

# Modeling Peptide Nucleic Acid Binding Enthalpies Using MM-GBSA

Jack Goodman,\* David Attwood, Janice Kiely, Pablo Coladas Mato, and Richard Luxton



Cite This: *J. Phys. Chem. B* 2022, 126, 9528–9538



Read Online

ACCESS |



Metrics & More

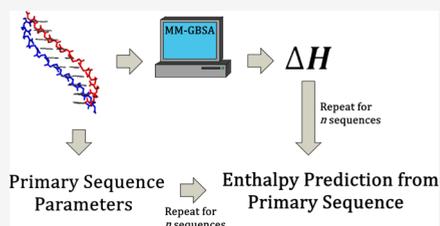


Article Recommendations



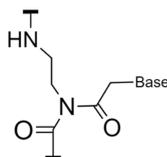
Supporting Information

**ABSTRACT:** The binding enthalpies of peptide nucleic acid (PNA) homoduplexes were predicted using a molecular mechanics generalized Born surface area approach. Using the nucleic acid nearest-neighbor model, these were decomposed into sequence parameters which could replicate the enthalpies from thermal melting experiments with a mean error of 8.7%. These results present the first systematic computational investigation into the relationship between sequence and binding energy for PNA homoduplexes and identified a stabilizing helix initiation enthalpy not observed for nucleic acids with phosphoribose backbones.



## INTRODUCTION

The binding of nucleic acids is a sequence-dependent process facilitated primarily through the Watson–Crick pairing of complementary base pairs. This is not a property unique to naturally occurring nucleic acids such as DNA or RNA. It is therefore possible to produce nucleic acid analogue molecules which adhere in a sequence-specific manner but with modified properties. Peptide nucleic acid (PNA) is an analogue wherein the phosphoribose backbone of DNA or RNA has been entirely replaced. Instead, PNA has a backbone composed of repeating *N*-(2-aminoethyl)glycine monomers (Figure 1).<sup>1</sup>

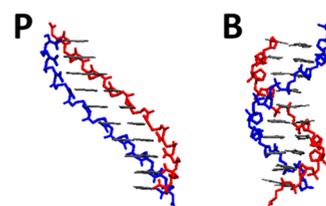


**Figure 1.** Schematic of an *N*-(2-aminoethyl)glycine monomer of a PNA backbone connected to a base through a carbonyl methylene linker.

The directionality of the backbone proceeds from an N-terminal amine to a C-terminal carboxyl, similar to a polypeptide, and PNA forms both stable homo- and heteroduplex structures in solution.<sup>2,3</sup>

PNA homoduplexes form P-form<sup>1,3</sup> double helices in solution. The P-form helix has a large 18 base-pair pitch and a wide 28 Å helical diameter (Figure 2). PNA/DNA heteroduplexes have a structure intermediate of the P- and B-forms, demonstrating that PNA is flexible enough to remain stable in its non-preferred helical conformation.<sup>4</sup>

Despite adopting non-preferred helical structures, PNA heteroduplexes are more thermally stable than DNA or RNA homoduplexes, though they are less stable than PNA homoduplexes.<sup>3</sup> Unlike DNA or RNA homoduplexes, PNA heteroduplex stability is largely independent of the concen-



**Figure 2.** 10 base-pair structures of P-form PNA (left) and B-form DNA (right).

tration of salt in solution.<sup>2,5</sup> This is due to the PNA neutral backbone since the net repulsion between the negatively charged phosphoribose backbones in DNA or RNA is not present.<sup>6</sup> As a result, the PNA backbone does not need to associate with counterions to stabilize the double helix.<sup>2</sup>

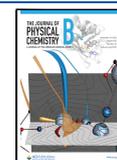
PNA heteroduplexes are also biostable since they are not substrates for nucleases and proteases.<sup>5</sup> The acidic conditions that would depurinate DNA or RNA homoduplexes also typically do not depurinate PNA.<sup>7</sup> Since PNA remains stable under a variety of conditions, it has attracted interest for applications ranging from antisense and antigene technologies to biosensing, genetic diagnostics, molecular biology, and medicinal chemistry. For example, PNA has been applied in vivo to silence gene transcription and translation.<sup>8–10</sup>

The PNA backbone can be modified to produce favorable properties in heteroduplexes. For example, the incorporation of lysine or arginine functional groups at the  $\alpha$  carbon of the backbone increased the electrostatic forces of attraction between a PNA strand and a DNA complement.<sup>11</sup> Similar

**Received:** August 4, 2022

**Revised:** October 14, 2022

**Published:** November 14, 2022



modifications could be applied to PNA homoduplexes, with the advantage that identical chemistry can be used for both strands. Given the higher thermal and environmental stabilities of PNA relative to other nucleic acids with biotechnological applications, it is feasible that PNA technologies based on such modifications could have desirable properties.

PNA homoduplex technologies, however, must be preceded by a deeper understanding of the relation between the PNA structure and binding properties. Currently, no model exists that decomposes the binding energies of PNA duplexes into sequence parameters. All-atom molecular dynamics is an appropriate means to develop such a model since any number of PNA structures can be easily generated, and the expense of specialist equipment such as synthesizers can be avoided.

To this end, molecular mechanics generalized Born surface area (MM-GBSA) analysis of simulated PNA structures was conducted. Entropy was estimated using a quasi-harmonic (QH) approximation. The current force field parameters designed by Jasiński et al.<sup>12</sup> were implemented. The computational results were benchmarked against thermal melting data obtained both from this work and the available literature.

Using a previously developed method,<sup>13</sup> it was determined that nearest-neighbor decomposition of PNA binding enthalpies was possible. Entropy and free energy could also be obtained from experiments, though benchmarking the QH entropy revealed that it was a poorer estimate. Finally, it was found that sequence parameters for binding enthalpies could be understood in terms of the dynamics of PNA homoduplexes through hydrogen bonding analysis.

## THEORY AND METHODS

**Literature Analysis.** Seven PNA homoduplexes with binding energies and entropies from thermal melting experiments were compiled from the literature.<sup>2,3,14–18</sup>

Experimental conditions used by the publications were inconsistent. Namely, the concentration of sodium ions and the use of a solubility-enhancing lysine tag were varied. Since prior research consistently agreed that PNA stability is negligibly affected<sup>2,5,6</sup> by a change in the concentration of sodium chloride, this was not considered critical.

To determine that lysine tagging did not affect the binding energies, the sequence GTAGATCACT, which appeared the most times in the published research, was categorized according to whether this tag was present or not. It was found that the binding energies and entropies for unmodified sequences were all within one standard error of the mean binding energies and entropies for sequences with the tag (Table S1).

Sequences were therefore grouped into a single data set regardless of these conditions.

**Thermal Melting Experiments.** Three PNA homoduplexes, CGATCG, AACGTT, and TAGCTA, were purchased from Eurogentec at 95% purity (Cat. Number BA-PN010-005) as determined by high-performance reverse-phase and ion-exchange liquid chromatography methods. Samples were diluted to a concentration of 1  $\mu$ M in deionized water and were then melted via an electric heater in a water bath from which samples were taken at 2 K intervals after 5 min equilibrations at the target temperature. The absorbances of the PNA strands at 260 nm were then monitored using an Agilent Technologies Cary 60 UV–vis spectrophotometer. Their absorbances increased during melting because single-

stranded nucleic acids are hyperchromic at 260 nm relative to the duplex.

Absorbances were normalized between zero and one to produce alpha curves, where  $\alpha = 1$  corresponds to a sample containing only double helices, while  $\alpha = 0$  contains only single strands. The melting point was defined at  $\alpha = 0.5$ . The equilibrium constant was then derived according to a two-state assumption for the transition using the equation for self-complementary helices.<sup>19</sup>

$$K = \frac{\alpha}{2C_T(1 - \alpha)^2} \quad (1)$$

where  $C_T$  is the total strand concentration. The natural logarithm of the equilibrium constant was then plotted against the reciprocal temperature to derive the standard enthalpies and entropies of the transition according to the linear form of the van't Hoff equation. This assumes that the entropy and enthalpy are temperature-independent.

$$\ln K = -\frac{\Delta H^\circ}{RT} + \frac{\Delta S^\circ}{R} \quad (2)$$

The free energy of binding of the strands was then calculated according to  $\Delta G = \Delta H - T\Delta S$ .

**Molecular Dynamics.** Simulated structures of PNA homoduplexes were generated using the proto-Nucleic Acid Builder.<sup>20</sup> A set of initial helical parameters<sup>21–23</sup> were obtained from published crystal or NMR structures. A weighted Monte Carlo conformational search was conducted to generate chemical structures based on these initial parameters and on thresholds such as bonding energies and interatomic distances set by the user. The structures generated by this were imported into the Gromacs program using the July 2021 edition of the CHARMM36m forcefield,<sup>24,25</sup> and the N- and C-termini were acetylated and amidated, respectively.

Simulated homoduplexes were energy minimized in vacuum over 50,000 simulation steps using a steepest-descent algorithm. Structures were then solvated in explicit waters in dodecahedral boxes using the simple point charge (SPC) water model spc216.<sup>26</sup> The distance between any atom of the solute and the edge of the box was set to a minimum of 15 Å. Periodic boundary conditions<sup>27</sup> were enforced to imitate a bulk solvent, and solvated structures were energy minimized using steepest-descent minimization over 50,000 simulation steps.

For each PNA homoduplex, three replicates were obtained from this point onward. Each replicate was heated from 30 to 298 K over 500 ps, and Newton's equations of motion were integrated using a leapfrog integrator.<sup>28,29</sup> Temperature was weakly coupled to an external bath using a velocity rescaling<sup>30</sup> algorithm. Harmonic restraints of 24 kcal mol<sup>-1</sup> Å<sup>-1</sup> were applied to non-hydrogen atoms, whereas hydrogen atoms were constrained using the LINCS<sup>31</sup> algorithm for the suppression of high-frequency hydrogen oscillations such that 2 fs timesteps could be used. During this and for all future simulations, the van der Waals interactions were handled using a switched cut-off scheme, with switching to zero occurring from 10 to 12 Å. The electrostatic nonbonded interactions were treated using a particle mesh Ewald (PME)<sup>32,33</sup> algorithm with quartic interpolation and a grid spacing of 1 Å. The short-ranged component of the PME was computed to 10 Å, and long-ranged components were handled using the fast Fourier transform library FFTW.<sup>34</sup>

After heating, position restraints were halved every 2 ns, with the equations of motion and temperature coupling thereon

handled by a stochastic integrator, which incorporated friction and noise terms.<sup>35</sup> After 10 ns of simulation, the non-hydrogen restraints were removed entirely, and structures were equilibrated for 10 ns in the canonical ensemble. They were then equilibrated for a further 10 ns in the isobaric–isothermal ensemble with pressure coupling at 1 bar handled by a Berendsen barostat.<sup>36</sup> 100 ns production runs were then carried out for each of the three replicas of each sequence using Parrinello–Rahman pressure coupling.<sup>37</sup>

**Trajectory Analysis.** Trajectories were analyzed by `gmx_mmpbsa`, which enables MM-GBSA estimations of enthalpy, entropy, and free energy to be calculated.<sup>38</sup> The free energy of a molecule computed using MM-GBSA<sup>39–41</sup> is given by the average molecular mechanics potential energy  $E_{MM}$ , the solvation energy  $G_{solv}$ , and the entropic term  $TS$ . The difference in free energies of two single PNA strands and the PNA homoduplex is then given by the difference of these.

$$\Delta G = \Delta\langle E_{MM} \rangle + \Delta\langle G_{solv} \rangle - T\Delta S \quad (3)$$

$E_{MM}$  is the sum of the bonded and nonbonded interactions of the molecule in vacuum, obtained from snapshots of the simulation trajectory, and accounts for the bond angle and dihedral energies between bonded atoms and the van der Waals and electrostatic interactions between nonbonded atoms.

$$\Delta\langle E_{MM} \rangle = \Delta\langle E_{bonded} \rangle + \Delta\langle E_{vdW} \rangle + \Delta\langle E_{elec} \rangle \quad (4)$$

In the single-trajectory approach used in this project, the individual PNA strands are isolated from the trajectory of the homoduplex, meaning that  $\Delta\langle E_{bonded} \rangle = 0$ . The single-trajectory approach assumes that the conformation of a PNA strand in the duplex is similar to a free PNA strand in solution but is computationally less expensive. The validity of the approach was tested by simulating three single-stranded PNAs, for a total of nine replicates, and calculating their enthalpies using the same generalized Born method. These enthalpies were then compared to the enthalpies of the single strands as evaluated by the single-trajectory approach, that is, when taken from the trajectory of the duplex. The difference in this conformation enthalpy<sup>42,43</sup> was limited ( $\sim 1\%$ ) in all three cases (Table S2).

The solvation energy difference  $\Delta\langle G_{solv} \rangle$  from eq 3 is calculated using a continuum solvent model, which approximates the explicit water solvent as a homogeneous medium with a dielectric constant of 78.5 at 298 K. It is composed of polar and nonpolar components. The polar term accounts for the electrostatic interactions between solute and solvent. The nonpolar component accounts for the work of formation of a cavity the size of the solute in the solvent and the short-ranged van der Waals interactions between the solute and the solvent.<sup>44</sup>

$$G_{solv} = G_{polar} + G_{nonpolar} \quad (5a)$$

$$G_{polar} \approx G_{GB} = -k \sum_{ij} \frac{q_i q_j}{f_{GB}} \quad (5b)$$

$$G_{nonpolar} \approx G_{SA} = \gamma SA + b \quad (5c)$$

where  $q_i$  and  $q_j$  are atomic charges and  $k$ , in the generalized Born model, is a function of the dielectric constants of the solvent and of free space, where it is unity.  $f_{GB}$  is a function of interatomic distances and Born radii of the  $i$ th and  $j$ th atoms,

wherein the Born radius is the spherically averaged distance between an atom of the solute and the solvent.  $G_{SA}$  is a linear function of the solute's solvent-accessible surface area (SA).<sup>45</sup>

The final contribution to the free energy change according to eq 4 is then  $T\Delta S$ , which was evaluated in this study using a QH approximation.<sup>46</sup>

Using the MM-GBSA approach, the binding energies for PNA homoduplexes were evaluated. Snapshots from each production trajectory were stripped of explicit water, and the first 20 ns were discarded. The remaining 80 ns were analyzed using MM-GBSA. The version of the generalized Born model, `igb = 5`,<sup>47</sup> was used for the analysis, and enthalpy change was taken to be  $\Delta\langle E_{MM} \rangle + \Delta\langle G_{solv} \rangle$  according to  $\Delta G = \Delta H - T\Delta S$ .

A standard state correction was applied to the free energy differences to correct for the box volumes and thereby standardize the concentrations.

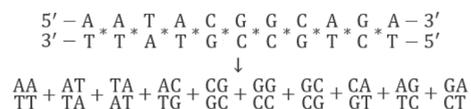
$$\Delta G^\circ = \Delta G - RT \ln \left( \frac{V_{com}}{V^\circ} \right) \quad (6)$$

where  $V^\circ = 1661 \text{ \AA}^3$  and  $V_{com}$  is the volume of the simulation box.<sup>48</sup>

The production trajectories with explicit waters were then analyzed using `gmx_rms` and `gmx_hbond`, which were used to compute the root-mean-square atomic deviations (rmsds) (Figure S1) over time and the number of Watson–Crick hydrogen bonds between the base pairs over time. Hydrogen bonds were defined as bonds between the atom triplets involved in Watson–Crick bonding with distances between donor and acceptor of less than 0.325 nm and angles of hydrogen-donor–acceptor of less than  $30^\circ$ .

To confirm that run durations were sufficient to obtain converged thermodynamic data, relative enthalpies, entropies, and free energies were plotted as functions of sequence length (Figures S2–S4). To determine whether simulations were equilibrated and repeatable, rmsd histogram distributions for the replicas of each sequence were overlapped (Figures S5 and S6).

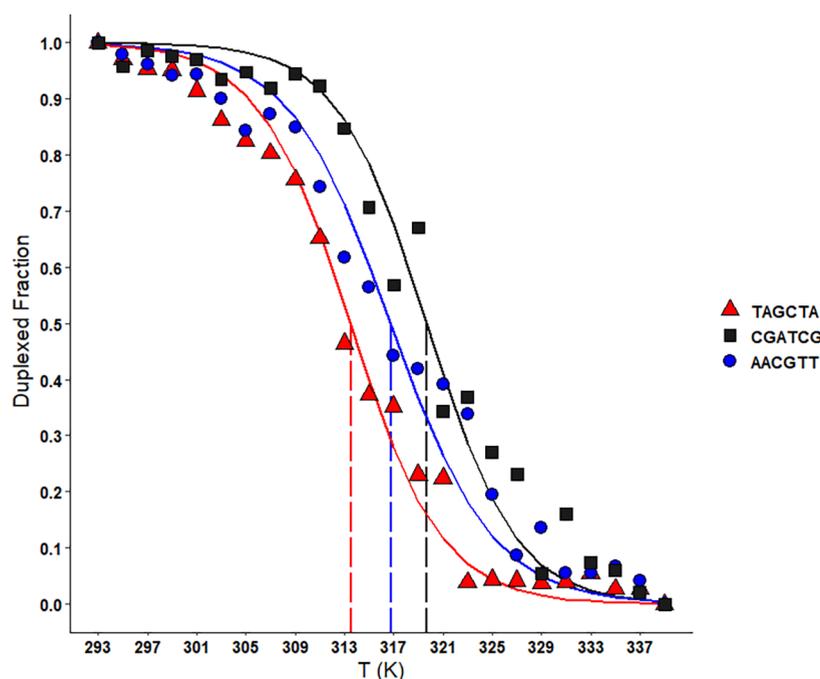
**Modeling Nucleic Acid Binding Energies.** A nucleic acid duplex can be thought of as a series of overlapping base pair stacks (Figure 3). A stack is two base pairs in sequence,



**Figure 3.** An oligonucleotide duplex broken down into a series of overlapping stacks, as indicated by the asterisks. Stacks are written below the duplex, as indicated by the arrow.

and there are 10 unique stacks for homoduplexes. Overlapping means that the first pair of the following stack is the second pair of the preceding stack. This is known as the nearest-neighbor model. Given that the stabilities of nucleic acid duplexes are impacted both by their base-pairing and base-stacking energies, the nearest-neighbor method is theoretically rigorous in that it accounts for both.<sup>49–51</sup>

The energies and entropies of binding are given by summing over the energies and entropies of these stacks. In addition, the helix initiation energy or entropy and the terminal guanine–cytosine (GC) energy or entropy are added. The binding



**Figure 4.** Normalized melting curves for the three PNA hexamers TAGCTA, CGATCG, and AACGTT, as measured by UV melting experiments. The duplexed fraction,  $\alpha$ , is indicated on the  $y$  axis and the temperature in kelvin on the  $x$  axis. Melting points are shown by the dashed vertical lines. Points represent the means of independent triplicates and were obtained at 2 K intervals.

energy or entropy of a sequence is then a sum over the stacks, the helix initiation term, and the terminal GC term.

$$B \begin{cases} 0 & n(\text{T. GC}) < 2 \\ 1 & n(\text{T. GC}) = 2 \end{cases} \quad (7a)$$

$$\Delta\chi = \Delta\chi_{\text{init}} + B\Delta\chi_{\text{T.GC}} + \sum_i j_i(\Delta\Delta\chi_i) \quad (7b)$$

where  $\Delta\chi$  is a binding energy or entropy.  $\Delta\chi_{\text{init}}$  is a helix initiation energy or entropy.  $\Delta\chi_{\text{T.GC}}$  is the energy or entropy added so long as the number of terminal GC base pairs,  $n(\text{T.GC})$ , is two, as indicated by eq 7a. Put differently,  $B\Delta\chi_{\text{T.GC}}$  evaluates to zero if there is a terminal adenine-thymine (AT) base pair.  $\Delta\Delta\chi_i$  is the incremental change in binding energy or entropy given by the  $i$ th unique base pair with occurrence  $j_i$ .

For self-complementary helices, the binding entropy is decreased by  $1.4 \text{ cal K}^{-1} \text{ mol}^{-1}$  to account for the C2 rotational symmetry of the double helix, which reduces the degrees of freedom upon binding.<sup>51</sup>

Equation 7b has the matrix form

$$\begin{pmatrix} \Delta\chi_1 \\ \vdots \\ \vdots \\ \vdots \\ \Delta\chi_N \end{pmatrix}_{\chi} = \begin{pmatrix} j_1^1 & \cdots & j_v^1 \\ \vdots & & \vdots \\ \vdots & \ddots & \vdots \\ \vdots & & \vdots \\ j_1^N & \cdots & j_v^N \end{pmatrix}_{\mathcal{S}} \cdot \begin{pmatrix} \Delta\chi_1 \\ \vdots \\ \Delta\chi_{v-2} \\ \Delta\chi_{\text{init}} \\ \Delta\chi_{\text{T.GC}} \end{pmatrix}_{\epsilon} \quad (8)$$

where  $\chi$  is a  $1 \times N$  column matrix containing either mean enthalpies, entropies, or free energies of hybridization of PNA homoduplexes.  $\mathcal{S}$  is an  $N \times v$  matrix containing the occurrences

$j$  of all stacked base pairs, terminal GC and initiation terms.  $\epsilon$  is a  $1 \times v$  column matrix containing the unknown energies or entropies of these terms. These unknowns are then evaluated by solving the linear least-squares problem, which minimizes the sum of the square of the residuals of  $\chi$  with respect to  $\mathcal{S}$ .

The problem is generalized by weighting each observable  $\Delta\chi$  by its reciprocal standard error.

$$\sigma^{-1} \cdot \chi = \sigma^{-1} \cdot \mathcal{S} \cdot \epsilon \quad (9)$$

where  $\sigma^{-1}$  is an  $N \times N$  diagonal matrix containing the reciprocal standard errors of each observable with non-diagonal elements of zero. The estimate  $\epsilon$  obtained by solving the multiple linear regression problem thus contains the error-weighted least-squares estimates of the energies or entropies of the terms on the right-hand side of eq 7b.

Equation 9 can be used to decompose MM-GBSA estimates of binding enthalpies, entropies or free energies for PNA homoduplexes into sequence parameters.

## RESULTS AND DISCUSSION

**Thermal Melting Data.** Melting points of the three PNA hexamers defined as  $\alpha = 0.5$  were observed at  $313.5 \pm 0.4$ ,  $316.8 \pm 0.4$ , and  $319.7 \pm 0.8 \text{ K}$  for TAGCTA, AACGTT, and CGATCG, respectively (Figure 4).

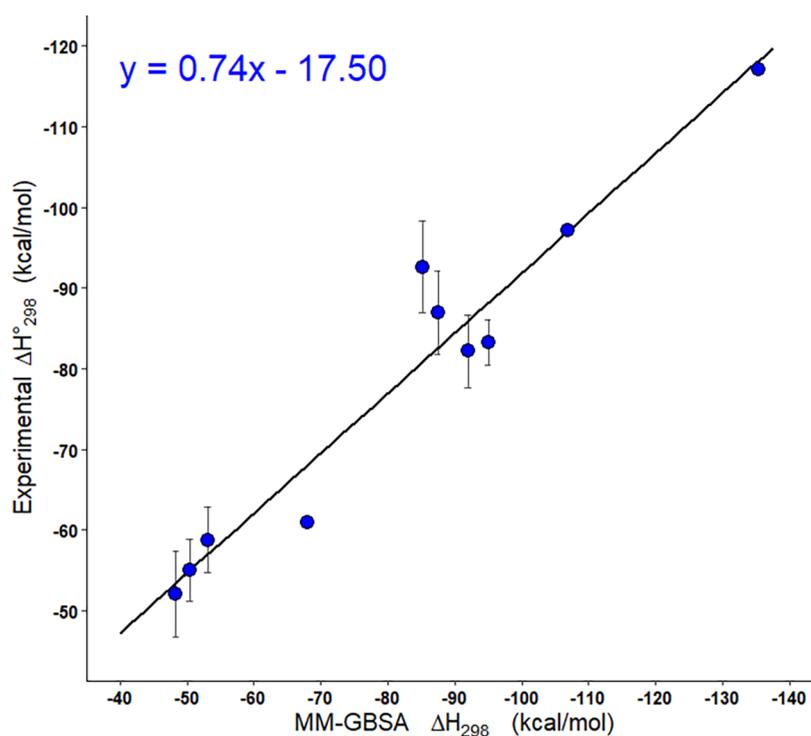
Using eqs 1 and 2, the standard binding free energies, enthalpies and entropies were determined for the three hexamers. The standard errors of the measured enthalpies and entropies for all three sequences overlapped, whereas the free energy of hybridization for CGATCG was approximately  $1 \text{ kcal mol}^{-1}$  more negative. This suggested that CGATCG was stabilized relative to the other two homoduplexes as was corroborated by their melting temperatures.

The binding energies and entropies of the three sequences were compiled alongside the binding energies and entropies obtained from the literature. This produced a data set totaling

**Table 1. Standard Free Energies, Enthalpies, and Entropies of Binding for the 10 PNA Homoduplexes from Thermal Melting Experiments<sup>a</sup>**

sequence	$-\Delta G_{298}^{\circ}$ (kcal mol <sup>-1</sup> )	$-\Delta H_{298}^{\circ}$ (kcal mol <sup>-1</sup> )	$-\Delta S_{298}^{\circ}$ (cal K <sup>-1</sup> mol <sup>-1</sup> )
(a) Experimental Thermal Melting Data			
CGATCG	13.64 ± 0.25	58.78 ± 4.09	151.42 ± 12.89
AACGTT	12.64 ± 0.39	52.05 ± 5.30	132.24 ± 16.45
TAGCTA	12.34 ± 0.29	55.08 ± 3.85	143.41 ± 11.95
(b) Literature Thermal Melting Data			
GTAGATCACT <sup>b,c,d,e,f,g</sup>	18.99 ± 0.87	86.92 ± 5.00	227.97 ± 14.48
TGTTACGACT <sup>h</sup>	21.08 ± 1.00	92.60 ± 5.70	240.00 ± 16.11
AGGTAACCAG <sup>d</sup>	18.76 ± 0.60	83.30 ± 2.80	216.60 ± 7.40
AGTGAAGCAG <sup>d</sup>	19.02 ± 0.82	82.15 ± 4.56	211.85 ± 12.86
TGATCTAC <sup>f</sup>	13.10 ± 0.00	60.90 ± 0.00	224.70 ± 0.00
GTAGATCACTGT <sup>f</sup>	21.40 ± 0.00	97.10 ± 0.00	253.80 ± 0.00
GTAGATCACTGTCACT <sup>f</sup>	26.40 ± 0.00	117.10 ± 0.00	304.30 ± 0.00

<sup>a</sup>Standard errors are as provided by the study author or are the standard errors from reported energies or entropies of multiple publications if a sequence is reported multiple times. Standard errors of zero indicate that they were not standard errors but provided by the study author. <sup>b</sup>Tomac, S.; Sarkar, M.; Ratilainen, T.; Wittung, P.; Nielsen, P.; Nordén, B.; Gräslund, A. Ionic Effects on the Stability and Conformation of Peptide Nucleic Acid Complexes. *J. Am. Chem. Soc.* **1996**, *118*, 5544–5552. <sup>c</sup>Sen, A.; Nielsen, P. On the stability of peptide nucleic acid duplexes in the presence of organic solvents. *Nuc. Acid. Res.* **2007**, *35*, 3367–3374. <sup>d</sup>Sen, A.; Nielsen, P. Unique Properties of Purine/Pyrimidine Asymmetric PNA-DNA Duplexes: Differential Stabilization of PNA-DNA Duplexes by Purines in the PNA Strand. *Biophys. J.* **2006**, *90*, 1329–1337. <sup>e</sup>Ratilainen, T.; Holmén, A.; Tuite, E.; Haaima, G.; Christensen, L.; Nielsen, P.; Nordén, B. Hybridization of Peptide Nucleic Acid. *Biochemistry.* **1998**, *37*, 12331–12342. <sup>f</sup>Totsingan, F. Synthesis and Applications of PNA and Modified PNA in Nanobiotechnology. Ph.D. Thesis, University of Parma, Parma, Italy, **2007**. <sup>g</sup>Sforza, S.; Haaima, G.; Marchelli, R.; Nielsen, P. Chiral Peptide Nucleic Acids (PNAs): Helix Handedness and DNA Recognition. *Eur. J. Org. Chem.* **1999**, 197–204. <sup>h</sup>Jasiński, M.; Miszkiewicz, J.; Feig, M.; Trylska, J. Thermal Stability of Peptide Nucleic Acid Complexes. *J. Phys. Chem. B.* **2019**, *123*, 8168–8177.

**Figure 5.** Correlation analysis for the enthalpy of binding of PNA homoduplexes from thermal melting experiments and from MM-GBSA analysis.

10 sequences, 7 of which were obtained from the literature,<sup>2,3,14–18</sup> against which computationally derived energies and entropies could be benchmarked (Table 1).

**GBSA Benchmarking.** The entropies, enthalpies, and free energies of binding for the 10 PNA homoduplexes were obtained using MM-GBSA. Means and standard errors were obtained from triplicates. Linear regression of the enthalpies of binding from thermal melting experiments against MM-GBSA

analysis produced an  $R^2$  of 0.93, indicating that they were well correlated. A high  $R^2$  would be anticipated given that there were few sequences of the same length, meaning a positive correlation requires only that the simulation and reality both report a higher enthalpy of binding for longer lengths, which is what was observed. The mean absolute difference, taken as the error of the prediction, between the enthalpy of binding from thermal melting and from MM-GBSA experiments was 7.87

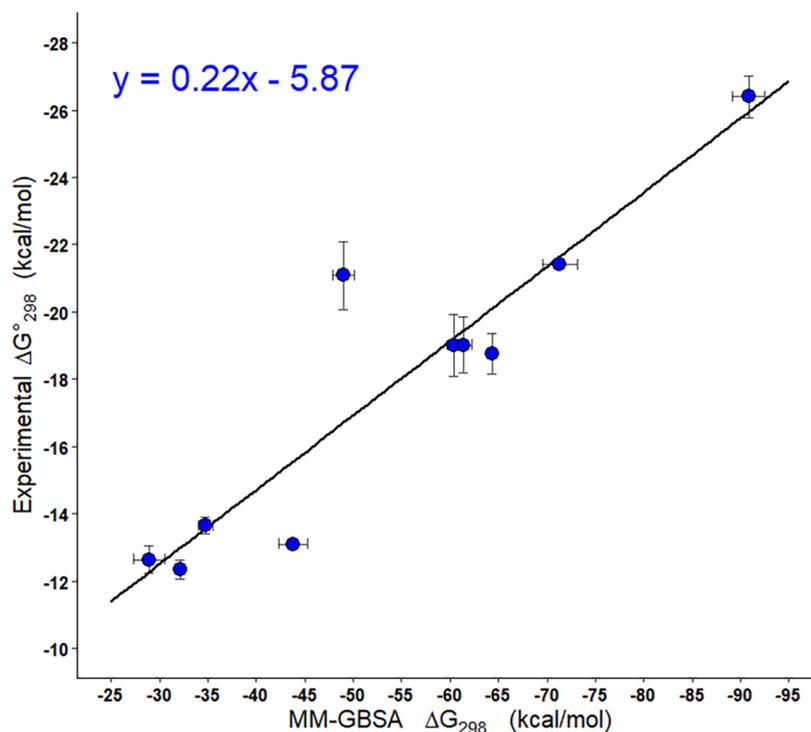


Figure 6. Correlation analysis for the Gibbs free energy of binding of PNA homoduplexes from thermal melting experiments and MM-GBSA.

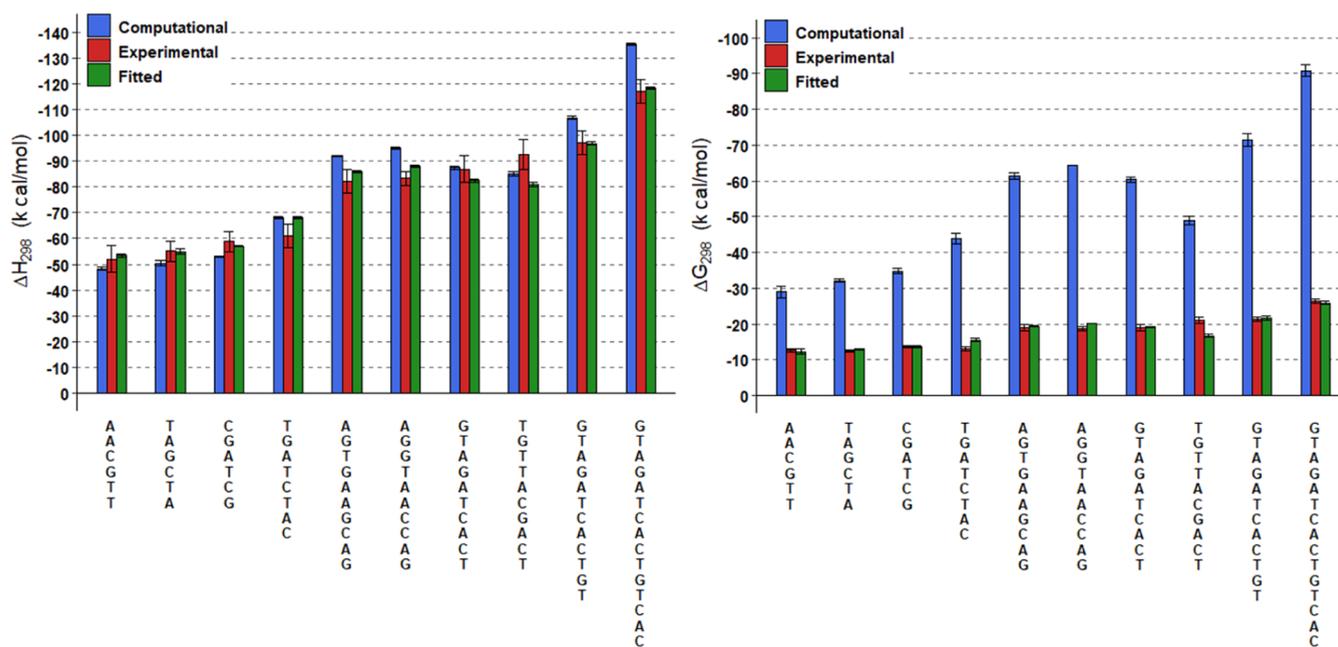


Figure 7. Enthalpies, on the left, and free energies, on the right, for the 10 PNA homoduplexes of the benchmarking set. Data from thermal melting experiments is in red. Data from MM-GBSA is in blue. Data after fitting according to  $y = ax + b$  is in green. Error bars indicate the standard errors. Standard errors for computational and fitted data are obtained from triplicate simulations.

$\text{kcal mol}^{-1}$ , and the mean relative error was 10.01%. The equation of the line between the thermal melting and MM-GBSA enthalpies (Figure 5) yielded a slope coefficient  $a = 0.74$  and a shift coefficient  $b = -17.50$ . The large shift coefficient indicated a deviation between the thermal melting and MM-GBSA analysis. This was likely a result of a combination both of the assumptions made during simulation and of the small size of the data set, with only 10 sequences with thermal

melting data being analyzed, causing the shift to be largely determined by only a few points.

Treatment of the binding enthalpies according to the equation of the line  $\Delta H_{\text{exp}}^{\circ} = 0.74(\Delta H_{\text{com}}) - 17.50$  reduced the relative and absolute errors of the estimates to 5.25% and 4.12  $\text{kcal mol}^{-1}$ , respectively.<sup>2,3,14–18</sup>

In contrast to the binding enthalpy, binding entropy and free energy were worse estimated, though they remained highly correlated with  $R^2$  of 0.84 and 0.86, respectively. The mean

absolute and relative errors were 31.49 kcal mol<sup>-1</sup> and 50.14% for  $T\Delta S$  and 35.95 kcal mol<sup>-1</sup> and 102.68% for the Gibbs free energy of binding. These errors are significantly higher than the error for the enthalpy and indicate that the configurational entropy as estimated by QH analysis does not accurately evaluate the entropy of binding from thermal melting experiments.

Similarly to enthalpy, treatment of free energy according to the equation of the line,  $\Delta G_{\text{exp}}^{\circ} = 0.22(\Delta G_{\text{com}}^{\circ}) - 5.87$  (Figure 6) reduced the relative and absolute errors to 6.53% and 1.16 kcal mol<sup>-1</sup>, respectively. This demonstrates that significant correction can be achieved via linear regression, enabling the prediction of experimental binding energies from computer simulations even when there is quantitative disagreement (Figure 7).

**Nearest-Neighbor Modeling.** To decompose binding energies into sequence parameters according to eqs 8 or 9, the coefficients in  $\epsilon$  must converge. For that to be the case, the number of binding energies in  $\chi$  must be large. The 10 PNA homoduplexes in Table 1 are not sufficient and so an additional 39 PNA homoduplexes were simulated (Table S3) and estimates of their binding enthalpies, entropies, and free energies were determined using MM-GBSA. A total of 49 PNA homoduplexes were used to produce the nearest-neighbor model. As a result, it is assumed that the errors described in the previous section are representative of the average error of the remaining 39 homoduplexes, which lack experimental data.

Benchmarking the subset of 10 sequences gave relative errors of 10.01 and 102.68% for the enthalpy and free energy of binding, respectively. A correction reduced these to 5.25 and 6.53%. This shows that thermal melting data can be accurately predicted using MM-GBSA without correction for enthalpy, though predictions of free energy require correction.

It is possible to develop a nearest-neighbor model with both corrected and uncorrected data. However, corrections according to the equations of the line,  $\Delta\chi_{\text{exp}} = a\Delta\chi_{\text{com}} + b$ , in Figures 5 and 6, impact the sequence parameters on the right-hand side of eq 7b in two ways

$$(\Delta\Delta\chi_i)_C = a(\Delta\Delta\chi_i) \quad (10a)$$

$$(\Delta\chi_{\text{init}})_C = a(\Delta\chi_{\text{init}}) + b \quad (10b)$$

where the subscript C indicates a term derived from solving the multiple regression problem using binding energies or entropies, which have been corrected according to  $\Delta\chi_{\text{exp}} = a\Delta\chi_{\text{com}} + b$ . The impact of the correction on the nearest-neighbor terms is not trivial. First, by scaling the stacked base pair energies according to the slope coefficient  $a$  (eq 10a), the equality  $\Delta\Delta G = \Delta\Delta H - T\Delta\Delta S$  only holds when  $a_1\Delta\Delta G = a_2\Delta\Delta H - a_3T\Delta\Delta S$ , where  $a_i$  is the slope coefficient for the  $i$ th state variable such as the enthalpy. This is unlikely to be the case since the source of error differs between enthalpy and entropy: for enthalpy, it arises from the evaluation of potential energy, whereas for entropy, it arises from the evaluation of the fluctuations in atomic coordinates. In addition to this, the shift coefficient  $b$  may be large compared to the actual helix initiation energy or entropy. The corrected helix initiation value on the left-hand side of eq 10b thus may largely be determined by a metric of computational error as opposed to any physically relevant quantity.

So long as these considerations are taken into account, the model can be highly predictive of real-world measurements.

This was the case for previous research on MM-GBSA estimates of DNA and RNA.<sup>13</sup>

In this research, corrected MM-GBSA data were not used for fitting the nearest-neighbor model so as to maximize the likelihood that sequence parameters were meaningful. Since the binding entropy, and therefore free energy, quantitatively disagreed with thermal melting data, a nearest-neighbor model was generated solely for the binding enthalpy. This was possible since, for the binding enthalpy, the experimental error was approximately equal to a typical experimental error of 10%.<sup>13</sup>

MM-GBSA enthalpies for all 49 sequences (Table S3) were used to solve the generalized multiple linear regression problem in eqs 5c–7a. Nearest-neighbor stacking, initiation, and terminal GC enthalpies were thus determined (Table 2).

**Table 2. Nearest-Neighbor Binding Enthalpies for PNA Homoduplexes<sup>a</sup>**

	$-\Delta H_{298}$ (kcal mol <sup>-1</sup> )	$-\langle\Delta H_{298}\rangle$ (kcal mol <sup>-1</sup> )
AA	8.65 ± 0.16	8.76 ± 0.31
AT	8.28 ± 0.32	
TA	9.34 ± 0.32	
AG	10.64 ± 0.27	9.72 ± 0.46
GA	8.91 ± 0.35	
AC	8.90 ± 0.19	
CA	10.31 ± 0.17	
GG	12.32 ± 0.19	10.91 ± 1.12
GC	8.69 ± 0.30	
CG	11.71 ± 0.34	
	$-\Delta H_{298}$ (kcal mol <sup>-1</sup> )	
T.GG	1.62 ± 0.35	
Init	2.27 ± 0.46	

<sup>a</sup>± indicates the standard error of the mean.

All parameters from Table 2 were significant to  $p < 10^{-5}$  in solving the multiple linear regression problem. Each stack was categorized according to the number of canonical Watson–Crick hydrogen bonds between them. These were six for GG, GC, and CG; five for AG, GA, AC, and CA; and four for AA, AT, and TA. On average, the stacking enthalpy became more negative as the number of Watson–Crick hydrogen bonds increased. This was in line with expectations since the hydrogen bonds stabilize the duplex.

The stacks GG and CG were more stabilizing relative to GC despite having the same number of hydrogen bonds. This difference could arise from the base stacking interaction, since base pairing only partly accounts for the energies of interaction between the strands. However, it may also reflect the sequence selection used in this study since bias can be introduced by particular stacks often coinciding or by a few sequences accounting for a large amount of the total occurrences of a stack.

Atypical for nucleic acid binding enthalpies is the negative helix initiation enthalpy. This enthalpy is typically positive for DNA and RNA duplexes.<sup>48,49</sup> The explanation for it being stabilizing in PNA is due to PNA neutral backbone. In DNA and RNA, negative backbones repel, introducing a barrier to the association of the two strands. For PNA, however, this net repulsion does not occur, and so it is feasible that helix initiation is more favorable which would explain the negative helix initiation enthalpy. Similarly to other nucleic acid nearest-

**Table 3. Experimental Binding Enthalpies  $\Delta H_{\text{exp}}^{\circ}$ , Nearest-Neighbor Predicted Binding Enthalpies  $\Delta H_{\text{NN}}^{\circ}$ , and Absolute  $|-\Delta H_{\text{exp}}^{\circ} + \Delta H_{\text{NN}}^{\circ}|$  and Relative  $|-\Delta H_{\text{exp}}^{\circ} + \Delta H_{\text{NN}}^{\circ}|/\Delta H_{\text{exp}}^{\circ}$  Differences for 10 Homoduplexes**

sequence	$-\Delta H_{\text{exp}}^{\circ}$ (kcal mol <sup>-1</sup> )	$-\Delta H_{\text{NN}}^{\circ}$ (kcal mol <sup>-1</sup> )	$ -\Delta H_{\text{exp}}^{\circ} + \Delta H_{\text{NN}}^{\circ} $ (kcal mol <sup>-1</sup> )	difference (%)
CGATCG	58.78 ± 4.09	53.40	5.38	9.15
AACGTT	52.05 ± 5.30	49.07	2.98	5.73
TAGCTA	55.08 ± 3.85	50.91	4.17	7.57
GTAGATCACT	86.92 ± 5.00	87.09	0.17	0.20
TGTTACGACT	92.60 ± 5.70	85.24	7.36	7.95
AGGTAACCAG	83.30 ± 2.80	92.28	8.98	10.78
AGTGAAGCAG	82.15 ± 4.56	89.95	7.80	9.49
TGATCTAC	60.90 ± 0.00	67.55	6.65	10.92
GTAGATCACTGT	97.10 ± 0.00	106.30	9.20	9.47
GTAGATCACTGTCAC	117.10 ± 0.00	136.03	18.93	16.17

neighbor models, the terminal GC term is negative, indicating that the absence of terminal AT base pairs stabilizes the double helix.

The parameters in Table 2 can be used to estimate the binding enthalpies of PNA duplexes according to eq 7b (Table 3). Predicted binding enthalpies had a mean relative error of 8.74% when compared to thermal melting binding enthalpies. This was a decrease from the 10.01% error from the direct comparison of MM-GBSA to thermal melting data for the subset of 10 sequences. Since the nearest-neighbor model was based on a total of 49 sequences, the decrease in error showed that the accuracy of prediction increased by accounting for unrelated sequences.

This indicated that binding enthalpy could be readily predicted as a function of sequence parameters and that MM-GBSA could derive these sequence parameters. It is consequently possible that the nearest-neighbor model developed herein could aid in the design and structural understanding of PNA homoduplexes.

**Hydrogen Bond Analysis.** Since the MM-GBSA nearest-neighbor predictions agree well with the available experimental data, it is possible to interpret them as having physical meaning. The first indication of this being the case is that, when the stacks are grouped according to the number of Watson–Crick hydrogen bonds involved, the binding enthalpy becomes, on average, more stable as the number of bonds increases.

Additional interpretations of physical meaning can be obtained through dynamics studies. The terminal GC term, T.GC, was investigated using hydrogen bond analysis. Watson–Crick hydrogen bonds of terminal and 3rd-position, classed as internal base pairs, were monitored. The number of Watson–Crick hydrogen bonds was normalized and these were expressed as percentages. These percentages were averaged over the simulation durations and for simulations involving sequences of the same length (Table 4).

The percentage of unbroken bonds, from Table 4, is independent of sequence length and always, on average, higher for internal than terminal base pairs. Terminal AT hydrogen bonds were broken more often than terminal GC hydrogen bonds, though both spent most of their time in the bound state. Additionally, the standard deviation in the percentage of unbroken bonds was greater, most often by around 5-fold but in one case over 15-fold, for terminal AT than GC base pairs. This indicated that the terminal AT hydrogen bonding behavior was characterized by fluctuations between bound and melted states.

**Table 4. Mean Normalized Number of Hydrogen Bonds, between Zero and One, Expressed as a Percentage of Terminal and Internal AT and GC Base Pairs of Varying Sequence Lengths<sup>a</sup>**

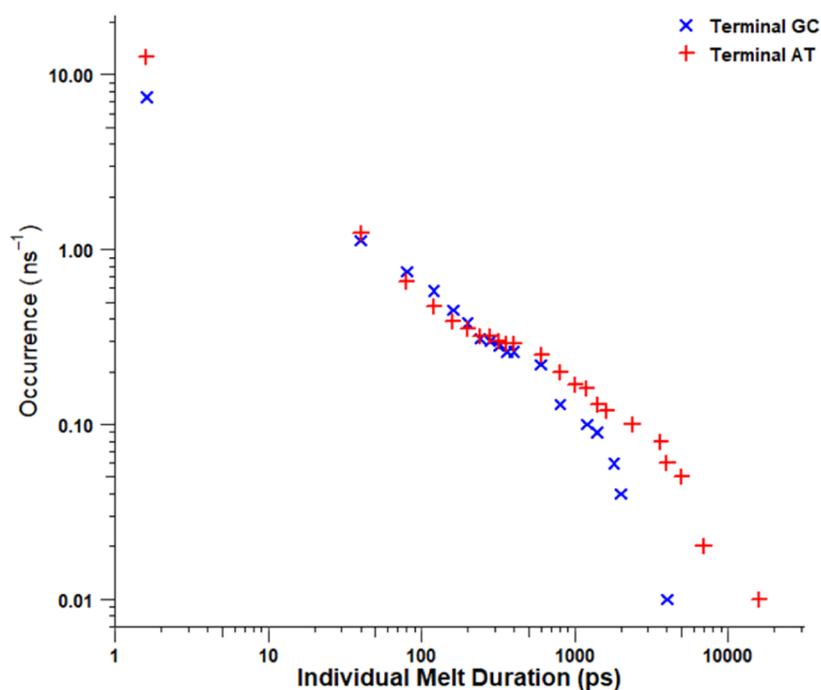
length	% H-bonded	
	AT	GC
		terminal
6	70.4 ± 27.1	83.3 ± 4.3
8	74.4 ± 12.0	88.4 ± 2.9
10	84.7 ± 13.2	89.1 ± 2.1
12	79.3 ± 10.6	83.2 ± 7.1
14	80.3 ± 7.8	84.1 ± 6.5
16	68.4 ± 29.4	88.4 ± 1.7
18	67.6 ± 20.6	82.9 ± 3.9
		internal
6	90.2 ± 2.4	95.7 ± 0.3
8	90.9 ± 0.5	94.0 ± 1.5
10	91.7 ± 0.9	95.3 ± 0.5
12	89.9 ± 1.6	96.0 ± 0.9
14	91.6 ± 0.7	96.0 ± 0.7
16	91.1 ± 0.4	95.0 ± 0.9
18	89.6 ± 2.4	95.8 ± 0.4

<sup>a</sup>Errors indicate standard deviations.

To determine the nature of these fluctuations, the duration of the melting events was plotted against the occurrence, per nanosecond, of a melting event of that duration being initiated (Figure 8). A melting event was defined as an uninterrupted period of time with at least one hydrogen bond broken.

From Figure 8, it was observed that, for both AT and GC terminal base pairs, melting events of a duration less than one nanosecond were similarly common. By contrast, melting events longer than this were more common for terminal AT than GC pairs. This indicated that the large deviations in the number of unbroken bonds for the terminal AT pairs were likely a consequence of long, persistent melting events as opposed to short-lived ones. Since long melting events of this manner may destabilize the duplex and decrease its binding enthalpy, this may partly explain why the T.GC enthalpy, which is present only in the absence of AT termini, is negative.

In addition, the total melting time, obtained by aggregating all individual melting events during a run, was obtained. Total melting times for terminal (Figure S2) and internal (Figure S3) AT and GC base pairs were then plotted against the occurrence of a run with at least that much melting time out of all runs.



**Figure 8.** Occurrences per nanosecond of an individual melting event of at least a duration of picoseconds initiating for terminal GC base pairs, in blue, and terminal AT base pairs, in red.

Consequently, the behavior observed in the hydrogen bonding analysis can corroborate the predictions of the nearest-neighbor model. This supports the hypothesis that an MM-GBSA nearest neighbor model contains meaningful quantities that relate energy with structure.

## CONCLUSIONS

Parameters for calculating the binding enthalpies of PNA duplexes as a function of their sequence have been developed. The difference between these predictions and PNA duplex binding enthalpies from thermal melting data was 8.74%. This was a reduction from an error of 10.01% from the direct comparison between MM-GBSA and thermal melting binding enthalpies. This suggested that the accuracy of the estimates improved when unrelated sequences were considered.

These results show that MM-GBSA can accurately decompose PNA binding enthalpies into parameters with structural meaning. This was demonstrated for the terminal GC term using a hydrogen bonding analysis, which showed that AT termini instabilities are characterized by persistent melting events.

Notably, the helix initiation enthalpy, which is positive for nearest-neighbor models of DNA and RNA,<sup>48,49</sup> was negative for PNA. A plausible reason for this is that the backbone of PNA is neutrally charged, and hence association does not have to overcome the like-charge repulsion that occurs between the phosphoribose backbones of DNA or RNA. Despite its approximate nature, MM-GBSA was able to resolve binding enthalpies with accuracy comparable to experimental error. This research demonstrates the usefulness of MM-GBSA for predicting thermodynamic quantities in cases of limited experimental data, as would be expected for studies into novel nucleic acids.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcc.2c05547>.

Binding energies and entropies for all models; convergence of MM-GBSA outputs; rmsd traces of example models and replicas; and rmsd histograms of all replicas (PDF)

## AUTHOR INFORMATION

### Corresponding Author

Jack Goodman – University of the West of England, Bristol BS16 1QY, U.K.; [orcid.org/0000-0003-4677-723X](https://orcid.org/0000-0003-4677-723X);  
Email: [jack2.goodman@live.uwe.ac.uk](mailto:jack2.goodman@live.uwe.ac.uk)

### Authors

David Attwood – University of the West of England, Bristol BS16 1QY, U.K.  
Janice Kiely – University of the West of England, Bristol BS16 1QY, U.K.  
Pablo Coladas Mato – GKN Aerospace, Bristol BS34 6FB, U.K.; [orcid.org/0000-0003-4347-6204](https://orcid.org/0000-0003-4347-6204)  
Richard Luxton – University of the West of England, Bristol BS16 1QY, U.K.

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jpcc.2c05547>

### Author Contributions

J.G. designed and conducted the research. J.G. and D.A. wrote the manuscript. The study was conceived by R.L. and J.K. All authors provided advice, reviewed the manuscript, and assisted in the direction of the research.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was supported through joint funding from the University of the West of England and GKN Aerospace.

## REFERENCES

- (1) Wittung, P.; Nielsen, P. E.; Buchardt, O.; Egholm, M.; Nordén, B. DNA-like double helix formed by peptide nucleic acid. *Nature* **1994**, *368*, 561–563.
- (2) Tomac, S.; Sarkar, M.; Ratilainen, T.; Wittung, P.; Nielsen, P.; Nordén, B.; Gräslund, A. Ionic Effects on the Stability and Conformation of Peptide Nucleic Acid Complexes. *J. Am. Chem. Soc.* **1996**, *118*, 5544–5552.
- (3) Jasiński, M.; Miszkiewicz, J.; Feig, M.; Trylska, J. Thermal Stability of Peptide Nucleic Acid Complexes. *J. Phys. Chem. B* **2019**, *123*, 8168–8177.
- (4) Nielsen, P. E. Peptide Nucleic Acid. A Molecule with Two Identities. *Acc. Chem. Res.* **1999**, *32*, 624–630.
- (5) Eriksson, M.; Nielsen, P. E. PNA-nucleic acid complexes. Structure, stability and dynamics. *Q. Rev. Biophys.* **1996**, *29*, 369–394.
- (6) Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. Sequence-Selective Recognition of DNA by Strand Displacement with a Thymine-Substituted Polyamide. *Science* **1991**, *254*, 1497–1500.
- (7) Shakeel, S.; Karim, S.; Ali, A. Peptide nucleic acid (PNA) - a review. *J. Chem. Technol. Biotechnol.* **2006**, *81*, 892–899.
- (8) Gupta, A.; Mishra, A.; Puri, N. Peptide nucleic acids: Advanced tools for biomedical applications. *J. Biotechnol.* **2017**, *259*, 148–159.
- (9) Ray, A.; Nordén, B. Peptide nucleic acid (PNA): its medical and biotechnical applications and promise for the future. *FASEB J.* **2000**, *14*, 1041–1060.
- (10) Pellestor, F.; Paulasova, P. The peptide nucleic acids (PNAs), powerful tools for molecular genetics and cytogenetics. *Eur. J. Hum. Genet.* **2004**, *12*, 694–700.
- (11) Moccia, M.; Adamo, M.; Saviano, M. Insights on chiral, backbone modified peptide nucleic acids: Properties and biological activity. *Artif. DNA PNA XNA* **2014**, *5*, No. e1107176.
- (12) Jasiński, M.; Feig, M.; Trylska, J. Improved force fields for peptide nucleic acids with optimized backbone torsion parameters. *J. Chem. Theory Comput.* **2018**, *14*, 3603–3620.
- (13) Golyshev, V.; Pyshnyi, D.; Lomzov, A. Calculation of Energy for RNA/RNA and DNA/RNA Duplex Formation by Molecular Dynamics Simulation. *Mol. Biol.* **2021**, *55*, 927–940.
- (14) Sforza, S.; Haaima, G.; Marchelli, R.; Nielsen, P. Chiral Peptide Nucleic Acids (PNAs): Helix Handedness and DNA Recognition. *Eur. J. Org. Chem.* **1999**, *1999*, 197–204.
- (15) Ratilainen, T.; Holmén, A.; Tuite, E.; Haaima, G.; Christensen, L.; Nielsen, P.; Nordén, B. Hybridization of Peptide Nucleic Acid. *Biochemistry* **1998**, *37*, 12331–12342.
- (16) Sen, A.; Nielsen, P. On the stability of peptide nucleic acid duplexes in the presence of organic solvents. *Nucleic Acids Res.* **2007**, *35*, 3367–3374.
- (17) Totsingan, F. Synthesis and Applications of PNA and Modified PNA in Nanobiotechnology. Ph.D. Thesis, University of Parma, Italy, 2007.
- (18) Sen, A.; Nielsen, P. Unique Properties of Purine/Pyrimidine Asymmetric PNA-DNA Duplexes: Differential Stabilization of PNA-DNA Duplexes by Purines in the PNA Strand. *Biophys. J.* **2006**, *90*, 1329–1337.
- (19) Marky, L.; Breslauer, K. Calculating Thermodynamic Data for Transitions of any Molecularity from Equilibrium Melting Curves. *Biopolymers* **1987**, *26*, 1601–1620.
- (20) Alenaizan, A.; Barnett, J.; Hud, N.; Sherrill, D.; Petrov, A. The proto-Nucleic Acid Builder: a software tool for constructing nucleic acid analogs. *Nucleic Acids Res.* **2020**, *49*, 79–89.
- (21) Yeh, J.; Pohl, E.; Truan, D.; He, W.; Sheldrick, G.; Du, S.; Achim, C. The Crystal Structure of Non-Modified and Bipyridine-Modified PNA Duplexes. *Chem.—Eur. J.* **2010**, *16*, 11867–11875.
- (22) Kiliszek, A.; Banaszak, K.; Dauter, D.; Rypniewski, R. The first crystal structures of RNA-PNA duplexes and a PNA-PNA duplex containing mismatches-toward anti-sense therapy against TREDs. *Nucleic Acids Res.* **2015**, *44*, 1937–1943.
- (23) He, W.; Hatcher, E.; Balaëff, A.; Beratan, D.; Gil, R.; Madrid, M.; Achim, C. Solution Structure of a Peptide Nucleic Acid Duplex from NMR Data: Features and Limitations. *J. Am. Chem. Soc.* **2008**, *130*, 13264–13273.
- (24) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhoe, G.; Mark, A.; Berendsen, H. GROMACS: Fast, Flexible and Free. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
- (25) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B.; Grubmüller, H.; MacKerell, A., Jr. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71–73.
- (26) Berendsen, H. J. C.; Postma, J. P. M.; Gunsteren, W. F.; Hermans, J. Interaction Models for Water in Relation to Protein Hydration. *Intermolecular Forces* **1981**, *14*, 331.
- (27) Bekker, H.; Dijkstra, E.; Renardus, M.; Berendsen, H. An Efficient, Box Shape Independent Non-Bonded Force and Virial Algorithm for Molecular Dynamics. *Mol. Simul.* **1995**, *14*, 137–151.
- (28) Hockney, R. W.; Goel, S. P.; Eastwood, J. W. Quiet High Resolution Computer Models of a Plasma. *J. Comput. Phys.* **1974**, *14*, 148–158.
- (29) Berendsen, H. J. C.; van Gunsteren, W. F. Practical algorithms for dynamics simulations. *Molecular-Dynamics Simulation of Statistical-Mechanical Systems*; Enrico Fermi Summer School, 1986; pp 43–65.
- (30) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
- (31) Hess, B.; Bekker, H.; Berendsen, H.; Fraaije, J. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- (32) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An  $N \log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (33) Essmann, U.; Perera, L.; Berkowitz, M.; Darden, T.; Lee, H.; Pedersen, L. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (34) Frigo, M.; Johnson, S. The Design and Implementation of FFTW3. *Proc. IEEE* **2005**, *93*, 216–231.
- (35) Goga, N.; Rzeplia, A.; de Vries, A.; Marrink, S.; Berendsen, H. Efficient Algorithms for Langevin and DPD Dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 3637–3649.
- (36) Berendsen, H.; Postma, J.; van Gunsteren, A.; DiNola, J.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (37) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- (38) Valdés-Tresanco, M. S.; Valdés-Tresanco, M. E.; Valiente, P.; Moreno, E. gmx\_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS. *J. Chem. Theory Comput.* **2021**, *17*, 6281–6291.
- (39) Still, C.; Tempczyk, A.; Hawley, R.; Hendrickson, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
- (40) Onufriev, A.; Case, D. Generalized Born Implicit Solvent Models for Biomolecules. *Annu. Rev. Biophys.* **2019**, *48*, 275–296.
- (41) Bashford, D.; Case, D. Generalized Born Models of Macromolecular Solvation Effects. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129–152.
- (42) Jawad, B.; Poudel, L.; Podgornik, R.; Ching, W. Thermodynamic Dissection of the Intercalation Binding Process of Doxorubicin to dsDNA with Implications of Ionic and Solvent Effects. *J. Phys. Chem. B* **2020**, *124*, 7803–7818.
- (43) Jawad, B.; Poudel, L.; Podgornik, R.; Steinmetz, N.; Ching, W. Molecular mechanism and binding free energy of doxorubicin intercalation in DNA. *Phys. Chem. Chem. Phys.* **2019**, *21*, 3877–3893.
- (44) Gonçalves, P.; Stassen, H. Calculation of the free energy of solvation from molecular dynamics simulations. *Pure Appl. Chem.* **2004**, *76*, 231–240.

(45) Otsuka, K.; Jinno, K.; Morikawa, A. Active and Selective Catalysts for the Synthesis of C<sub>2</sub>H<sub>4</sub> and C<sub>2</sub>H<sub>6</sub> via Oxidative and Coupling of Methane. *J. Catal.* **1986**, *100*, 353–359.

(46) Janežič, D.; Venable, R. M.; Brooks, B. R. Harmonic analysis of large systems. III. Comparison with molecular dynamics. *J. Comput. Chem.* **1995**, *16*, 1554–1566.

(47) Onufriev, A.; Bashford, D.; Case, D. Exploring Protein Native States and Large-Scale Conformational Changes With a Modified Generalized Born Model. *Proteins* **2004**, *55*, 383–394.

(48) General, I. A Note on the Standard State's Binding Free Energy. *J. Chem. Theory Comput.* **2010**, *6*, 2520–2524.

(49) SantaLucia, J., Jr.; Allawi, H.; Seneviratne, P. Improved Nearest-Neighbor Parameters for Predicting DNA Duplex Stability. *Biochemistry* **1996**, *35*, 3555–3562.

(50) Xia, T.; SantaLucia, J., Jr.; Burkard, M.; Kierzek, R.; Schroeder, S.; Jiao, X.; Cox, C.; Turner, D. Thermodynamic Parameters for an Expanded Nearest-Neighbor Model for Formation of RNA Duplexes with Watson–Crick Base Pairs. *Biochemistry* **1998**, *37*, 14719–14735.

(51) SantaLucia, J., Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Biochemistry* **1998**, *95*, 1460–1465.

### NOTE ADDED AFTER ASAP PUBLICATION

Due to a production error, the version of this paper that was published ASAP November 14, 2022, contained an error in Table 2. This error was corrected and the paper was reposted November 15, 2022.

## Recommended by ACS

### RNA Captures More Cations than DNA: Insights from Molecular Dynamics Simulations

Sergio Cruz-León and Nadine Schwierz

OCTOBER 19, 2022  
THE JOURNAL OF PHYSICAL CHEMISTRY B

READ 

### Prediction of Aptamer–Small-Molecule Interactions Using Metastable States from Multiple Independent Molecular Dynamics Simulations

Alan Fernando Rodríguez Serrano and I-Ming Hsing

SEPTEMBER 22, 2022  
JOURNAL OF CHEMICAL INFORMATION AND MODELING

READ 

### Molecular Dynamics Simulations of Protein RNA Complexes by Using an Advanced Electrostatic Model

Zhifeng Jing and Pengyu Ren

SEPTEMBER 15, 2022  
THE JOURNAL OF PHYSICAL CHEMISTRY B

READ 

### Force-Field-Dependent DNA Breathing Dynamics: A Case Study of Hoogsteen Base Pairing in A6-DNA

Sharon Emily Stone, Ioan Andricioaei, *et al.*

SEPTEMBER 01, 2022  
JOURNAL OF CHEMICAL INFORMATION AND MODELING

READ 

Get More Suggestions >