MAR09-2008-004771

Abstract for an Invited Paper
for the MAR09 Meeting of
the American Physical Society

Statistical mechanics of chromatin: Inferring free energies of nucleosome formation from high-throughput data sets[1]

ALEXANDRE MOROZOV, Department of Physics & Astronomy, Rutgers University

Formation of nucleosome core particles is a first step towards packaging genomic DNA into chromosomes in living cells. Nucleosomes are formed by wrapping 147 base pairs of DNA around a spool of eight histone proteins. It is reasonable to assume that formation of single nucleosomes *in vitro* is determined by DNA sequence alone: it costs less elastic energy to wrap a flexible DNA polymer around the histone octamer, and more if the polymer is rigid. However, it is unclear to which extent this effect is important in living cells. Cells have evolved chromatin remodeling enzymes that expend ATP to actively reposition nucleosomes. In addition, nucleosome positioning on long DNA sequences is affected by steric exclusion - many nucleosomes have to form simultaneously without overlap. Currently available bioinformatics methods for predicting nucleosome positions are trained on *in vivo* data sets and are thus unable to distinguish between extrinsic and intrinsic nucleosome positioning signals. In order to see the relative importance of such signals for nucleosome positioning *in vivo*, we have developed a model based on a large collection of DNA sequences from nucleosomes reconstituted *in vitro* by salt dialysis. We have used these data to infer the free energy of nucleosome formation at each position along the genome. The method uses an exact result from the statistical mechanics of classical 1D fluids to infer the free energy landscape from nucleosome occupancy. We will discuss the degree to which *in vitro* nucleosome occupancy profiles are predictive of *in vivo* nucleosome positions, and will estimate how many nucleosomes are sequence-specific and how many are positioned purely by steric exclusion. Our approach to nucleosome energetics should be applicable across multiple organisms and genomic regions.

[1]AM acknowledges financial support from the BioMaPS Institute for Quantitative Biology.