Comment

How much free energy is absorbed upon breaking DNA base pairs? Comment on "DNA melting and energetics of the double helix" by Maxim Frank-Kamenetskii et al.

John SantaLucia Jr.

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

Received 1 March 2018; accepted 5 March 2018

A fundamental question in biology is how much Gibbs free-energy needs to be absorbed (i.e. ΔG°_{37}) to break the Watson–Crick base pairs in a section of a double helix or to completely break all base pairs of a duplex. In this issue, Alexander Vologodskii and Maxim Frank-Kamenetskii provide a historical retrospective of this question and offer new insights into the quantification of DNA stability and deducing the origin of sequence dependence [1]. Such knowledge is critical to understanding the processes of opening the DNA double helix (i.e. melting) to form an internal loop (sometimes called a "bubble") and then translocating such internal loops during DNA replication, RNA transcription, and DNA repair. Other important applications of DNA melting involve the design of oligonucleotide primers and probes for disease diagnostics [2]. Paul Doty's group made initial observations regarding the thermal denaturation of DNA detected by UV absorbance and the dependence of the melting temperature, T_m, on G + C content [3]. Tinoco's group then applied the nearest-neighbor model to RNA duplex formation and single-stranded RNA folding [4] and refined nearest-neighbors were later published by Turner's Lab [5]. Advances were made in the application of statistical mechanics of DNA melting [6] that was used by a variety of groups to study the melting of polymeric double-stranded DNA [7–11]. There was also important theoretical and experimental work done on counterion screening of polyelectrolytes [12,13].

Over a period of 16 years (1981–1997), at least 11 different groups published nearest-neighbor parameters for predicting the ΔG°_{37} helix formation for DNA [8–10,14–22]. The parameters from 7 of those groups were analyzed in detail in 1998 [23]. Surprisingly, the parameters from the 7 different groups differed substantially in both the trends and magnitudes of the NN ΔG°_{37} contributions (Fig. 1). One group [9], led by Maxim Frank-Kamenetskii, however, reported essentially the correct answer for polymers in 1984. For more than a decade, this historical fact was not widely appreciated. It is fair to say that the discrepancies among the literature studies represented the confusion in the field at that time. My own group published a set of nearest-neighbor parameters in 1996 based on analysis of oligonucleotides [18]. Shortly after our publication, Sugimoto's group published an independent set of parameters based on analysis of oligonucleotides [19].

In 1997, my group combined the experimental data for 108 duplexes from throughout the literature [18,19] and performed singular value decomposition (SVD) to produce what we dubbed the "Unified" NN parameters [20,23]. Importantly, the unified NN parameters published [20] the standard errors (which none of the previous publications
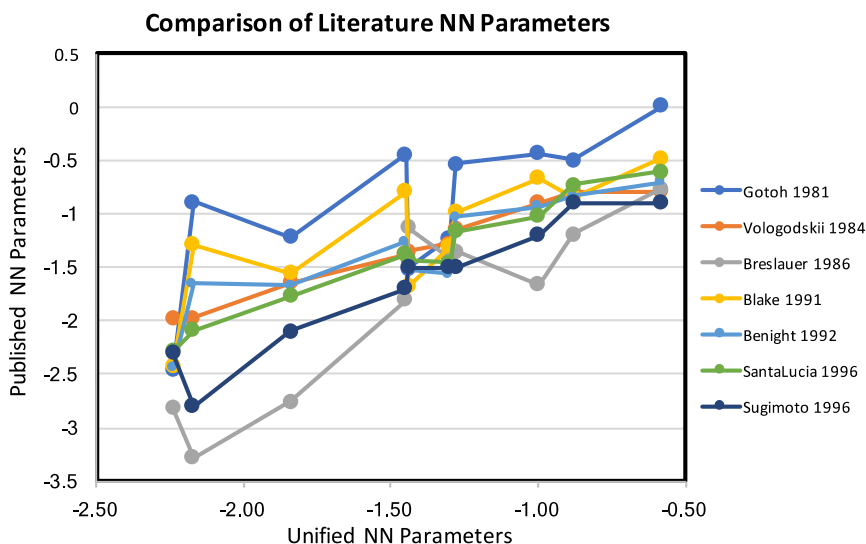
Comparison of Literature NN Parameters

Fig. 1. Comparison of literature nearest-neighbor ΔG_{37}° parameters (Y-axis) to the Unified NN parameters (X-axis). All parameters are given in kcal/mol. The order of the NN from left to right are: GC/CG (-2.2 kcal/mol), CG/GC, CC/GG, CA, AC, TC, CT, AA/TT, AT/TA, TA/AT (-0.58 kcal/mol). Data are from references [8–10,14,15,18,19] after recasting to ΔG°_{37} as described in [23]. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

had done) in the SVD derived NN parameters and showed that the NN parameters were precise to the second decimal point (e.g. standard errors in the mean ranging from 0.01 to 0.06 kcal/mol for AA/TT and TA/AT, respectively [20]). These small errors suggest that further measurements or even more accurate measurements are unlikely to substantially change the NN parameters. My group has subsequently compiled [17,20] a larger set of measurements for 317 oligonucleotide duplexes and derived refined NN parameters that agree with the published Unified values within 0.03 kcal/mol (J. SantaLucia, unpublished), indicating that indeed the Unified parameters have essentially converged to the limits of the NN model itself.

Based upon the structural similarity of the base pairs and backbones in DNA polymers and oligomers, it seems reasonable to expect that *a priori* there *should be* concordance in the different studies of duplex stability and sequence dependence thereof. Despite the unification of parameters for oligonucleotides, concordance was *not evident* for polymers [8–10]. Thus, the field was ripe for someone to figure out the origins of the discrepancies from the different studies. In addition to the very different lengths of oligonucleotides (4 to 20 bp) vs. polymers (i.e. thousands of bp), the different groups used different salt concentrations, some groups presented melting temperature stacking contributions, "T_{MN}" [8,10,17] rather than ΔG_{37}°, and the different groups used different numerical methods for determining the NN parameters (i.e. iterative numerical methods vs. analytical methods such as SVD). There was also some debate about the effect of end-fraying on the determination of NN propagation parameters. To address the issue of helix fraying, Benight's lab introduced the study of circular DNA dumbbells [15,16]. The use of dumbbells itself, however, introduced an additional unknown, namely the entropy of circle formation, which Benight dealt with properly. Thus, Benight's group also got a valid set of NN parameters, though their parameter for ΔG_{37}°(CG/GC) differed from the Unified value by more than 0.5 kcal/mol [23].

Thus, in 1998 I decided to perform a critical analysis of 7 literature NN sets from the field [23] (see Fig. 1). The work by Goldstein and Benight [16] was critical in revealing that the polymer studies were rank deficient and they suggested that the best method for determining the NN parameters was SVD. Thus, I decided to reanalyze all of the literature data sets using SVD. This immediately revealed that the polymer studies had a rank of 8, indicating that it is impossible to determine a unique set of 10 NN parameters using polymers alone. Frank-Kamenetskii's lab fully realized this in their 1984 publication [9] and they explicitly presented a set of 8 polymer invariants and suggested the assumptions of ΔG_{37}°(AT/TA) = ΔG_{37}°(TA/AT) and ΔG_{37}°(GC/CG) = ΔG_{37}°(CG/GC). The current publication by Vologodskii and Frank-Kamenetskii provides a set of 8 polymer invariants and explicitly gives the two constraints on polymer NN composition [1]. Interestingly, the 2 NN parameters that could be determined uniquely in the polymer