The Three-Dimensional Architecture and Dynamics of the Immunoglobin Heavy-Chain and other Loci - its Functional Implications for Genome Organization

Tobias A. Knoch

With

Cornelis C. Murre & Frank G. Grosveld

Biophysical Genomics & Erasmus Computing Grid

Dept. Cell Biology & Genetics, Erasmus Medical Center, Rotterdam, The Netherlands.

BioQuant Center & German Cancer Research Center (DKFZ), Heidelberg, Germany.
The Three-Dimensional Architecture and Dynamids of the Immunoglobin Heavy-Chain and other Loci - its Functional Implications for Genome Organization

Tobias A. Knoch
With
Cornelis C. Murre & Frank G. Grosveld

Biophysical Genomics & Erasmus Computing Grid
Dept. Cell Biology & Genetics, Erasmus Medical Center, Rotterdam, The Netherlands.
BioQuant Center & German Cancer Research Center (DKFZ), Heidelberg, Germany.
The Three-Dimensional Architecture and Dynamics of the Immunoglobulin Heavy-Chain and other Loci - its Functional Implications for Genome Organization

Towards a Holistic Understanding of Genomes!
Dynamic and Hierarchical Genome Organization

The different organization levels of genomes bridge several orders of magnitude concerning space and time. How all of these organization levels connect to processes like gene regulation, replication, embryogenesis, or cancer development is still unclear.
Dynamic and Hierarchical Genome Organization

The different organization levels of genomes bridge several orders of magnitude concerning space and time. How all of these organization levels connect to processes like gene regulation, replication, embryogeneses, or cancer development is still unclear.

Only a Holistic Genome View will increase the knowledge of these Complex Processes.
The process of cytogenetic analysis requires proper patient and sample analysis as well as a comprehensive evaluation of the results.
The Complexity of Cytogenetic Diagnostics

The process of cytogenetic analysis requires proper patient and sample analysis as well as a comprehensive evaluation of the results.

Better Cytogenetic Diagnostics is based on Advanced Genome Understanding!
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.
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

A: courtesy K. Richter; B: courtesy K. Greulich-Bode

© 1996-2011 All Rights Reserved
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 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.

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.

RW/GL model, loop size 126 kbp, after ~80,000 MC and 1000 relaxing BD steps. Here rosettes form subcompartments as separated organizational and dynamic entities.

MLS model, loop size 126 kbp, after ~50,000 MC and 1000 relaxing BD steps. Large loops intermingle freely thus forming no distinct features like in MLS model.
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.
3D Architecture of the Prader-Willi Region

Fluorescence *in situ* hybridization with various protocols of small probes within the Prader-Willi region combined with spectral precision distance confocal laser scanning microscopy and comparison with large-scale computer simulations shows a Multi-Loop Subcompartiment organization of the Prader-Willi region.
Fluorescence in situ hybridization with various protocols of small probes within the Prader-Willi region combined with spectral precision distance confocal laser scanning microscopy and comparison with large-scale computer simulations shows a Multi-Loop Subcompartment organization of the Prader-Willi region.
3D Architecture of the Prader-Willi Region

Fluorescence *in situ* hybridization with various protocols of small probes within the Prader-Willi region combined with spectral precision distance confocal laser scanning microscopy and comparison with large-scale computer simulations shows a Multi-Loop Subcompartment organization of the Prader-Willi region.
3D Architecture of the Prader-Willi Region

Fluorescence in situ hybridization with various protocols of small probes within the Prader-Willi region combined with spectral precision distance confocal laser scanning microscopy and comparison with large-scale computer simulations shows a Multi-Loop Subcompartment organization of the Prader-Willi region.
3D Architecture of the Prader-Willi Region

Fluorescence *in situ* hybridization with various protocols of small probes within the Prader-Willi region combined with spectral precision distance confocal laser scanning microscopy and comparison with large-scale computer simulations shows a Multi-Loop Subcompartiment organization of the Prader-Willi region.

**Figure A**
- **A1**: Graph showing mean spatial distance (µm) vs. genomic distance [Mbp].
- **A2**: Graph showing cumulative frequency vs. spatial distance (50nm bins) [µm].
- **A3**: Graph showing frequency vs. genomic distance between markers [Mbp].

**Figure B**
- **B1**: Average PSF:
  - FWHM: axial: 970nm, lateral: 280nm

**Figure C**
- **C1**: Graph showing mean compaction [×10] vs. genomic distance [Mbp].

**Figure D**
- **D1**: Graph showing mean compaction [×10] vs. genomic distance [Mbp].

**Figure E**
- **E1**: Graph showing mean compaction [×10] vs. genomic distance [Mbp].

**Figure F**
- **F1**: Graph showing mean compaction [×10] vs. genomic distance [Mbp].

© 1996-2011 All Rights Reserved
3D Architecture & Function of the IgH Locus

Fluorescence *in situ* hybridization of the IgH locus combined with spectral precision distance epifluorescence microscopy, analytical trilateration and comparison with computer simulations shows again a Multi-Loop Subcompartment organization of the IgH locus with functional relevant distances.
Fluorescence *in situ* hybridization of the IgH locus combined with spectral precision distance epifluorescence microscopy, analytical trilateration and comparison with computer simulations shows again a Multi-Loop Subcompartiment organization of the IgH locus with functional relevant distances.
3D Architecture & Function of the IgH Locus

Fluorescence in situ hybridization of the IgH locus combined with spectral precision distance epifluorescence microscopy, analytical trilateration and comparison with computer simulations shows again a Multi-Loop Subcompartiment organization of the IgH locus with functional relevant distances.
3D Architecture & Function of the IgH Locus

Fluorescence in situ hybridization of the IgH locus combined with spectral precision distance epifluorescence microscopy, analytical trilateration and comparison with computer simulations shows again a Multi-Loop Subcompartiment organization of the IgH locus with functional relevant distances.
“Synoptic” 3D Architecture of Various Loci

A history “synoptic” comparison of the spatial distance mapping from their original background and aim, FISH methodological protocols, via microscopic imaging and restoration analysis procedures, to their interpretation, reveals that with time Multi-Loop Subcompartment models are favoured.

<table>
<thead>
<tr>
<th>Study</th>
<th>Location</th>
<th>Preparation of Cells</th>
<th>FISH</th>
<th>Microscopy</th>
<th>Image acquisition</th>
<th>Fit to model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. 3B, Trask '89</td>
<td>DHFR</td>
<td>UA41 G1-cf 75 dropped MAA 3:1 FM 50 %</td>
<td>F</td>
<td>Biotin 1 20-37 photo, wall</td>
<td>RWGL 0.08 J RWGL 1.0</td>
<td></td>
</tr>
<tr>
<td>Fig. 3B, Lawrence '90</td>
<td>Dystrophin</td>
<td>W138F G1 75 dropped MAA 3:1 FM 50 %</td>
<td>F</td>
<td>Biotin 1 20-60 photo, wall</td>
<td>RWGL 0.5-1</td>
<td></td>
</tr>
<tr>
<td>Fig. 3A, Trask '91</td>
<td>Xq28</td>
<td>F G1-cf 75 dropped MAA 3:1 FM 50 %</td>
<td>F</td>
<td>Biotin 1 30-60 photo, wall</td>
<td>RWGL 0.7 J RWGL 2.0- &gt;50</td>
<td></td>
</tr>
<tr>
<td>Fig. 3B, Trask '91</td>
<td>Xq28</td>
<td>F G1-cf 75 dropped MAA 3:1 FM 50 %</td>
<td>F</td>
<td>Biotin 2 30-60 photo, wall</td>
<td>RWGL 1.0-3.0</td>
<td></td>
</tr>
<tr>
<td>Fig. 3, v.d. Engl '92 or Fig. 5A, Trask '93</td>
<td>4p16.3</td>
<td>F G1-cf 75 dropped MAA 3:1 FM 50 %</td>
<td>F</td>
<td>Biotin 2 ? photo, d-board</td>
<td>Lq &lt; 0.1 for GS &lt; 0.5 &lt; RWGL &lt; 5.0</td>
<td></td>
</tr>
<tr>
<td>Fig. 5B, Trask '93</td>
<td>6p21</td>
<td>F G1-cf 75 MAA 3:1 FM 50-70 %</td>
<td>Biotin 2 ? photo, d-board</td>
<td>RWGL &lt; 7.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fig. 5, Sanger '93</td>
<td>MHC 6p21.31</td>
<td>HFF G1-cf ? ?</td>
<td>F</td>
<td>Biotin 1 &gt; 30 photo, wall</td>
<td>MLS Lq=0.12 Lls=0.25 RWGL 0.1-0.5</td>
<td></td>
</tr>
<tr>
<td>Fig. 5, Sanger '93</td>
<td>MHC 6p21.31</td>
<td>HFF G1-cf ? ?</td>
<td>F</td>
<td>Biotin 2 &gt; 30 photo, wall</td>
<td>MLS Lq=0.1 Lls=0.18 RWGL 0.1-0.5</td>
<td></td>
</tr>
<tr>
<td>Fig. 1, Warrington '94</td>
<td>4p16.3</td>
<td>F G1-cf 75 MAA 3:1 FM 50 %</td>
<td>D</td>
<td>Biotin 2 ? photo, d-board</td>
<td>RWGL &gt; 5.0</td>
<td></td>
</tr>
<tr>
<td>Ish 1, Warrington '94</td>
<td>5q31-33</td>
<td>L ? ?? ?? ?? CLSM BioRad</td>
<td>F</td>
<td>Biotin 2 photo, d-board</td>
<td>RWGL &gt; 5.0</td>
<td></td>
</tr>
<tr>
<td>Fig. 2B, Yokota '95</td>
<td>4p16.3</td>
<td>F G1-cf 40 dropped MAA 3:1 FM 70 %</td>
<td>D</td>
<td>Biotin 2 40-360 photo, d-board</td>
<td>RWGL 2.0-4.0</td>
<td></td>
</tr>
<tr>
<td>Fig. 3B, Yokota '95</td>
<td>4p16.3</td>
<td>F G1-cf - PFA 4 %</td>
<td>D</td>
<td>Biotin 2 40-350 photo, d-board</td>
<td>MLS Lq=0.1 Lls=0.125</td>
<td></td>
</tr>
<tr>
<td>Fig. 2A, Yokota '97</td>
<td>4p16.3 R-band F G1-cf 40 dropped MAA 3:1 FM 70 %</td>
<td>D</td>
<td>Biotin 2 37-178 photo, d-board</td>
<td>RWGL 2.0-3.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fig. 2B, Yokota '97</td>
<td>6p21.3 R-band F G1-cf 40 dropped MAA 3:1 FM 70 %</td>
<td>D</td>
<td>Biotin 2 37-178 photo, d-board</td>
<td>RWGL 4.0-5.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fig. 2C, Yokota '97</td>
<td>21q22.2 G-band F G1-cf 40 dropped MAA 3:1 FM 70 %</td>
<td>D</td>
<td>Biotin 2 37-178 photo, d-board</td>
<td>RWGL 1.0-2.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fig. 2D, Yokota '97</td>
<td>Xp21.3 G-band F G1-cf 40 dropped MAA 3:1 FM 70 %</td>
<td>D</td>
<td>Biotin 2 37-178 photo, d-board</td>
<td>RWGL 0.5-9.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fig. 2D, Yokota '97</td>
<td>Xq28 G-band F G1-cf 40 dropped MAA 3:1 FM 70 %</td>
<td>D</td>
<td>Biotin 2 37-178 photo, d-board</td>
<td>RWGL 1.0-5.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fig. 2A, Yokota '97</td>
<td>Xp21.3 G-band F - PFA 4 %</td>
<td>D</td>
<td>Biotin 2 37-178 photo, d-board</td>
<td>RWGL 1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fig. 2A, Yokota '97</td>
<td>Xq28 G-band F - PFA 4 %</td>
<td>D</td>
<td>Biotin 2 37-178 photo, d-board</td>
<td>RWGL 0.25 MLS Lq=0.12 MLS Lls=0.200</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fig. 4B, Yokota '97</td>
<td>Xp21.3 G-band HeLa 40 dropped MAA 3:1 FM 70 %</td>
<td>D</td>
<td>Biotin 2 37-178 photo, d-board</td>
<td>RWGL 1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fig. 4B, Yokota '97</td>
<td>Xq28 G-band HeLa 40 dropped MAA 3:1 FM 70 %</td>
<td>D</td>
<td>Biotin 2 37-178 photo, d-board</td>
<td>RWGL 1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Momier '97</td>
<td>11q13</td>
<td>F - PFA 4 %</td>
<td>D</td>
<td>Biotin 1 22-69 CLSM</td>
<td>MLS Lq=0.12 MLS Lls=0.180</td>
<td></td>
</tr>
<tr>
<td>Momier '97</td>
<td>11q13</td>
<td>L - PFA 4 %</td>
<td>D</td>
<td>Biotin 1 22-69 CLSM</td>
<td>MLS Lq=0.1 Lls=0.0.24</td>
<td></td>
</tr>
<tr>
<td>Knoch '98/ Ranch '99</td>
<td>15q11-21 F - PFA 4 %</td>
<td>D</td>
<td>Biotin 1 &amp; 2 60-120 CLSM</td>
<td>MLS Lq=0.1 Lls=0.0.125</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
“Synoptic” 3D Architecture of Various Loci

A history “synoptic” comparison of the spatial distance mapping from their original background and aim, FISH methodological protocols, via microscopic imaging and restoration analysis procedures, to their interpretation, reveals that with time Multi-Loop Subcompartment models are favoured.
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 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 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 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.
Quantified TSA induced Morphology Changes

Trichostatin A induced histone acetylation can be quantified by \textit{in vivo} H2A-GFP confocal images and image correlation spectroscopy (iFCS), which is a scaling analysis, and reveals the opening of chromatin, and thus reorganization changes on scales from 0.2 to \(\sim 1 \mu m\), consistent with MLS models.
Quantified TSA induced Morpholoy Changes

Trichostatin A induced histone acetylation can be quantified by *in vivo* H2A-GFP confocal images and image correlation spectroscopy (iFCS), which is a scaling analysis, and reveals the opening of chromatin, and thus reorganization changes on scales from 0.2 to ~1µm, consistent with MLS models.

![Graph showing changes in fluorescence intensity over length](image)
Counting nucleosomes *In Vivo*

Counting nucleosomes in living cells with a combination of fluorescence correlation spectroscopy and confocal laser scanning imaging reveals not only the free unbound histone component but also the concentration in absolute numbers of bound histones. Thus, the absolute concentration distribution of histones can be determined and reveals again the typical expected distribution of aggregated chromatin loops.
**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).

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

<table>
<thead>
<tr>
<th>Nuclear diameter [µm]</th>
<th>Nuclear Volume [µm^3]</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>

© 1996-2011 All Rights Reserved
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 breakage probability depends on the detailed folding topology of the chromatin fiber and the RW/GL and MLS models differ largely.
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 breakage probability depends on the detailed folding topology of the chromatin fiber and the RW/GL and MLS models differ largely.

© 1996-2011 All Rights Reserved
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 breakage probability depends on the detailed folding topology of the chromatin fiber and the RW/GL and MLS models differ largely.

© 1996-2011 All Rights Reserved
DNA Sequence Organization

Determination of the concentration fluctuation function $C(l)$ and its local slope the correlation coefficient $\delta(l)$ are an indication for the i) degree of long-rang scaling behaviour, ii) general multi-scaling, and iii) fine-structure features, which all are connected to all levels of genome organization and especially also the three-dimensional genome architecture.

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

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

\[
C(l) = \frac{1}{Ll} \sqrt{\frac{1}{L-l} \sum_{s=1}^{L-l} \left[ \left( \sum_{k=1}^{l} Ln \right) - \left( \sum_{k=1}^{L} lN \right) \right]^2}
\]
Long-Range Correlations in DNA Sequences

Determination of the concentration fluctuation function $C(l)$ and its local slope the correlation coefficient $\delta(l)$ show that genomes show characteristic long-range correlations up to $10^8$ bp, i.e. in principle over their complete length. Beyond, the show a specific multi-scaling behaviour, as well as a characteristic fine-structure. All correlates with the 3D-architecture of genomes.
Long-Range Correlations in DNA Sequences

Determination of the concentration fluctuation function $C(l)$ and its local slope the correlation coefficient $\delta(l)$ show that genomes show characteristic long-range correlations up to $10^8$ bp, i.e. in principle over their complete length. Beyond, they show a specific multi-scaling behaviour, as well as a characteristic fine-structure. All correlates with the 3D-architecture of genomes.

![Graphs showing correlation coefficients and concentration fluctuation functions for various genomes.](image)
Sequence Organization and Phylogeny

The concentration fluctuation function $C(l)$ and its local slope the correlation coefficient $\delta(l)$ characteristic long-range fine-structured multi-scaling behaviour, which is specific. Tree construction thus leads to characteristic groups, which are similar to those suggested by classic phylogenetic trees. Thus, the sequence organization represents evolutionary lines.
Sequence Organization and Phylogeny

The concentration fluctuation function $C(l)$ and its local slope the correlation coefficient $\delta(l)$ characteristic long-range fine-structured multi-scaling behaviour, which is species specific. Tree construction thus leads to characteristic groups, which are similar to those suggested by classic phylogenetic trees. Thus, the sequence organization represents evolutionary lines.
Systems Biological Result Integration via the GLOBE 3D Genome Platform

All results will be integrated using our GLOBE 3D Genome Platform, established for analysis, manipulation and understanding of multi-dimensional complex genome wide data. Thus in reiterative cycles between experiments and simulations a systems biological/medical genome model will be achieved.
Systems Biological Result Integration via the GLOBE 3D Genome Platform

All results will be integrated using our GLOBE 3D Genome Platform, established for analysis, manipulation and understanding of multi-dimensional complex genome wide data. Thus in reiterative cycles between experiments and simulations a systems biological/medical genome model will be achieved.
Systems Biological Result Integration via the GLOBE 3D Genome Platform

All results will be integrated using our GLOBE 3D Genome Platform, established for analysis, manipulation, and understanding of multi-dimensional complex genome wide data. Thus in reiterative cycles between experiments and simulations a systems biological/medical genome model will be achieved.

© 1996-2011 All Rights Reserved
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 \textit{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

Biophysical Genomics, Cell Biology, Erasmus MC
Biophysics, BTZ Dresden
Thomas Weidemann
Cell Biology & Biophysics
EMBL
Malte Wachsmuth
Molecular Biophysics
KIP
Katalin Fejes-Toth
Karsten Rippe

Biological Sciences, UCSD
Suchit Jhunjhunwala
Menno van Zelm
Cornelis Murre
Clinical Genetics
Erasmus MC
Bert Eussen
Annelies de Klein

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

Biophysics of Macromolecules
DKFZ
Gabriele Müller
Waldemar Waldeck
Jörg Langowski

Erasmus Medical Center and BioQuant & German Cancer Research Center
High-Performance Computing Center Stuttgart, University of Stuttgart; Supercomputing Center Karlsruhe, University of Karlsruhe; Computing Center, Deutsches Krebsforschungszentrum Heidelberg (DKFZ)

© 1996-2011 All Rights Reserved
The 3D Architecture and Dynamics of the Immunoglobin Heavy-Chain and other Loci and its Functional Implications for Genome Organization

Knoch, T. A., Murre, C. & Grosveld, F. G.


Abstract

The general 3D architecture of the immunoglobin heavy-chain (Igh) and other loci was determined by a novel interdisciplinary combination of high-resolution FISH and high-resolution epifluorescence spectral distance microscopy with analytical analysis, computer simulations, as well as trilateration (Cell 133, 265-279, 2008). The Igh locus is organized into distinct regions that contain multiple variable (V_{H}), diversity (D_{H}), joining (J_{H}) and constant (C_{H}) coding elements. Determination of distance distributions between genomic markers across the entire locus showed that the Igh locus is organized into compartments consisting of small loops separated by linkers with in detail dynamic functional relevance: V_{H}, D_{H}, J_{H}, and C_{H} elements showed striking conformational changes involving V_{H} and D_{H}-J_{H} elements during early B cell development, culminating in a merger and juxtaposition of the entire repertoire of V_{H} regions to the D_{H} elements in pro-B cells allowing long-range genomic interactions with relatively high frequency. This is in agreement with our recent study of the Prader-Willi/Angelmann region using a similar approach (Differentiation 76, 66-82, 2008) and in agreement with the Multi-Loop-Subcompartment (MLS) model of chromosome organization predicting 60-150 kbp loop aggregates separated by a similar linker (Knoch, ISBN 3-00-009959-X, 2002). With a new technology we are also able to proof, that this holds for other loci as well. Synopsis with previous spatial distance measurement studies and combination with sequence correlation analysis of the DNA sequence, fine-structure multi-scaling analysis of the chromatin fiber topology or in vivo morphology of entire cell nuclei, electron microscopy of chromosome spreading studies and even the diffusion behaviour within the cell nucleus, are all suggesting such an MLS architecture. This framework reveals a consistent picture of genome organization joining structural and dynamical aspects ranging from the DNA sequence to the entire nuclear morphology level with functional aspects of gene location and regulation. Many previously contradictory viewpoints are resolved by this framework as well. Consequently, the determination of the general 3D architecture of the Igh and other loci has beyond its major functional relevance, huge implications for the understanding of the entire genome understanding in a holistic system-biological manner.

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 statistical mechanics, genomic uncertainty principle, 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, chromatin cross-linking, chromosome conformation capture (3C), selective high-resolution high-throughput chromosome interaction capture (T2C), confocal laser scanning microscopy, fluorescence correlation spectroscopy, super resolution microscopy, spatial precision distance microscopy, auto-fluorescent proteins, CFP, GFP, YFP, DiRed, 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**


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


