DNA Power: Love thy Neighbor

Tod A. Pascal, Dr. Nagarajan Vaidehi, Dr. William A. Goddard III

Material and Process Simulation Center, California Institute of Technology, Pasadena, CA 91125
City of Hope, Beckman Research Institute, Duarte, CA 91010

The stability of the DNA structures has been actively investigated since Watson and Crick proposed the double helix as the molecular structure: where A is hydrogen bonded with T and G is hydrogen bonded with C. A number of qualitative factors have been identified as contributing to this thermodynamic stability: base stacking interactions (van da Waals), Watson-Crick hydrogen bond interactions and interactions with water molecules and metal ions (electrostatics). Different levels of water and ion concentration lead to the three distinct forms of DNA found in nature: A-DNA, B-DNA and Z-DNA. B-DNA is the predominant form, commonly occurring in well solvated system and the presence of sodium and magnesium ions at relatively low concentrations (0.1 M).

Quantitative measures of DNA duplex stabilities have been attempted in order to explain everything from binding in protein-DNA complexes to DNA melting. Future applications such as DNA computers, DNA based nano-sensing arrays, and DNA as a component in molecular electronics demand a fundamental and quantitative understanding of DNA structure and the factors affecting it. To this end, a number of experiments have been performed, the most famous results of which was pioneered by the SantaLucia group. Using NMR spectroscopy, they determined the sequence and solution dependence of the DNA double helix; the data was show to closely fit the Nearest-Neighbor (NN) interaction model. This model assumes that the stability of a DNA duplex can be depends on the identity and orientation of neighboring base pairs. This approximation therefore reduces the free energy to a base-pair to it’s interaction with the neighboring bases. Ten different nearest-neighbor interactions are therefore possible in any Watson-Crick DNA duplex structure.

We present the first every theoretical study testing the nearest neighbor model purely from fully atomistic simulations. Using a set of 13 DNA double helices, of random sequence, and variable length, the 10 NN parameters was determined by single value decomposition (SVD) of the per-atom enthalpies of the entire set. We observe the same trends as the SantaLucia experimental enthalpies, namely: GC > CG > GT ~ AA > GA > CA > GG > AT > TA > CT. Further, there is a >99% correlation between the predicted helical enthalpies from our parameters and that of SantaLucia. More importantly, our parameters allow us to give a priori estimations of the simulation enthalpy of a random sequence of B-DNA to a confidence level 0.5%.

Figure 1. B-DNA Schema
Plot of Total Helix Energy for Santa Lucia vs WAG NN

$y = 15.017x - 125.62$

$R^2 = 0.9979$

Plot 1. Correlation between WAG NN and SantaLucia NN for set of 13 DNA Double Helices