The nucleotide sequence and the local electronic structure

2007 J. Phys.: Condens. Matter 19 181001

(http://iopscience.iop.org/0953-8984/19/18/181001)

IOP PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **19** (2007) 181001 (2pp) doi:10.1088/0953-8984/19/18/181001

VIEWPOINT

The nucleotide sequence and the local electronic structure

Nestor Norio Oiwa

Núcleo de Neurociências e Comportamento, Departamento de Psicologia Experimental, Instituto de Psicologia, Universidade de São Paulo, Brazil
and
Departamento de Física Geral, Instituto de Física, Universidade de São Paulo, Brazil

Received 7 March 2007
Published 11 April 2007
Online at stacks.iop.org/JPhysCM/19/181001

At present, GenBank contains more than one hundred billion nucleotide sequences [1]. Despite the huge amount of data, understanding of the DNA interaction with the molecular machine of the cell is still limited.

Since Li's critical review in 1997 [2], progress has been made in order to link the hidden structure of the genome with the biochemical function in the cell. The most popular is the ratio of the cytosine and guanine (CG) content, where we count the number of CGs in a segment of the DNA chain. We can demonstrate the existence of the mosaic structure of the genome, because there are long domains with constant CG ratio, called isochors [3, 4]. Although popular, the relation of the isochors with biochemical function is unclear [2]. Other suggestions for the DNA landscape linking the genomic content with some local property have been proposed, most of them based on scores and energetic aspects of the nucleotides [5–8]. They localize the genomic biochemical features, but the criteria for such local properties lack realism, because the problem is oversimplified. They do not explain the changes of the electronic surface of the DNA chain with the nucleotide chemical composition, for example. However, studying the cellular metabolism, we have some clues to describe the local properties of genome.

We know that proteins spot very specific genomic sites without unzipping the double helix [9]. The usual mechanisms for DNA transcription in the textbook are energy cost processes, because they open the double strain. This means that we need to break the hydrogen bond between the nucleotides in opposite chains, demanding adenosine 5′-triphosphate (ATP). So, we expect the existence of a more subtle way for DNA reading. In fact, the protein works with other molecules recognizing the local charge distribution, an energy saving way to interact with the neighbour. Despite evidence of the protein–DNA interactions [9], the mechanism describing how the nucleotide sequences change the electron local distribution was unknown.

In a recent *Journal of Physics Condensed Matter* paper, Zhu *et al* apply the Peyrard–Bishop–Holstein theory in the P5 viral promoter nucleotide sequence, computing the electron local density of states [10]. The authors describe the quantum charge behaviour based on the Peyrard and Bishop portrayal of the electronic degree of freedom along the DNA chain and the transverse stretching for the hydrogen bonds between the nucleotides [11] as well as the Holstein's base pair (local) electronic stretching description [12, 13]. Then, they report

the differences of the spectrum of energy level and local density of states (LDOS) for the electrons of the homopolymer $(A)_{68}$ and the P5 promoter. Promoters are nucleotide sequences responsible for the transcription regulation. Specialized proteins spot such genomic sites without unzipping the double helix. Then, these proteins attach themselves to these particular nucleotide sequences starting the biochemical pathway for the DNA transcription. In particular, the P5 promoter is a 68-nucleotide sequence including TATA and CGCC, both common in the human genome, and, the adeno-associated virus P5 is a 4500 base pairs long genome, used as vector in gene therapy.

The approach adopted by Zhu *et al* will allow us to understand the LDOS of many biologically relevant nucleotide sequences: protein binding sites, promoters and repressors, etc. In addition, we may study damage in the DNA coding by studying the thymine dimers which occur due to the chromosomal exposure to ultraviolet irradiation, for example.

Acknowledgments

This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo, Brazil, under grant number 2005/59668-4. The author thanks Dora Fix Ventura for a critical reading of the manuscript.

References

[1] Benson D A, Karsch-Mizrachi I, Lipman D J, Ostell J and Wheeler D L 2007 *Nucleic. Acids Res.* **35** D21
[2] Li W 1997 *Comput. Chem.* **21** 257
[3] Bernardi G, Olofsson B, Filipski J, Zerial M, Salinas J, Cuny G, Meunier-Rotival M and Rodier F 1985 *Science* **228** 953
[4] Churchill G A 1989 *Bull. Math. Biol.* **51** 79
[5] Peng C-K, Buldyrev S V, Goldberger A L, Havlin S, Sciortino F, Simons M and Stanley H E 1992 *Nature* **356** 168
[6] Bussemaker H J, Li H and Siggia E D 2000 *Proc. Natl Acad. Sci. USA* **97** 10096
[7] Audit B, Thermes C, Vaillant C, d'Aubenton-Carafa Y, Muzy J F and Arneodo A 2001 *Phys. Rev. Lett.* **86** 2471
[8] Oiwa N N and Glazier J A 2004 *Cell Biochem. Biophys.* **41** 41
[9] Watson J D, Hopkins N H, Roberts J W, Steiz J A and Weiner A M 1987 *Molecular Biology of The Gene* (Menlo Park, CA: Benjamin/Cummings)
[10] Zhu J-X, Rasmussen K O, Balatsky A V and Bishop A R 2007 *J. Phys.: Condens. Matter* **19** 136203
[11] Peyrard M and Bishop A R 1989 *Phys. Rev. Lett.* **62** 2755
[12] Holstein T 1959 *Ann. Phys.* **8** 325
[13] Turkevich L A and Holstein T D 1987 *Phys. Rev.* B **35** 7474