DNA in chromatin : what can we learn from a multi-scale wavelet analysis of DNA sequences ?

A. Arneodo

Laboratoire Joliot Curie and Laboratoire de Physique, Ecole Normale Supérieure de Lyon,
46 Allée d'Italie, 69364 Lyon Cedex 07, France

Understanding how chromatin is spatially and dynamically organized in the nucleus of eukaryotic cells and how this affects genome functions is one of the main challenges of cell biology. Recent technical progress in live cell imaging have confirmed that the structure and dynamics of chromatin play an essential role in regulating many biological processes, such as gene activity, DNA replication, recombination and DNA damage repair. The emerging view is that genomes are compartimentalized at the level of chromosome territories in mammalian nuclei, into subchromosal structural domains that are likely to be fundamental functional units that coordinate the spatial organization and timing of replication and transcription. To which extent one may learn about the higher order structure and dynamics of chromatin directly from the primary DNA sequence and its functional landmarks, is a question of fundamental and practical importance.

In the first part of this talk, we use the space-scale decomposition provided by the continuous wavelet transform (WT) to characterize the scale invariance properties of genomic sequences. We show the existence of long-range correlations over distances up to 20 -30 kb as the signature of the nucleosomal structure of the 30 nm chromatin fiber. In a second part, we explore the large-scale compositional heterogeneity of several large (tens of megabases) contigs within human chromosomes through the optics of the WT microscope. We show that the GC content displays relaxational nonlinear oscillations with two main frequencies corresponding to 100 kb and 400 kb which are well recognized characteristic sizes of chromatin loops and loop domains involved in the hierarchical folding of the chromatin fiber. These frequencies are also remarkably similar to the size of mammalian replicons and replicon clusters. When further investigating deviations from intrastrand equimolarities between A and T and between G and C, we corroborate the existence of these two fundamental frequencies as the footprints of the replication and/or transcription mutation biais and we show that the observed nonlinear oscillations enlighten a remarkable cooperative organization of gene location and orientation. In the third part of this talk, we use the singularity tracking ability of the WT to develop a methodology to detect the origins of replication. We report the discovery of 486 putative origins of replications in the human genome (only 9 were

experimentally identified so far). We present a model of replication with well positioned replication origins and random terminations that accounts for the observed oscillatory behavior. These results strongly suggest that replication and gene expression are likely to be regulated by the structure and dynamics of the chromatin fiber.