

Deciphering the conformational code behind the indirect readout of DNA sequences

SY07-10

P.D. Dans^I, M. Orozco^I

^IInstitute for Research in Biomedicine (IRB Barcelona), Barcelona, Spain

The correct recognition of DNA sequences by proteins is crucial for important biological events as DNA replication, transcription, and regulation of gene function. Several cases of binding can be explained by a direct readout mechanism, where the proteins recognize a specific DNA sequence, interacting at the molecular level with specific nucleobases in specific positions. A second model has been suggested to explain protein-DNA recognition for those cases where the binding was confirmed but it was impossible to find a consensus sequence: The indirect readout mechanism. In this cases specific conformational features of the DNA, which are local in nature and sequence-dependent, give to DNA a particular shape recognized by the protein. The shape adopted reflects DNA's ability to exist in diverse conformational sub-states at the backbone and base level giving rise to structural polymorphisms.

Combining the new parmBSC1 force field and the latest knowledge in the area of DNA conformations in the helical space, we present the most complete and unified description, at the tetranucleotide level, of the different existing polymorphisms and their interconnections. For this purpose we ran microsecond-long Molecular Dynamics (MD) simulations on a newly designed an optimized sequence library containing all the unique 136 tetranucleotides. Using this dataset, and the previous μ ABC data (from the Ascona B-DNA Consortium), we carried out our analysis on more than 75 μ s of accumulated simulation time, involving a total of 52 different sequences, and at least 6 replicas of each tetranucleotide. Our results reconcile with unprecedented accuracy the major theoretical and experimental sources of knowledge on DNA structural polymorphisms (X-ray crystallography, H^1/P^{31} -NMR, data mining of databases, and MD simulations), and reveal the possible sub-state combinations that DNA could adopt and which ultimately conform the basic code needed for understanding the indirect readout mechanism.