

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



A Consistent Systems Mechanics Model of the 3D Architecture and Dynamics of Genomes

Tobias A. Knoch

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.89836>

Abstract

Already for thousands of years mankind is aware of inheritance and its manipulation by mating and breeding. The discovery of the cell nucleus by A. van Leeuwenhook in the 17th century marks a start to elucidate the epically discussed evolutionary transfer of information in detail. Now after more than 170 years of research on the 3D architecture and dynamics of genomes and the co-evolved interaction networks of regulatory elements creating genome function - i.e. the storage, replication, and expression of genetic information—a consistent systems statistical mechanics genomics framework emerges for the first time. Obviously the structure and function of genomes co-evolved as an inseparable system allowing the physical storage, expression, and replication of genetic information. The DNA double helix and the nucleosome had been determined structurally at the very highest level already, including genome sequences and epigenetic histone modifications. That chromosomes form territories with functional relevant positioning within the cell nucleus and that chromosomal subdomains exist has been also determined to a fair degree of detail. Only recently, however, we were finally able to fill the much debated gap in-between by establishing that nucleosomes compact into a quasi-fibre folded into stable loops which form stable multi-loop aggregates/rosettes connected by linkers and hence creating chromosome arms and entire chromosomes. Interestingly, this has lead immediately to a consistent and cross-proven systems statistical mechanics genomic framework which is balancing stability/flexibility ensuring genome integrity, enabling expression/regulation of genetic information, as well as genome replication - all this in evolutionary perspectives as the natural outcome of Darwinian natural selection and Lamarkian self-referenced manipulation. Thus, genotype and phenotype are multilistcly entangled and beyond are embedded in genome ecology i(!)n- and environments. This not only opens the door to a true universal sequencing of genetic information, but also is the key for a general understanding of genomes, their function and evolution, as well as for applied diagnostics and treatment of disease, for future genome manipulation and engineering efforts, as far as the creation of artificial or extra-terrestrial live contexts.

Keywords: chromatin structure, genome organization, systems genomics, genomic statistical mechanics, genotype phenotype entanglement, evolution, genome ecology

1. Introduction to the History and State of the Art

Inheritance has always played a central part in the quest for elucidating the origin of nature, life and mankind. Beyond the epic mythical assumptions, it also has been obvious for millennia that the evolutionary transfer of information plays a key role during the manipulation of inheritance by mating and breeding. Already in antique times many a "theory" was devoted to the apparent, as well as especially to the obvious fact that nature seemed to be composed of small, similar, and consistent subcomponents—so called atoms. With the description of the tissue of plants (including its substructures of vesicles and bubbles) by Robert Hooke or in the case of the cell nucleus by Anton van Leeuwenhook, in the 17th century new momentum entered the field. Nevertheless, it took until 1830 when Robert Brown defined the cell nucleus as such and until 1939 when Theodor Schwann established the cell as the fundamental unit of all plant and animal tissues while linking to the assumed fundamental design principle of life as well as nature in general. Despite fast growing microscopic resolutions there were huge challenges: not only staining and visualization methods were lacking, but also huge preparatory issues were faced especially concerning the "notorious" hard to stain cell nucleus. With the development of the natural sciences many a discovery was made culminating in the structural description of the DNA double helix [1] and the discovery of the nucleosome [2–4] at the atomic level, full genome sequences and finally histone modifications defining epigenetic landscapes. It also became obvious that the structure and function of genomes co-evolved as an inseparable system allowing the physical storage, replication, and expression of genetic information [5–7].

However, the immense size and structural complexity of genomes spanning many orders of magnitude has always imposed huge experimental challenges. Thus, the higher-order architecture has been and still is widely discussed with many interesting details yet to be described. Already how nucleosomes are spaced, positioned, remodelled, and whether and how nucleosome chains fold into fibres at physiological salt concentrations have been matters of continuing debate: e.g. Finch and Klug [8] proposed a relatively regular solenoid and *in vivo* neutron scattering experiments revealed a fibre diameter of 30 ± 5 nm as a dominant nuclear feature [9–12]. In contrast more recent work suggested no compaction at all (rev. [13, 14]), and highly polymorphic, nucleosome position- [15] and dynamic function-dependent structures [16, 17], which are essential to explain nucleosome concentration distributions [18–20], or dynamic and functional properties such as the nuclear diffusion of macromolecules. Moreover, the fine-structured multi-scaling long-range correlation behaviour of the DNA sequence also predicts a compacted chromatin fibre [21–24]. With a novel chromatin interaction technique—*T2C*—we were, however and indeed, able to show that nucleosomes form in general a quasi-fibre with a differential compaction of $\sim 5 \pm 1$ nucleosomes/11 nm [25, 26], which is in agreement with a novel *in vivo* fluorescence correlation spectroscopy (FCS) approach measuring the dynamics of chromatin [27].

The higher-order chromatin architecture has been a matter of even greater debate: Pioneering light microscopy studies by Rabl [28] and Boveri [29] hinted towards a hierarchical self-similar, territorial organization. Electron microscopy suggested a more random

interphase organization as in the models of Comings [30, 31] or Vogel and Schroeder [32]. In the radial-loop-scaffold model of Paulson and Laemmli [33] ~60 kbp-sized chromatin loops attached to a nuclear matrix/scaffold explained the condensation degree of metaphase chromosomes. According to Pienta and Coffey [34], these loops persisted in interphase and formed stacked rosettes in metaphase. Micro-irradiation studies by C. Cremer and T. Cremer [35, 36] and fluorescence *in situ* hybridization (FISH) by Lichter [37] as well as C. Cremer and T. Cremer [38] and publications thereafter [39, 40], confirmed a territorial organization of chromosomes, their arms, and stable sub-chromosomal domains during interphase, including their structural persistence during metaphase (de-)condensation. The assumption since then has been that the ~850 G, Q, R, and C ideogram bands [41, 42] split into and thus also consist actually of ~2500 subchromosomal interphase domains. Chromatin rosettes explaining a (sub-)territorial folding were first visualized using electron microscopy by Jekatrina Erenpreisa [43] and others [44] but remained unappreciated, until Belmont and Bruce proposed the EM-based helical hierarchy chromonema fibre (CF) model [45]. Spatial distance measurements between small FISH-labelled genetic regions, led to the Random-Walk/Giant-Loop (RW/GL) model with the first analytical looped polymer description by Sachs [46–48]. Here, 1 to 5 Mbp loops are attached to a non-protein backbone, following the line of Pienta and Coffey [34]. Later, a combination of distance measurements using structure-preserving FISH protocols, high-resolution microscopy, and huge parallel polymer simulations of chromosomes and entire cell nuclei, only were compatible with the rosette-like Multi-Loop-Subcompartment (MLS) model in which around 60 to 120 kbp loops form rosettes connected by similar sized linkers [7, 21–24, 49, 50]. Thereafter, the RW/GL model has then been discussed in terms of methodological “demolition” of the architecture [21, 22, 51, 52]. This is also in agreement with studies on replication (see [39] and thereafter). Again *in vivo* FCS measurements of nucleosome concentration distributions and dynamic and functional properties such as the diffusion of macromolecules are only in agreement with a small multi-loop aggregate/rosette-like chromatin folding [18–20, 22, 53, 54]. The fine-structured multi-scaling long-range correlations of the DNA sequence once again also predict this [22–24, 55].

To further distinguish between the different architecture proposals, proximity crosslinking techniques (developed and used already in the last century) were further developed into a family of interaction capture techniques such as 3C [56, 57], 3C-qPCR [58], 4C [59], 3C-seq/4C-seq [60], 5C [61], and Hi-C [62]. They once more confirmed the existence of looping and sub-chromosomal domains, now inconsistently referred to as topologically associated domains (TAD; [63]) with a somewhat higher localization accuracy when compared to FISH. These approaches also led to a number of - although by the underlying (raw) data basically unsupported - conjectures (Imam et al., in preparation), e.g. the fractal globule model [62], the loop array architecture of mitotic chromosomes [64], and the highly dynamic loop formation based on single-cell experiments [65] or in a genome wide assay [66]. In contrast, with the introduction of targeted chromatin capture T2C [25, 67–69], we were able to show that the chromatin quasi-fibre forms small stable loops of ~30-100 kbp which form stable multi-loop aggregates/rosettes connected by linkers of similar sizes as the loops [25, 26]. The development of our novel *in vivo* FCS approach came to the same conclusion [27].

2. Finalizing the 3D Genome Architecture & Dynamics

Heuristically, it is very instructive how the central part of the 3D genome architecture and dynamics could now be determined by us in detail, and how out of this process immediately an also evolutionary consistent model (**Figure 1**) arises in agreement with the entire history and heuristics of the field. This has been achieved by a highly integrated systems approach linking holistically: i) a novel high-quality selective high-throughput high-resolution chromosome interaction capture (T2C) technique [25, 26, 67–69] (elucidating the structure with unprecedented resolution of some base pairs), ii) a novel *in vivo* FCS approach [27] exploring the structure and dynamics by measuring chromatin movement, and iii) a novel analytical approach [27] and improvement of super-computer simulations of individual chromosomes and entire cell nuclei [7, 21–24, 26, 49–52, 70] to predict, analyse, and interpret the 3D architecture and dynamics from a theoretical standpoint, and combining all these with iv) scaling analysis of the 3D-architecture [21, 22, 26] and the DNA sequence itself [22, 24, 26] since the architecture and its dynamics leaves sequence "footprints" due to the co-evolutionary entanglement of structure and sequence. The combination of these resulted not only in a consistent

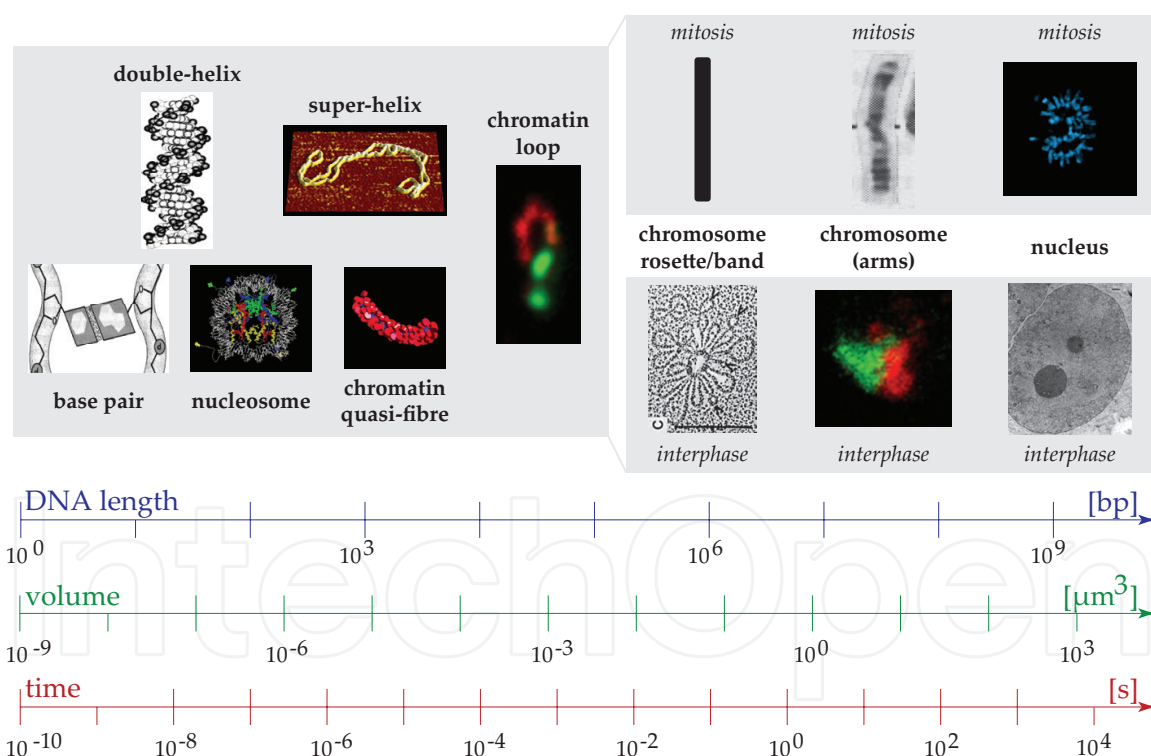


Figure 1. Overview on the size and time scaling of genome organization: The scaling and the levels of organization range over 9, 12, and 14 orders of magnitude! Initially base pairs are formed composing the DNA double helix (image see [22]), forming with a histone core complex the nucleosome (image from [22]), which condense into a chromatin quasi-fibre (simulation image; courtesy G. Wedemann). The DNA double helix forms also superhelices (AFM image of plasmid DNA; courtesy K. Rippe). The next compaction step consists of stable chromatin loops (FISH image; courtesy P. Fransz) forming stable loop aggregates/rosettes connected by a linker (EM image from [44]), which make up interphase chromosome arms and territories (FISH image; courtesy S. Dietzel) and the metaphase ideogram bands (image see [22]). 46 chromosomes compose the human nucleus and are decondensed in interphase (EM image; courtesy K. Richter) and condensed for separation during mitosis (image from [22]).

model for genome organization, but re-evaluation of the development of the entire field in the last ~170 years fostered this conclusion also tremendously and directly resulted in an evolutionary consistent model of genome organization in general.

2.1. Detailed Structure Determination by T2C

To finally determine and structurally sequence with highest resolution, signal-to-noise ratio, interaction frequency range, and statistical significance the 3D genome architecture we developed targeted chromatin capture (T2C) - a chromatin interaction technique though with far-better quality specifically addressing the needs for genome architectural "sequencing" [25, 26 67–69]. Briefly: i) after chromatin crosslinking, ii) cell permeabilization for intra-nuclear enzymatic DNA restriction, iii) the extracted and largely diluted cross-linked DNA is re-ligated primarily within the crosslinked complexes. After iv) de-crosslinking, purification, and final shortening to <500 bp of the chimeric DNA ligates, v) a purified region-specific DNA interaction fragment library is selected by using DNA capture arrays, before finally vi) high-throughput sequencing, mapping to the reference genome, interaction partner determination and visual/quantitative analysis is conducted (**Figure 2**). Notably, we use only uniquely mapped sequences without applying any other corrections

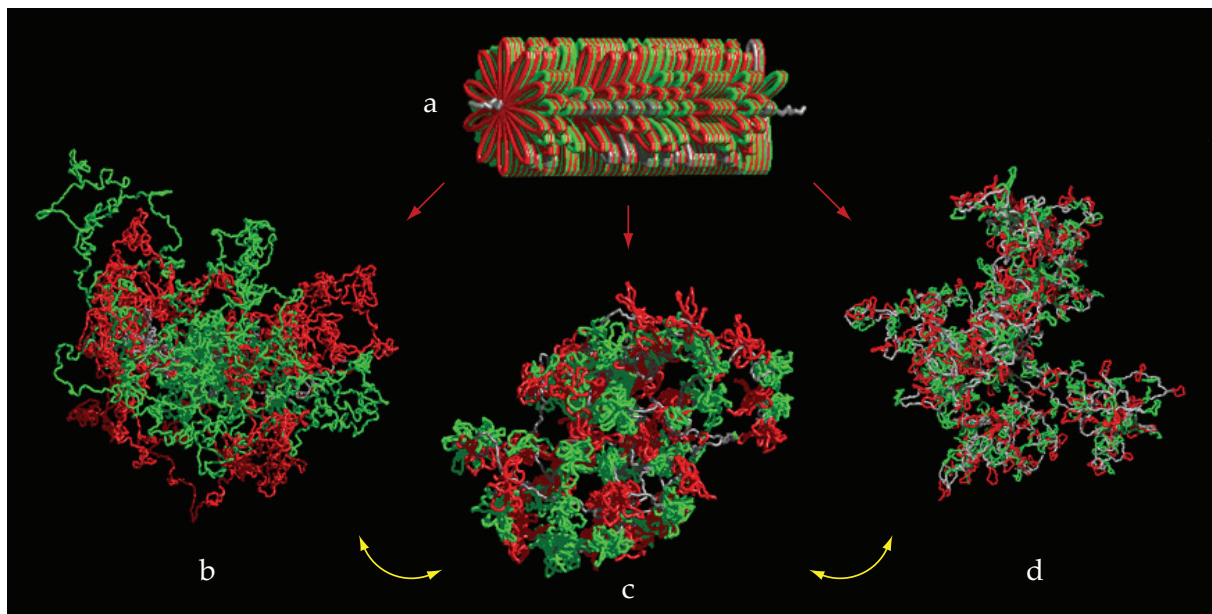


Figure 2. Simulated chromosome models [7, 21–23, 26, 49–52]: Volume rendered images of simulated Random-Walk/Giant-Loop (RW/GL) and Multi-Loop-Subcompartment (MLS) models. As a starting conformation with metaphase chromosome form and size (top), rosettes were stacked (a). Thereof, interphase chromosomes in thermodynamic equilibrium, were decondensed by Monte-Carlo and relaxing Brownian Dynamics. The simulated RW/GL model containing here large 5 Mbp loops notably shows that the large loops do not form distinct structures but intermingle freely (b). In contrast, in the MLS model with 126 kbp loops and linkers, the rosettes form distinct subchromosomal domains and chromatin territories in which the loops do not intermingle freely (c). In an RW/GL model with 126 kbp loops and 63 kbp linkers, again distinct chromatin territories are formed but in contrast to the MLS model without subchromosomal domains (d). It is obvious that the MLS model not only balances stability and flexibility considerations in storage and transcriptional respects, but also is optimal for replication due to its in essence two-dimensional topology allowing controlled duplication and separation during mitosis.

bearing information loss due to the very nature of T2C. This specific setup is not only far superior due to its improvement of 3 to 4 orders of magnitude compared to other interaction approaches (see Introduction), but also allows nearly unlimited opportunities e.g. such as multiplexing for complex research and diagnostics.

Most importantly, however, T2C allows reaching fundamental resolution limits where "genomic" statistical mechanics and uncertainty principles apply [26]: With fragment length and thus resolutions of a couple of base pairs, a high interaction frequency range, and high signal-to-noise ratio, not only molecular resolution is reached and thus the fundamental limits of cross-linking techniques, but also the mechanism of observation is now on the same scale as the observables (in analogy to classic and quantum mechanics). Actually due to the stochastics following the bias of the system behaviour, the observables, the observation, and thus the measured values are constrained by what we call "genomic" statistical mechanics with corresponding uncertainty principles. This originates from the individual complexity of each highly resolved interaction with a unique but coupled individual probabilistic fragment setting in each cell at a given time. Hence, the actual conditions and components can be determined only partially with high accuracy while with low accuracy otherwise and are eventually even entirely destroyed by the measurement. Thus, the central limit theorem applies with an overlap of system inherent and real noise stochastics, and hence in the end only probabilistic analyses and statements can be drawn as hitherto is well known from classical mechanics, and more so from quantum (mesoscopic) systems. Consequently, population based or multiple single-cell experiments have to be interpreted and understood in a "genome" statistical mechanics manner with uncertainty principles due to the inseparability of factors/parameters also seen there. Thus, in practical terms, valid results are obtained when the statistical limit is reached, i.e. when scaling up the experiment does not narrow down the distribution any further and does not lead to fundamental (overall) changes anymore in observables. Nevertheless, if the statistical limit is reached and if the quality parameters like resolution, frequency range, and signal-to-noise ratio are sound, conclusions could be drawn as in the many cases of classic mechanics, and more so of quantum (mesoscopic) systems.

Consequently, due to this sensitivity of T2C, we [26] were able to determine finally the missing parts of the 3D architecture on scales where a "genomic" statistical mechanics applies with stable reproducibility as one can already see visually in colour coded interaction maps (**Figure 2**): Not only are rare interactions stably detected within an unprecedented frequency range spanning 5-6 orders of magnitude, but also the maps are reproducibly mostly empty (<10% of possible signals are taken). Both interactions and non-interactions show clearly dedicated interaction patterns on all spatial scales within and between domains, including their re-emergence as attenuated repetition on other scales since obviously genomes are scale-bridging systems [22, 23]—all of which can be immediately identified as structural features - briefly (**Figure 2**):

- i. On the largest genomic and thus spatial scale, subchromosomal domains are visible as square-like interaction domains (often unfortunately called TADs; [63]) featuring in general a higher average uniform interaction degree compared to interactions between domains, with a sharp drop at the edge of domains, as well as a clear linker region

between the domains that connects them. The borders of the domains can be determined down to the single fragment level and thus a very high resolution (see below). The interaction of domains with each other and a closer inspection of the interactions in the vicinity of the linker interacting often more frequently compared to other domain parts are mainly due to the breaking of spatial isotropy.

- ii. At intermediate scales within the subchromosomal domains, the interaction pattern shows clearly distinct gaps and a quantifiable grid-like arrangement of interactions, which also continues outside and “crosses” with the linear pattern originating from sequentially subsequent domain(s). These interactions on scales of tens of kilo base pairs are doubt-free originating from stable chromatin loops, forming a stable loop aggregate/rosette like architecture, due to several consecutive loops coinciding.
- iii. On the smallest scale, a dense and high interaction frequency pattern is observed in the region from 3 to 10 kbp (i.e. $< \sim 5$ -15, and ~ 50 nucleosomes, respectively) along the diagonal. It varies independently of the local fragment size with distinct interactions and non-interacting “gaps”. This suggests, that there are defined stable interactions on the nucleosome scale forming an irregular yet locally defined and compacted structure, i.e. a quasi-fibre with average properties (e.g. an average linear mass density).

A detailed quantification [26, 27] of several regions leads to a quasi-fibre compaction of 5 ± 1 nucleosomes per 11 nm, with an average chromatin quasi-fibre persistence length of ~ 80 to 120 nm, loops and linkers of ~ 30 to 100 kbp, forming multi-loop aggregates/rosettes with typically 300 kbp to 1.5 Mbp subchromosomal domain sizes. Different cell types, species, or functional conditions showed only a relatively small variation of this theme [26, 27].

All this is consistent with a variety of previous observations and predictions such as compacted fibre structures described throughout the literature (see e.g. [16, 17]), the internal structure of subchromosomal domains [7, 21, 22, 24, 38–40, 43, 49, 50] agreeing on all structural levels with the absolute nucleosome concentration distributions [18, 19], the dynamic and functional properties such as the architectural stability and movement of chromosomes [7, 22, 54, 71, 72], chromatin dynamics [73], as well as the diffusion of molecules inside nuclei (e.g. [22, 54, 72]), and recent genome wide *in vivo* FCS measurements of the chromatin quasi-fibre dynamics [27] also suggesting such a chromatin quasi-fibre with variable, function-dependent properties. Beyond, other hypothesis (see Introduction; [26, 27]) about the 3D genome organization on these scales can clearly be ruled out: e.g. no-compaction or a highly-regular chromatin fibre, unstable/dynamic loops or unstable/dynamic loop aggregates/rosettes can clearly be ruled out, because they simple would lead to other interaction patterns and the intrinsic chromatin fibre dynamics with movements on the milli-second scale (Movies 1, 2 [26]) would lead to immediate structural dissolution. Most importantly no other model leads to a consistent functional framework bridging consistently the here described scales as can also be shown by the agreement with scaling analysis of the 3D-architecture [21, 22, 26] and the DNA sequence itself [22, 24, 26]. Beyond, not only functional aspects as the easy (de-)condensation during mitosis can be easily explained, but we were also able to find this organization in the data of others across species and even across specie-kingsdoms (Imam et al., in preparation).

2.2. Dynamics and Structure Revealed by FCS

To investigate the 3D genome architecture and dynamics also by an orthogonal genome wide and *in vivo* approach, a novel *in vivo* FCS technique exploring the structure and dynamics by measuring chromatin movement combined with a novel analytical approach was introduced [27]. It is based on the fact that a specific chromatin quasi-fibre and its higher-order architecture directly influences its intrinsic dynamics. Thus, the concept dissects intra-molecular polymer dynamics from fluorescence intensity fluctuations measured with FCS to investigate meso-scale chromatin dynamics in living cells and connects this to the underlying three-dimensional organization. Besides, the classical analytical polymer models were extended to include dynamics, physical properties, and accessibility. As primary tracer protein for chromatin movement a linker histone H1.0-EGFP construct was chosen [18, 19, 22]. On the one hand, H1.0 decorates chromatin globally and reflects its density. On the other hand, it binds only transiently such that photobleached molecules are constantly replaced by fluorescent ones, and thus chromatin dynamics becomes amenable to FCS analysis (see also [20, 54]): Here, topologically and dynamically independent chromatin domains of 500 kbp to 1.5 Mbp in size were identified that are best described by a compacted chromatin fibre and a loop-cluster polymer model under theta-solvent conditions. In more detail again the formation of stable loops and stable multi-loop aggregates/rosettes from a chromatin fibre with certain density and flexibility properties emerged as prominent structural feature of dynamically independent domains - and this throughout the cell nucleus in living cells! The detailed quantitative values for the involved parameters again lead in essence to the same values as found already in the T2C data: a quasi-fibre compaction of 5 ± 1 nucleosomes per 11 nm, with an average persistence length of ~ 80 to 120 nm, and loops and linkers of ~ 30 to 100 kbp [27]. Notably, it cannot be stressed enough that the loops and multi-loop aggregates/rosettes form *stable* entities on the time scales which were approachable by FCS (between 10 μ s and 10 to 20 s) and do neither open, close, or in any other way reform (longer timescale up to hours are historically known). This not only moves many an assumption currently proposed (see Introduction) into the realm of fairy tales—conceptually and by hard experimental facts in agreement with the research of the last ~ 30 years (e.g. [18–20, 22, 54, 71]). Visualization of simulated structures illustrates this clearly (Movies 1, 2 [26]): structures described consistently throughout the literature would dissolve immediately - what has never been observed (though attempted to be measured) - and also in consistent agreement with the T2C results measured at the limit of resolution. Beyond, also characteristic variations were found between eu- and heterochromatin: Hydrodynamic relaxation times and gyration radii of independent chromatin domains are larger for open (161 ± 15 ms, 297 ± 9 nm) than for dense chromatin (88 ± 7 ms, 243 ± 6 nm) and increase globally upon chromatin hyperacetylation or ATP depletion. Thus, functional changes are a variation of a basic theme, e.g. more compact heterochromatic domains have a larger inaccessible volume fraction than more open euchromatic ones. Nevertheless, molecular diffusion is fast enough to roam a complete domain within few microseconds, during which the domain itself appears static. Relaxation of domains in the 100 ms range affects genome access in a protein concentration-dependent manner: highly abundant molecules at several 100 nM concentrations ‘fill’ the fluctuating domain so that a larger volume fraction than for a static TAD becomes adiabatically accessible. In contrast,

for low-abundance molecules encounters with specific loci within a domain are diffusion-limited. They sense a higher inaccessible volume fraction. Thus, domain dynamics result in a concentration-dependent differential accessibility that is more pronounced in heterochromatin than in euchromatin due to its shorter relaxation times [20, 22, 27, 54]. In this manner the FCS approach can be extended to acquire complete nuclear maps and thus to "sequence" the dynamic organization of nuclei in living cells.

2.3. Analytical and Computer Simulations Theoretic Evaluation

To better understand the 3D genome organisation suggested e.g. by the above results, to evaluate hypotheses, and to plan future experiments, we were the first who have - since 1996 - developed polymer models with pre-set conditions for *in silico* super-computer simulations (i.e. without attempting to fit data; [7, 21–23, 26, 49–52, 70]) and later also an analytical mathematics framework [27]. The simulations use a stretchable, bendable, and volume excluded polymer (hydrodynamic) approximation of the 30 nm chromatin fibre consisting of individual homogenous segments with a resolution of ~1.0 to 2.5 kbp while combining Monte Carlo and Brownian Dynamics approaches (**Figures 2–4**). The analytical polymer approach extends and applies for the first time Gaussian chain and Kratky-Porod model descriptions in combination with the Rouse and Zimm models for polymer dynamics to complex star and rosette topologies under real excluded volume conditions as well as dilute and semi-dilute solvent conditions [27]. Whereas the analytical model is exact, the simulations explore emerging effects not explicitly introduced into the analytical model.

Simulations (**Figure 2**) of the Random-Walk/Giant-Loop model in which large individual loops (0.5–5.0 Mbp) are connected by a linker resembling a flexible backbone, as well as the Multi-Loop Subcompartment (MLS) model with rosette-like aggregates (0.5–2 Mbp) with smaller loops (60–250 kbp) connected by linkers (60–250 kbp), have already predicted that only an MLS model, i.e. a compacted quasi-fibre forming stable loops and stable loop aggregates/rosettes connected by a linker, can properly explain the formation of chromosome arms and territories [22], the spatial distances measured both using fluorescence *in situ* hybridization (FISH) experiments [7, 21–23, 26, 49–52, 70], and beyond even the general morphology of nuclei *in vivo* using histone fluorescence fusion proteins [22, 51], nucleosome concentration distributions, as well as dynamic and functional properties such as the diffusion of macromolecules [18, 19, 22, 53, 54]. These models also contained already enough information/aspects to cover other architectures such as free random-walks, random or fractal globules as well as their stability and dynamics. Additionally, the visualization (**Figures 2–4**, Movies 1, 2 [26]) creates an immediate feeling for the behaviour of genomes in 3D - a fact which already by pure visual inspection rules out many of the introduction mentioned obscure suggestions immediately.

With the unprecedented quality of both the interaction mapping by T2C and the FCS dynamic measurements (see above) the introduction of simulation and analytical models complex enough to approximate the 3D genome organization adequately showed even more clearly that only a quasi-fibre, stable loop, stable loop aggregate/rosette-like architecture is compatible with the measurements: In essence the simulations and analytical models describe even the slightest details of the T2C and FCS measurements correctly including many at first sight

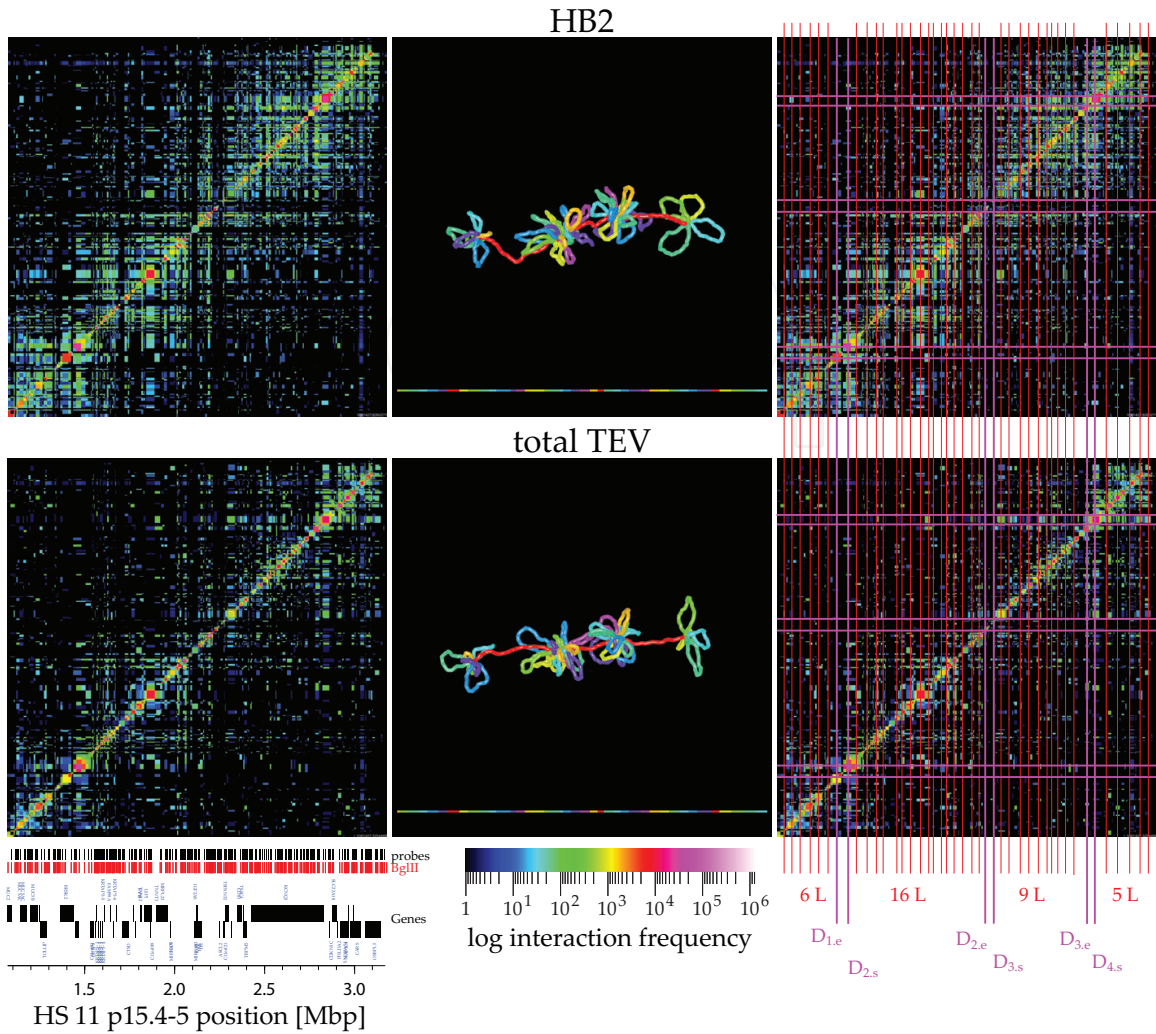


Figure 3. Determination of the 3D architecture in the IGF/H19 11p 15.5-15.4 region by T2C interaction mapping and computer simulations: Interaction matrices (logarithmic and colour coded scale; left & right) in HB2 and HEK293T TEV cells [26] show in unprecedented clarity the formation of a quasi-chromatin fibre, folding into stable loops (red lines; xL: number of loops), forming due to the grid-like pattern stable multi-loop aggregates/rosettes, i.e. subchromosomal domains separated by a linker (borders: pink lines, right; D1s, D1e: start and end of domains). A grid-like pattern is also visible in the interactions between the domains and corresponds to trans-domain loop interactions. The aggregation into a chromatin quasi-fibre is visible near the diagonal and loop internal structures are also detectable. Between different cell types or functional states only some local differences are visible resulting in a consensus architecture and allowing simulation of the 3D architecture (middle; resolution < ~1 kbp). Note that the simulation is driven by the dominant consensus architecture.

paradoxical results as e.g. i) that high numbers of especially small loops in a rosette result due the high density in steric exclusion and thus stretched loops eventually even “shielding” inner-rosette parts, ii) that inter-domain interactions are influenced by the connecting linker, loop size and numbers, and how non-equilibrium effects would appear, as well as iii) the isotropy breaking of consecutive subchromosomal domains as seen in the interactions at the border of domains and the domain-domain interactions. On a more general level the simulations support also the large and at first sight remarkable emptiness of interaction matrices and its link to the existence of a dedicated chromatin quasi-fibre. Additionally, the simulations hint

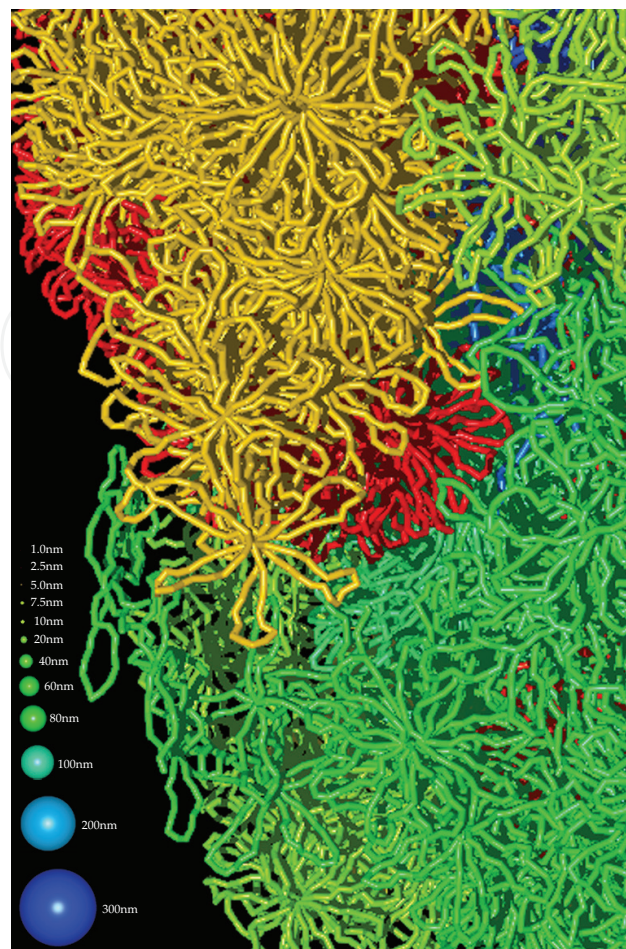


Figure 4. Insight into the spatial and dynamic/diffusional properties and morphology of the 3D organization of entire nuclei: The detailed view from the outside into a simulation for an MLS model with 126 kbp loops and linkers [22] shows the structure and low overlap of chromosome territories, the rosette like subchromosomal domains, and that the mean spacing between quasi-fibres ranges at least from 50 to 100 nm. Hence, the obstruction of diffusing particles (see spherical legend) is proportional to their size. Thus, small molecules as nucleotides and most (subunits of) proteins or gene transcripts reach every location of the nucleus by moderately obstructed diffusion. Consequently, active transport of molecules should be restricted to few exceptions and a channel like network for transportation (proposed by the Inter Chromosomal Domain model) is not necessary. Nevertheless, obviously the rosette core is denser leading to a general diffusion limited access. Thus, the interplay between accessibility and obstruction while considering the fast Brownian dynamics of the quasi-fibre and the entire system is of functional relevance.

to a relatively low crosslink probability, radius, and frequency in experiments comparing the clearly visible fine-structure (such as the (anti-)parallel neighbouring of the chromatin quasi-fibre at loop bases [26]). Also both the simulation and analytical approach describe in detail every aspect of the experimentally found multi-scaling behaviour with a fine-structure not only of the architecture and dynamics, but also of the DNA sequence (see below) to a degree of detail even we are still astonished about. The stability of the architecture with respect to the intrinsic chromatin fibre dynamics can also be illustrated by e.g. the decondensation from a mitotic chromosome into interphase (Movie 1 [26]) or just in a normal interphase state (Movie 2 [26]). This also shows that any 3D architecture would dissolve within seconds if it would not be stabilised. Consequently, both theoretic approaches came with old and new data

consistently to the same conclusion whatever orthogonal high-quality method is used and thus are a theoretical framework for the understanding, test, and engineering of genomes.

2.4. DNA-Sequence Fine-Structured Multi-Scaling

Since what is near in physical space should also be near (i.e. in terms of similarity) in DNA sequence space and this presumably genome-wide [22–24, 55], and because evolutionary surviving mutations of all sorts will be biased by the genome architecture itself and vice versa, the correlation and thus scaling behaviour of the DNA sequence [22–24, 26, 55] and its connection to the 3D genome architecture scaling - either from T2C interaction mapping [26] or from simulations [21–23] - allows for comprehensive investigation of genome organization in a unified scale-bridging manner from a few to the mega base pair level. Using to this end, the perhaps simplest correlation analysis possible (to avoid information loss or biases), we calculated the mean square deviation of the base pair composition (purines/pyrimidines) within windows of different sizes and calculating the function $C(l)$ and its local slope $\delta(l)$, which measures the correlation degree, or in more practical lay-men terms, is similar to a spectral measure [22–24, 26]: in relation to mammalian genome organization for each of two different human and mouse strains i) long-range power-law correlations were found on almost the entire observable scale, ii) with the local correlation coefficients showing a species specific multi-scaling behaviour with close to random correlations on the scale of a few base pairs, a first maximum from 40 bp to 3.6 kbp, and a second maximum from 8×10^4 to 3×10^5 bp, and iii) an additional fine-structure is present in the first and second maxima. The correlation degree and behaviour within the species are nearly identical comparing different chromosomes (with larger differences for the X and Y chromosomes). The behaviour on all scales is equivalent concerning the different measures used to investigate the long-range multi-scaling of the genome architecture with the transitions of behaviours even at similar scaling positions [26] and can be associated with a single base pair resolution i) the nucleosome, ii) the compaction into a quasi-fibre, iii) the chromatin fibre regime, iv) the formation of loops, v) subchromosomal domains, and vi) their connection by linkers. Additionally, the already previously proven association to nucleosomal binding on the fine-structural level [22–24] is not only found again, but also is in agreement with the fine-structure found in the interaction scaling. Since the correlation analysis is genome-wide (in contrast to the T2C analysed regions so far) and since individual chromosomes show a highly similar scaling this clearly shows the genome-wide validity of the 3D organization. Moreover, the existence and details of this behaviour show the stability and persistence of the architecture since sequence reshuffling or other destructive measures would result in a loss of this pattern. This would also be the case for an unstable architecture, which would not leave a defined footprint within the sequence. This is again in agreement with our simulations of the dynamics or the genome wide *in vivo* FCS measurements [27]. Consequently, this shows not only by two analysis of completely independent “targets” (the T2C interaction experiments and the analysis of the DNA sequence) the compaction into a chromatin quasi-fibre and a stable multi-loop aggregate/rosette genome architecture again, but proved here also the long discussed notion that what is near in physical space is also near, i.e. more similar, in sequence space. Hence, the 3D architecture and DNA sequence organization are co-evolutionarily tightly entangled (review

of previous notions in [22, 24]). Thus, in the future from the DNA sequence and other higher-order codes (e.g. the epigenetic code) most architectural genome features can be potentially determined, since most structural/architectural features left a footprint on the DNA sequence and other code levels and vice versa as one would expect from a stable scale bridging systems genomic entity.

3. Systems Consistency of the 3D Genome Organization

The above described holistic combination of several new orthogonal approaches [26, 27] including the heuristics of the field leads interestingly undoubtedly to a consistent picture of genome architecture, dynamics, and in general organization, by establishing that nucleosomes compact into a quasi-fibre folded into stable loops, forming stable multi-loop aggregates/rosettes connected by linkers creating chromosome arms and entire chromosomes. Nevertheless, the heuristics of the field immediately questions whether i) we really now have an evolutionary consistent picture of genome organization, ii) whether this is the unavoidable outcome of Darwinian natural selection and Lamarkian self-referenced manipulation (what we introduce here), and iii) finally whether we can understand now genome organization in its systems context within cells, organs, and the entire organism? This in essence already relates back to the fundamental question of how life emerged from the primordial soup [5, 6, 22]; see details in following sections) but in the context discussed here can be addressed by first reflecting on the existing major functions of genomes, thus setting the stage: i) genomes need to stably store genetic information, ii) the information needs to be differentially read out to give rise to and regulate the molecular machinery, and iii) genomes need to replicate and mutate to spread and evolve:

- i. Obviously the by far most important function is to stably store over long periods of time genetic information though with enough flexibility including mutations - or in short: without proper storage neither information retrieval, nor replication, nor evolutionary development exist. This involves obviously being resistant against physical/chemical and/or in- or external mechanical destruction. Whereas, the first act mainly as from the bottom up involving one or a group of chemical bonds in proximity by direct interactions in the molecular soup, the latter depends on the large-scale structure of the basic molecular components and thus acts indirectly top-down on chemical bonds, i.e. that in- or external global stress is transferred and eventually accumulated via the global structure down to molecular levels while leading to mechanical failure. Both this physico-chemical and structural conformation-based destruction paradigms, influence genome architecture on all its levels under evolutionary pressure. They can be formulated such that a) mechanical failure rates are minimized regarding very long time spans, and b) in- or external mechanical failure rates reach an optimum due to the right balance between internal stability increasing with scale (for sensible ranges) and external stress decreasing the stability with increasing scale. From the well known average DNA breaking length of ~300–500 bp after already relatively severe sonication, this translates right away to the nucleosome and chromatin quasi-fibre level assuming that internal

nucleosomal attachment increases the stability and elongating it by a factor 146 bp to 200 bp (repeat length), i.e. the average breakage length of an uncompacted chromatin fibre is 44 kbp or in the extreme 100 kbp balancing the quasi-fibre internal stability increase by further compaction counteracted by the bigger mechanical susceptibility due to local compaction clusters. Thus, the found loops size of 30-100 kbp as well as its chromatin quasi-fibre persistence length of 80-120 nm is just what one would theoretically expect as the evolutionary outcome. The same holds for the formation of stable multi-loop aggregates/rosettes where the major player is internal stability, which is a function of quasi-fibre compaction, loops sizes, and loop numbers [51, 52], giving rise to the natural found size distribution between ~0.3-1.5 Mbp [21–24, 26, 27, 40–42]. Also on the entire chromosome level again in- and external stability criteria have reached an optimum during evolution concerning the number of subchromosomal domains as well as their total size and number within a genome which again would just fit what one would theoretically expect: subchromosomal domain linkers are in the ballpark of loop sizes, the number of subchromosomal domains is <200-300 which just is the optimum size where mechanical stress does not too much destruct mitotic chromosomes under normal conditions. Consequently, the stability criteria are clearly satisfied while obviously still allowing enough flexibility by variation of this theme within the relatively broad boundary limits and various levels compensating individual stretching of limits (e.g. bigger loops might be stabilised by higher quasi-fibre compaction). Beyond, destruction of a complete structural element (e.g. nucleosome, loop) in relation to the characteristic scale seems never really to exceed 1-5% - an important criterion for overall system resilience.

- ii. Access to and obstruction of genetic information, i.e. genetic information retrieval in a regulated fashion is, of course, next to pure storage the major task for a genome, although without a stable information storage retrieval gets arbitrarily complicated whether replication takes place or not. Since the information is readout with similar means as the storage itself, i.e. in a molecular way in contrast e.g. to an optical readout, this relies in principle on two major conditions: a) the physical space for the regulation of the 3D architecture needed that a readout takes place, and b) accessibility/obstruction to the genetic information for the readout-machinery as well as post-processing and transport of the transcribed information. For the first the DNA, nucleosomes, chromatin quasi-fibre, loops and loop aggregates/rosettes, need to have the space to be modified and get rearranged, i.e. that a volume several times bigger than the actual structure exists for ease of change. This involves, naturally a certain compaction, since a homogenous soup would not allow this. Since the regulation and readout is done by molecular mechanisms, it is also obvious that a low spatial occupancy allows moderately obstructed diffusional access of both the regulation and readout machinery only for DNA with a certain compaction degree. For such a scenario the volume occupancy of the architecture in aqueous solution should be well (!) below the limit of ~50% (model depending) as known from percolation studies [74], i.e. in terms of the performance expected for genomes, volume occupancy should be <10% since both the genomic architecture as well as the machinery should be able to access it for regulation by modification as well as readout. For chromatin, experimental values are between 2.5% to ~8% with a homogenous mesh

spacing ranging from 115 to 65 nm ([22] and literature cited therein). Together with other factors and molecules in the cell nucleus like proteins and RNA, which all have a similar density, the volume occupancy is still <25%. These percolation assumptions hold, of course, also for the dynamics of the structure itself as pointed out above. At first sight this seems to be a dense system but the architecture is moving constantly by Brownian motion like in a spaghetti soup with additional floating components [18–20, 22, 27, 53, 54]. For chemical reactions this is well known for diffusion limited aggregation processes [75] as well as for percolating systems [75]. Due to the described consistent multi-layered 3D organizations showing also a multi-scaling of its volume occupancy as well as the space in-between this creates now even more and especially a scale dependent accessibility and obstruction to enhance the theoretic predictions of homogeneous though compacted systems with percolating space. Thus, under such conditions the necessary machinery for transcription as well as transcript transport is based mainly on moderately obstructed diffusion and despite of its high overall concentrations acts as an adequate multi-scale space [22, 53]. Consequently, similar to diffusion limited (catalytic) processes modification of the intrinsic architecture and dynamics of the entire genome organization is used for locally or globally fine-tuning of processes and thus functional regulation. Concerning, the stability of the 3D architecture only a quasi-fibre with stable loop aggregates/rosettes allows in terms of stability and flexibility local containment of large-scale interactions during the initiation of transcription e.g. by enhancer promoter interactions. For knot-free replication of the genome these (spatial) arguments also apply: whereas accessibility allows access of the machinery and space for the duplication, spatial obstruction protects the structural integrity. Interestingly, none of the described alternative architectures and dynamics hypothesis (see Introduction) agree to even a sufficient degree with these fundamental necessities to guaranty genome function.

- iii. Replication and extinction of genetic information is the most crucial intervention into genome organization, since in contrast to the readout and regulation of genetic information by transcription, the entire structure and dynamics are affected by copying every single component of the organization. Here, an exact copy within a constrained space not only sequence wise, but also of its 3D architecture and dynamics as well as its disentanglement are the crucial parameters while still allowing structural stability/flexibility and even the access/obstruction of genetic information. From protein folding it is well known, that already during the amino-acid chain synthesis in the ribosome folding takes place, leading to a different 3D folding compared to the relaxation of finished and stretched out amino-acid chains. Obviously, also chromosome replication is such an adiabatic process (also chromosomes never fold from scratch, i.e. *de novo*, and always go continuously from one state to another), which takes place in parallel in the entire cell nucleus. And here again, genome architecture and dynamics are enabling replication to take place easily in principle only compatible with a chromatin quasi-fibre arranged in stable multi-loop aggregates/rosettes. This is due to the fact this architecture on the level of stable multi-loop aggregates/rosettes follows a knot-free two-dimensional topology. Of course, genome architecture is not a simple two-dimensional object in space considering the DNA-double helix and nucleosomal twist and writhe but nevertheless

in terms of replication disentanglement it is. Consequently, replication origins can be situated and start replication everywhere in each chromatin loop with replication forks leading towards both directions until they hit a loop base (which is the reason for the bidirectional CTCF sites functioning as linear DNA markers for the directional oriented replication machinery). During this procedure even the twist and writhe are copied and need to be untangled as in the case of transcription. While hitting the loop bases then the two forks coming from two loops have to be joined and untangled, but no complex network of knots as they would appear even in a Random-Walk/Giant-Loop or even more so in a fractal globule like replication scenario would have to be cut and re-joined. Again here theoretical predictions for loop size and loop numbers are just fitting the experimental findings (see e.g. [39] and thereafter). Due to the two-dimensional topology of the multi-loop aggregates/rosettes, they can just be separated very easily in 3D space (this idea was proposed and illustrated to the author by his at the time 6 year old son Leander Aurelius!). And again the compaction and volume occupancy in the cell nucleus play an important role: the compaction into a chromatin fibre reduces not only the formation of DNA knots largely (perhaps almost to zero), but also provides with the volume occupancy in the cell nucleus the room for undisturbed replication, with the right flexibility provided by the intrinsic dynamics, allowing the disentanglement of replicated structures with minimal e.g. topoisomerase/decatenase driven active processes.

In summary, the above proves even further and especially in a holistic combination with the presented new orthogonal approaches [26, 27] and including the heuristics of the field, that indeed the described 3D genome organization - DNA forming nucleosomes compacted into a quasi-fibre folded into stable loops, forming stable multi-loop aggregates/rosettes connected by linkers creating chromosome arms and entire chromosomes (**Figure 1**) - presents without doubt a consistent scale bridging systems statistical mechanics genomics fulfilling the functional conditions necessary for storage, transcription, and replication. Additionally, the actual values found for the various parameters involved are just found in those "regions" one would expect as the unavoidable outcome of Darwinian natural selection and Lamarkian self-referenced manipulation (see below).

4. A Systems Genomics Statistical Mechanics

The heuristics leading to the here described consistent 3D genome organization has also resulted in another fundamental breakthrough besides merely clarifying the missing gap(s): the emergence of a multilistic systems statistical mechanics with uncertainty principles by reaching the fundamental resolution limits (see Section 2.1 above; [26]). Hence, this allows directly not only i) to extend the atomic theory based on ancient Greek philosophy and the notion of Theodor Schwann of cells being the fundamental atomic unit of tissues to the mesoscopic scale of genome architecture/dynamics, but also ii) to analyse and to describe how from the collective behaviour of these elements a holistic meta level, i.e. a phenotype, emerges. Thus, by reaching fundamental resolution limits now the statistical and uncertainty properties of each architectural/dynamic level can be determined both by experimental

measurements as well as theoretical descriptions. Hence, from each of these "atomistic" basic units/elements their collective behaviour can be derived by a statistical mechanics on each individual level as well as a complex interwoven scale-bridging, i.e. a hierarchic back referencing networked systems statistical mechanics - which obviously exists - can now be established in detail. This exceeds and is much more complex than establishing the statistical mechanics at the turn of the 20th century where from the individual components e.g. gas molecules a statistical mechanics established the collective properties of the entire system, e.g. the entire gas, because genome organization is not only a simple dualistic system of e.g. two levels but a complex multilistic network system with back references: In detail this means determining experimentally the behaviour of a genome structural/dynamic level precisely with its entire statistics and then doing the same on the level emerging from the underlying level. In principle this is what we have started already by setting up an experimental and theoretic framework over the past 20 years to elucidate genome organization [7, 18–24, 26, 27, 49, 50], although only now with the complete description of the general 3D genome architecture/dynamics it is possible to fill the existing lack of knowledge in detail, determine the values for parameters with high precision, and in constant cycles of refinement adjust the description to an ever higher degree of approximation. Thus, the difference to the development of statistical mechanics in classical and later quantum physics at the turn to the 20th century is that in biology many and also much higher levels still are determined by and also act back even on the very first level to a much higher degree. This also immediately unites the at first sight contradicting theoretic descriptions of living systems of Ilia Prigogine [75], stating that living systems are far away from thermodynamic equilibrium, with those proposed by Georgi Gladyshev [76] stating that hierarchic substance stability is locally in thermodynamic equilibrium. Actually, these descriptions are even extended due to the multilistic statistical systems mechanics, i.e. manifold recursive hierarchically back-referencing, which are until now not described but e.g. envisioned in efforts to extend quantum mechanics to higher order complexities [77]. Consequently, a genomic multilistic statistical systems mechanics allows not only to describe and test basic properties of life, but also to answer perhaps the most fundamental questions of life as e.g. whether life time-wise can be extended beyond the currently obvious or thought of limits by manipulated engineering in one of its most central parts - the genome - a quest of epic dimensions appearing already at least between the lines in "What Is Life ?" by Erwin Schrödinger [78].

5. Genotype-Phenotype-Entanglement and Genome Ecology

The most important implication from the findings described above is most likely the multilistic entanglement between genotype and phenotype being the natural outcome of Darwinian natural selection and Lamarkian self-referenced manipulation in a genome ecology framework, which is connected directly to the origin of genomes and life itself: While entropy grows like an inexorable river, local disturbances lead to ever more ordered self-organizing and self-sustaining resistors, more complex structures, and finally life. In the 1970s Manfred Eigen [5, 6] showed how from the primordial soup autocatalytic chemical reaction-networks emerged and how they form ever more complex cooperatively organized networks and

systems of so called hypercycles. With environmental separation by the emergence of units as cells and specialization of subunits, then genomes have developed as specialized keepers of the blueprint needed to maintain, regulate, and develop this syntropic machinery. Since genetic information is physically stored in molecular structures with dedicated architecture and dynamics, it is thus also obvious that the material carrier for the storage, usage, and replication of genetic information co-evolved inseparably. Yet another inevitable consequence of our results leading to the consistent statistical systems genome mechanic framework is indeed our proof [26, 27] that architecture, dynamics, and DNA sequence are co-evolutionary unseparably entangled (in a quantum mechanical sense): All architecture/dynamics levels have not only left a footprint on the DNA sequence level but beyond also all levels have left a footprint on all other levels with an astonishing degree of detail (see Section 2.4). Consequently, the co-evolution of all levels has also co-evolved not only to a higher degree than previously thought, but also indeed as an entire system where all levels are (equally ?) determinant (**Figure 5**).

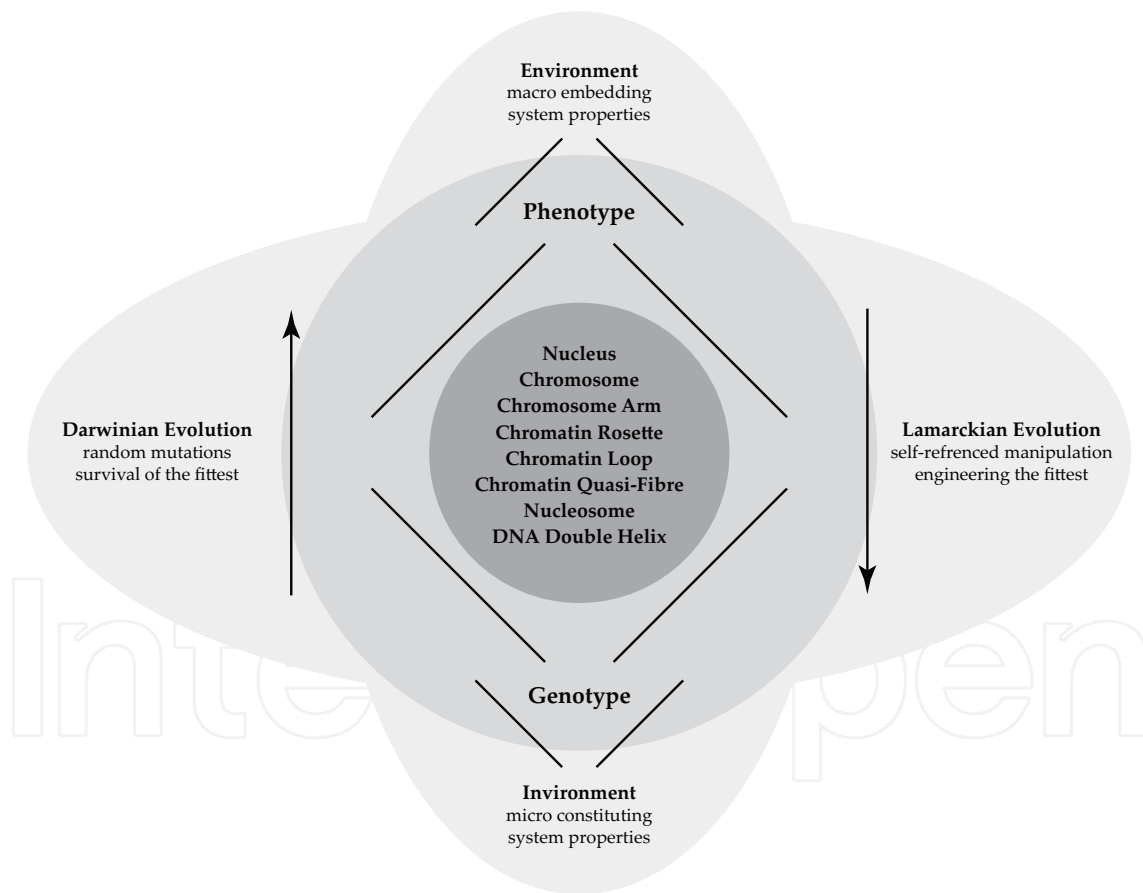


Figure 5. Genome ecology emerging from the system mechanics of genomes in relation to the genotype-phenotype entanglement and its embedding in- and environment: Genomes are interwoven holistic multi-scale hierarchic systems entities in which all organizational levels are also manifest, i.e. fingerprinting, on all other levels. Thus, immediately each level is a phenotype of its underlying genotype immediately conditioning back on it recursively. Thus, both genotype and phenotype are entangled inseparably in a (due to the involvement and entanglement of all levels) multilistic manner. In consequence this not only unites Darwinian and Lamarckian evolutionary paradigms, but also embeds and relates genomes with their in- and environment, and thus giving rise to a general genome ecology.

In evolutionary terminology the genotype (i.e. the double helix) creates a phenotype (the nucleosome) and this phenotype recursively conditions the genotype (i.e. again the double helix). The nucleosome is also a genotype conditioning the quasi-fibre phenotype, recursively conditioning the nucleosome and DNA, etc. Since this is happening with all levels simultaneously this inseparable dualism extends in the present genome organisation to a multilism, shaping evolutionary development in hierarchical terms from bottom to top by Darwinian natural selection as well as from top to bottom by Lamarckian self-referenced manipulation. Thus, our finding that indeed all genome architecture/dynamic levels are tightly entangled with each other also immediately resolves the falsely assumed paradoxes between Darwinian and Lamarckian evolution by uniting them at least on the genome level. This is remarkable not only in historic terms considering the even politically and religiously extremely hot debates/fights about "man evolving from apes" as well as the "intentionally planed long neck of giraffes", but also heuristically, since the in principle relatively simple final completion of the 3D genome architecture/dynamics at the limit of the resolution leads not only to a consistent 3D genome organization and statistical systems genome mechanics, but beyond reveals in one go some and perhaps the most important fundamentals of life (**Figure 5**).

Beyond, this strong entanglement over several orders of magnitude (**Figures 1, 2**) within the genome, the described genotype-phenotype-entanglement can be driven conceptually even further considering the influence of both the a) hierarchically constituting elements giving rise to the system, i.e. chemical molecular base, atomic, and subatomic units, which will be called here i(!)nvironment, and b) the hierarchical higher levels, i.e. tissues, organs, animal etc., which are the environment. Although this may seem far fetched, but influences from both "directions" are well known (see e.g. Section 3), although due to their complexity this is often hard to track down in a reductionistic manner, thus hence their degree of influence is just emerging. In this respect the found entanglements bridging so many multi-scale levels and orders of magnitude in space and time, are on the one hand already astonishing in terms of the obviously wrong assumption that such influences would die-off very fast, while on the other hand this has general implications for all hierarchic systems showing that complex inter-, cross-, and even multi-cross-level influences are much more frequent and far reaching. Actually, the here shown multilistic genotype-phenotype entanglement shows a highly interwoven, networked, and recursive structure: instead of more or less separate hierarchic layers where only first or at the most secondary neighbour layers are connected, there are also influential connections to more distant layers at least locally if not in every part of the layer space. Thus, the genotype-phenotype entanglement embedded within an i(!)n- and environment actually results in a genome ecology in direct analogy to e.g. human ecology, autopoieses of social systems, or just any kind of systems theoretic entity [77–82].

6. Conclusion

Nature has created ever more complex forms of life by creating structural and dynamical islands of systems with specialized organelles such as genomes being responsible for storage, access, and replication of the information for their persistence and development. Despite the epic quest to determine the details and origin of inheritance, only recently we were finally able to fill the debated gaps of the central part of genome architecture and dynamics - despite the pioneering

works of the last 170 years - by establishing that nucleosomes compact into a quasi-fibre which is folded into stable loops, forming stable multi-loop aggregates/rosettes connected by linkers creating chromosome arms and entire chromosomes [26, 27]. Although the heuristics of the field leads already to a sound basis, this could only be achieved - as we summarized here - by a highly integrated systems approach linking holistically i) a by far superior selective chromosome interaction (T2C) technique, ii) a novel *in vivo* FCS dynamic method, iii) a novel analytical approach and improved super-computer simulations, and iv) finally scaling analysis of the 3D-architecture and the DNA sequence itself. Including the heuristics of the field this leads to a consistent picture of genome organization, which match all the criteria necessary for storage, transcription, and replication as one would expect them as the outcome of Darwinian natural selection and Lamarkian self-referenced manipulation as shown here. In parallel, a multilistic systems statistical mechanics with uncertainty principles has emerged while reaching the fundamental resolution limits in the above holistic approach, which represents a theoretical framework which also reunites the overall far from thermodynamic equilibrium notion with local hierarchic substance stability. Beyond, the tight entanglement of genome levels having left footprints on all levels, has not only shown that genomes have evolved as an entire system, but also the multilistic entanglement between genotype and phenotype. Hence, the natural outcome of Darwinian natural selection and Lamarkian self-referenced manipulation is united in a genome ecology framework, which we consider a major step in the systems theory of life. Thus, this not only leads to a solid basis for sequencing genetic information holistically and thus for applied diagnostics and treatment of disease, as well as future genome manipulation and engineering efforts, but more importantly paves the path to a true understanding of genomes, their function and evolution, and thus of life in general - earthbound, extra-terrestrial, or artificial.

Acknowledgements

For supporting and influencing this long lasting work of T.A.K thanks go to: M. Wachsmuth, T. Weidemann, K. Fejes-Toth, M. Göker, R. Lohner, M. Stör, E. Spiess, K. Rippe, W. Waldeck, C. Cremer, T. Cremer, K. Erenpreisa, A. Ollins, D. Ollins, K. Sullivan, C. C. Murre, J. Skok, A. M. A. Imam, F. G. Grosveld, K. Egger, O. Zimina, and last but not least L. A. Knoch, as well as the German and International Societies for Human Ecology. T2C was invented by T.A.K. and F. G. Grosveld, with many thanks to M. Lesnussa, N. Kepper, A. Abuseiris, P. Kolovos, Jessica Zuin, R. W. W. Brouwer, H. J. G. van de Werken, W. F. J. van IJken, and Kerstin S. Wendt. This work was also part of the EpiGenSys consortium setup and coordinated by T.A.K., funded by ERASysBio+/FP7 and the national funding organizations (the Dutch Ministry for Science and Education, the Netherlands Science Organization, the UK Biotechnology and Biological Sciences Research Council, and the Bundesministerium für Bildung und Forschung (BMBF)). Further support came from the BMBF under grants # 01 KW 9602/2 (Heidelberg 3D Human Genome Study Group, German Human Genome Project), #01AK803A (German MediGRID), #01IG07015G (Services@MediGRID), as well as the Erasmus Medical Centre and the Hogeschool Rotterdam. The High-Performance Computing Center Stuttgart (HLRS; grant HumNuc), the Supercomputing Center Karlsruhe (SCC; grant ChromDyn), and the Computing Facility of the German Cancer Research Center (DKFZ) are thanked for access to their CRAY T3E and IBM SP2s in the initial part of this work. Thanks

also go to all those institutions, universities, and companies providing us computational grid resources: the German D-Grid, the European Grid Initiative EGEE, as well as the Erasmus Computing Grid the Almere Grid, and all the unnamed computing grids there is access through via these. Very special thanks go also to all the world-wide distributed and unnamed donors of desktop computer power of our world-wide Correlizer@home BOINC grid!

Legends Movie 1 and 2

Video files available at: <https://bit.ly/34KCuxe>

Movie 1. Brownian Dynamics simulated decondensation from a metaphase starting configuration of a simulated Multi-Loop-Subcompartment model with 126 kbp loops and linkers with segment length of 50 nm (~5.2 kb) [26]. The whole 750 ms long movie shows how abruptly the metaphase chromosome expands due to its high density while opening the linker, which is constrained/condensed/pulled into a loop in metaphase. Nevertheless, the rosettes form distinct chromatin territories in which the loops do not intermingle freely in contrast to other models (see Introduction) such as the RW/GL model (**Figure 2**). The final shape and form in a whole nucleus would be determined by the limitations the other adjacent chromosomes provide (for more details see [22]). The different densities during decondensation also resemble nicely the conditions of shorter linkers, general genome regions with higher densities, or also the variation of nuclear volumes. Notably, the intrinsic movement of the chromatin fibre is clearly taking place on the millisecond scale, and hence, obviously a topological preformed architecture would dissolve within seconds if it would not be stable [26, 27].

Movie 2. Brownian Dynamics simulation of the consensus architecture (i.e. with the real measured loop and linker sizes) of the of the IGF/H19 region at HS11p15.5–15.4 (**Figure 3**), with a segment length of 20 nm (~2.0 kbp; colours of loops like in **Figure 3** middle, with additional linkers at the beginning and end of the region in red; for details see [26]). The whole movie encompasses 146 ms and shows the high intrinsic dynamics of the loops and the loop aggregate/rosette. Obviously, the single subchromosomal domains are constrained by the subsequent subchromosomal domains. Hence, and also obviously a topological preformed architecture would dissolve within seconds if it would not be stable [26, 27]. Nevertheless, the loop aggregates/rosettes form distinct subchromosomal domains in which the loops do not intermingle freely in contrast to other models (see Introduction) such as the RW/GL model (**Figure 2**). The final shape and form in a whole nucleus would be determined by the limitations adjacent chromosomes provide (for more details see [22]).

Author details

Tobias A. Knoch

Address all correspondence to: ta.knoch@taknoch.org

Biophysical Genomics, Department of Cell Biology and Genetics, Erasmus MC, Rotterdam, The Netherlands

References

- [1] Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*. 1953;**171**(4356):737-738
- [2] Kornberg RD, Klug A. The nucleosome. *Scientific American*. 1981;**2**:28-44
- [3] Olins AL, Olins DE. Spheroid chromatin units (v bodies). *Science*. 1974;**183**:330-332
- [4] Luger C, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Science*. 1997;**389**:251-260
- [5] Eigen M, Winkler-Oswatitsch R. Transfer-RNA, an early gene? *Naturwissenschaften*. 1981;**68**:217-228
- [6] Eigen M, Winkler-Oswatitsch R. Transfer-RNA: The early adapter. *Naturwissenschaften*. 1981;**68**:217-228
- [7] Knoch TA, Munkel C, Langowski J. Three-dimensional organization of chromosome territories and the human cell nucleus—About the structure of a self replicating nano fabrication site. In: Foresight Institute—Article Archive. Palo Alto, CA, USA: Foresight Institute; 1998. pp. 1-6. Available at: <http://www.foresight.org>
- [8] Finch JT, Klug A. Solenoidal model for the superstructure in chromatin. *Proceedings of the National Academy of Sciences of the United States of America*. 1976;**73**:1897-1901
- [9] Baudy P, Bram S. Chromatin fiber dimensions and nucleosome orientation: A neutron scattering investigation. *Nucleic Acids Research*. 1978;**5**(10):3697-3714
- [10] Baudy P, Bram S. Neutron scattering on nuclei. *Nucleic Acids Research*. 1979;**6**(4):1721-1729
- [11] Ibel K. Neutron diffraction of interphase nuclei. *Journal of Molecular Biology*. 1982;**160**(1):77-85
- [12] Notbohm H. Small angle scattering of cell nuclei. *European Biophysics Journal*. 1986;**13**(6):367-372
- [13] Dubochet J. Cryo-EM—The first thirty years. *Journal of Microscopy*. 2012;**245**(3):221-224
- [14] Eltsov M, Maclellan KM, Maeshima K, Frangakis AS, Dubochet J. Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ. *PNAS*. 2008;**105**(50):19732-19737
- [15] Müller O, Kepper N, Schöpflin R, Ettig R, Rippe K, Wedemann G. Changing chromatin fiber conformation by nucleosome repositioning. *Biophysical Journal*. 2014;**107**(9):2141-2150
- [16] Kepper N, Foethke D, Stehr R, Wedemann G, Rippe K. Nucleosome geometry and inter-nucleosomal interactions control the chromatin fiber conformation. *Biophysical Journal*. 2008;**95**(8):3677-3691

- [17] Stehr R, Schöpfling R, Ettig R, Kepper N, Rippe K, Wedemann G. Exploring the conformational space of chromatin fibers and their stability by numerical dynamic phase diagrams. *Biophysical Journal*. 2010;**98**(6):1028-1037
- [18] Weidemann T, Wachsmuth M, Knoch TA, Müller G, Waldeck W, Langowski J. Counting nucleosomes in living cells with a combination of fluorescence correlation spectroscopy and confocal imaging. *Journal of Molecular Biology*. 2003;**334**(2):229-240
- [19] Wachsmuth M, Weidemann T, Müller G, Hoffmann-Rohrer UW, Knoch TA, Waldeck W, et al. Analyzing intracellular binding and diffusion with continuous fluorescence photo-bleaching. *Biophysical Journal*. 2003;**84**(5):3353-3563
- [20] Capoulade J, Wachsmuth M, Hufnagel L, Knop M. Quantitative fluorescence imaging of protein diffusion and interaction in living cells. *Nature Biotechnology*. 2011;**29**(9):835-839
- [21] Knoch TA. Dreidimensionale Organisation von Chromosomen-Domänen in Simulation und Experiment. (Three-dimensional organization of chromosome domains in simulation and experiment.). Mannheim, Germany: TAK Press; 1998. ISBN 3-00-010685-5
- [22] Knoch TA. Approaching the Three-Dimensional Organization of the Human Genome: Structural-, Scaling- and Dynamic Properties in the Simulation of Interphase Chromosomes and Cell Nuclei, Long-Range Correlations in Complete Genomes, In Vivo Quantification of the Chromatin Distribution, Construct Conversions in Simultaneous Co-transfections. Mannheim, Germany: TAK Press; 2002. ISBN 3-00-009959-X
- [23] Knoch TA. Towards a holistic understanding of the human genome by determination and integration of its sequential and three-dimensional organization. In: Krause E, Jäger W, Resch M, editors. *High Performance Computing in Science and Engineering, High-Performance Computing Center (HLRS)*. Stuttgart, Berlin-Heidelberg-New York: University of Stuttgart, Springer; 2003. pp. 421-440. ISBN 3-540-40850-9
- [24] Knoch TA, Göcker M, Lohner R, Abuseiris A, Grosveld FG. Fine-structured multi-scaling long-range correlations in completely sequenced genomes—Features, origin and classification. *European Biophysics Journal*. 2009;**38**(6):757-779
- [25] Knoch TA, Grosveld FG. Method for analysing the interaction of nucleotide sequences in a three-dimensional DNA structure. GB Patent Application GB20130020351 and International Patent Application WO2014IB02485 20141118
- [26] Knoch TA, Wachsmuth M, Kepper N, Lesnussa M, Abuseiris A, Imam AMA, et al. The detailed 3D multi-loop aggregate/rosette chromatin architecture and functional dynamic organization of the human and mouse genomes. *Epigenetics & Chromatin*. 2016;**9**(58):1-22
- [27] Wachsmuth M, Knoch TA, Rippe K. Mapping properties of dynamic chromatin domains by fluorescence correlation spectroscopy. *Epigenetics & Chromatin*. 2016;**9**(57):1-22
- [28] Rabl C. Über Zellteilung. *Morphologisches Jahrbuch*. 1885;**10**:214-330
- [29] Boveri T. Die Blastomerenkerne von *Ascaris megalocephala* und die Theorie der Chromosomenindividualität. *Archiv für Zellforschung*. 1909;**3**:181-268

- [30] Comings DE. The rationale for an ordered arrangement of chromatin in the interphase nucleus. *American Journal of Human Genetics*. 1968;**20**:440-460
- [31] Comings DE. Mechanisms of chromosome banding and implications for chromosome structure. *Annual Review of Genetics*. 1978;**20**:440-460
- [32] Vogel F, Schroeder TM. The internal order of the interphase nucleus. *Humangenetik*. 1974;**25**(4):265-297
- [33] Paulson JR, Laemmli UK. The structure of histone-depleted metaphase chromosomes. *Cell*. 1980;**12**:817-828
- [34] Pienta KJ, Coffey DS. A structural analysis of the role of the nuclear matrix and DNA loops in the organization of the nucleus and chromosome. *Journal of Cell Science*. Supplement. 1984;**1**:123-135
- [35] Cremer C, Zorn C, Cremer T. An ultraviolet microbeam for 257 nm. *Microscopy Acta*. 1974;**75**:331-337
- [36] Cremer T, Cremer C, Baumann H, Luedtke EK, Sperling K, Teubner V, et al. Rabl's model of the interphase chromosome arrangement, tested in Chinese hamster cells by premature chromosome condensation and laser-UV-microbeam experiments. *Human Genetics*. 1982;**60**:46-46
- [37] Lichter P, Cremer T, Borden J, Manuelidis L, Ward DC. Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. *Human Genetics*. 1988;**80**:224-234
- [38] Cremer T, Cremer C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature Reviews. Genetics*. 2001;**2**:292-301
- [39] Berezney R, Dubey DD, Huberman JA. Heterogeneity of eukaryotic replicons, replicon clusters, and replication foci. *Chromosoma*. 2000;**108**(8):471-484
- [40] Cremer T, Cremer M. Chromosome territories. *Cold Spring Harbor Perspectives in Biology*. 2010
- [41] Yunis JJ. Mid-prophase human chromosomes. The attainment of 2000 bands. *Human Genetics*. 1981;**56**:293-298
- [42] Francke U. Digitized and differentially shaded human chromosome ideograms for genomic applications. *Cytogenetics and Cell Genetics*. 1994;**65**:206-219
- [43] Erenpreisa J. Large rosettes—The element of the suprachromonemal organisation of interphase cell nucleus. *Proceedings of the Latvian Academy of Sciences Series*. 1989; **B7**:68-71 (Russ)
- [44] Reznik NA, Yampol GP, Kiseleva EV, Khristolyubova NB, Gruzdev AD. In: Harris JR, Zbarsky IB, editors. *Possible Functional Structures in the Chromomere, Nuclear Structure and Function*. New York-London: Plenum Press; 1990. pp. 27-29

- [45] Belmont AS, Bruce K. Visualization of G1 chromosomes: A folded, twisted, supercoiled chromonema model of interphase chromatid structure. *The Journal of Cell Biology*. 1994;**127**(2):287-302
- [46] Sachs RK, van den Engh G, Trask B, Yokota H, Hearst JE. A random-walk/giant-loop model for interphase chromosomes. *Proceedings of the National Academy of Sciences of the United States of America*. 1995;**92**:2710-2714
- [47] Yokota H, Singer MJ, van den Engh GJ, Trask BJ. Regional differences in the compaction of chromatin in human G0/G1 interphase nuclei. *Chromosome Research*. 1997;**5**(3):157-166
- [48] Yokota H, van den Engh G, Hearst J, Sachs RK, Trask BJ. Evidence for the organization of chromatin in megabase pair-sized loops arranged along a random walk path in the human G0/G1 interphase nucleus. *The Journal of Cell Biology*. 1995;**130**(6):1239-1249
- [49] Jhunjhunwala S, van Zelm MC, Peak MM, Cutchin S, Riblet R, van Dongen JJM, et al. The 3D-structure of the immunoglobulin heavy chain locus: Implications for long-range genomic interactions. *Cell*. 2008;**133**(2):265-279
- [50] Rauch J, Knoch TA, Solovei I, Teller K, Stein S, Buiting K, et al. Lightoptical precision measurements of the Prader-Willi/Angelman Syndrome imprinting locus in human cell nuclei indicate maximum condensation changes in the few hundred nanometer range. *Differentiation*. 2008;**76**(1):66-82
- [51] Knoch TA. Simulation of different three-dimensional polymer models of interphase chromosomes compared to experiments—An evaluation and review framework of the 3D genome organization. *Seminars in Cell & Developmental Biology*. 2019;**90**:19-42
- [52] Knoch TA. Simulation of different three-dimensional polymer models of whole interphase nuclei compared to experiments. submitted
- [53] Knoch TA, Waldeck W, Müller G, Alonso A, Langowski J. DNA-Sequenz und Verfahren zur in vivo Markierung und Analyse von DNA/Chromatin in Zellen. German Patent Application 10013204.9-44 and International Patent Application PCT/DE01/01044. 2000
- [54] Baum M, Erdel F, Wachsmuth M, Rippe K. Retrieving the intracellular topology from multi-scale protein mapping in living cells. *Nature Communications*. 2014;**5**:4494-4506
- [55] Knoch TA, Göcker M, Lohner R. Methods for the analysis, classification and/or tree construction of sequences using correlation analysis. US Patent Application 60/436.056 and International Patent Application PCT/EP03/14854
- [56] Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science*. 2002;**295**:1306-1311
- [57] Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Molecular Cell*. 2002;**10**:1453-1465

- [58] Hagege H, Klous P, Braem C, Splinter E, Dekker J, Cathala G, et al. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nature Protocols*. 2007;**2**:1722-1733
- [59] Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nature Genetics*. 2006;**38**:1348-1354
- [60] Stadhouders R, Kolovos P, Brouwer R, Zuin J, van den Heuvel A, Kockx C, et al. Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. *Nature Protocols*. 2013;**8**:509-524
- [61] Dostie J, Dekker J. Mapping networks of physical interactions between genomic elements using 5C technology. *Nature Protocols*. 2007;**2**:988-1002
- [62] 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**:289-293
- [63] 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
- [64] Naumova N, Imakaev M, Fudenberg G, Zhan Y, Lajoie BR, Mirny LA, et al. Organization of the mitotic chromosome. *Science*. 2013;**342**(6161):948-953
- [65] Giorgetti L, Galupa R, Nora EP, Piolot T, Lam F, Dekker J, et al. Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell*. 2014;**157**(4):950-963
- [66] 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**:1-16
- [67] Kolovos P, van de Werken HJ, Kepper N, Zuin J, Brouwer RW, Kockx CE, et al. Targeted chromatin capture (T2C): A novel high resolution high throughput method to detect genomic interactions and regulatory elements. *Epigenetics & Chromatin*. 2014;**7**:10
- [68] Kolovos P, Brouwer RWW, Kockx CEM, Lesnussa M, Kepper N, Zuin J, et al. Investigation of the spatial structure and interactions of the genome at sub-kilobase-pair resolution using T2C. *Nature Protocols*. 2018;**13**(3):459-477
- [69] Knoch TA. A guided protocol for array based T2C—A high-quality high-resolution high-throughput chromosome interaction capture. *Current Protocols in Human Genetics*. 2018;**99**(1):1-35
- [70] Knoch TA, Lesnussa M, Kepper FN, Eussen HB, Grosveld FG. The GLOBE 3D genome platform—Towards a novel system-biological paper tool to integrate the huge complexity of genome organization and function. *Studies in Health Technology and Informatics*. 2009;**147**:105-116

- [71] Gerlich D, Beaudouin J, Kalbfuss B, Daigle N, Eils R, Ellenberg J. Global chromosome positions are transmitted through mitosis in mammalian cells. *Cell*. 2003;**112**:751-764
- [72] Dross N, Spriet C, Zwerger M, Muller G, Waldeck W, Langowski J. Mapping eGFP oligomer mobility in living cell nuclei. *PLoS One*. 2009;**4**:e5041
- [73] Belmont AS. Visualizing chromosome dynamics with GFP. *Trends in Cell Biology*. 2001;**11**:250-257
- [74] Bunde A, Havlin S. *Fractals and Disordered Systems*. 2nd ed. Berlin Heidelberg, New York: Springer Verlag; 1996. ISBN 3-540-56219-2
- [75] Prigogine IR. *From Being to Becoming*. Freemann; 1980 ISBN 0-7167-1107-9
- [76] Gladyshev GP. Hierarchical thermodynamics: Foundation of extended Darwinism. *Imperial Journal of Interdisciplinary Research*. 2017;**3**(2):1-14
- [77] Mann F, Mann C. In: Fischer S, editor. *Es werde Licht: Die Einheit von Geist und Materie in der Quantenphysik*. 2017 ISBN 97803103972450
- [78] Schrödinger E. *What Is Life?/Mind and Matter*. 1967th ed. Cambridge University Press; 1944
- [79] Maturana HR, Varela F. *Tree of Knowledge*. Shambhala; 1992. ISBN 0877736421, ISBN 978-0877736424
- [80] Luhmann N. *Soziale Systeme: Grundriß einer allgemeinen Theorie*. Suhrkamp; 2008. ISBN 3518282662, ISBN 978-3518282663
- [81] Knoch TA, Baumgärtner V, de Zeeuw LV, Grosveld FG, Egger K. E-human grid ecology: Understanding and approaching the inverse tragedy of the commons in the e-grid society. *Studies in Health Technology and Informatics*. 2009;**147**:269-276
- [82] Knoch TA. Sustained renewability: Approached by systems theory and human ecology. In: Nayeripour M, Keshti M, editors. *Renewable Energy 2*. Intech; 2011. pp. 21-48. ISBN: 978-953-307-573-0. <https://www.intechopen.com/books/sustainable-growth-and-applications-in-renewable-energy-sources/sustained-renewability-approached-by-systems-theory-and-human-ecology>

