

Nearest-neighbor thermodynamics of deoxyinosine pairs in DNA duplexes

Norman E. Watkins, Jr and John SantaLucia, Jr*

Department of Chemistry, Wayne State University, Detroit, MI 48202, USA

Received August 11, 2005; Revised and Accepted October 4, 2005

ABSTRACT

Nearest-neighbor thermodynamic parameters of the 'universal pairing base' deoxyinosine were determined for the pairs I·C, I·A, I·T, I·G and I·I adjacent to G·C and A·T pairs. Ultraviolet absorbance melting curves were measured and non-linear regression performed on 84 oligonucleotide duplexes with 9 or 12 bp lengths. These data were combined with data for 13 inosine containing duplexes from the literature. Multiple linear regression was used to solve for the 32 nearest-neighbor unknowns. The parameters predict the T_m for all sequences within 1.2°C on average. The general trend in decreasing stability is I·C > I·A > I·T \approx I·G > I·I. The stability trend for the base pair 5' of the I·X pair is G·C > C·G > A·T > T·A. The stability trend for the base pair 3' of I·X is the same. These trends indicate a complex interplay between H-bonding, nearest-neighbor stacking, and mismatch geometry. A survey of 14 tandem inosine pairs and 8 tandem self-complementary inosine pairs is also provided. These results may be used in the design of degenerate PCR primers and for degenerate microarray probes.

INTRODUCTION

Inosine is biologically found in tRNA (1), and as an adenosine deamination product in DNA where it must be repaired to maintain genomic fidelity (2,3) and in RNA editing (4). Inosine's ability to act as a 'universal pairing base' was recognized soon after the sequences of many tRNA's became available and inosine was observed in the first anti-codon position, which pairs with the third codon position on mRNA (1). This observation led to the 'wobble hypothesis' (1). Despite the widespread acceptance of the notion of inosine as a universal pairing base, there has been relatively little scrutiny of this hypothesis. Figure 1 shows common structures for inosine pairs (5–7). While it is established that all of the I·X pairs

can form structures with 2-hydrogen bonds, the thermodynamic stabilities of the pairs have not been quantified.

More recently, the genomics revolution has resulted in important applications of inosine in DNA molecular biology. The most common application of inosine is in the determination of a mRNA sequence using degenerate hybridization probes (8–10). These probes have been shown to reduce the number of primers needed while increasing the priming efficiency (5). In degenerate primers, particular emphasis is placed on the wobble position to allow annealing of the primer to a variety of related sequences. For instance, if the wobble position of the codon is a purine, like arginine (AGR), Lysine (AAR) or glutamic acid (GAR), then the degeneracy is 2-fold where R is A or G. If the nature of the wobble position is completely ambiguous, like glycine (GGN), alanine (GCN) and valine (GUN), then the degeneracy is 4-fold where N is any of the four coding bases. If inosine is used in every ambiguous instance then the degeneracy is reduced to 1. The presence of multiple degenerate positions results in an exponential growth in total degeneracy. If one primer has five positions made up of any mixed base, then the degeneracy would be 4⁵ or 1024 primers. The inosine thermodynamic parameters presented here enable a more accurate design of primers and probes to protein coding regions, or other templates with ambiguous sequences (11–13).

Amplification of ambiguous sequences can be done in degenerate PCR using inosine containing degenerate primers (14,15). Previously, degenerate PCR was often done with primers with mixed bases at degenerate positions (16,17). This approach, however, results in relatively inefficient amplification due to imperfect hybridization (18). It is now routine to isolate a protein of interest and sequence the N-terminal amino acids (13,16). Using the amino acid sequence a degenerate primer can be designed so the gene of interest may be amplified and sequenced (15). Inosine may also be substituted in difficult guanine rich PCR primers to reduce G-quartet formation as well as primer-dimer artifacts (19).

Electrochemical based DNA-hybridization biosensors require guanine free probes (20). This is necessitated because of the low oxidation potential of guanine. This technology

*To whom correspondence should be addressed. Tel: +1 313 577 0101; Fax: +1 313 577 8822; Email: jsl@chem.wayne.edu

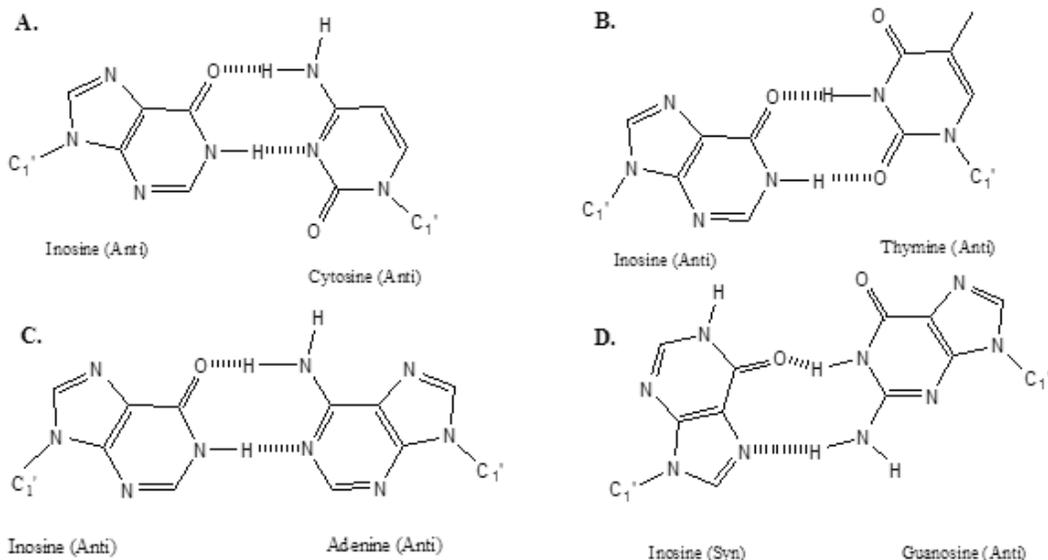


Figure 1. The proposed inosine-Watson-Crick configurations (5–7).

exploits the electrochemical activity of nucleic acids, by using the guanine oxidation signal. If both probe and target contain the base guanine, then hybridization does not cause a large change in electrochemical signal. In these electrochemical applications, inosine is usually substituted for guanine because it has similar pairing properties as guanine, but with a much higher oxidation potential. Thus, electrochemical signal is only observed upon hybridization of the natural guanosine-containing target to the inosine-containing probe (20).

For DNA microarray applications, inosine may be used to increase the stability of a library of oligos without increasing the diversity of the library (5,13–16). For instance, if one were to build a library of just octamers there would be 4^8 or 65 536 octamers possible. Also, shorter probes can result in relatively weak binding and in poor specificity for large genomes. Longer probes would result in exponential growth in the size of the probe library. Thus, addition of inosine may be used to increase the length and specificity of the probe without increasing the diversity of the library. The result is an increase of stability of a library of oligos without increasing its diversity. Chizhikov *et al.* (21) have exploited this stability in the design of a rotavirus classification microarray using inosine at ambiguous sites. This result can also be of use in identifying important novel RNA editing sites using microarrays in non-coding Alu repetitive clusters in mammals (4).

The parameters reported here may also be used in the ‘lab-on-a-chip’ technology currently under development (22). This technology involves microfluidics devices that are highly portable and reliable that facilitate rapid DNA analysis of blood, air and water samples for pathogens and chemicals using electrochemical or other detection as suggested by Wang *et al.* (20). Degenerate primers will be needed for unambiguous identification of the sample genotype using PCR amplification. Technology such as this could prove to be indispensable for pathogen identification for public health and biodefense.

Alternative universal bases are 3-nitropyrrole 2′ deoxynucleoside and 5-nitroindole 2′ deoxynucleoside (9,10,23). It has been shown by Bergstrom and coworkers (9) that sequencing ladders using primers containing contiguous 3-nitropyrrole and 5-nitroindole were indistinguishable from the normal natural base control primers. However, when these bases were dispersed throughout the primer, their efficiency fell per base added (10). Also, as reviewed by Loakes (23), non-hydrogen bonding universal bases are not efficiently extended by DNA polymerase. Also 5-nitroindole is heterogeneously iodinated during the oxidation step of phosphoramidite synthesis resulting in a mixture of compounds whose thermodynamics are unknown (24).

One advantage of inosine over 3-nitropyrrole and 5-nitroindole is that it does not have to occur consecutively in the primer to be effective, and inosine containing oligos are more stable than oligos with 3-nitropyrrole and 5-nitroindole where the T_m is decreased by 15–35°C on average (25). The position dependence for addition of universal bases was tested across six wobble positions in a 20mer oligonucleotide primer. Overall, in PCR experiments neither 3-nitropyrrole nor 5-nitroindole were as effective in the wobble position as deoxyinosine (10). With more detailed knowledge of inosine thermodynamics, a more efficient scheme for primer design can be devised, particularly for cases where the degeneracy is partially known.

In this paper, accurate thermodynamics were determined for 84 oligonucleotide dimers containing inosine. These data were combined with the reported thermodynamics of 13 oligonucleotide dimers from the literature (5,26). The sequences were designed to yield a unique determination of all possible pairs of inosine, I-C, I-A, I-T, I-G and I-I, in all possible nearest-neighbor contexts of G-C, C-G, A-T and T-A. The data reported here are the most comprehensive set of inosine nearest-neighbor thermodynamics reported to date.

MATERIALS AND METHODS

DNA synthesis and purification

Integrated DNA Technologies (Coralville, IA) synthesized the oligodeoxyribonucleotides on a 1 μ M or 50 nM scale, on solid support using standard phosphoramidite chemistry. Samples were dissolved in 250 μ l of ddH₂O and purified on a Si500F thin-layer chromatography (TLC) plate (Baker) by eluting for \sim 6 h with n-propanol/ammonia/ddH₂O (55:35:10 by volume) (27). Bands were visualized with an ultra-violet (UV) lamp where the least-mobile most-intense band was cut out and eluted three times with a total volume of 3 ml of ddH₂O. The silica gel from the TLC plate was then pelleted with a clinical centrifuge and the supernatant removed and evaporated to dryness. A Sep-pak C-18 cartridge (Waters) was used to further desalt and purify the DNA. Sample loading and washing was done with a 10 mM ammonium bicarbonate solution adjusted to pH 4.5. The elution buffer consisted of 30% acetonitrile, by volume, in ddH₂O. The samples were collected in two fractions of 5.0 ml and evaporated to dryness. Each oligomer was then dissolved in 1.0 ml of ddH₂O and dialyzed against a salt solution of 0.1 M NaCl and 0.1 mM EDTA overnight followed by ddH₂O overnight using a 1000 molecular weight cut-off membrane (Spectra-Por), which was previously washed with 1 mM EDTA, for a period of \sim 1 h. Optical density at 260 nm was then determined and the oligomers were then divided into 2.5 OD aliquots. All samples were then evaporated to dryness.

Measurement of melting curves

Absorbance versus temperature profiles (melting curves) were measured with an Aviv 14DS UV-VIS spectrophotometer with a five-cuvette thermoelectric controller as described previously (27–30). The buffer used in each experiment was 1.0 M NaCl, 10 mM sodium cacodylate and 0.5 mM Na₂EDTA (pH 7). Each oligonucleotide was dissolved in a volume to yield an absorbance reading just below 2.00 in a 0.1 cm pathlength cuvette. A typical volume was \sim 120 μ l. The total concentration of each single strand oligonucleotide was calculated from the absorbance reading taken at a wavelength of 260 nm at 85°C (27). It is important to measure the absorbance at high temperature since strands often form a self-structure at low temperatures resulting in a hypochromic absorbance (27,28). Raising the temperature to 85°C also degases the samples. Care was taken not to allow the total absorbance to rise above 2.00 and thus to be outside of the linear region of Beer's Law.

The extinction coefficients for the inosine containing oligomers were calculated using the nearest-neighbor method (28,31). Unfortunately, nearest-neighbor extinction coefficients for inosine have not been reported and thus an approximation was developed for this work. The inosine monomer UV extinction at 265 nm was reported, on average, to be 7500 M⁻¹ cm⁻¹ while that of thymine and cytidine, at 260 nm, is 8400 and 7050 M⁻¹ cm⁻¹, respectively (32–35). Thus it appears that the inosine monomer extinction is approximately the average of cytidine and thymine. We also assume that the hypochromicity of inosine in the middle of a DNA strand is similar to that of cytidine or thymine. Thus we estimated the extinctions of inosine containing oligos by averaging the

extinctions of equivalent sequences with inosine substituted with cytosine or thymine as shown in Equation 1.

$$\epsilon_{260}(\text{CTAIGCA}) = \frac{[\epsilon_{260}(\text{CTATGCA}) + \epsilon_{260}(\text{CTACGCA})]}{2} \quad 1$$

This method is expected to provide extinctions with an error of \sim 10–15%. The extinction coefficients for the A, C, G and T containing oligos were calculated using the extinction coefficient calculator provided in MELTWIN v. 3.0 melt curve program. A new web-server that automatically calculates the inosine contribution using Equation 1 is available at: <http://ozone.chem.wayne.edu> (K. Metani and J. SantaLucia, unpublished data).

The individual oligonucleotide concentrations were used to mix equal concentrations of the non-self-complementary strands to an initial volume of 160 μ l using Equations 2 and 3 (27).

$$\text{Volume}_{(\text{oligos1})} = \frac{160 \mu\text{l}[\text{oligos2}]}{[\text{oligos2}] + [\text{oligos1}]} \quad 2$$

$$\text{Volume}_{(\text{oligos2})} = 160 \mu\text{l} - \text{Volume}_{(\text{oligos1})} \quad 3$$

Two dilution series were used to allow for an 80- to 100-fold concentration range of the samples, over a total of 10 microcuvettes, or two dilution series. The samples were 'annealed' and degassed by raising the temperature to 85°C when the total strand concentration (C_T) in each microcuvette was calculated using the absorbance measured at a wavelength of 260 nm. The average of the two individual strand extinction coefficients was used in the calculation of C_T . The samples were then allowed to cool to -1.6°C and the melt was performed as described previously (28).

Determination of thermodynamic parameters

Thermodynamic parameters for duplex formation were obtained from melting curve data using the program MELTWIN v. 3.0 melt curve program (29). This program uses the Marquardt non-linear least squares method to solve for ΔH° , ΔS° and upper and lower baselines from absorbance versus temperature profiles (28,36).

$$T_m^{-1} = \frac{R \ln(C_T/4)}{\Delta H^\circ} + \frac{\Delta S^\circ}{\Delta H^\circ} \quad 4$$

ΔH° and ΔS° can also be found from the slope and intercept as the plot of T_m^{-1} versus $\ln C_T/4$ as given in Equation 4. Both of the above methods assume a two-state model and $\Delta C_p^\circ = 0$ for the transition equilibrium (36). In fact, there was no statistically significant temperature dependence of ΔH° observed in the experiments reported herein. We report free energies at 37°C to be consistent with the literature (27). In fact, however, the main application for these parameters is for PCR which involves temperatures above 50°C, which is close to the mean T_m of the experiments described here. Thus, the determined parameters are most accurate near 50°C. The two-state approximation was assumed to be valid for sequences in which the ΔH° values derived from the two methods agreed within 15% (31,37). The error weighted

average thermodynamic parameters were used for subsequent nearest-neighbor analysis (30).

Design of sequences

Oligonucleotide nonamers and dodecamers were designed to have melting temperatures between 30 and 65°C and to minimize formation of undesired hairpin or slipped-duplex conformations that might result in a non-two-state transition (27,31). End fraying and counter-ion effects were controlled by the use of terminal G/C pairs and 1 M NaCl buffer, respectively (38,39). The convention used to describe nearest-neighbors in this paper is with a slash, where the top sequence direction is 5' to 3' and the bottom sequence direction is 3' to 5', read from left to right. The eight nearest-neighbor sequences occur in this study with the following frequencies: AI/TN = 33, TI/AN = 39, AN/TI = 28, TN/AI = 35, CI/GN = 28, GI/CN = 26, CN/GI = 35 and GN/CI = 32. Here the variable base 'N' is A, T, G, C or I. In addition, 14 measurements of tandem inosine duplexes (i.e. I I/NN) and 8 tandem self-complementary pairs (i.e. IN/NI) were made.

Determination of the I-X mismatch contribution to duplex stability

The internal inosine mismatch contribution to duplex stability cannot be directly measured. As described previously, the total energy change for duplex formation can be approximated by a nearest-neighbor model that is the sum of energy increments for helix initiation, helix symmetry (for self-complementary sequences), and nearest-neighbor interactions between base pairs (40). For example

$$5' - \text{GCAT} \underline{\text{IATCG}} - 3' = \text{initiation} + \text{GC} + \text{CA} + \text{AT} + \underline{\text{TI}} + \underline{\text{IA}} + \text{AT} + \text{TC} + \text{CG} \\ 3' - \text{CGTACTAGC} - 5' \quad \text{CG GT TA } \underline{\text{AC}} \quad \underline{\text{CT}} \text{ TA AG GC} \quad 5$$

Thus, the inosine mismatch contribution to duplex stability can be derived by rearranging Equation 5.

$$\underline{\text{TI}} + \underline{\text{IA}} = \text{GCAT} \underline{\text{IATCG}} = \text{initiation} - \text{GC} - \text{CA} - \text{AT} - \text{AT} - \text{TC} - \text{CG} \\ \underline{\text{AC}} \quad \underline{\text{CT}} \quad \text{CGTACTAGC} \quad \text{CG GT TA TA AG GC} \quad 6$$

$$\underline{\text{TI}} + \underline{\text{IA}} = -8.52 - 2(0.98) - (-2.24) - (-1.45) - (-0.88) - (-0.88) \\ \underline{\text{AC}} \quad \underline{\text{CT}} \quad -(-1.30) - (-2.17) = -1.56 \text{ kcal/mol} \quad 7$$

The nearest-neighbors TI/AC and IA/CT in Equation 7 are the unknowns. Similarly, calculations for ΔH° and ΔS° mismatch contributions can be made. Alternatively, the inosine contribution may be computed as in Equation 8:

$$\Delta G_{37\text{trimer}}^\circ = \Delta G_{37\text{exp}}^\circ - \Delta G_{37\text{core}}^\circ + \Delta G_{37\text{NN}}^\circ \quad 8$$

These values are presented as trimers where $\Delta G_{37\text{NN}}^\circ$ is the central nearest-neighbor in the core sequence that is not present in the actual inosine duplex (31).

Regression analysis

The nearest-neighbor free energy parameters for I-C, I-A, I-T, I-G and I-I were solved with MATHEMATICA v 4.0 using equations similar to Equation 7, as described previously (40). Similar equations were used to calculate nearest-neighbor ΔH° values. Nearest-neighbor ΔS° values were calculated from

ΔG_{37}° and ΔH° . The ΔG_{37}° values for the tandem inosine oligonucleotides (I I/NN) and the tandem self-complementary pairs (I N/N I) were solved by subtracting all the known nearest-neighbor parameters. The core free energy in Equation 9 was calculated using the HYTHER server (<http://ozone.chem.wayne.edu>), so symmetry and initiation terms are already included. The inosine free energy term is for the inosine-Watson-Crick nearest-neighbors that are present on either side of the self-complementary dimer being solved for in a heteroduplex.

$$\Delta G_{37\text{tandem}}^\circ = \Delta G_{37\text{exp}}^\circ - \Delta G_{37\text{core}}^\circ - 2 * \Delta G_{37\text{inosine}}^\circ \\ - \Delta G_{37\text{symmetry}}^\circ - \Delta G_{37\text{initiation}}^\circ \quad 9$$

The constraint imposed on all sequences in this study is that all inosine mismatches are internal. For the non-symmetrical pairs, I-C, I-A, I-T and I-G, this results in one less parameter that can be derived from any set of linearly independent sequences (41). Thus, the maximum number of unique linearly independent nearest-neighbor parameters that can be derived for internal I-C pairs is 7 not 8. This was verified from the column rank of the stacking matrix, which is 7 for our data set. These 7 uniquely determined parameters are linear combinations of the 8 I-N dimers in each set of I-C, I-A, I-T and I-G. An example of a set of 7 linearly independent parameters for I-C nearest-neighbors is provided in the Supplementary Table S1. For more details on analysis of linearly independent sequences see the work of Allawi and SantaLucia (31,40). The nearest-neighbor contributions for I-I were uniquely determined.

Outlier determination

Outlier determination in this study was performed using Meltwin and Mathematica. If the Meltwin determination of ΔH° from the two fitting methods is not within 15%, then the melt was not considered to be two-state. If the melt, used as an equation in Mathematica, had a residual value over 1 kcal/mol, then the melt was considered to have competing equilibria and was not included in this study. Supplementary Table S2 lists every non-two-state determination, or data considered to be outliers in the Mathematica determination for each dimer. Since the unknowns are overdetermined removal of these outlier points has little effect on the determined parameters.

Error analysis

The sampling errors reported in Supplementary Table S2 for T_m^{-1} versus $\ln C_T/4$ plots and curve fits were obtained using standard methods in MELTWIN 3.0. These errors were then combined in a weighted average given by Equation 10:

$$\sigma_{\text{weighted average}} = \sqrt{\frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}} \quad 10$$

where σ_1 represents the precision of the data obtained in the curve fit analysis and σ_2 is the precision of the data obtained in the T_m^{-1} analysis (30). These results were then combined with the Watson-Crick dimer errors to compute the square-root of the sum of squares, which were the error values used in the

Table 1. Experimental (Expt) and predicted (Pred) thermodynamics of oligonucleotides with inosine mismatches in 1 M NaCl

DNA duplex	ΔG_{37}° (kcal/mol)		ΔH° (kcal/mol)		ΔS° (cal/K mol)		T_m ($^{\circ}\text{C}$)	
	Expt	Pred	Expt	Pred	Expt	Pred	Expt	Pred
I/C								
CGCIGAACIGGC	-14.33	-14.31	-94.4	-89.9	-258.2	-243.7	64.9	66.4
CCGICTGTIGCG	-14.07	-13.90	-91.3	-89.5	-249.0	-243.8	64.9	64.8
CGAITCCAITCC	-12.73	-12.94	-91.5	-94.9	-254.0	-264.3	59.5	59.5
CGAITCCTIACC	-11.82	-12.11	-88.1	-87.5	-245.9	-243.1	56.8	58.1
CAAACAAG ⁴	-6.08	-6.12	-58.0	-62.4	-167.4	-181.5	34.6	35.0
CAAATAAG ⁴	-6.51	-6.32	-66.0	-61.5	-191.8	-177.9	36.9	35.9
CGCIAATTGCGG ²⁴	-12.60	-12.77	-77.2	-84.2	-208.3	-230.3	63.5	63.7
GTGICTTCIGTC	-11.43	-11.20	-80.4	-82.7	-222.4	-230.5	57.1	55.6
GCAITATCG	-8.52	-8.01	-68.6	-62.1	-193.7	-174.4	46.3	44.6
CGGITCAGITGC	-13.91	-13.27	-97.2	-92.4	-268.5	-255.1	62.5	61.4
GAAICCTAICCG	-12.18	-12.11	-86.1	-85.3	-238.3	-236.0	58.8	58.7
GAAGIACGIAGG	-10.75	-11.25	-78.1	-81.9	-217.2	-227.8	54.7	56.0
CCAIGTGAICCG	-13.74	-13.40	-94.3	-91.1	-259.6	-250.5	62.7	62.3
CCTIGTGTICCG	-12.61	-12.37	-86.2	-84.6	-237.3	-232.9	60.5	60.0
CGTICCAITGGG	-11.82	-12.37	-81.2	-83.6	-223.7	-229.7	58.6	60.3
CCAICTGTICCG	-12.83	-13.07	-82.3	-84.4	-224.0	-230.0	62.7	63.1
CCAIGTGTIGCG	-13.07	-13.25	-88.7	-93.4	-243.8	-258.4	61.7	61.0
CGCITCACITGG	-13.97	-13.99	-92.2	-90.2	-252.2	-245.7	64.2	65.0
I/A								
CGCIGAACIGGC	-12.49	-12.17	-76.2	-71.3	-205.4	-190.6	63.3	63.6
CCGICAAGICCG	-15.26	-14.93	-94.3	-92.0	-254.8	-248.5	68.6	68.2
CCGICTGTIGCG	-14.50	-13.88	-95.1	-89.3	-259.9	-243.2	65.4	64.8
CGTACCTIACC	-8.86	-9.07	-67.8	-68.8	-190.0	-192.6	48.0	48.9
CGAITCCAITCC	-10.13	-10.02	-78.4	-77.5	-220.1	-217.6	51.9	51.6
CGAITCCTIACC	-9.48	-9.39	-77.1	-72.7	-218.0	-204.1	49.3	49.7
CAAATAAG ⁴	-4.61	-4.56	-48.0	-51.8	-139.9	-152.3	25.1	25.6
CAAATAAG ⁴	-5.28	-5.16	-63.0	-57.3	-186.1	-168.1	31.0	29.8
CGCIAATTGCGG ²⁴	-9.45	-9.43	-54.7	-61.0	-145.9	-166.3	54.5	54.9
GCAITATCG	-7.00	-6.75	-57.5	-56.0	-162.8	-158.8	39.6	38.2
GCACIGTCG	-9.22	-9.03	-61.8	-58.1	-169.5	-158.2	51.1	50.9
GCAGIGTCG	-9.73	-9.93	-64.3	-67.9	-175.9	-186.9	53.2	53.3
GCAAICTCG	-9.13	-9.32	-64.3	-68.5	-177.9	-190.8	50.1	50.2
GAAGIACGIAGG	-10.35	-10.59	-75.1	-76.7	-208.8	-213.2	53.6	54.3
CCAIGTGAICCG	-12.61	-12.65	-88.9	-91.2	-246.0	-253.3	59.8	59.3
CCTIGTGTICCG	-11.19	-11.42	-80.5	-80.9	-223.5	-224.0	56.1	57.0
CCAICTGTICCG	-12.47	-12.59	-86.2	-86.5	-237.7	-238.3	60.0	60.4
CCAIGTGTIGCG	-11.85	-12.03	-83.2	-87.7	-230.0	-244.0	58.2	57.8
CGCITCACITGG	-9.75	-10.05	-57.1	-58.0	-152.7	-154.6	55.5	57.0
I/T								
CGCIGAACIGGC	-10.17	-9.97	-69.9	-64.3	-192.6	-175.2	54.0	54.5
CCGICAAGICCG	-11.10	-11.25	-80.9	-76.8	-225.1	-211.3	55.6	57.3
CGTACCTIACC	-7.43	-7.69	-76.8	-74.0	-223.7	-213.8	40.7	41.9
CGAITCCAITCC	-7.55	-7.38	-66.9	-69.7	-191.4	-201.0	41.8	40.8
CGAITCCTIACC	-7.26	-7.38	-68.5	-71.4	-197.5	-206.4	40.3	40.7
CAAATAAG ⁴	-4.04	-4.19	-50.0	-59.8	-148.2	-179.3	22.3	25.3
CAAATAAG ⁴	-3.82	-3.52	-58.0	-48.0	-174.7	-143.4	23.2	18.7
CGCIAATTGCGG ²⁴	-8.66	-8.75	-58.9	-67.2	-162.0	-188.4	48.6	49.7
CAAITGCIACG	-8.11	-8.42	-69.0	-68.4	-196.3	-193.4	44.3	45.8
GTGICTTCIGTC	-7.96	-7.83	-55.4	-65.9	-153.0	-187.2	45.2	43.2
CAITGGCIGCG	-9.79	-10.19	-73.5	-69.0	-205.4	-189.7	51.4	54.4
GCAITATCG	-6.30	-6.06	-61.4	-58.6	-177.7	-169.4	35.8	34.5
GCACIGTCG	-7.58	-7.93	-47.9	-54.6	-130.0	-150.5	43.9	45.2
GCAGIGTCG	-7.88	-7.70	-64.1	-56.5	-181.3	-157.3	43.7	43.6
GAAGIACGIAGG	-7.53	-7.65	-63.8	-67.1	-181.4	-191.7	41.9	42.3
CCTIGTGTICCG	-8.93	-8.91	-74.9	-76.7	-212.7	-218.6	47.3	46.9
CCAICTGTICCG	-9.89	-9.52	-74.7	-74.2	-209.0	-208.6	51.6	50.0
CCAIGTGTIGCG	-9.89	-8.18	-82.5	-67.8	-234.1	-192.3	50.2	44.7
CGAICCAAIGGG	-8.61	-8.19	-71.2	-62.2	-201.8	-174.2	46.3	45.5
CGCITCACITGG	-10.41	-10.01	-84.8	-75.0	-239.9	-209.5	51.9	52.1
I/G								
CCGICAAGICCG	-11.72	-11.47	-78.4	-78.6	-215.0	-216.4	59.0	57.8
CCGICTGTIGCG	-10.44	-10.17	-87.8	-68.3	-249.4	-187.4	51.5	54.5
CGTACCTIACC	-6.17	-6.57	-60.8	-61.2	-176.1	-176.1	35.2	37.2
CGAITCCAITCC	-8.05	-7.90	-71.8	-69.1	-205.5	-197.3	43.7	43.3
CGAITCCTIACC	-6.94	-7.08	-65.0	-64.7	-187.2	-185.8	39.0	39.7
CAAAGAAG ⁴	-3.83	-3.36	-52.0	-49.5	-155.3	-148.8	21.7	18.3
CAAATAAG ⁴	-4.27	-4.05	-57.0	-51.6	-170.0	-153.3	25.2	22.8

Table 1. Continued

DNA duplex	ΔG_{37}° (kcal/mol)		ΔH° (kcal/mol)		ΔS° (cal/K mol)		T_m ($^{\circ}\text{C}$)	
	Expt	Pred	Expt	Pred	Expt	Pred	Expt	Pred
CGC <u>I</u> AATTGGCG ²⁴	-6.47	-6.97	-40.8	-47.4	-110.7	-130.4	36.5	42.8
GTG <u>I</u> CTTCIGTC	-7.77	-7.08	-52.9	-53.0	-145.5	-148.1	44.4	40.3
GCAT <u>I</u> ATCG	-5.58	-5.50	-54.9	-52.2	-159.0	-150.6	31.7	31.0
GCAC <u>I</u> GTCCG	-7.25	-7.07	-49.2	-40.8	-135.3	-108.8	41.6	41.2
CGG <u>I</u> TCCG <u>I</u> TCC	-8.30	-9.40	-58.4	-63.4	-161.5	-174.1	46.7	51.7
GAAG <u>I</u> ACG <u>I</u> AGG	-7.43	-7.15	-59.0	-58.7	-166.3	-166.2	41.8	40.3
CCA <u>I</u> GTGA <u>I</u> CCG	-9.44	-9.69	-69.9	-71.6	-194.9	-199.6	50.5	51.3
CCT <u>I</u> GTGT <u>I</u> CCG	-7.86	-8.18	-61.9	-65.1	-174.2	-183.5	43.8	45.1
CCA <u>I</u> CTGT <u>I</u> CCG	-9.95	-10.21	-80.0	-81.6	-225.9	-230.2	50.8	51.6
CCA <u>I</u> GTGT <u>I</u> GCG	-7.97	-8.21	-49.2	-57.2	-132.9	-158.0	46.3	46.4
I/I								
CGC <u>I</u> GAAC <u>I</u> GCG	-8.81	-8.83	-61.0	-57.3	-168.3	-156.3	49.0	50.0
CCG <u>I</u> CAAG <u>I</u> CCG	-10.39	-10.15	-66.9	-63.2	-182.2	-171.0	56.0	55.8
CGA <u>I</u> TCCA <u>I</u> TCC	-7.53	-7.64	-74.3	-72.7	-215.3	-209.8	41.2	41.8
CGA <u>I</u> TCCT <u>I</u> ACC	-6.37	-6.79	-65.1	-65.4	-189.4	-189.0	36.2	38.2
CAA <u>I</u> AAAAG ⁴	-4.04	-3.56	-47.0	-50.9	-138.5	-152.6	21.4	19.9
GTG <u>I</u> CTTCIGTC	-7.19	-6.71	-56.4	-55.6	-158.7	-157.6	40.7	38.0
CGG <u>I</u> CTTT <u>I</u> AGG	-6.32	-7.01	-51.2	-58.6	-144.7	-166.3	35.7	39.6
GCAT <u>I</u> ATCG	-5.51	-5.34	-56.9	-51.1	-165.7	-147.5	31.5	29.9
GCAC <u>I</u> GTCCG	-7.13	-7.36	-47.9	-51.1	-131.5	-141.0	40.9	42.1
GAAG <u>I</u> ACG <u>I</u> AGG	-6.95	-6.91	-59.5	-59.3	-169.4	-168.9	39.2	39.0
Non-two state behavior or outliers in nearest neighbor determination								
I/C								
CCG <u>I</u> CAAG <u>I</u> CCG	-10.69	-13.65	-60.4	-84.8	-160.4	-229.4	59.7	65.8
CCA <u>I</u> TGGC <u>I</u> GCG	-12.24	-15.14	-74.1	-94.5	-199.6	-255.2	62.7	68.9
GCAC <u>I</u> GTCCG	-11.84	-10.10	-79.1	-67.4	-217.0	-184.6	59.1	54.6
CGA <u>I</u> CCAA <u>I</u> GGG	-11.33	-13.35	-70.2	-90.1	-189.7	-247.1	59.9	62.8
I/A								
CGA <u>I</u> CCAA <u>I</u> GGG	-10.95	-12.60	-73.2	-90.2	-200.7	-249.9	56.9	59.7
CGT <u>I</u> CCAT <u>I</u> GGG	-10.02	-11.42	-69.1	-79.9	-190.4	-220.7	53.6	57.3
I/T								
CCG <u>I</u> CTGT <u>I</u> GCG	-8.30	-10.59	-49.6	-77.7	-133.2	-216.2	48.4	54.3
CGT <u>I</u> GCCT <u>I</u> GCG	-7.78	-9.55	-48.7	-75.0	-131.8	-210.9	45.5	50.2
CCA <u>I</u> GTGA <u>I</u> CCG	-9.11	-8.24	-92.3	-63.2	-267.7	-177.3	46.5	45.5
I/G								
CGT <u>I</u> GCCT <u>I</u> GCC	-6.64	-9.97	-30.8	-60.4	-77.8	-162.7	38.4	55.5
CGA <u>I</u> CCAA <u>I</u> GGG	-7.33	-9.64	-42.4	-58.8	-113.0	-163.3	43.1	45.8
CGT <u>I</u> CCAT <u>I</u> GGG	-6.38	-8.18	-37.8	-56.5	-101.6	-156.1	35.0	45.8
I/I								
CCG <u>I</u> CTGT <u>I</u> GCG	-13.60	-9.64	-86.8	-63.1	-236.0	-172.4	64.5	53.0

Underlined residues are paired with inosine. The opposing strands (not shown) are complementary except across from inosine.

MATHEMATICA calculation. The Watson–Crick dimer standard errors were reported previously (40). For literature oligonucleotides standard errors of ΔG_{37}° , ΔH° and ΔS° were assumed to be 4, 7 and 8%, respectively.

RESULTS AND DISCUSSION

Thermodynamic data

Table 1 contains the weighted average experimental thermodynamic values compared to the predicted thermodynamic values. The inosine nearest-neighbor parameters in Table 2 along with the previously determined parameters for Watson–Crick pairs were used to predict the thermodynamics of all of the duplexes listed in Table 1 with average deviations of 3.5, 4.8 and 5.0%, and 1.2 $^{\circ}\text{C}$ for ΔG_{37}° , ΔH° and ΔS° and T_m , respectively (40). This level of agreement between experimental and predicted oligos is within the limits of what can be expected for the nearest-neighbor model (41,42). This good agreement can also be seen in Figure 2 where the plot of all

two-state experimental versus predicted free energy changes has an R^2 of 0.99.

Stability trends

By averaging the nearest-neighbors in Table 2 that have the same I·X pair but different neighboring pairs, a general trend in decreasing stability is obtained. The stability trend is I·C > I·A > I·T > I·G > I·I, with numerical averages of -0.86, -0.46, +0.16, +0.29 and +0.37 kcal/mol of free energy, respectively. This result indicates that inosine does not pair indiscriminately and thus it is not really appropriate to classify it as a 'universal pairing base' *per se*, though its pairing energy is less dependent than other modified nucleotides. Importantly, the neighboring Watson–Crick pairs have a large influence on the stability of I·X pairs. The stability trend for the base pair 5' of the I·X pair is G·C > C·G > A·T > T·A, with averages of -0.60, -0.28, -0.19 and +0.19 kcal/mol, respectively. The stability trend for the base pair 3' of I·X is G·C >> C·G > A·T > T·A, with averages of -0.92, -0.18, +0.03 and +0.15 kcal/mol, respectively. These trends indicate a complex interplay

Table 2. Non-unique nearest-neighbor thermodynamics of inosine mismatches in 1 M NaCl

Dimer sequence	ΔG_{37}° (kcal/mol)	ΔH° (kcal/mol)	ΔS° (cal/K mol)
I-C pairs			
<u>A</u> I/ <u>T</u> C	-0.96 ± 0.06	-8.9 ± 1.4	-25.5 ± 1.4
<u>T</u> I/ <u>A</u> C	-0.46 ± 0.06	-5.9 ± 1.5	-17.4 ± 1.5
<u>A</u> C/ <u>T</u> I	-0.89 ± 0.06	-8.8 ± 1.4	-25.4 ± 1.4
<u>T</u> C/ <u>A</u> I	-0.59 ± 0.07	-4.9 ± 1.6	-13.9 ± 1.6
<u>C</u> I/ <u>G</u> C	-1.14 ± 0.08	-5.4 ± 1.6	-13.7 ± 1.6
<u>G</u> I/ <u>C</u> C	-0.86 ± 0.08	-6.8 ± 1.6	-19.1 ± 1.6
<u>C</u> C/ <u>G</u> I	-0.88 ± 0.09	-8.3 ± 1.7	-23.8 ± 1.7
<u>G</u> C/ <u>C</u> I	-1.07 ± 0.08	-5.0 ± 1.7	-12.6 ± 1.7
I-A pairs			
<u>A</u> I/ <u>T</u> A	-0.51 ± 0.06	-8.3 ± 1.5	-25.0 ± 1.5
<u>T</u> I/ <u>A</u> A	0.09 ± 0.06	-3.4 ± 1.4	-11.2 ± 1.4
<u>A</u> A/ <u>T</u> I	0.12 ± 0.07	-0.7 ± 1.7	-2.6 ± 1.7
<u>T</u> A/ <u>A</u> I	0.12 ± 0.06	-1.3 ± 1.4	-4.6 ± 1.4
<u>C</u> I/ <u>G</u> A	-0.18 ± 0.08	2.6 ± 1.6	8.9 ± 1.6
<u>G</u> I/ <u>C</u> A	-1.24 ± 0.08	-7.8 ± 1.7	-21.1 ± 1.7
<u>C</u> A/ <u>G</u> I	-0.77 ± 0.07	-7.0 ± 1.6	-20.0 ± 1.6
<u>G</u> A/ <u>C</u> I	-1.33 ± 0.08	-7.6 ± 1.6	-20.2 ± 1.6
I-T pairs			
<u>A</u> I/ <u>T</u> T	0.71 ± 0.06	0.49 ± 1.5	-0.7 ± 1.4
<u>T</u> I/ <u>A</u> T	0.36 ± 0.06	-6.5 ± 1.5	-22.0 ± 1.3
<u>A</u> T/ <u>T</u> I	0.22 ± 0.07	-5.6 ± 1.7	-18.7 ± 1.6
<u>T</u> T/ <u>A</u> I	0.54 ± 0.05	-0.8 ± 1.4	-4.3 ± 1.4
<u>C</u> I/ <u>G</u> T	-0.26 ± 0.07	-1.0 ± 1.5	-2.4 ± 1.5
<u>G</u> I/ <u>C</u> T	-0.19 ± 0.08	-3.5 ± 1.8	-10.6 ± 1.8
<u>C</u> T/ <u>G</u> I	0.41 ± 0.07	0.1 ± 1.5	-1.0 ± 1.5
<u>G</u> T/ <u>C</u> I	-0.54 ± 0.08	-4.3 ± 1.8	-12.1 ± 1.7
I-G pairs			
<u>A</u> I/ <u>T</u> G	0.02 ± 0.06	-4.9 ± 1.6	-15.8 ± 1.6
<u>T</u> I/ <u>A</u> G	0.76 ± 0.06	-1.9 ± 1.5	-8.5 ± 1.5
<u>A</u> G/ <u>T</u> I	0.65 ± 0.07	0.1 ± 1.7	-1.8 ± 1.7
<u>T</u> G/ <u>A</u> I	0.70 ± 0.06	1.0 ± 1.4	1.0 ± 1.4
<u>C</u> I/ <u>G</u> G	0.47 ± 0.10	7.1 ± 2.3	21.3 ± 2.3
<u>G</u> I/ <u>C</u> G	-0.10 ± 0.07	-1.1 ± 1.5	-3.2 ± 1.5
<u>C</u> G/ <u>G</u> I	0.54 ± 0.09	5.8 ± 2.0	16.9 ± 2.0
<u>G</u> G/ <u>C</u> I	-0.74 ± 0.08	-7.6 ± 1.8	-22.0 ± 1.8
I-I pairs			
<u>A</u> I/ <u>T</u> I	0.40 ± 0.05	-3.3 ± 1.2	-11.9 ± 3.9
<u>T</u> I/ <u>A</u> I	0.81 ± 0.05	0.1 ± 1.4	-2.3 ± 4.3
<u>C</u> I/ <u>G</u> I	0.36 ± 0.06	1.3 ± 1.4	3.0 ± 4.2
<u>G</u> I/ <u>C</u> I	-0.09 ± 0.06	-0.5 ± 1.2	-1.3 ± 3.7

The I-C, I-A, I-T and I-G parameters are a linear least-squares fit of the data for a singular matrix with a rank of 7. The parameters for I-I are not rank deficient. These parameters were used to make the predictions found in Table 1. These parameters are for internal Inosine pairs only (note that these parameters do not apply to terminal helix inosine pairs). Errors shown are standard deviations computed by error propagation. Underlined residues are paired.

between H-bonding, nearest-neighbor stacking and mismatch geometry on the observed stability of an I-N pair.

Context dependence of I-C mismatch thermodynamics

Trimer stabilities were analyzed for I-C mismatches. The data in Table 2 can be used to predict the thermodynamics of I-C mismatches in all 10 different trimer contexts closed by Watson-Crick pairs. The most stable context is CIC/GCG, which contributes -2.21 kcal/mol to duplex free energy at 37°C . The least stable context is AIA/TCT, which contributes $+0.37$ kcal/mol. The general trend for the nucleotide on the 5' side of inosine of an I-C pair in order of decreasing stability is: C·G > A·T > G·C > T·A, with averages of -1.14 , -0.96 , -0.86 and -0.46 kcal/mol, respectively. On the 3' side of inosine, the stability order is: G·C > A·T \approx C·G > T·A with averages of

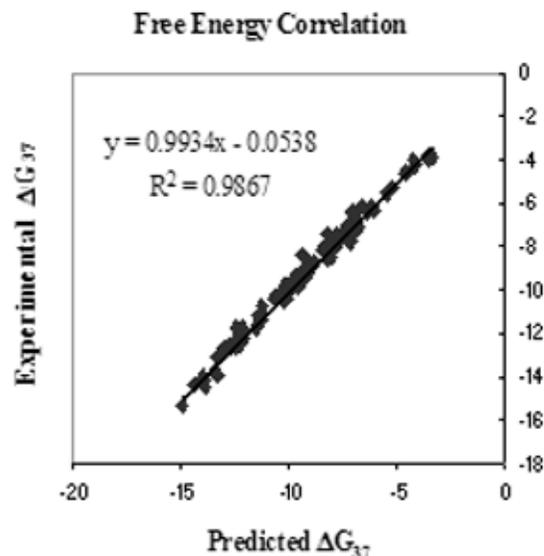


Figure 2. Comparison of experimental versus predicted free energies of all two-state duplexes from Table 1.

-1.07 , -0.89 , -0.88 and -0.59 kcal/mol, respectively. This illustrates that there is a nearest-neighbor context dependence for I-C pairs.

A comparison of I-X pair and G-X pair thermodynamics

A comparison between I-C, I-A, I-T and I-G was done versus the corresponding guanosine pairs are shown in Supplementary Figure S6 (31,40,43,44). Overall, the best correlation observed is for I-C vs G-C with an R^2 of 0.78. Guanosine pairs appear to have a larger context dependence than do inosine pairs where the general trend is G·C \gg G·G \approx G·T \approx G·A (44). This stability trend shows an average free energy of -1.70 , -0.23 , $+0.03$ and $+0.09$ kcal/mol, respectively. This corresponds to a 1.79 kcal/mol stability range versus 1.15 kcal/mol for inosine pairs in the same context. On average, the free energy of G·C was 0.84 kcal/mol more stable than I-C, which is attributable to the extra H-bond in G-C pairs as well as the presumed differences in inosine H-bonding, hydration and stacking as discussed previously for RNA (45–47). The result for I-A is surprising where G·A is 0.55 kcal/mol less stable than I-A. Other comparisons were: G·T is 0.13 kcal/mol more stable than I-T and G·G is 0.52 kcal/mol more stable than I-G. One interpretation of these differences is that guanosine stacking plays a more significant thermodynamic role in dimer mismatches than inosine stacking does. The conclusion drawn here is that it is not effective to approximate inosine as guanosine. Similar results are observed if inosine were approximated as an adenosine (data not shown).

Tandem inosine thermodynamics

Tandem inosine thermodynamic contributions are shown in Table 3. These were calculated from the raw data in Supplementary Table S3 by subtracting the Watson-Crick and inosine nearest-neighbors from Table 2 (40). Of the sixteen duplexes containing all possible tandem inosine pairs, only twelve were two-state. The non-two-state tandem pairs were I

Table 3. The thermodynamics contribution of tandem internal Inosine-Watson-Crick mismatch pairs

Propagation sequence	ΔG_{37}° (kcal/mol)	ΔH° (kcal/mol)	ΔS° (cal/K mol)
GGTA <u>II</u> GTGTCG			
II/CC	-0.64 ± 0.93	-9.3 ± 5.8	-28.0 ± 17.1
II/AC	0.27 ± 0.25	-3.1 ± 5.8	-11.0 ± 16.8
GGTT <u>II</u> GAGTCG			
II/CA	0.44 ± 0.27	-8.7 ± 6.2	-29.3 ± 18.2
II/AA	-0.27 ± 0.27	-2.1 ± 5.7	-6.1 ± 18.1
II/TA	0.83 ± 0.26	2.3 ± 6.1	4.4 ± 18.0
II/GA	0.30 ± 0.25	4.2 ± 5.8	-14.8 ± 17.1
GGTGC <u>II</u> GGTTCG			
II/CT	0.33 ± 0.30	-14.5 ± 6.1	-47.8 ± 17.5
II/AT	0.19 ± 0.30	-17.8 ± 6.1	-58.1 ± 17.6
II/TT	1.69 ± 0.30	-7.0 ± 6.1	-28.1 ± 17.5
II/GT	0.13 ± 0.31	-19.4 ± 6.3	-62.8 ± 17.7
GGTT <u>II</u> AGTTCG			
II/TG	0.03 ± 0.22	13.3 ± 6.8	40.6 ± 21.3
II/TC	-1.30 ± 0.18	0.3 ± 6.3	2.1 ± 19.6
GGTA <u>II</u> GTGTCG			
II/II	(-0.7 ± 0.25)	(-13.8 ± 5.6)	(-42.1 ± 17.4)
GGTT <u>II</u> AGTTCG			
II/II	(1.74 ± 0.23)	(-7.5 ± 6.2)	(-29.0 ± 19.5)

Values in parentheses are non two-state. Values reported were calculated by subtracting Watson-Crick nearest-neighbors and inosine nearest-neighbors from the raw experimental thermodynamics seen in Supplementary Table S3. Errors shown are standard deviations determined by error propagation. Underlined residues are paired with inosine. The opposing strands (not shown) are complementary except across from inosine.

I/AG, I I/CG, I I/GC and I I/TC. The most stable context was AIIC/TGGG at -3.33 kcal/mol and the least stable was TIIA/ATTT at $+0.64$ kcal/mol, which is a larger range than observed for single inosine pairs. Addition of an A·T pair to the 5' end of the inosine dimer may relieve some of the strain on the backbone. This may be either due to the increased flexibility of an A·T pair compared to a G·C pair, or due to the steric hindrance of the 2-amino group of the 3' G·C pair (46). Alternatively, stacking of G·C on I·X pairs may be inherently weaker than A·T stacking, as suggested in some RNA contexts (47). In any case, it is evident that tandem inosine pairs have complex behavior which is not well understood currently.

Tandem internal self-complementary inosine mismatch pairs

To complete the set of inosine thermodynamic parameters eight duplex melts were performed to solve for the self-complementary inosine mismatch pairs IC/CI, CI/IC and so on. The individual thermodynamic contributions for these dimers were found by Equation 9. Supplementary Tables S4 and S5 contain the raw thermodynamic parameters and the weighted averages for each data set, respectively. Table 4 contains the contributions for each individual mismatch pair. The general stability trend for dimers where inosine 5' is: IA/AI \gg IC/CI \gg IG/GI \gg IT/TI. Where inosine occurs 3' in the dimer the stabilities are: AI/IA \gg CI/IC $>$ TI/IT \gg GI/IG. These parameters were compared to those reported in the literature by Kawase *et al.* (48). To make a direct comparison of the contribution of the data it is necessary to calculate ΔG_{37}° at 37°C and 1 M NaCl using Equation 11. Also, there was a symmetry correction of 0.4 kcal/mol and an initiation correction of 1.96 kcal/mol for the free energy

Table 4. Thermodynamic contributions of other tandem internal Inosine mismatch pairs

Dimer sequence	ΔG_{37}° (kcal/mol)	ΔH° (kcal/mol)	ΔS° (cal/K mol)
IC/CI	-0.85 ± 0.23	-12.1 ± 4.2	-36.3 ± 6.6
CI/IC	0.06 ± 0.23	-1.8 ± 4.4	-6.4 ± 7.1
IA/AI	-1.43 ± 0.21	-13.9 ± 4.4	-40.5 ± 8.0
AI/IA	-0.56 ± 0.22	-9.5 ± 4.2	-29.1 ± 6.8
IT/TI	2.03 ± 0.22	-7.6 ± 4.3	-31.1 ± 7.5
TI/IT	0.61 ± 0.28	-14.7 ± 5.1	-49.4 ± 11.4
IG/GI	1.18 ± 0.19	3.2 ± 4.4	6.1 ± 8.3
GI/IG	1.12 ± 0.19	-4.2 ± 4.1	-39.8 ± 7.1

Values reported were calculated by subtracting Watson-Crick and Inosine nearest-neighbors from the raw experimental thermodynamics given in Supplementary Table S4.

Table 5. Hydrogen-bond free energy increments at 37°C for internal I-C pairs

I-C pairs	$\Delta \Delta G_{HB}^{\circ}$
5'-CGCIGAACIGGC-3'	-1.99
5'-CCGICTGTIGCG-3'	-2.05
5'-CGAITCCAITCC-3'	-0.87
5'-CGAITCCTIACC-3'	-1.29
5'-CAAACAAAAG-3'	-1.54
5'-CAAIAAAAAG-3'	-1.03
5'-CGCIAATTCGCG-3'	-1.74
5'-GTGICTTCIGTC-3'	-2.07
5'-GCATIATCCG-3'	-1.70
5'-CGGTCAGITGC-3'	-1.53
5'-GAAICCTAICCG-3'	-1.49
5'-GAAGIACGIAGG-3'	-1.69
5'-CCAIGTGAICCG-3'	-1.28
5'-CCTIGTGTICCG-3'	-2.06
5'-CGTICCATIGGG-3'	-2.06
5'-CCAICTGTICCG-3'	-1.83
5'-CCAIGTGTIGCG-3'	-1.62
5'-CGCITCACITGG-3'	-1.58

The complementary strand is not shown.

calculations as reported previously (40). The salt dependence of nucleic acid pairs have been shown to be independent of sequence (40). Thus, we are making the assumption that the same salt dependence applies to inosine pairs as for the normal Watson-Crick pairs in DNA (40,42).

$$\Delta G_{37}^{\circ}[0.1M NaCl] = \Delta G_{37}^{\circ}[1.0M NaCl] - [0.114 * 11 * \ln(0.11)]$$

11

The free energy contributions of the dimers IC/CI, IA/AI and IG/GI, as calculated using raw data from Kawase *et al.* (48) are -2.4 , -2.0 and -0.7 kcal/mol, respectively. These values do not compare well with those reported in Table 4. A possible reason for this disparity may be competing equilibria between hairpins and the desired heteroduplex. The sequence used in the study by Kawase was GGGAAXYTTCCC where X was deoxyinosine and Y was either C, A or G, which has the potential to form competing hairpin and homo-duplexes.

Hydrogen bond values in inosine containing sequences

Previous studies have investigated the H-bond contribution of the 2-amino group of guanosine by substituting inosine (45,47). The values for $\Delta \Delta G_{HB}^{\circ}$ were calculated as described previously (47) with Equation 12:

$$\Delta \Delta G_{HB}^{\circ} = [\Delta G^{\circ}(CGC/GCG) - \Delta G^{\circ}(CIC/GCG)]$$

12

Table 6. The effect of inosine position on hybridization thermodynamics for 5' ATCATCGTACG 3' paired with sequences listed

Complementary duplex	Mismatch Type	ΔG_{37}° (kcal/mol)	T_m (°C) (1×10^{-4} M C_T)
AGTAGCATG	Control	-9.06	48.6
AGTAGCITG	TI	-7.22	40.3
AGTAICATG	CI	-7.32	41.0
AGIA \overline{G} CATG	AI	-8.01	43.7
AGTAGIATG	GI	-4.89	27.2

The calculations were done using HYTHER Module 3, which uses the nearest neighbor parameters from Table 2 and those from reference 40. Inosines are in boldface and underlined for emphasis.

The results are tabulated in Table 5 for I-C pairs. By Equation 13, the 2-amino group of a G-C pair contributes from -0.8 to -2.0 kcal/mol. These H-bond values in DNA are consistent with those reported previously in RNA by Turner *et al.* (47) who observed a $\Delta\Delta G_{HB}^{\circ}$ range of -0.5 to -1.9 kcal/mol (47). It should be noted, however, that the hydrogen bond energies calculated here may contain a term for cooperativity due to the bimolecularity of the system used in this study as pointed out by Moody and Bevilacqua (49).

Implications for DNA probe design

To allow accurate predictions, the inosine nearest-neighbor parameters reported here have been added to the program HYTHER at <http://www.ozone2.chem.wayne.edu/HYTHER/hythermenu.html>. Thus, it is now possible to design degenerate probes containing inosine with optimized thermodynamic parameters using HYTHER. Replacement of a guanine or adenine with inosine can, depending upon the nearest-neighbors, affect the hybridization stability as needed to potentially improve hybridization specificity as seen in Table 6. For instance, it is now possible to design an inosine probe for a guanine rich target to have a net favorable ΔG° where our I/G dimers reveal <1.5 kcal/mol of instability as seen in Table 2. As Table 3 indicates, tandem inosine effects may be more or less favorable, again, depending upon the nearest-neighbors, in the range of ± 2 kcal/mol of ΔG° . An investigator can now use HYTHER to try all possible inosine locations and choose the one with the best binding energy as shown in Table 6. Knowledge of inosine's nearest-neighbor parameters will allow one to design a probe of optimal stability. Use of these parameters would allow the design of the most stable probes possible in the construction of an oligos library. This information allows more control in the use of inosine in such applications as degenerate PCR and the sequencing of ambiguous sequences. Martin *et al.* (5) gives an excellent discussion of selected Watson-Crick ambiguities when paired with inosine containing probes (5).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr Rostem Irani and Dr Nic Peyret for their assistance in this project. We also thank Kamal Mettani for his computer

support and Raviprasad Aduri and Larry Clos for useful discussions. We also thank Integrated DNA Technologies for providing inosine oligonucleotides at a reduced price. This work was supported by NIH grant HG02020 and by a Michigan Life Sciences Corridor grant LSC1653. Funding to pay the Open Access publication charges for this article was provided by Wayne State University.

Conflict of interest statement. None declared.

REFERENCES

- Crick, F.H.C. (1966) Codon-anticodon pairing: the wobble hypothesis. *J. Mol. Biol.*, **19**, 548-555.
- Zimmerman, F.K. (1977) Genetic effects of nitrous acid. *Mutat. Res.*, **39**, 127-148.
- Karran, P. and Lindahl, T. (1980) Hypoxanthine in deoxyribonucleic acid: generation by heat-induced hydrolysis of adenine residues and release in free form by a deoxyribonucleic acid glycosylase from calf thymus. *Biochemistry*, **19**, 6008-6011.
- Clutterbuck, D.R., Leroy, A., O'Connell, M.A. and Semple, C.A.M. (2005) A bioinformatic screen for novel A-I RNA editing sites reveals recoding editing in BC10. *Bioinformatics*, **21**, 2590-2595.
- 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.
- Case-Green, S.C. and Southern, E.M. (1994) Studies on the base pairing properties of deoxyinosine by solid phase hybridisation to oligonucleotides. *Nucleic Acids Res.*, **22**, 131-136.
- Uesugi, S., Oda, Y., Ikehara, M., Kawase, Y. and Ohtsuka, E. (1987) Identification of I:A mismatch base-pairing structure in DNA. *J. Biol. Chem.*, **262**, 6965-6968.
- Corfield, P.W.R., Hunter, W.N., Brown, T., Robinson, P. and Kennard, O. (1987) Inosine adenine base pairs in a B-DNA duplex. *Nucleic Acids Res.*, **15**, 7935-7949.
- Nichols, R., Andrews, P.C., Zhang, P. and Bergstrom, D.E. (1994) A universal nucleoside for use at ambiguous sites in DNA primers. *Nature*, **369**, 492-493.
- Loakes, D., Brown, D.M., Linde, S. and Hill, F. (1995) 3-Nitropyrrole and 5-nitroindole as universal bases in primers for DNA sequencing and PCR. *Nucleic Acids Res.*, **23**, 2361-2366.
- Takahashi, Y., Kato, K., Hayashizaki, Y., Wakabayashi, T., Ohtsuka, E., Matsuki, S., Ikehara, M. and Matsubara, K. (1985) Molecular cloning of the human cholecystokinin gene by use of a synthetic probe containing deoxyinosine. *Proc. Natl Acad. Sci. USA*, **82**, 1931-1935.
- Ohtsuka, E., Matsuki, S., Ikehara, M., Takahashi, Y. and Matsubara, K. (1985) An alternative approach to deoxyoligonucleotides as hybridization probes by insertion of deoxyinosine at ambiguous codon positions. *J. Biol. Chem.*, **260**, 2605-2608.
- Miura, N., Ohtsuka, E., Yamaberi, N., Ikehara, M., Uchida, T. and Okada, Y. (1985) Use of the deoxyinosine-containing probe to isolate and sequence cDNA encoding the fusion (F) glycoprotein of Sendai virus (HVJ). *Gene*, **38**, 271-274.
- Patil, R. and Dekkar, E. (1990) PCR amplification of an *Escherichia Coli* gene using mixed primers containing deoxyinosine at ambiguous positions in degenerate amino acid codons. *Nucleic Acids Res.*, **18**, 3080.
- Palva, A., Vidgren, G. and Paulin, L. (1994) Application of PCR with oligonucleotide primers containing deoxyinosine for gene detection, isolation and sequencing. *J. Microbiol. Methods*, **19**, 315-321.
- Suggs, S.V., Wallace, R.B., Hirose, T., Kawashima, E.H. and Itakura, K. (1981) Use of synthetic oligonucleotides as hybridization probes: isolation of cloned cDNA sequences for human β_2 -microglobulin. *Proc. Natl Acad. Sci. USA*, **78**, 6615-6617.
- Wallace, R.B., Johnson, M.J., Hirose, T., Miyake, T., Kawashima, E.H. and Itakura, K. (1981) The use of synthetic oligonucleotides as hybridization probes. II. Hybridization of oligonucleotides of mixed sequence to rabbit beta-globin DNA. *Nucleic Acids Res.*, **9**, 879-894.
- Steger, G. (1994) Thermal denaturation of double-stranded nucleic acids: prediction of temperatures critical for gradient gel electrophoresis and polymerase chain reaction. *Nucleic Acids Res.*, **22**, 2760-2768.

19. Cheng, A. and Van Dyke, M.W. (1997) Oligodeoxyribonucleotide length and sequence effects on intramolecular and intermolecular G-quartet formation. *Gene*, **197**, 253–260.
20. Wang, J., Rivas, G., Fernandes, J.R., Paz, J.L.L., Jiang, M. and Waymire, R. (1998) Indicator-free electrochemical DNA hybridization biosensor. *Anal. Chim. Acta*, **375**, 197–203.
21. Chizhikov, V., Wagner, M., Ivshina, A., Hoshino, Y., Kapikian, A.Z. and Chumakov, K. (2002) Detection and genotyping of human group A rotaviruses by oligonucleotide microarray hybridization. *J. Clin. Microbiol.*, **40**, 2398–2407.
22. Burns, M.A., Johnson, B.N., Brahmasandra, S.N., Handique, K., Webster, J.R., Krishnan, M., Sammarco, T.S., Man, P.M., Jones, D., Heldsinger, D., Mastrangelo, C.H. and Burke, D.T. (1998) An integrated nanoliter DNA analysis device. *Science*, **282**, 484.
23. Loakes, D. (2001) The applications of Universal DNA base analogues. *Nucleic Acids Res.*, **29**, 2437–2447.
24. Loakes, D. and Brown, D.M. (1994) 5-Nitroindole as an universal base analogue. *Nucleic Acids Res.*, **22**, 4039–4043.
25. Bergstrom, D.E., Zhang, P. and Johnson, W.T. (1997) Comparison of the base pairing properties of a series of nitroazole nucleobase analogs in the oligodeoxyribonucleotide sequence 5'-d(CGCXAATTYGCG)-3'. *Nucleic Acids Res.*, **25**, 1935–1942.
26. Johnson, W.T., Zhang, P. and Bergstrom, D.E. (1997) The synthesis and stability of oligodeoxyribonucleotides containing the deoxyadenosine mimic 1-(2'-deoxy- β -D-ribofuranosyl) imidazole-4-carboxamide. *Nucleic Acids Res.*, **25**, 559–567.
27. SantaLucia, J., Jr (2000) The use of spectroscopic techniques in the study of DNA stability. In Gore, M.G. (ed.), *Spectrophotometry and Spectrofluorimetry*. University Press, Oxford, NY, pp. 329–354.
28. SantaLucia, J., Jr, Allawi, H.T. and Seneviratne, P.A. (1996) Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry*, **35**, 3555–3562.
29. McDowell, J.A. and Turner, D.H. (1996) Investigation of the structural basis for thermodynamic stabilities of tandem GU mismatches: solution structure of (rGAGGUCUC)₂ by two-dimensional NMR and simulated annealing. *Biochemistry*, **35**, 14077–14089.
30. Bommarito, S., Peyret, N. and SantaLucia, J., Jr (2000) Thermodynamic parameters for DNA sequences with dangling ends. *Nucleic Acids Res.*, **28**, 1929–1934.
31. Allawi, H.T. and SantaLucia, J., Jr (1997) Thermodynamics and NMR of internal G-T mismatches in DNA. *Biochemistry*, **36**, 10581–10594.
32. Dedman, J.R. and Means, A.R. (1977) Characterization of a spectrophotometric assay for cAMP phosphodiesterase. *J. Cyclic Nucleotide Res.*, **3**, 139–152.
33. Mura, U., Sgarella, F. and Ipaten, P. (1977) Spectrometric determination of ribose-1-phosphate. *Anal. Biochem.*, **82**, 210–216.
34. Kalckar, H.M. (1947) Differential spectrophotometry of purine compounds by means of specific enzymes. *J. Biol. Chem.*, **167**, 445.
35. Oliver, J.D., Parker, K.A. and Suggs, J.W. (2001) Effect of the universal base 3-nitropyrrole on the selectivity of neighboring natural bases. *Org. Lett.*, **3**, 1977–1980.
36. Petersheim, M. and Turner, D.H. (1983) Base-stacking and base-pairing contributions to helix stability: thermodynamics of double-helix formation with CCGG, CCGGp, CCGGAp, ACCGGp, CCGGUp, and ACCGGUp. *Biochemistry*, **22**, 256–263.
37. Freier, S.M., Diersek, R., Jaeger, J.A., Sugimoto, N., Caruthers, M.H., Neilson, T. and Turner, D.H. (1986) Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl Acad. Sci. USA*, **83**, 9373–9377.
38. SantaLucia, J., Jr, Kierzek, R. and Turner, D.H. (1990) Effects of GA mismatches on the structure and thermodynamics of RNA internal loops. *Biochemistry*, **29**, 8813–8819.
39. Olmsted, M.C., Anderson, C.F. and Record, M.T., Jr (1989) Monte Carlo description of oligoelectrolyte properties of DNA oligomers: range of the end effect and the approach of molecular and thermodynamic properties to the polyelectrolyte limits. *Proc. Natl Acad. Sci. USA*, **86**, 7766–7770.
40. SantaLucia, J., Jr (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl Acad. Sci. USA*, **95**, 1460–1465.
41. Gray, D.M. and Tinoco, I., Jr (1970) A new approach to the study of sequence-dependent properties of polynucleotides. *Biopolymers*, **9**, 223–244.
42. SantaLucia, J., Jr and Hicks, D. (2004) The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.*, **33**, 415–440.
43. Allawi, H.T. and SantaLucia, J., Jr (1998) Nearest-neighbor thermodynamic parameters for internal G-A mismatches in DNA. *Biochemistry*, **37**, 2170–2179.
44. Peyret, N., Seneviratne, P.A., Allawi, H.T. and SantaLucia, J., Jr (1999) Nearest-neighbor thermodynamics and NMR of data sequences of internal A-A, C-C, G-G, and T-T Mismatches. *Biochemistry*, **38**, 3468–3477.
45. SantaLucia, J., Jr, Kierzek, R. and Turner, D.H. (1991) Functional group substitutions as probes of hydrogen bonding between GA mismatches in rna internal loops. *J. Am. Chem. Soc.*, **113**, 4313–4322.
46. Gaffney, B.L., Marky, L.A. and Jones, R.A. (1984) The influence of the Purine 2-Amino Group on DNA Conformation and Stability. *Tetrahedron*, **40**, 3–13.
47. Turner, D.H., Sugimoto, N., Kierzek, R. and Dreiker, S.D. (1987) Free energy increments for hydrogen bonds in nucleic acid base pairs. *J. Am. Chem. Soc.*, **109**, 3783–3785.
48. Kawase, Y., Iwai, S., Inoue, H., Miura, K. and Ohtsuka, E. (1986) Studies on nucleic interactions I. Stabilities of mini-duplexes (dG₂A₄XA₄G₂ (dC₂T₄YT₄C₂) and self-complementary d(GGGAAXYTTCCC) containing deoxyinosine and other mismatched bases. *Nucleic Acids Res.*, **14**, 7727–7736.
49. Moody, E.M. and Bevilacqua, P.C. (2003) Folding of a stable DNA motif involves a highly cooperative network of interactions. *J. Am. Chem. Soc.*, **125**, 16285–16293.