Oscar Gonzalez, University of Texas, Austin, TX, USA
Daiva Petkeviciute and John Maddocks, EPFL, Lausanne, Switzerland

A Rigid-Base Model for DNA Structure Prediction

The sequence-dependent curvature and flexibility of DNA are critical for its packaging into the cell, recognition by other molecules, and conformational changes during biochemical processes. Is it possible to predict this curvature and flexibility from the basepair sequence? A model is presented which can predict these properties, and more specifically the relative position, orientation and coupling of every base in an arbitrary DNA oligomer provided it is close to the B-form structural family. The model is of the rigid-base type in which each individual base is treated as an independent rigid body. The model is founded upon a hierarchy of sequence-dependent local energies that describe physically distinct interactions, involves only local parameters, can capture important local features that are below the resolution of other coarse-grained models, and can also capture important non-local features as have been observed in various investigations. A novelty of the model is its ability to account for the intrinsic, pre-existing stress in an oligomer. A complete parameter set for double-stranded, B-form DNA, in standard environmental conditions, has been estimated using an extensive database of atomic-resolution, explicit-solvent MD data produced by a consortium of groups. In this talk, an overview of the model and various mathematical issues associated with its parameterization will be discussed, as well as various examples illustrating some features and limitations of the model.

For any given oligomer, our model delivers an internal energy and a Gaussian probability density function on the associated internal configuration space, where the degrees of freedom are the relative displacement and rotation of each rigid base on each strand of the oligomer. The internal energy model is based on a hierarchy of sequence-dependent local energies that describe physically distinct interactions between various groups of proximal bases. Consistent with a nearest-neighbor assumption, we consider only the first two members of the hierarchy that describe the local interactions between the two bases in a monomer and the four bases in a dimer. Moreover, we characterize these interactions by a finite set of parameters that depend only on the local monomer and dimer sequence. The internal energy of an arbitrary oligomer of any length is then defined by a construction rule in which the local interaction energies are superimposed. We show that an internal energy constructed in this way provides a natural model for the intrinsic curvature and flexibility of an oligomer. Indeed, we show that these properties are determined by the local parameters in a non-trivial way through the construction rule. Moreover, our internal energy also provides a natural model for the intrinsic pre-existing stress in an oligomer. This stress arises from the fact that each base cannot simultaneously minimize all its local interactions and must instead find a compromise. As a consequence, our model predicts that the intrinsic or ground-state curvature of an oligomer depends non-locally on its sequence. That is, local mutations of the sequence produce non-local changes in shape. The description of such non-local behavior using only local parameters is unique to our model and is consistent with recent observations in the literature.

A complete parameter set for our model was estimated using a maximum relative entropy approach on the space of normalized probability density functions. Specifically, we sought to minimize an objective functional defined as the sum of Kullback-Leibler divergences between the model and observed probability density functions for each oligomer in the MD database. The numerical treatment of this problem was complicated by various constraints on the set of admissible parameters: some of the parameters are symmetric, positive-semi-definite matrices of different sizes, and various independent superpositions of these matrices must be positive-definite. Through a detailed study of this system of constraints and the construction rule for the internal energy and hence probability density function for an oligomer, we constructed an analytical characterization for an approximate minimizer of the Kullback-Leibler objective functional. Using an initial guess based on this characterization, we were able to successfully minimize the objective functional using a constrained gradient flow procedure and thereby obtain a best-fit parameter set for our model. Various predictions using this parameter set have been compared with existing data on B-form DNA, both experimental and simulated, and both sequence-averaged and sequence-specific. Our comparisons show that the model predictions are consistent with accepted properties of B-form DNA, and that the model can successfully predict properties such as the non-local effects of single nucleotide permutations and the non-local context effects of various structural degrees of freedom.