Minireview

New approaches to the mapping of chromosomal domains*

Jan Filipski[a], Ekaterina Svetlova[a], Nathalie Avril-Fournout[a], Patric Deschavanne[a] and Michel Bellis[b]

[a]*Institut J. Monod, 2 place Jusieu tour 43, 75251, Paris,* [b]*Institut de Biologie UPR 9008 CNRS, U294 INSERM, Montpellier, France*

Key words: MAR; SAR; LCR; chromosomal domains; four-stranded DNA; GC-rich streches

Although it is generally accepted that the chromosome is divided into elementary subunits, the structural and functional domains, the organisation of these structures at the molecular level is not well understood. In particular, the domain boundaries are not easily identifiable. Several possible candidates such as MARs/SARs, insulators, LCRs, palindromic sequences, or easily melting sequences have been found in the regions having properties one would except for boundaries. None of these elements, however, has been found in all of the constructs functioning as boundaries in tests *in vivo*. Recent work suggests that the common denominator might be the presence of GC-rich oligonucleotide streches and the formation of the chromatin hypersensitive sites. A model is discussed in which "unusual" structures, in particular the four-stranded DNA sequence elements containing unpaired bases, play the role of domain boundaries.

DNA in eucaryotic chromosomes is organised into structural and functional domains several dozens kilobases long. The structural chromosomal domains have been identified as DNA loops observed by electron microscopy upon extraction of the soluble chromosomal proteins [1]. Boundaries of these loops contain the Matrix Associated Regions or Scaffold-Attachment Regions (MARs or SARs, respectively [2, 3]. The transcriptional domains of tissue specific genes consist of segments of DNA carrying one or more genes and their regulatory sequences. The copy number-related and developmentally regulated expression of genes in these domains depends on several kilobases long

Locus Control Regions (LCR) [4]. An LCR contains the sequence elements responsible for insulation of the domain from the influence of the neighbour domains [5] and can be considered its boundary. The boundaries of transcriptional domains of housekeeping genes are probably located within the genomic regions of high GC content called CpG islands (the CpG dinucleotides are unusually abundant in these regions [6]). Replication of eucaryotic DNA occurs also by domains, the replicons, containing one or several origins of replication and flanked by zones of arrest of replication [7]. The existence of DNA repair domains has also been suggested [8] but boundaries of these domains in

---

Abbreviations: MAR, matrix associated region; SAR, scaffold-attachement region; LCR, locus control region; HSS, hypersensitive sites; YAC, yeast artificial chromosome.

the eucaryotic organisms are unknown as opposed to *E. coli.* The structure and function of the boundaries of repair domains have been well characterised in this organism [9]. Eucaryotic genomes may also be considered to be mosaics of recombinational domains, with the hot spots of recombination representing their boundaries. The structure of some of these hot spots has been studied in detail in yeast [10, 11].

The attempts of identification and characterisation of the boundaries of chromosomal domains involved identification of the regions forming strong complexes with nuclear matrix proteins. These regions are supposedly located at the base of the DNA loops. The localisation experiments produce usually complex patterns of DNA regions interactiong with the matrix. Some of these regions are not related to the anchorage of the DNA loops but rather represent the DNA undergoing replication or transcription. To avoid these artefacts new techniques of mapping are being developed in several laboratories, based on the properties of the domain boundaries other than the DNA protein complex formation.

Treatment of mammalian cells or nuclei with topoisomerase II inhibitors or with agents stimulating endogenous nucleases brings about fragmentation of DNA into large molecules approximately 50 kb long [12]. DNA molecules of similar size are also released from the nuclei at the early stages of apoptosis. These findings suggested that (a) strong toposites and hypersensitive sites (HSS) are relatively regularly distributed in the chromatin fibre and (b) the spacing between these sites is in the range of chromosomal domains. Razin and collaborators [13, 14] used this property to map the hypersensitive sites and toposites in multicopy genes present in mammalian cells in the form of several hundred kilobases long repetitive DNA regions. They have mapped HSS, toposites and nuclease S1 sensitive sites in the region carrying ribosomal genes and in the c-myc amplicon. Typically they obtained in these experiments (which involved pulse field gel electrophoresis and indirect end-labelling) ladders of bands corresponding to DNA loops and to their multimers excised at the matrix attachment regions. This technique has been recently successfully applied to the mapping of the chromosomal loops in a 800 kb long, single copy region in the genome of *Drosophila mela-* *nogaster* [15]. Sensitivity of this technique is, however, not high enough to allow the long range mapping of HSS or toposites in single copy regions in mammalian cell nuclei.

In order to overcome this problem we determined the hypersensitive sites in the chromatin assembled *in vivo* on segments of human DNA cloned as yeast artificial chromosomes (YACs). The HSS observed in the DNA segments studied were distributed with an average spacing of about 35 kb [16]. Separately we mapped the CGCG tetranucleotides in the same YACs using a restriction enzyme. The CGCG tetranucleotides are relatively rare in mammalian genomes except for the region rich in GC content. They are particularly overrepresented in CpG islands [17]. It appeared that the CGCG clusters are also distributed relatively regularly within the chromosomal segments studied [18]. Most of them were coincident with the HSSs in the studied region of the human chromosome.

The coincidence of the pattern of CGCG with the pattern of hypersensitivity suggests that in vertebrates, similarly as in bacteria [19], the long range compositional pattern of DNA composition correlates with functional features of genomes. In addition, it suggests that the boundaries of the chromosomal domains might be located within the GC rich genomic regions. These regions are characterised by the presence of G-rich oligonucleotides which are able to form "unusual" DNA secondary structures, in particular the four-stranded structures (4G) [20–22]. The presence of such structures has been proposed to explain the properties of eucaryotic telomers [23]. It has also been suggested that the CG-rich DNA segments containing the sequence elements which are able to form the 4G structures are distributed periodically along the eucaryotic chromosomes. Such a distribution may facilitate recognition and alignment of the homologous chromosomes during meiosis [24]. These elements might form the domain boundaries also in mitotic chromosomes (see Fig. 1). There are several observations suggesting such a possibility. The sequence elements capable of forming the four stranded structures are involved in the maintenance of CpG dinucleotides in unmethylated state in a CpG island [25]. Similar sequence elements are responsible for arrest of DNA replication [26] and have been found in the hot spots of recombination [27]. The GC-
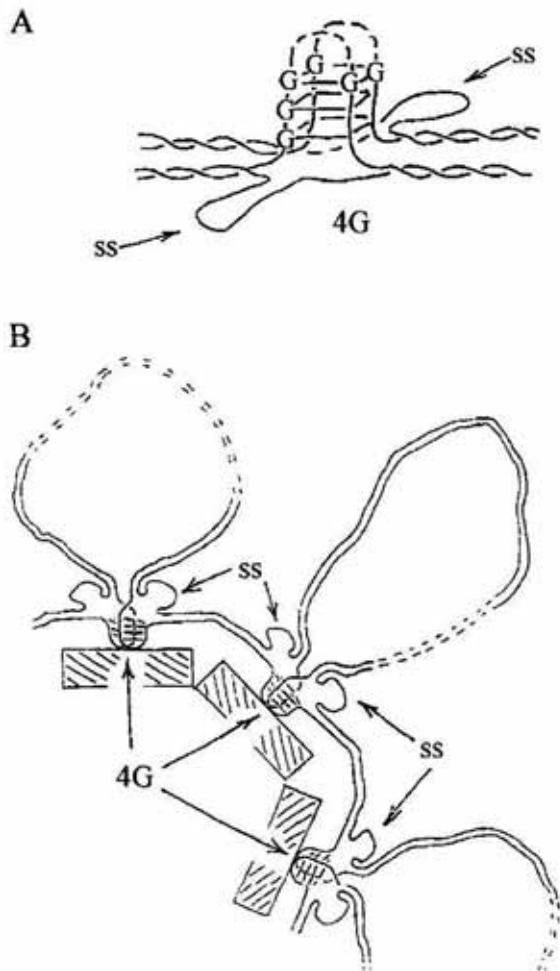
A

B

Fig. 1. *Model of chromosomal higher order structures with the four-stranded (4G) sequence elements as boundaries of the chromosomal domains.*
A, Example of a 4G structure; B, segment of a chromosomal rosette; ss, DNA single-stranded regions; 4G, four stranded structures at the domain boundaries.

rich sequence elements are recognised by topoisomerase II in the satellite III [28], react with specific protein in yeast [29], and have an affinity to the matrix [30]. The four-stranded structures are accompanied by single-stranded portions of DNA which are hypersensitive to DNase I and to nuclease S1. The four-stranded structures fulfill thus requirements for the domain boundaries and are good candidates for organising DNA into metaphase chromosomes.

REFERENCES

1. Paulson, J.R. & Laemmli, U.K. (1977) The structure of histone-depleted metaphase chromosomes. *Cell* **12**, 817–827.

2. Gasser, S.M. & Laemmli, U.K. (1977) A glimps at chromosomal order. *Trends Genetics* **3**, 16–21.

3. Cockerill, P.N. & Garrard, W.T. (1986) Chromosomal loop anchorage sites appear to be evolutionary conserved. *FEBS Lett.* **204**, 5–7.

4. Collis, P., Antoniou, M. & Grosveld, F. (1990) Definition of the minimal requirements within the human beta-globin gene and the dominant control region for high level expression. *EMBO J.* **9**, 233–240.

5. Chung, J.H., Whiteley, M. & Felsenfeld, G. (1993) A 5' element of the chicken beta-globin domains serves as an insulator in human erythroid cells and protects against position effect in *Drosophila*. *Cell* **74**, 505–514.

6. Bird, A., Taggart, M., Frommer, M., Miller, O.J. & Macleod, D. (1985) A fraction of the mouse genome that is delivered from islands of nonmethylated CpG-rich DNA. *Cell* **40**, 91–99.

7. Little, R.D., Platt, T.H.K. & Schildkraut, C.L. (1993) Initiation and termination of DNA replication in human rRNA genes. *Mol. Cell. Biol.* **13**, 6600–6613.

8. Bohr, V.A., Phillips, D.H. & Hanawald, P.C. (1987) Heterogeneous DNA damage and repair in the mammalian genome. *Cancer Res.* **47**, 6426–6436.

9. Radman, M. & Wagner, R. (1986) Mismatch repair in *Escherichia coli*. *Annu. Rev. Genet.* **20**, 523–538.

10. Liu, J., Wu, T. & Lichten, M. (1995) The location and structure of double-strand breaks induced during yeast meiosis: Evidence for a covalently linked DNA-protein intermediate. *EMBO J.* **14**, 4599–4608.

11. Ohta, K., Shibata, T. & Nicolas, A. (1994) Changes in chromatin structure at recombination initiation sites during yeast meiosis. *EMBO J.* **13**, 5754–5763.

12. Filipski, J., Leblanc, J., Youdale, T., Sikorska, M. & Walker, P.R. (1990) Periodicity of DNA folding in higher order chromatin structures. *EMBO J.* **9**, 1319–1327.

13. Razin, S.V., Hancock, R., Iarovaya, O., Westergaard, O., Gromova, I. & Georgiev, G.P. (1993) Structural-functional organization of chromosomal DNA domains. *Cold Spring Harbor Symp. Quant. Biol.* **58**, 25–35.

14. Gromova, I.I., Thompsen, B. & Razin, S.V. (1995) Loop-size spacing between CGCG clusters in long segments of human DNA. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 102–106.

15. Iarovaia, O., Hancock, R., Lagarkova, M., Miassod, R. & Razin, S.V. (1996) Mapping of genomic DNA loop organization in a

500-kilobase region of Drosophila X chromosome using the topoisomerase II-mediated DNA loop excission protocol. *Mol. Cell. Biol.* **16**, 302–308.

16. Filipski, J. & Bellis, M. (1993) Nuclease hypersensitive sites in yeast artificial chromosomes. *Int. Congress of Genetics,* Birmingham. Abst. p. 185.

17. Filipski, J. Salinas, J. & Rodier, F. (1987) Two distinct compositional classes of vertebrate gene-bearing streches, their structures and possible evolutionary origin. *DNA* **6**, 109–118.

18. Avril, N., Deschavanne, P., Bellis, M. & Filipski, J. (1995) Loop-size spacing between CGCG clusters in long segments of human DNA. *Biochem. Biophys. Res. Commun.* **213**, 147–153.

19. Deschavanne, P. & Filipski, J. (1995) Correlation of GC content with replication timing and repair mechanisms in weakly expressed *E. coli* genes. *Nucleic Acids Res.* **23**, 1350–1353.

20. Sen, D. & Gilbert, W. (1988) A sodium--potassium switch in the formation of four-stranded G4-DNA. *Nature (London)* **334**, 364–366.

21. Borisova, O.F., Shcholkina, A.K., Timofeev, E.N. & Florentiev, V.L. (1992) Evidence for the tetraplex structure of the $d(GT)_n$ repetitive sequences in solution. *FEBS Lett.* **306**, 140–142.

22. Fry, M. & Loeb, L.A. (1994) The fragile X syndrom $d(CGG)_n$ nucleotide repeats form a stable tetrahelical structures. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 4950–4954.

23. Murchie, A.I.H. & Lilley, D.M.J. (1994) Tetraplex folding of telomere sequence and inclusion of adenine bases. *EMBO J.* **13**, 993–1001.

24. Sen, D. & Gilbert, W. (1988) Formation of parallel four stranded complexes by guanine-rich motifs in DNA and its implication for meiosis. *Nature (London)* **334**, 364–366.

25. Macleod, D., Charlton, J., Mullins, J. & Bird, A.P. (1994) Sp1 sites in the mouse *aprt* promoter are required to prevent methylation of the CpG island. *Genes & Devel.* **8**, 2282–2292.

26. Woodford, K.J., Howell, R.M. & Usdin, K. (1994) A novel K(+)-dependent DNA synthesis arrest site in a commonly occuring sequence motif in eukaryotes. *J. Biol. Chem.* **269**, 27029–27035.

27. Shimizu, A. & Honjo, T. (1984) Immonoglobulin class switching. *Cell* **36**, 801–803.

28. Käs, E. & Laemmli, U.K. (1992) *In vivo* topoisomerase II cleavage of the *Drosophila* histone and satellite III repeats: DNA sequence and structural characterization. *EMBO J.* **11**, 705–716.

29. Liu, Z., Frantz, D.J., Gilbert, W. & Tye, B.-K. (1993) Identification and characterization of a nuclease activity specific for the G4 tetrastrand DNA. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 3157–3161.

30. de Lange, T. (1992) Human telomerases are attached to the nuclear matrix. *EMBO J.* **11**, 717–724.