows: the changes in hybridization enthalpy and entropy upon LNA substitution were measured for a set of 67 oligonucleotides that included all 64 possible trinucleotides with a single LNA in the center, hybridized to an all-DNA complement. The results provided nearest-neighbor $\Delta\Delta H°$ and $\Delta\Delta S°$ parameters for LNA−DNA dinucleotides, and different parameter sets were then evaluated for their accuracy in predicting the thermodynamics of 33 test oligonucleotides. We find that the sequence context on both the 5′ and 3′ sides of the LNA must be considered for accurate prediction of hybridization thermodynamics. Quantitative analysis of the enthalpy and entropy changes due to LNA substitution suggests that decreased entropy loss upon hybridization is the usually the dominant contributor to its enhanced stability, but that the most stabilizing LNA substitutions are characterized by favorable enthalpy changes and unfavorable entropy changes.

Accurate predictions of the change in melting temperature can be obtained using the $\Delta\Delta H°$ and $\Delta\Delta S°$ parameters derived here. The individual predictions for $\Delta\Delta H°$ and $\Delta\Delta S°$ are less accurate, but the errors are highly correlated. These parameters offer a substantial improvement over the available thermodynamics on LNA.

Sequence Design. We assume that most of the sequence dependence for the effect of an LNA substitution on hybridization thermodynamics can be captured by considering the nearest neighbors of the LNA. There are 64 possible trinucleotides 5′-M$\mathbf{X^L}$N-3′, where M and N = A, C, G, or T and $\mathbf{X^L}$ = LNA-A, LNA-C, LNA-G, or LNA-T. There are 32 possible nearest-neighbor pairs 5′-M$\mathbf{X^L}$-3′ and 5′-$\mathbf{X^L}$N-3′. In this work, we restrict our attention to these 64 trinucleotides, hybridized to their perfect-match complements, and to singly substituted DNA hybridizing to all-natural DNA (e.g., 5′-A$\mathbf{C^L}$G-3′•5′-CGT-3′). Furthermore, LNA is always at least two nucleotides away from the 5′ or 3′ end of the molecule. The case of a modified strand hybridizing to native DNA is the most relevant to the use of LNA as a probe or therapeutic.

LNA in other sequence contexts could be treated analogously. An LNA at or near the 5′ or 3′ end can be treated similarly to internal substitutions. A complete description of the 12 mismatched LNA•DNA base pairs in context requires 192 trinucleotides (e.g., 5′-A$\mathbf{C^L}$G-3′•5′-CAT-3′, •5′-CCT, and •5′-CTT). LNA•LNA hybridization, multiple LNA substitutions, mismatches adjacent to an LNA, and dangling ends can all be addressed systematically as well.

The 64 LNA-containing trinucleotides were embedded in 15 sequence contexts, so that several different singly substituted oligonucleotides could be hybridized to a common DNA complement and compared to a common reference DNA. Initial reference DNA sequences were chosen from those used in determining DNA nearest-neighbor rules (*13*), as they were known to exhibit two-state melting. Subsequently, new sequences were designed to include the remaining LNA-containing trinucleotides and for the test data set. In each case, the sequence was evaluated for possible secondary structure using MFOLD (*33*), the all-DNA duplex was synthesized, and its melting was characterized before the more expensive LNA oligonucleotides were synthesized. DNA sequences with substantial non-two-state character in their melting curves were redesigned. This set of 67 oligonucleotides will be referred to as the "training data set."

To test the predictions (derived as described below) from analysis of the training data set, we designed a "test data set" of 33 oligonucleotides, in nine sequence contexts. The

Table 2: Average Thermodynamic Parameters for Each LNA Nucleotide[a]

		$\Delta\Delta H^\circ$ (kcal/mol)	$\Delta\Delta S^\circ$ (cal/mol K)	$\Delta\Delta G^\circ_{37}$ (kcal/mol)	ΔT_M (°C)
LNA-A	avg ± std dev	3.64 ± 5.04	12.60 ± 15.09	−0.27 ± 0.49	2.11 ± 1.30
	range observed	−4.71 to 12.93	−13.84 to 39.36	−1.47 to 0.75	−0.28 to 5.74
LNA-C	avg ± std dev	1.93 ± 5.52	8.74 ± 16.44	−0.78 ± 0.48	4.44 ± 1.46
	range observed	−8.05 to 11.62	−21.45 to 37.18	−1.72 to 0.35	2.25 to 7.26
LNA-G	avg ± std dev	0.74 ± 3.99	4.17 ± 11.51	−0.56 ± 0.54	2.83 ± 1.75
	range observed	−6.07 to 10.68	−13.70 to 32.28	−1.90 to 0.83	−0.62 to 6.88
LNA-T	avg ± std dev	1.04 ± 4.53	5.33 ± 13.53	−0.62 ± 0.44	3.21 ± 1.41
	range observed	−7.97 to 10.22	−22.01 to 33.11	−1.22 to 0.57	0.62 to 6.17

[a] Data are averaged over the entire combined data set of 100 oligonucleotides.

test data set included two examples of each of the 32 possible nearest-neighbor pairs 5′-M**X**$^{\mathbf{L}}$-3′ and 5′-**X**$^{\mathbf{L}}$N-3′. Half of the test data set oligonucleotides included the corresponding 5′-M**X**$^{\mathbf{L}}$N-3′ trinucleotide from the training data set in the same pentanucleotide context as in the training data set (for example, 5′-T**G**$^{\mathbf{L}}$C appears in the training data set in 5′-CCATT**G**$^{\mathbf{L}}$CTACC, L5c, and in the test set in 5′-TT**G**$^{\mathbf{L}}$-CTCGATGT, LT8a), whereas half of them had the trinucleotide in a new context (for example, 5′-C**G**$^{\mathbf{L}}$A appears in 5′-GTC**G**$^{\mathbf{L}}$AACAGC, L1a, in the training data set and in 5′-TATTAAGC**G**$^{\mathbf{L}}$ACCACACATAA, LT4c, in the test data set). To check for length effects, two of the oligonucleotide contexts in the test set were 20-mers rather than the 8−12-mers of the training data set. After evaluation of the accuracy of predictions made for the test set, the final parameter sets were derived from the "combined data set" of all 100 oligonucleotides, as described below. All of the oligonucleotide sequences and their thermodynamic properties are listed in Table 1. This training/test/combined data set approach, with the same oligonucleotide sequences, should be generally applicable to any modified chemistry where the nearest neighbors are likely to dominate the thermodynamics. Extensions to cases such as terminal LNA substitutions, mismatches, and dangling ends can employ many of the same sequences.

Thermodynamic Effects of LNA Substitution. The effect of LNA is represented as the difference in the ΔH° and ΔS° for hybridization between LNA-containing and natural DNA, $\Delta\Delta H^\circ \equiv \Delta H^\circ_{\text{LNA+DNA:DNA}} - \Delta H^\circ_{\text{DNA:DNA}}$ and $\Delta\Delta S^\circ \equiv \Delta S^\circ_{\text{LNA+DNA:DNA}} - \Delta S^\circ_{\text{DNA:DNA}}$. A positive $\Delta\Delta H^\circ$ represents a less favorable enthalpy decrease and a positive $\Delta\Delta S^\circ$ represents a less unfavorable entropy loss upon hybridization. The central challenge of this work is that $\Delta\Delta H^\circ$ and $\Delta\Delta S^\circ$ values are small differences between large numbers. Therefore, they are subject to large uncertainties, generally ±2−3 kcal/mol for $\Delta\Delta H^\circ$ and ±6−9 cal/mol K for $\Delta\Delta S^\circ$, of the same order as the $\Delta\Delta H^\circ$ and $\Delta\Delta S^\circ$ values themselves (Table 1). The $\Delta\Delta G^\circ_{37}$ and ΔT_M values are much more precise, but they do not allow straightforward predictions for the effect of an LNA substitution on T_M in a new context. However, the overall trends in $\Delta\Delta H^\circ$, $\Delta\Delta S^\circ$, $\Delta\Delta G^\circ_{37}$, and ΔT_M are quite informative as to the qualitative origin of LNA-induced duplex stabilization and the degree of sequence dependence.

In general, the absorbance melting curve analysis on all 100 oligonucleotides of the combined set demonstrates that LNA purines contribute significantly less enhanced stability than do pyrimidines. LNA-A in particular has the smallest effect of the four; in terms of $\Delta\Delta G^\circ_{37}$ the average stability increments are ordered $\mathbf{A}^{\mathbf{L}} \ll \mathbf{G}^{\mathbf{L}} < \mathbf{T}^{\mathbf{L}} < \mathbf{C}^{\mathbf{L}}$. The averages and ranges for $\Delta\Delta H^\circ$, $\Delta\Delta S^\circ$, $\Delta\Delta G^\circ_{37}$, and ΔT_M are broken down according to LNA base in Table 2. Superimposed on the LNA base dependence there is substantial dependence on the neighboring sequence for all of the LNA bases. For example, the observed $\Delta\Delta G^\circ_{37}$ due to LNA-A is −0.27 ± 0.49 kcal/mol (average ± standard deviation), vs. −0.78 ± 0.48 for LNA-C, but the observed range of values is −1.47 to +0.75 for LNA-A and −1.72 to +0.35 for LNA-C. The observed ΔT_M ranges from −0.6 °C (GTC**G**$^{\mathbf{L}}$AACAGC) to 7.8 °C (GGTG**C**$^{\mathbf{L}}$CAA). Thus, the identity of the LNA base alone is clearly not sufficient for making accurate stability predictions. The context effects are described more fully below, based upon the derived nearest-neighbor parameters.

Despite the large errors associated with $\Delta\Delta H^\circ$ and $\Delta\Delta S^\circ$, some trends in these parameters are clear, as illustrated in Figure 3. The majority of the LNA-containing oligonucleotides have positive $\Delta\Delta H^\circ$ values, so that enthalpy effects actually tend to destabilize most of the LNA−DNA duplexes. If there is any enhanced base stacking in these duplexes, it is apparently counterbalanced by destabilizing enthalpic effects, which could include an enthalpic cost for junctions between A-like DNA around the LNA and B-like DNA elsewhere, or unfavorable steric interactions induced by the methylene bridge. These LNA+DNA:DNA duplexes are stabilized by positive $\Delta\Delta S^\circ$ values, confirming that an important stabilizing feature of the LNA substitution is the reduced entropic cost of forming the duplex, presumably due to preorganization of the single-stranded LNA+DNA reactant (*26*).

Figure 3B shows, however, that the most stabilizing LNA substitutions (those with the most negative $\Delta\Delta G^\circ_{37}$, for example, L8a1 and L11b) are characterized by negative $\Delta\Delta H^\circ$ and $\Delta\Delta S^\circ$ values. In these cases, there may be a dominant effect of improved stacking, which also decreases the entropy of the duplex. These results suggest that different LNA-containing oligonucleotides may have significantly different structural and dynamic properties. Note that the intercept of the $\Delta\Delta S^\circ$ vs $\Delta\Delta H^\circ$ plot in Figure 3A is near zero, indicating that stabilization is primarily entropic or enthalpic, but not both at the same time for the same oligonucleotide.

Figure 3 also shows that some sequences, mainly in the LT4 and LT5 families of longer oligonucleotides, have unusually large $\Delta\Delta H^\circ$ and $\Delta\Delta G^\circ_{37}$ values. A $\Delta\Delta H^\circ$ of 10 kcal/mol seems greater than what is physically likely, since the ΔH° for formation of a base pair is only −7 to −10 kcal/mol. It is more likely that ΔH° and ΔS° were underestimated for these sequences than for the shorter ones, since experimental noise and the increased likelihood of multiple premelted states will tend to blur the sharpness of the melting transition. We have also observed more multistate behavior

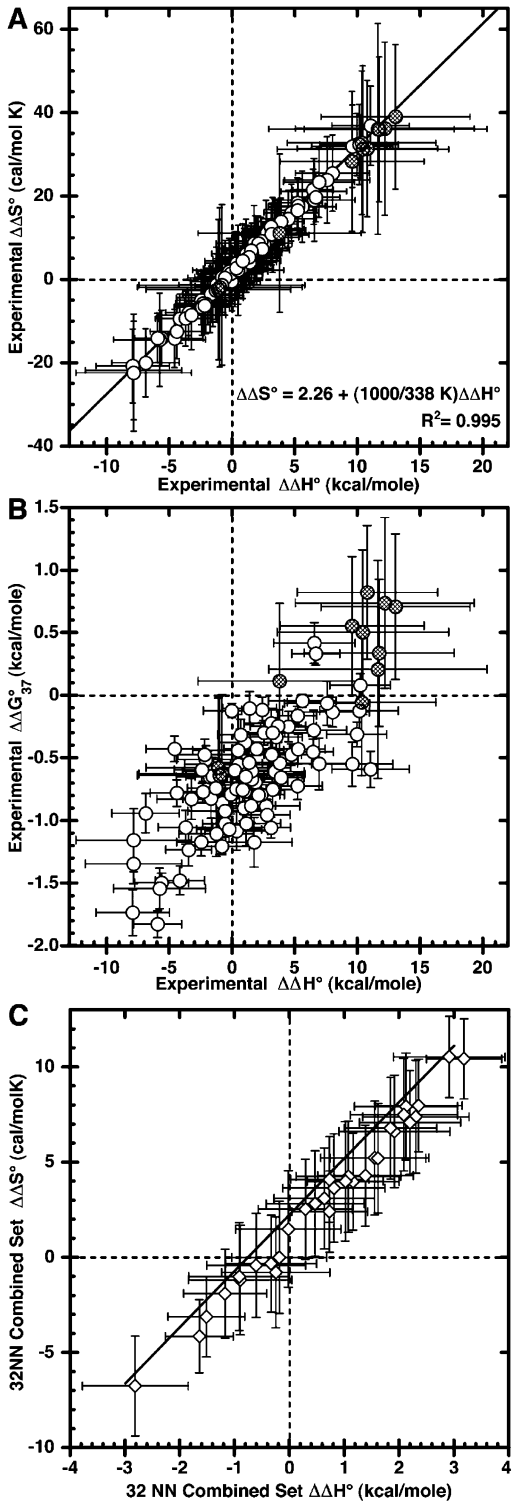

FIGURE 3: LNA thermodynamics and enthalpy−entropy compensation. Experimental $\Delta\Delta S°$, $\Delta\Delta H°$, and $\Delta G°_{37}$ values (circles) and estimated uncertainties for the oligonucleotides of the combined data set are from Table 1. The hatched symbols correspond to the LT4 and LT5 families of longer oligonucleotides, for which the measured $\Delta\Delta H°$ and $\Delta\Delta G°_{37}$ are probably overestimated. Panel A illustrates strong apparent enthalpy−entropy compensation in the combined data set, and shows that the majority of the LNA-modified oligonucleotides exhibit positive $\Delta\Delta H°$ and $\Delta\Delta S°$ values. Panel B illustrates correlation between $\Delta\Delta G°_{37}$ and $\Delta\Delta H°$, showing that the most stabilizing substitutions are characterized by negative $\Delta\Delta H°$ (and therefore negative $\Delta\Delta S°$). Panel C illustrates enthalpy−entropy compensation in the 32NN parameter set from Table 4, with the linear regression line obtained from the data in panel A superimposed.

like that of Figure 2 in LNA-containing DNA, especially if the LNA is near one end. In this situation, the measured $\Delta\Delta H°$ is likely to be greater than the true value, and the compensating errors in $\Delta\Delta H°$ and $\Delta\Delta S°$ then lead to a larger measured $\Delta\Delta S°$ than the true value. Since the T_M is ~70 °C for the longer oligonucleotides, $\Delta\Delta G°_{37}$ is obtained by extrapolation over a large temperature difference from the $\Delta\Delta G°(T_M)$, so an artifactually large $\Delta\Delta S°$ will lead to an artifactually large $\Delta\Delta G°_{37}$. We debated omitting the LT4 and LT5 families from further analysis, but retained them because they represent the best models for oligonucleotide lengths that would be used in applications. The results for these oligonucleotides do not affect the quantitative evaluation of the quality of fit to T_M very much, because the predicted ΔT_M values for these long sequences are relatively small and therefore insensitive to errors.

Entropy-Enthalpy Compensation. The results for $\Delta\Delta H°$ and $\Delta\Delta S°$ (Figure 3) illustrate remarkable apparent enthalpy−entropy compensation, meaning that $\Delta\Delta H°$ and $\Delta\Delta S°$ are proportional, and therefore that $\Delta\Delta G°$ for the set of sequences is relatively small and constant whereas $\Delta\Delta H°$ and $\Delta\Delta S°$ vary substantially. Part of the compensation may be artifactual, since covariation of errors in $\Delta\Delta H°$ and $\Delta\Delta S°$ gives apparent compensation. This area has been authoritatively reviewed (*34*); based on the suggested tests that compensation should be apparent when $\Delta H°$ and $\Delta S°$ are plotted with appropriate error bars and that they should correlate with $\Delta G°_{37}$, we believe that the compensation found here does reflect an underlying physical reality. If compensation were solely due to experimental error, then the plot of $\Delta\Delta G°_{37}$ vs $\Delta\Delta H°$ in Figure 3B would show randomly scattered $\Delta\Delta H°$ values instead of the observed positive correlation. The observed compensation has an apparent isoequilibrium temperature (the temperature at which all the sequences have the same $\Delta\Delta G°_T$ and hence the same stabilizing effect) of 65 °C, but an isoequilibrium relationship has not been demonstrated experimentally. True enthalpy−entropy compensation is often qualitatively explained as resulting from the tendency of enthalpically more stable interactions to decrease entropic freedom of motion more than weaker interactions. Even though the root cause of quantitative compensation is not well understood, it has been used to evaluate theoretical models for base pairing (*35*).

Deriving Predictive Parameter Sets for $\Delta\Delta H°$ and $\Delta\Delta S°$. One goal of this work is to extract predictive parameters from measured thermodynamics, and to provide the results in a form that can be readily incorporated into standard prediction algorithms. Therefore, the approach to the reduction of the data depends on the commonly used nearest-neighbor models. We extracted nearest-neighbor LNA+DNA parameters from the training data set using the principle of maximum likelihood, evaluated their predictions against the test data set, and then used the combined data sets to generate the final parameters.

The "unified view" DNA:DNA nearest-neighbor model of SantaLucia (*12*) gives $\Delta H°$, $\Delta S°$, and $\Delta G°_{37}$ by expressions of the following form:

$$\Delta H° = \sum_i n_i \Delta H°_i + n_{t,GC}\Delta H°_{t,GC} + n_{t,AT}\Delta H°_{t,AT} \quad (10)$$

where n_i is the count of the ith nearest-neighbor dinucleotide of the 10 unique DNA-DNA neighbors, ΔH°_i is ΔH° of the ith nearest-neighbor, $n_{t,GC}$ is the number of G:C termini with the corresponding $\Delta H^\circ_{t,GC}$, and $n_{t,AT}$ is the number of A:T initiation pairs with corresponding $\Delta H^\circ_{t,AT}$. Replacement of ΔH° by ΔS° or ΔG°_{37} gives corresponding equations. Since none of the LNA+DNA:DNA duplexes in question here can be self-complementary, no symmetry correction terms for ΔS° and ΔG°_{37} are needed. All of our analysis assumes that ΔH° and ΔS° do not change with temperature. The concentration range explored here may not be large enough to show the effects of heat capacity changes that have been observed by others (*36*).

The simplest way to extend the above equations to account for the incremental effect of a single LNA incorporation is to add a trinucleotide term as follows:

$$\Delta H^\circ = \sum_i n_i \Delta H^\circ_i + n_{t,GC}\Delta H^\circ_{t,GC} + n_{t,AT}\Delta H^\circ_{t,AT} + \Delta\Delta H^\circ_{LNA-MX^LN} \quad (11)$$

in which $\Delta\Delta H^\circ_{LNA-MX^LN}$ is the $\Delta\Delta H^\circ$ value determined for the training data set oligonucleotide that contains the M$\mathbf{X^L}$N sequence. Since there are 64 LNA-M$\mathbf{X^L}$N parameters and 67 sequences in the training data, each of the sequences would contribute to the estimate for one, and only one, trinucleotide parameter, and each parameter would be determined by one, and only one, sequence. There would be no internal indication of the likely accuracy or precision of any prediction for a new sequence, and the large errors in the experimental $\Delta\Delta H^\circ$ and $\Delta\Delta S^\circ$ values make it likely that some of the predictions would be incorrect by a wide margin, especially if the unknown sequence had a T_M significantly different from that of the training data set oligonucleotide. Deriving a reduced set of nearest-neighbor parameters uses more of the experimental data to compute each parameter, and the data enter in various combinations: random errors have a decreased effect and the parameters are determined more precisely. However, too great an extent of reduction, such as using just one $\Delta\Delta H^\circ$ and one $\Delta\Delta S^\circ$ for each LNA base, fails to model the true variation among sequences.

The most straightforward nearest-neighbor approach, referred to as 32NN here, partitions $\Delta\Delta H^\circ$ into contributions due to the corresponding M$\mathbf{X^L}$ and $\mathbf{X^L}$N dinucleotides of the M$\mathbf{X^L}$N triplet. This assumes that the contribution of a nearest-neighbor pair is independent of the adjacent nearest neighbors. The resulting expression is given below:

$$\Delta H^\circ = \sum_i n_i \Delta H^\circ_i + n_{t,GC}\Delta H^\circ_{t,GC} + n_{t,AT}\Delta H^\circ_{t,AT} + \Delta\Delta H^\circ_{LNA-MX^L} + \Delta\Delta H^\circ_{LNA-X^LN} \quad (12)$$

where $\Delta\Delta H^\circ_{LNA-MX^L}$ and $\Delta\Delta H^\circ_{LNA-X^LN}$ are the incremental effects of a single LNA substitution, for dinucleotides with 3′ or 5′ LNA substitutions, respectively. This model has 32 $\Delta\Delta H^\circ$ ($\Delta\Delta S^\circ$, $\Delta\Delta G^\circ_{37}$) parameters, each of the 67 training data set sequences contributes to two dinucleotide parameters, and each dinucleotide parameter is determined from at least four different sequences. The final 32NN parameters in Table 4 were computed based on the combined data set, and each parameter appears six times in the set (seven times for eight of the parameters).

There are many possible parameter sets (Table 3), ranging from the 64 M$\mathbf{X^L}$N trinucleotides to simply the four LNA bases. Even the 32NN set has an uncomfortably large ratio of parameters to data points (1:2). The simplest way to reduce the number of parameters while retaining a connection to structure is to consider purine vs pyrimidine neighbors as a group, yielding the 16(RY)XL(RY) set of (R or Y)$\mathbf{X^L}$ + $\mathbf{X^L}$(R or Y) nearest neighbors. The two nearest-neighbor sets were also split into their component halves (for example, 32NN was split into the 16MXL and 16XLN sets) to evaluate whether the 5′ or 3′ neighbors of an LNA have a greater influence on the thermodynamics. In fitting to the training data set, half-sets must necessarily fit less well than the parent set, but they can give better agreement to the test data set if the parent set parameters are strongly affected by fitting to the noise in the data. Reduced sets that considered the LNA purines vs pyrimidines (e.g., 16 N($\mathbf{R+Y}$)$^\mathbf{L}$ + ($\mathbf{R+Y}$)$^\mathbf{L}$M) were also evaluated, with comparable but not superior results, not shown.

For each parameter set, we compute the values of the N parameters that minimize the error-weighted difference between prediction and experiment for the M members of the training data set. The training data set "stacking matrix" is the $M\times N$ matrix in which row j (of 67) contains a 1 in each column corresponding to an LNA dinucleotide that appears in the jth oligonucleotide, and 0's elsewhere. The error-weighted "design matrixes" for $\Delta\Delta H^\circ$, $\Delta\Delta S^\circ$, and $\Delta\Delta G^\circ_{37}$ (denoted $\mathbf{H}_\sigma$, $\mathbf{S}_\sigma$, $\mathbf{G}_\sigma$) are then obtained by dividing each row j of the stacking matrix by $\sigma^2(\Delta\Delta H^\circ_{j,expt})$, the experimental variance in $\Delta\Delta H^\circ$ (or $\Delta\Delta S^\circ$, or $\Delta\Delta G^\circ_{37}$) for oligonucleotide j. The M-element vector $\mathbf{b_H}$ (or $\mathbf{b_S}$, or $\mathbf{b_G}$) of experimental $\Delta\Delta H^\circ$ (or $\Delta\Delta S^\circ$, or $\Delta\Delta G^\circ_{37}$) is similarly weighted. The desired parameters are then the N-element vectors $\mathbf{h_{NN}}$, $\mathbf{s_{NN}}$, and $\mathbf{g_{NN}}$ of nearest-neighbor parameters that minimize χ^2 error expressions of the following form:

$$\chi^2_{train}(\Delta\Delta H^\circ) = \sum_j \frac{(\Delta\Delta H^\circ_{j,pred} - \Delta\Delta H^\circ_{j,expt})^2}{\sigma^2(\Delta\Delta H^\circ_{j,expt})} = |\mathbf{H}_\sigma\cdot\mathbf{h_{NN}} - \mathbf{b_H}|^2 \quad (13)$$

where j indexes the 67 members of the training data set, $\Delta\Delta H^\circ_{j,pred}$ is the sum of the computed $\Delta\Delta H^\circ$ values for the LNA-containing dinucleotides in the jth oligonucleotide, and χ^2_{train} is calculated with respect to the training data set experimental values.

SVD was used to compute values for the nearest-neighbor parameters that minimize the χ^2_{train} errors above (*37*). SVD was done independently for $\Delta\Delta H^\circ$, $\Delta\Delta S^\circ$, and $\Delta\Delta G^\circ_{37}$, essentially as described by Xia et al. (*11*), with the results shown in Table 3. Although the input free energies are derived from the enthalpy and entropy, the weighting of errors in the SVD means that the final parameter estimates for $\Delta\Delta G^\circ_{37}$ obey $\Delta\Delta G^\circ_{37} = \Delta\Delta H^\circ - 310.15\ \text{K} \times \Delta\Delta S^\circ$ approximately but not exactly. Note that the nearest-neighbor sets 32NN and 16(RY)XL(RY) can only be determined to within four arbitrary constants (for example, adding a constant to the 5′-R$\mathbf{G^L}$ and 5′-Y$\mathbf{G^L}$ parameters and subtracting the same constant from 5′-$\mathbf{G^L}$R and 5′-$\mathbf{G^L}$Y gives the

Table 3: Evaluation of the Accuracy of Prediction of LNA−DNA Hybridization Thermodynamics

Parameter Sets Computed from the Training Data Set, Evaluated against the Training Data Set

set name	form	sample	no./64[a]	χ^2_{train} ($\Delta\Delta H°$)	χ^2_{train} ($\Delta\Delta S°$)	χ^2_{train} ($\Delta\Delta G°_{37}$)	χ^2_{train} (ΔT_M)	$Q(\chi^2/4;f)^b$ (ΔT_M)	$Q(\chi^2/4;f)^{b,c}$ (ΔT_M; 120% e.t.)
64Trinuc	M$\mathbf{X^L}$N	C$\mathbf{G^L}$A	1	(0)	(0)	(0)	(0)	(1)	(1)
32NN	M$\mathbf{X^L}$ + $\mathbf{X^L}$N	C$\mathbf{G^L}$ + $\mathbf{G^L}$A	4	118	113	391	257	7.14×10^{-11}	9.05×10^{-5}
16MXL	M$\mathbf{X^L}$	C$\mathbf{G^L}$	4	148	141	660	482	4.65×10^{-20}	1.98×10^{-8}
16XLN	$\mathbf{X^L}$N	$\mathbf{G^L}$A	4	145	135	715	449	2.08×10^{-18}	1.12×10^{-11}
16(RY)XL(RY)	(R/Y)$\mathbf{X^L}$ + $\mathbf{X^L}$(R/Y)	Y$\mathbf{G^L}$ + $\mathbf{G^L}$R	8	160	152	635	407	2.39×10^{-16}	9.56×10^{-7}
8(RY)XL	(R/Y)$\mathbf{X^L}$	Y$\mathbf{G^L}$	8	165	155	751	467	2.78×10^{-18}	2.28×10^{-7}
8XL(RY)	$\mathbf{X^L}$(R/Y)	$\mathbf{G^L}$R	8	170	159	862	566	2.98×10^{-23}	1.30×10^{-9}
4XL	$\mathbf{X^L}$	$\mathbf{G^L}$	16	174	162	1014	669	8.75×10^{-28}	1.51×10^{-11}

Parameters Sets Computed from the Training Data Set, Evaluated against the Test Data Set

set name	χ^2_{test} ($\Delta\Delta H°$)	χ^2_{test} ($\Delta\Delta S°$)	χ^2_{test} ($\Delta\Delta G°_{37}$)	χ^2_{test} (ΔT_M)	$Q(\chi^2/4;f)^d$ (ΔT_M)	$Q(\chi^2/4;f)^{c,d}$ (ΔT_M; 120% e.t.)
64Trinuc	110.0	110.2	202.0	131	6.71×10^{-5}	0.0370
32NN	99.8	101.4	142.4	71.7	0.0218	0.345
16MXL	91.8	91.6	160.0	88.5	0.00469	0.198
16XLN	92.4	94.8	98.2	62.9	0.0466	0.447
16(RY)XL(RY)	87.3	88.7	101.3	60.0	0.0592	0.484
8(RY)XL	81.3	83.1	97.3	70.6	0.0240	0.357
8XL(RY)	90.1	91.3	115.1	61.7	0.0516	0.463
4XL	83.8	85.6	112.4	68.2	0.0295	0.384

Parameter Sets Recomputed from the Combined Data Set, Evaluated against the Combined Data Set

set name	χ^2_{comb} ($\Delta\Delta H°$)	χ^2_{comb} ($\Delta\Delta S°$)	χ^2_{comb} ($\Delta\Delta GG°_{37}$)	χ^2_{comb} (ΔT_M)	$Q_{comb}(\chi^2/4;f)^e$ (ΔT_M)	$Q_{comb}(\chi^2/4;f)^{c,e}$ (ΔT_M; 120% e.t.)
32NN	195	191	509	319	4.31×10^{-10}	0.00136
16MXL	227	220	799	542	1.00×10^{-18}	8.19×10^{-7}
16XLN	226	218	806	487	3.36×10^{-16}	9.36×10^{-6}
16(RY)XL(RY)	240	233	727	431	1.14×10^{-13}	0.000101
8(RY)XL	247	238	956	609	4.66×10^{-21}	1.31×10^{-7}
8XL(RY)	250	241	862	512	1.51×10^{-16}	9.96×10^{-6}
4XL	256	245	1125	724	4.80×10^{-26}	1.20×10^{-9}

[a] "no./64" is the number of times each parameter in the set appears in the set of 64 M$\mathbf{X^L}$N trinucleotides. [b] The number of degrees of freedom f used in computing $Q(\chi^2/4;f)$ is $(64-\nu)/4$, where ν, the number of parameters, is given at the beginning of the parameter set name. [c] The $Q(\chi^2/4;f)$ value is calculated assuming an error threshold of 120% of the minimum RSS for $\Delta H°$ and $\Delta S°$ vs the 110% used elsewhere, as described in Materials and Methods. [d] The degrees of freedom f is 33/4, since no parameters were adjustable. [e] The degrees of freedom f is $(100-\nu)/4$.

same predictions for any actual sequence). The SVD identifies such combinations of linearly dependent parameters as weight matrix columns with vanishing small corresponding weights, and zeroing the contribution of these columns then sets the arbitrary increments to zero (*12*, *37*). All of the other weights were comparable to each other, as expected given the uniform distribution of training data set sequences.

Evaluation of Parameter Set Quality and Predictive Accuracy. A small χ^2_{train} value for a parameter set does not necessarily mean that it is the best set, since a larger number of parameters may simply allow more precise fitting to noise in the data. To evaluate the predictive value of the different derived parameter sets, we used each set to predict $\Delta\Delta H°$, $\Delta\Delta S°$, $\Delta\Delta G°_{37}$, and ΔT_M for all of the test data set oligonucleotides, and then calculated χ^2_{test} for each of the four experimental observables. The results in Table 3 show that, as expected, the set of 64 trinucleotide parameters used directly from the training data set gives poor agreement with the test data set results. The agreement for the 32NN set is significantly better, but not as good as the more reduced representations, including its own 16XLN subset. This suggests that the training set data are not sufficient to define all 32 nearest neighbors precisely. Parameter sets that consider both the 5′ and 3′ neighbors give the best predictions, suggesting that both neighbors affect the environment of the LNA.

The quality of the parameter sets can also be judged by the "*Q* value" for each parameter set, which is the probability that a value of χ^2 at least as large (poor) as the observed value would be obtained by chance, given that there are f degrees of freedom (f = no. of oligonucleotides − no. of nearest-neighbor parameters) in the fit. The value of Q is a sensitive indicator of whether the model is appropriate to the data, and of whether additional parameters in a fit are justified (*11*, *37*). $Q(\chi^2; f)$ is given by the upper incomplete gamma function as follows:

$$Q(\chi^2; f) = \frac{1}{\Gamma(f/2)} \int_{\chi^2/2}^{\infty} e^{-t} t^{(f-2)/2} dt \qquad (14)$$

where the gamma function $\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt$. The LNA-containing oligonucleotides used here each have only a single LNA, so the nearest-neighbor parameters referring to an LNA base (for example, R$\mathbf{A^L}$, Y$\mathbf{A^L}$, $\mathbf{A^L}$Y, and $\mathbf{A^L}$R in the 16-(RY)XL(RY) set) never appear in the same stacking matrix row as the parameters for any other LNA base. Thus, the

stacking and design matrices are block-diagonal, so each SVD analysis is essentially four separate problems. Therefore, χ^2 and f are each divided by 4 in calculating Q.

We evaluate the quality of the parameter sets primarily based on χ^2 and Q values for ΔT_M, because ΔT_M is the most important and most precisely measured experimental number and because it depends on both $\Delta\Delta H°$ and $\Delta\Delta S°$. The $\chi^2_{train}(\Delta T_M)$ values are larger than $\chi^2_{train}(\Delta\Delta H°)$ and $\chi^2_{train}(\Delta\Delta S°)$, because the errors in the experimental ΔT_M are smaller. High values of χ^2 (low values of Q) suggest that the estimated variances for ΔT_M are too small. As discussed in Materials and Methods, increasing the error threshold in the $\Delta H°$ and $\Delta S°$ fits to 120% rather than 110% of the minimum RSS would decrease $\chi^2(\Delta T_M)$ by a factor of 2, and the recalculated Q is in the last column of Table 3. These Q values are much larger; 0.001 is considered acceptable if the experimental errors may not be normally distributed (*11*, *37*). The maximum Q value is observed for the 32NN set.

The key test for predictive value, however, is the quality of predictions for the test data set, which did not enter into the computation of the parameters. As shown in Table 3, the χ^2_{test} values confirm that the 32NN parameter set is not well-defined by the training data set alone, since the 16(RY)XL(RY) set gives a smaller χ^2_{test} and larger Q. The Q values are much larger for the test data set than for the training data set because the number of degrees of freedom is larger for the test data set, but the χ^2_{test} values for $\Delta\Delta H°$ and $\Delta\Delta S°$ are similar to the χ^2_{train} values on a per oligonucleotide basis. The χ^2_{test} values for $\Delta\Delta G°_{37}$ and ΔT_M are surprisingly small relative to χ^2_{train}, mainly because experimental uncertainties are larger for the test data set (the LT4 and LT5 families in particular). Figure 4(A) compares the experimental test data set $\Delta\Delta H°$, $\Delta\Delta G°_{37}$, and ΔT_M with the predictions of the 32NN and 16(RY)XL(RY) parameter sets. Agreement is seen to be quite good, except for the $\Delta\Delta H°$ and $\Delta\Delta G°_{37}$ of the LT4 and LT5 families, for which the experimental $\Delta\Delta H°$ and $\Delta\Delta G°_{37}$ are probably overestimated (see above).

Comparison Among Different Methods for T_M Prediction. The goal for these methods is the accurate prediction of T_M values for the test set data oligonucleotides. At this point in the analysis, since the 16(RY)XL(RY) parameter set derived from the training data set gave the minimum χ^2_{test} values (and smaller parameter uncertainties than the 32NN set), we used the 16(RY)XL(RY) set to compare the nearest-neighbor predictions with other available methods. For an unbiased comparison, the four prediction methods are derived from input data that did not include the test data set. (Similar results in terms of the overall fit errors were obtained using the 32NN parameter set or the parameters subsequently derived for the combined data set.) The four prediction methods evaluated are as follows:

(1) Calculation of T_M and ΔT_M using a prediction algorithm from Exiqon (*32*), which is based on a large database of LNA oligonucleotide melting curves and on nearest-neighbor parameters for DNA, but uses an algorithm that is not described in detail. The T_M values were obtained from www.lna-tm.com, with the $[Na^+] = 999$ mM.

(2) Calculation of ΔT_M by simply assuming that it transfers directly from the corresponding training set oligonucleotide with the same M**X**$^{\mathbf{L}}$N trinucleotide. As discussed in the introduction, this approach ignores the temperature and context dependence of $\Delta\Delta G°$, but it does not use the more uncertain $\Delta\Delta H°$ and $\Delta\Delta S°$ values.

(3) Calculation of ΔT_M for LNA-containing oligonucleotides using the 16(RY)XL(RY) $\Delta\Delta H°$ and $\Delta\Delta S°$ values for LNA and the measured $\Delta H°$ and $\Delta S°$ from the test data set DNA reference oligonucleotides DTn.

(4) Calculation of T_M and ΔT_M using the 16(RY)XL(RY) $\Delta\Delta H°$ and $\Delta\Delta S°$ for LNA and the unified set $\Delta H°$ and $\Delta S°$ parameters for DNA (*12*).

The results of all of these methods are compared in Figure 5 for predictions of T_M and ΔT_M. Figure 5A shows that the Exiqon algorithm, method 1, overestimates T_M by an average of 1 °C, and that errors occur for all lengths, not only the LT4 and LT5 families of longer sequences. The RMSDs for experiment vs prediction (for LNA oligonucleotides only) for T_M and ΔT_M are 2.41 and 2.15 °C, respectively, showing that the errors for LNA T_M are not due to errors in the reference DNA T_M. Exiqon quotes a standard deviation of 5 °C for T_M predictions (*32*), but presumably this estimate includes multiply substituted oligonucleotides. Figure 5B shows that method 2, as expected, gives poor predictions for ΔT_M, which illustrates the quality of prediction available before this work and that of Exiqon. Direct transfer of $\Delta\Delta H°$ and $\Delta\Delta S°$ values was shown above (Table 3) to give much poorer predictions than method 3.

Method 3 gives much better ΔT_M estimates, with RMSD$(\Delta T_M) = 1.06$ °C and a maximum error of 2.31 °C. Method 4 provides the best predictions, giving T_M within 2.6 °C of experiment for all 33 test sequences except LT3c (predicted 66.0 °C, observed 62.7 °C), ΔT_M within 1.6 °C of the experiment for all except LT2b (predicted 3.94 °C, observed 5.88 °C), and RMSD$(T_M) = 1.47$ °C and RMSD$(\Delta T_M) = 0.85$ °C. Calculated reference $\Delta H°$ and $\Delta S°$ values presumably give better predictions because of errors in the measured $\Delta H°$ and $\Delta S°$.

In summary, we have demonstrated that the nearest-neighbor thermodynamic approach provides accurate predictions for the hybridization of LNA-containing oligonucleotides, using the $\Delta\Delta H°$ and $\Delta\Delta S°$ values obtained in this work in concert with the unified thermodynamic parameters for natural DNA. In addition, we have spot-checked the salt concentration dependence of the LNA−DNA thermodynamics (data not shown). We find that the $\Delta H°$ and $\Delta S°$ both change with salt, but that the standard approach of absorbing the salt dependence into $\Delta S°$ (*12*) gives accurate predictions for the change in T_M due to salt.

Refinement and Application of Predictive Parameter Sets. Having established the predictive value of the LNA parameter sets, we repeated the SVD analysis on the combined data set of 100 oligonucleotides, to derive nearest-neighbor values from all of experimental the data. The χ^2_{comb} and Q values of the fits obtained are given in Table 3, Table 4 lists the $\Delta\Delta H°$, $\Delta\Delta S°$, and $\Delta\Delta G°_{37}$ values for the different dinucleotides, and Figure 4B illustrates the agreement between experiment and prediction.

The 32NN parameter set, with the addition of the test data set to the training data set, now gives the smallest χ^2 values vs the four experimental observables, as well as the highest Q values (Table 3). Figure 4B shows that it provides excellent agreement with the experimental T_M, with RMSD$(\Delta T_M) = 0.88$ °C. The predictions for $\Delta\Delta H°$ and $\Delta\Delta G°_{37}$ are accurate except for the LT4 and LT5 families, as discussed above.

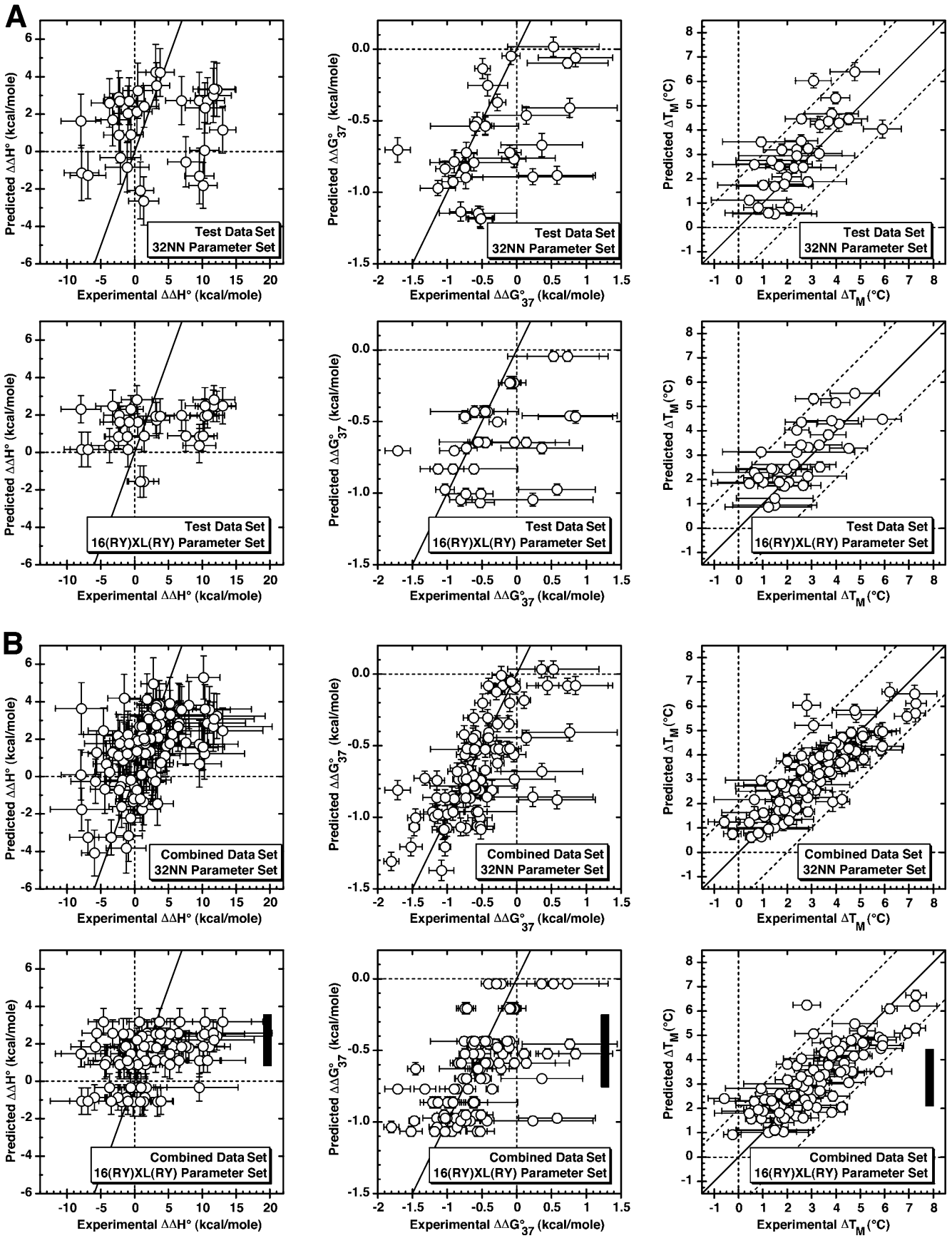

FIGURE 4: Predicted vs observed $\Delta\Delta H°$, $\Delta\Delta G°_{37}$, and ΔT_M for 32 NN and 16RYLRY parameter sets. Solid lines in each graph indicate perfect agreement, and the diagonal dashed lines in the ΔT_M panels delineate prediction within 2 °C of experiment. Panel A shows the results for the 33 test data set oligonucleotides, with predictions from the 32NN and 16(RY)XL(RY) parameter sets derived only from the training data set. Panel B shows the results for the 100 combined data set oligonucleotides, with predictions from the 32NN and 16(RY)-XL(RY) parameter sets recomputed from the combined data. The gray bars at the right indicate the range of predictions using the average parameters for the four LNA bases.

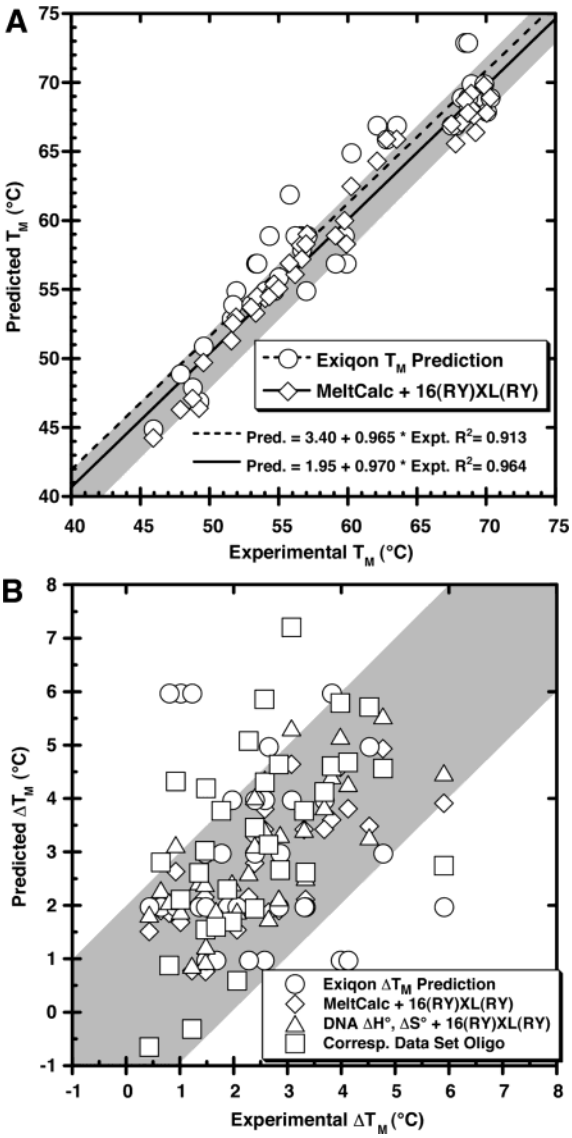

FIGURE 5: Comparison of methods for predicting the T_M and ΔT_M of test data set LNA oligonucleotides. The gray areas on each panel represent agreement to within 2°C between prediction and experiment. Panel A shows T_M predictions vs experiment for the Exiqon (*32*) algorithm (method 1 in Results, circles) and for the 16(RY)-XL(RY) parameter set (method 4, diamonds). The $\Delta\Delta H°$ and $\Delta\Delta S°$ parameters were derived solely from the training data set, and $\Delta H°$ and $\Delta S°$ were derived from the unified DNA parameters (*12*) using MeltCalc (*38*). Panel B shows the ΔT_M values obtained using methods 1 and 4 as in Panel A, as well as ΔT_M values from using the 16RYXLRY $\Delta\Delta H°$ and $\Delta\Delta S°$ parameters with measured $\Delta H°$ and $\Delta S°$ values for reference DNA oligonucleotides (method 3, triangles), and ΔT_M values simply transferred from the corresponding training data set oligonucleotide (method 2, squares). See Results for details.

Since the LT4 and LT5 families comprise only 12% of the data, and the SVD procedure underweights their relatively imprecise values for $\Delta\Delta H°$ and $\Delta\Delta S°$, the inclusion of the LT4 and LT5 data does not skew the values of the final combined data set parameters substantially.

The 16(RY)XL(RY) predictions for $\Delta\Delta H°$, $\Delta\Delta G°_{37}$, and ΔT_M underestimate the observed extent of variation (note compression of predicted values on the *y*-axes of Figure 4). The RMSD(ΔT_M) of 1.02 °C is low, but ΔT_M is systematically overestimated for small ΔT_M and underestimated for large ΔT_M. Apparently, the sequence variation in $\Delta\Delta H°$ and $\Delta\Delta G°_{37}$ cannot be represented by 16 nearest neighbors. (The 16(RY)XL(RY) set gives only 16 different MX^LN trinucleotide values, vs 64 different although not independent trinucleotides for the 32 NN set.) Figure 4B shows that even the 32NN set shows some compression of predicted $\Delta\Delta H°$, $\Delta\Delta G°_{37}$, and ΔT_M vs experiment, suggesting that a 64-parameter trinucleotide or even longer-range model may be necessary to capture the full sequence variation; this could be evaluated with more experimental data. The predictions cannot simply be rescaled to improve accuracy: since, a priori, we do not know which of the predictions are overestimates or underestimates, scaling introduces unpredictable errors across the whole set. Finally, using the averaged $\Delta\Delta H°$, $\Delta\Delta S°$, $\Delta\Delta G°_{37}$ or ΔT_M values for the four LNA bases underestimates the true extent of variability much more (bars beside the bottom three plots in Figure 4B), and gives RMSD(ΔT_M) = 1.37 °C.

Proposed LNA Incorporation Algorithm. The 32NN parameter set derived from the combined data set gives the best predictions for the T_M of a new singly LNA-substituted sequence. To use it, calculate $\Delta H°$ and $\Delta S°$ of the corresponding all-DNA duplex (at the relevant strand and salt concentrations) using the unified DNA parameters (*12*), conveniently done within MeltCalc (*38*). Then add the $\Delta\Delta H°$ and $\Delta\Delta S°$ values from Table 4 for the two LNA nearest-neighbor pairs, and calculate the new T_M using eq 5. To choose which base of a probe to substitute, step through a candidate sequence, predict the T_M for LNA substitution at each position, and choose the position that brings the T_M closest to a predetermined ideal value. To predict the comparative stabilities of LNA-containing DNA duplexes at 37 °C, use the $\Delta\Delta G°_{37}$ values in Table 4.

DISCUSSION

Nucleic acid analogues can improve on the stability and specificity of natural DNA and RNA. A variety of analogues, with accompanying design software, may be necessary to industrialize nucleic acid testing, with the goals of reducing the frequency of design failures, the need for expensive laboratory optimization, and the need for extensive training in probe design. LNA is one of the most promising nucleic acid analogues.

This work is the first systematic exploration of sequence-dependent LNA thermodynamics presented in terms of enthalpy and entropy increments ascribable to LNA. The results show that the effects of LNA are strongly context-dependent. LNA pyrimidines provide much more enhancement of stability than purines, but neighboring purines tend to enhance stability. The origin of increased stability is usually entropic stabilization of the duplex, but the most stabilizing substitutions are often enthalpically stabilized, especially by guanine neighbors. Strong apparent enthalpy−entropy compensation is observed. The 32NN set of nearest-neighbor $\Delta\Delta H°$ and $\Delta\Delta S°$ parameters in Table 4 provides accurate predictions for the stability of singly LNA-containing DNA. The values can be used within the framework of standard primer design algorithms.

Perfect-match nearest-neighbor parameters do not complete the description of LNA thermodynamics. The most useful extensions would be to LNA:DNA mismatches, the cooperative effects of multiple substitutions, and LNA at the termini.

Table 4: Nearest-Neighbor Thermodynamic Parameter Sets for LNA Incorporation

32NN Parameter Set Derived from Combined Data Set[a]

nearest neighbors	$\Delta\Delta H°$ (kcal/mol)	$\Delta\Delta S°$ (cal/mol K)	$\Delta\Delta G°_{37}$ (kcal/mol)
5′ **A**$^{\mathbf{L}}$A 3′	0.707 ± 0.694	2.477 ± 2.135	−0.092 ± 0.028
A$^{\mathbf{L}}$C	1.131 ± 0.840	4.064 ± 2.550	−0.122 ± 0.042
A$^{\mathbf{L}}$G	0.264 ± 0.858	2.613 ± 2.635	−0.561 ± 0.030
A$^{\mathbf{L}}$T	2.282 ± 0.972	7.457 ± 2.972	−0.007 ± 0.050
C$^{\mathbf{L}}$A	1.049 ± 0.957	4.320 ± 2.915	−0.270 ± 0.049
C$^{\mathbf{L}}$C	2.096 ± 0.941	7.996 ± 2.807	−0.457 ± 0.050
C$^{\mathbf{L}}$G	0.785 ± 0.930	3.709 ± 2.837	−0.332 ± 0.034
C$^{\mathbf{L}}$T	0.708 ± 0.715	4.175 ± 2.251	−0.666 ± 0.043
G$^{\mathbf{L}}$A	3.162 ± 0.686	10.544 ± 2.101	−0.072 ± 0.041
G$^{\mathbf{L}}$C	−0.360 ± 0.831	−0.251 ± 2.551	−0.414 ± 0.039
G$^{\mathbf{L}}$G	−2.844 ± 0.964	−6.680 ± 2.627	−0.700 ± 0.046
G$^{\mathbf{L}}$T	−0.212 ± 0.865	0.073 ± 2.952	−0.194 ± 0.040
T$^{\mathbf{L}}$A	−0.046 ± 0.960	1.562 ± 3.061	−0.563 ± 0.049
T$^{\mathbf{L}}$C	1.893 ± 1.010	6.685 ± 2.957	−0.208 ± 0.043
T$^{\mathbf{L}}$G	−1.540 ± 0.701	−3.044 ± 2.112	−0.548 ± 0.040
T$^{\mathbf{L}}$T	1.528 ± 0.986	5.298 ± 2.990	−0.130 ± 0.036
5′ A**A**$^{\mathbf{L}}$ 3′	0.992 ± 0.691	4.065 ± 3.128	−0.396 ± 0.034
A**C**$^{\mathbf{L}}$	2.890 ± 1.018	10.576 ± 2.142	−0.390 ± 0.035
A**G**$^{\mathbf{L}}$	−1.200 ± 0.758	−1.826 ± 2.327	−0.603 ± 0.045
A**T**$^{\mathbf{L}}$	1.816 ± 0.844	6.863 ± 2.673	−0.309 ± 0.038
C**A**$^{\mathbf{L}}$	1.358 ± 0.875	4.367 ± 2.664	0.046 ± 0.032
C**C**$^{\mathbf{L}}$	2.063 ± 0.977	7.565 ± 2.958	−0.404 ± 0.027
C**G**$^{\mathbf{L}}$	−0.276 ± 0.988	−0.718 ± 2.914	−0.003 ± 0.047
C**T**$^{\mathbf{L}}$	−1.671 ± 0.619	−4.070 ± 1.919	−0.409 ± 0.040
G**A**$^{\mathbf{L}}$	0.444 ± 0.897	2.898 ± 2.757	−0.437 ± 0.053
G**C**$^{\mathbf{L}}$	−0.925 ± 0.944	−1.111 ± 2.880	−0.535 ± 0.046
G**G**$^{\mathbf{L}}$	−0.943 ± 0.918	−0.933 ± 2.842	−0.666 ± 0.054
G**T**$^{\mathbf{L}}$	−0.635 ± 0.901	−0.342 ± 2.734	−0.520 ± 0.046
T**A**$^{\mathbf{L}}$	1.591 ± 0.872	5.281 ± 2.877	0.004 ± 0.038
T**C**$^{\mathbf{L}}$	0.609 ± 0.916	3.169 ± 2.663	−0.396 ± 0.046
T**G**$^{\mathbf{L}}$	2.165 ± 0.922	7.163 ± 2.715	−0.106 ± 0.040
T**T**$^{\mathbf{L}}$	2.326 ± 0.790	8.051 ± 2.438	−0.212 ± 0.044

Average Values Derived from 32NN Parameter Set

class	$\Delta\Delta H°$ (kcal/mol)	$\Delta\Delta S°$ (cal/mol K)	$\Delta\Delta G°_{37}$ (kcal/mol)
Average Effects of Different LNA Bases[b]			
M**A**$^{\mathbf{L}}$N (LNA-A)[b]	2.192	8.305	−0.391
M**C**$^{\mathbf{L}}$N (LNA-C)[b]	2.319	10.100	−0.863
M**G**$^{\mathbf{L}}$N (LNA-G)[b]	−0.127	1.843	−0.689
M**T**$^{\mathbf{L}}$N (LNA-T)[b]	0.917	5.251	−0.725
all LNA bases[b]	1.325	6.375	−0.667
Average Effects of Different 5′ Neighbors			
A**X**$^{\mathbf{L}}$N	1.125	4.920	−0.424
C**X**$^{\mathbf{L}}$N	0.368	1.786	−0.192
G**X**$^{\mathbf{L}}$N	−0.515	0.128	−0.540
T**X**$^{\mathbf{L}}$N	1.673	5.916	−0.178
Average Effects of Different 3′ Neighbors			
M**X**$^{\mathbf{L}}$A	1.218	4.726	−0.249
M**X**$^{\mathbf{L}}$C	1.190	4.624	−0.300
M**X**$^{\mathbf{L}}$G	−0.834	−0.851	−0.535
M**X**$^{\mathbf{L}}$T	1.077	4.251	−0.249

16(RY)XL(RY) Parameter Set Derived from Combined Data Set[a]

nearest neighbors	$\Delta\Delta H°$ (kcal/mol)	$\Delta\Delta S°$ (cal/mol K)	$\Delta\Delta G°_{37}$ (kcal/mol)
5′ **A**$^{\mathbf{L}}$Y 3′	1.742 ± 0.434	5.886 ± 1.329	−0.077 ± 0.018
A$^{\mathbf{L}}$R	0.435 ± 0.611	2.381 ± 1.863	−0.249 ± 0.030
C$^{\mathbf{L}}$Y	1.215 ± 0.579	5.590 ± 1.766	−0.584 ± 0.028
C$^{\mathbf{L}}$R	0.847 ± 0.475	3.752 ± 1.449	−0.290 ± 0.020
G$^{\mathbf{L}}$Y	−0.084 ± 0.598	0.572 ± 1.829	−0.322 ± 0.021
G$^{\mathbf{L}}$R	0.137 ± 0.445	1.875 ± 1.307	−0.410 ± 0.029
T$^{\mathbf{L}}$Y	1.810 ± 0.594	6.255 ± 1.808	−0.167 ± 0.030
T$^{\mathbf{L}}$R	−1.041 ± 0.443	−1.477 ± 1.400	−0.553 ± 0.026
5′ Y**A**$^{\mathbf{L}}$ 3′	1.473 ± 0.565	4.756 ± 1.726	0.048 ± 0.027
R**A**$^{\mathbf{L}}$	0.703 ± 0.619	3.512 ± 1.894	−0.374 ± 0.031
Y**C**$^{\mathbf{L}}$	1.395 ± 0.418	5.558 ± 1.332	−0.401 ± 0.009
R**C**$^{\mathbf{L}}$	0.666 ± 0.514	3.784 ± 1.519	−0.473 ± 0.020
Y**G**$^{\mathbf{L}}$	1.014 ± 0.472	3.492 ± 1.461	−0.108 ± 0.022
R**G**$^{\mathbf{L}}$	−0.962 ± 0.505	−1.045 ± 1.432	−0.623 ± 0.028
Y**T**$^{\mathbf{L}}$	0.021 ± 0.556	1.031 ± 1.747	−0.305 ± 0.029
R**T**$^{\mathbf{L}}$	0.749 ± 0.580	3.747 ± 1.795	−0.415 ± 0.019

[a] The uncertainties given are the estimated errors in the value of the parameter from SVD analysis, not a standard deviation in the population. [b] These averages are multiplied by 2, as each LNA has 5′ and 3′ neighbors.

LNA increases the specificity of allele-specific primers when singly incorporated at the 3′ end of an allele-specific primer (*39*). All-LNA capture probes have been used for the genotyping of SNPs in apolipoprotein E (*40*) and for the detection of the Factor V Leiden mutation (*41*).

The Origin of Stabilization Induced by LNA. Some of the quantitative results reported here have been anticipated by qualitative work in the literature. Preorganization, improved stacking, or both have generally been advanced as likely sources for stabilization by LNA. It is recognized that LNA substitution gives a smaller ΔT_M in longer oligonucleotides. Thermodynamics other than ΔT_M measurements, however, have been reported only infrequently. Obika et al. measured a large positive $\Delta\Delta S°$ and negligible $\Delta\Delta H°$ for a gapmer with seven LNA (*24*). Koshkin et al. measured $\Delta\Delta H°$ and $\Delta\Delta S°$ values for two LNA+DNA oligonucleotides with three or six LNA nucleotides (*29*). The former showed a small positive $\Delta\Delta H°$ and the latter an extremely large negative $\Delta\Delta H°$, but both had a positive $\Delta\Delta S°$. Christensen et al. measured small positive $\Delta\Delta H°$ and $\Delta\Delta S°$ values for oligonucleotides containing one to three LNA-T residues (*42*). LNA residues at the termini of an oligonucleotide have a smaller effect on T_M than those in the center (*28*, *43*). LNA in "mixmers" with substitutions scattered throughout a sequence has a greater effect per substitution than in "gapmers" with clustered substitutions (*28*), which Wengel (*3*) has termed "structural saturation." There is little published information on sequence dependence, although data reviewed (*4*) and obtained (*43*) by Corey's group hint at decreased stabilization by LNA purines. Exiqon provides a web-based tool for T_M prediction based on 1400 LNA-containing sequences, but gives little detail on how the prediction is made (*32*). We show above that the thermodynamics derived here give superior results for singly substituted oligonucleotides, but the tools at www.lna-tm.com allow for a wider variety of substitution patterns.

Structural studies have not reached clear conclusions on the origin of increased stability, and use exclusively **T**$^{\mathbf{L}}$. The NMR studies of Petersen et al. (*26*) showed that the structural perturbations caused by LNA are quite local, and interpreted chemical shift data as indicating both preorganization and improved stacking as stabilizing forces: our results show that either enthalpy or entropy, but not both for the same sequence, may contribute to stability. Nielsen et al. (*44*) suggested that a decreased helical rise observed in an (LNA+DNA):DNA duplex reflects improved stacking, but Jensen et al. (*45*) subsequently reported no obvious improved stacking or other structural causes for stabilization in a similar (LNA-DNA):DNA duplex, and concluded that stabilization is largely entropic. Multiple LNA substitutions can convert

an entire LNA-DNA mixmer to A form geometry when hybridized to RNA (*46*). An X-ray structure shows complete conversion of DNA to A form due to just one LNA per strand (*25*). All of the NMR results are complicated by mixed populations of C3′-endo and C2′-endo sugar puckers in the DNA nucleotides next to an LNA, which makes it impossible to obtain a unique structure. This suggests the novel possibility that stabilization can arise because the entropy of an LNA+DNA:DNA duplex is larger than that of DNA:DNA. The diversity of structural results suggests substantial plasticity in LNA-containing oligonucleotides, allowing different modes of stabilization and enthalpy−entropy compensation.

Thermodynamic Parameters for Other Modified Nucleic Acids. Martin et al. (*5*) investigated the nearest-neighbor energies for the G analogue deoxyinosine (dI) by incorporating a single dI into a DNA strand. They extracted thermodynamic parameters for 8 of the 64 possible (DNA+I):DNA triplets, from which it is possible to estimate 16 of the 32 possible nearest-neighbor parameters. In this study, each test sequence contributed to two nearest-neighbor parameters, and each nearest-neighbor parameter estimate was derived from a single sequence.

Vallone and Benight (*6*) have studied the universality of 5-nitroindole by incorporating a single 5-nitroindole into a DNA strand that formed a hairpin structure with a 7-bp stem and a 4-bp loop. They extracted thermodynamic parameters for two of the 64 possible (DNA+N):DNA triplets, finding that 5-nitroindole has a substantial destabilizing effect, decreasing the T_M by an average of 18.0°. They also observed enthalpy−entropy compensation.

PNA has been investigated more thoroughly than any other nucleic acid analogue, because its high stability and uncharged backbone make it attractive as a therapeutic. Jensen et al. (*47*) investigated a single PNA:PNA, PNA:RNA, and PNA:DNA perfect match duplex, as well as 12 single-base mismatches. They found that perfect match PNA:PNA was most stable, followed by PNA:RNA, PNA:DNA, and DNA:DNA, and that PNA:RNA and PNA:DNA mismatches were more destabilizing than identical DNA:DNA mismatches. Igloi (*48*) characterized the thermodynamics of single PNA:DNA mismatches using one 11-mer PNA strand and 33 DNA complements, each with a single-base mismatch. The 33 PNA:DNA duplexes thus represented all three possible single-base mismatches at each of the 11 positions. Mismatches nearer the center had greater destabilizing effects than mismatches toward the ends.

Giesen et al. (*20*) developed a linear regression model for predicting PNA:DNA T_M from a data set of 316 sequences of known T_M. The best predictive model included an intercept term and coefficients applied to the equivalent DNA nearest-neighbor T_M, a the fractional pyrimidine content, and the length. In an independent set of 44 PNA:DNA duplexes, this model was able to predict T_M within ∼5° for ∼90% of the sequences.

Griffin and Smith (*22*) modeled PNA:DNA stability by augmenting the DNA:DNA nearest-neighbor parameters with terms for dangling ends, for the number of PNA:DNA nearest-neighbors in the duplex, and for ionic strength. As noted by Sugimoto et al. (*21*), the term for the number of PNA:DNA nearest-neighbors captures the overall effect of PNA as compared to DNA and is not a nearest-neighbor parameter. This model made T_M predictions that differed by as much as 13.9° from the observed T_M in an independent test set of duplexes, possibly because the PNA parameters were derived from one 9-bp PNA:DNA that included only 7 of the 16 possible nearest neighbors (*21*).

Sugimoto et al. (*21*) investigated the applicability of the nearest-neighbor model to PNA by testing whether duplexes with different sequences but the same nearest-neighbor composition have the same values for the thermodynamic parameters. DNA:RNA duplexes and short PNA:DNA duplexes with equal numbers of nearest-neighbor dinucleotides gave $\Delta H°$, $\Delta S°$, and $\Delta G°_{37}$ values that differed on average by less than 6%. However, PNA:DNA duplexes of length 10 or more did not conform well to the nearest-neighbor model. In this data set of 20 sequences, the T_M prediction methods of Griffin and Smith (*22*) and Giesen (*20*) had RMSDs of 19.7 and 6.0 °C, respectively.

Prospects for Complete Thermodynamics of Nucleic Acid Analogues. This work is the first to provide complete nearest-neighbor thermodynamics for a modified nucleotide in DNA. The T_M predictions for singly modified LNA+DNA hybridized to DNA are nearly as accurate as those for unmodified DNA. Limiting the work to singly modified oligonucleotides made it possible to treat sequence context in detail, and in many cases even a single LNA may enable a probe to hybridize at a desired temperature.

Thermodynamic studies of nucleic acid analogues face similar challenges to those faced in early studies of DNA and RNA, and some new ones. The synthesis of homogeneous and chimeric oligonucleotides can be technically challenging and expensive. Homogeneous single strands such as PNA may have aggregation or solubility problems. The choice of model is not always clear: duplexes may not exhibit two-state transitions, or nearest-neighbor models may not apply, or they may apply only to very short duplexes. The new challenge for nucleic acid analogues is that the number of parameters needed to describe them may be very large, as there are many ways to incorporate even one type of modification, and there are hundreds of commercially available nucleic acid analogues (Glen Research).

The large number of combinations means that a thermodynamic study must include a large number of duplexes, even to cover only a subset of possible incorporation strategies. Modifications give different results at internal positions and ends. Parameters for internal incorporations should apply to multiple incorporations if they are spaced far enough apart, but closely spaced incorporations are unlikely to have simple independent and additive effects. This complexity results in the conundrum of a promising chemistry perhaps not being widely adopted because of a lack of design rules, and a lack of design rules because the chemistry has not been widely adopted. In this work, we have provided a framework that allows the extraction of useful parameters with a reasonable investment of time and money. It can be applied to new chemistries and incorporation strategies during their initial development.

ACKNOWLEDGMENT

We are grateful to Lawrence Kessner (Celadon) and members of the Kahn laboratory for support, and to Alex Amiet of PrOligo for responsive LNA oligonucleotide

synthesis. Patent protection for the thermodynamic values derived has been applied for under no. 10/685,711.

REFERENCES

1. Kutyavin, I. V., Afonina, I. A., Mills, A., Gorn, V. V., Lukhtanov, E. A., Belousov, E. S., Singer, M. J., Walburger, D. K., Lokhov, S. G., Gall, A. A., Dempcy, R., Reed, M. W., Meyer, R. B., and Hedgpeth, J. (2000) 3′-minor groove binder-DNA probes increase sequence specificity at PCR extension temperatures. *Nucleic Acids Res. 28*, 655−661.
2. Znosko, B. M., Barnes, T. W., 3rd, Krugh, T. R., and Turner, D. H. (2003) NMR studies of DNA single strands and DNA:RNA hybrids with and without 1-propynylation at C5 of oligopyrimidines. *J. Am. Chem. Soc. 125*, 6090−6097.
3. Petersen, M., and Wengel, J. (2003) LNA: A versatile tool for therapeutics and genomics. *Trends Biotechnol. 21*, 74−81.
4. Braasch, D. A., and Corey, D. R. (2001) Locked nucleic acid (LNA): Fine-tuning the recognition of DNA and RNA. *Chem. Biol. 8*, 1−7.
5. Martin, F. H., Castro, M. M., Aboul-ela, F., and Tinoco, I., Jr. (1985) Base pairing involving deoxyinosine: Implications for probe design. *Nucleic Acids Res. 13*, 8927−8938.
6. Vallone, P. M., and Benight, A. S. (1999) Melting studies of short DNA hairpins containing the universal base 5-nitroindole. *Nucleic Acids Res. 27*, 3589−3596.
7. Lok, C. N., Viazovkina, E., Min, K. L., Nagy, E., Wilds, C. J., Damha, M. J., and Parniak, M. A. (2002) Potent gene-specific inhibitory properties of mixed-backbone antisense oligonucleotides comprised of 2′-deoxy-2′-fluoro-D-arabinose and 2′-deoxyribose nucleotides. *Biochemistry 41*, 3457−3467.
8. Kutyavin, I. V., Lokhov, S. G., Afonina, I. A., Dempcy, R., Gall, A. A., Gorn, V. V., Lukhtanov, E., Metcalf, M., Mills, A., Reed, M. W., Sanders, S., Shishkina, I., and Vermeulen, N. M. J. (2002) Reduced aggregation and improved specificity of G-rich oligodeoxyribonucleotides containing pyrazolo[3,4-*d*]pyrimidine guanine bases. *Nucleic Acids Res. 30*, 4952−4959.
9. Demidov, V. V., Protozanova, E., Izvolsky, K. I., Price, C., Nielsen, P. E., and Frank-Kamenetskii, M. D. (2002) Kinetics and mechanism of the DNA double helix invasion by pseudocomplementary peptide nucleic acids. *Proc. Natl. Acad. Sci. U.S.A. 99*, 5953−5958.
10. Zuker, M. (2000) Calculating nucleic acid secondary structure. *Curr. Opin. Struct. Biol. 10*, 303−310.
11. Xia, T., SantaLucia, J., Jr., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X., Cox, C., and Turner, D. H. (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson−Crick base pairs. *Biochemistry 37*, 14719−14735.
12. SantaLucia, J., Jr. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. U.S.A. 95*, 1460−1465.
13. SantaLucia, J., Jr., Allawi, H. T., and Seneviratne, P. A. (1996) Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry 35*, 3555−3562.
14. Sugimoto, N., Nakano, S., Katoh, M., Matsumura, A., Nakamuta, H., Ohmichi, T., Yoneyama, M., and Sasaki, M. (1995) Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry 34*, 11211−11216.
15. Borer, P. N., Dengler, B., Tinoco, I., Jr., and Uhlenbeck, O. (1974) Stability of ribonucleic acid double-stranded helices. *J. Mol. Biol. 86*, 843−853.
16. Peyret, N., Seneviratne, P. A., Allawi, H. T., and SantaLucia, J., Jr. (1999) Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A·A, C·C, G·G, and T·T mismatches. *Biochemistry 38*, 3468−3477.
17. Bommarito, S., Peyret, N., and SantaLucia, J., Jr. (2000) Thermodynamic parameters for DNA sequences with dangling ends. *Nucleic Acids Res. 28*, 1929−1934.
18. Tuerk, C., Gauss, P., Thermes, C., Groebe, D. R., Gayle, M., Guild, N., Stormo, G., d'Aubenton-Carafa, Y., Uhlenbeck, O. C., Tinoco, I., Jr., Brody, E. N., and Gold, L. (1988) CUUCGG hairpins: Extraordinarily stable RNA secondary structures associated with various biochemical processes. *Proc. Natl. Acad. Sci. U.S.A. 85*, 1364−1368.
19. Chen, X., Kierzek, R., and Turner, D. H. (2001) Stability and structure of RNA duplexes containing isoguanosine and isocytidine. *J. Am. Chem. Soc. 123*, 1267−1274.
20. Giesen, U., Kleider, W., Berding, C., Geiger, A., Ørum, H., and Nielsen, P. E. (1998) A formula for thermal stability (Tm) prediction of PNA/DNA duplexes. *Nucleic Acids Res. 26*, 5004−5006.
21. Sugimoto, N., Satoh, N., Yasuda, K., and Nakano, S. (2001) Stabilization factors affecting duplex formation of peptide nucleic acid with DNA. *Biochemistry 40*, 8444−8451.
22. Griffin, T. J., and Smith, L. M. (1998) An approach to predicting the stabilities of peptide nucleic acid:DNA duplexes. *Anal. Biochem. 260*, 56−63.
23. Singh, S. K., Nielsen, P., Koshkin, A. A., and Wengel, J. (1998) LNA (locked nucleic acids): Synthesis and high-affinity nucleic acid recognition. *Chem. Commun. 4*, 455−456.
24. Obika, S., Nanbu, D., Hari, Y., Andoh, J.-i., Morio, K.-i., Doi, T., and Imanishi, T. (1998) Stability and structural features of the duplexes containing nucleoside analogues with a fixed N-type conformation, 2′-O,4′-C-methyleneribonucleosides. *Tetrahedron Lett. 39*, 5401−5404.
25. Egli, M., Minasov, G., Teplova, M., Kumar, R., and Wengel, J. (2001) X-ray crystal structure of a locked nucleic acid (LNA) duplex composed of a palindromic 10-mer DNA strand containing one LNA thymine monomer. *J. Chem. Soc., Chem. Commun. 7*, 651−652.
26. Petersen, M., Nielsen, C. B., Nielsen, K. E., Jensen, G. A., Bondensgaard, K., Singh, S. K., Rajwanshi, V. K., Koshkin, A. A., Dahl, B. M., Wengel, J., and Jacobsen, J. P. (2000) The conformations of locked nucleic acids (LNA). *J. Mol. Recognit. 13*, 44−53.
27. Braasch, D. A., Liu, Y., and Corey, D. R. (2002) Antisense inhibition of gene expression in cells by oligonucleotides incorporating locked nucleic acids: Effect of mRNA target sequence and chimera design. *Nucleic Acids Res. 30*, 5160−5167.
28. Kurreck, J., Wyszko, E., Gillen, C., and Erdmann, V. A. (2002) Design of antisense oligonucleotides stabilized by locked nucleic acids. *Nucleic Acids Res. 30*, 1911−1918.
29. Koshkin, A. A., Nielsen, P., Meldgaard, M., Rajwanshi, V. K., Singh, S. K., and Wengel, J. (1998) LNA (locked nucleic acid): An RNA mimic forming exceedingly stable LNA:LNA duplexes. *J. Am. Chem. Soc. 120*, 13252−13253.
30. Torigoe, H., Hari, Y., Sekiguchi, M., Obika, S., and Imanishi, T. (2001) 2′-O,4′-C-methylene bridged nucleic acid modification promotes pyrimidine motif triplex DNA formation at physiological pH: Thermodynamic and kinetic studies. *J. Biol. Chem. 276*, 2354−2360.
31. Braasch, D. A., and Corey, D. R. (2002) Novel antisense and peptide nucleic acid strategies for controlling gene expression. *Biochemistry 41*, 4503−4510.
32. Tolstrup, N., Nielsen, P. S., Kolberg, J. G., Frankel, A. M., Vissing, H., and Kauppinen, S. (2003) OligoDesign: Optimal design of LNA (locked nucleic acid) oligonucleotide capture probes for gene expression profiling. *Nucleic Acids Res. 31*, 3758−3762.
33. Zuker, M. (1989) On finding all suboptimal foldings of an RNA molecule. *Science 244*, 48−52.
34. Liu, L., and Guo, Q.-X. (2001) Isokinetic relationship, isoequilibrium relationship, and enthalpy−entropy compensation. *Chem. Rev. 101*, 673−695.
35. Florián, J., Sponer, J., and Warshel, A. (1999) Thermodynamic parameters for stacking and hydrogen bonding of nucleic acid bases in aqueous solution: Ab initio/Langevin dipoles study. *J. Phys. Chem. B 103*, 884−892.
36. Wu, P., Nakano, S., and Sugimoto, N. (2002) Temperature dependence of thermodynamic properties for DNA/DNA and RNA/DNA duplex formation. *Eur. J. Biochem. 269*, 2821−2830.
37. Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1989) *Numerical Recipes: The Art of Scientific Computing (FORTRAN version)*, 1st ed.; Cambridge University Press, Cambridge.
38. Schütz, E., and von Ahsen, N. (1999) Spreadsheet software for thermodynamic melting point prediction of oligonucleotide hybridization with and without mismatches. *Biotechniques 27*, 1218−1224.
39. Latorra, D., Campbell, K., Wolter, A., and Hurley, J. M. (2003) Enhanced allele-specific PCR discrimination in SNP genotyping using 3′ locked nucleic acid (LNA) primers. *Hum. Mutat. 22*, 79−85.
40. Jacobsen, N., Bentzen, J., Meldgaard, M., Jakobsen, M. H., Fenger, M., Kauppinen, S., and Skouv, J. (2002) LNA-enhanced detection of single nucleotide polymorphisms in the apolipoprotein E. *Nucleic Acids Res. 30*, e100.

41. Ørum, H., Jakobsen, M. H., Koch, T., Vuust, J., and Borre, M. B. (1999) Detection of the factor V Leiden mutation by direct allele-specific hybridization of PCR amplicons to photoimmobilized locked nucleic acids. *Clin. Chem. 45*, 1898−1905.
42. Christensen, U., Jacobsen, N., Rajwanshi, V. K., Wengel, J., and Koch, T. (2001) Stopped-flow kinetics of locked nucleic acid (LNA)-oligonucleotide duplex formation: Studies of LNA-DNA and DNA-DNA interactions. *Biochem. J. 354*, 481−484.
43. Braasch, D. A., Jensen, S., Liu, Y., Kaur, K., Arar, K., White, M. A., and Corey, D. R. (2003) RNA interference in mammalian cells by chemically modified RNA. *Biochemistry 42*, 7967−7975.
44. Nielsen, K. E., Singh, S. K., Wengel, J., and Jacobsen, J. P. (2000) Solution structure of an LNA hybridized to DNA: NMR study of the d($CT^{L}GCT^{L}T^{L}CT^{L}GC$):d(GCAGAAGCAG) duplex containing four locked nucleotides. *Bioconjugate Chem. 11*, 228−238.
45. Jensen, G. A., Singh, S. K., Kumar, R., Wengel, J., and Jacobsen, J. P. (2001) A comparison of the solution structures of an LNA:DNA duplex and the unmodified DNA:DNA duplex. *J. Chem. Soc., Perkin Trans. 2*, 1224−1232.
46. Petersen, M., Bondensgaard, K., Wengel, J., and Jacobsen, J. P. (2002) Locked nucleic acid (LNA) recognition of RNA: NMR solution structures of LNA:RNA hybrids. *J. Am. Chem. Soc. 124*, 5974−5982.
47. Jensen, K. K., Ørum, H., Nielsen, P. E., and Nordén, B. (1997) Kinetics for hybridization of peptide nucleic acids (PNA) with DNA and RNA studied with the BIAcore technique. *Biochemistry 36*, 5072−5077.
48. Igloi, G. L. (1998) Variability in the stability of DNA-peptide nucleic acid (PNA) single-base mismatched duplexes: Real-time hybridization during affinity electrophoresis in PNA-containing gels. *Proc. Natl. Acad. Sci. U.S.A. 95*, 8562−8567.

BI035976D