bioRxiv preprint doi: https://doi.org/10.1101/096487; this version posted May 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

10.05.2017

Chromatin domain formation: a new model

Giorgio Bernardi

gbernardi@uniroma3.it

Science Department, Roma Tre University, Viale Marconi 446, 00146 Rome, Italy, and

Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Naples, Italy.

bioRxiv preprint doi: https://doi.org/10.1101/096487; this version posted May 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

The mechanisms of formation of LADs, the lamina associated domains, and TADs, the topologically associating domains of mammalian chromatin, were investigated here by using as a starting point the observation that chromatin architecture relies on an isochore framework and by doing a new analysis of both isochore structure and the isochore/chromatin domain connection. This approach showed that LADs correspond to isochores from the very GC-poor, compositionally homogeneous L1 family and from the "low-heterogeneity" L2 (or $L2^-$) sub-family; in fact, LADs are compositionally flat, flexible chromatin structures (because of the local wider nucleosome spacing associated with the high frequency of oligo-A's) that attach themselves to the nuclear lamina in self-interacting clusters. In contrast, TADs correspond to the increasingly GC-richer isochores from the "high-heterogeneity" L2 (or $L2^+$) sub-family and from the H1, H2 and H3 families. These isochores, making the framework of the individual chromatin loops or of the chromatin loop ensembles of TADs, were found to consist of single or multiple GC peaks. The self-interacting single or multiple loops of TADs appear to be shaped by the property that accompany the increasing levels of GC and CpG islands in their isochore peak backbones, namely by an increasing bendability (and increasing nuclease accessibility) due to decreasing nucleosome density and decreasing supercoiling. In conclusion, chromatin architecture appears not only to be encoded but also to be molded by isochores, the DNA units of genome organization. This "isochore molding" model of chromatin domains (both LADs and TADs) is essentially based on the differential binding of nucleosomes by isochore sequences and on the role of architectural proteins in closing TAD loops and ensuring TAD insulation.

bioRxiv preprint doi: https://doi.org/10.1101/096487; this version posted May 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Isochores, the long (>200 Kb), compositionally "fairly homogeneous" DNA sequences, were discovered (1) during an effort to better understand the large-scale organization of the mammalian genome (2). This organization had been revealed by compositional genomics, a strategy based on the frequencies of short sequences (3; or GC levels as a proxy) which led to a number of biologically important insights (4-6).

As far as the correlation of isochore composition with chromatin structure is concerned, it was proposed, more than thirty years ago, that "the different GC levels of isochores, their different ratios of CpG to GpC, and the accompanying differences in potential methylation sites are bound to be associated with differences in DNA and chromatin structure and, possibly, with differences in the regulation of gene expression" (4). Many significant advances along this line were made since, for instance by finding correlations between isochore composition and the structural/functional properties of DNA and chromatin (6,7).

In interphase nuclei, chromatin comprises two sets of domains that are largely conserved in mammals: LADs, the lamina associated domains (~0.5Mb median size), that are scattered over all chromosomes and correspond to GC-poor sequences (8-10), and TADs, the topologically associating domains (0.2-2 Mb in size), a system of GC-rich loops (11-14); many TADs can be resolved into contact domains (0.185 Mb median size; 13). Very recently, a comparison of maps of chromatin domains with maps of isochores from mouse and human chromosomes showed that isochores are the genomic units that underlie chromatin domains (15).

In spite of the recent, impressive advances in our understanding of chromatin structure (see refs. 16-22, for reviews), the problem of the formation mechanism(s) of LADs and TADs is still unsolved. Interesting models have been proposed for TADs (23-32), but no satisfactory solution has been reached so far. Three models will be considered here: 1) in the "handcuff model" (32), the

bioRxiv preprint doi: https://doi.org/10.1101/096487; this version posted May 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

two ends of a loop are brought together by architectural proteins such as the sequence-specific, zinc-finger insulation protein CTCF (the CCCTC-binding factor) that bind to each loop boundary and recruit the cohesin complex; 2) in the "extrusion model" (24, 26-28), a loop is generated dynamically by a pair of tethered CTCF units that attach to the chromatin fiber and travel along in opposite directions until they reach converging CTCF-binding sites; 3) in the "insulation-attraction model" (17) the insulation observed at TAD boundaries is visualized as resulting from stiffness of the chromatin fiber caused by functional elements (CTCF binding sites, highly active transcription starting sites etc) associated with increased nucleosome density, while specific local chromatin interactions are due to "attractive forces" within the domain. The problems of each one of these models will be discussed in the final section of this article.

Here, the issue of chromatin domain formation was approached by taking into account the observation that isochores make up the framework of chromatin domains (15) and by having a new look at isochore structure and at the isochore/chromatin domain connection. This approach was applied to human chromosome 21 which was chosen because this chromosome 1) is a good representative of human chromosomes, allowing an extension of the results to all other chromosomes (as investigated in ref. 15); and 2) is the smallest human chromosome, allowing a more expanded graphical presentation of data.

As far as nomenclature is concerned, although TADs comprise, by definition, all the topologically associating domains, in the context of this article TADs will indicate the chromatin domains other than LADs. The main reason for this choice is that the mechanisms of formation of the two sets of domains are different, even if based on the same fundamental DNA property, differential nucleosome binding.

bioRxiv preprint doi: https://doi.org/10.1101/096487; this version posted May 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Materials and methods

The sources of the data presented in the panels of Fig. 1 A, B, C, E, are given in the Figure legend. Fig. 1D and 1F are point-by-point plots of the 100Kb sliding window profiles of Fig. 1B and Fig. 1E, respectively. Fig. 2 is just an enlarged view of Fig. 1D. The panels of Fig. 3 and 4 (except for Fig. 4D) are from ref. 15, the vertical lines showing, however, new correlations (see ref. 15 for additional information).

Results

Isochore structure: a new analysis

Figure 1A shows the compositional profile of the DNA sequence from human chromosome 21 as seen through non-overlapping 100Kb windows. This window size was used because 100Kb is a plateau value under which the composition of DNA segments shows an increasing variance with decreasing size (33) due to several factors (for instance, the distribution of interspersed repeated sequences).

Figures 1B and 1C display the isochore profile as obtained from the chromosome 21 sequence using either a sliding window approach (34,35; Fig. 1B) or a fixed window approach (33,36; Fig. 1C). Both approaches flatten the compositional profiles by averaging, in two different ways, the fluctuating values of the large regions characterized by “fairly homogeneous” composition, the isochores. In the case of the sliding window approach, remnants of the fluctuations can still be seen as spikes in GC-rich regions, whereas in the case of the fixed window approach the fluctuations disappear because of the strict averaging procedure applied.

A new, simpler, in fact elementary, approach was used here, namely plotting (see Fig. 1D) the GC values of 100Kb DNA blocks of Fig. 1B as individual points. This approach was suggested by two recent results: 1) correlations hold between isochore properties and the

bioRxiv preprint doi: https://doi.org/10.1101/096487; this version posted May 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

properties of chromatin domains (7); and, more precisely, 2) the framework of TADs and LADs is made up by GC-rich and GC-poor isochores, respectively (15). One may therefore imagine a possible topological similarity between the flat structure of LADs and the loops of TADs on the one hand, and the compositionally flat GC-poor and the compositionally heterogeneous GC-rich isochores, respectively, on the other. If such is the case, a simple 100Kb point-by-point profile of GC levels is clearly preferable not only to both sliding and fixed window approaches, that flatten the compositional profile, but also to the color bar plot of 100Kb DNA segments (Fig. 1A) for reasons of graphical clarity.

The point-by-point profile (see Fig. 1D) expectedly showed the compositionally flat region 2 and the H1 and L2 peaks (**a** to **f**) of regions 1 and 3, that were already evident in Figs. 1A, 1B and 1C. It also led, however, to the discovery that the sequences of isochores from H1 (region 4), H2 (region 5) and H3 (region 6) isochores were not simply fluctuating within the compositional borders of the corresponding families (see Supplementary Table S1), but consisted, in fact, of sets of GC peaks. Upon close inspection, these very evident peaks can be seen to correspond to the minute peaks of Fig. 1B, that were flattened by the sliding window approach. Expectedly, the peaks of the point-by-point plots covered a broader GC range compared with the flattened peaks of the sliding window approach, as shown by comparing Fig. 1D with Fig. 1B, and Fig. 1F with Fig. 1E, which displays the high resolution compositional profiles of a multi-peak H2 isochore from chromosome 20.

In purely compositional terms (see Fig. 2 for a larger-scale presentation of the data of Fig. 1D), three different situations were found: 1a) a series of single peaks (regions 1 and 3) corresponding to an H1 isochore (**a**), and to several L2 isochores (**b** to **f**), in which latter case very few points were slightly beyond the “fixed” isochore family borders, but still within the “extended” borders of Supplementary Table S1; 1b) several very sharp H3 single peaks (region

bioRxiv preprint doi: https://doi.org/10.1101/096487; this version posted May 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

6), in which case an overall GC range of 18% was reached; 2) a homogeneous L1 isochore (region 2), in which the overall GC range barely reached 4% and all points were within the "fixed" GC borders of isochore family L1; and 3) two series of GC-rich multi-peak isochores that belonged to H1 (region 4) and H2 (region 5) families in which, again, very few points were slightly beyond the "fixed" isochore family borders. The striking difference between the compositional profile of L1 and H3 isochores is shown in Fig. 2A. Expectedly, when using a higher resolution windows (50Kb; see Supplementary Figure S1), the compositional profiles of isochore peaks became broader in the GC level gradients and more complex, because of the presence of interspersed repeated sequences and CpG islands (see ref. 33).

Isochores and LADs

It is well established (see refs. 8-10) that LADs 1) may cover ~35% of the human genome; 2) comprise 1,100-1,400 discrete domains demarcated by CTCF sites and CpG islands; 3) have a median size of ~0.5Mb; 4) are scattered over all chromosomes; 5) can be subdivided into cLADs, *i.e.*, "constitutive" LADs present in the four cell types originally tested and fLADs "facultative" LADs, only present in some cell types (in fact, only ~15% of the genome is involved in "stable contacts" present in most cells); 6) are characterized, in the case of cLADs, by conserved positions in syntenic regions of human and mouse chromosomes; 7) show a correspondence of cLADs and ciLADs (the "constitutive inter-LADs") with GC-poor and GC-rich isochores, respectively.

As shown in Figs. 3A (and 4A,4B), the major LAD of chromosome 21 corresponds to a large L1 isochore (which, incidentally, includes an exceptional GC-poor interLAD; see also the interval in the self-interacting domains corresponding to the L1 isochore of Fig. 4C). The other LADs correspond to the L1 isochores that separate the L2 peaks (to be described below and in

bioRxiv preprint doi: https://doi.org/10.1101/096487; this version posted May 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the following section), and to a "valley" L2 isochore comprised between two H1 isochores (on the right side of Fig. 3A). Moreover, two LADs flank the centromere; in fact, this appears to be the rule for all human chromosomes, as judged by looking at the results of ref. 15.

In chromosome 20 (Fig. 3B), the largest LAD corresponds to an L2 isochore (interrupted by an interLAD) while several other LADs correspond to L2 valley isochores flanked by H1 isochores; among faint LADs, one (extreme right) corresponds to an H1 isochore comprised between two H2 isochores and two other ones flank the centromere. In the very GC-rich chromosome 19 (Fig. 3C), two LADs correspond to two H1 isochores flanking an H2 isochore and two other LADs correspond to L2 isochores flanking an H1 isochore; finally, two faint LADs flank the centromere.

These results show that LADs correspond not only to L1 isochores that represent ~19% of the genome (incidentally, not too far from the ~15% involved in "stable contacts" that are extremely gene-poor; see ref. 10), but also to L2 isochores and even to H1 isochores in the rare case of very GC-rich chromosomes (*e.g.*, chromosome 19) and chromosomal regions (*e.g.*, the telomeric end of the short arm of chromosome 1).

As far as L2 isochores are concerned, it appears (see Supplementary Table 1 and Fig. 1) that 1) some isochores belong to a "low-heterogeneity" L2 sub-family that may be called $L2^-$, show a flat profile (see, for example, the largest LAD of chromosome 20); and 2) some other isochores belong to a "high-heterogeneity" L2, or $L2^+$, sub-family that are higher in average GC and are in the shape of single peaks (see Figs. 1 A-D and 3A). Now, as shown in Fig. 3, $L2^-$ isochores correspond to LADs, whereas $L2^+$ isochores correspond to interLADs and TADs (see the following section). The remaining L2 isochores are generally present as valleys flanked by

bioRxiv preprint doi: https://doi.org/10.1101/096487; this version posted May 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

GC-richer isochores (see Fig. 3A,3B,3C; the relative amounts of L2 sub-families are presented in Supplementary Table S1).

Isochores and TADs

It should be recalled, as a preliminary remark, that the isochores from the five families (L1, L2, H1, H2 and H3) of the human genome (and other mammalian genomes; 38) are characterized not only by increasing GC levels and different short-sequence frequencies, but also by increasing compositional heterogeneities, increasing levels of CpG, CpG islands and Alu sequences and by decreasing levels of LINE sequences and of 5mC/CpG ratios (33,39-43; G. Lamolle, H. Musto and G. Bernardi, paper in preparation). Moreover, at the chromatin level, GC increases are correlated with higher bendability (44), higher nuclease accessibility (45,46), lower nucleosome density (47) and lower supercoiling (48), all properties linked to DNA sequences.

The connection of the isochores of chromosome 21, as seen in Figs. 1D and 2, with chromatin loops can be described as follows (see Fig. 4A,4B,4C): 1) regions 1 and 3 show a series of H1 (**a**) and L2 (**b** to **f**) isochores in which latter case at least some of their single peaks appear to correspond to individual self-interactions; 2) region 2 is the GC-poorest L1 isochore which corresponds to two self-interactions separated by an exceptional interLAD; 3) the multi-peak H1 isochores of region 4 correspond to a large interLAD region and to several self-interactions; the two short X and Y sequences, corresponding to LADs, separate region 4 from regions 3 and 5; 4) the small multi-peak H2 isochore (region 5) seems to correspond to a single self-interaction; this may be due to a lack of resolution (see below); 5) a series of H3 isochores (red points comprised between two red arrows) correspond to a series of self-interactions comprised between the two red lines on the heat map; in this case, the six H3 isochore peaks correspond to at least three chromatin loops (see below). In conclusion, the two classes of

bioRxiv preprint doi: https://doi.org/10.1101/096487; this version posted May 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

isochores, single-peak and multi-peak, essentially correspond to two classes, single-loop and multi-loop, respectively, of TADs, both of which also show inter-chromosomal interactions (15).

The correspondence between isochore peaks and self-interactions is improved at a higher resolution of the heat map (compare the high-resolution Fig. S2A with the low resolution Fig. S2B). Likewise, a very good match of isochore peaks with chromatin loops can be seen in the high resolution heatmap of the multipeak H2 isochore of chromosome 20 (see Supplementary Fig. S3). Finally, 34 TADs/LADs were identified by molecular imaging on chromosome 21 (49).

A very interesting correlation is shown in Fig. 4D, in that GC-rich regions 1 and 3-6 correspond to A compartments (open chromatin; see ref. 23) whereas GC-poor region 2 and X and Y sequences correspond to B compartments (closed chromatin). More precisely, the A compartment corresponds to multi-peak isochore TADs (regions 1,3,4,5,6), the B compartment to individual LADs (region 2,X,Y), the former being predominant in the telomeric region.

Discussion and Conclusions

Encoding of chromatin domains by isochores

Very recent investigations showed that GC-poor and GC-rich isochores should be visualized as the framework of chromatin architecture or, in other words, as the DNA units that underlie LADs and TADs, respectively (15). This was a crucial step towards the idea that isochores encode chromatin domains. The present results provide a conclusive evidence for this idea 1) by showing in more detail a match between the chromatin domains and the isochores of chromosome 21, including the isochores from the newly discovered $L2^+$ and $L2^-$ sub-families; 2) by finding that isochores mold chromatin domains (see the following subsection); and 3) by generalizing these results to all human chromosomes.

bioRxiv preprint doi: https://doi.org/10.1101/096487; this version posted May 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Indeed, the compositional profiles, the TAD heatmaps and the LAD maps (15) show that: 1) the isochores from the L1 family and the $L2^-$ sub-family correspond to LADs in all human chromosomes; 2) $L2^+$ peaks emerging from an L1 background and corresponding to interLADs and TADs are also found in other human chromosomes, although less frequently than in chromosome 21; likewise, H3 isochores, encompassing a broad GC range and corresponding to interLADs and TADs are present in most human chromosomes; 3) the spikes of the compositional profiles of H1 and H2 isochores of Fig. 1B, that reflect the peaks of Fig. 1D, are regularly present in H1 and H2 isochores from all human chromosomes and correspond to the peaks of the point-by-point profiles (P. Cozzi et al., paper in preparation), as well as to self-interactions; finally, weak LADs flank centromeres in all chromosomes. This general match is important in that the only alternative to the encoding proposed here is that the match of the thousands of LADs and TADs with the corresponding isochores is just a coincidence (and this cannot be *quia absurdum*).

Molding of LADs and TADs by isochores

A preliminary remark should be made to stress that the models under discussion here concern the basic evolutionarily stable chromatin domains, since, as it is well known, epigenetic modifications and environmental factors may cause changes in chromatin architecture; indeed, while self-associating domains are stable in mammals, chromatin interactions within and between domains may change during differentiation (see refs.16-22 for reviews) and evolution (the latter subject will be discussed elsewhere).

The present results solve an important open problem, namely the mechanism of formation of chromatin domains. Indeed, LADs should be visualized as chromatin structures corresponding to GC-poor isochores that are flexible, because of the local wider nucleosome spacings linked to

bioRxiv preprint doi: https://doi.org/10.1101/096487; this version posted May 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the richness of oligo-A sequences in the corresponding isochores (39,43,50,51; G. Lamolle, H. Musto, G. Bernardi; paper in preparation). LADs only twist and bend in the three dimensions in order to adapt and attach themselves to (and even embed in) the lamina, which is reassembled after mitosis (10). Expectedly, this leads to self-interactions (see Fig. 4), as well as to interactions with other LADs from the same chromosomes (15; see, for example, the two X and Y LADs bracketed by black lines in Fig. 4).

In the case of TADs, the GC gradient within each GC-rich isochore peak is accompanied by increasing levels of CpG, CpG islands and Alu sequences that lead to increasing nucleosome spacing, bendability and accessibility, as well as to decreasing supercoiling (44-48). These factors constrain the corresponding chromatin to fold into loops, the tips of the loops corresponding to the highest GC levels, the GC peaks; the architectural proteins (CTCF, cohesin) play an important role in closing the loops and in ensuring loop insulation. It should be stressed that the term loop is just a conventional definition for the basic structure of TADs, since the folding of chromatin involves supercoiled structures that are increasingly underwound with increasing GC levels (48).

Figs. 5A and 5B display the "isochore molding models" developed here for the formation of LADs and TADs. In these new models, a crucial role is played by isochore sequences, the differential nucleosome bindings and the corresponding local (in LADs) or extended (in TADs) flexibilities of the chromatin fibers. Needless to say, the "isochore molding models" are in agreement with the encoding of chromosome domains by isochores.

Interestingly, the "isochore molding model" of chromatin domains represents a paradigm shift compared to previously proposed models (discussed in the following sub-section). Indeed, the new model 1) is essentially based on the physico-chemical properties of isochore sequences and on their differential binding of nucleosomes; and 2) provides the same basic explanation for

bioRxiv preprint doi: https://doi.org/10.1101/096487; this version posted May 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

both TADs and LADs. As already mentioned, the model still involves the interactions of architectural proteins visualized by other models.

Finally, it should be stressed that the "isochore molding model" is compatible 1) with both the absolute requirement of CTCF binding to close chromatin loops into insulated TADs (52) or the lack of such requirement (53); 2) with both a partial or an absolute need of cohesin for TAD formation (54); and 3) with the interaction of topoisomerase II beta with cohesin and CTCF at topological domain borders (55).

The "isochore molding model" and the other models of TAD formation

Figs. 6A, 6B and 6C present the existing models for the formation of TADs, the "handcuff model" (32), the "extrusion model" (24, 26-28) and the "insulation-attraction model" (17). The first model has two problems: 1) the practical impossibility for the boundary of a given loop to find the other boundary in the very complex nuclear space; and 2) the lack of an explanation for the folding process itself. The main problem of the "extrusion model" is the fact that the source of energy required is unknown. In the third model, the problem is that chromatin flexibility is seen as modulated 1) by nucleosome spacing not visualized, however, as directly linked to DNA base composition; 2) and/or by other properties.

The "isochore molding model" of chromatin domains can solve the problems of both the "handcuff model" and the "insulation-attraction model", and add a new element to the "extrusion model" of TAD formation. Indeed, in the case of the "handcuff model", the "isochore molding model" accounts for the formation of loops as due to the GC gradient and the accompanying increasing nucleosome depletion and flexibility, as well as for a central start (relative to loop boundaries) of the folding process, this symmetry favoring (together with supercoiling) the joining of the ends of the corresponding loops (Fig. 7A). In the case of the "insulation-attraction

bioRxiv preprint doi: https://doi.org/10.1101/096487; this version posted May 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

model", the "isochore molding model" shows that chromatin flexibility is directly linked to DNA base composition and to its accompanying properties (such as CpG islands density). In conclusion, both the "handcuff" and the "insulation-attraction" models can merge into the "isochore molding model", if the properties of the latter are taken into consideration (Fig. 7A).

In the case of the "extrusion model" (the object of a very recent discussion; 56), the "isochore molding model" suggests that extrusion may start at chromatin sites corresponding to isochore GC peaks. In this case loop molding would help extrusion and would lead to a "central" extrusion process. Moreover, CTCF, the component of the extrusion factor which directly interacts with DNA, has a binding site, CCCTC, which is very likely to be present in GC-rich isochore peaks (Fig. 7B). Incidentally, this modified "extrusion model" would require less energy because extrusion would concern a molded (and supercoiled) loop so avoiding the pulling of chromatin fibers from increasingly longer distances. In contrast, when extrusion starts at random sites, as predicted by the model, isochore molding would be a serious obstacle to extrusion, because it would create chromatin folds that should be resolved by the extrusion process.

At this point, one may have the impression that the "isochore molding model" and the "extrusion model" as just described are equivalent in terms of explaining the formation of TADs. This is not so for at least two major reasons.

In the "isochore molding model" cohesin provides the final closure of the (supercoiled) loop already started by the CTCF interactions. In the "extrusion model" the cohesin complex (cohesin + CTCF) provides the initial step (as well as the following ones) of loop formation. According to what was just mentioned above, the (supercoiled) loop is already present as a result of the "isochore molding" and the extrusion does not concern any more the formation of the loop but simply its extrusion, at the expense of an unknown source of energy. In other words, extrusion is useless while still requiring energy .

bioRxiv preprint doi: https://doi.org/10.1101/096487; this version posted May 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The second reason concerns what happens in the case of the "mitotic memory", namely the rapid and precise re-establishment of the original interphase chromatin domains at the exit from mitosis. In the case of the "isochore molding model", the basic information required for such quick re-establishment is already present in the sequences of isochores (with the additional possible help of CTCF if retained over the interphase to mitosis transition; see ref. 7). In the case of the "extrusion model", the formation of thousands of loops involves the simultaneous attachment of loop extruding factors (the cohesin complex) and extrusion processes requiring a source of energy which is still unknown. It is obvious that the "isochore molding model" which relies on well-known intrinsic physical properties of DNA (that also account for the formation of LADs) and on "classical" interactions with architectural proteins is more parsimonious and more likely to be correct than the second one, which, again, has no real purpose because of the preceding re-establishment of chromatin domains, while still requiring energy.

Isochores as functional genome units

The present results lead to a new vision of isochores since they not only correspond to a fundamental level of genome structure and organization (6), but also to a set of functional genome units that encode and mold LADs and TADs. Several observations support the above conclusion: 1) the evolutionary conservation of the GC-rich and GC-poor isochore patterns in mammals (38) matches that of TADs and LADs (8-14); 2) GC-rich isochores are replication units characterized by all early replicons (57) and TADs from cells of adult organisms are stable units of early replication timing (58); 3) TADs show a visible similarity between sperm cells and fibroblasts of the mouse in spite of the replacement of histones with protamines (59); 4) codon usage is tightly linked to isochore composition (4, 6) as it is to chromatin architecture (60); 5) previous investigations led to the idea of "genomic diseases", defined as diseases due to sequence

bioRxiv preprint doi: https://doi.org/10.1101/096487; this version posted May 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

alterations that do not affect genes or classical regulatory sequences, but other sequences that "cause regional changes in chromatin structure" (61), and we now know that alterations of the architecture of chromatin domains (both LADs and TADs), known to lead to senescence and diseases, are due to changes in their isochore framework (16-22).

Interestingly, the isochore/chromatin domain connection has some practical implications. Indeed, screening even a small human population in terms of chromatin structure in view of detecting "genomic diseases" is simply not feasible at least at the present time; the availability of LADs and TADs maps along sequenced human chromosomes may allow, however, to link alterations at the DNA sequence level with chromatin structure alterations. For instance, the maps of insertions, deletions and SNPs of Venter's chromosomes (62) in combination with reference chromatin structure maps, might lead (hopefully not) to detect problems in Venter's chromatin domains.

Finally, the fact that alterations in chromatin architecture, not affecting genes or classical regulatory sequences, lead to problems in transcription represents the strongest and final objection to the idea that non-coding DNA is "junk DNA".

Other levels of organization of the human genome

We can now consider a higher level of isochore and chromatin organization. At the DNA level, two "genome spaces" were defined on the basis of gene density (63,64): the gene-poor "genome desert" (L1+L2 isochores; the L2 isochores should now be split into the $L2^+$ and $L2^-$ sub-families; see Supplementary Table 1) and the gene-rich "genome core" (H1+H2+H3 isochores. In the interphase nucleus, the chromatin corresponding to the genome core showed an internal location and an open structure, whereas the chromatin corresponding to the genome desert showed a peripheral location and a closed structure (see Table 1); moreover, the former

bioRxiv preprint doi: https://doi.org/10.1101/096487; this version posted May 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

showed a preference for (generally GC-rich) telomeric regions, the latter for centromeric regions, this preference explaining the polarity of chromosomes (65; see also ref. 66).

The recently proposed chromosome compartments, A and B, characterized by open and closed chromatin, respectively (23), show properties very similar to those just described (see Table 1). Indeed, the two compartments, A and B, appear to correspond to the two genome spaces, the "genome core" and the "genome desert", as well as to TADs and LADs. This conclusion is supported by the comparison of sub-compartments with isochore profiles, in which case A1 sub-compartments correspond to H2/H3 isochores (sometimes including flanking isochores from the H1 and even from the L2 family), A2 sub-compartments to H1 and L2 isochores, B1-B3 sub-compartments to L2 and L1 isochores (15). The compartmentalization of the two genome spaces appears to be due to their different locations on interphase chromosomes and on chromosome folding (see the compartment profile of chromosome 21 in Fig. 4D).

Another, lower, level of organization concerns differences among isochores from the same families. Differences in the frequencies of oligo A's, of short sequences, of CpG islands and of repeated sequences make such isochores different from each other with repercussions on nucleosome density and on the corresponding LADs and TADs. In other words, there is a spectrum of properties within the isochores from the same family and, as a consequence, within the corresponding LADs and TADs.

The genomic code

The encoding of chromatin domains by isochores deserves the name of "genomic code". This definition was originally coined (67,68; see also ref. 6) for two sets of compositional correlations 1) those that hold among genome sequences (for instance, between coding and

bioRxiv preprint doi: https://doi.org/10.1101/096487; this version posted May 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

contiguous non-coding sequences and among the three codon positions of genes) and that reflect isochore properties; and 2) those that link isochores with all the structural/functional properties of the genome (6,7,15), the latter now including the properties of TADs and LADS. Here it is proposed that the definition of "genomic code" be applied to the encoding of chromatin domains by isochores, since this is in fact the basis for the second set of the correlations just mentioned.

Interestingly, the genomic code may be visualized as the fourth, and last, pillar of molecular biology, the first one being the double helix (1951-1953), the second the regulation of gene expression in *E. coli* (1957-1961), and the third the genetic code (1961-1966). In contrast with the other pillars, the genomic code took decades to be established.

Acknowledgements

The author thanks Paolo Ascenzi for hospitality, Giacomo Bernardi, Oliver Clay, and, especially, Kamel Jabbari for critical reading, comments and discussions as well as Caterina Nuvoli for excellent technical help. This research was supported by the Kimura Prize for Molecular Evolution and Evolutionary Genomics conferred to the author (Tokyo, June 2016).

bioRxiv preprint doi: https://doi.org/10.1101/096487; this version posted May 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References:

1. Macaya G, Thiery JP, Bernardi G. An approach to the organization of eukaryotic genomes at a macromolecular level. J Mol Biol. 1976;108: 237–254.

2. Filipski J, Thiery JP, Bernardi G An analysis of the bovine genome by Cs_2SO_4-Ag density gradient centrifugation. J Mol Biol. 1973; 80: 177-197.

3. Corneo G, Ginelli E, Soave C, Bernardi G. Isolation and characterization of mouse and guinea pig satellite deoxyribonucleic acids. Biochemistry. 1968. 7(12): 4373-4379.

4. Bernardi G, Olofsson B, Filipski J, Zerial M, Salinas J, Cuny G, et al. The mosaic genome of warm-blooded vertebrates. Science. 1985;228: 953–8.

5. Bernardi G. The Isochore Organization of the Human Genome. Annu Rev Genet. 1989;23: 637–659.

6. Bernardi G (2004) *Structural and evolutionary genomics: natural selection in genome evolution.* (Elsevier, Amsterdam). This book is freely available at www.giorgiobernardi.eu

7. Bernardi G. Chromosome architecture and genome organization. PLoS One. 2015;10. doi:10.1371/journal.pone.0143739

8. Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature. 2008;453: 948–51.

9. Meuleman W, Peric-Hupkes D, Kind J, Beaudry JB, Pagie L, Kellis M, et al. Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence. Genome Res. 2013;23: 270–280.

10. Kind J, Pagie L, De Vries SS, Nahidiazar L, Dey SS, Bienko M, et al. Genome-wide Maps of Nuclear Lamina Interactions in Single Human Cells. Cell. 2015;163: 134–147.

11. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012;485: 376–380.

12. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature. 2012;485: 381–385.

13. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014;159: 1665–1680.

14. Vietri Rudan M, Barrington C, Henderson S, Ernst C, Odom DT, Tanay A, et al. Comparative Hi-C Reveals that CTCF Underlies Evolution of Chromosomal Domain Architecture. Cell Rep. 2015;10: 1297–1309.

bioRxiv preprint doi: https://doi.org/10.1101/096487; this version posted May 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

15. Jabbari K, Bernardi G. An isochore framework underlies chromatin architecture. Plos One 2017http://dx.doi.org/10.1371/journal.pone.0168023

16. Bonev B, Cavalli G. Organization and function of the 3D genome. Nat Rev Genet. Nature Research; 2016;17: 661–678.

17. Dixon JR, Gorkin DU, Ren B. Chromatin Domains: The Unit of Chromosome Organization. Mol Cell. 2016;62: 668–680.

18. Ghirlando R, Felsenfeld G. CTCF: Making the right connections. Genes and Development. 2016. pp. 881–891.

19. Gonzalez-Sandoval A, Gasser SM. On TADs and LADs: Spatial Control Over Gene Expression. Trends Genet. 2016; doi:10.1016/j.tig.2016.05.004

20. Merkenschlager M, Nora EP. CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. Annu Rev Genomics Hum Genet. 2016;1-27

21. Pueschel R, Coraggio F, Meister P, Ahmed K, Dehghani H, Rugg-Gunn P, et al. From single genes to entire genomes: the search for a function of nuclear organization. Development. 2016. 143:910-23

22. Solovei I, Thanisch K, Feodorova Y. How to rule the nucleus: divide et impera. Curr Opin Cell Biol. 2016;40: 47–59.

23. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009;326: 289–93.

24. Alipour E, Marko JF. Self-organization of domain structures by DNA-loop-extruding enzymes. Nucleic Acids Res. Oxford University Press; 2012;40: 11202–12.

25. Barbieri M, Chotalia M, Fraser J, Lavitas LM, Dostie J, Pombo A, et al. Complexity of chromatin folding is captured by the strings and binders switch model. Proc Natl Acad Sci U S A. 2012;109: 16173–16178.

26. Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. Formation of Chromosomal Domains by Loop Extrusion. Cell Rep. 2016;15: 2038–2049.

27. Sanborn AL, Rao SSP, Huang S-C, Durand NC, Huntley MH, Jewett AI, et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc Natl Acad Sci USA. National Acad Sciences; 2015;112: E6456–E6465.

28. Dekker J, Mirny L. The 3D Genome as Moderator of Chromosomal Communication. Cell. 2016;164: 1110–1121.

bioRxiv preprint doi: https://doi.org/10.1101/096487; this version posted May 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

29. Di Pierro M, Zhang B, Lieberman-Aiden E, Wolynes PG, Onuchic JN. Transferable model for chromosome architecture. Proc Natl Acad Sci. 2016; 201613607.

30. Rowley MJ, Corces VG. The three-dimensional genome: Principles and roles of long-distance interactions. Current Opinion in Cell Biology. 2016. pp. 8–14.

31. Ulianov S V, Khrameeva EE, Gavrilov AA, Flyamer IM, Kos P, Mikhaleva EA, et al. Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. Genome Res. 2016;26: 70–84.

32. Vietri Rudan M, Hadjur S (2015) Genetic Tailors: CTCF and Cohesin Shape the Genome During Evolution. *Trends Genet* 31(11):651–660.

33. Costantini M, Clay O, Auletta F, Bernardi G. An isochore map of human chromosomes. Genome Res. 2006;16: 536–541.

34. Pavlícek A, Paces J, Clay O, Bernardi G. A compact view of isochores in the draft human genome sequence. FEBS Lett. 2002;511: 165–9.

35. Pačes J, Zíka R, Pačes V, Pavlíček A, Clay O, Bernardi G. Representing GC variation along eukaryotic chromosomes. Gene. 2004;333: 135–141.

36. Cozzi P, Milanesi L, Bernardi G. Segmenting the human genome into isochores. Evol Bioinforma. 2015;11: 253–261.

37. Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature. 2008;453: 948–51.

38. Costantini M, Cammarano R, Bernardi G. The evolution of isochore patterns in vertebrate genomes. BMC Genomics. 2009;10: 146. doi:10.1186/1471-2164-10-146

39. Costantini M, Bernardi G. The short-sequence designs of isochores from the human genome. Proc Natl Acad Sci U S A. 2008;105: 13971–6.

40. Meunier-Rotival M, Soriano P, Cuny G, Strauss F, Bernardi G. Sequence organization and genomic distribution of the major family of interspersed repeats of mouse DNA. Proc Natl Acad Sci U S A. 1982;79: 355–9.

bioRxiv preprint doi: https://doi.org/10.1101/096487; this version posted May 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

41. Soriano P, Meunier-Rotival M, Bernardi G. The distribution of interspersed repeats is nonuniform and conserved in the mouse and human genomes. Proc Natl Acad Sci U S A. 1983;80: 1816–20.

42. Varriale A, Bernardi G. Distribution of DNA methylation, CpGs, and CpG islands in human isochores. Genomics. 2010;95: 25–28.

43. Arhondakis S, Auletta F, Bernardi G. Isochores and the regulation of gene expression in the human genome. Genome Biol Evol. 2011;3: 1080–9.

44. Vinogradov AE. DNA helix: The importance of being GC-rich. Nucleic Acids Research. 2003. pp. 1838–1844.

45. Di Filippo M, Bernardi G. Mapping DNase-I hypersensitive sites on human isochores. Gene. 2008;419:62-65.

46. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. The accessible chromatin landscape of the human genome. Nature. 2012;489: 75–82.

47. Fenouil R, Cauchy P, Koch F, Descostes N, Cabeza JZ, Innocenti C, et al. CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. Genome Res. 2012;22: 2399–2408.

48. Naughton C, Avlonitis N, Corless S, Prendergast JG, Mati IK, Eijk PP, et al. Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. Nat Struct Mol Biol. 2013;20: 387–95.

49. Wang S, Su J-H, Beliveau BJ, Bintu B, Moffitt JR, Wu C -t., et al. Spatial organization of chromatin domains and compartments in single chromosomes. Science (80-). 2016;353: 598–602.

50. Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, et al. The DNA-encoded nucleosome organization of a eukaryotic genome. Nature. 2009;458: 362–366.

51. Frenkel ZM, Bettecken T, Trifonov EN. Nucleosome DNA sequence structure of isochores. BMC Genomics. 2011;12: 203.

bioRxiv preprint doi: https://doi.org/10.1101/096487; this version posted May 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

52. Nora EP, Goloborodko A, Valton AL, Gibcus J, Uebersohn A, et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from higher-order genomic compartmentalization. http://biorxiv.org/content/early/2016/12/21/095802
53. Kubo N, Ishii H, Gorkin D, Meitinger F, Xiong X, Fang R, et al. Preservation of Chromatin Organization after Acute Loss of CTCF in Mouse Embryonic Stem Cells. bioRxiv. Cold Spring Harbor Labs Journals; 2017; 118737.
54. Schwarzer W, Abdennur N, Goloborodko A, Pekowska A, Fudenberg G, Loe-Mie Y, et al. Two independent modes of chromosome organization are revealed by cohesin removal. bioRxiv. 2016; Available: http://biorxiv.org/content/early/2016/12/15/094185
55. Uusküla-Reimand L, Hou H, Samavarchi-Tehrani P, Rudan MV, Liang M, et al. Topoisomerase II beta interacts with cohesin and CTCF at topological domain borders. Genome Biol. 2016; 17: 1–22.
56. Dolgin E. A loop of faith. Nature. 2017;544: 284–286.
57. Costantini M, Bernardi G. Replication timing, chromosomal bands, and isochores. Proc Natl Acad Sci U S A. 2008;105: 3433–3437.
58. Pope BD, Ryba T, Dileep V, Yue F, Wu W, Denas O, et al. Topologically associating domains are stable units of replication-timing regulation. Nature. 2014;515: 402–405.
59. Battulin N, Fishman VS, Mazur AM, Pomaznoy M, Khabarova AA, Afonnikov DA, et al. Comparison of the three-dimensional organization of sperm and fibroblast genomes using the Hi-C approach. Genome Biol. BioMed Central; 2015;16: 77.
60. Diament A, Pinter RY, Tuller T. Three-dimensional eukaryotic genomic organization is strongly correlated with codon usage expression and function. Nat Commun. 2014;5: 5876.
61. Bernardi G. The neoselectionist theory of genome evolution. Proc Natl Acad Sci U S A. 2007;104: 8385–90.
62. Costantini M, Bernardi G. Mapping insertions, deletions and SNPs on Venter's chromosomes. PLoS One. 2009;4. doi:10.1371/journal.pone.0005972
63. Bernardi G. The Vertebrate Genome : Isochores and Evolution. Mol Biol Evol. 1993;10: 186–204.

bioRxiv preprint doi: https://doi.org/10.1101/096487; this version posted May 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

64. Bernardi G. Isochores and the evolutionary genomics of vertebrates. Gene. 2000. 3–17.

65. Saccone S, Federico C, Bernardi G. Localization of the gene-richest and the gene-poorest isochores in the interphase nuclei of mammals and birds. Gene. 2002;300: 169–78.

66. Gilbert N, Boyle S, Fiegler H, Woodfine K, Carter NP, Bickmore WA. Chromatin architecture of the human genome: Gene-rich domains are enriched in open chromatin fibers. Cell. 2004;118: 555–566.

67. Bernardi G. Le génome des vertébrés : organisation, fonction et évolution. Biofutur. 1990. 94 : 43-46.

68. Bernardi G, Bernardi G. Compositional properties of nuclear genes from cold-blooded vertebrates. J. Mol. Evol. 1986. 33:57-67.

69. Naumova N, Imakaev M, Fudenberg G, Zhan Y, Lajoie BR, Mirny LA, et al. Organization of the mitotic chromosome. Science. 2013;342: 948–53.

70. Consortium IHGS. Initial sequencing and analysis of the human genome. Nature. 2001;409: 860–921.

71. Bernardi G. Misunderstandings about isochores. Part 1. Gene. 2001. pp. 3–13.

72. Jabbari K, Nürnberg P. A genomic view on epilepsy and autism candidate genes. Genomics. 2015. 108:31-36.

bioRxiv preprint doi: https://doi.org/10.1101/096487; this version posted May 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

TABLE 1. GENOME ORGANIZATION: DNA AND CHROMATIN (a)		
2 Genome spaces (61,62)	**Genome core: GC-rich, gene-rich isochores**	**Genome desert: GC-poor, gene-poor isochores**
2 Nuclear locations (63,64)	**Internal**	**Peripheral**
2 Chromatin structures (64)	**Open**	**Closed**
2 Replication timings (57,58)	**Early**	**Late**
2 Genome compartments (23)	**A, open chromatin**	**B, closed chromatin**
6 Isochore families (6,8)/**sub-families**[(b)]	**L2$^+$, H1, H2, H3**	**L1, L2$^-$**
2 Chromatin domains (10-14)	**TADs**	**LADs**

(a) See sub-section "The large-scale organization of the human genome" in the Discussion/Conclusions section of the manuscript for a description of this Table.

(b) Present work

bioRxiv preprint doi: https://doi.org/10.1101/096487; this version posted May 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure Legends

Figure 1. A: Compositional profile (from ref. 36) of human chromosome 21 (release hg38) as seen through non-overlapping 100-Kb windows. DNA stretches from isochore families L1 to H3 are represented in different colors, deep blue, light blue, yellow, orange, red, respectively. The left-side ordinate values are the minima GC values between the isochore families listed on the right side (see Supplementary Table S1); **a-f** correspond to an H1 peak (**a**) and to several $L2^+$ peaks (**b** to **f**; the thin yellow bars in peaks **c**, **e** and **f** correspond to single 100Kb blocks that are assigned to the H1 family if using "fixed" isochore borders (as shown in the Figure), but to L2 isochores if "extended" borders (see Supplementary Table 1) are used. Six regions numbered 1 to 6 (top and bottom of the Figure) of the compositional profile are separated by vertical lines or double lines, X and Y.

B. Isochore profile of human chromosome 21 using the matched b37 assembly of ref. 13 and a sliding window approach (34,35) with "fixed" isochore borders (from ref. 15). The color convention is as in Fig. 1A. This profile is slightly more extended on the centromeric (left) side than those of Figs. 1A and 1C.

C. Isochore profile of human chromosome 21 (release hg38) using a non-overlapping 100Kb window and the isoPlotter program (from ref. 36).

D. GC levels of 100Kb windows of human chromosome 21. Data of Fig. 1B are plotted as points. This figure shows that individual isochores from the $L2^+$ to H3 families are in the form of peaks. The GC gradients of the peaks do not appear in a clear way in the standard presentation of compositional profiles of chromosomes (Fig. 1A), except for the broad, isolated H1 (**a**) and $L2^+$ peaks (**b** to **f**). Upon close inspection, however, the peaks of the H1, H2 and H3 isochores of this

bioRxiv preprint doi: https://doi.org/10.1101/096487; this version posted May 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure show a correspondence with the small peaks of isochores belonging to H1, H2 and H3 family present in Fig. 1B. Blue and red arrows delimit regions 2 and 6, respectively. Black, blue and red arrows, as well as X,Y double lines (see Fig. 4 and its legend for additional information) separate regions 3,4 and 5; horizontal lines correspond to minima GC values between isochore families (see legend of Fig. 1A and Supplementary Table 1).

E.F. The compositional map of a 2.1Mb region of human chromosome 20, using a sliding window approach (34,35) and the "extended" family borders of Supplementary Table 1 (Fig. 1E), is aligned with the corresponding "point-by-point map" (Fig. 1F, in the Fig. 1D style). The correspondence of five peaks and valleys from the latter (red lines) precisely correspond to the peaks and valleys of the compositional profile; several additional peaks and valleys show a less precise, yet unequivocal correspondence (blue lines). The GC range of the peaks from the bottom profile, 1F, is much larger (about double) than the corresponding one of the top profile, 1E. The whole region may be visualized as a multipeak H2 isochore on the basis of the bottom profile (see also Supplementary Fig. S3).

Figure 2A. GC levels of 100Kb windows from L1 and H3 regions of human chromosome 21 comprised between the blue and red arrows (regions 2 and 6) of Fig. 1D, respectively, are displayed at a higher magnification. Horizontal lines correspond to "fixed" minima GC values between isochore families (see Supplementary Table 1). The abscissa scale is in 100Kb units.

B.C. GC levels of 100Kb windows from the regions 1,3,4 and 5 of human chromosome 21. See the legend of Fig. 2A for other indications. **C.** Region 4 is delimited by two GC-poor sequences, X and Y, that correspond to LADs (see Fig. 4 and its legend for additional information).

bioRxiv preprint doi: https://doi.org/10.1101/096487; this version posted May 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 3. A. The isochore profile of human chromosome 21 (from ref. 15; see also legend of Fig. 3A) is compared with the (inverted) LAD profile (from ref. 10) to show the correspondence of LADs 1) with L1 isochores (blue lines; two broken blue lines bracket the largest L1 isochore, in which case the LAD is (exceptionally) interrupted by an interLAD); 2) with one L2 "valley" isochore (blue line, last on the right side); and 3) with two LADs flanking the centromere. Four "high heterogeneity" $L2^+$ "peak" isochores (red lines) correspond to interLADs, or TADs; the H1 peak right of the centromere corresponds to the same interLAD as the first $L2^+$ peak. Two LADs flank the centromere. The multicolored bar on the right is the color code for isochore families.

B. The isochore profile of human chromosome 20 (from ref. 15; see also legend of Fig. 3A) is compared with the (inverted) LAD profile (from ref. 10) to show the correspondence of LADs with a large low-heterogeneity L2 ($L2^-$) isochore (bracketed by broken blue lines; on the left of the panel) which includes an interLAD (red line) and with several L2 "valley" isochores (blue lines), as well as the correspondence of interLADs with GC-rich isochores (red lines); two faint LADs flank the centromere.

C. The isochore profile of human chromosome 19 (from ref. 15; see also legend of Fig. 3A) is compared with the (inverted) LAD profile (from ref. 10) to show that two LADs correspond to H1 isochores (blue lines) flanking a H2 isochore, an interLAD (red line); two other LADs (on the left) correspond to L2 "valley" isochores (blue lines); two faint LADs flank the centromere.

Figure 4. The compositional profile of human chromosome 21 (**B,** from Fig. 1D), is compared 1) with the corresponding LAD profile (**A**; an inverted scale from ref.10 is shown on the top right of the figure); 2) with the heat map of chromatin interactions (**C**; from ref. 15) and with Fig. 1B on the right side; 3) with the A/B compartment profile (**D**; data for mid G1 from ref. 37). Regions 1

bioRxiv preprint doi: https://doi.org/10.1101/096487; this version posted May 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and 3 correspond to multiple interactions and to open chromatin; region 2 to multiple interactions and closed chromatin, as well as to an interLAD. Two double vertical black lines, X and Y, corresponding to LADs (characterized by both self- and intra-chromosomal interactions) and to closed chromatin separate region 4, a multi-peak H1 isochore block corresponding to multiple interactions and open chromatin, from regions 3 and 5. The small multi-peak H2 isochores (region 5) corresponds to a single self-interaction and to open chromatin (see Text). A telomeric block of H3 isochores (region 6) defined by red points, arrows and lines corresponds to several (at least three) self-interactions on the heat map and to open chromatin (see Text).

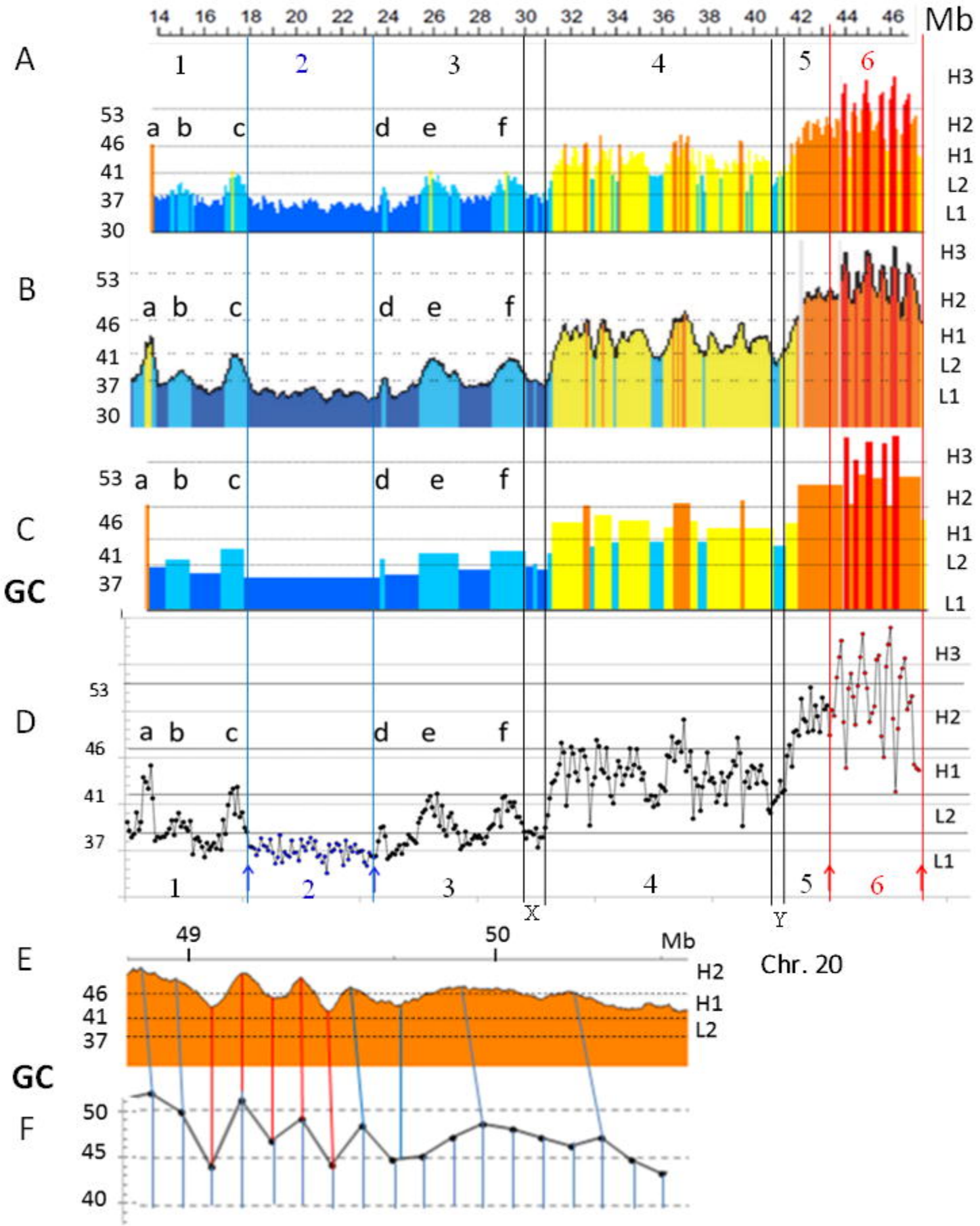
Figure 5. The "isochore molding model" for the formation of LADs and TADs. **A**. A GC-poor chromatin fiber corresponding to an L1 isochore; blue bar) bounded by CTCF binding sites (green boxes) attaches to the lamina, forming a LAD; the wavy profile indicates the three-dimensional physical adaptation by bending and twisting to (and embedding in) the lamina due to local nucleosome depletions associated with oligo-A sequences (yellow bands). **B**. A GC-rich chromatin fiber corresponding to an H3 isochore folds upon itself forming a single-loop TAD bounded by CTCF-binding sites (green boxes); the yellow to red color gradient indicates an increasing GC level with increasing frequencies of CpG islands and repeated sequences. These properties are responsible for the increasing flexibility due to the increasing nucleosome depletion and for the folding. The supercoiling of the loop (see Text) is not shown.

Figure 6. Models for the formation of TADs. **A.** The "handcuff model" of TAD formation (according to the graphical abstract of ref. 14): the two ends of a loop are brought together by CTCF units (yellow boxes) that bind to their binding sites (blue arrows) located on the loop

bioRxiv preprint doi: https://doi.org/10.1101/096487; this version posted May 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

boundaries and recruit the cohesin complex (red circle). **B.** The "extrusion model" of TAD formation (according to ref. 56). A loop is generated dynamically by extrusion factors (a pair of CTCF units and cohesin), that attach to the chromatin fiber and travel along in opposite directions until they reach converging CTCF binding sites (not shown). **C.** The "insulation-attraction model" (from ref. 17). Chromatin stiffness at boundaries is associated with closer nucleosome spacing due to the presence of CTCF sites, promotors etc (and/or with other properties).

Figure 7. Versions of the previous models as revised on the basis of the "isochore molding model". **A.** The "handcuff" and the "insulation-attraction model" merge into the "isochore molding model". **B.** The "extrusion model" of ref. 56 is modified on the basis of the "isochore molding model". Extrusion starts not at random positions but at the GC-richest chromatin peaks and involves a supercoiled chromatin (not shown).


Mb
14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46
A
B
C
GC
D
E
F
1 2 3 4 5 6
a b c d e f
53 46 41 37 30
H3 H2 H1 L2 L1
X
Y
49 50
Chr. 20
50 45 40

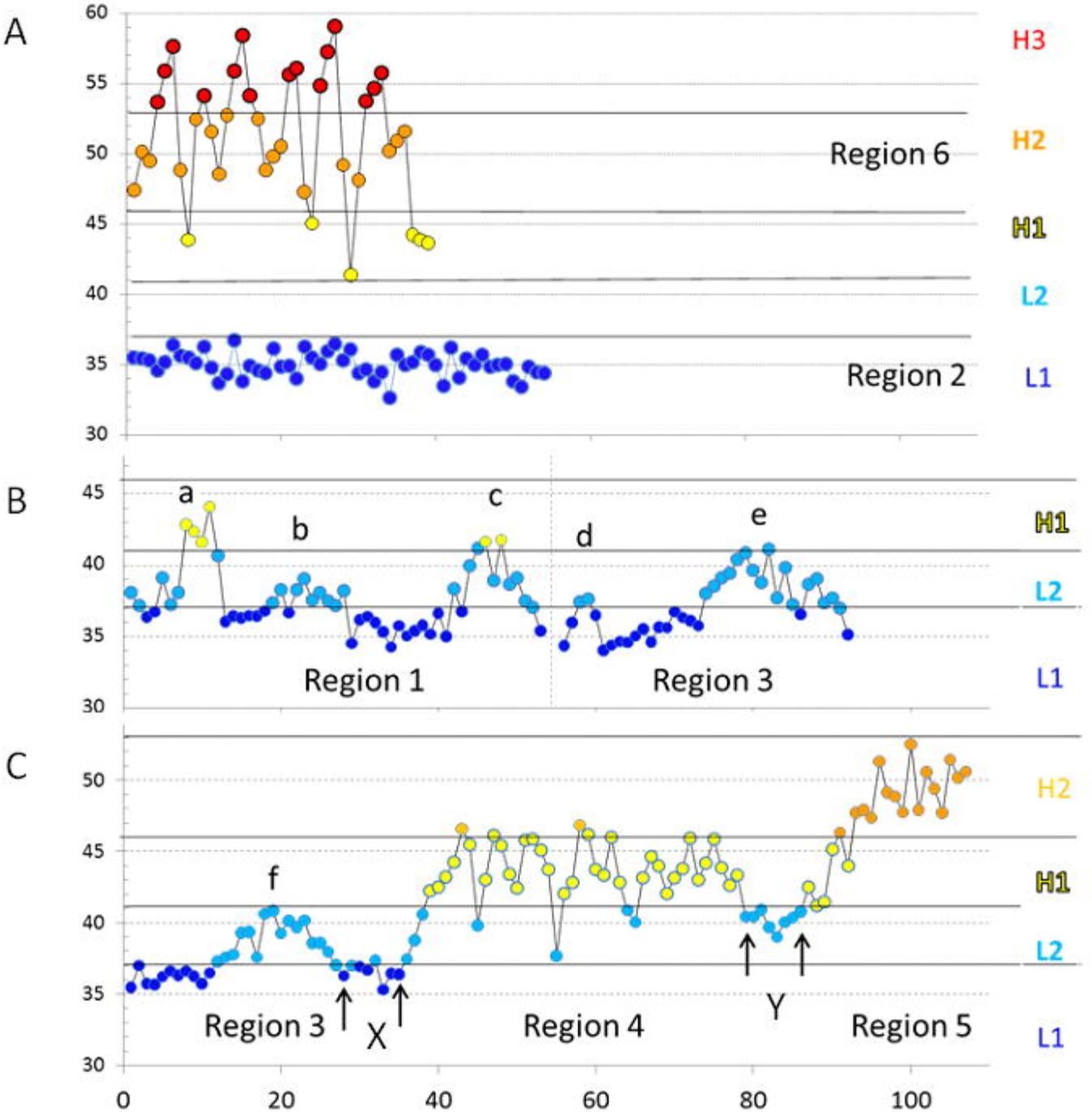

A
60
55
50
45
40
35
30
H3
H2
H1
L2
L1
Region 6
Region 2
B
45
40
35
30
a
b
c
d
e
H1
L2
L1
Region 1
Region 3
C
50
45
40
35
30
f
H2
H1
L2
L1
Region 3
X
Region 4
Y
Region 5
0
20
40
60
80
100

Fig. 2

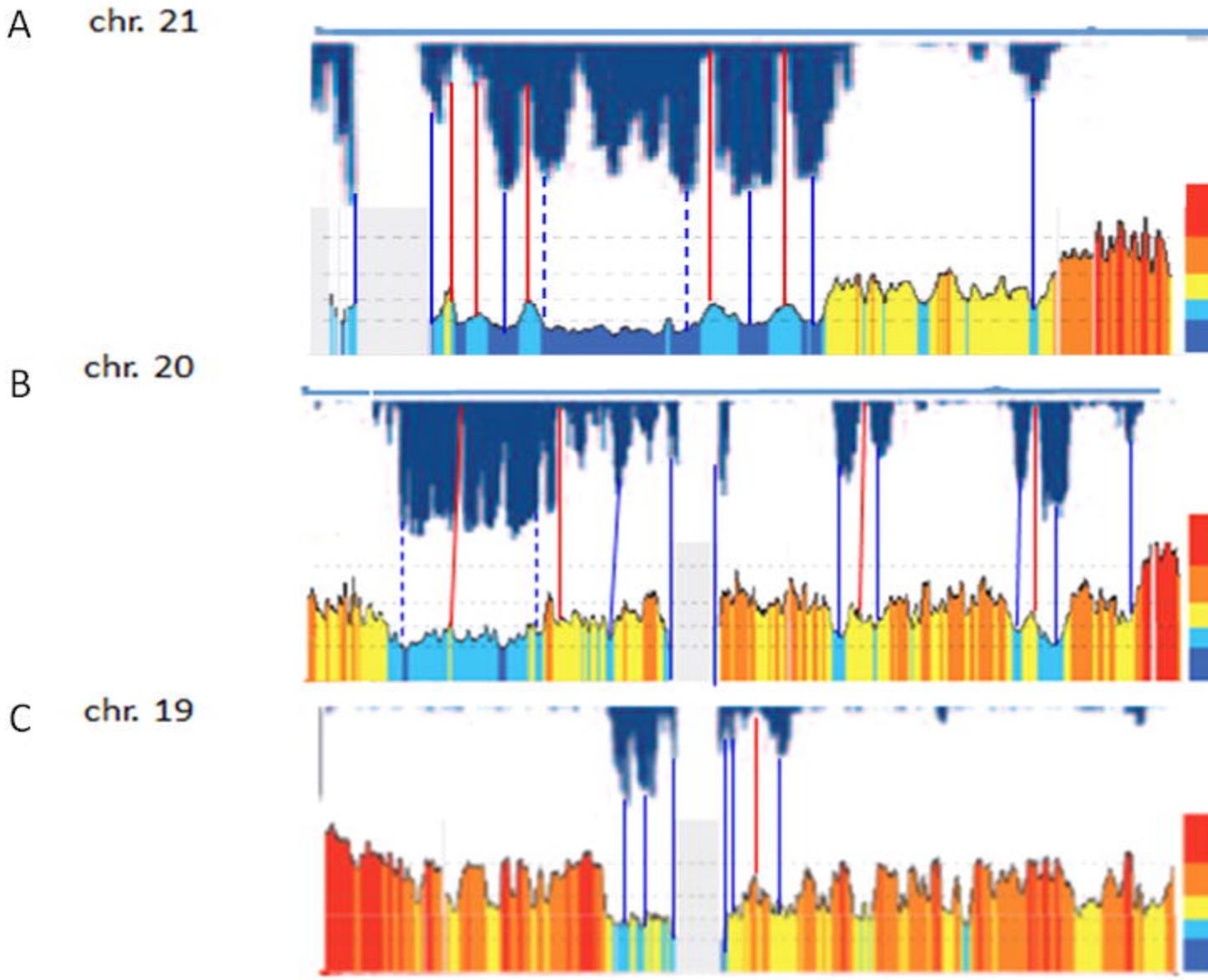
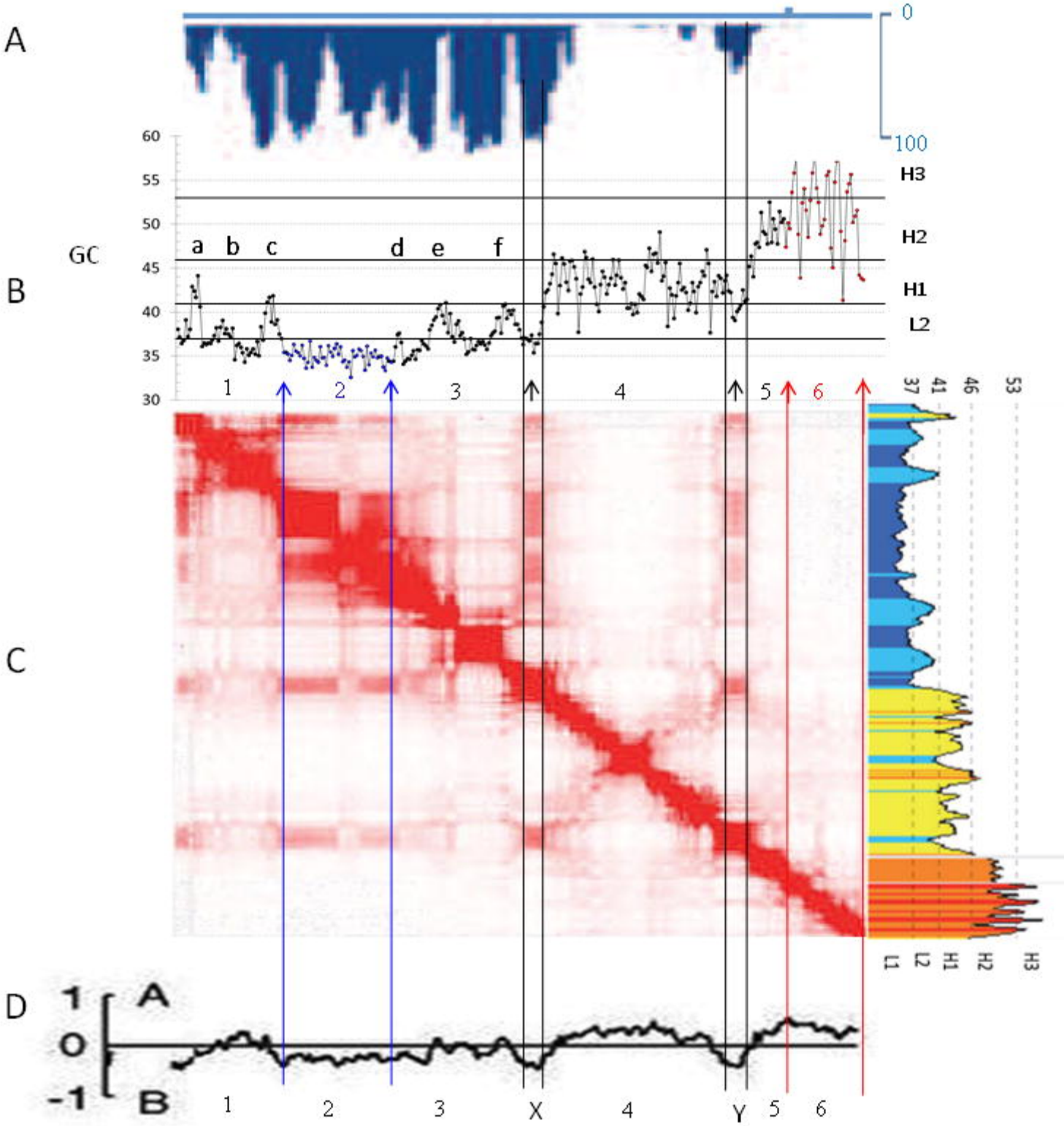


Fig. 3


A
0
100
B
GC
60
55
50
45
40
35
30
a b c d e f
H3
H2
H1
L2
1 2 3 4 5 6
C
37 41 46 53
L1 L2 H1 H2 H3
D
1
0
-1
A
B
1 2 3 X 4 Y 5 6

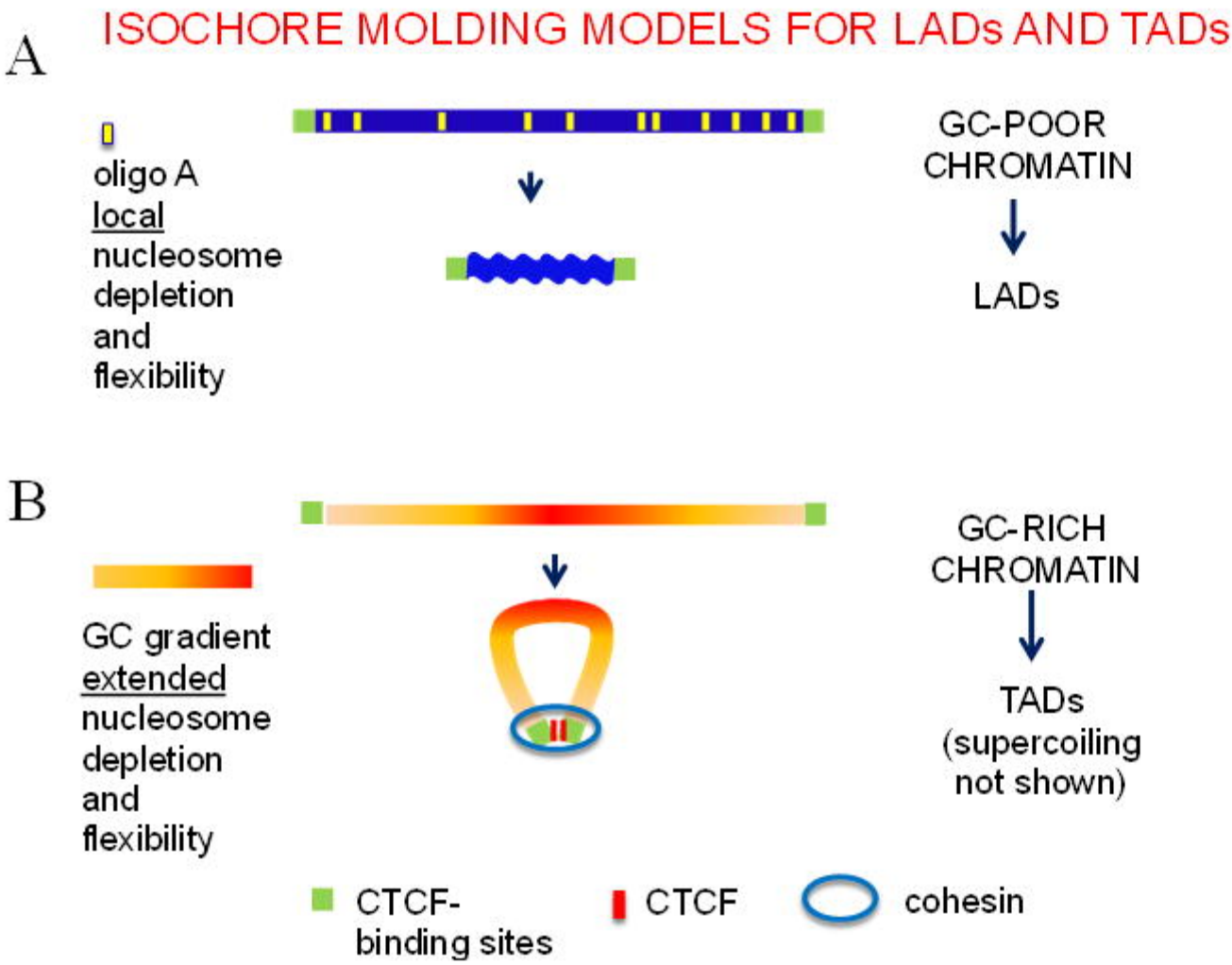
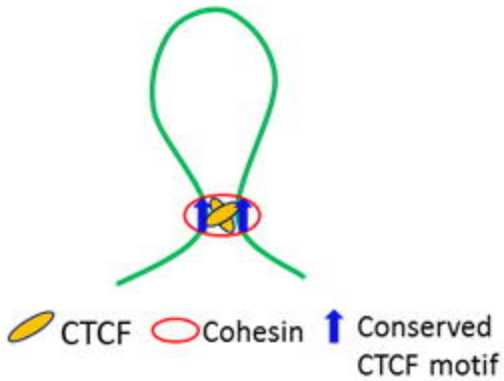

A
ISOCHORE MOLDING MODELS FOR LADs AND TADs
oligo A
local
nucleosome
depletion
and
flexibility
GC-POOR
CHROMATIN
LADs
B
GC gradient
extended
nucleosome
depletion
and
flexibility
GC-RICH
CHROMATIN
TADs
(supercoiling
not shown)
CTCF-
binding sites
CTCF
cohesin

Fig. 5

MODELS FOR TADs

A

HANDCUFF



PROBLEMS

1. FOLDING
2. CLOSING THE CORRESPONDING BOUNDARIES

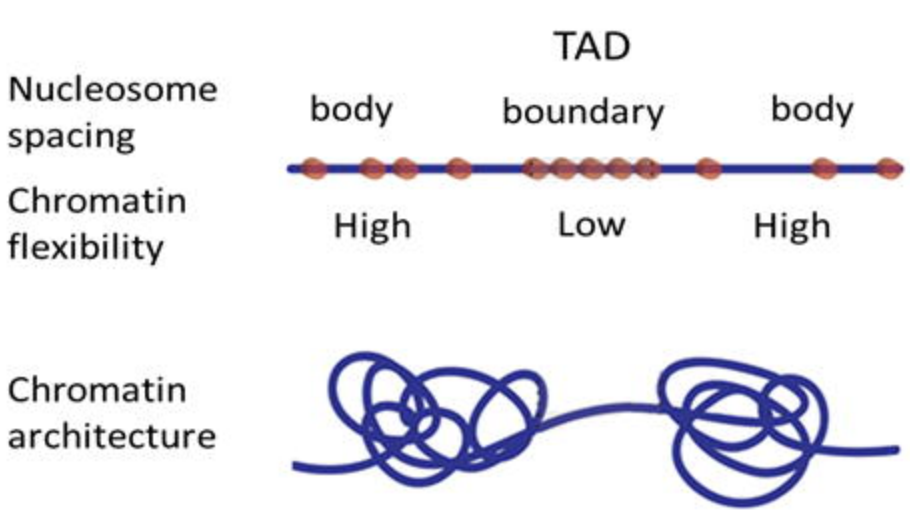
B

EXTRUSION

1. ENERGY SOURCE
2. RANDOM INITIATION

C

INSULATION-ATTRACTION



FLEXIBILITY DUE

1. TO NUCLESOME SPACING (CTCF SITES, PROMOTORS)
2. AND/OR TO OTHER PROPERTIES

Fig. 6

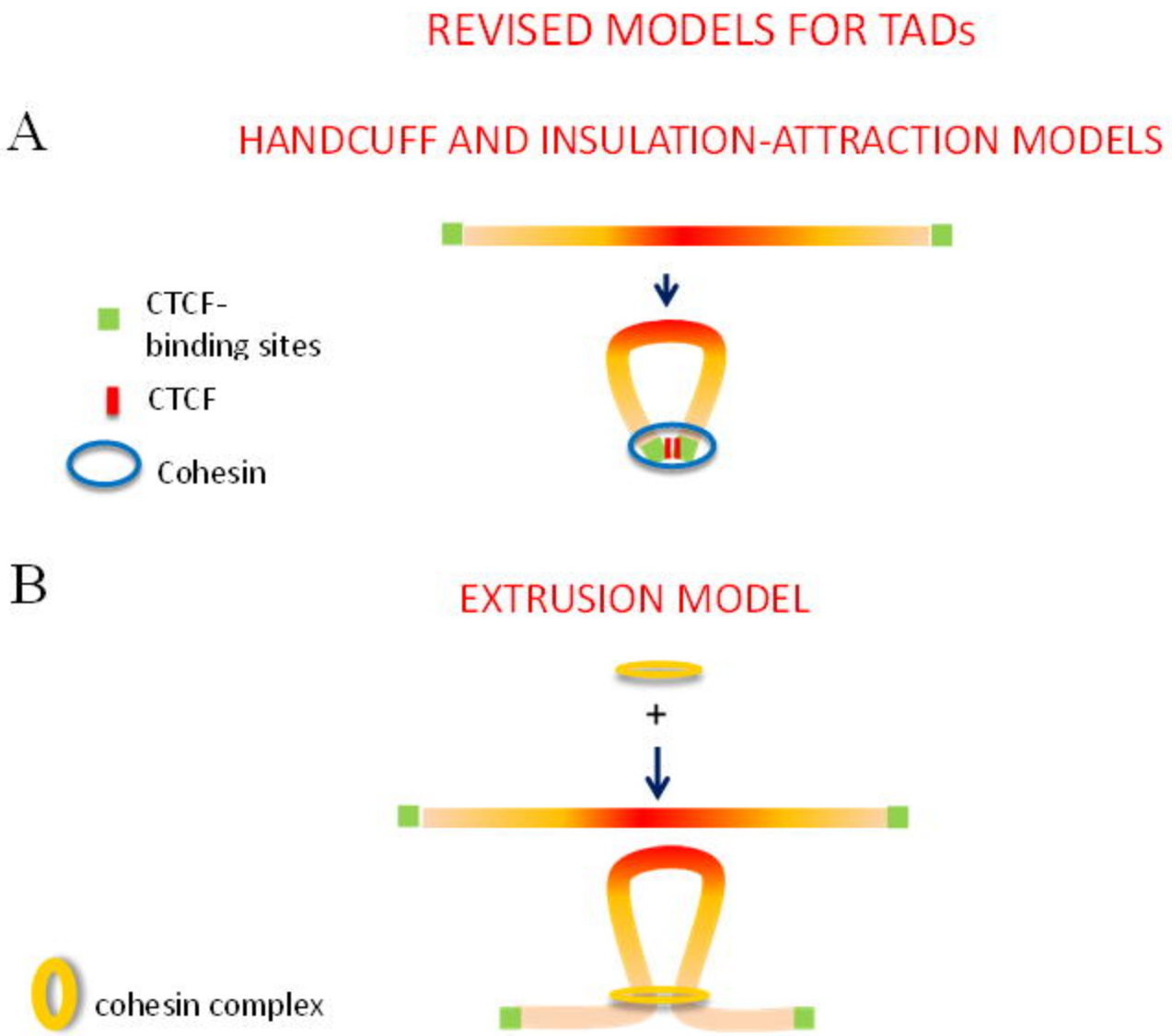

REVISED MODELS FOR TADs
A
HANDCUFF AND INSULATION-ATTRACTION MODELS
CTCF-
binding sites
CTCF
Cohesin
B
EXTRUSION MODEL
+
cohesin complex

Fig. 7