Journal of Statistical Physics, Vol. 101, Nos. 1/2, 2000

Statistical Dynamics of Clustering in the Genome Structure

A. Provata[1] **and Y. Almirantis**[2]

Received November 24, 1999

Clustering and long-range correlations in the nucleotide sequences of different categories of organisms are studied. As a result of clustering, the size distribution of coding and non-coding DNA regions is estimated analytically using the Generalised Central Limit Theorem.

The alternation of coding regions (which follow a short range size distribution) with non-coding regions (which follow a long range size distribution in higher organisms) leads to DNA structures which have a striking resemblance to random Cantor Fractals. For lower organisms (such as viruses, procaryotes etc.) long-range correlations are sporadically observed and the DNA structures do not present fractality.

Statistical models are proposed based on biologically motivated dynamical mechanisms (such as aggregation of oligonucleotides, influx and DNA length reduction), which can account for the above statistical features.

REFERENCES

C. K. Peng, S. V. Buldyrev, A. L. Goldberger, S. Havlin, F. Sciortino, M. Simons and H. E. Stanley, *Nature* **356**:168 (1992).

W. Li and K. Kaneko, *Europhys. Lett.* **17**:655 (1992).

Y. Almirantis and A. Provata, *J. Stat. Phys.* **97**:233 (1999).

A. Provata and Y. Almirantis, *Fractals* **8**:15 (2000).

---

[1] Institute of Physical Chemistry, NRCPS Demokritos, 15310 Athens, Greece.

[2] Institute of Biology, NRCPS Demokritos, 15310 Athens, Greece.

0022-4715/00/1000-0709$18.00/0