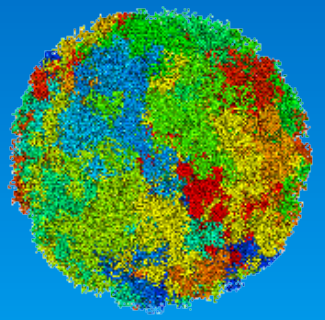
Towards a Holistic Understanding of the Human Genome by Determination and Integration of its Three-Dimensional and Sequential Organization



-

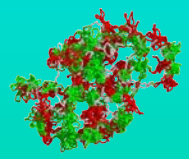
Structural-, Scaling and Dynamic-Properties
in the Simulation of
Interphase Chromosomes and Cell Nuclei

Long-Range Correlations in Complete Sequenced Genomes

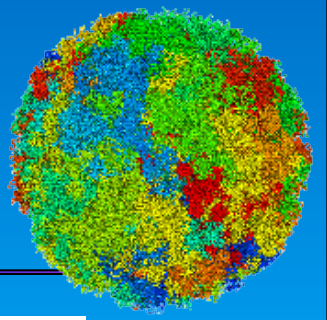
by

Tobias A. Knoch

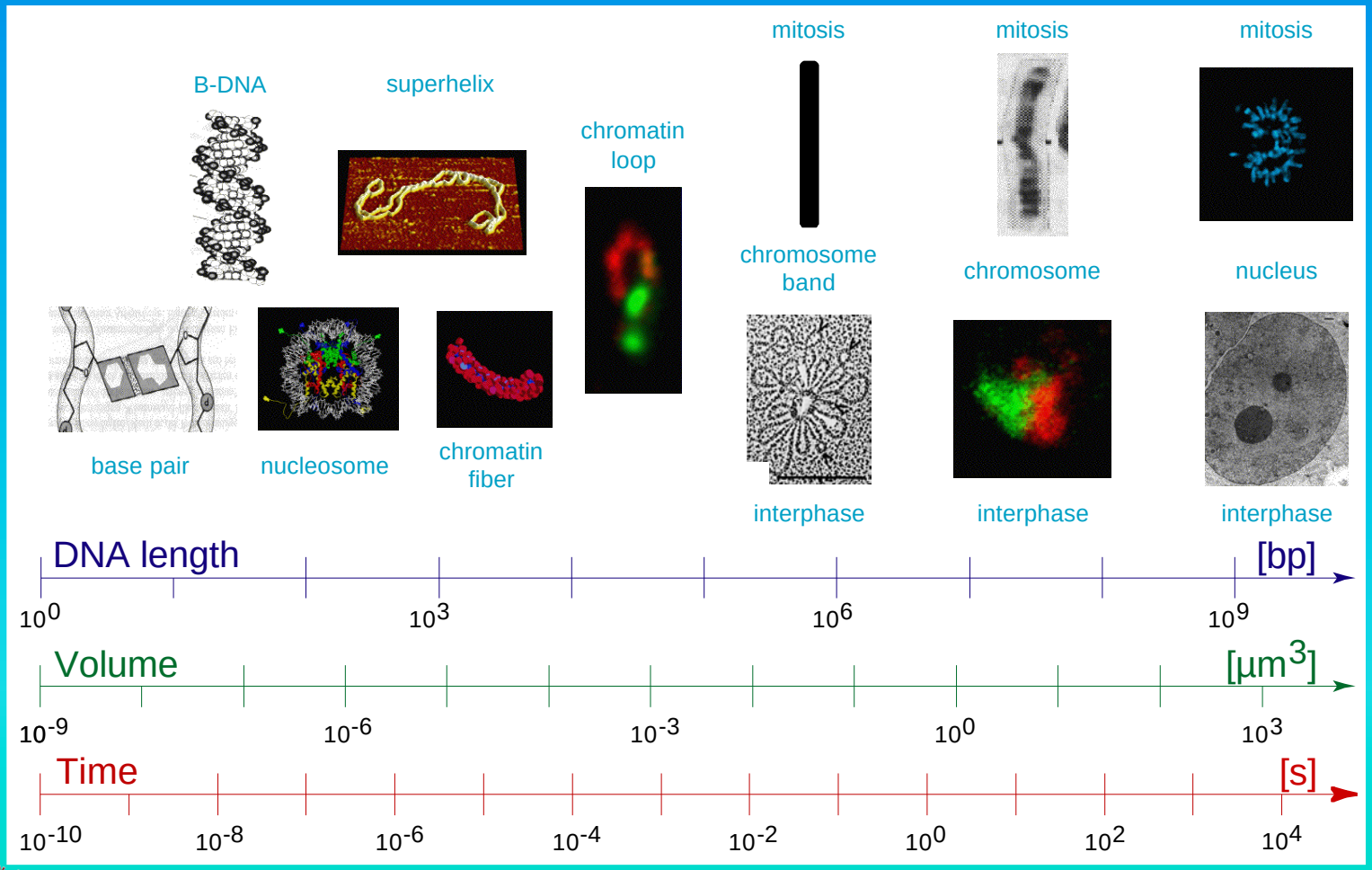
Kirchhoff Institut for Physics, Ruperto-Carola University
&
Deutsches Krebsforschungszentrum (DKFZ)
Heidelberg
Germany

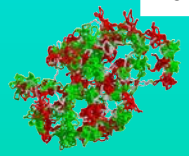


Dynamic and Hierarchical Genome Organization



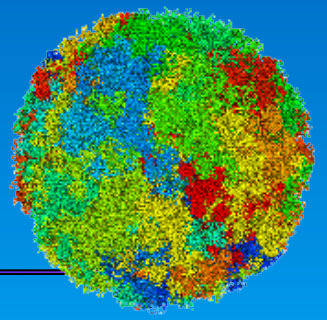
10 and 13 orders of magnitude concerning length and time scales are bridged.
Are and how are all of these organization levels connected to fullfill their obvious functions, e. g. gene regulation or replication, since they are optimized by evolution ?

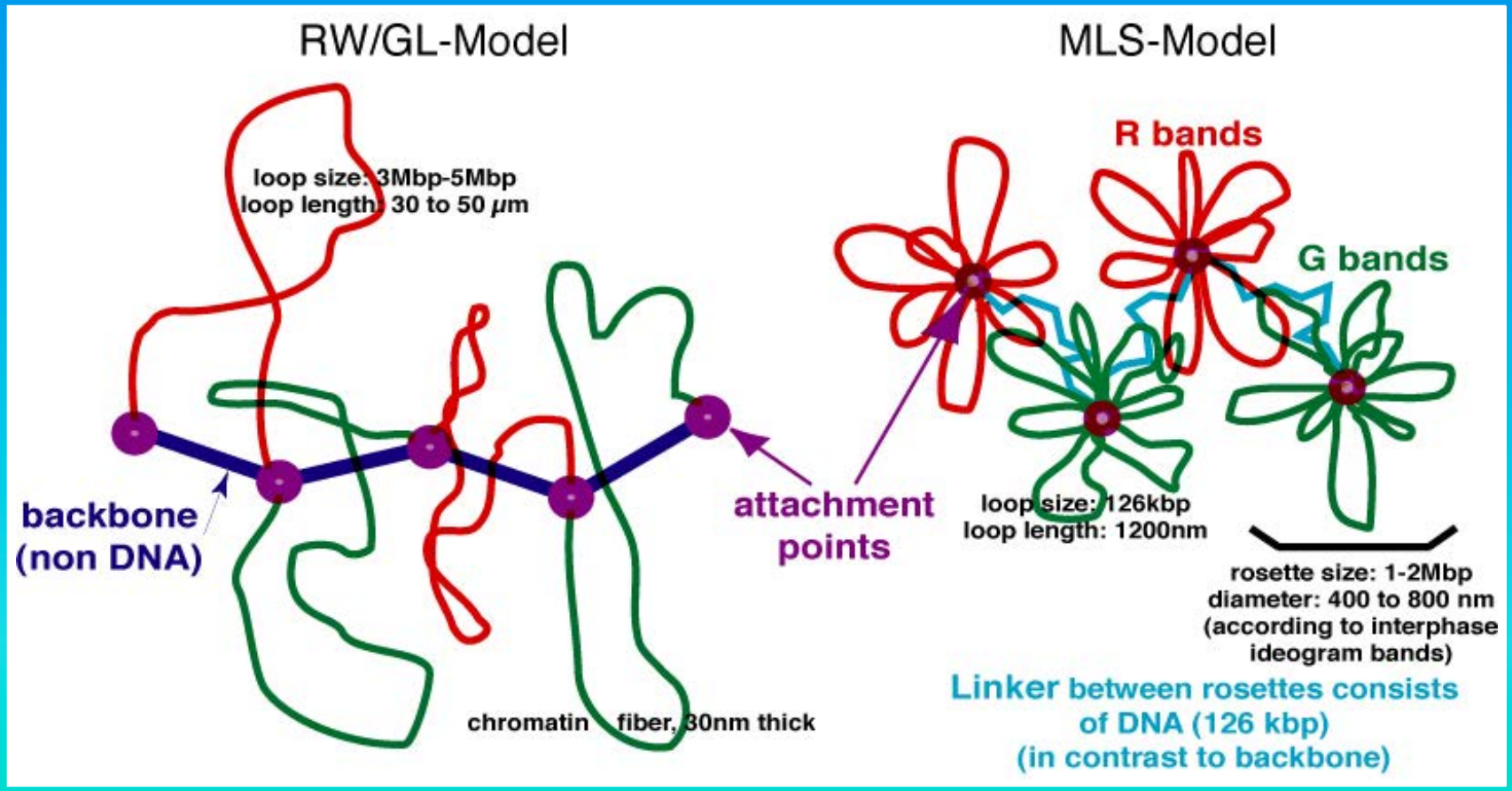


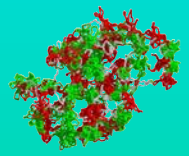


Simulated Interphase Chromosome Models

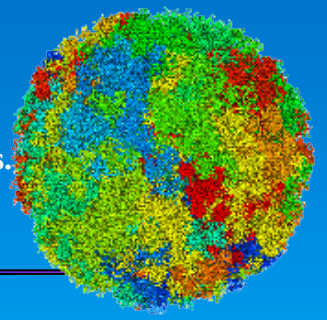
Random-Walk/Giant-Loop (RW/GL) and Multi-Loop-Subcompatment (MLS) Model







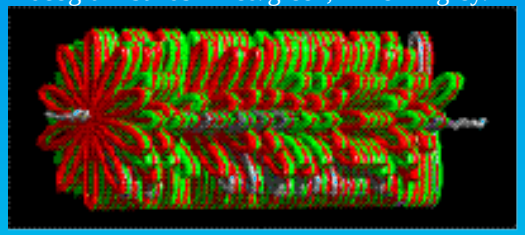
Simulation of Single Chromosomes

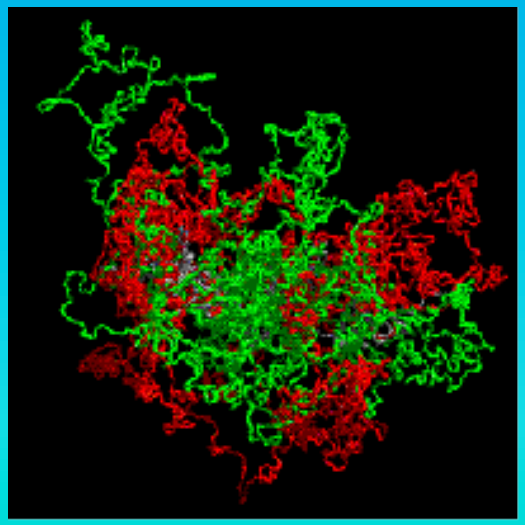


The 30 nm chromatin fiber is modeled as a polymer chain with stretching, bending, and excluded volume interactions. Monte Carlo and Brownian Dynamic methods lead to thermodynamical equilibrium configurations.

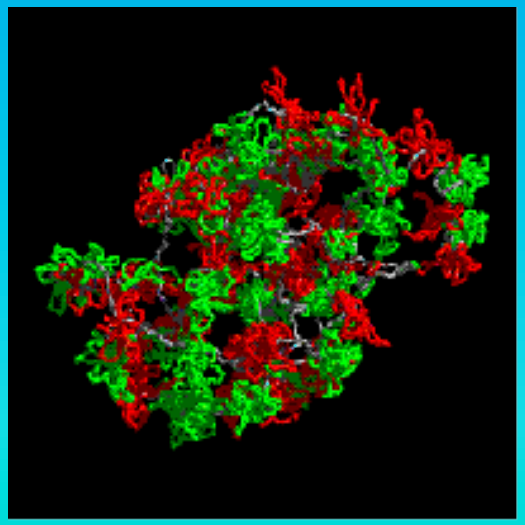
All models form chromosome territories with big voids and different chromatin morphologies. Experimental territory and subcompartment diameters agree best with an MLS model with 80 to 120 kbp loops and linkers.

Metaphase starting configuration with ideogram bands in red/green, linker in grey.

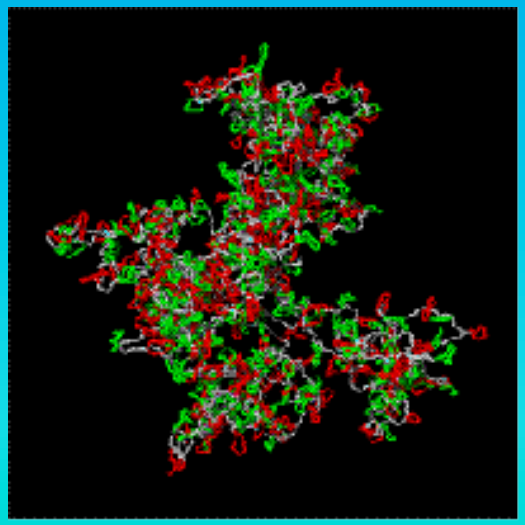




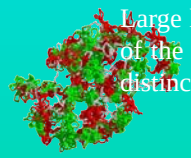
RW/GL model, loop size 5 Mbp, after ~80.000 MC and 1000 relaxing BD steps. Large loops intermingle freely and reach out of the chromsome territory, thus forming no distinct features like in MLS model.



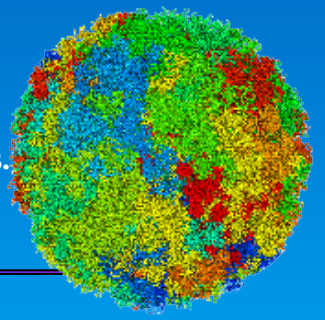
MLS model, loop size 126kbp, linker size 126 kbp, after ~50.000 MC and 1000 relaxing BD steps. Here rosettes form subcompartments as separated organizational and dynamic entities.



RW/GL model, loop size 126 kbp, after ~80.000 MC and 1000 relaxing BD steps. Large loops intermingle freely thus forming no distinct features like in MLS model.



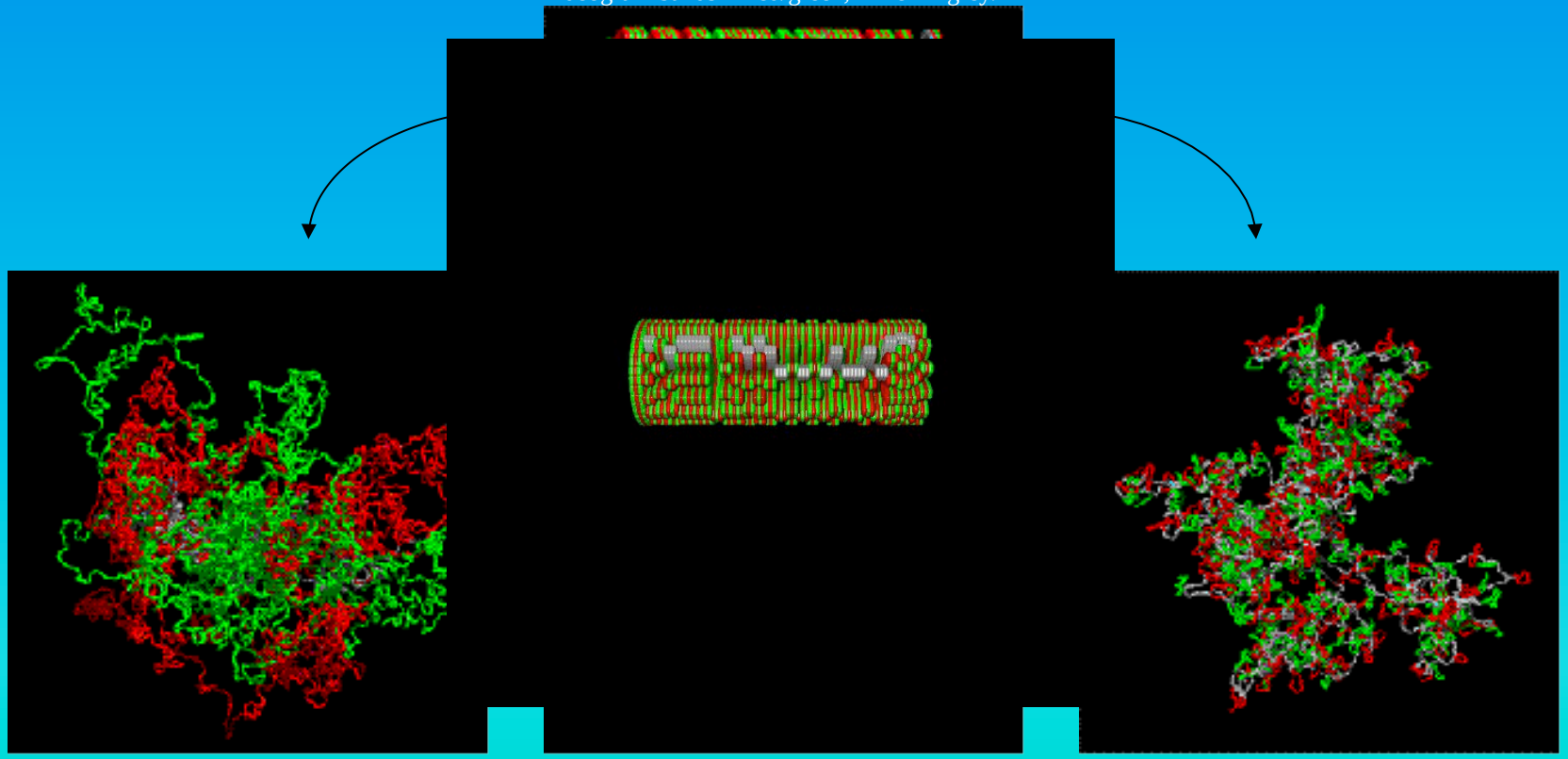
Simulation of Single Chromosomes



The 30 nm chromatin fiber is modeled as a polymer chain with stretching, bending, and excluded volume interactions. Monte Carlo and Brownian Dynamic methods lead to thermodynamical equilibrium configurations.

All models form chromosome territories with big voids and different chromatin morphologies. Experimental territory and subcompartment diameters agree best with an MLS model with 80 to 120 kbp loops and linkers.

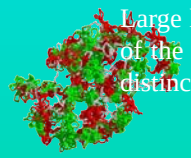
Metaphase starting configuration with ideogram bands in red/green, linker in grey.



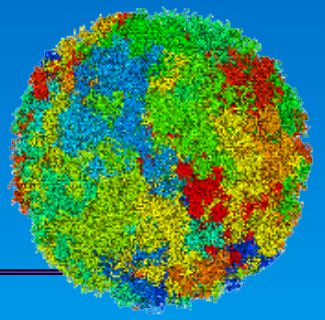
RW/GL model, loop size 5 Mbp, after ~80.000 MC and 1000 relaxing BD steps. Large loops intermingle freely and reach out of the chromsome territory, thus forming no distinct features like in MLS model.

MLS model, loop size 126kbp, linker size 126 kbp, after ~50.000 MC and 1000 relaxing BD steps. Here rosettes form subcompartments as separated organizational and dynamic entities.

RW/GL model, loop size 126 kbp, after ~80.000 MC and 1000 relaxing BD steps. Large loops intermingle freely thus forming no distinct features like in MLS model.

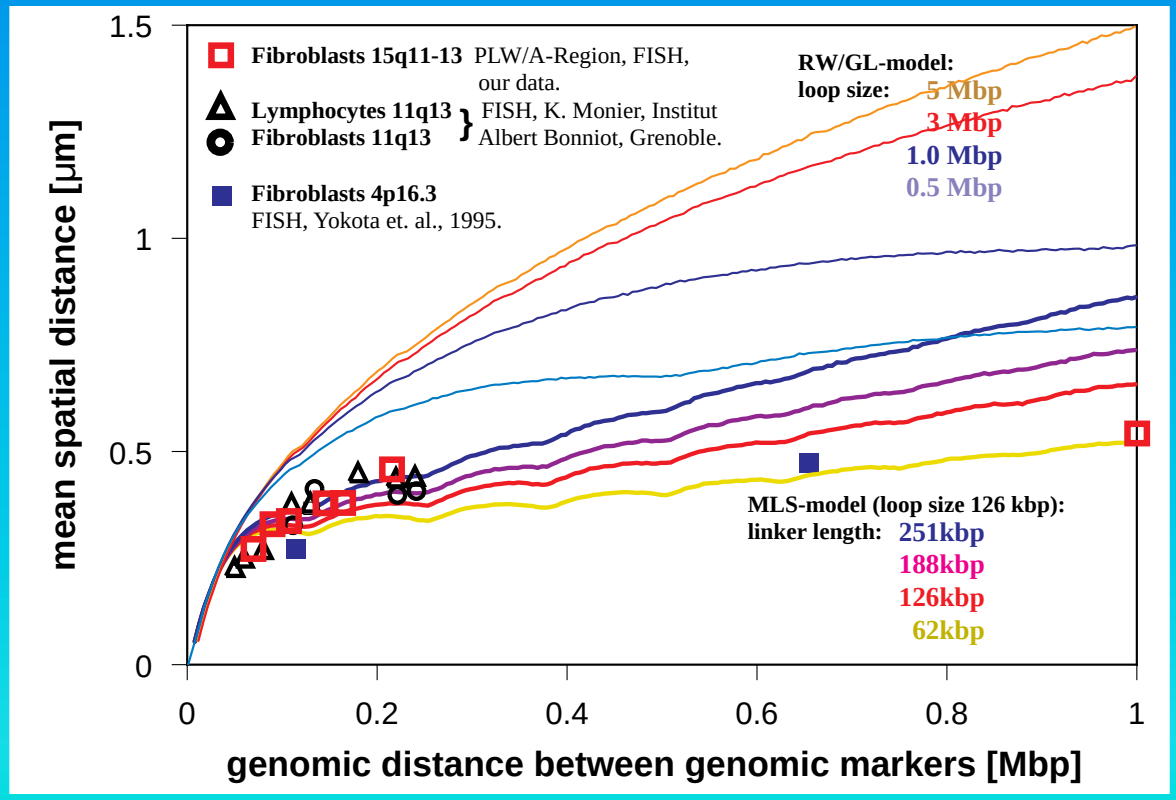


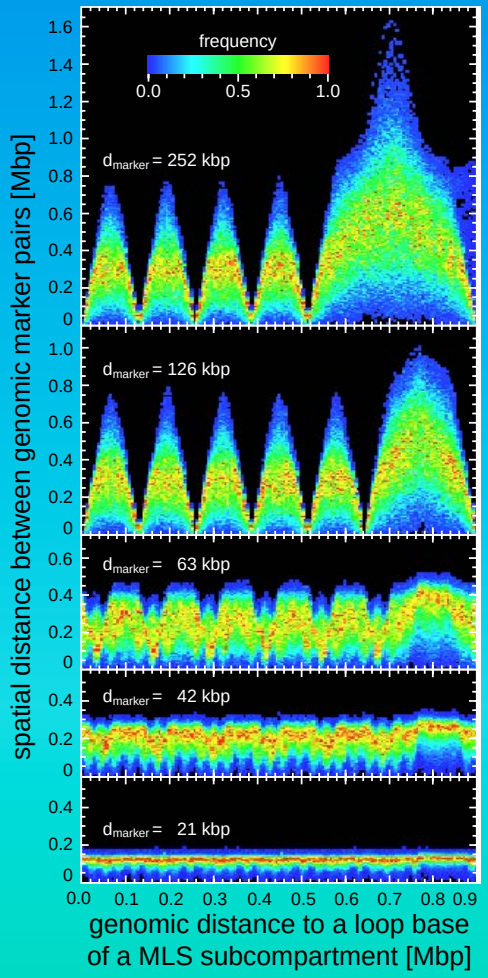
Spatial Distances between Genetic Markers

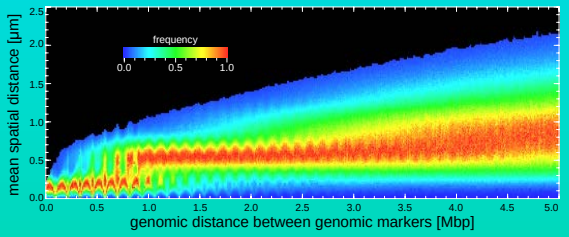


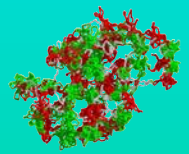
Simulated spatial distances between random genetic markers as function of their genetic separation leads to best agreement in a comparison to experiments for an MLS model with 80 to 120 kbp loops and linkers.

The spatial distance distributions are also model characteristic and show in a set of markers as function of their relative position to the chromatin fiber topology characteristic variation, strongly connected.

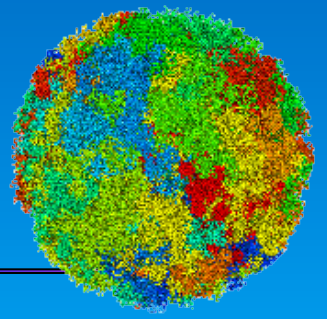






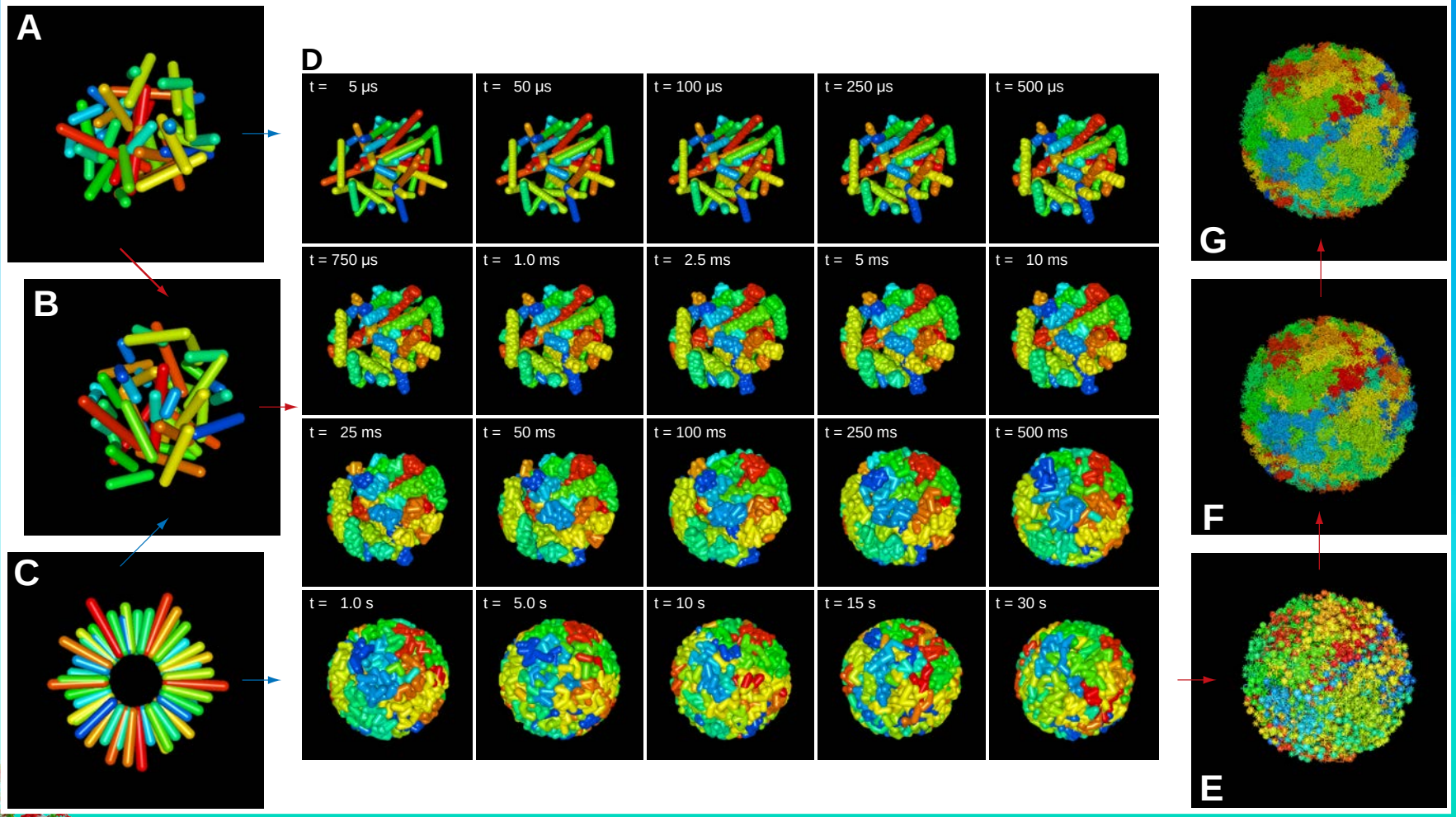


Simulation of Whole Nuclei with all 46 Chromosomes

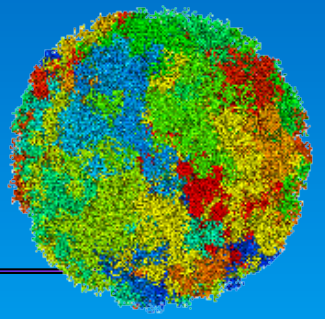


Starting with some metaphase arrangement of cylindrical chromosomes, interphase nuclei with a 30 nm fiber resolution and at thermodynamical equilibrium are created in 4 steps using simulated annealing and Brownian Dynamics methods with stretching, bending, excluded volume and a spherical boundary interactions.

The chromosome territory position depends on their metaphase position and is reasonably stable.

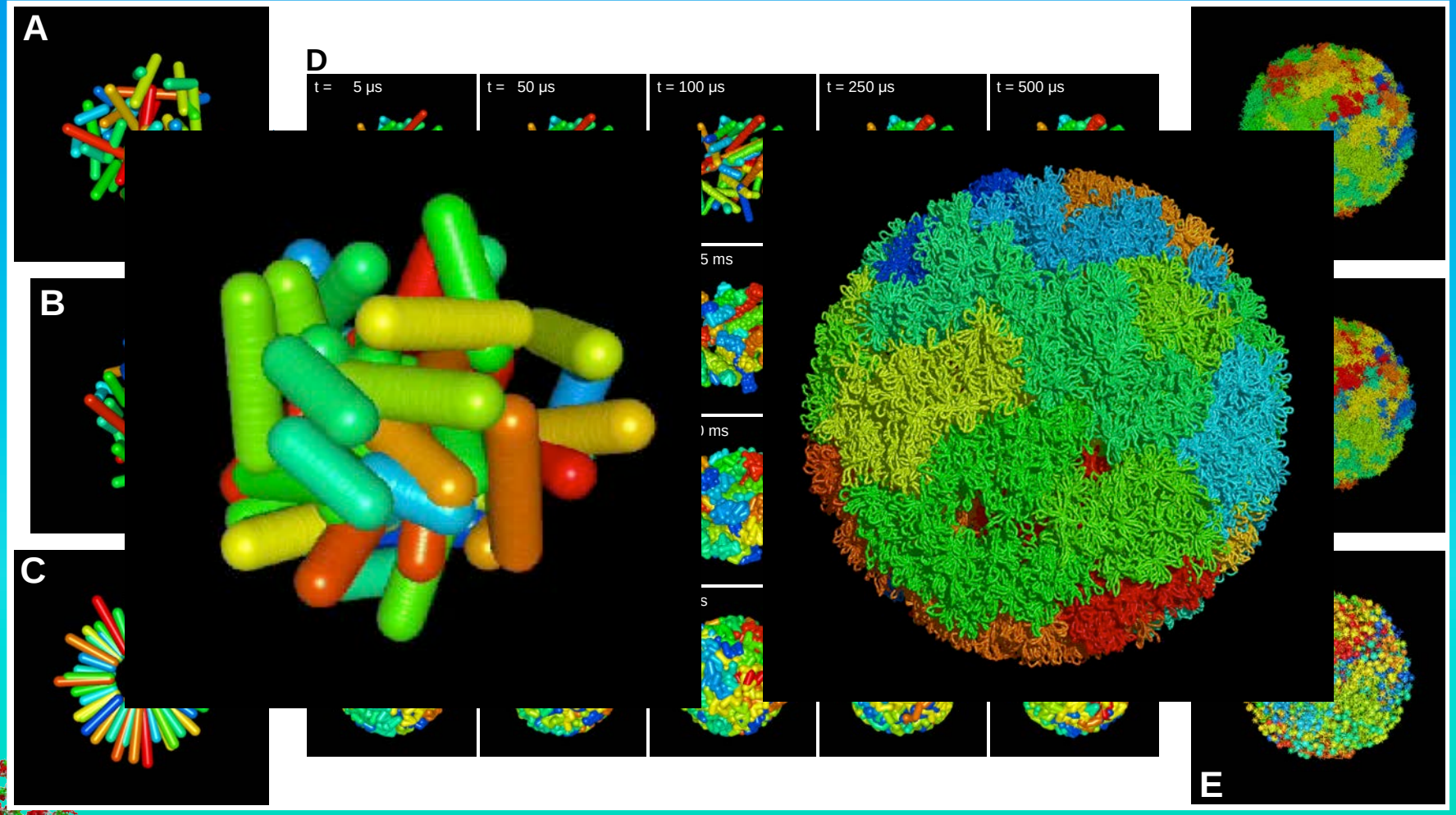


Simulation of Whole Nuclei with all 46 Chromosomes

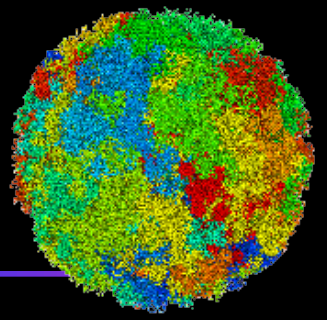


Starting with some metaphase arrangement of cylindrical chromosomes, interphase nuclei with a 30 nm fiber resolution and at thermodynamical equilibrium are created in 4 steps using simulated annealing and Brownian Dynamics methods with stretching, bending, excluded volume and a spherical boundary interactions.

The chromosome territory position depends on their metaphase position and is reasonably stable.



From Fiber Topology to Nuclear Morphology



Chromosome territories form in the RW/GL and the MLS model. However, only the MLS model leads distinct subcompartments and low chromosome and subcompartment overlap. Best agreement is reached for an MLS model with 80 to 120 kbp loops and linkers in nuclei with 8 to 10 µm diameter.

The simulated nuclear morphology reflects the chromosome fiber topology of different models in detail.

rendering

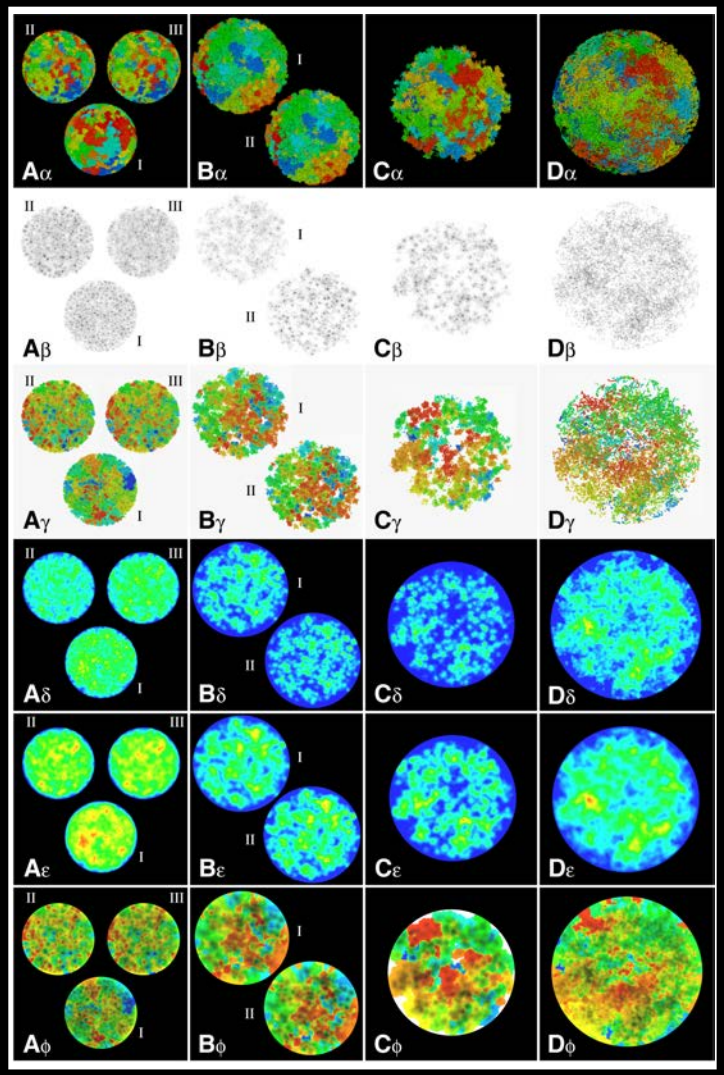
electron microscopy

electron microscopy territory painting

confocal microscopy
100x objective, theoretic resolution

confocal microscopy
63x objective, real resolution

confocal microscopy
territory painting

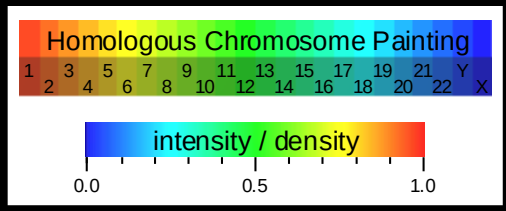


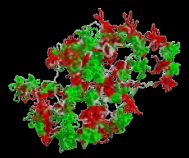
A: MLS in 6 µm nucleus
I: 63 kbp loops, 63 kbp linkers
II: 63 kbp loops, 252 kbp linkers
III: 126 kbp loops, 252 kbp linkers

B: MLS in 8 µm nucleus
I: 126 kbp loops, 126 kbp linkers
II: 84 kbp loops, 126 kbp linkers

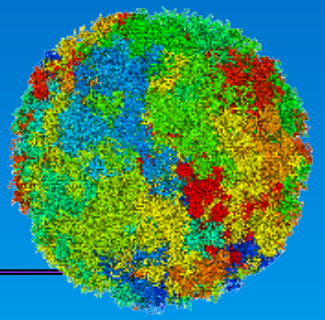
C: MLS in 10 µm nucleus
126 kbp loops, 126 kbp linker,
not totally relaxed

D: RW/GL in 12 µm nucleus
5 Mbp loops
not totally relaxed





In vivo Morphology & Chromatin Distribution

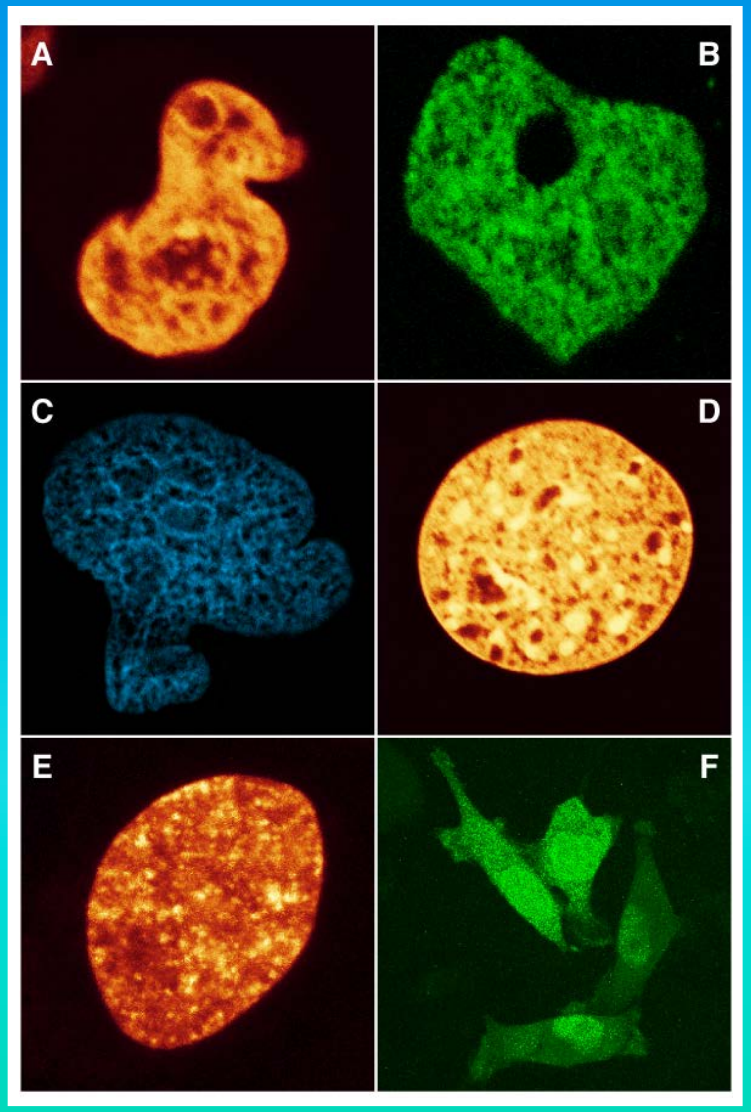


The stable expression of fusions between histones and autofluorescent proteins and the integration into nucleosomes allows the minimal invasive investigation of the structure and dynamics of chromatin.

The clustered morphology in detail favour an MLS like chromatin topology.

HeLa, H2A-YFP

Cos7, H1.0-GFP

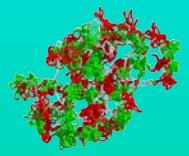


LCLC 103H, H2A-CFP

ID13, H2A-YFP

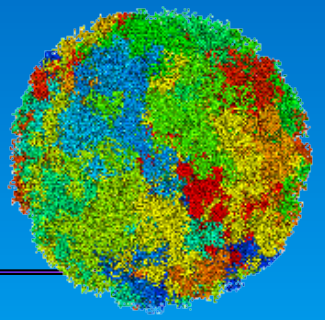
HeLa, mH2A1.2-YFP

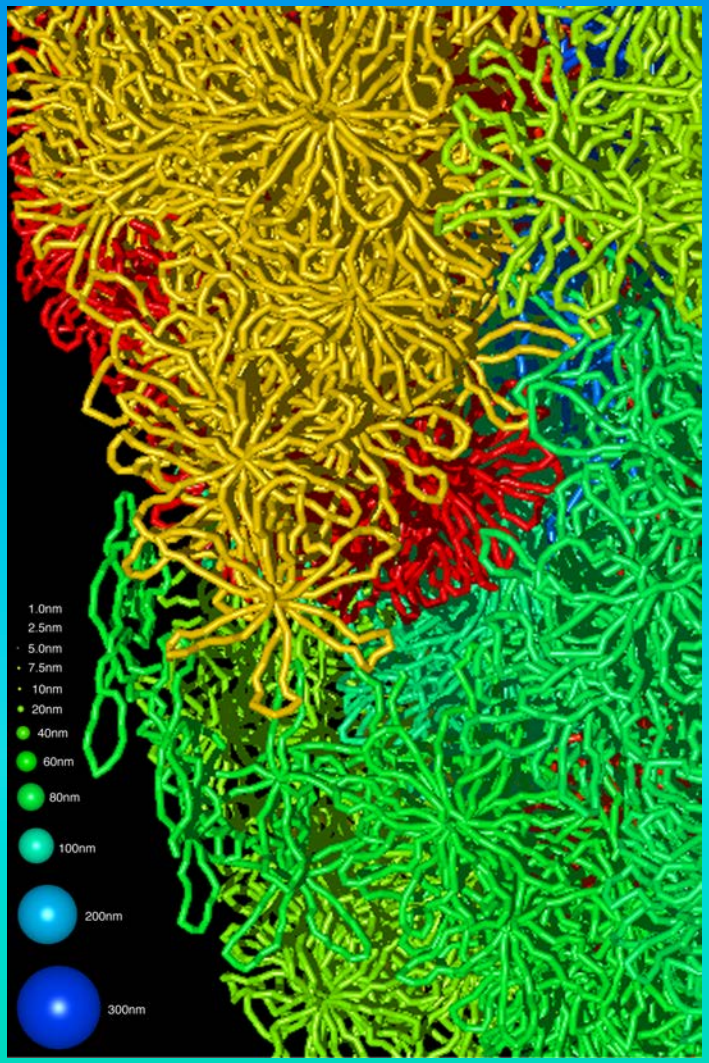
HeLa, H2A-YFP, natural promoter

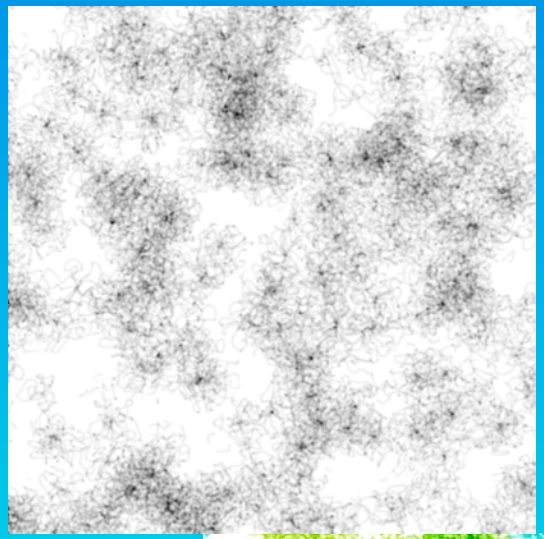


Fine Morphology of Nuclei

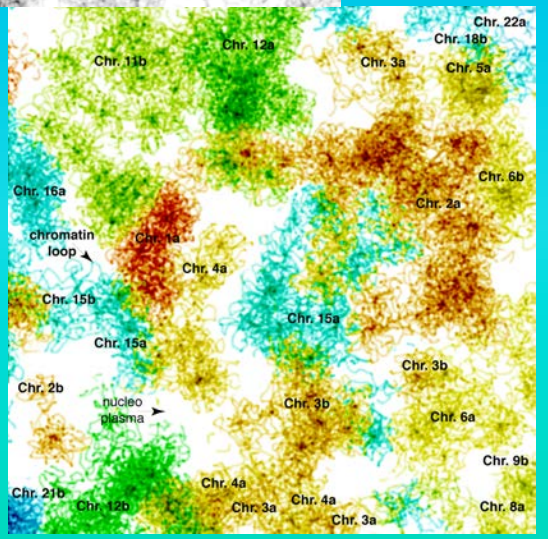
High resolution rendering and simulated electron microscopy including territory painting reveal not only again the model details but also that any location in the nucleus is accessible to biological molecules <15 nm in diameter and that even the Extended Interchromosomal Domain hypothesis is oversimplified.

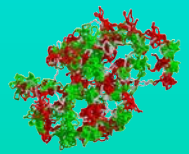






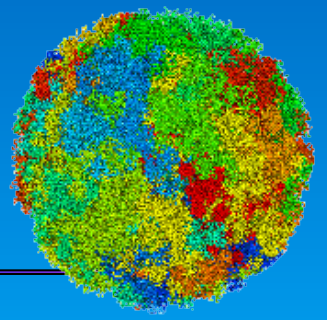
MLS models model with 126 kbp loops and linkers in a 10 µm nucleus.

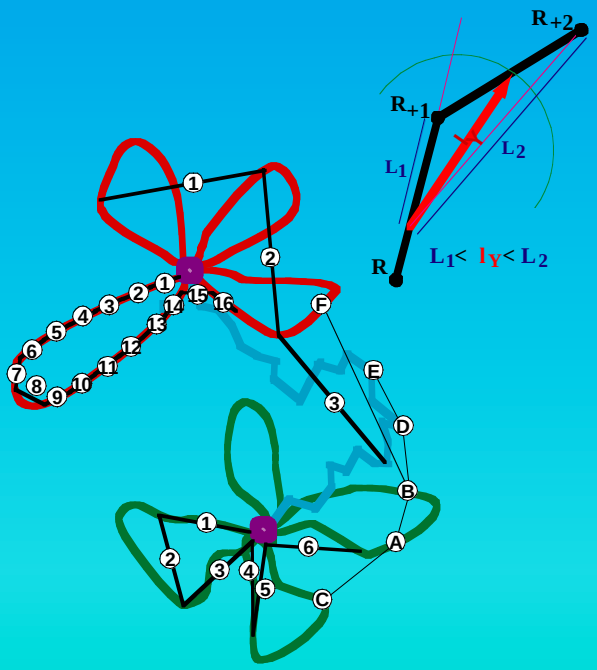


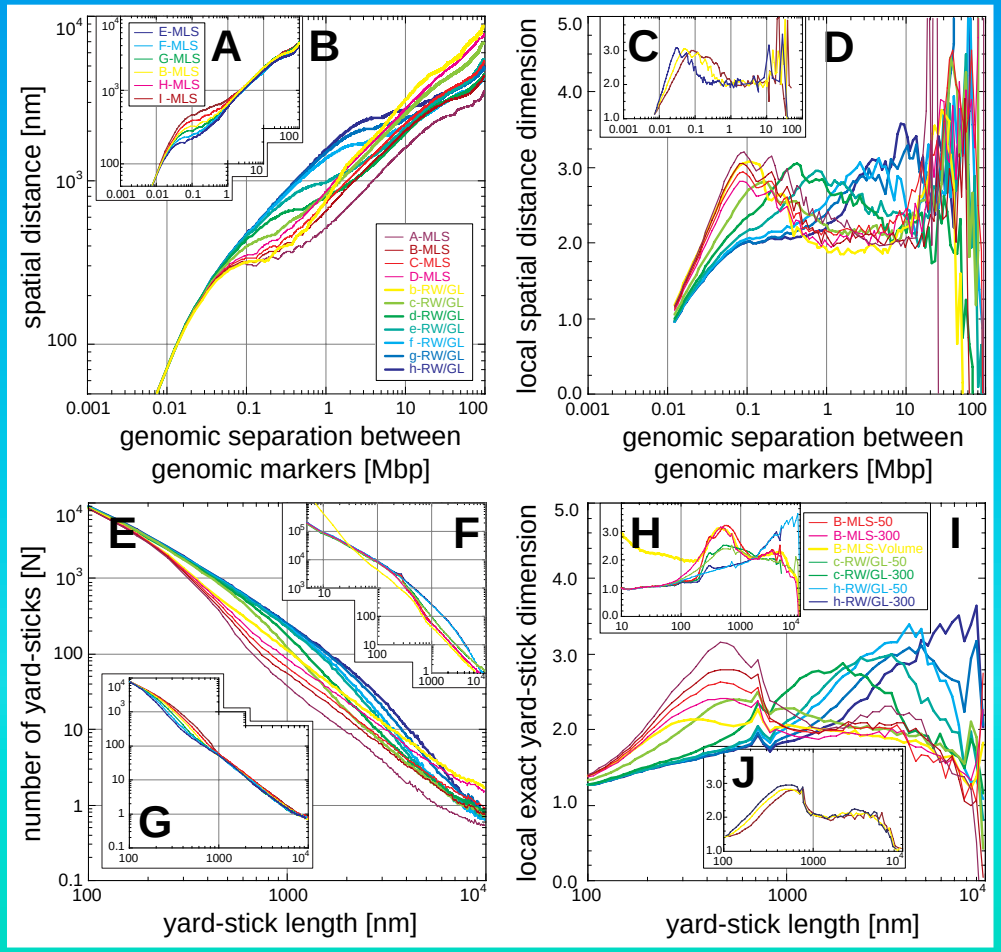


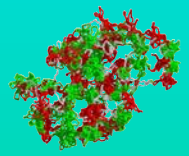
Scaling of the Chromatin Fiber Topology

The spatial-distance and exact yard-stick dimension distinguish between the simulated models in detail.The MLS model shows a globular and fine-structured multi scaling behaviour due to the loops froming rosettes. This agrees with DNA fragmentation by Carbon ion irradiation and the appearance of fine-structured multi-scaling long-range correlations found in the sequential organization of genomes.

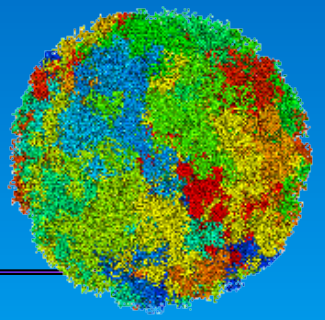






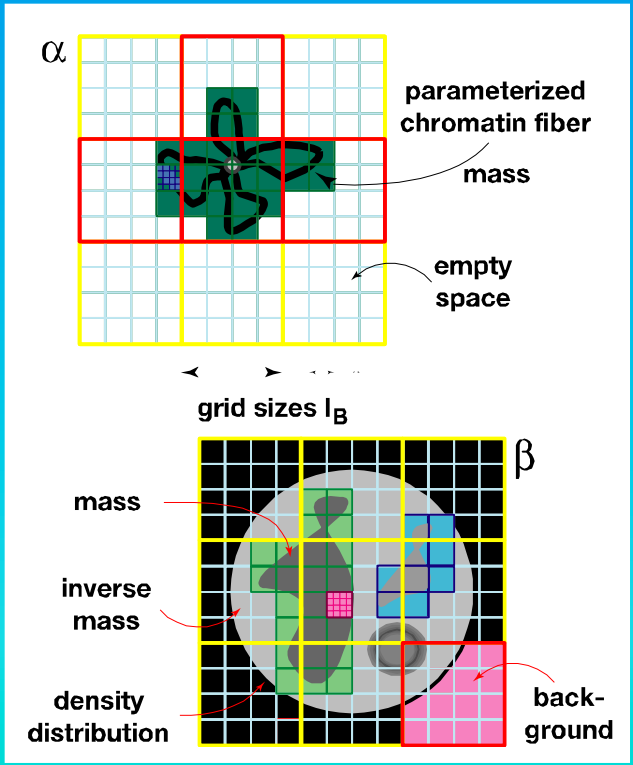


Scaling of the Chromatin Morphology & Distribution

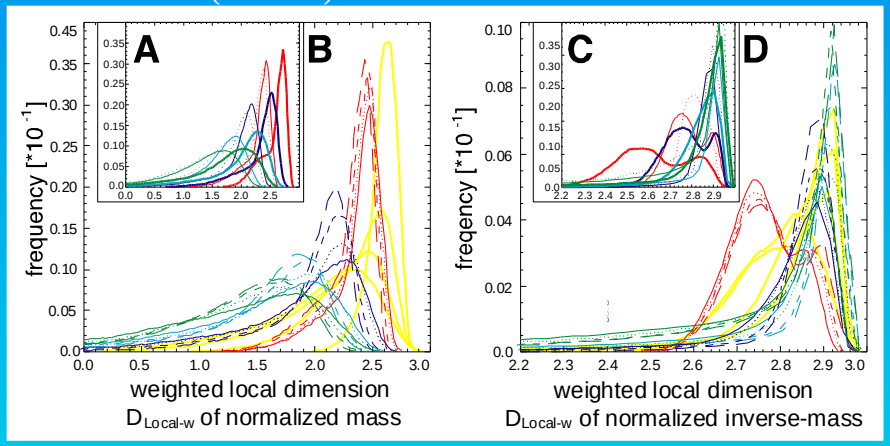


The local (inverse-) mass dimension distribution distinguishs between the models in detail and show also a multi-scaling behaviour with globular feature for the MLS model like the scaling of the fiber topology. With the mass dimension as function of intensity separates very well between different nuclei *in vivo*.

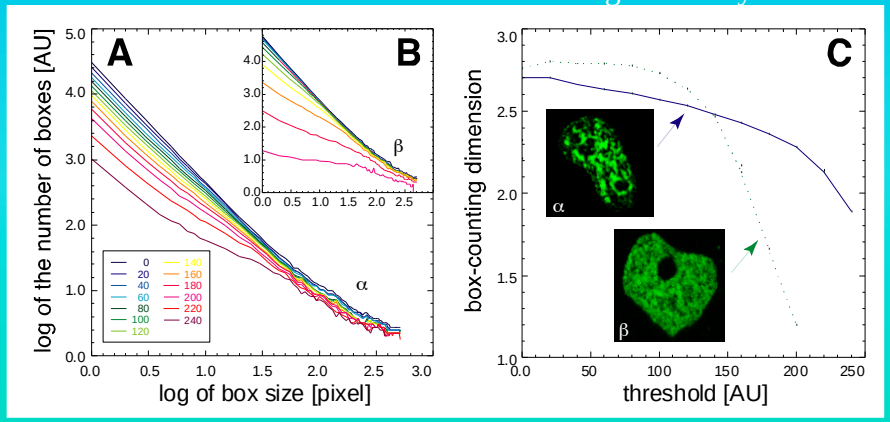
Consequently, the chromatin morphology is causally and quantitatively connected to the fiber topology.

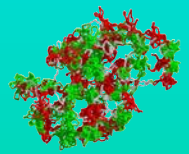


(inverse-) mass dimension distribution

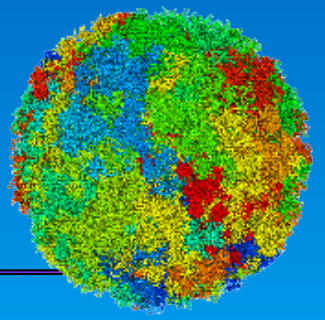


mass dimension as function of image intensity



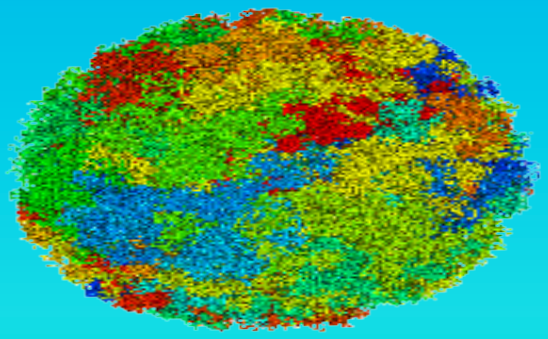


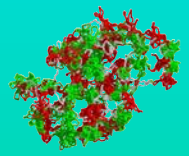
DNA Fragment Distribution after Ione-Irradiation



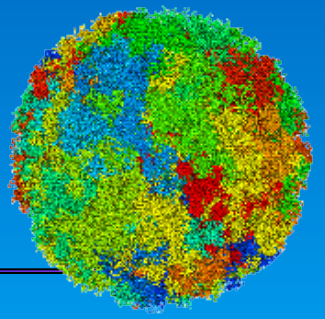
The length distribution of DNA fragments after irradiation with e. g. C or Ca with an inhomogeneous spatial double strand breackage probability depends on the detailed folding topology of the chromatin fiber and the RW/GL and MLS models differ largely.

Experiments always agree best with the MLS model independent of the irradiation conditions.



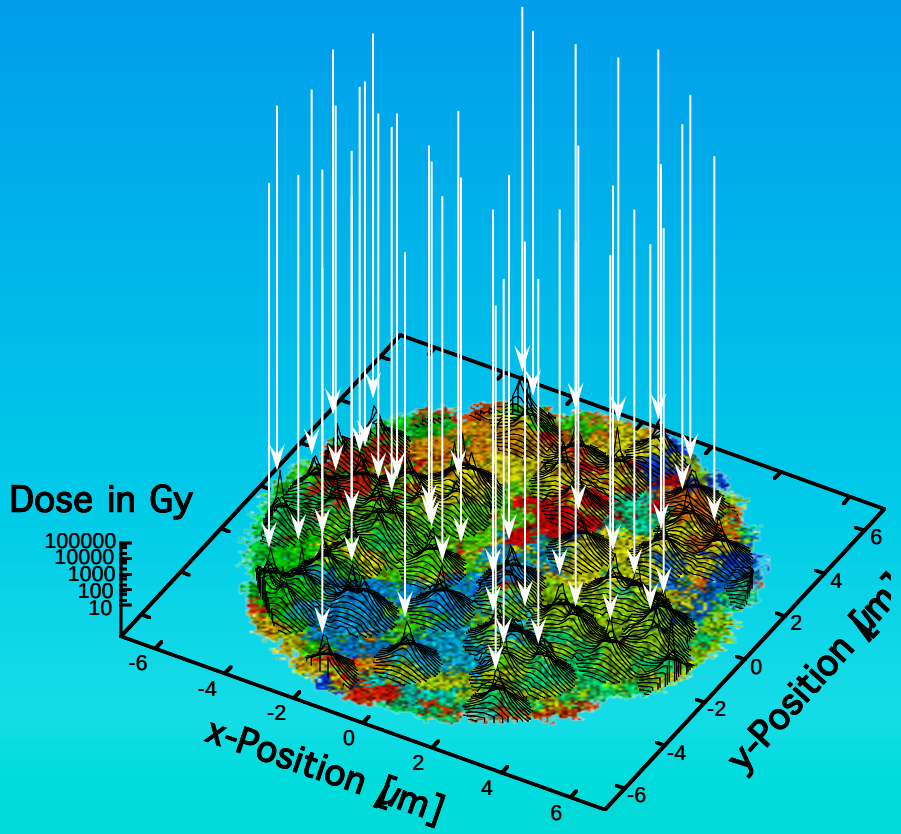


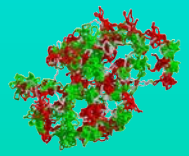
DNA Fragment Distribution after Ione-Irradiation



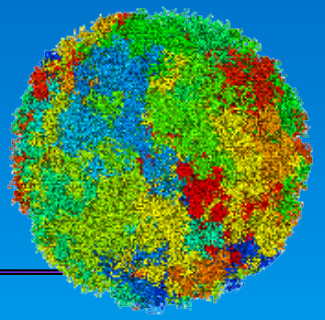
The length distribution of DNA fragments after irradiation with e. g. C or Ca with an inhomogeneous spatial double strand breackage probability depends on the detailed folding topology of the chromatin fiber and the RW/GL and MLS models differ largely.

Experiments always agree best with the MLS model independent of the irradiation conditions.



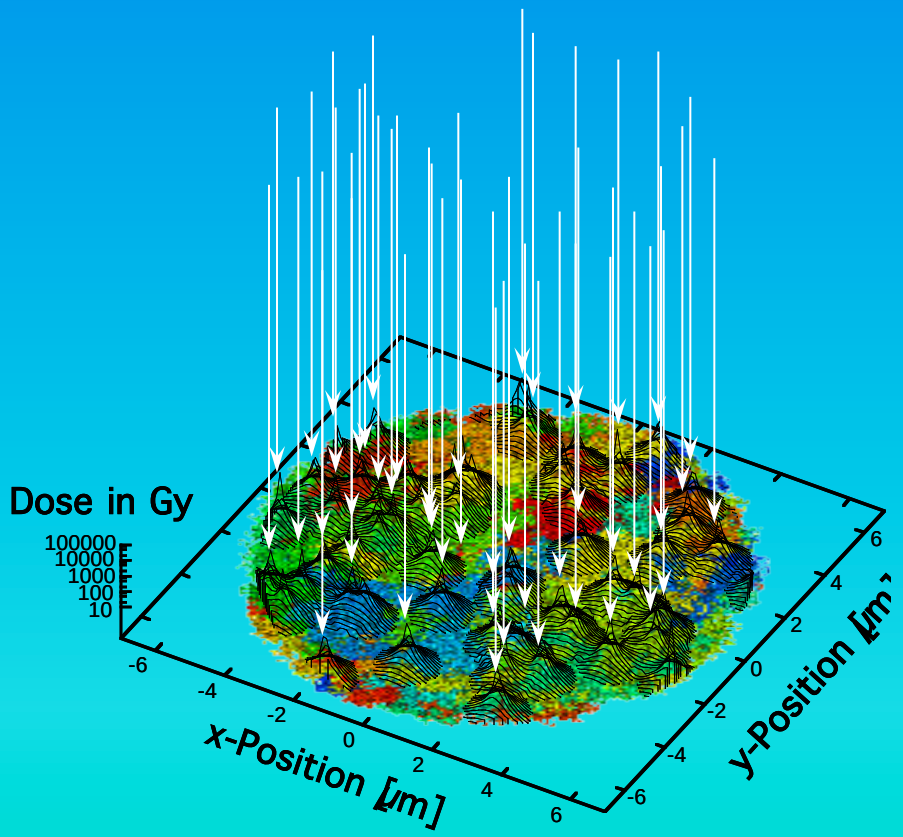


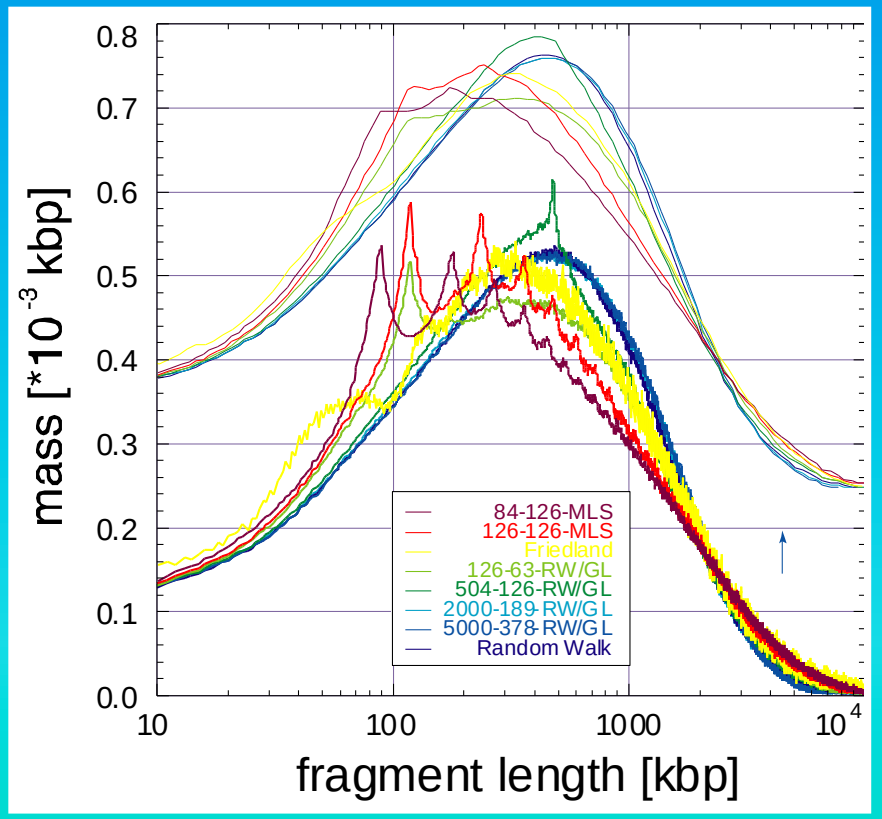
DNA Fragment Distribution after Ione-Irradiation

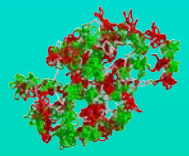


The length distribution of DNA fragments after irradiation with e. g. C or Ca with an inhomogeneous spatial double strand breackage probability depends on the detailed folding topology of the chromatin fiber and the RW/GL and MLS models differ largely.

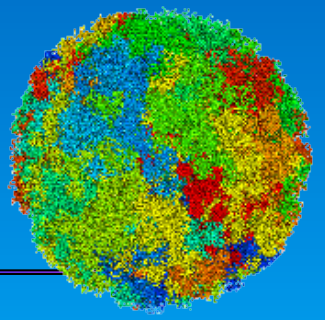
Experiments always agree best with the MLS model independent of the irradiation conditions.







Diffusion of Particles in the Nucleus

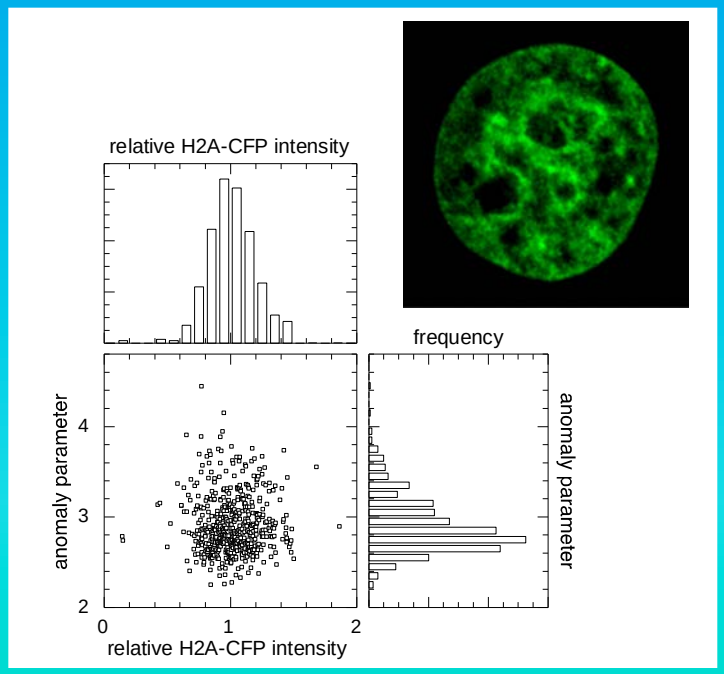


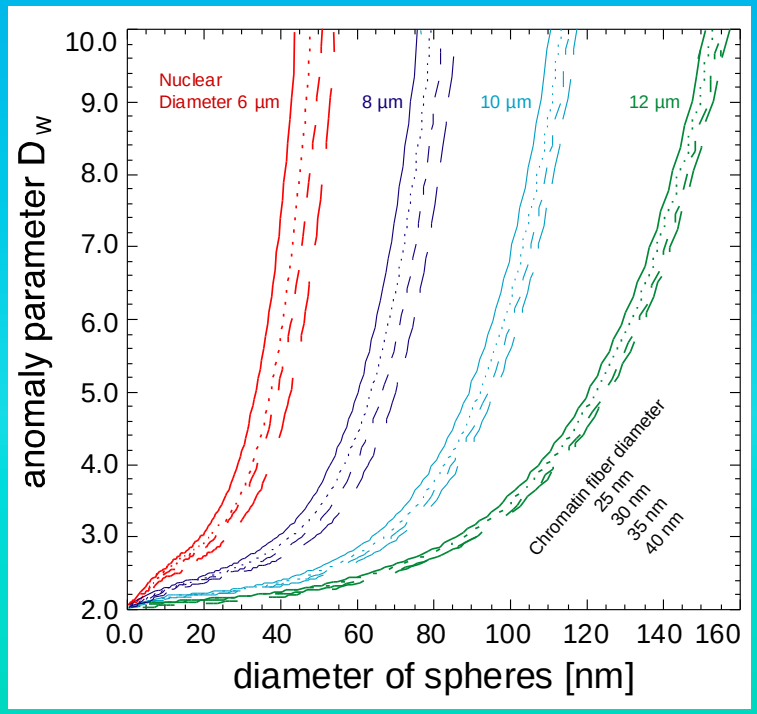
Due to the volume and spatial relation ships in the nucleus typical particles reach almost any location in the nucleus by moderately obstructed diffusion: a 10 nm particle moves 1 to 2 µm within 10 ms.

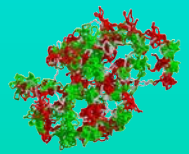
The structural influence on the obstruction degree is random for Alexa 568 as function of the chromatin distribution visualized by H2A CFP in vivo and measured by fluorescence correlation spectroscopy (FCS)

$$\langle r^2 \rangle \propto t^{2/D_w}$$

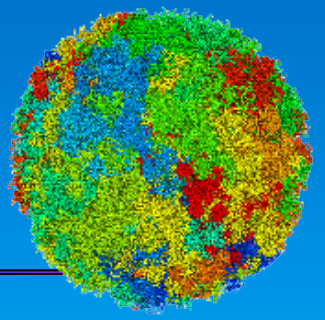
Nuclear diameter [µm]	Nuclear Volume [µm³]	Mean Nucloesome Concentration [µM]	Chromatin Volume Fraction [%]	Mean Isotropic Mesh Spacing [nm]
6	115	251	20.1	41
8	268	107	8.6	64
10	523	55	4.4	90
12	904	32	2.6	117







Sequential Organization of Genomes



Determination of the concentration fluctuation function C(l) and its local slope the correlation coefficient δ(l) reveal multi-scaling long-range correlation up to 10^6 to 10^7 bp in *Homo sapiens* which clearly deviate from random sequences with high significance (decreasing the nearer to the cut-off).

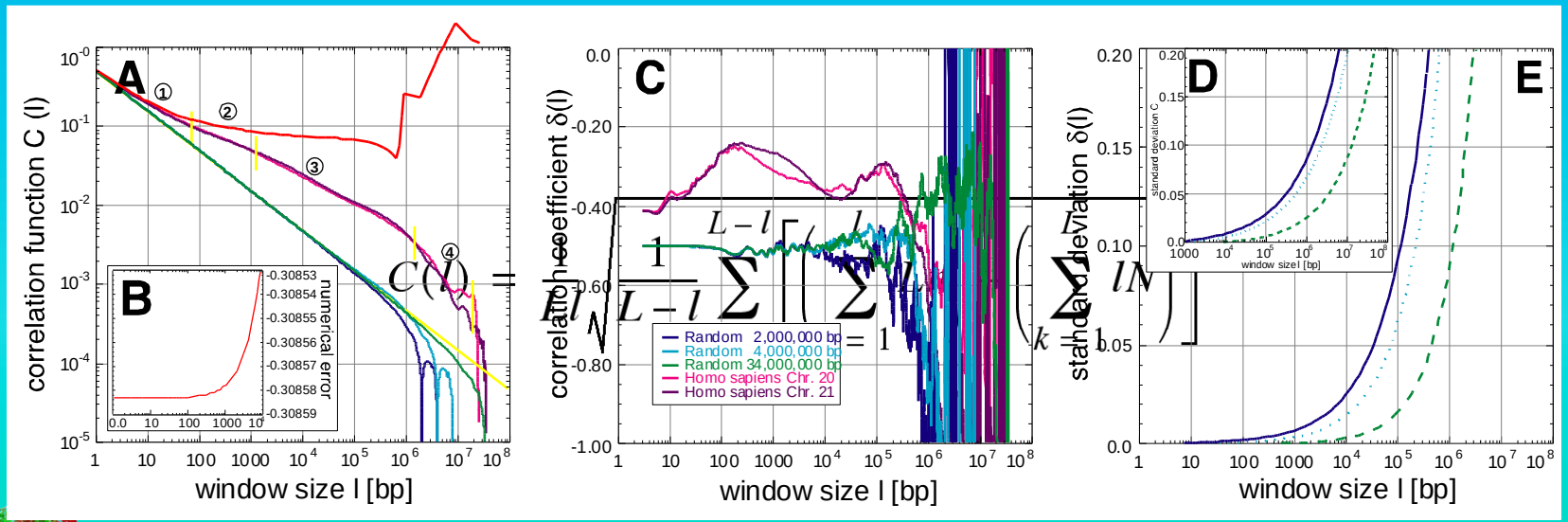
On large scales this might only be due to a strong and definite three-dimensional genome organization.

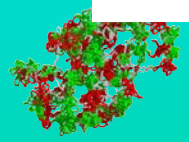
$$C(l) = \sqrt{\left\langle (c_l - \bar{c}_L)^2 \right\rangle_s}$$

numerically unstable

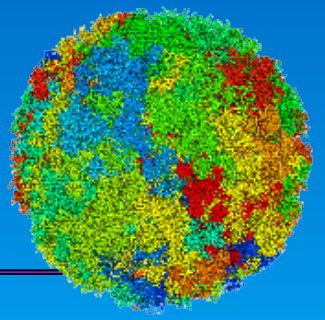
$$C(l) = \sqrt{\frac{1}{L-l+1}\sum_{s=1}^{L-l}\left(\frac{1}{l}\sum_{k=1}^{l} n - \frac{1}{L}\sum_{k=1}^{L} N\right)^2}$$

numerically stable





Fine-Structured Multi-Scaling Long-Range Correlations of *Homo sapiens*

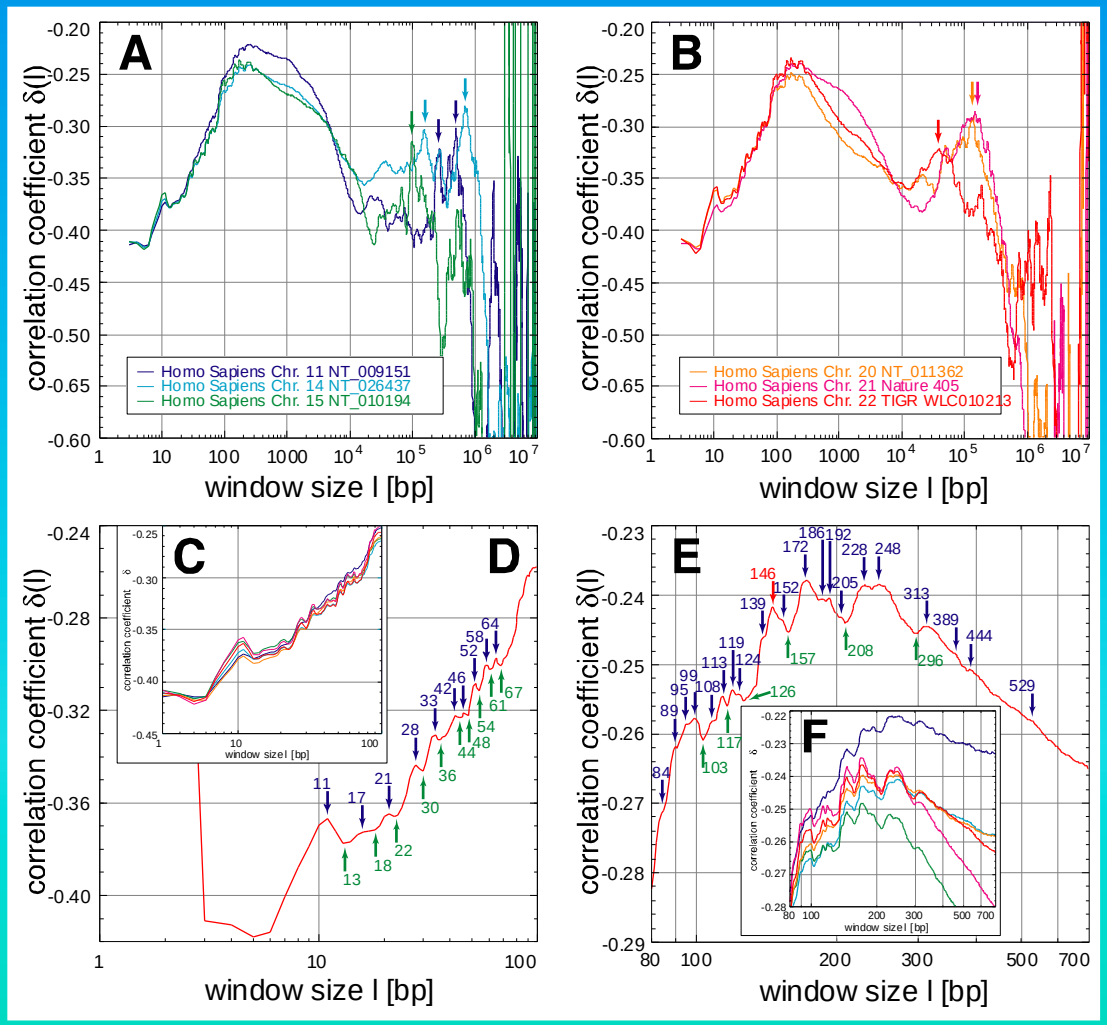


The general behaveour is characterized by first maximum of the correlation coefficient d(l) at ~250 bp and at 1x10^5 to 3x10^5 bp, both due to a globular block structure of genomes. Due to their fine structure the first is attributable to nucleosomal binding and the latter due to aggregation of chromatin loops as in the MLS model.

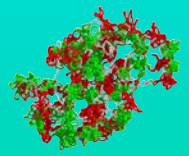
Thus, the sequential organization is closely connected to the three-dimensional organization of genomes.

general behaviour

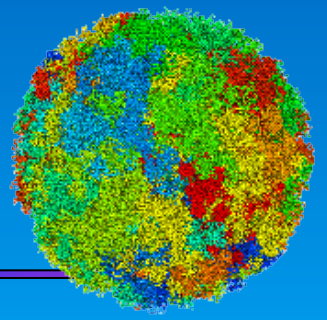
fine structure



the fine structure survives averaging over several human chromosomes.

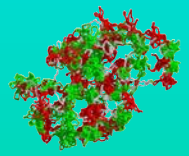


Conclusion

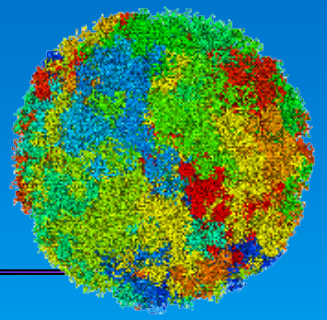


Every structural level of nuclear organization including its dynamics is connected and represented in all the other levels.

- **Only the MLS model leads to chromosome territories with subcompartments agreeing qualitatively and quantitatively with experiments.**
- **Comparison between simulated and experimental spatial distances between genetic markers favours and MLS model with 80 to 120 kbp loops and linkers.**
- **The nuclear morphology or chromatin distribution is tightly connected to the folding topology of the chromatin fiber.**
- **Scaling analysis of the chromatin fiber topology and nuclear morphology reveals a fine-structured multi-scaling behaveour and allows a detailed description model changes.**
- **Most biological particles (molecules, proteins…) could reach almost any location in the nucleus by only moderately obstructed diffusion in agreement with *in vivo* experiments.**
- **The DNA fragment distribution after ion irradiation reflects the chromatin fiber topology not only in detail but also favours always an MLS model.**
- **The sequential organization of genomes is characterized by fine-structured multi-scaling long-range correlations, which are specie specific and tightly connected to the three-dimensional organization of genomes. On large-scales again an MLS model is favoured.**



Acknowledgements



Biophysics of Macromolecules DKFZ

Thomas Weidemann

Gabriele Müller

Waldemar Waldeck

Jörg Langowski

Molecular Biophysics KIP

Katalin Fejes-Toth

Malte Wachsmuth

Karsten Rippe

Biomedical Structure Analysis DKFZ

Felix Bestvater

Eberhard Spiess

The Cremer Labs

Joachim Rauch

Irina Solovei

Michael Hausmann

Christoph Cremer

Thomas Cremer

Molecular Genetics DKFZ

Karsten Richter

Peter Lichter

University Tübingen

Markus Göker,

LMU Munich

Peter Quichen

Anna Friedl

Supercomputin Center Karlsruhe

Rudolph Lohner

Scripps Research Institute

Karin Monier

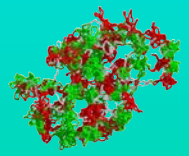
Kevin Sullivan

Others from the DKFZ:
Monika Stöhr, Michael Stöhr, Andreas Hunziker, Angel Alonso

High-Performance Computing Center Stuttgart, University of Stuttgart; Supercomputing Center Karlsruhe, University of Karlsruhe; Computing Center, Deutsches Krebsforschungszentrum Heidelberg (DKFZ)

Bundesministerium für Forschung und Technology (BMFT) 01 KW 9602/2
(3D-Human Genome Study Group Heidelberg, German Human Genome Projekt)

Deutsches Krebsforschungszentrum (DKFZ)



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

Knoch, T. A.

Result and Review Workshop, High-Performance Computing-Center Stuttgart (HLRS), Stuttgart, Germany, 6th - 7th October, 2003.

Abstract

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. To achieve a deeper unterstanding of the human genome the three-dimensional organization of the human cell nucleus, the structural-, scaling- and dynamic properties of interphase chromosomes and cell nuclei were simulated and combined with the analysis of long-range correlations in completely sequenced genomes as well as the chromatin distribution in vivo. Using Monte Carlo and Brownian Dynamics methods, the 30 nm chromatin fiber was folded according to the Multi-Loop-Subcompartment (MLS) model, in which ~100 kbp loops form rosettes, connected by a linker, and the Random-Walk/Giant-Loop (RW/GL) topology, in which 1-5 Mbp loops are attached to a flexible backbone. Both the MLS and the RW/GL model form chromosome territories but only the MLS rosettes result in distinct subcompartments visible with light microscopy and low overlap of chromosomes, -arms and subcompartments. This morphology and the size of subcompartments agree with the morphology found by expression of histone auto-fluorescent protein fusions and fluorescernce *in situ* hybridization (FISH) experiments. Even small changes of the model parameters induced significant rearrangements of the chromatin morphology. Thus, pathological diagnoses based on this morphology, are closely related to structural changes on the chromatin level. The position of interphase chromosomes depends on their metaphase location, and suggests a possible origin of current experimental findings. The chromatin density distribution of simulated confocal (CLSM) images agrees with the MLS model and with recent experiments. The scaling behaviour of the chromatin fiber topology and morphology of CLSM stacks revealed fine-structured multi-scaling behaviour in agreement with the model prediction and correlations in the DNA sequence. 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 to 126 kbp. Simulation of the fragment distribution after ion-irradiation revealed also best agreement with the MLS model by comparison with experimental distributions. Correlation analyses of completely sequenced Archaea, Bacteria and Eukarya chromosomes revealed fine-structured positive long-range correlation due to codon, nucleosomal or block organization of the genomes, allowing classification as well as tree construction. This shows a complex sequential organization of genomes closely connected to their three-dimensional organization. Visual inspection of the morphology reveals also big spaces allowing high accessibility to nearly every spatial location, due to the chromatin occupancy <30% and a mean mesh spacing of 29 to 82 nm for nuclei of 6 to 12 µm diameter. The simulation of diffusion agreed with this structural prediction, since the mean displacement for 10 nm sized particles of ~1 to 2 µm takes place within 10 ms. Therefore, the diffusion of biological relevant tracers is only moderately obstructed, with the degree of obstruction ranging from 2.0 to 4.0 again in experimental agreement. 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.

Corresponding author email contact: TA.Knoch@taknoch.org

Keywords:

Genome, genomics, genome organization, genome architecture, structural sequencing, architectural sequencing, systems genomics, coevolution, holistic genetics, genome mechanics, genome function, genetics, gene regulation, replication, transcription, repair, homologous recombination, simultaneous co-transfection, cell division, mitosis, metaphase, interphase, cell nucleus, nuclear structure, nuclear organization, chromatin density distribution, nuclear morphology, chromosome territories, subchromosomal domains, chromatin loop aggregates, chromatin rosettes, chromatin loops, chromatin fibre, chromatin density, persistence length, spatial distance measurement, histones, H1.0, H2A, H2B, H3, H4, mH2A1.2, DNA sequence, complete sequenced genomes, molecular transport, obstructed diffusion, anomalous diffusion, percolation, long-range correlations, fractal analysis, scaling analysis, exact yard-stick dimension, box-counting dimension, lacunarity dimension, local nuclear dimension, nuclear diffuseness, parallel super computing, grid computing, volunteer computing, Brownian Dynamics, Monte Carlo, fluorescence in situ hybridization, confocal laser scanning microscopy, fluorescence correlation spectroscopy, super resolution microscopy, spatial precision distance microscopy, autofluorescent proteins, CFP, GFP, YFP, DsRed, fusionprotein, in vivo labelling.

Literature References

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, Ruperto-Carola University, Heidelberg, Germany, 1998, and TAK Press, Tobias A. Knoch, Mannheim, Germany, ISBN 3-00-010685-5 and ISBN 978-3-00-010685-9 (soft cover, 2rd ed.), ISBN 3-00-035857-9 and ISBN 978-3-00-0358857-0 (hard cover, 2rd ed.), ISBN 3-00-035858-7, and ISBN 978-3-00-035858-6 (DVD, 2rd ed.), 1998.

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., Münkel, C. & Langowski, J. Three-Dimensional Organization of Chromosome Territories and the Human Interphase Nucleus. *High Performance Scientific Supercomputing*, editor Wilfried Juling, 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*, editors Krause, E. & Jäger, W., High-Performance Computing Center (HLRS) Stuttgart, University of Stuttgart, Springer Berlin-Heidelberg-New York, ISBN 3-540-66504-8, 229-238, 2000.

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

Knoch, T. A. (editor), Backes, M., Baumgärtner, V., Eysel, G., Fehrenbach, H., Göker, M., Hampl, J., Hampl, U., Hartmann, D., Hitzelberger, H., Nambena, J., Rehberg, U., Schmidt, S., Weber, A., & Weidemann, T. Humanökologische Perspectiven Wechsel - Festschrift zu Ehren des 70. Geburtstags von Prof. Dr. Kurt Egger. Human Ecology Working Group, Ruperto-Carola University of Heidelberg, Heidelberg, Germany, 2002.

Knoch, T. A. 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. *Dissertation*, Ruperto-Carola University, Heidelberg, Germany, and TAK†Press, Tobias A. Knoch, Mannheim, Germany, ISBN 3-00-009959-X and ISBN 978-3-00-009959-5 (soft cover, 3rd ed.), ISBN 3-00-009960-3 and ISBN 978-3-00-009960-1 (hard cover, 3rd ed.), ISBN 3-00-035856-9 and ISBN 978-3-00-010685-9 (DVD, 3rd ed.) 2002.

Knoch, T. A. Towards a holistic understanding of the human genome by determination and integration of its sequential and three-dimensional organization. *High Performance Computing in Science and Engineering 2003*, editors Krause, E., Jäger, W. & Resch, M., High-Performance Computing Center (HLRS) Stuttgart, University of Stuttgart, Springer Berlin-Heidelberg-New York, ISBN 3- 540-40850-9, 421-440, 2003.

Wachsmuth, M., Weidemann, T., Müller, G., Urs W. Hoffmann-Rohrer, **Knoch, T. A.**, Waldeck, W. & Langowski, J. Analyzing intracellular binding and diffusion with continuous fluorescence photobleaching. *Biophys. J.* 84(5), 3353-3363, 2003.

Weidemann, T., Wachsmuth, M., **Knoch, T. A.**, Müller, G., Waldeck, W. & Langowski, J. Counting nucleosomes in living cells with a combination of fluorescence correlation spectroscopy and confocal imaging. *J. Mol. Biol.* 334(2), 229-240, 2003.

Fejes Tóth, K., **Knoch**, **T. A.**, Wachsmuth, M., Frank-Stöhr, M., Stöhr, M., Bacher, C. P., Müller, G. & Rippe, K. Trichostatin A induced histone acetylation causes decondensation of interphase chromatin. *J. Cell Science* 177, 4277-4287, 2004.

Ermler, S., Krunic, D., **Knoch, T. A.**, Moshir, S., Mai, S., Greulich-Bode, K. M. & Boukamp, P. Cell cycle-dependent 3D distribution of telomeres and telomere repeat-binding factor 2 (TRF2) in HaCaT and HaCaT-myc cells. *Europ. J. Cell Biol.* 83(11-12), 681-690, 2004.