Towards a Holistic Understanding of the Human Genome by Determination and Integration of Its Sequential and Three-Dimensional Organization

Tobias A. Knoch

Deutsches Krebsforschungszentrum (DKFZ),
Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany
E-mail: TA.Knoch@DKFZ-Heidelberg.de or TA.Knoch@taknoch.org

In Cooperation With: F. Bestvater[1,2], C. Cremer[6], T. Cremer[4], K. Fejes-Toth[1,6], A. Friedel[4], M. Göcker[9], A. Kellerer[4], R. Lohner[5], K. Monier[8], J. Langowski[1], P. Lichter[1], P. Quicken[4,7], J. Rauch[6], K. Rippe[1,6], E. Spiess[1], K. Sullivan[8], W. Waldeck[1], M. Wachsmuth[3], T. Weidemann[1]
[1]Deutsches Krebsforschungszentrum, Heidelberg, Germany; [2]Febit GmbH, Mannheim, Germany;
[3]Leica Microsystems GmbH, Mannheim, Germany; [4]Ludwig-Maximillians Universität, Munich, Germany;
[5]Supercomputing Center Karlsruhe, Universität Karlsruhe, Karlsruhe, Germany;
[6]Ruprecht-Karls Universität, Heidelberg, Germany; [7]Symantec Education Services, Ismaning, Germany;
[8]The Scripps Institute, La Jolla, California, USA; [9]Universität Tübingen, Tübingen, Germany

Summary. Genomes are one of the major foundations of life due to their role in information storage, process regulation and evolution. However, the sequential and three-dimensional structure of the human genome in the cell nucleus as well as its interplay with and embedding into the cell and organism only arise scarcely from the unknown, despite recent successes e. g. in the linear sequencing efforts and growing evidence for seven genomic organization levels. To achieve a deeper understanding of the human genome the structural, scaling and dynamic properties in the simulation of interphase chromosomes and cell nuclei are determined and combined with the analysis of long-range correlations in completely sequenced genomes as well as the analysis of the chromatin distribution *in vivo*: This integrative approach reveals that the chromatin fiber is most likely folded according to the Multi-Loop-Subcompartment (MLS) model in which the chromatin fiber bents into 63–126 kbp big loops aggregated to rosettes connected by again 63–126 kbp linkers. The MLS model exhibits fine-structured multi-scaling and predicts correctly the transport of molecules by moderately obstructed/anomalous diffusion. On the basic sequence level, genomes show fine-structured positive long-range correlations, allowing classification and tree construction. This, DNA fragment distributions after carbon ion irradiation and on the highest structural level, the nuclear morphology visualized by histone autofluorescent protein fusions *in vivo*, agrees again best with the MLS model. Thus, the local, global and dynamic characteristics of cell nuclei are not only tightly inter-connected, but also are integrated holisticly to fulfill the overall function of the genome.

1 Introduction

In human cells the genetic information controlling most processes from the cellular level, over embryogeneses to cognitive ability, manifests in a diploid set of 23 DNA molecules, the chromosomes. They consist of $\approx 7\times10^9$ base pairs (bp) storing around 1.4×10^{10} Bit or 1.75 GByte, and add to a molecular length totalling ≈ 2 m, which are kept in comparably small cell nuclei with typical diameters of $\approx 10\,\mu$m or volumes of $500\,\mu m^3$. This corresponds to a compaction factor of $\approx 2\times10^5$. Consequently, beyond pure compaction, the structuring of the genetic information in several organizational levels seems obviously to allow on the one hand sufficient performance during information transcription in interphase and on the other hand replication of the information and segregation of the chromosomes into the daughter cells during mitosis. Additionally, the abundant mutations need to be continuously found, controlled and repaired to avoid the inevitable course of entropy.

Considering the huge length and time scales, which bridge 10^{-9}–10^{-5} m and 10^{-10}–10^4 s, the genetic information of the human genome involves seven organizational levels according to current believe (Fig. 1): the DNA double helix (i), winds around a protein complex forming the nucleosome (ii), which condenses irregularly to the 30 nm chromatin fiber (iii), which is folded into chromatin loops (iv), which aggregate to chromosomal subdomains (v), which constitute a chromosome (vi), being non-randomly arranged in the cell nucleus (vii).

The DNA double helix and the nucleosome structure are known to atomic precision, but already the detailed nucleosome conformation in the 30 nm chromatin fiber is still debated. The latter holds even more for the higher-order structures having born many a hypothesis: Whereas light microscopic studies by Rabl (1885) and Boveri (1909) proposed territorial chromosomes with a hierarchical, self-similar organization of chromatin fibers in the late 19 th century, electron microscopy suggested thereafter a random interphase chromatin organization in the models of Comings (1968, 1978) and of Vogel & Schroeder (1974). To explain the high condensation degree of metaphase chromosomes and their stainable ideogram bands, chromatin loops attached to a nuclear matrix scaffold were suggested in the Radial-Loop-Scaffold model by Paulson & Laemmli (1980). According to Pienta & Coffey (1984) these loops persisted in interphase forming stacked rosettes in metaphase. Microirradiation and fluorescence *in situ* hybridization (FISH) finally proved a territorial organization of chromosomes, of their arms, and of subchromosomal domains and led to the Inter-Chromosomal Domain (ICD) model hypothesizing an interchromosomal channel network (Zirbel *et al.*, 1993, Cremer & Cremer, 2001). For the intraterritorial chromatin folding, the chromonema fiber (CF) model by Belmont & Bruce (1994) postulated a helix hierarchy, whereas in the Random-Walk/Giant-Loop (RW/GL) model 1–5 Mbp loops are attached to a non-protein backbone (Sachs *et al.*, 1995) and in the Multi-Loop-

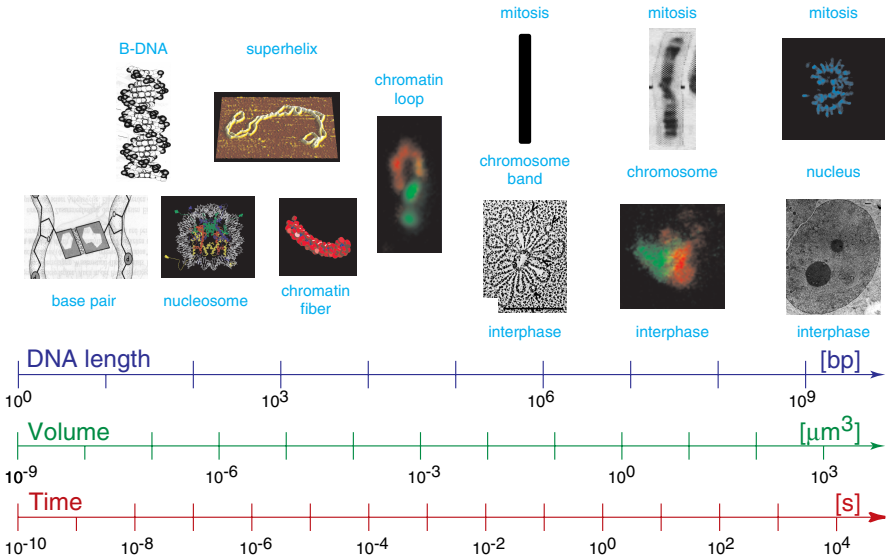


Fig. 1. *Overview on the Size and Time Scaling of the Human Genome.* The scaling and the levels of organization range over 9 decades for base pairs, 12 decades for the volume (≈4 length decades) and 14 decades for the time: At the initial stage base pairs are formed composing the DNA double helix (Voet & Voet, 1995) which winds around the histone core complex building the nucleosome (callotte model according to Luger *et al.*, 1997), which condense into the 30 nm chromatin fiber (simulation image, Wedemann & Langowski, 2002). The DNA double helix forms also superhelices (scanning force microscopic image of plasmid DNA, with courtesy of K. Rippe, Division Molecular Biophysics, Kirchhoff Institute for Physics, Ruprecht-Karls-Universität, Heidelberg, Germany). The next compaction step consists of chromatin loops (fluorescence *in situ* hybridization (FISH) image, with courtesy of P. Fransz, Swammerdam Institute for Life Sciences, BioCentrum Amsterdam, Amsterdam, The Netherlands) possibly forming rosettes (electron microscopic image, Reznik *et al.*, 1990), which make up interphase chromosome arms and territories (FISH image, with courtesy of S. Dietzel) and the metaphase ideogram bands (giemsa chromosome staining from Alberts *et al.*, 1994). 46 chromosomes compose the human nucleus and are decondensed in interphase (electron microscopic image, with courtesy of K. Richter, Division Molecular Genetics, Deutsches Krebsforschungszentrum, Heidelberg, Germany) and condensed as shown for separated metaphase plates (Fig. 10 B).

Subcompartment (MLS) model 60–120 kbp loops form rosettes connected by a similar linker (Münkel & Langowski, 1998; Münkel *et al.*, 1999).

However, despite all these successes, the sequential and three-dimensional organization as well as their connection are still known only rudimentarily. Therefore, aspects from all nuclear scales as well as simulations and experiments were combined to determine and integrate the sequential and three-dimensional organization of the human genome (Knoch, 2002).

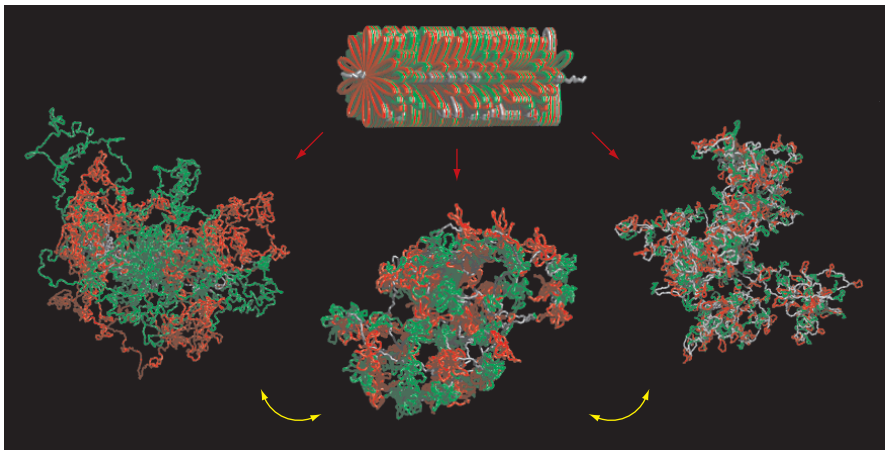


Fig. 2. *Volume Rendered Images of Simulated Chromosome Models for Chromosome XV.*
From the typical startconfiguration for simulations with the form and size of a metaphase chromosome (above), interphase chromosomes in thermodynamical equilibrium are decondensed by Monte Carlo (MC) steps, followed by relaxing Brownian Dynamics (BD) steps. In a Random-Walk/Giant-Loop (RW/GL) model, with 5 Mbp loops and 378 kbp, the large loops intermingle freely not forming distinct features like in the Multi-Loop-Subcompartment (MLS) model (8×10^4 MC, 10^3 BD, left). In an MLS model with both 126 kbp loops and linkers the rosettes form subcompartments as separated organizational and dynamic entities (5×10^4 MC, 10^3 BD, middle). In a RW/GL model with 126 kbp loops and 63 kbp linkers, the small loops neither intermingle freely nor form distinct subcompartments (9×10^4 MC, 10^3 BD, right). Obviously, the MLS model can easily be transformed in both RW/GL models by minor topological changes, indicating possible biological mechanism or the origin of possible experimental preparation artefacts. Consecutive loops of the RW/GL or MLS rosettes are painted in red and green. The fiber diameter is 30 nm in all images.

2 Simulation of Single Chromosomes and Cell Nuclei

To investigate the folding of the 30 nm chromatin fiber into chromosome territories, their morphology and experimental distinguishability, single chromosomes based on the Multi-Loop-Subcompartment (MLS) model and the Random-Walk/Giant-Loop (RW/GL) topology, were simulated for various loop and linker sizes. The 30 nm chromatin fiber was modelled as a polymer chain with stretching, bending and excluded volume interactions. A spherical boundary potential simulated the confinement of other chromosomes and the nucleus. Monte Carlo and Brownian Dynamics methods were applied to generate chain configurations at thermodynamic equilibrium (Fig. 2). These simulations of single chromosomes were extended to nuclei of diploid human cells containing all 46 chromosomes, to determine the chromosome arrangement and the related microscopic morphology, besides the validation of the results

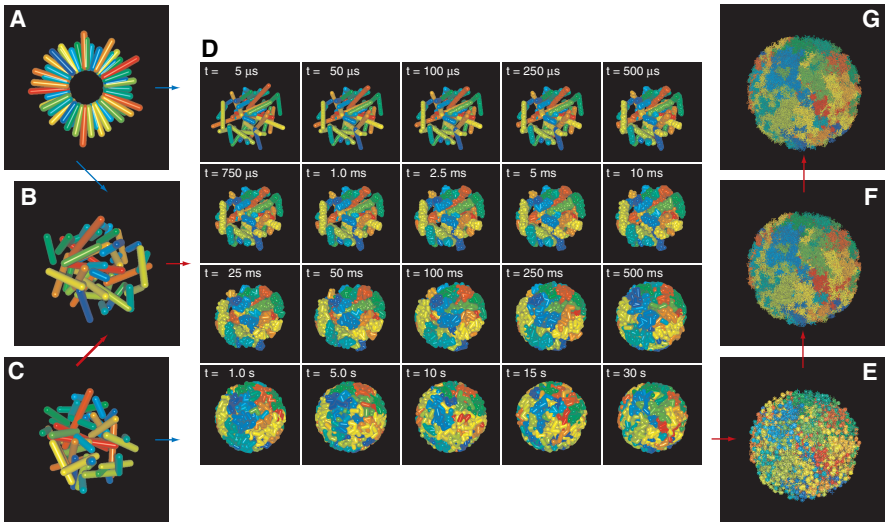


Fig. 3. *Startconfigurations and Decondensation into Interphase of Simulated Cell Nuclei.*
Metaphase chromosomes were placed in a metaphase plate (A) or as cylinders randomly (C) into a spherically constrained potential before optional relaxing by simulated annealing avoiding unphysical configurations (B). The cylinders were split into spheres and decondensed into interphase by $\leq 6\times10^6$ Brownian Dynamic (BD) steps of 5 µs, i.e. 30 s (D). Then the 30 nm chromatin fiber was placed into the spheres with scaled down chain segments avoiding concatenation (E) and softly relaxed with 10^3 BD steps linearly increasing from 0.01–0.5 µs (F). Segment sizes were 300 nm or smaller that chromatin loops consisted of ≥ 4 segments. Finally the resolution was increased to 50 nm segments, and relaxed further by 5×10^3 BD steps of 0.5 µs to ensure proper morphologic and quantitative analysis (G). Shown for an MLS model with 126 kbp loops and linkers and 5 µm nuclear radius.

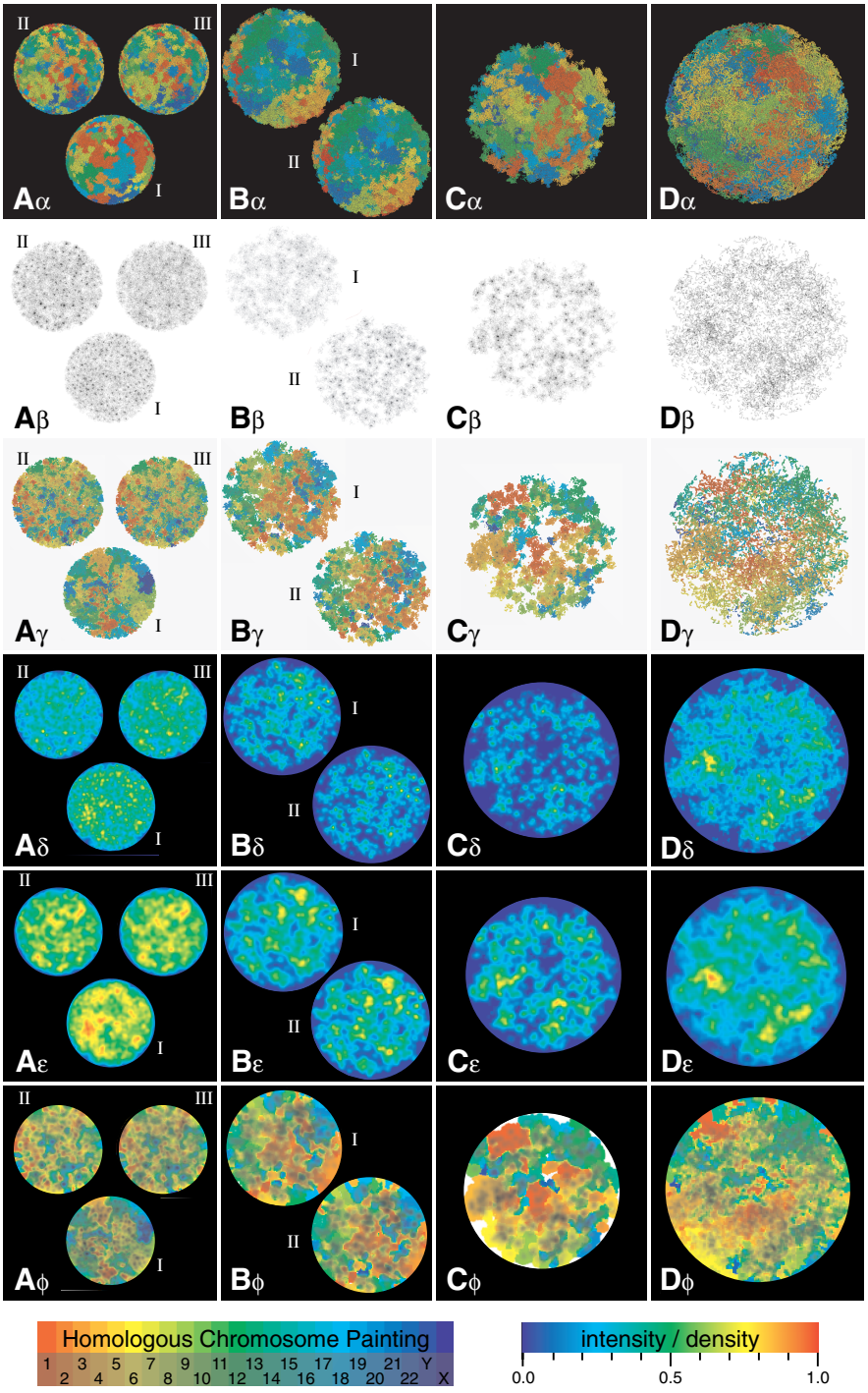
of the simulation of single chromosomes. The chromatin fiber was simulated as in the case of single chromosomes. Since the computer power increased by a factor of 46, in this case simulated annealing and Brownian Dynamics methods as well as a four step decondensation procedure from metaphase (Fig. 3) were applied to generate interphase configurations again at thermodynamic equilibrium (Fig. 4). The simulation program VirtNuc as well as various helper and analysis programs were written in the object oriented C++ language and parallelized using the message passing interface (MPI) standard. By application of a very efficient linked cell algorithm for the calculation of the physical interactions and by implementation of dynamic load balancing, the parallelization efficiency has been improved to 95 % using 16 processors. The simulation and analysis of single chromosomes, totalled $\approx$ 96,000 CPUh ($\approx$ 11 years) on a single R6000 processor with 60 MHz and those of whole nuclei totalled $\approx$ 260.000 h ($\approx$ 30 years) on a single R6000 processor with 120 MHz.

Both the MLS and the RW/GL model form chromosome territories with different morphologies: The MLS rosettes result in distinct subcompartments visible with light microscopy (Fig. 2– 4). This morphology and the size of these subcompartments agree with the morphology found by expression of histone autofluorescent protein fusions (see below) and FISH experiments. In contrast, the big RW/GL loops lead to a homogeneous chromatin distribution. Even small changes of the model parameters induced significant rearrangements of the chromatin morphology as can be seen in simulated electron microscopic (EM, Fig. 7 A&C) and confocal laser scanning microscopic (CLSM) images (Fig. 4). Thus, pathological diagnoses of e. g. cancer based on the nuclear morphology, might be related to structural changes on the chromatin level. The position of chromosome territories in interphase depends on their metaphase location (Fig. 3), and suggests a possible origin of current experimental findings. Only the MLS model leads to a low overlap of chromosomes, their arms and subcompartments, again in agreement with experiments. The chromatin density distribution in CLSM image stacks of the MLS model but not the RW/GL model reveals a bimodal behaviour in agreement with recent experiments (Weidemann, 2002). Review and comparison of experimental to simulated spatial distance measurements between genomic markers as function of their genomic separation also favour an MLS model with loop and linker sizes of 63–126 kbp (Fig. 5).

3 Scaling of the Simulated Chromatin Topology and Nuclear Morphology

In order to characterize the huge time and length scale bridging several levels of packaging of the genome, the scaling behaviour of the 30 nm chromatin

Fig. 4. *Morphology of Cell Nuclei by Rendering, EM and CLSM Images.* All differences of chromosome models 63-63-MLS^3 (AI), 63-252-MLS^3 (AII), 126-252-MLS^3 (AIII), 126-126-MLS^4 (BI), 84-126-MLS^4 (BII), 126-126-MLS^5 (C), 1320-RW/GL^6 (D), are visible in the three-dimensional rendering of the actual 30 nm chromatin fiber (α), the electron microscopic (EM) image (β), the EM image chromosome map (γ; homologous chromosome painting legend below), the confocal laser scanning microscopic (CLSM) image with a high resolution 100× 1.4 oil immersion PL APO objective (δ), the CLSM image with a lower resolution 60× 1.2 water immersion PL APO objective (ε; colour intensity coding below), and the CLSM image chromosome map (ϕ): The rosettes of the MLS model form subcompartments being visible as separated organizational and dynamic entities (A, B, C). In contrast, the large loops of the RW/GL model intermingle freely, neither forming distinct features like in the MLS model (D), nor forming clearly separated chromosome territories due to high overlap. The EM and CLSM images were normalized to highest intensity, representing absolute no intensities. Nomenclature: L_S-L_{IS}-model^r; L_S: loop size; L_{IS}: linker size; r: nuclear radius.


II
III
I
Aα
Bα
Cα
Dα
Aβ
Bβ
Cβ
Dβ
Aγ
Bγ
Cγ
Dγ
Aδ
Bδ
Cδ
Dδ
Aε
Bε
Cε
Dε
Aϕ
Bϕ
Cϕ
Dϕ
Homologous Chromosome Painting
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 Y X
intensity / density
0.0
0.5
1.0

fiber topology (Fig. 7 A-C) and the scaling behaviour of the morphology of simulated confocal laser scanning microscopic (CLSM) image stacks was determined (Fig. 4). Both were obtained from simulations of single chromosomes and whole nuclei using the RW/GL and MLS chromatin fiber topologies. For the analysis, the following scaling/fractal dimensions were calculated: The scaling of the one-dimensional axis of the 30 nm chromatin fiber was analysed with the exact spatial-distance and yard-stick dimensions and its voluminous scaling was investigated by the pure box-counting dimension. The scaling behaviour of the (inverse-, iso-) mass distribution of CLSM image stacks was investigated with the weighted box-counting, lacunarity and local mass dimensions. To support the latter, additionally, the local diffuseness, skewness and kurtosis distributions were calculated. The scaling of the chromatin fiber revealed different power-law behaviours on different scales (Fig. 6 A&B). This multi-scaling is created by the random walk behaviour of the fiber, the globular nature and the arrangement of loops or rosettes. Within the multi-scaling regime a fine structure was present for the MLS model arising from the rosette loops. A similar fine-structured multi-scaling behaviour was also found in the correlation behaviour on the level of the DNA sequence of human chromosomes (see below). Thus, the sequential and three-dimensional organization of genomes are closely interconnected. The scaling of CLSM image stacks also reflected the model and imaging properties in detail and are also able to characterize the morphology found in experiments using histone autofluorescent protein fusions *in vivo* (Fig. 6 C-E). Thus, the chromatin fiber topology is also closely connected to nuclear morphology. Therefore, scaling analyses of the nuclear morphology are a suitable approach to differentiate between different cell states, e. g. during the cell cycle, due to malignancy, in apoptosis or in response to drugs. Consequently, the scaling behaviour shows that all nuclear organization levels are connected.

4 Simulation of the DNA Fragment Distribution after Ion Irradiation

Since the chromatin fiber is inhomogeneously distributed and exhibits fine-structured multi-scaling behaviour in cell nuclei also the length distribution of DNA fragments after irradiation with an inhomogeneous nuclear radiation dose distribution and thus the sites of double strand breakage depend on the spatial arrangement of the chromatin fiber (Friedl, 1994; Knoch, 2000). Quantitatively, the fragment distribution of carbon ions was calculated by simulating irradiation with carbon ions of single chromosome interphase configurations by extending the usual random breakage model of the DNA to a dose dependent mixed Poisson processes (Fig. 8 B): The inhomogenously distributed single ion tracks were mapped against the chromatin fiber assuming an energy distribution and thus destruction probability of the track according to Kraft/Scholz and Chatterjee. The resulting fragment distributions revealed

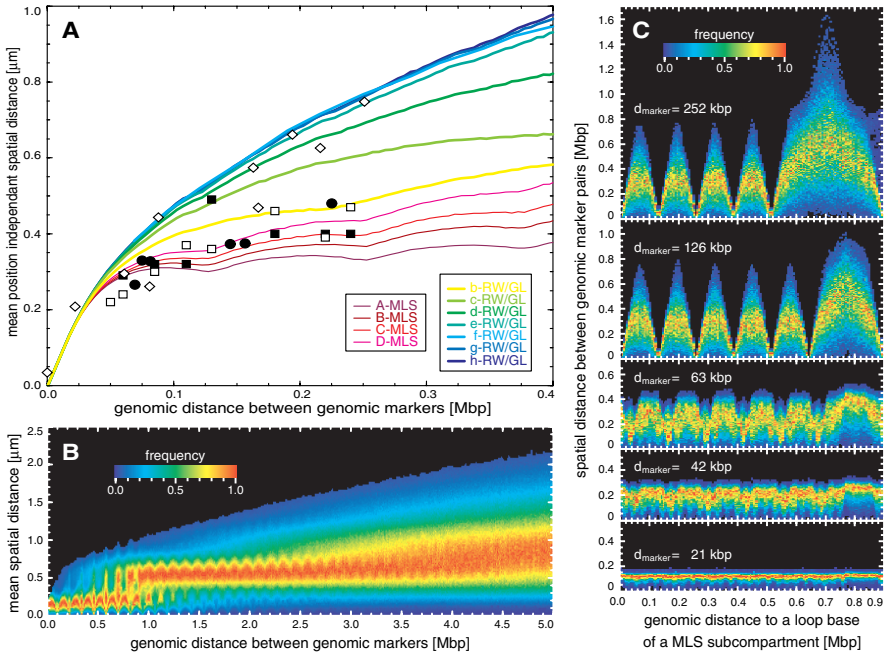


Fig. 5. *Simulated Spatial Distance Dependencies and Comparison to Experiments.*
Simulated spatial distances between position independent placed genomic markers as function of their genomic distance reflect the chromatin fiber topology in detail and yields the better agreement with the MLS model and experiments the better their latter preparation and imaging methods are (A, single chromosome simulations). Thin lines are A-D MLS models with 126 kbp loops and 63, 126, 189 and 252 kbp linkers. Thick lines: a-h RW/GL models with 0.252, 0.504, 1.0, 2.0, 3.0, 4.0 and 5.0 Mbp. Data are: full circles: Knoch (1998, 1999, 2000) and Rauch (1999); data from Monier (1997): full squares: fibroblasts 11q13, open squares: lymphocytes 11q13; open rhombi: Trask *et al.* (1989). The independent spatial distance distributions (B, for a 126-126-MLS^5 nucleus) reveal loop introduced oscillations up to ≈3 Mbp and distance spikes for loop size multiples due to the higher probability of markers located in the linker. The visible shifts into new rosettes (e. g. 0.75 Mbp) and the oscillation decrease suggests higher rosette than loop flexibility concerning phase space distributions. The distributions of position dependent spatial distances of a marker set with different genomic separations as function of their shift (5.2 kbp) relative to a loop basepoint of a rosette show characteristic oscillations and correlations as expected for fixed loop and rosette structures (C, for a B-MLS model).

better agreement for the Chatterjee track assumption with experiments and represent the chromatin fiber topology not only in its general behaviour but by the appearance of a fine structure also in its detail, e. g. loop size or loop arrangement in rosettes (Fig. 8 A). Thus, the fine-structured multi-scaling of the chromatin folding is present also in the fragment distribution. A compar-

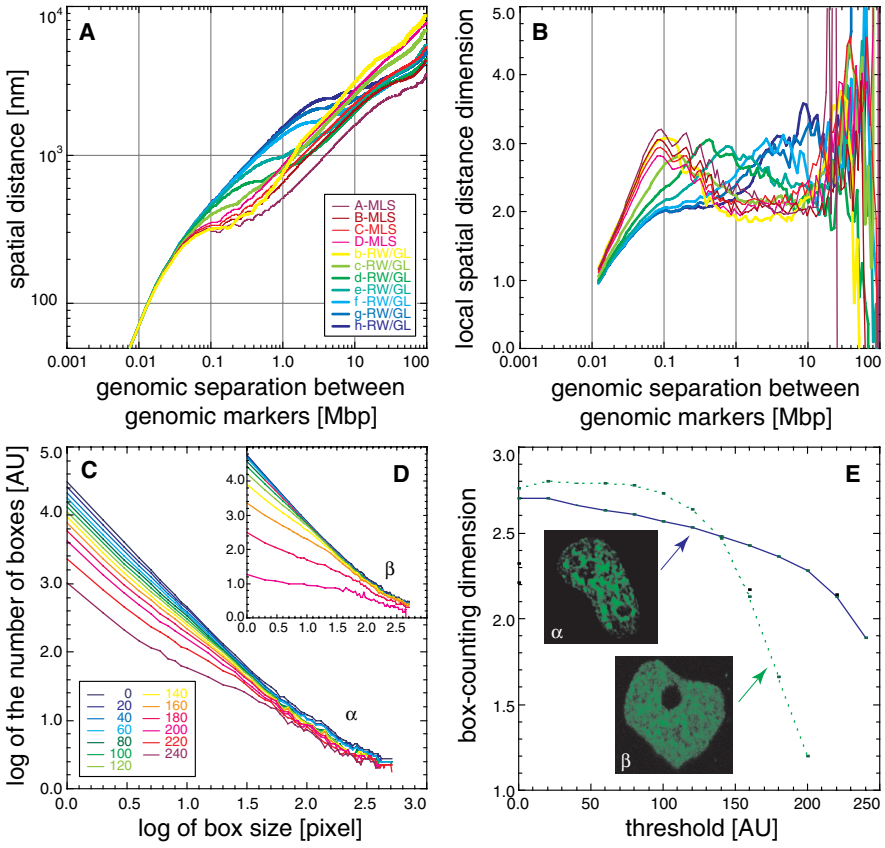


Fig. 6. *Scaling of the Chromatin Fiber of Simulated Chromosomes and of the Nuclear Morphology in vivo.*

In simulated chromosomes the spatial-distance function (A, double log plot of Fig. 5 A) and its local slope the spatial-distance dimension (B) shows a fine-structured multi-scaling behaviour of the chromatin fiber topology below the cut-off greater ≈ 80 Mbp or $\approx 8\,\mu$m due to the finite chromosome size. The general multi-scaling behaviour is characterized by an increase from 1.0 on small scales characterizing the stiff chromatin segments, over values ≈ 2.0 as for the random walk of the segments to a maximum of 3.0 stating the ring-shaped loops of both the MLS and the RW/GL model and globular state of the rosettes of the MLS models. In the MLS model thereafter again local dimensions ≈ 2.0 are reached describing the random organization of the rosettes relative to each other. The fine structure is attributable to the loops aggregated in rosettes for MLS models and reflects the models in detail. The weighted box-volume function of the nuclear chromatin morphology in H2A-YFP expressing cells (Fig. 10) shows distinct power-law behaviour for two nuclei for different intensity thresholds (C, D). Its slope the box-counting dimension as function of these thresholds differs clearly for both nuclei. Thus not only the morphology can be quantified but in comparison with the simulated scaling behaviour also an MLS model is more likely supported.

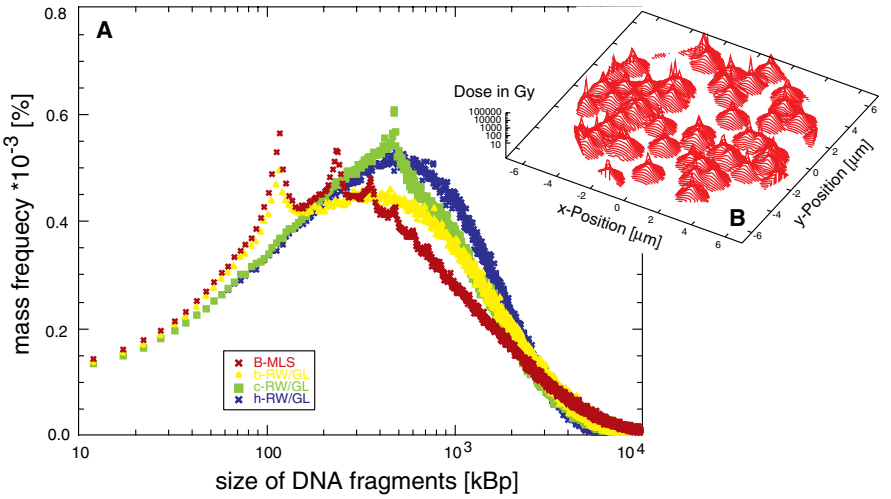


Fig. 7. *DNA Fragmentation after Irradiation with Carbon Ions.* Simulation of the length distribution of DNA fragments after irradiation with carbon ions with an inhomogeneous nuclear radiation dose distribution and thus the sites of double strand breakage (B) depend on the spatial arrangement of the chromatin fiber and reflect the used interphase topology of simulated chromosomes like loop size or loop arrangement in rosettes in detail (A). Best agreement of the simulations with experiments is reached for the MLS model.

ison to experimental fragment distribution results again in best agreement with an MLS topology with loops and linkers of 63–126 kbp.

5 Dynamics in Simulated Interphase Cell Nuclei

To determine the impact of the three-dimensional genome organization on molecular mobility, the accessability of nuclear loci and the hypothesis of the Inter-Chromosomal Domain (ICD) model, the diffusion of spheres was simulated by Brownian Dynamics in computer generated nuclei with an MLS chromatin fiber topology. The tracers interacted with the static fiber by an excluded volume potential. Visual inspection of the morphology of simulated chromosomes or nuclei revealed big spaces allowing high accessibility to nearly every spatial location. A channel like network for molecular transport between chromosome territories, as postulated by the ICD model, was not apparent in the simulations (Fig. 7 A-C). The big spaces are supported by estimating the nuclear volume occupied by chromatin of $\leq 30\,\%$, leaving $\geq 70\,\%$ of space for diffusion with an average mesh spacing of 29–82 nm for nuclei of 6–12 µm diameter. This agrees with the simulated mean displacement for 10 nm sized particles of ≈ 1–$2\,\mu$m within 10 ms. Therefore, the diffusion of biological relevant tracers is only moderately obstructed. The anomaly parameter char-

acterizing the degree of obstruction ranged from 2.0 (obstacle free diffusion) to 4.0 (Fig. 7 C), in agreement with experiments in which the obstruction of fluorophores was measured as function of the chromatin density labelled with histone autofluorescent protein fusions *in vivo* (Fig. 7 C; Wachsmuth *et al.*, 1998, Wachsmuth, 2001; Wachsmuth *et al.*, 2003). The degree of obstruction was proportional to the nuclear density, the fiber diameter, the interaction hardness and the tracer size. Different fiber topologies had no effect on the average particle displacement. Consequently, molecules and proteins might reach every nuclear location by energy independent diffusion without a special channel like network without necessity of an ICD model, which can therefore be refuted.

6 Long-Range Correlations in DNA Sequences

The sequential organization, i. e. the relations within DNA sequences, and its connection to the three-dimensional organization of genomes was investigated by correlation analyses of 113 completely sequenced chromosomes of 0.5×10^6 to 3.0×10^7 bp from Archaea, Bacteria, *Arabidopsis thaliana*, *Saccharomyces cerevisae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens*. The analyses are based on the concentration profile of single bases along the DNA sequence (Knoch *et al.*, patent): The square root of the mean-square deviation between the concentration of bases c_l in a window of length l and the concentration c_L of bases in the entire DNA sequence with length L was calculated

$$C(l) = \sqrt{\langle (c_l - c_L)^2 \rangle_{s=L-l-1}} \,.$$

For a fractal self-similar sequence like a random walk the concentration fluctuation function $C(l)$ shows power-law behaviour whose exponent characterizes the scale dependent degree of correlation. To avoid numeric instabilities infinitely exact calculation tools provided by the GNU multiple precision package GMP were implemented. Calculations were performed on PCs and an IBM SP2, using in total ≈ 5000 h CPU time. On the latter the analyses were split into jobs of a few minutes computing single or few windows, thus being an extremely efficient filler for the unavoidable gaps in batch mode of big parallel machines. The analyses would also be predestined for grid computing, e. g. a screen saver application.

All sequences revealed long-range power-law correlations almost on the entire observable scale (Fig. 9). The local correlation coefficient shows close to random correlations on the scale of a few base pairs, a first maximum from 40–3400 bp (for *Arabidopsis thaliana* and *Drosophila melanogaster* divided in two submaxima), and often a region of one or more second maxima from 10^5–3×10^5 bp. This multi-scaling behaviour was species specific. Computer generated random sequences assuming a block organization of genomes reproduced such multi-scaling. Within this multi-scaling behaviour an additional

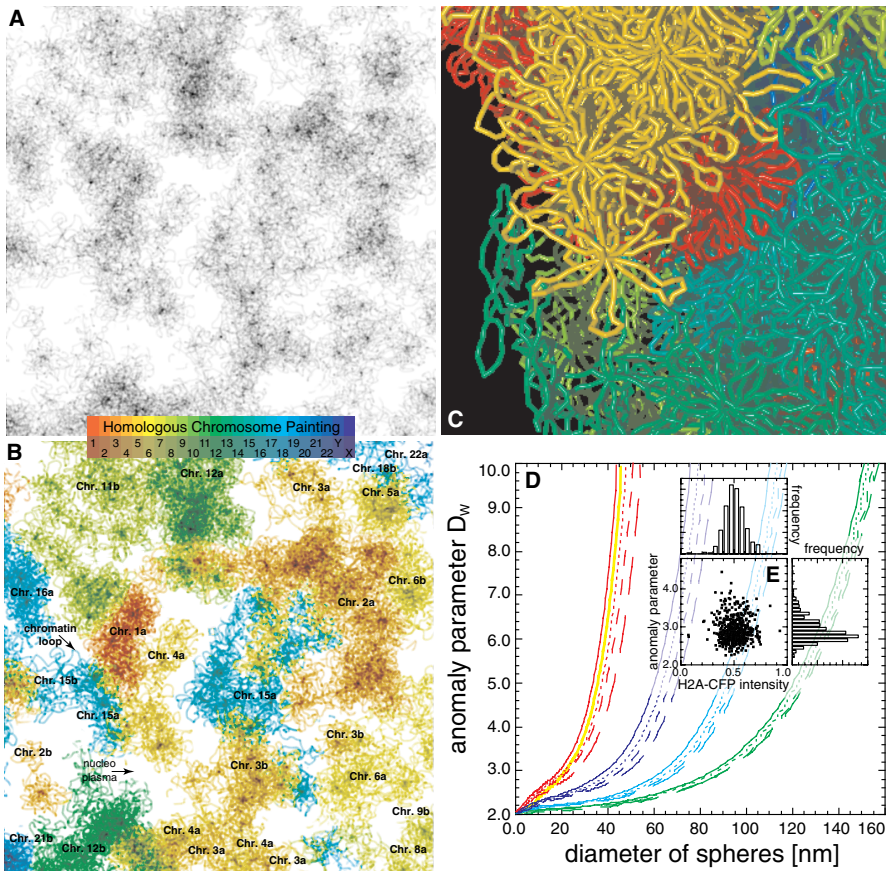


Fig. 8. *Dynamic Properties of Simulated Nuclei: Diffusion of Particles.* The simulated electron microscopic image (A) and chromosome map (C) for a 126-126-MLS[5] nucleus (deliberately unrelaxed, thus the voids between chromosome territories are an artefact) and the detailed rendered view of a 126-126-MLS[6] nucleus demonstrate the low overlap of chromosome territories and the MLS rosettes. The mean spacing between chromatin fibers ranges from 50–100 nm, thus small chemical substances as nucleotides or ATP molecules reach every location in the nucleus and most relevant proteins or protein subunits should only be obstructed moderately, regarding the scale. This refutes the Inter Chromosomal Domain (ICD) hypothesis. The anomaly parameter characterizing the degree of obstruction of diffusing spheres in simulated nuclei (D), agrees with this semi-quantitative view and is proportional to the nuclear radius and therefore density (nuclear radii of 3, 4, 5, and 6 µm are red, blue, light blue and green). Changes of the chromatin fiber diameter leads to a smaller proportionality (diameters of 25, 30, 35 and 40 nm are solid, dotted, dashed, and long dashed). This agrees with the small obstruction and its minor dependence of the chromatin morphology, determined *in vivo* for Alexa–568 fluorescence molecules by two colour fluorescence correlation spectroscopy (FCS) as function of the chromatin density labelled by H2A-CFP fusionproteins (E, Wachsmuth, 2001; the anomaly parameter is bigger than expected due to binding effects of Alexa–568).

fine structure is present and attributable to the codon usage in all except the human sequences. Here it is connected to nucleosomal binding. Computer generated random sequences assuming the codon usage and nucleosomal binding agree with these results. Mutation by sequence reshuffling destroyed all correlations, thus their stability seems evolutionary tightly controlled and connected to the spatial genome organization on large scales. This is supported by the scaling behaviour of the chromatin topology (see above). The correlation behaviour was used to construct trees, which were similar to the corresponding phylogenetic trees for β-Tubulin genes of *Oomycetes* and Eukarya genomes. For Archaea and Bacteria tree construction led to a new classification system with four major tree branches/classes. In summary, these findings suggest a complex sequential organization of genomes closely connected to their three-dimensional organization.

7 "Chromatin Alive": *In vivo* Labelling of the Chromatin Morphology

The *in vivo* morphology and dynamics of chromatin is difficult to assess by electron microscopy, fluorescence *in situ* hybridization (FISH) and *in vivo* stains since these methods require fixation or produce artefacts. To overcome these limitations a novel *in vivo* technique for chromatin labelling was established (Knoch *et al.*, patent): DNA vectors encoding the fusion proteins of all histones H1.0, H2A, mH2A1.2, H2B, H3, H4 and the autofluorescent proteins CFP, GFP, YFP, DsRed1 DsRed2 were developed and expressed stably in HeLa, LCLC103H, Cos7 and ID13 cells. 2.6 to $\approx 20\,\%$ of the nucleosomes carry a label. No apparent influence of the cell cycle status, the proliferation rate or the AFP fluorescent excitation/emission spectra, but recently a somewhat increased nucleosomal repeat length was detected (Weidemann, 2002). With this approach the structure and dynamics of histones, nucleosomes, chromatin, chromosomes and whole nuclei during cell cycle, differentiation, and apoptosis could be investigated artefact-free *in vivo* (Fig. 10). The interphase morphology showed globular structures as predicted by the Multi-Loop-Subcompartment model. All stages of mitosis as well as apoptosis were clearly distinguishable (Gil-Parado *et al.*, 2002, 2003). Deacetylase inhibitors led to a smoothing of the interphase morphology. With this *in vivo* chromatin label the interphase morphology and changes thereof could be investigated by quantitative scaling and statistical analyses (Fig. 6 C-E). The technique could also be applied for cell culture control and counterstaining, or *in organo* and *in organismo* by creation of transgenic animals.

This now widely used *in vivo* method led also to the discovery of construct conversions in simultaneous co-transfections (Bestvater *et al.*, patent, 2002), a convenient and widely used approach in multicolour labelling experiments *in vivo*, using green fluorescent proteins with distinct spectral characteristics. These co-transfections can cause false positive results due to con-

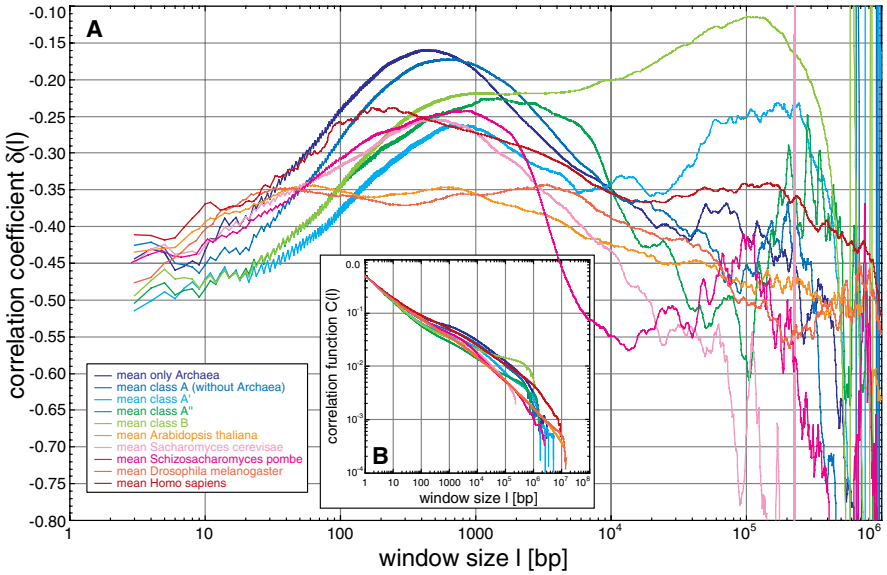


Fig. 9. *Fine-Structured Multi-Scaling Long-Range Correlations in Completely Sequenced Genomes.*
The average over the single analysed chromosomes (A) for each of the Eukarya genomes, the Archaea and the Bacteria classes A, A', A" and B, reveals that only *Homo sapiens* does not show the zig-zag pattern due to the codon usage, although it shows a nucleosome dependent fine structure not present in the other sequences. All genomes show a maximum between window sizes of 100–1000 bp of which only the maxima of *Homo sapiens* is connected to the nucleosome. The classes A' and B show a second maximum after a decrease of $\delta(l)$ with very high correlations for window lengths of $\approx 10^5$ bp. Beyond, only *Homo sapiens* shows a distinct second peaked maximum which is washed out in the average but existing in the of the single analysed human chromosomes and indicating a looped and clustered chromatin topology. For comparison the means of the concentration fluctuation function for the same averages are shown (B).

version of their spectral properties. Standard transfection result in $\approx 8\,\%$, depending on the treatment of the DNA up to 26 %, of the cells expressing altered fusion proteins. This could lead to severe misinterpretation of the results. The conversion is independent of the transfection method and the cell type. The results show that conversion is based on homologous recombination/repair/replication (RRR) events occurring between the nucleotide sequences of the fluorescent proteins. Conversion can be avoided by consecutive transfection or by fluorescent constructs with low sequence similarities. The appearance of conversion allows to easily exchange spectral properties in fusion proteins, to create libraries or to assemble DNA fusion constructs *in vivo*. The detailed quantification of the conversion rate could be used to investigate RRR processes in general.

8 Conclusions

To achieve a deeper understanding of the human genome the structural, scaling and dynamic properties in the simulation of interphase chromosomes and cell nuclei were determined and combined with the analysis of long-range correlations in completely sequenced genomes as well as the analysis of the chromatin distribution *in vivo*: Integration of these different aspects suggests that the cell nucleus can be viewed as an optimized bioreactor in which the sequential and three-dimensional organization coevolved:

Advancing the interphase nucleus from the cellular level reveals a globular morphology. Staining of the single chromosomes reveals that these form chromosome territories, which are arranged non-randomly. The globular morphology is created by aggregates of chromatin loops within the chromosomes. Increasing the resolution reveals that the underlying chromatin fiber consists of nucleosomes around which the DNA is wound. Analysing the DNA base pair sequence shows a complex organization which can be linked to the codon usage, the nucleosome and the chromatin fiber topology on larger scales. Thus, every structural level of nuclear organization is connected and represented in all the other levels. Features present on one scale are reflected on other scales, and changes on one scale might either reflect or induce changes on other scales. These structural links are best described by scaling analyses.

Beyond the structural also the dynamics of the three-dimensional organization itself, i. e. chromosomes or chromatin loops, or the mobility of particles inbetween is scale dependent. Chromosomes or large protein complexes move slowly, in contrast to small and highly mobile molecules. Due to the low volume occupancy of the three-dimensional topology the mobility of medium sized molecules is only moderately obstructed. Thus, most molecules and proteins can reach nearly every location in the nucleus by simple diffusion very quickly and can commit to their function. Therefore, the dynamics is also closely connected to the underlying or surrounding structure, i. e. structural changes shape also the accessibility by molecules.

Consequently, the local, global and dynamic characteristics of cell nuclei, are tightly inter-connected, obviously due to their coevolution. Beyond, however, this view of the nucleus as an entity stresses, that its function can only be fulfilled by the integrated whole and that the information for processes from the cellular level, over embryogenesis to cognitive ability is present in this integrated whole holisticly.

Acknowledgements

The Supercomputing Center Karlsruhe (SCC; grant ChromDyn), and the Computing Facility of the Deutsches Krebsforschungszentrum (DKFZ) are thanked for access to their IBM SP2s. This work was supported by the Bundesministerium für Bildung und Forschung (BMBF) under grant 01

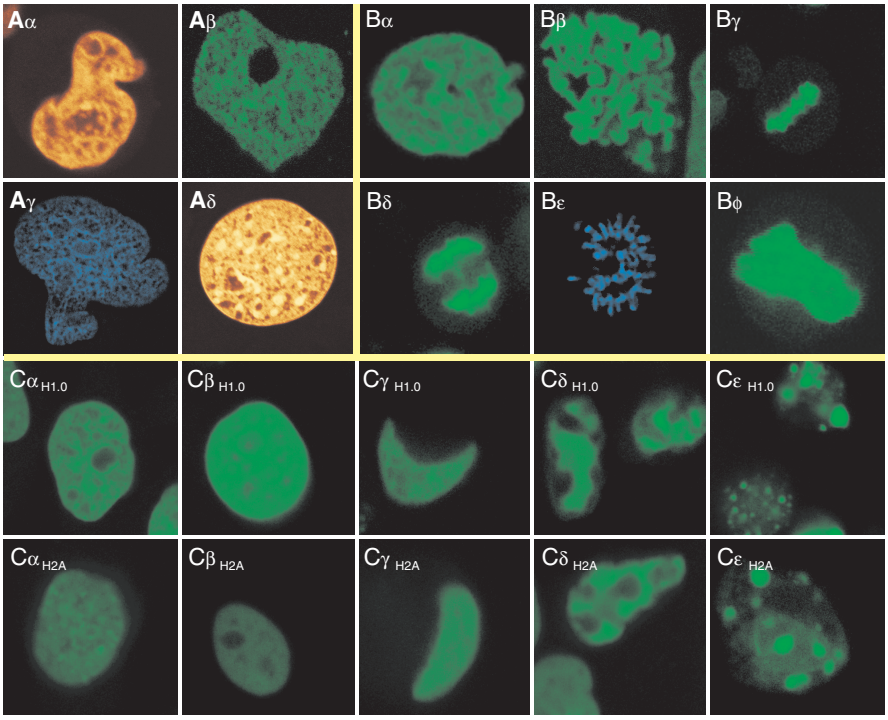


Fig. 10. *In vivo Labelling of the Chromatin Morphology in Interphase, Mitotic and Apoptotic Nuclei.*
Expression of H2A-YFP fusionproteins reveals different interphase morphologies and thus different genome organizations in human HeLa-H2A-YFP (Aα), the primate Cos7-H1-GFP (Aβ), the human LCLC103H-H2A-CFP (Aγ) and the mouse ID13-H2A-YFP (Aδ) cell lines. During mitosis all stages of the cell division are visible: At the beginning of prophase chromosomes start condensing (Bα) into cylindrical chromosomes (Bβ), pair in prometaphase and arrange in the metaphase plate in metaphase (Bγ). Then the chromosomes are draged appart into the daughter cells in anaphase (Bδ; Bε, H2A-CFP), before decondensation again into interphase during telophase (Bϕ). The induction of apoptosis by the deacetylation inhibitor Sodiumbutyrat in Hela cells expressing H1.0-YFP or H2A-YFP indicates interphase changes and apoptosis stages until total cellular destruction: In contrast to control nuclei (Cα), treated nuclei show rapidly a lower granularity (Cβ), before apoptosis sets in with the transformation to "apoptotic half moon" shaped nuclei and first chromatin agglomerations (Cγ). Then the chromatin condenses totally (Cδ) until apoptotic fragmentation of nuclei (Cε). Image sidelength: 20 µm: Aα Aβ; 25 µm: Aδ, Bα, Bβ, Bϕ; 30 µm: Aγ; 35 µm: Bγ, Bδ, Bε, C.

KW 9602/2 (Heidelberg 3D Human Genome Study Group, German Human Genome Project). T. A. Knoch was kindly provided with a dissertation grant of the Deutsches Krebsforschungszentrum (DKFZ).

References

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. & Watson, J. D. Molecular Biology of the Cell. 3rd ed., *Garland Publications Inc., New York & London*, ISBN0-8240-7283-9, 1994.

Bestvater, F., Knoch, T. A. & Spiess, E. Nachweisverfahren für homologe Rekombinationsereignisse. *German Patent Application 10116826.8* and *International Patent Application PCT/DE02/01207*.

Bestvater, F., Knoch, T. A., Langowski, J. & Spiess, E. GFP-Walking: Artificial construct conversions caused by simultaneous cotransfection. *BioTechniques 32(4)*, 844-854, 2002.

Boveri, T. Die Blastomerenkerne von *Ascaris meglocephala* und die Theorie der Chromosomenindiviualität. *Archiv für Zellforschung 3*, 181-268, 1909.

Comings, D. E. The rationale for an ordered arrangement of chromatin in the interphase nucleus. *Am. J. Hum. Genet. 20*, 440-460, 1968.

Comings, D. E. Mechanisms of chromosome banding and implications for chromosome structure. *Annu. Rev. Genet. 20*, 440-460, 1978.

Cremer, T. & Cremer, C. Chromosome territories, nuclear architecture, and gene regulation in mammalian cells. *Nat. Rev. Gen. 2(4)*,292, 2001.

Friedl, A. A. Development of simulation methods and their application in the qualitative and quantitative analysis of DNA double strand breaks in lower and higher Eukarya. *Dissertation*, Ludwig-Maximilians Universität, Munich, Germany, 1994.

Gil-Parado, S., Fernàndez-Montalvàn, A., Assfalg-Machleidt, I., Popp, O., Bestvater, F., Holloschi, A., Knoch, T. A., Auerswald, E. A., Welsh, K., Reed, J. C., Fritz, H., Fuentes-Prior, P., Spiess, E., Salvesen, G. & Machleidt, W. Ionomycin-activated calpain triggers apoptosis: A probable role for Bcl-2 family members. *J. Biol. Chem. 277(30)*, 27217-27226, 2002.

Gil-Parado, S., Popp, O., Knoch, T. A., Zahler, S., Bestvater, F., Felgenträger, M., Holoshi, A., Fernàndez-Montalvàn, A., Auerswald, E. A., Fritz, H., Fuentes-Prior, P., Machleidt, W. & Spiess, E. Subcellular localization and subunit interactions of over-expressed human μ-calpain. *J. Biol. Chem.*, 2003. (in press)

Knoch, T. A., 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*.

Knoch, T. A., Münkel, C. & Langowski, J. Three-dimensional organization of chromosome territories and the human cell nucleus - about the structure of a self replicating nano fabrication site. *Foresight Institute - Article Archive*, Foresight Institute, Palo Alto, CA, USA, http://www. foresight.org/, 1-6, 1998.

Knoch, T. A. Dreidimensionale Organisation von Chromosomen-Domänen in Simulation und Experiment. (Three-dimensional organization of chromosome domains in simulation and experiment.) *Diploma Thesis*, Faculty for Physics and Astronomy, Ruperecht-Karls Universität Heidelberg, Heidelberg, Germany, 1998 and TAK Press, Tobias A. Knoch, Mannheim, Germany, ISBN 3-00-010685-5, 2003.

Knoch, T. A., Münkel, C. & Langowski, J. Three-Dimensional Organization of Chromosome Territories and the Human Interphase Nucleus. *High Performance Scientific Supercomputing*, Wilfried Juling (ed.), Scientific Supercomputing Center (SSC) Karlsruhe, University of Karlsruhe (TH), 27-29, 1999.

Knoch, T. A., Münkel, C. & Langowski, J. Three-dimensional organization of chromosome territories in the human interphase nucleus. *High Performance Computing in Science and Engineering 1999*, Egon Krause and Willi Jäger (eds.), Springer Berlin-Heidelberg-New York, ISBN 3-540-66504-8, 229-238, 2000.

Knoch, T. A. Approaching the Three-Dimensional Organization of the Human Genome. *Dissertation*, Ruprecht-Karls Universität Heidelberg, Heidelberg, Germany, and TAK Press, Tobias A. Knoch, Mannheim, Germany, ISBN 3-00-009959-X (soft cover, 3rd ed.), ISBN 3-00-009960-3 (hard cover, 3rd ed.), 2002.

Knoch, T. A., Göcker, M. & Lohner, R. Methods for the analysis, classification and/or tree construction of sequences using correlation analysis. *US Patent Application 60/436.056*.

Luger, C., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature 389*, 251–260, 1997.

Monier, K. Cartographie linéaire et tridimensionnelle du génome humain par hybridation *in situ* fluorescente et imagerie microscopique digitale. *Dissertation*, Institut Albert Bonniot, Université Joseph Fourier Grenoble I, Grenoble, France, 1997.

Münkel, C. & Langowski,J. Chromosome structure described by a polymer model. *Phys. Rev. E 57 (5-B)*, 5888-5896, 1998.

Münkel, C., Eils, R., Zink, D., Dietzel, S., Cremer, T. & Langowski, J. Compartmentalization of interphase chromosomes observed in simulation and experiment. *J. Mol. Biol. 285 (3)*, 1053-1065, 1999.

Paulson, J. R. & Laemmli, U. K. The structure of histone-depleted metaphase chromosomes. *Cell 12*, 817-828, 1980.

Pienta, K. J. & Coffey, D. S. A structural analysis of the role of the nuclear matrix and DNA loops in the organization of the nucleus and chromosome. *J. Cell. Sci. Suppl. 1*, 123-135, 1984.

Rabl, C. Über Zellteilung. *Morphologisches Jahrbuch 10*, 214-330, 1885.

Rauch, J. Spektrale Präzisionsdistanzmikroskopie zur Untersuchung der 3D-Topologie ausgewählter DNA-Punktmarker. *Dissertation*, Faculty for Physics and Astronomy, Ruprecht-Karls Universität Heidelberg, Heidelberg, Germany, 1999.

Reznik N. A., G. P. Yampol, Kiseleva, E. V., Khristolyubova, N. B. & Gruzdev, A. D. Possible functional Structures in the chromomere, *Nuclear Structure and Function.* by Harris, J. R. & Zbarsky, I. B. (eds.), Plenum Press, New York - London, 27-29, 1990.

Sachs, R. K., van den Engh, G., Trask, B., Yokota, H. & Hearst, J.E. A random-Walk/Giant-Loop model for interphase chromosomes. *Proc. Nat. Acad. Sci. USA 92*, 2710-2714, 1995.

Trask, B.J., Allen, S., Massa, H., Fertitta, A., Sachs, R., van den Engh, G. & Wu, M. Studies of metaphase and interphase chromosomes using fluorescence *in situ* hybridization. *Cold Spring Harb. Symp. Quant. Biol. 58*, 767-775, 1993.

Voet, D & Voet, J. G. Biochemistry. 2nd ed., *John Wiley & Sons Inc.*, 1995.

Vogel, F. & Schroeder, T. M. The internal order of the interphase nucleus. *Humangenetik 25 (4)*, 265-97, 1974.

Wachsmuth, M, Waldeck, W. & Langowski, J. Anomalous diffusion of fluorescent probes inside living cell nuclei investigated by spatially-resolved fluorescence correlation spectroscopy. *J. Mol. Biol. 298 (4)*, 677-686, 2000.

Wachsmuth, M. Fluorescence fluctuation microscopy: Design of a prototype, theory and measurements of the mobility of biomolecules in the cell nucleus. *Dissertation*, Faculty for Physics and Astronomy, Ruprecht Karls Universität Heidelberg, Heidelberg, Germany, 2001.

Wachsmuth, M., Weidemann, T., Müller, G., Urs W. Hoffmann-Rohrer, Knoch, T. A., Waldeck, W. & Langowski, J. Molecules between binding and diffusion: A quantitative approach continuous photobleaching. *Biophys. J. 84 (5)*, 3353–3363, 2003.

Wedemann, G. & Langowski, J. Computer simulation of the 30-nanometer chromatin fiber. *Biophys. J. 82(6)*, 2847-2859, 2002.

Weidemann, T. Quantitative investigation of distribution, mobility, and binding of fluorescently labeled histones *in vitro* and *in vivo* with fluorescence fluctuation microscopy. *Dissertation*, Faculty for Physics and Astronomy, Ruprecht-Karls Universität Heidelberg, Heidelberg, Germany, 2002.

Zirbel, R. M., Mathieu, U. R., Kurz, A., Cremer, T. & Lichter, P. Evidence for a nuclear compartment of transcription and splicing located at chromosome domain boundaries. *Chrom. Res. 1 (2)*, 93-106, 1993.