From Sequence to Morphology

Towards a Holistic Understanding of Genomes

History and Perspectives

by

Tobias A. Knoch

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

Erasmus MC, Rotterdam, The Netherlands
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 fulfill their obvious functions, e.g., gene regulation or replication, since they are optimized by evolution?
WE KNOW THAT WE HARDLY KNOW ANYTHING

From the sequence to the morphology we just have begun to elucidate the organization of genomes and we also just have begun to understand that it is not only the detailed knowledge about one organizational level but beyond the holistic entity of the whole cell nucleus or genome which makes genomes function!

**Nucleus:**
- Nuclear organization: chromosome arrangement, morphology?
- Nuclear Code: information content, regulation, variability?

**Chromosome:**
- Chromosome organization: loops, loop aggregates, extension?
- Chromosome code: information content, regulation, variability?

**Chromatin:**
- Chromatin fiber organization: prevalence, variation, dynamics?
- Chromatin code: coding, regulation, modification?

**Nucleosome:**
- Nucleosome organization: tail position, mobility, modification?
- Histon code: coding, regulation, modification?

**DNA Sequence:**
- DNA local structure: bending, melting, stability, modification?
- General sequence organization: coding, regulating and the rest?
Chromatin Conformation and Higher-Order Topologies

It becomes increasingly clearer, that the chromatin conformation is a random organization of nucleosomes, which depending on external or modification conditions has different condensation degrees, with a prevalence for the 30nm fiber with ~6 nucleosomes per 11nm. This seems to make loops which further cluster to form aggregates more or less rosette-like which then constitute the chromosome.

A-C: Voeit & Voit; D: Reznik et al.

Courtesy P. Fransz, Amsterdam

Integral Models of Cell Nuclear Organization

Already Rabl and Boveri were aware of the obvious fact that the organization of genomes has to be consistent from the sequence level to the morphology of the whole cell nucleus. Although they might be different in detail their common seem is recursive folding and clustering thereof with variation/ modification and dynamics accounting for different nuclear states and function.
Integral Models of Cell Nuclear Organization

The biggest advantage of integral models is the again obvious and simple fact, that they allow the validation from the consistency of different levels of organization from the other levels. Thus, e.g. the so called Interchromosomal Domain Model can be ruled out by simple voluminous thought…

Random-Walk/Giant-Loop Multi-Loop-Subcompartment Model

RW/GL-Model

MLS-Model

backbone (non DNA)

attachment points

loop size: 3Mbp-5Mbp
loop length: 30 to 50 µm

Linker between rosettes consists of DNA (126 kbp)
(in contrast to backbone)

chromatin fiber: 80nm thick

R bands

G bands

rosette size: 1-2Mbp
diameter: 400 to 800 nm
(according to interphase ideogram bands)
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 chromosome territory, thus forming no distinct features like in MLS model.

MLS model, loop size 126 kbp, 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.

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 chromosome territory, thus forming no distinct features like in MLS model.

MLS model, loop size 126 kbp, 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.

- **Fibroblasts 15q11-13** PLW/A-Region, FISH, our data.
- **Lymphocytes 11q13** FISH, K. Monier, Institut Albert Bonniot, Grenoble.
- **Fibroblasts 4p16.3** FISH, Yokota et. al., 1995.

**RW/GL-model:**
- Loop size: 5 Mbp, 3 Mbp, 1.0 Mbp, 0.5 Mbp

**MLS-model (loop size 126 kbp):**
- Linker length: 251kbp, 188kbp, 126kbp, 62kbp

**Spatial distance between genomic markers [Mbp]**

**Frequency distribution:**
- Genomic distance to a loop base of a MLS subcompartment [Mbp]

**Genomic distance to a loop base [Mbp]**
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.
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.

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.
Fine Morphology of Nuclei

High resolution rendering and simulated electron microscopy including territory painting reveal not only 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 forming 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 distinguishes between the models in detail and shows also a multi-scaling behavior 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.
Diffusion of Particles in the Nucleus

Due to the volume and spatial relationships 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 a function of the chromatin distribution visualized by H2A CFP in vivo and measured by fluorescence correlation spectroscopy (FCS).

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

<table>
<thead>
<tr>
<th>Nuclear diameter [µm]</th>
<th>Nuclear Volume [µm³]</th>
<th>Mean Nucleosome Concentration [µM]</th>
<th>Chromatin Volume Fraction [%]</th>
<th>Mean Isotropic Mesh Spacing [nm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>115</td>
<td>251</td>
<td>20.1</td>
<td>41</td>
</tr>
<tr>
<td>8</td>
<td>268</td>
<td>107</td>
<td>8.6</td>
<td>64</td>
</tr>
<tr>
<td>10</td>
<td>523</td>
<td>55</td>
<td>4.4</td>
<td>90</td>
</tr>
<tr>
<td>12</td>
<td>904</td>
<td>32</td>
<td>2.6</td>
<td>117</td>
</tr>
</tbody>
</table>

Nuclear Diameter 6 µm 8 µm 10 µm 12 µm

Chromatin fiber diameter 25 nm 30 nm 35 nm 40 nm
Sequential Organization of Genomes

Determination of the concentration fluctuation function $C(l)$ and its local slope the correlation coefficient $\delta(l)$ 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.
Sequential Organization of Genomes

Determination of the concentration fluctuation function \( C(l) \) and its local slope the correlation coefficient \( \delta(l) \) 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.

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

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

<p>|</p>
<table>
<thead>
<tr>
<th>Numerically unstable</th>
</tr>
</thead>
<tbody>
<tr>
<td>stable</td>
</tr>
</tbody>
</table>
Sequential Organization of Genomes

Determination of the concentration fluctuation function $C(l)$ and its local slope the correlation coefficient $\delta(l)$ 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.
Fine-Structured Multi-Scaling Long-Range Correlations of *Homo sapiens*

The general behaviour is characterized by first maximum of the correlation coefficient $d(l)$ at ~250 bp and at $1 \times 10^5$ to $3 \times 10^5$ 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.

---

The figures illustrate the correlation coefficient $d(l)$ for different human chromosomes, showing the general behavior and the 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 behaviour 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 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

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

The Cremer Labs
Joachim Rauch
Irina Solovei
Michael Hausmann
Christoph Cremer
Thomas Cremer

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

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

University Tübingen
Markus Göker

Molecular Genetics
DKFZ
Karsten Richter
Peter Lichter

LMU Munich
Peter Quichen
Anna Friedl

Scripps Research Institute
Karim Monier
Kevin Sullivan

Clinical Genetics
Erasmus MC
Bert Eussen

Cell Biology
Erasmus MC
Frank Grossveld
Wouter de Laat

Bioinformatics
Erasmus MC
Michael Moorhouse

Jesse Tjang
Wouter de Breejn
Michael Lesnussa
Marten Muettegert

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), Universität Heidelberg, Erasmus MC
Acknowledgements

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

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

Biophysics of Macromolecules

DKFZ
- Thomas Weidemann
- Gabriele Müller
- Waldemar Waldeck
- Jörg Langowski

The Cremer Labs
- Joachim Rauch
- Irina Solovei
- Michael Hausmann
- Christoph Cremer
- Thomas Cremer

Biomedical Structure Analysis

DKFZ
- Felix Bestvater
- Eberhard Spiess

Molecular Biophysics

KIP
- Katalin Fejes-Toth
- Malte Wachsmuth
- Karsten Rippe

University Tübingen
- Markus Göker

Supercomputing Center Karlsruhe
- Rudolph Lohner

Molecular Genetics

DKFZ
- Karsten Richter
- Peter Lichter

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

Deutsches Krebsforschungszentrum (DKFZ), Universität Heidelberg, Erasmus MC

Clinical Genetics

Erasmus MC
- Bert Eussen

Bioinformatics

Erasmus MC
- Michael Moorhouse

Cell Biology

Erasmus MC
- Frank Grossveld
- Wouter de Laat

Jesse Tjang
- Wouter de Breejn
- Michael Lesnussa
- Marten Muetgeert

Scripps Research Institute
- Karin Monier
- Kevin Sullivan

LMU Munich
- Peter Quichen
- Anna Friedl

THE REAL FUN IS YET TO COME!
From Sequence to Morphology

Towards a Holistic Understanding of Genomes

History and Perspectives

Knoch, T. A.

Riga Meeting on Comprehensive Cell Biology, Riga, Latvia, 2nd - 4th June, 2005.

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. Various models were put forward through history. To achieve a deeper understanding of the human genome the three-dimensional organization of the human cell nucleus, here, the structural-, scaling- and dynamic properties of interphase chromosomes are evaluated on all scales from a single base pair to the nuclear morphology level in respect to achieving a consistent holistic framework of genome organization.

Thus, by using Monte Carlo and Brownian Dynamics methods, the 30 nm chromatin fiber was simulated 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 one finds that 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. The MLS morphology, the size of subcompartments and chromatin density distribution of simulated confocal (CLSM) images agree with the expression of fusionproteins from the histones H1, H2A, H2B, H3, H4 and mH2A1.2 with the autofluorescent proteins CFP, GFP, YFP, DsRed-1 and DsRed-2 which also revealed different interphase morphologies for different cell lines. Even small changes of the model parameters induced significant rearrangements of the chromatin morphology. Thus, pathological diagnoses, 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 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. Simulated and experimental DNA fragment distribution after ion-irradiation revealed also best agreement with such an MLS. 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 between the chromatin fiber allowing high accessibility to nearly every spatial location, due to the
chromatin occupancy <30% and a mean mesh spacing of 29 - 82 nm for nuclei of 6 - 12 µm diameter. This agrees with a simulated displacement of 10 nm sized particles of ~1 - 2 µm takes place within 10 ms, i.e. a moderately obstructed diffusion of biological molecules in experimental agreement.

These results show that the local, global and dynamic characteristics of cell nuclei are not only tightly interconnected, but also are integrated holistically to fulfill the overall function of the genome. Future experiments and simulations benefit from taking this ever further completed consistent framework into account in respect to project planning as well as evaluation and hypothesis creation. Consequently, such a detailed understanding and knowledge of genomes will ultimately lead to full functional virtual genomes in silico allowing predictions for diagnostics, disease treatment and the modification of genomes.

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, auto-fluorescent proteins, CFP, GFP, YFP, DsRed, fusion protein, in vivo labelling, information browser, visual data base access, holistic viewing system, integrative data management, extreme visualization, three-dimensional virtual environment, virtual paper tool.

Literature References


