DISSERTATION

submitted to the

Combined Faculties for the
Natural Sciences and for Mathematics

of the
Ruperto-Carola University of Heidelberg
Germany

for the Degree of
Doctor of the Natural Sciences

presented by

Dipl.-Phys. Tobias A. Knoch

born in Mannheim

Date of oral examination: 11th July, 2002

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* Analysis of the Chromatin Distribution**

Construct Conversions in Simultaneous Co-Transfections

Referees: **Prof. Dr. Jörg Langowski**
Prof. Dr. Dr. Christoph Cremer

Zusammenfassung

Um sich an die immer noch weitgehend unbekannte sequentielle und dreidimensionale Organisation des menschlichen Zellkernes anzunähren, wurden strukturelle, skalierungs und dynamische Eigenschaften von Chromosomen und Zellkernen in der Interphase auf der Ebene der 30 nm Chromatinfaser mit Monte Carlo, Brownschen Dynamic und parallelen Computer Methoden simuliert. Unterschiede zwischen den benutzten Modellen erklären verschiedene experimentelle Bedingungen, die ein Multi-Loop-Subcompartment Modell mit 63-126 kbp Schleifen, die zu Rosetten aggregiert und durch einen 63-126 kbp Linker verbunden sind, favorisieren. Dieses Modell sagt auch die Mobilität von Molekülen durch moderat behinderte Diffusion korrekt vorher. Korrelations Analysen von komplett sequenzierten Archaea, Bacteria und Eukarya Chromosomen zeigten fein-strukturierte positive langreichweitige Korrelationen aufgrund von Codon-, Nucleosomen- und Blockorganisation dieser Genome, die Klassifikation und Baumkonstruktion ermöglichen. Konstruktion und Expression von Fusionsproteinen aus den Histonen H1, H2A, H2B, H3, H4 und mH2A1.2 mit den autofluoreszenten Proteinen CFP, GFP, YFP, DsRed-1 und DsRed-2 erlaubte die Untersuchung der Chromatinstruktur in vivo während der Interphase, der Mitose oder Apoptose und zeigte verschiedene Interphase-Morphologien, quantifizierbar mit Skalierungsanalysen. Schließlich, wurden Konstrukt-Konversionen in simultanen Ko-Transfektionen basierend auf Rekombination/Reparatur/Replikation in ≤25% der Zellen entdeckt.

Abstract

To approach the still largely unknown sequential and three-dimensional organization of the human cell nucleus, the structural-, scaling- and dynamic properties of interphase chromosomes and cell nuclei were simulated on the 30 nm chromatin fiber level with Monte Carlo, Brownian Dynamics and parallel computing methods. Differences between used models explain various experimental conditions, favouring a Multi-Loop-Subcompartment model with 63-126 kbp loops aggregated to rosettes and connected by 63-126 kbp linkers, and predict correctly the mobility of molecules by moderately obstructed diffusion excluding the Inter-Chromosomal Domain hypothesis. 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 and tree construction. By construction and 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, the chromatin morphology could be investigated in vivo during interphase, mitosis or apoptosis and revealed different interphase morphologies for cell lines quantifiable by scaling analyses. Finally, construct conversions in simultaneous cotransfections due to recombination/repair/replication were discovered in ≤25% of cells.

Dr. Tobias A. Knoch
Offenburgerstr. 100
D-68239 Mannheim
Germany

Dissertation Edition
Dissertation Edition (soft cover, 1st edition)
Special Dissertation Edition (hard cover, 2nd edition)
ISBN 3-00-009959-X (soft cover, 3rd edition)
ISBN 3-00-009960-3 (hard cover, 3rd edition)

Printing: German Cancer Research Center (DKFZ), Heidelberg
Binding: Dyroff, Heidelberg

Printed in Germany on Acid-Free Paper

Acknowledgements

Prof. Dr. Jörg Langowski is thanked for the possibility to conduct this thesis under ideal conditions in his interdisciplinary group Biophysics of Macromolecules, German Cancer Research Center (DKFZ), Heidelberg.

Prof. Dr. Dr. Christoph Cremer is thanked for many discussions and for representing this thesis at the Combined Faculty for the Natural Sciences and Mathematics of the University Ruperto-Carola, Heidelberg.

Prof. Dr. Josef Bille and ***Prof. Dr. Reimer Kühn*** are thanked for being members of the disputation jury.

Dr. Thomas Weidemann is grateful acknowledged for his supportive friendship, pluralistic debates and his constructive criticism.

Dr. Malte Wachsmuth is regarded for many ideas and discussions as well as for his great motivation and cooperation in experiments.

Dr. Markus Hammermann is not only thanked for sharing the same laboratory and many thoughts but also for his friendship and many theatre visits.

Dr. Volkmar Baumgärtner must be tributed for his critical impulses, creative discourses and his friendship. Thanks go also to ***Dipl.-Arch. Miriam Baumgärtner*** for her great encouragement.

Dr. Georg Eysel is recognised for his support, distraction from lab work and for organizing several flats in the end phase of this thesis.

Secretary Erica Stolte is thanked for very her friendly and happy support.

Dr. Christian Münkel is appreciated for writing parallel simulation code and initiating the work leading to Chapter 2 “Simulation of Single Chromosomes”, later succeeding into Chapter 3 “Simulation of Interphase Nuclei”, and Chapter 5 “Simulation of the Dynamics in Interphase Nuclei”.

Dr. Gregor Kreth should be recognized for many constructive questions and ideas.

Dipl.-Biol. Markus Göker is thanked for his friendship, many constructive ideas, his expertise and cooperation in the work of Chapter 6 “Long-Range Correlations in DNA Sequences” as well as for initiating a cooperation leading to fractal analyses of foraging trails of leaf cutting ants.

PD Dr. Rudolf Lohner is greatly appreciated for his ready enthusiasm and support with exact numerics, programming, implementation as well as job management on the IBM SP2 concerning Chapter 6 “Long-Range Correlations in DNA Sequences”.

Dr. Waldemar Waldeck is greatefully acknowledged for many fruitful scientific conversations leading to Chapter 7 “Chromatin Alive” as well as other projects. ***TA Gabriela Müller*** is thanked for introducing me to cell culture and some molecular techniques as well as for her general technical assistance.

Dr. Katalin Fejes-Toth and ***PD Dr. Karsten Rippe*** for valuable advice and dyn-amic cooperation on chromatin regulation, related to Chapter 7 “Chromatin Alive”.

Prof. Dr. Angel Alonso is thanked for providing some of the coding sequences for the development of histone-AFP fusionproteins and many information and ***Prof. Dr. Detlef Doenecke*** is thanked for providing plasmids encoding histones with their natural promoter for the histone-AFP fusionproteins.

Dr. Michael Stoehr is not only honoured for his cooperation in fluorescence activated cell sorting (FACS) measurements, cell cycle analysis and the many another help but also his original open-mindedness and instrumental conversations. Great thanks go also to ***TA Monika Stoehr*** for her technical assistance and proceeding of the cell cycle analysis as well as for her friendly motivation.

Dipl.-Ing. Andreas Hunziker is thanked for extremely well control sequencing of plasmids encoding the histone-AFP fusionproteins, for the difficult sequencing of PCR products proving the “GFP-Walking” of Chapter 7 and many a relaxing chat.

Prof. Dr. Peter Lichter is regarded for providing the confocal laser scanning microscope facilities used in Chapter 7 “Chromatin Alive” and helpful discussions.

Dr. Karsten Richter and ***Dipl.-Biol. Michaela Reichenzeller*** are thanked for many fruitful discussions on the three-dimensional organization of the cell nucleus and cooperative advancement of the confocal laser scanning microscope facilities.

Dipl.-Biol. Felix Bestvater has to be honoured for making the initial observation and for the excellent cooperation on the equally shared project leading to Chapter 8 “GFP-Walking” as well as for his dear friendship.

Prof. Dr. Eberhard Spiess needs not only to be regarded for his ready cooperation leading to Chapter 8 “GFP-Walking” but also for his motivation, the many discussions and the great support of the Diploma/PhD student initiative I worked in.

Dipl.-Chem. Nina Baudendistel is thanked for continuing the AP1-AFP project as her PhD thesis and for her great cooperation.

Dr. Karin Monier and ***Prof. Dr. Kevin Sullivan*** are thanked for many important discussions and suggestions, e. g. concerning Chapter 7 “Chromatin Alive”.

Besides thanking all other members of the group Biophysics of Macromolecules, all those from the DKFZ who supported me, I would like to appreciate all of those who integrated my results and propositions into cooperative research projects: ***Dr. Peter Quicken***, ***Dr. Anna Friedel*** and ***Prof. Dr. A. Kellerer*** integrated simulations from Chapter 1 “Simulation of Single Chromosomes” into “Simulation of Fragment Distributions caused by Carbon-Ion Irradiation”. ***Dr. Rainer Wirth’s*** and ***Dipl.-Biol. Christian Kost’s*** foraging trails of leaf-cutting ants were subject to fractal analyses. ***Dr. Gabriela Westphal*** cooperated me into her Nerve Growth Factor (NGF) and Erythropietin and ***Dr. Shirley Gil Parado*** into her µ-calpain research counting on the histone-AFPs and *in vivo* confocal laser scanning microscopy.

The ***German Cancer Research Center (DKFZ)*** is thanked for providing me with a dissertation grant, besides its infrastructure especially parallel high performance computing on various IBM SP2 and HP machines. I would also like to thank the ***Federal Ministry of Research and Technology (BMFT)*** for its support under grant 01 KW 9602/2 (3D-Heidelberg Human Genome Study Group, German Human Genome Project). The ***High-Performance Computing Center Stuttgart (HLRS) and the Supercomputing Center (SCC) Karlsruhe*** guaranteeing a frictionless access to their CRAY T3E and IBM SP2s are also thanked for their many a support. The ***Molecular Graphics and Modelling Society (MGMS)*** as well as the ***German Cancer Research Center (DKFZ)*** are thanked for honouring my interdisciplinary research approaches by awards for a talk and a poster.

Great thanks go to those who supported me in the endphase of this thesis with their constructive criticism: ***Dr. Malte Wachsmuth***, ***Dr. Thomas Weidemann***, ***Dr. Waldemar Waldeck***, ***Dipl.-Biol. Markus Göker***, ***PD Dr. Rudolf Lohner***, ***Dr. Katalin Fejes-Toth***, ***PD Dr. Karsten Rippe*** and ***Prof. Dr. Eberhard Spiess.***

Finally, I would like to thank deeply **Prof. Dr. Kurt Egger** and all people of the **Human Ecology (HÖ) Working Group** at the University of Heidelberg for their support, inspirations, insights and happy moments I was so lucky to receive.

To my parents ***Waltraut Knoch*** and ***Wolfgang F. Knoch***, I owe a debt to great to be repaid for their constant support, encouragement and love!

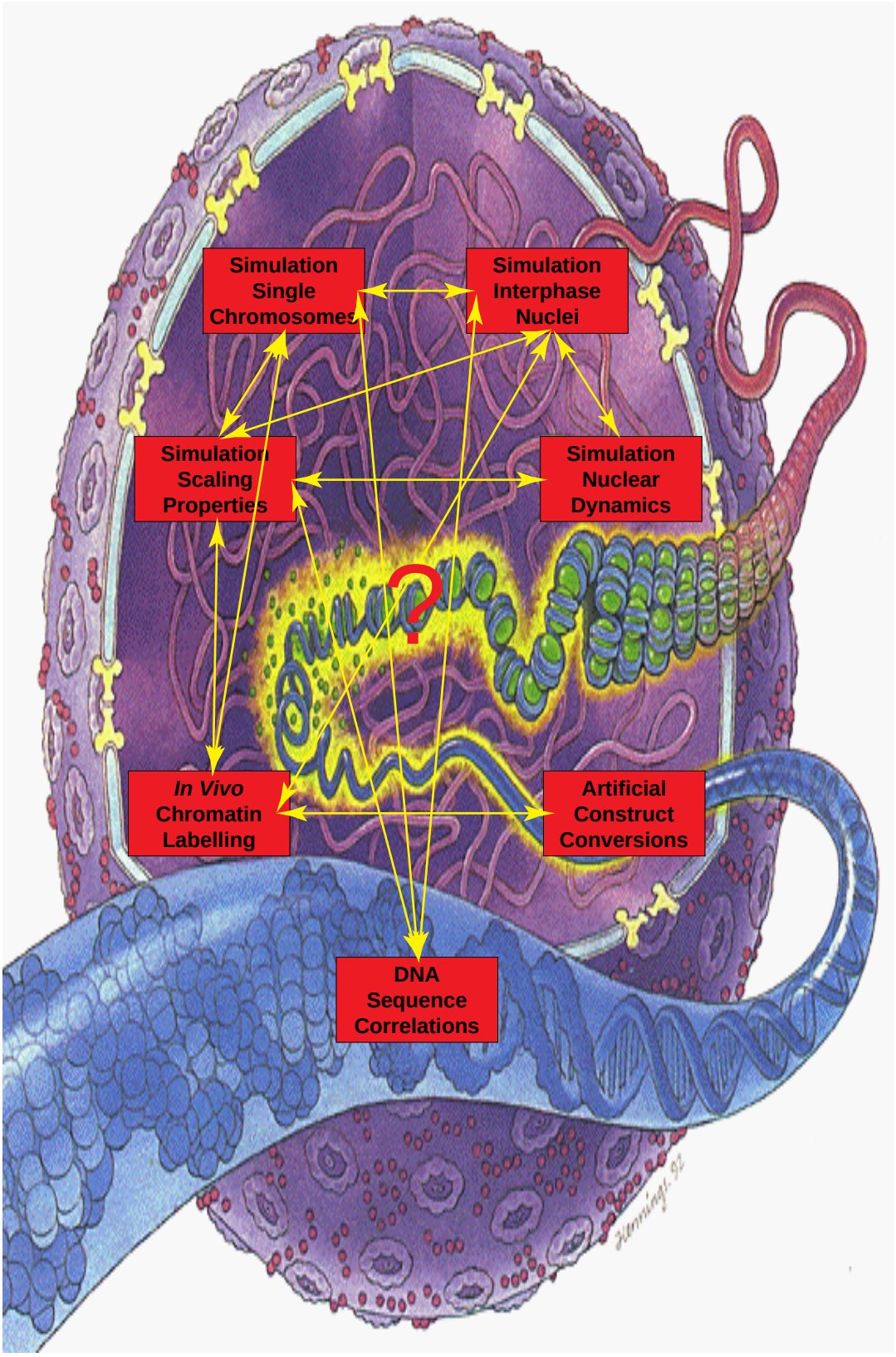
Heidelberg, 15th May, 2002,

Tobias Aurelius Knoch

Contents

List of Figures

List of Tables


Simulation
Single
Chromosomes
Simulation
Interphase
Nuclei
Simulation
Scaling
Properties
Simulation
Nuclear
Dynamics
?
In Vivo
Chromatin
Labelling
Artificial
Construct
Conversions
DNA
Sequence
Correlations

1 Introduction

1.1 Intention

In human cells the genetic information controlling most processes from the cellular level, over embryogeneses to cognitive ability, manifests in a diploid set of 23 DNA molecules, the chromosomes. They consist of ~$7x10^9$ base pairs (bp) storing around $1.4x10^{10}$ Bit or 1.75 GByte. This whole genome, whose added molecular length totals ~2 m, is kept in comparably small cell nuclei with typical diameters of ~10 µm or volumes of 500 $µm^3$. This corresponds to a compaction factor of ~$2x10^5$. Consequently, beyond pure compaction, the structuring of the genetic information in several organizational levels seems obvious to allow on the one hand sufficient performance during information transcription in interphase and on the other hand replication of the information and segregation of the chromosomes into the daughter cells in metaphase. Additionally, the abundant mutations need to be continuously found, controled and repaired to avoid the inevitable course of entropy. Considering the huge length and time scales, which bridge 10^{-9} to 10^{-5} m and 10^{-10} to 10^4 s, the genetic information of the human genome involves seven organizational levels according to current believe (Fig. 1.2, 1.2, 1.3): the DNA double helix (i) winds around a protein complex forming the nucleosome (ii), which condenses irregularly to the 30 nm chromatin fiber (iii), which is folded into chromatin loops (iv), which aggregate to chromosomal subdomains (v), which constitute a chromosome (vi), being nonrandomly arranged in the nucleus (vii). The DNA double helix (1.2.1) and the nucleosome (1.2.2) structure are known to atomic precision, but already the detailed nucleosome conformation in the 30 nm chromatin fiber is still debated

---

Fig. 1.1 Approaching the Three-Dimensional Organization of the Human Genome
To approach the three-dimensional organization of the human genome holisticly from different aspects covering its entire length and time scale, the structural-, scaling- and dynamic properties in the simulation of interphase chromosome and cell nuclei were analysed, long-range correlations in complete genomes were investigated, a method for the *in vivo* quantification of the chromatin distribution was developed and construct conversions in simultaneous co-transfections were discovered (thesis chapters are red and connected green). (Image: V. Hennings in Wolffe, 1993.)

(1.2.3). The latter holds even more for the higher-order structures having born many a hypothesis: Whereas light microscopic studies by Rabl and Boveri proposed territorial chromosomes with a hierarchical, self-similar organization of chromatin fibers in the late 19th century (1.3.1), electron microscopy suggested thereafter a random interphase chromatin organization in the models of Comings and of Vogel & Schroeder (1.3.2). To explain the high condensation degree of metaphase chromosomes and their stainable ideogram bands, chromatin loops attached to a nuclear matrix scaffold were suggested in the Radial-Loop-Scaffold model by Paulson & Laemmli (1.3.3). According to Pienta & Coffey these loops persisted in interphase forming stacked rosettes in metaphase (1.3.4). Microirradiation and fluorescence *in situ* hybridization (FISH) finally proved a territorial organization of chromosomes, of their arms, and of subchromosomal domains and led to the Inter-Chromosomal Domain (ICD) model hypothesizing an interchromosomal channel network (1.3.5). For the intraterritorial chromatin folding, the chromonema fiber (CF) model by Bruce & Belmont postulated a helix hierarchy (1.3.6), whereas in the Random-Walk/Giant-Loop (RW/GL) model 1 to 5 Mbp loops are attached to a non-protein backbone (1.3.7) and in the Multi-Loop-Subcompartment (MLS) model 60 to 120 kbp loops form rosettes connected by similar linker (1.3.8).

The intention of this thesis was to approach the debated sequential and three-dimensional organization of the human genome integrating aspects from all nuclear scales (Fig. 1.1). Therefore, different RW/GL and MLS topologies of single chromosomes were simulated on the level of the 30 nm chromatin fiber, to determine whether chromosome territories form, whether different morphologies appear, and whether these models can be distinguished experimentally (Chapter 2). The simulations were extended to whole nuclei containing all 46 chromosomes, to confirm the results of Chapter 2, to investigate the nuclear arrangement of chromosomes and the hypothesis of the ICD model (Chapter 3). Since from the above simulated chromatin fiber topologies to simulated confocal images of whole nuclei four closely connected compaction levels are bridged, their scaling behaviour was determined (Chapter 4). To explore the influence of the three-dimensional organization on the mobility of particles, their diffusion was simulated within nuclei, and their obstruction and accessibility to nuclear loci compared to the ICD predictions (Chapter 5). The coevolutive relation between the sequential and the three-dimensional genome organization was approached by analyzing the sequential correlation properties of completely sequenced Archaea, Bacteria and Eukarya genomes, including their multi-scaling, fine-structure and species specifity (Chapter 6). To overcome the limitations for the *in vivo* investigation of the morphology and dynamics of chromatin, a novel technique by labelling chromatin through expression of histone-autofluorescent fusionproteins was established and the interphase morphology and the course of mitosis and apoptosis were investigated (Chapter 7). This now widely used standard technique led to the discovery of construct conversions in simultaneous co-transfections, the clarification of their origin and appearance.

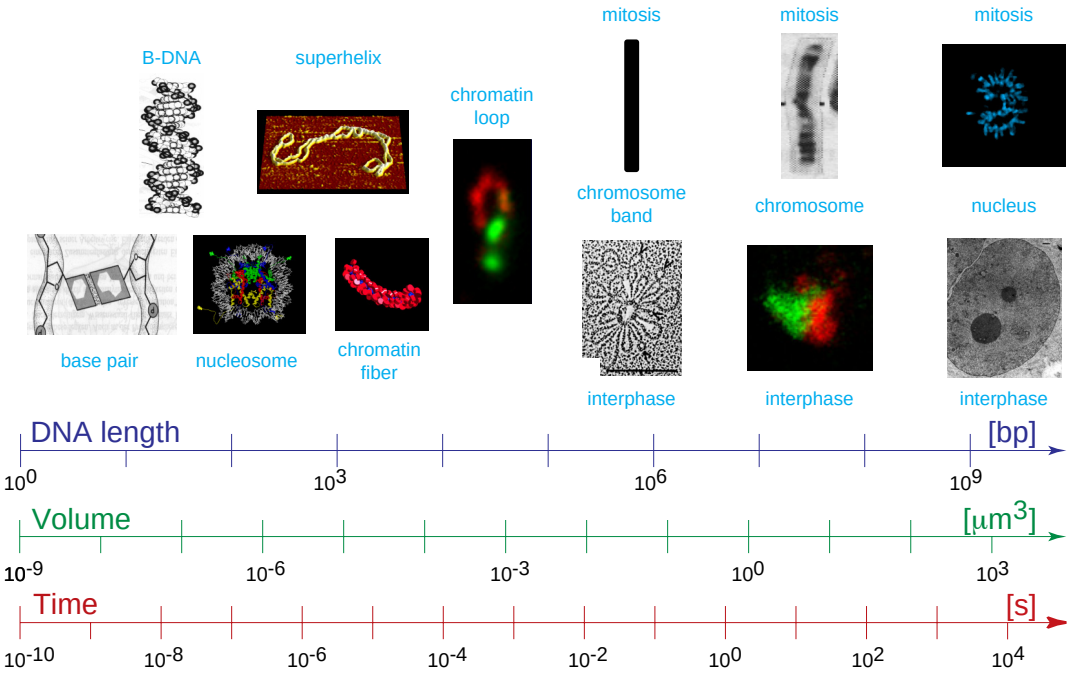


Fig. 1.2 Overview on the Size and Time Scaling of the Human Genome Organization
The scaling and the levels of organization range over 9 decades for base pairs, 12 decades for the volume (≈4 length decades) and 14 decades for the time: At the initial stage base pairs are formed composing the DNA double helix (Voet & Voet, 1995) which winds around the histone core complex building the nucleosome (Fig. 1.4), which condense into the 30 nm chromatin fiber (simulation image with courtesy of G. Wedemann, Division Biophysics of Macromolecules, German Cancer Research Center (DKFZ), Heidelberg, Germany). The DNA double helix forms also superhelices (scanning force microscopic plasmid image with courtesy of K. Rippe, Division Biophysics of Macromolecules, German Cancer Research Center (DKFZ), current: Division Physics of Molecular Processes, Kirchhof Institute for Physics, University of Heidelberg, Germany). The next compaction step consists of chromatin loops (Fig. 1.5) possibly forming rosettes (Fig. 1.6), which make up interphase chromosome arms and territories (Fig. 1.12) and the metaphase ideogram bands (Fig. 1.9). 46 chromosomes compose the human nucleus and are decondensed in interphase (Fig. 1.13) and condensed as shown for separated metaphase plates (Fig. 7.10). This thesis involves all length and time scales.

1.2 The First Three Compaction Levels of the Human Genome

The first three compaction levels are the evolutionary oldest. The initial storage of the genetic information, the DNA double helix, is even common to all three regia, the Archaea, the Bacteria and the Eukarya. Already histones and nucleosomes are only common to a few Bacteria presumably to their coevolution with Eukarya where nucleosomes are a common feature. Whether the pure formation of nucleosomes is directly connected with the condensation into the 30 nm chromatin fiber remains evolutionary unclear. Considering the huge length and times scales covered by the

whole genome, the genetic information is stored in the DNA double helix in the most direct manner compared to the higher compaction levels.

1.2.1 Deoxyribonucleicacid (DNA) Structure

Deoxyribonucleicacid (DNA) was isolated from pus cell nuclei and fish sperm by Miescher in 1869 shortly after he isolated chromatin in 1868. Although DNA was already suspected to be the carrier of the genetic information, this assumption seemed impossible regarding the variety of species and their complexity. This hypothesis remained unproven until transformation experiments on a strain of Pneumoccocus bacteria by O. T. Avery, C. M. MacLeod and M. McCarty in 1944, except for the much earlier chemical characterization: DNA, the presumably longest fibrous macromolecule, is a polymer consisting of four nucleotides (Fig. 1.3A), consisting of a ß-D-2´-Deoxyribose a phosphat group and one of four heterocyclic bases, the purines Adenin (A) and Guanin (G), as well as the pyrimidines Thymin (T) and Cytosin (C). These monomers are coupled to an unbranched single strand through a phosphat-sugar backbone. Consequently, the sequence of nucleotides with different bases primarily code for the genetic information, in contrast to the Deoxyribose and the Phosphat group being of structural importance.

The three-dimensional structure of DNA was discovered by J. D. Watson, F. H. C. Crick, L. C. Pauling and R. E. Franklin by X-ray diffraction in 1953: Two DNA single strands with antiparallel sense of direction, pair to a right handed double helix (Fig. 1.3A&B). Between Adenin and Thymin two and between Guanin and Cytosin three hydrogen bonds are formed. The bases are directed to the interior of the double helix and the sugar-phosphate backbone are directed to the exterior. The double helix forms only between complementary bases and complementary single strands. This, together with the higher elasticity module, ~50nm persistence length, and higher structural stability, is of fundamental importance for transcription, replication and repair and in consequence the use and evolutionary stability of the genetic information. In organisms, the so called DNA double helix is covered by a hydrate hull, has a diameter of ~2.4nm, 10.4 base pairs per helical turn and 3.4nm of helical pitch. Due to the antiparallel sense of the single strand pairing, the glycosidic base binding to the Deoxyribose does not lay directly across, thus a minor and major groove forms whose deepness depend on the base pair tilts and turns against neighbouring bases and the helical diameter. Different base pair sequences lead not only to regions with different stability due to the hydrogen bonding, but also to curvature due to summation of different tilts and turns (Fig. 1.3C), besides the sequence motives for genes, their regulation as well as other patterns on various scales (1.3.3, Chapter 6, Lewin, 2000). Consequently, the sequence shapes the three-dimensional structure of the double helix, which is important for general protein binding, nucleosome formation and positioning as well as other structural and regulatory functions.

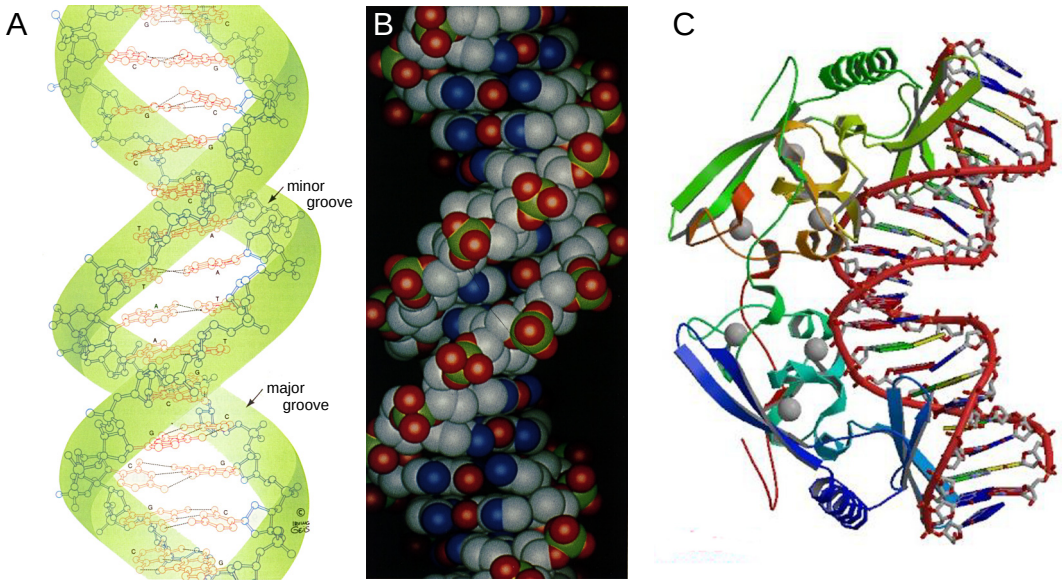


Fig. 1.3 DNA Double Helix Structure and Protein Binding to DNA
In the DNA double helix the bases of a two antiparallel and complementary DNA single strands coupled between the bases Adenin and Thymin by two and between the bases Guanin and Cytosin by three hydrogen bonds shown for the palindromic sequence GGTATACC (**A**: structural description; **B**: callot description; images: Voet & Voet, 1995). Due to the spatial arrangement of bases a minor and major grove form in the double helix. Since the relative position of bases relative to each other varies according to the DNA sequence, the double helix could also be curved being of regulatory impact or lead to improved binding for proteins which could also bent DNA for regulatory purposes and vice versa (**C**: Binding of the intron-encoded homing endonuclease I-PpoI to DNA; Flick *et al.*, 1998).

1.2.2 The Nucleosome

The existence of the nucleosome as the primary level of DNA compaction was first suggested by micrococcal DNAase digestion experiments of nuclear chromatin isolations, leading to DNA fragments with a minimal size ~146 bp. Therefore, the DNA must have been protected sterically from further digestion in contrast to pure isolated DNA (Clark & Felsenfeld, 1971). The existence of nucleosomes was confirmed by electron microscopy manifesting a protein complex associated to the DNA like perls on a string (Kornberg, 1973; Olins & Olins, 1974). However, only recently the three-dimensional structure of the whole nucleosome was resolved in detail by X-ray diffraction on crystalized nucleosomes (Fig. 1.4, Luger *et al.*, 1997).

In the cylindrical nucleosome with 5.5 nm height and 11 nm diameter, ~146 bp of DNA are wound in 1.75 turns and around a protein octamer core consisting of two times the histones H2A, H2B, H3, and H4. Additionally, one histone H1 or H5 is involved in the entry and exit of the DNA to the nucleosome. Specie-depending, two nucleosomes are connected by a ~50 to 110 bp linker, thus defining a repeat length of nucleosomes of 196 to 256 bp on the DNA double helix. The histone protein sequences and structures belong to the best evolutionary conserved and consist of two smaller α-helices flanking a central one in the middle part. Spatially, two H2A-

H2B heterodimers and one H3-H4 tetramer are clamped together with an antiparallel histone orientation and form the nucleosome core. The DNA double helix is attached by the polarity of the flanking α-helices and defined hydrogen bonds to the DNA phosphate backbone, by structural clamping of Arginins into the minor groove of the double helix, and a variety of non-polar interactions with the Deoxyriboses as well as further hydrogen and salt junctions. The binding probabilities and forces are not only directly connected to the three-dimensional structure of the DNA double helix, but also to the base pair sequence, thus specific nucleosomal binding sequences exist (Ambrose *et al.*, 1990; Blank & Becker, 1996; Liu & Stein, 1997; Lowary & Widom, 1998; Baily *et al.*, 2000). The existence of binding sequences could also indicate a locally different nucleosomal repeat length (Fig. 1.5D).

Besides the nucleosomal core, ~28% of the histone amino acid sequence belong to the C- and N-terminal histone tails, which reach out of the core partly with binding to the DNA. The major part of the tails are not connected to DNA and the exact spatial tail positions are unknown. Nevertheless, they are involved in the chromatin fiber formation and are a substrate for posttranslational modifications as acetylation, methylation, phosphorylation, ribosylation and ubiquitinilization and binding of proteins. These modifications are currently discussed as a second coding level called histon-code for the genetic information in addition to the base pair sequence.

Recently, the 42kDa heavy H2A derivate macroH2A (mH2A), was discovered (Pehrson *et al.*, 1992; Vijay-Kumar *et al.*, 1995; Pehrson *et al.*, 1997; Constanzi & Pehrson, 1998; Pehrson *et al.*, 1998; Lee *et al.*, 1998; Csankovszki *et al.*, 1999; Mermoud *et al.*, 1999; Rasmussen *et al.*, 1999; Rasmussen *et al.*, 2000). mH2A which locates preferentially in the inactive X chromosome, consists of two major parts: The N-terminal one has ~50% similarity to the usual H2A. The C-terminal one consists of a region with 57% similarity to H1 and a region common for DNA binding zink-finger proteins. This second part acts as C-terminal histone tail and might play a major role in the inactivation of the X chromosome.

1.2.3 The Chromatin Fiber

On the third level of DNA compaction, the chain of nucleosomes (1.2.2) is further compacted to the 30nm chromatin fiber under physiological conditions due to interactions of the histones and/or the total nucleosome (Fig. 1.5). The nucleosome arrangement within the 30nm chromatin fiber (Fig. 1.5C) as well as the local degree of compaction is still under debate (Woodcock *et al.*, 1993; van Holde & Zlatanova, 1995; Woodcock & Horowitz, 1995; Ehrlich *et al.*, 1997; Hammermann *et al.* 2000): In the solenoid model proposed by Finch & Klug (1976) the nucleosomes are confined to a solenoid with nucleosomes stretching to the exterior and the linker between the nucleosomes crossing the fiber interior. However, many *in vitro*, scanning force, and cryo-electron microscopic studies, favour a zig-zag arrangement of nucleosomes (Leuba *et al.*, 1994; Horowitz *et al.*, 1994). Nevertheless, under physiological conditions the average density of 6 nucleosomes per 11nm (e.g. ≈

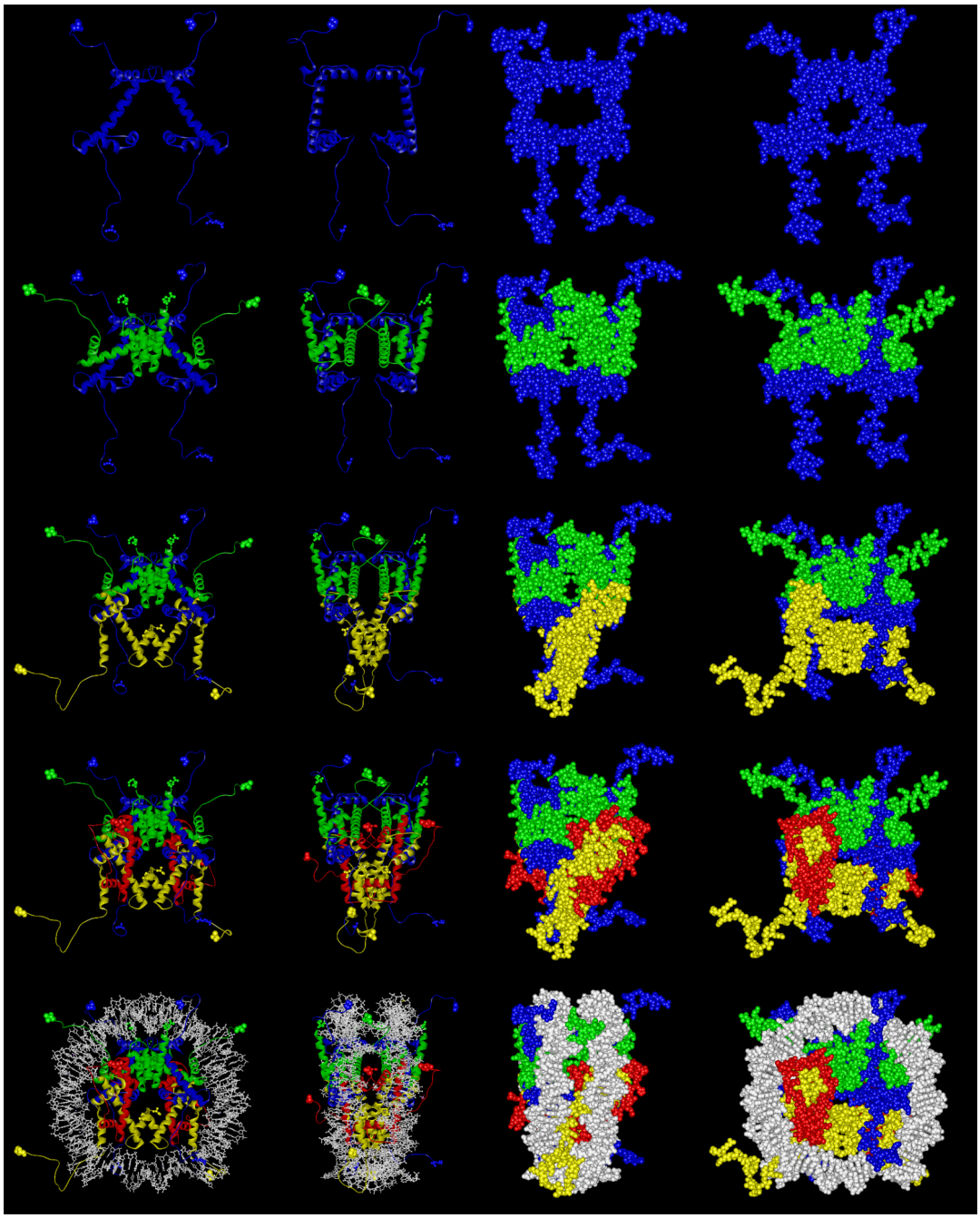


Fig. 1.4 Structure of the Nucleosome
The nucleosome is composed of two of each of the histones H2A(blue), H2B (green), H3 (yellow), H4 (red) around which the DNA double helix (white) is wound 1.75 times. The autofluorescent proteins (AFP) used in chapter 7 “Chromatin Alive” were attached to the C-terminus of the histones (C-terminus: small spheres; N-terminus: big spheres; both in left two columns only). Due to the movement of the histone tails their position could not be located in X-ray diffraction analysis. The position of the histone H1 is unknown and was therefore not included. (Atomic positions due to Protein Data Bank (PDB) entry 1EQZ, 06.04.2000, visualized with WebLeb Viewer-Lite, Molecular Simulations Inc., USA).

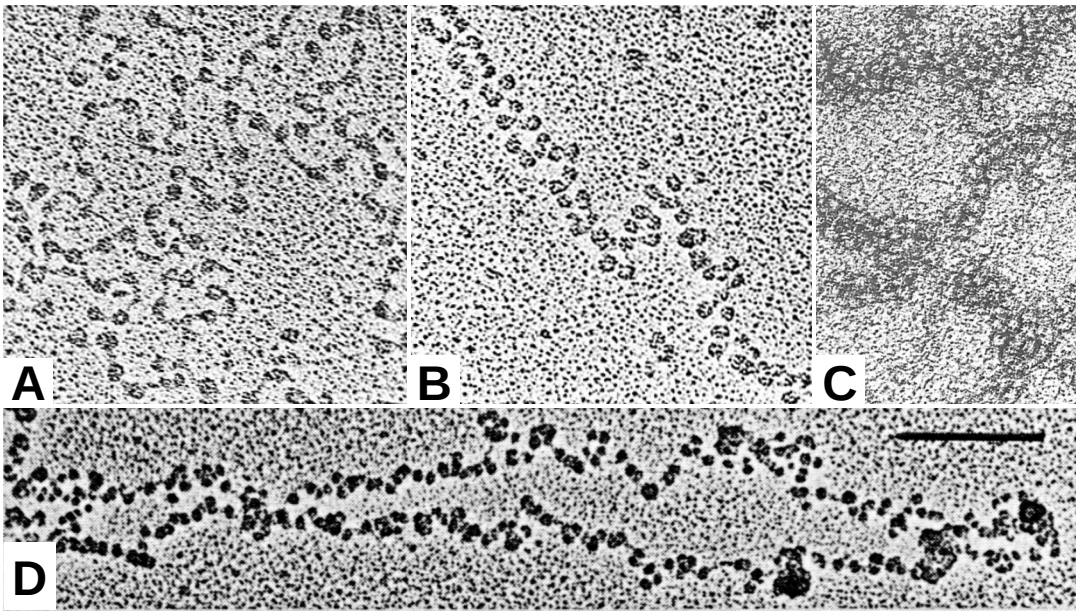


Fig. 1.5 Structure and Loops of the 30 nm Chromatin Fiber
Electron microscopy of chromatin reveals an open perl on a string motive structure without the histone H1 (**A**; 10 mM salt), which is condensed to the 30 nm chromatin fiber if the histone H1 is present (**B**, **C**; both 10 mM salt). The chromatin fiber is folded into loops, thus being the next stage in the higher organization of the human genome (**D**). Due to the partial decondensation of the 30 nm chromatin fiber because of the preparation, the uneven distribution of nucleosomes along the DNA sequence could be estimated. (images **A**, **B**, **C** from Voet & Voet, 1995; **D** from Reznik *et al.*, 1990).

105 bp/nm) is common for both models (Wolffe, 1995). Besides *in vitro* and electron microscopic studies (Horowitz *et al.*, 1994;Woodcock, 1994; Horowitz *et al.*, 1997; Woodcock & Horowitz, 1997; Woodcock & Horowitz, 1998; Bednar & Woodcock, 1999), neutron scattering experiments on intact nuclei of living cells revealed not only a fiber diameter of 30±5 nm but also that it is a dominant feature in the cell nucleus (Baudy & Bram, 1978; Baudy & Bram, 1979; Ibel, 1982; Notbohm, 1986). Recent cryo-electron microscopic studies reveal also density variations possibly combined with heterogeneous nucleosome positioning within the chromatin fiber (Fig. 1.7). Thus, also the 30 nm chromatin fiber, the third level of DNA compaction, plays an important role in the storage and regulation of the genetic information. Therefore, from the DNA, over the nucleosome to the chromatin compaction level, two chemical coding and three structural regulation schemes exist.

1.3 Hypotheses of the Higher-Order Chromatin Structure

In contrast to the DNA double helix (1.2.1) and the nucleosome core (1.2.2), the structure of which is known to atomic detail, all higher compaction levels require a

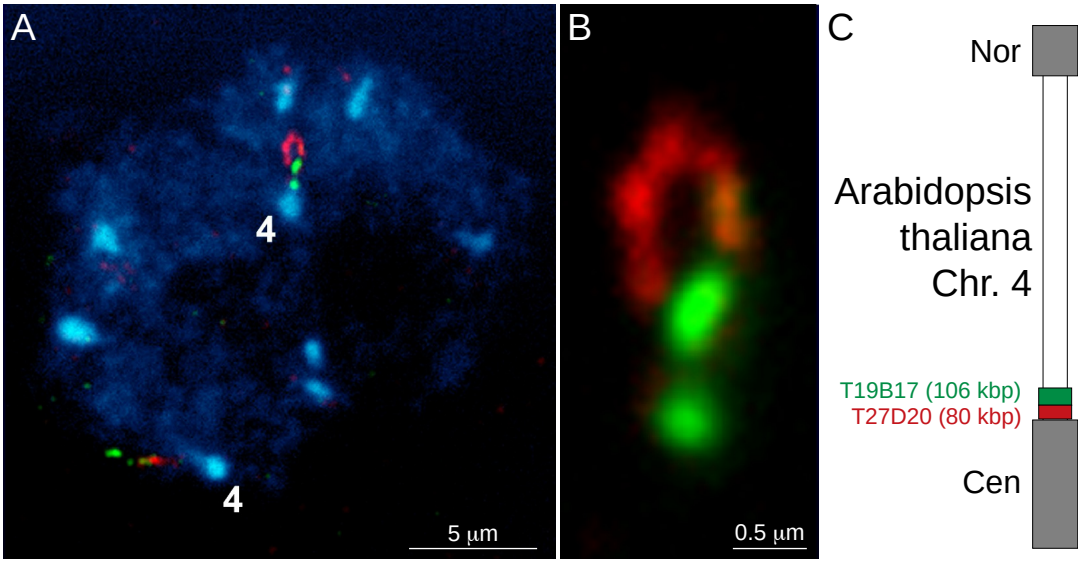


Fig. 1.6 Loop Structures in the *Arabidopsis thaliana* Genome visualized by FISH
A, Chromatin loop in a nucleus of a parenchyma cell of immature flowerbuds from Arabidopsis thaliana fixed with ethanol/acetic acid (3:1) and subsequent staining of two BAC (bacterial artificial chromosome) regions T19B17 (length: ~106 kbp; green) and T27D20 (length: ~80 kbp; red) using FISH. The loop is only visible in one of the two homologous chromosomes. The BAC locate near the centromer (Cen) on the top arm (Nor: telomer) of chromosome 4 (**C**). The DNA was counterstained with DAPI (blue). **B**, Enlargement of the chromatin loop reveals its length of 1.5 to 2 µm, thus the chromatin density ranges from 40 to 55 kbp/µm which is relatively low for Arabidopsis chromatin (100 to 200 kbp/µm; human: ~100 kbp/µm).This low density could be due to the harsh fixation procedure which is known to change the nucleosome status (sour extraction of histones!). (images with courtesy of P. Fransz, Swammerdam Institute for Life Sciences, BioCentrum Amsterdam, Amsterdam, The Netherlands; see also Fransz *et al.*, submitted).

more stochastic way of description due to increase of structural possibilities. The hypotheses for the higher-order chromatin organization range from highly random to very well defined, with a wide spectrum of combinations on different scales. The bigger the scale or the higher the organizational level, the more important the interplay between randomness and order gets.

1.3.1 The Rabl and Boveri Interphase Model

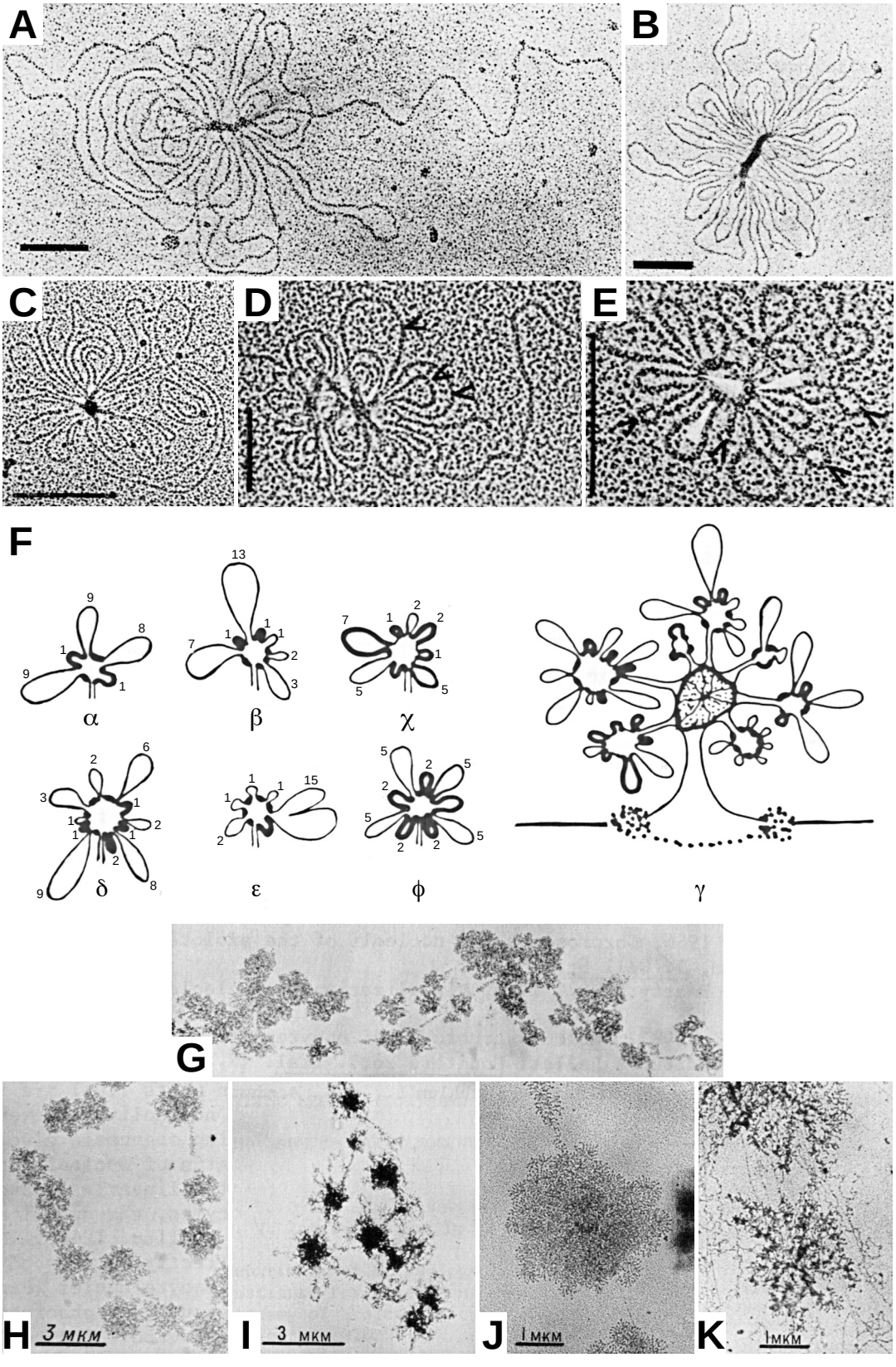
In the 19th century the high condensation and density degree of metaphase chromosomes were already assumed (Waldeyer, 1888), but the three-dimensional interphase organization of chromatin was highly speculative leading to a variety of hypotheses: e. g., the chromosomes should consist of chromatin spheres (Pfitzner, 1881) consisting of a kind of plasma with an unknown but nevertheless precisely defined structure (according to A. Weismann, for details and a general overview see Cremer, 1985). Based on light microscopic investigations on *Samandra maculata* and Proteus cells C. Rabl (1885) described that chromosomes kept their ana- and

telophase arrangement also during interphase. Beyond, from primary nuclear fibers, secondary and ternary fibers should extend and form the chromatin network of interphase nuclei (Fig. 1.8A). The primary nuclear fibers themselves should extend from a so called pole field, located on one side of the cell nucleus and containing the centromers, to the other anti-pole field located at the other side of the nucleus. T. Boveri (1909) extended this model by postulating that the chromosomes are organized in territories in interphase like in metaphase and that both conditions only differed in their degree of condensation.

1.3.2 The Interphase Models of Comings and of Vogel & Schroeder

With the advent of electron microscopy and the possibility to investigate the three-dimensional organization of cell nuclei with high resolution (Wischnitzer, 1973), the model of Rabl & Boveri (1885, 1909) and their territory or compartimentalization hypothesis of the nucleus seemed to be outdated. According to the new observations the metaphase organization is totally decondensed (Fig. 1.13A), thus according to Comings (1968) the chromatin fiber could roam the nucleus freely despite some few attachment points at the nuclear matrix (a hypothetical nuclear protein network postulated from existenceof the insoluble rest after biochemical extraction procedures of nuclei; Pienta & Coffey, 1974), the nucleolus and the nuclear membrane (Fig. 1.8B). This is in agreement with the localisation of so called euchromatin (less dense and more active chromatin) in the inner nuclear regions in contrast to the closer nuclear membrane localisation of so called heterochromatin. The condensation degree thereby should be regulatedby the density of chromatin attachment sites. In the relatively similar model of Vogel & Schroeder (1974) the density of attachment points depends on the number anddensity of the nuclear pores in the nuclear membrane. The chromatin fiber should stretch more or less linearly between these attachment points. Consequently, both models favour an unterritorialized and very random organization of chromosomes in interphase.

Fig. 1.7 Folding of the Chromatin Loops into Rosettes/Chromomeres in EM Images
Loops of the 30nm chromatin fiber form rosettes or also called chromomeres which are bound together by interacting granules at their bases (**A**-**E**). Depending on the preparation method the chromatin fiber including the nucleosomes and thus histones are preserved (**D**, **E**). **F**, The nucleosome distribution within a chromatin loop in a chromomere depends for genes on the composition of introns (thin lines) and exons (thick lines), their length fitting multiples of the nuleosomal repeat length as shown in schematic models (**α**: human preproglucagon gene, 6455bp; **β**: mouse MHC class II H2A-Ia-beta haplotype-b gene, 5801bp; **χ**: rat beta-actin gene, 2758bp; **δ**: chicken ovalbumin gene, 5280; **ε**: rabbit Ig germline kappa isotype K1 allotype b4 gene, 4545bp; **ϕ**: unknown; **γ**: sea urchin histone complex, H1, H4, H2B, H3 and H4; from Reznik *et al.*, 1990; see also Reznik *et al.*, 1991). The rosettes/chromomers are connected by linker chromatin to form whole chromosomes (**G**-**K**, see also linker chromatin in **A** and **D**). Within the rosettes also transcription takes place shown by the visibility of lambrushes (**J**, **K**). (Images: **A**, **B**:Avramova *et al.*, 1990; **C**: Salganik *et al.*, 1990; **D**-**F**: fromReznik *et al.*, 1990; **G**-**K**: Tsvetkov & Parvenov, 1990; see also Tsvetkov & Parvenov, 1995; scale bars: **A**: 160nm; **B**: 160nm; **C**: 500nm; **D**:125nm; **E**: 250nm; **G**: 8.0µm; **H**: 3.0µm; **I**: 3.0µm; **J**: 1.0µm; **K**: 1.0µm).


A
B
C
D
E
F
α
β
χ
δ
ε
ϕ
γ
G
H 3 мкм
I 3 мкм
J 1 мкм
K 1 мкм

1.3.3 Ideogram Banding and The Radial-Loop-Scaffold Metaphase Model

Metaphase chromosomes being the biggest nuclear structures were first described by C. W. Nägli (1842) and W. Hofmeister (1848), and later named with the greek word for stainable bodies by Waldeyer (1888). The visualization of the cylindrical metaphase chromosomes with light microscopy reveal their high degree of condensation. A compaction factor of ~1000 is estimated taking into account the average length of a linearized chromosome of ~4.5 cm and of a metaphase chromosome of ~4.3 µm (Fig. 1.9 A). The centromer where the two sister chromatids are attached in the metaphase plate before cell division and where the spindle fibers attach for their transport were already described.

The development of staining techniques led to the finding of the so called ideogram banding pattern classified as G-, Q-, R- and C-bands according to the used protocol (although the staining mechanism is still unknown; Fig. 1.9 B, Fig. 1.10): Giemsa or Quinakrin dark stained bands (G-/Q-bands) overlap with GC-rich chromosome regions containing up to 97% inactive genes (Goldman *et al.*, 1984). In contrast, Giemsa light or reversely stained bands (R-bands) are characteristic for AT-rich regions containing genes active through most of the cell cycle (Comings, 1978; Holmquist, 1992; Craig & Bickmore, 1993). Staining with Giemsa after a denaturing-renaturing preparation leads to a third type of bands (C-bands) mainly specific for in general AT-rich DNA regions and attributed to the so called constitutive heterochromatin containing repetitive and transcriptionally inactive DNA regions. According to their content of active genes, R-bands are replicated before G-bands and these before C-bands during the S-phase (Camargo & Cervenka, 1982). In summary, in the 24 different chromosomes of the human genome ~850 ideogram bands are present and split during the decondensation into interphase in ~2500 bands before they cannot be resolved anymore (Fig. 1.10; Francke, 1994).

Consequently, the chromatin fiber within metaphase chromosomes needs to be folded tightly and presumably into 30 to 120 kbp sized loops. According to the Radial-Loop-Scaffold model derived from electron microscopic images of histone-depleted chromosomes, these loops are attached to a protein scaffold forming the axis of the chromosome (Fig. 1.9 C; Paulson & Laemmli, 1977). The attachment was expected to be mediated by AT-rich scaffold-associated DNA regions possibly bound by Topoisomerase II or histone H1. The scaffold itself should consist of non-histone proteins, mostly Topoisomerase II (Earnshaw & Laemmli, 1983; Earnshaw & Heck, 1985). Based on immunofluorescent labelling, a helical topology with opposite isomery between sister chromatids was proposed (de la Tour & Laemmli, 1988; Rattner & Lin, 1985). The Radial-Loop-Scaffold was extended by assuming that the scaffold is in some regions parallel to the chromatid axis and helical in others (Saitho & Laemmli, 1994). The formation of ideogram bands would then correspond to different organizations of the scaffold and presumably also different loop sizes. The hypothesis, however, held only for metaphase chromosomes, although it seemed reasonable that regions containing inactive genes remained in this conden-

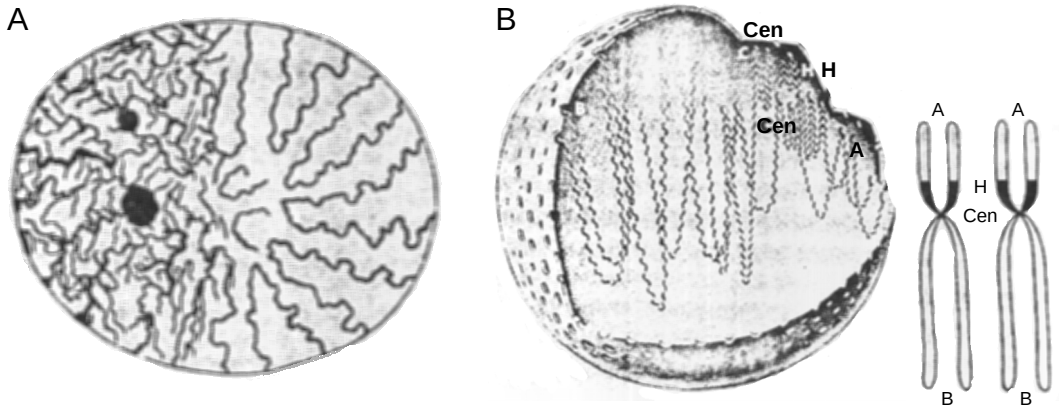


Fig. 1.8 Model of Rabl and Boveri, and Model of Comings
In the model of Rabl (1985) and Boveri (1909) the position and orientation of chromosomes is defined by a primary nuclear-fiber (right in **A**) from which secondary and tertiary fibers originate. In the model of Comings (1968) the chromosomes are attached to the nuclear matrix and the nuclear membrane (**B**). The attachment points are closer in the heterochromatin (H) than in the euchromatin (E). The chromosomes are polarised, that means that the centromers (Cen) are on one side of the nucleus whereas the telomers are on the other (A, B).

sated state during interphase. Thus, this model is not totally random but has a more definite higher-order chromatin structure.

1.3.4 The Pienta and Coffey Interphase-Metaphase Model

Based on a variety of experimental results Pienta & Coffey (1984), proposed a much more definite organization for the interphase topology of the 30 nm chromatin fiber than hypothesized by the models of Comings and of Vogel & Schroeder: Spriting of chromosomes leaves chromatin aggregations (e. g. Fig. 1.7) in which according to Vogelstein *et al.* (1980) the chromatin fiber was still attached to the nuclear matrix. Staining with ethidium bromid caused a diffuse, halo like visualization of these aggregates. Increased addition of ethidium bromid resulted in larger halo extensions (5 µg/ml), reversible at even higher concentrations (100 µg/ml; Benyajati & Worcel, 1976). The effect was explained by the detentioning and tensioning of superhelical DNA/chromatin. Quantification of the amount of DNA in the halos and their size increase suggested a looped organization of the chromatin fiber with loops of ~90 kbp. The discovery of replication taking place in these halos (Pardoll *et al.*, 1980) and the postulation of loops from electron microscopy (Paulson & Laemmli, 1977) supported these proposal. Consequently, Pienta and Coffey (1984) hypothesized that the loops exist as free DNA/nucleosome loops only during replication (Fig. 1.5 D; a recent visualization: Fig. 1.6), but else consist of condensed loops of the chromatin fiber (Fig. 1.11 A&B). At this time such decondensed DNA/nucleosome loops were often seen in electron microscopic spriting experiments. A comparison of further experimental data and their extrapolation to metaphase

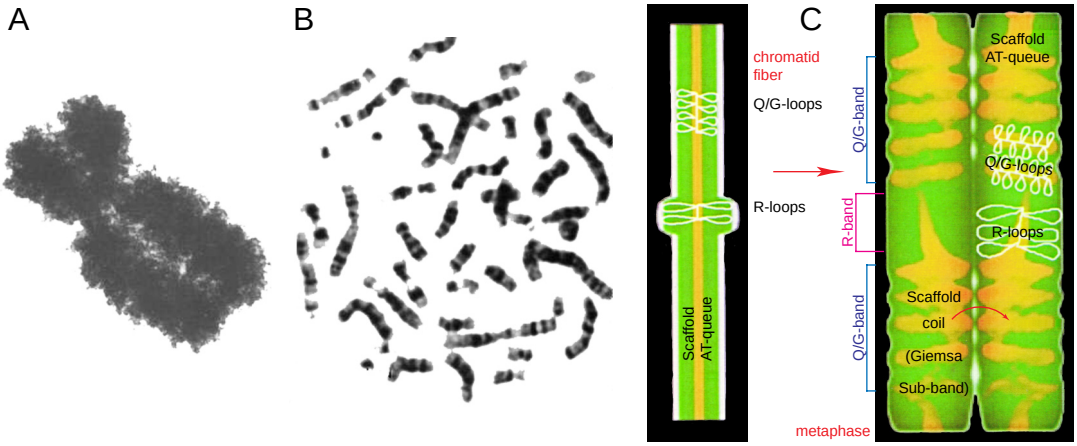


Fig. 1.9 Ideogram Banding and the Radial-Loop-Scaffold Model of Metaphase Chromosomes
In metaphase the chromosomes are condensed into cylinders, with a centromer separating the so called shorter p-arm from the longer q-arm and at which the sister chromatids after replication are tied together and where the spindle fibers attach for dragging the chromosomes into the daughter cells during cell division (**A**, image from Bloom & Fawcett). Staining with Giemsa or Quinakrin reveals a chromosome specific chromosome banding pattern: dark bands are called Q/G-bands, are GC-rich and contain mostly inactive genes in contrast to the light R-bands which are AT-rich and contain active genes (**B**, human metaphase spread from Alberts *et al.*, 1994). Condensation of the chromatin fiber into metaphase seems to involve chromatin loops of 30 to 120 kbp as proposed by electron microscopy (**A**). This led together with the ideogram banding to the Radial-Loop-Scaffold metaphase model (**C**, Paulson & Laemmli, 1977): Here the loops are attached to a protein scaffold whose folding parallel or helical to the chromosome axis creates R- or Q/G-banding.

chromosomes were only consistent with models proposed by Marsden & Laemmli (1979) and Adolf & Kreismann (1983). Therefore, Pienta and Coffey put forward a model with smaller chromatin loops of ~60 kbp, extending ~250 nm and attached to the nuclear matrix (Pienta & Coffey, 1974). Condensation of ~18 loops by folding of the nuclear matrix should result in a rosette like organization, the so called mini-bands, in metaphase chromosomes. Therefore, chromosome IV consisted of ~104 minibands whose stacking using the diameter of the chromatin fiber of 30 nm results also in the approximate length of the metaphase chromosome. Counting the numbers of loops in electron microscopic images turned out similar numbers (Dupraw, 1970; Laemmli, 1979; Utsumi, 1981). According to Cook *et al.* (1976) these mini-bands should dissolve in interphase, although the chromatin loops should still exist. In summary, Pienta & Coffey proposed a model with a mechanism for the transition from interphase to metaphase chromosomes integrating the different length and time scales from the chromatin fiber to the whole nucleus.

1.3.5 The (Extended) Inter-Chromosomal Domain (ICD) Interphase Model

The interphase models of Comings and of Vogel & Schroeder as well as that of Pienta & Coffey still proposed a very random organization of the 30 nm chromatin

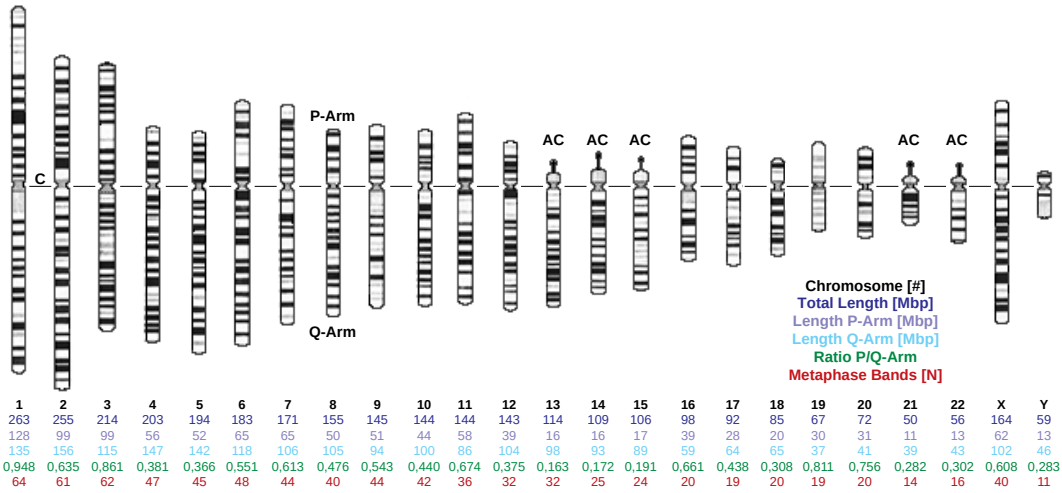


Fig. 1.10 Ideogram Banding and Data of the Human Genome
Above, The ideogram banding pattern of metaphase chromosomes shows differences in the size and distribution of the AT-rich R-bands (black), GC-rich G-bands (white), and the C-bands (grey). Below: The base pair content totals around 3.3 Gbp and varies between 263 and 50 Mbp for each chromosome with a mean of 137±61 Mbp/chromosome. The shorter P-arm contains between 128 and 13 Mbp with a mean of 46±30 Mpb/chromosome and the longer Q-arm contains between 156 and 37 Mbp with a mean of 91±35 Mbp/chromosome. The ratio P-arm/Q-arm varies from 0.948 to 0.283 and to 0.163 considering the acrocentric chromosomes (AC)13, 14, 15, 21 and 22. The number of metaphase ideogram bands in the haploid set of chromosomes totals ~850 and varies from 64 to 11 with a mean of 34±16 per chromosome, thus having a mean size of 4.03±1.9 Mbp/band. During decondensation into interphase the metaphase bands split up in around three subclusters, leading to ~2500 interphase bands for the haploid and ~5000 bands for the diploid set of chromosomes. (C: position of centromer; banding from Francke, 1994)

fiber roaming the whole nucleus unlike the interphase model of Rabl & Boveri. However, ultraviolet microbeam irradiation of cell nuclei damaged only a few chromosomes, compatible only the organization in chromosome territories since otherwise most of the chromosomes should have been damaged (Cremer *et al.*, 1974; Zorn *et al.*, 1976; Cremer *et al.*, 1982a; Cremer *et al.*, 1982b).

Only after the development of fluorescence *in situ* hybridization (FISH), in which fluorescently labelled DNA probes are hybridized to their complement in the nucleus after fixation and melting of the native DNA double helix, was the existence of chromosome territories proven (Fig. 1.13B; Cremer *et al.*, 1988; Lichter *et al.*, 1988; Pinkel *et al.*, 1988). The territories are more or less of round shape with little overlap (Eils *et al.*, 1995) and seemed to be very compact with a sharp border defining a surface on that level of observation. Thus, a hypothetical space between the territories, the Inter-Chromosomal Domain (ICD) was proposed into which chromatin loops from the dense chromosome body reach (Cremer *et al.*, 1993; Zirbel *et al.*, 1993). At the chromosomal surface and within the ICD space many important processes were expected to take place e. g. transcription, splicing, replication and repair as well as transport of e. g. DNA nucleotides, protein complexes, and mRNA.

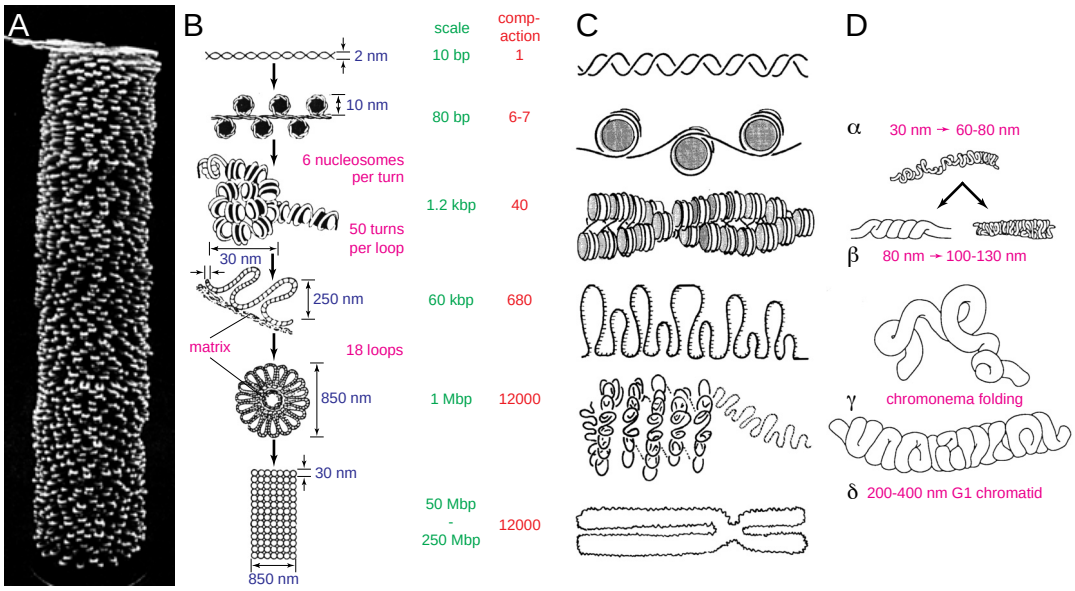


Fig. 1.11 The Pienta and Coffey as well as the Chromonema Fiber Model of Chromosomes
In the model of Pienta & Coffey (1984; **A**, **B**) the 30 nm chromatin fiber (which is a highly ordered solenoid) is folded into loops bound to a nuclear protein network, the nuclear matrix, during interphase. The loops contain ~60 kbp and extend ~250 nm. Condensation during mitosis leads to formation of rosettes by transformation of the nuclear matrix. The rosettes contain ~18 loops, have a diameter of ~850 nm and contain ~1 Mbp. The rosettes are stacked on top of each other to form a metaphase chromosome. A "real" model in which the white fiber has the same proportions as the 30 nm fiber is shown for a chromosome of 3.2 µm (**A**; the model has a length of 25 cm). In another unnamed model appearing in many textbooks and being a crossover model between the Pienta & Coffey and the Chromonema Fiber model, the helical chromatin fiber forms loops which are attached to the nuclear matrix but instead of forming rosette the chain of loops spiral up and form the sister chromatid (**C**, Alberts *et al.*, 1994). The chromonema fiber model being based on electron microscopic images of the higher-order structure of the chromatin fiber was proposed by Belmont & Bruce (1994, **D**): Here the 30 nm chromatin fiber is folded into the 60 to 80 nm wide chromonema fiber (**α**). The 30 nm chromatin fiber as well as the chromonema fiber form another fibrous structure with a diameter of 100 to 130 nm (**β**). These fibers fold again (**γ**) to form the interphase chromosome (**δ**).

Therefore, active genes should be preferentially located at the chromosomal periphery. Although this hypothesis was first confirmed for the ANT2 and ANT3 genes (Dietzel *et al.*, 1999) recent results showed that active genes are located throughout the territory (Tajbakhsh *et al.*, 2000). Unfortunately, the ICD space has not yet been visualized neither by electron nor by light microscopy. Recently, the fiber formation and localization of human Vimentin (carrying a nuclear localisation sequence since it is usually unpresent in the nucleus) was visualized by immunolabelling and by *in vivo* expression of fusionproteins with autofluorescent proteins led to a localization between the chromosome territories (Fig. 7.16, Reichenzeller *et al.*, 2000). Electron microscopy, however, shows no space extending beyond the Vimentin fibers.

The further refinement of FISH techniques revealed that also chromosome arms and the ideogram bands occupy distinct subregions of the interphase nucleus (Fig. 1.12; Cremer *et al.*, 1996; Dietzel *et al.*, 1998a; Dietzel *et al.*, 1998b). *In vivo*

puls labelling by incorporation of base analogs e. g. BrdU into the double helix during replication, led to the observation of sub-regions with ~1 Mbp of DNA and diameters of 300 to 800 nm (Jackson & Pombo, 1998; Ma *et al.*, 1999; Bornfleth *et al.*, 1999b). These so called foci are stable throughout the cell cycle (Zink & Cremer, 1998; Zink *et al.*, 1998; Zink *et al.*, 1999). Double pulse labelling early and late during replication showed that the foci correspond to R- and G- bands. As predicted the overlap between the foci was <10% and thus very low (Bornfleth, 1998; Zink *et al.*, 1999). Consequently, due to these results and simulation of chromosomes and nuclei (Knoch, 1998, Knoch *et al.*, 1998; Münkel & Langowski, 1998; Knoch *et al.*, 1999; Münkel *et al.*, 1999; Knoch *et al.*, 2000; Knoch, *et al.*, 2002) the initial Inter-Chromosomal Domain (ICD) model was refined and the surface and space between the territories extended to foci surface and the space inbetween (Cremer *et al.*, 2000). Recently, the investigation of particle diffusion in nuclei showed moderate obstructed diffusion behaviour at every spatial position (Wachsmuth *et al.*, 2000; Misteli *et al.*, 2001; Wachsmuth, 2001). Concerning, the arrangement of whole chromosome territories also a unrandom localization was found recently (v. Hase, 2000; Kreth, 2001; Habermann, 2001; Cremer, 2001; Tanabe *et al.*, 2002).

Consequently, the (extended-) ICD hypothesis increases the degree of order and compartimentalization to a higher degree than expectable from a polymerous distribution of the chromatin fiber and concerning the nuclear length and time scales.

1.3.6 The Chromonema Fiber Interphase Model

Bruce & Belmont (1994) proposed a helical folding of the 30 nm chromatin fiber resulting in the so called chromonema fiber with diameters 60 to 80 nm condensing further to a 100 to 130 nm fiber (Fig. 1.11 C). The latter folds irregularly to the interphase territory. The model is a direct consequence of the hypothetical solenoidal nucleosome arrangement in chromatin fiber, combined with the formation of chromosome territories. Using the genomic integration of a Lac-Repressor repeat DNA sequence and expressing a Lac-Repressor-autofluorescent fusionprotein *in vivo*, resulted in light microscopic images in agreement with the chromonema fiber and are consistent with the original results obtained by electron microscopy (Robinett *et al.*, 1996; Li *et al.*, 1998). The chromonema fiber agrees also with helical metaphase chromosome models (Fig. 1.11 D; Sedat & Manuelidis, 1977; Bak *et al.*, 1979).

1.3.7 The Random-Walk/Giant-Loop (RW/GL) Interphase Model

Again to explain the detailed topology of the 30 nm chromatin fiber within these territories, two-dimensional spatial distance measurements between genomic markers as function of their genomic separation revealed a strong monotonous proportionality of the spatial distance to the genomic separation (Lawrence *et al.*, 1988; Lawrence *et al.*, 1989; Lawrence *et al.*, 1990; Trask *et al.*, 1989; Trask *et al.*, 1991; for a

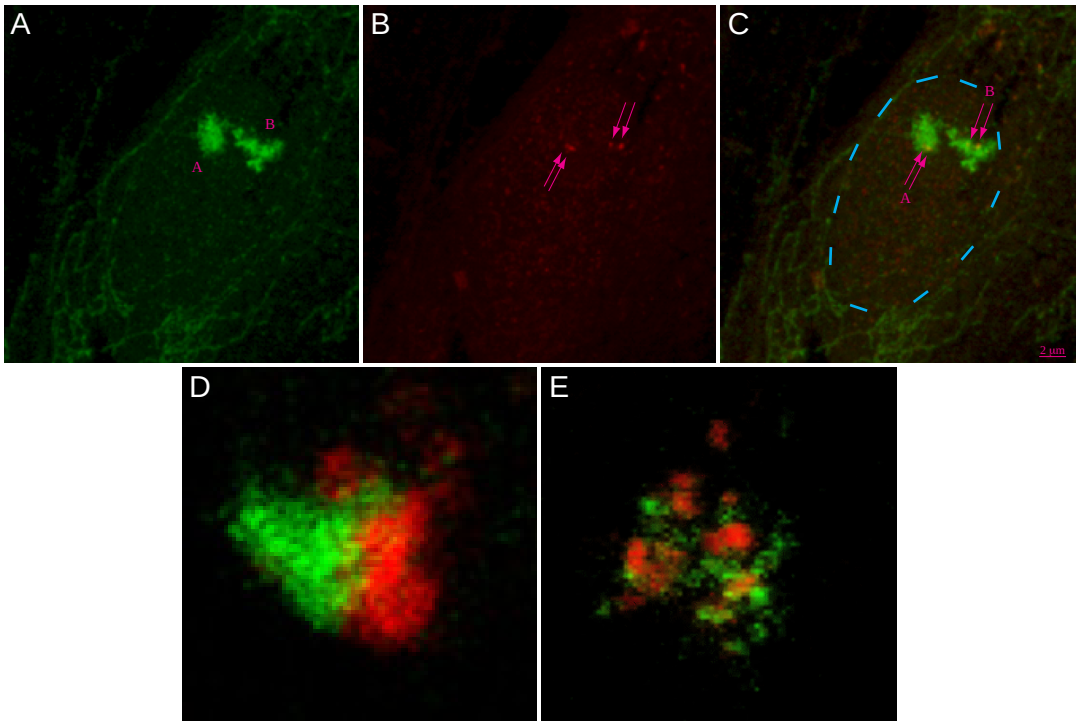


Fig. 1.12 FISH of Chromosomes, Chromosome Arms, Subcompartments and Genetic Markers
By chance the two homologous chromosomes XV neighbour each other in a FISH painting using fluorescein as fluorophore (**A**). Labelling of the genetic markers YAC48 and YAC60 with a genetic distance of 1 Mbp and using CY3 as fluorophore are spatially distinct (**B**). Overlay of **A** and **B** showing that the genetic markers lie within the chromosome 15 territory (nuclear membrane: blue, **C**). Painting of the p and q arm of chromosome 6 shows little overlap (**D**; with courtesy of S. Dietzel). Labelling of the ideogram bands/subcompartments of chromosome 15 shows their globular structure as predicted by the Multi-Loop-Subcompartment model (**E**; with courtesy of D. Zink, Ludwig-Maximilian University, Munich, Germany).

detailed discussion see 2.9). This was interpreted as a random walk behaviour of the chromatin fiber (van den Engh *et al.*, 1992). However, this hypothesis is incompatible with the formation of chromosome territories and their low overlap. Measurements using hypotonically swollen nuclei, thereafter revealed a biphasic behaviour of these distance measurements with different behaviours below and above ~2 Mbp (Trask *et al.*, 1993; Yokota *et al.*, 1995): Below ~2 Mbp the chromatin fiber seemed to extend rapidly with the genomic separation, and above ~2 Mbp a random walk behaviour was found. Therefore, Sachs *et al.* (1995) proposed the so called Random-Walk/Giant-Loop (RW/GL) model (Fig. 1.14), assuming loops of 1 to 5 Mbp being attached to a non-DNA and presumably protein backbone. This is in agreement with the hypothesis of a nuclear matrix to which the DNA or chromatin fiber is attached. The size of the backbone should be responsible for the extension of the chromosome territory. For quantitative analyses the analytical relation between the spatial distance and the separation of the genetic markers located on a fiber

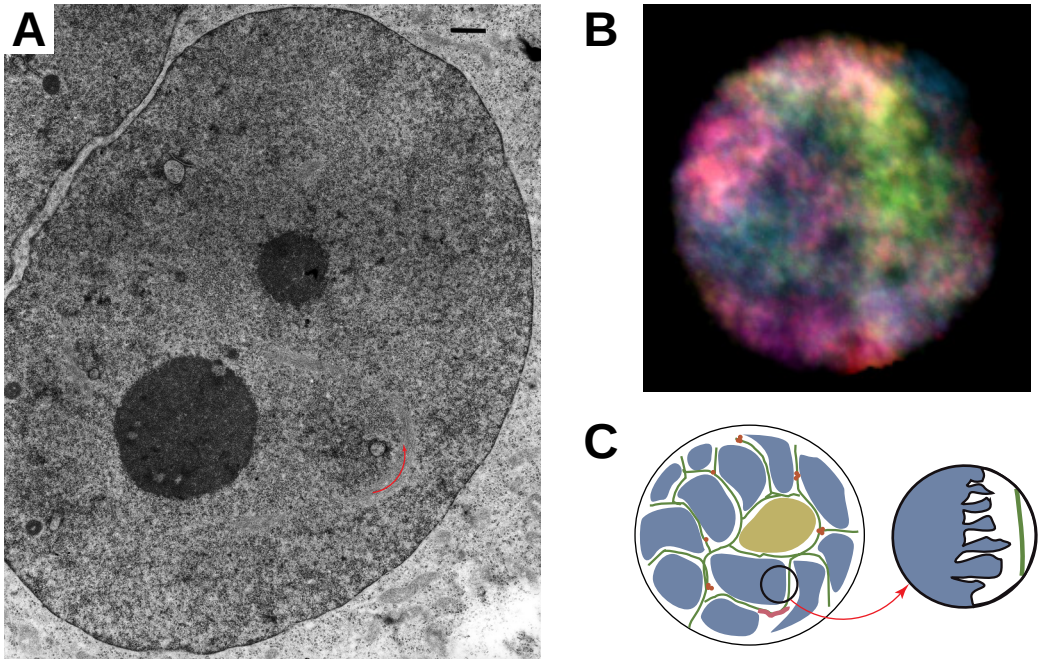


Fig. 1.13 The Inter-Chromosomal Domain (ICD) Model of Interphase Nuclei
In electron microscopic images despite chromatin density variations no compartimentalization into chromosome territories is visible (**A**, note the good visible nucleoli, with courtesy of K. Richter, Division Molecular Genetics, German Cancer Research Center (DKFZ), Heidelberg, Germany). In contrast multi-colour fluorescence *in situ* hybridization shows the 46 chromosomes organized into domains as shown in a fluorescence microscopic image of male human lymphocytes obtained from peripheral blood (**B**, with courtesy of K. M. Greulich-Bode, Division Genetics of Skin Carcinogenesis, German Cancer Research Center (DKFZ), Heidelberg, Germany). The latter led to the proposal of the Inter-Chromosomal Domain (ICD) model (**C**) of interphase nuclei in which the chromosome territories form compact domains being separated by the Inter-Chromosomal Domain space into which chromatin loops reach and in which transcription, splicing, replication and repair as well as transport of e. g. DNA nucleotides, protein complexes and m-RNA. To test the ICD-space hypothesis recently the fiber formation and localization of human Vimentin was analysed by immunolabelling and by *in vivo* expression of fusion proteins between human Vimentin and auto-fluorescent proteins (Fig. 7.16). The fibers localized between chromosome territories. Electron microscopy, however, reveals that beyond the human Vimentin fibers no space extends (**A**, vimentin fiber: red arrow).

(Equ. 2.4) could be extended to the case where loops are attached to a random walk. This analytical description also resulted in a Rayleigh probability distribution for distances between different parts of the fiber as in the case for the classic random walk. Fitting the experimental distance measurements (Trask *et al.*, 1993; Yokota *et al.*, 1995) with this extended analytical model revealed loop sizes of ~3 Mbp and separations of the attachment points of the giant loops of 620 nm. Unfortunately, this analytical description is not inversive definate thus at least two backbones could lead to the same interpretation (Liu & Sachs, 1997). Thus, the RW/GL model proposes a much more random organization of the chromatin fiber than expected from the ICD hypothesis.

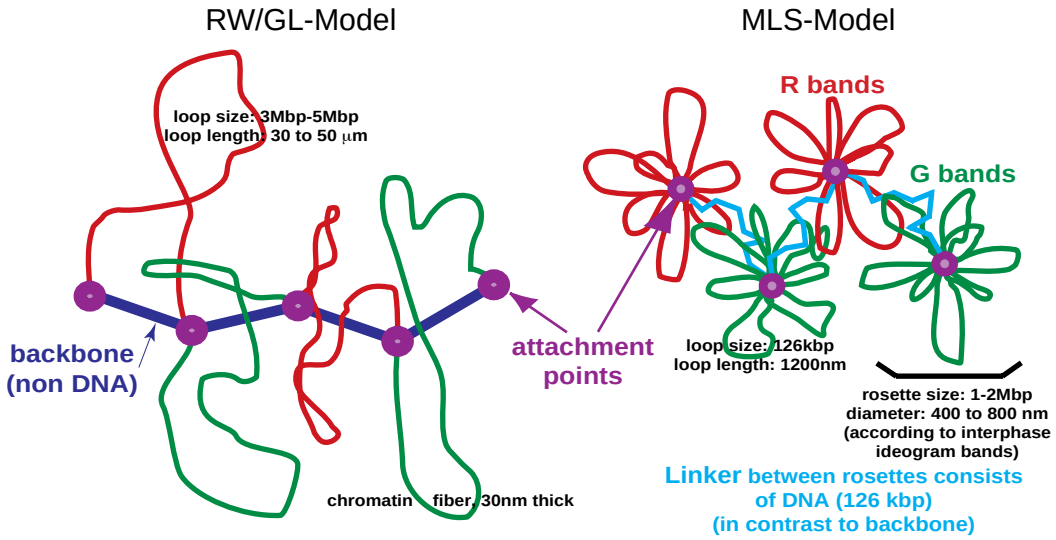


Fig. 1.14 Random-Walk/Giant-Loop (RW/GL) and Multi-Loop-Subcompartment (MLS) Model
In the RW/GL model by Sachs *et al.* (1995; **left**) big chromatin loops of ~3 Mbp are attached to a non-DNA backbone, presumably the nuclear matrix. The large loops intermingle freely not forming distinct features like in the MLS model. In the MLS model by Münkel and Langowski (1998; **right**) loops of ~ 100 kbp form rosettes which are connected by chromatin to make up the whole chromosome. The rosettes form subcompartments as separated organizational and dynamic entities.

1.3.8 The Multi-Loop-Subcompartment Interphase-Metaphase Model

More accurate spatial distance measurements between genetic markers as function of their genomic separation (Fig. 1.12) using a more structure preserving protocol also showed a biphasic behaviour. However, the increase in the spatial distance was slower with growing separations below ~2 Mbp (Yokota *et al.*, 1995) and therefore implied much smaller loops than in the RW/GL model (1.3.7). The behaviour above ~2 Mbp remained that of a random walk. Thus, the smaller loops need to be aggregated in connected clusters, which for larger genetic separations are arranged like a random walk. A detailed interpretation of these findings in combination with establishes interphase and metaphase hypothesis of chromosomes (1.3.1, 1.3.3, 1.3.4, 1.3.6) resulted in the proposal of the integrative Multi-Loop-Subcompartment (MLS) interphase-metaphase model (Münkel *et al.*, 1998): Here small loops of 60 to 120 kbp form rosettes in interphase which are linked by chromatin linkers of again ~120 kbp and responsible for size of the chromosome territory (Fig. 1.14). During metaphase the linker is contracted to a loop leading to stacking of the rosettes on top of each other. The base pair content of the rosettes corresponds to interphase ideogram bands which result from a division by three of each of the 850 ideogram bands of the haploid metaphase chromosome set (1.3.3). Thus, a rosette contains ~1 Mbp in agreement with replication foci (1.3.5). The low overlap of chromosome foci, arms and the whole territory can be explained because small loops explicitly exclude a high overlap. Recently such small chromatin loops were also

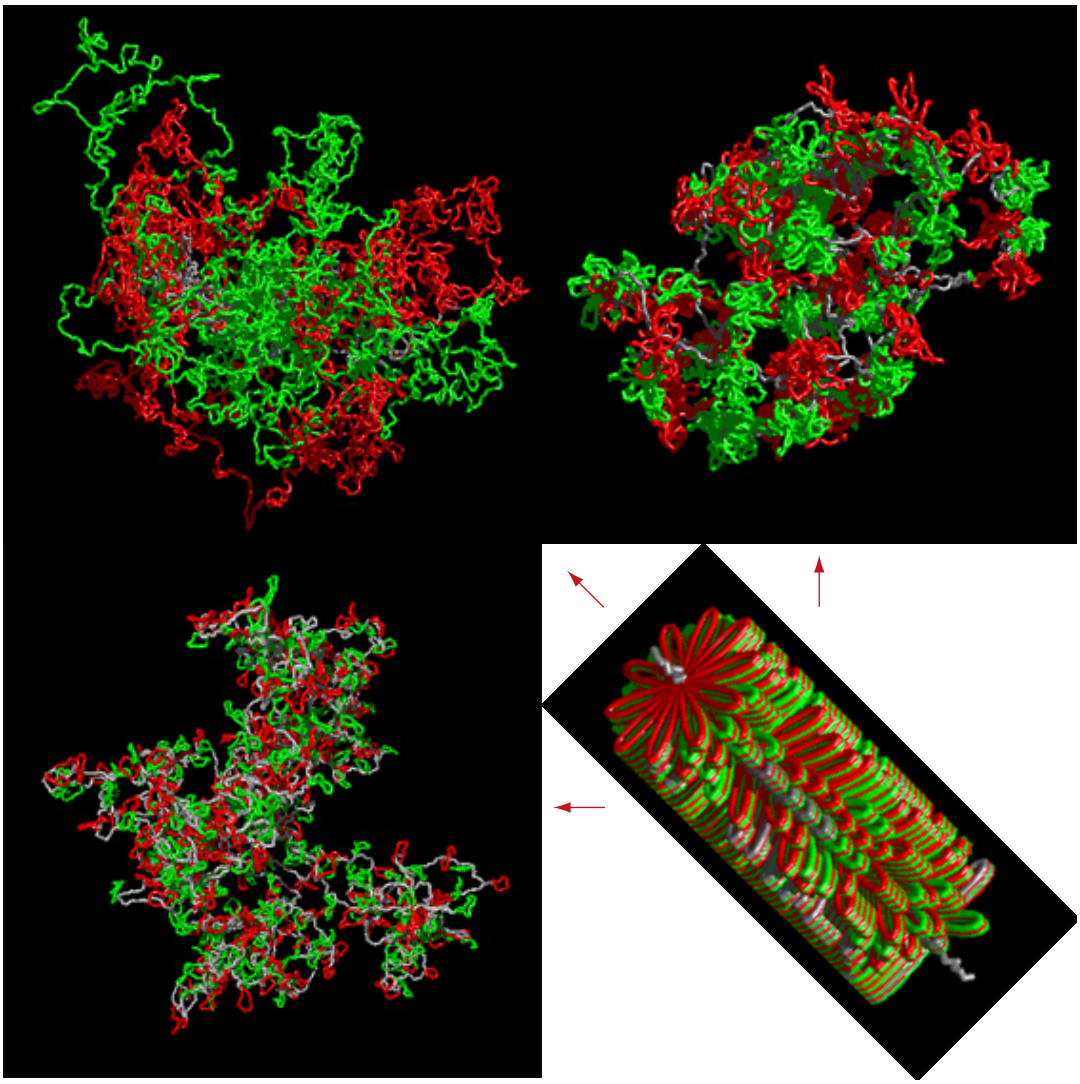
found by FISH labelling (Fig. 1.6). The model is also conform with electron microscopic images of chromosome spriting experiments showing chromatin loop aggregates and even images of chromatin rosettes, despite the harsh preparation conditions (Fig. 1.7).

In summary, the Multi-Loop-Subcompartment model integrates not only the territory hypothesis of Rabl & Boveri (1.3.1), the ideogram banding pattern and the Radial-Loop-Scaffold metaphase model (1.3.3), but also the interphase model of Pienta & Coffey (1.3.4), the (extended) Inter-Chromosomal Domain (ICD).

1.4 Questions Leading to This Thesis

Since many contrasting hypothesis exist about the sequential and three-dimensional organization of genomes, the major unresolved questions are be summarized here:

- Do the Random-Walk/Giant-Loop (RW/GL) and the Multi-Loop-Subcompartment (MLS) models form chromosome territories with different microscopic morphology and by which parameters can they be distinguished experimentally?
- What is the scaling behaviour of the chromatin fiber conerning the RW/GL and the MLS model since they bridge many length and times scales? Does the scaling behaviour posses multi-scaling properties or a fine structure?
- How is the mobility of particles influenced by the three-dimensional organization of the genome, i. e. is a channel like network necessary for transport as proposed by the Inter-Chromosomal Domain (ICD) hypothesis, or can particles access most nuclear loci with little obstruction as proposed by the RW/GL or MLS models?
- Can a stabel label hardly influencing the cells be developed to investigate the three-dimensional organization and dynamics of the genome *in vivo* to overcome the current experimental limitations?
- Is there a general sequential organization of genomes, i. e. are there long-range correlations in completely sequenced genomes? Does the correlation behaviour exhibit multi-scaling properties or a fine structure, is this species specific and related to their phylogenetic relationship?
- To what degree is the sequential and the three-dimensional organization of genomes connected concerning their coevolutive developement and to what view of of the nucleus does the integration of concerning results lead?



2 Simulation of Single Chromosomes

2.1 Introduction

The folding of the 30 nm chromatin fiber into chromosome territories is still a largely unresolved problem. To investigate this three-dimensional organization the Multi-Loop-Subcompartment (MLS) model, in which small loops form rosettes, connected by a linker, and the Random-Walk/Giant-Loop (RW/GL) model, in which large loops are attached to a flexible bachbone, were simulated for various loop and linker sizes. The 30 nm chromatin fiber was modelled as a polymer chain with stretching, bending and excluded volume interactions. A spherical boundary potential simulated the confinement of other chromosomes and the nucleus. Monte Carlo and Brownian Dynamics methods were applied to generate chain configurations at thermodynamical equilibrium. Both the MLS and the RW/GL model form chromosome territories with different morphologies: The MLS rosettes result in distinct subcompartments visible with light microscopy. The size of these subcompartments is in agreement with experiments. In contrast, the big RW/GL loops lead to a homogeneous chromatin distribution. Only the MLS model leads to a low overlap of chromosomes, arms and subcompartments, again in agreement with experiments. Review and comparison of experimental to simulated spatial distance measurements between genomic markers as function of their genomic separation agrees with dif-

---

Fig. 2.1 Volume Rendered Images of Simulated Chromosome Models for Chromosome XV
Random-Walk/Giant-Loop model, with 5 Mbp loop size and 378 kbp linker size (h-RW/GL, Tab 2.1) after $8x10^4$ Monte Carlo (MC) and 10^3 relaxing Brownian Dynamics (BD) steps (**A**, upper left). The large loops intermingle freely not forming distinct features like in the MLS model. Multi-Loop-Subcompartment model with 126 kbp loop size and 126 kbp linker size (B-MLS, Tab. 2.1), after $5x10^4$ MC and 10^3 relaxing BD steps (**B**, upper right). The rosettes form subcompartments as separated organizational and dynamic entities. A-RW/GL model (Tab. 2.1), loop size 126 kbp, linker size 63 kbp, after $9x10^4$ MC and 10^3 relaxing BD steps (**C**, lower left). The small loops neither intermingle freely nor form distinct subcompartments. The typical startconfiguration for simulations has the approximate form and size of a metaphase chromosome (**D**, lower right). Consecutive loops of the RW/GL or MLS rosettes are painted in red and green. The fiber diameter is 30 nm in all images.

ferent preparation conditions and favour an MLS model. Due to the presence of large spaces between the chromatin fibers, small molecules and typical proteins can reach nearly every location in the nucleus by moderately obstructed diffusion. Therefore, the assumptions of the Inter-Chromosomal Domain (ICD) model in which transport takes place in channels between territories, seem simplified. In summary, a Multi-Loop-Subcompartment like model with loop and linker sizes of 63 to 126 kbp is favoured. Additionally, the local and global characteristics of chromatin in cell nuclei are tightly inter-connected. This reveals that morphologic changes can be related to structural changes on the chromatin level.

2.2 Simulation Methods

The simulation of chromosome topologies is based on the folding of the 30 nm chromatin fiber. Although the arrangement of the nucleosomes within this fiber is still under discussion (Woodcock *et al.*, 1993; van Holde & Zlatanova, 1995; Woodcock & Horowitz, 1995; Ehrlich *et al.*, 1997; Hammermann *et al.* 2000), some properties like the mass density of 6 nucleosomes per 11 nm (e.g. ≈ 105 bp/nm, Wolffe, 1995) are known experimentally. Besides *in vitro* and electron microscopic studies (Horowitz *et al.*, 1994; Woodcock, 1994; Horowitz *et al.*, 1997; Woodcock & Horowitz, 1997; Woodcock & Horowitz, 1998; Bednar & Woodcock, 1999), neutron scattering experiments on intact nuclei of living cells indicated a 30 nm fiber diameter of 30±5 nm and that it is the dominant feature in the cell nucleus (Baudy & Bram, 1978; Baudy & Bram, 1979; Ibel, 1982; Notbohm, 1986). Here, in first approximation, the 30 nm chromatin fiber was simulated as a linear chain of segments. To model structural properties of chromosome topologies with high resolution and to reduce the necessary computer power a parallelized simulation approach joining Metropolis Monte Carlo and Brownian Dynamics methods was used. These methods allow to create chromosome configurations at thermal equilibrium.

2.2.1 Chain Properties

It was only necessary to assign three properties to the chain of segments: a stretching potential accounting for length fluctuations and numerical stability of the simulations with Brownian Dynamic algorithms (2.2.3), a bending potential controlling the bending rigidity, and an excluded volume potential keeping the chain from self crossing. A torsional potential was not introduced due to the dominant bending potential (see below), due to the existence of proteins relaxing torsional stress and to save computer power. For the stretching, a harmonic potential

$$U_s(l) = \frac{k_B T}{2\delta^2}(l - l_0)^2 \qquad (2.1)$$

with the segment length l, the equilibrium length l_0 and Boltzmann's constant k_B was used. All simulations were done at a temperature of 310K. As the stretching constant is not known exactly, δ was set to 0.1, allowing length fluctuations of ±10% and changes in the fiber diameter of ±2nm. This agrees with known variations. The same potential with l_0 ~10nm held chromatin loops (introduced by the chromosome models) together at their basepoints (2.3).

The bending potential was also harmonic:

$$U_B(\beta) = \frac{k_B T}{2\Psi^2}\beta^2 \tag{2.2}$$

with the angle between two segments β. The bending constant Ψ was obtained by estimating the Kuhn length L_K of the chromatin fiber which is related to ψ via

$$L_K = 2L_P \approx {b_0}^2\psi^2. \tag{2.3}$$

L_K is the length at which a chain of L_K long segments behaves like a Gaussian random chain (Doi & Edwards, 1986). L_K is twice the persistence length L_P, which describes the stiffness of a fiber. L_K can be calculated from experimental spatial distance measurements between genomic markers as function of their genomic separation. The dependence of the mean end-to-end spatial distance R of a chain with N segments of length L_K conducting a free and self crossing random walk can be used:

$$\langle R^2 \rangle = {L_K}^2 N^{2\nu}, \nu = 0.5. \tag{2.4}$$

Non-intersecting random walks with separated chain segments e. g. by an excluded volume potential, yield a $\nu = 0.59$. Substitution of N with the marker separation x in base pairs and expanding with the density of base pairs d results in

$$\langle R^2 \rangle = L_K\left(\frac{1}{d}d\right)L_K N = \frac{L_K}{d}x \tag{2.5}$$

Distance measurements of van den Engh (1992) and Trask (1993) allowed to determine L_K for the chromatin fiber assuming local random coil behaviour in fixed cells (2.9). L_K equals 260±70nm for the above chromatin density which is in agreement with earlier estimates of 200 to 359nm (Ostashevsky & Lange, 1994). Other distance measurements (2.9, Fig. 2.7) might lead to different results since no the distances show other than random walk behaviour. Bending force measurements of whole chromosomes (Houchmandzadeh *et al.*, 1997; Houchmandzadeh & Dimitrov, 1999; Poirier *et al.*, 2000) and computer simulations of the chromatin fiber on the nucleosomal level (Wedemann, 1999) also support these results. Thus, L_K was set to 300nm, or ≈31kbp.

To account for the forces between chromatin fibers and to keep the chain segments from self crossing, a short-ranged excluded volume potential (Fig. 5.2)

$$U_{EV}(r) = U_0 k_B T\left(1 + \frac{r^4 - 2{r_c}^2 r^2}{{r_c}^4}\right) \tag{2.6}$$

with the distance r to the axis of the chromatin fiber and the range [0, $r_c = 30\,\text{nm}$] was introduced. This potential kept chain segments at distances where van der Waals or electrostatic interactions are negligible most of the time. Variation of U_0 in the potential allows chain crossing to a certain extent. This parallels the natural process of stress reduction or disentanglement of chromatin fibers mediated by Topoisomerase-IIα and β (Gasser *et al.*, 1986; Sikorav & Jannink, 1994; Nitiss, 1998; Berger, 1998; Knopf & Waldeck, 2001). Such processes play an important role during replication or the chromosomal condensation/decondensation in mitosis (Duplantier *et al.*, 1995; Jannink *et al.*,1996). The excluded volume potential was also used to account for adjacent chromosome territories in the simulation of different nuclear volumes (2.3.4).

2.2.2 Monte Carlo Algorithm

In the first step of simulations, Monte Carlo methods were used to sample the phase space of chain configurations around thermal equilibrium at a constant temperature. Because no probability distribution creating the phase space exists a priori, so called importance sampling or Metropolis Monte Carlo (Metropolis *et al.*, 1953) was used: A given chromatin fiber configuration, according to the models used here for chromosomes (2.3), is changed by random displacements of single chain segments or groups of segments, e. g. whole loops. To reach thermal equilibrium, such a displacement must take place with a transition probability

$$P_{transition} = \frac{p_n}{p_m} = \frac{e^{\left(\frac{-H_n}{k_B T}\right)}}{e^{\left(\frac{-H_m}{k_B T}\right)}} = e^{\left(\frac{-\Delta H}{k_B T}\right)} \tag{2.7}$$

with the probability of single states p_n and p_m, the internal energy H, Boltzmann's constant k_B and the temperature T. Assuming the transition probability to be only energy dependent, a displacement is accepted if the internal energy of the whole chain $\Delta H = H_n - H_m$ decreased or if $P_{transition}$ is higher than a random number from the interval [0,1]. This Markovian process leads to a sampling by "systematical" exploration of the phase space if $P_{transition}$ is normalized, if there exists a positive $P_{transition}$ for every two points of phase space (ergodicity) and if a state of phase space leads to another only by applying the corresponding $P_{transition}$, assuring the equilibrium of a certain state.

The following displacement moves were used (Fig. 2.2 A-F): Local changes were performed by random translation of one segment and rotation by a random angle of one segment around the axis between the start of this and the end of the following segment (Verdier & Stockmayer, 1962; Baumgärtner & Binder, 1979). Loops of segments were rotated randomly around the loop base using a random vector from the unit sphere. To relax the relative positions of rosettes of the MLS model or loops of the RWGL model faster (2.3), a collective move of two neighbouring

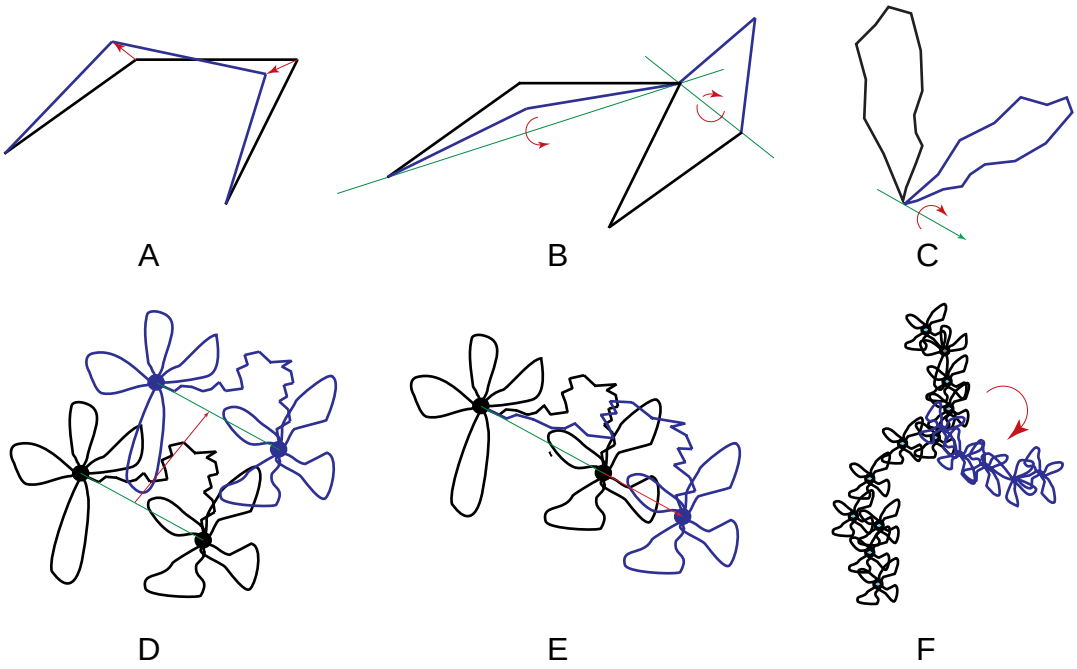


Fig. 2.2 Monte Carlo Moves on Various Length Scales
(**A**) Random displacement of one segment. (**B**) Random rotation of one segment around the baseline of the two adjectant segments. (**C**) Random rotation of a loop around its base. (**D**) Random displacement of the linker with two rosettes. (**E**) Random stretching of a linker along the connection between the two connected rosettes. (**F**) Random turn of on part of the chain of rosettes against the rest.

rosettes and stretching of the connecting linker was applied. Global repositioning was achieved by turning all the segments of one part of the chain around one segment whereas the other part of the chain remained unchanged (so called pivot move; Freire & Horta, 1976; Meal & Sokal, 1988).

During one Monte Carlo step 3% of the segments and loops were displaced, whereas global repositionings were done only once. The energy calculation used the potentials described above. The translation sizes and rotation angles were set automatically such that the acceptance rate of the displacements yielded 40 to 60%. To avoid stress in parts of the fiber configuration, each Monte Carlo step was followed by a relaxing Brownian Dynamics step (2.2.3). Up to $6\text{x}10^5$ Monte Carlo steps were necessary for adequate sampling and achieving a sufficient number of statistically independent configurations for analysis. The independence of configurations was monitored by comparing every 1000th configuration by visual inspection and by correlation analysis of the radius of gyration. The latter is one of the slowest relaxing parameters since it describes a single chromosome on the largest possible scale.

To speed up the sampling by the Monte Carlo algorithm, L_K was first set to 300 nm corresponding to a resolution of ≈31,000 bp. In this case, a 126 kbp loop in the MLS model consists of only four segments and loops of 63 kbp cannot be modelled adequately anymore without reducing L_K to 200 or 100 nm.

2.2.3 Brownian Dynamics Algorithm

To introduce a higher resolution and thereby further relax the Monte Carlo configurations Brownian Dynamics was used. In contrast to Monte Carlo methods (2.2.2), Brownian Dynamics describes the real dynamics caused by thermal fluctuations of particles which are far bigger than the particles in the surrounding fluid medium. The constraints created by the molecular impacts causing the big particles to diffuse can be described by pure statistics for observation time scales (scale of diffusion) far bigger than the correlation time of the impacts. Consequently, the detailed description of every impact can be replaced by a stochastic interaction.

A system of N particles can be described in Brownian Dynamics by the Langevin equation

$$m_i \frac{\partial^2 r_i}{\partial t^2} = \sum_j \mu_{ij} \frac{\partial r_{ij}}{\partial t} + F_i + \sum_j a_{ij} f_j \tag{2.8}$$

with the N particle coordinates r_i, the particle masses m_i, the sum of internal and external forces F_i, the stochastic forces $a_{ij} f_j$ caused by thermal fluctuations and indices i, j $(1 < i, j \leq 3N)$. For a classical Brownian system, f_i are Gaussian distributed with moments

$$\langle f_i \rangle = 0, \text{ and } \langle f_i(t) f_j(t) \rangle = 2D_i \delta_{ij} \delta(t - t'). \tag{2.9}$$

with the diffusion coefficients D_i. The a_{ij} are connected to the friction tensor via

$$\mu_{ij} = \frac{1}{k_B T} \sum_l a_{il} a_{jl} \tag{2.10}$$

where μ_{ij} is in turn proportional to the inverse diffusion tensor D_{ij}:

$$\sum_j \mu_{ij} D_{jl} = \sum_j D_{ij} \mu_{jl} = k_B T \delta_{il} \tag{2.11}$$

with Boltzmann's constant k_b and the temperature T. In first approximation the hydrodynamics of a chain polymer can be described by spheres with radius σ at position $\boldsymbol{r}_i$ (e. g. Chirico & Langowski, 1994). In this case, the tensor of diffusion can be formulated as the Rotne-Prager tensor using the viscosity η:

$$D_{ij} = \delta_{ij} \frac{k_B T}{6\pi\eta\sigma} - (\delta_{ij} - 1) \frac{k_B T}{8\pi\eta r_{ij}^2} \left[\left(1 + \frac{r_{ij} r_{ij}}{r_{ij}^2} \right) + \frac{2\sigma^2}{r_{ij}^2} \left(\frac{1}{3} - \frac{r_{ij} r_{ij}}{r_{ij}^2} \right) \right]. \tag{2.12}$$

Since the momentum distribution relaxes far faster than the equilibrium distribution of the position distribution of particles (Ermak & McCammon, 1978) Equ. 2.8 can be reformulated as

$$r_i(t + \Delta t) = r_i(t) + \sum_j \frac{\partial D_{ij}}{\partial r_j} \Delta t + \sum_j \frac{\partial D_{ij} F_{ij}}{\partial k_B T} \Delta t + R_i(\Delta t) \tag{2.13}$$

with stochastic displacements R_i, being again Gaussian distributed with moments:

$$\langle R_i(\Delta t) \rangle = 0, \langle R_i(\Delta t) R_j(\Delta t) \rangle = 2D_{ij} \Delta t. \tag{2.14}$$

For spherical beads (Iniesta & Garcia de la Torre, 1990) Equ. 2.13 can be transformed into

$$\boldsymbol{r}_i(t + \Delta t) = \boldsymbol{r}_i(t) + \frac{D}{k_B T}\boldsymbol{F}_i \Delta t + \boldsymbol{R}_i(\Delta t) \tag{2.15}$$

where the diffusion constant D for spheres with radius a is due to Stokes:

$$D = \frac{k_B T}{6\pi\eta a}. \tag{2.16}$$

As mentioned in 2.2.2, loops with four stiff segments do form more an edgy rhombus than a smooth loop. To increase the resolution, around 100 to 150 uncorrelated configurations were taken from the ~6x10^5 Monte Carlo configurations and the segment size was reduced to 50 nm. Each configuration was then relaxed with 1000 or 2000 Brownian Dynamics steps according to the method described above. Forces used were obtained from the derivative of the potentials introduced above (2.2.1).

2.2.4 Parallelization of Simulation Code

The simulation code itself was parallelized for two reasons: First, a typical chromosome such as chromosome XV is very large and consists of ≈3,500 segments of 300 nm or 21,000 segments of 50 nm length. Second, the complexity and dominance of the pairwise excluded volume interaction is computation intensive. For parallelization the very efficient linked cell algorithm was applied. It divides the three-dimensional space into cubes into which the chain segments are mapped (Allen & Tildesley, 1989). The size of the cubes was chosen such that only interactions on the scale of less than two sidelengths produced non-negligible contributions. Consequently, only interactions between chain segments in the same cell and to half of the 26 neighbour cells had to be considered. This three-dimensional grid of cubes was divided and mapped to a two-dimensional processor grid. Since whole chromosomes do not fill the three-dimensional grid of cubes homogeneously, the load of the different processors was equalized by changing the assignment between the cubes and processor grid every tenth Monte Carlo or every hundredth Brownian Dynamics step (Münkel & Langowski, 1998). With this so called dynamic load balancing a parallelization efficiency of 84% was reached for a simulation on 16 processors.
The object oriented simulation code (VirtNuc), various helper programs (ChromCreate which creates starting configurations, KuhnToBending which reduces segment sizes and allows changes in parameters) and analysis code (Geometry for general analysis, FracTAK for various fractal analysis, VirtMic for visualizations) were written in C++ (Fig. 2.3). Parallelization used the message passing interface (MPI) standard. The simulations of single chromosomes presented here, total about 96,000 CPUh (~11 years) on a single R6000 processor with 60 MHz.

2.3 Simulated Models and Their Properties

To evaluate the different proposed models as well as to generate statistically significant data for comparison with experiments, the mean sized chromosome XV with a base pair content of 106 Mbp was simulated.

2.3.1 Random-Walk/Giant-Loop (RW/GL) Model

In the RW/GL model, big loops are assumed to be attached to a non-DNA backbone with no closed loop base (Sachs *et al.*, 1995; (1.3.7), Fig. 1.14). Here the RW/GL model was simulated with loops attached at basepoints, which are connected by a linker. The loop size was varied between 5 Mbp and 126 kbp, thus the number of loops changed from 20 to 561 (Tab. 2.1). For loops smaller than 500 kbp the RW/GL term 'giant' seems, inappropriate and is more similar to a Pienta & Coffey (1984) like model of interphase organization (1.3.4, Fig. 1.11 A&B). The mean extension of a chromosome territory depends on the total backbone/linker length which itself conducts a random walk (Equ. 2.4). To keep the mean extension comparable between the different models, the linker size had to be inversely proportional to the loop size and number. Dividing the total number of linker segments by the number of loops minus one, results, however, in broken segment numbers. To reduce the computer power only non-broken linker segments with a Kuhn length of L_K = 300 nm could be used in general (Tab. 2.1). A real flexible linker required a minimum of two segments. For a loop size of 126 kbp and thus 561 loops this results in an average chromosome extension of 10 µm, thus spanning a whole nucleus. Since chromosomes do not extend through the whole nucleus (1.3.4), such an RW/GL topology was also simulated with linkers of three 40 nm long segments from the beginning of the simulation resulting in a mean extension of 5.9µm (Tab. 2.1).

2.3.2 Multi-Loop-Subcompartment (MLS) Model

In the MLS model, small loops are forming rosettes linked by a piece of chromatin fiber (1.3.8, Fig. 1.14). The rosettes form substructures of a chromosome like those found in interphase studies on transcription and replication (Berezney *et al.*, 1995; Zink & Cremer, 1998; Zink *et al.* 1998, Fig. 1.14). In metaphase chromosomes these rosettes could be related to the ideogram banding pattern (Pienta & Coffey, 1984; Laemmli, 1994). The loop and linker size was varied from 63 to 252 kbp (Tab. 2.1). The loop number in one rosette is proportional to the total DNA content of a rosette divided by the loop size. To simulate different DNA contents of the rosettes, which is the case in real chromosomes, the 850 band metaphase ideogram banding pattern of Francke (1994, Fig.1 10) was used. Each metaphase band was divided by three, since during decondensation into interphase the bands split up into ~2550 to 3000 (Yunis, 1981). Since the DNA in a linker is taken from the DNA con-

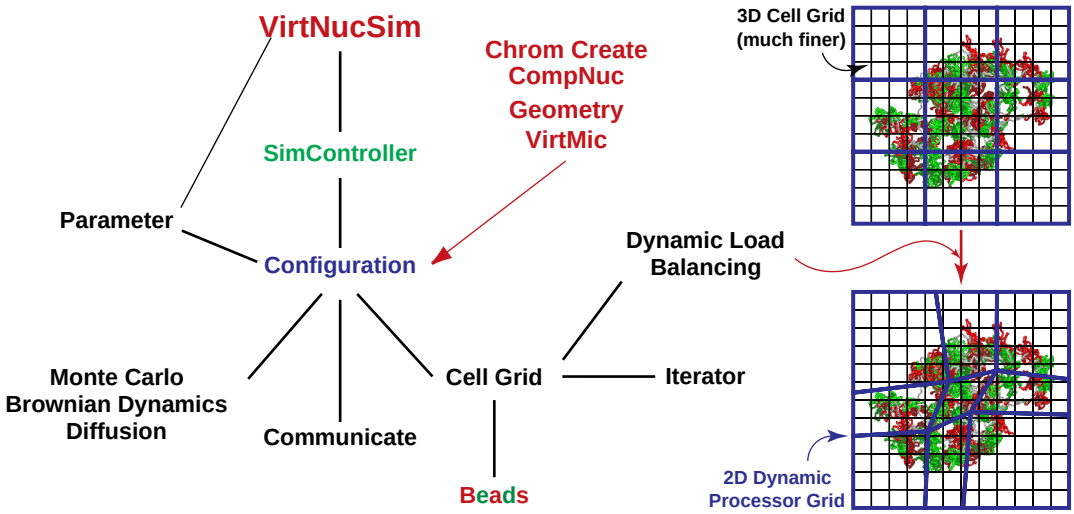


Fig. 2.3 Structure of the Simulation Programmes and Dynamic Load Balancing
Programs were object oriented using C++. The main program ***VirtNucSim*** is connected to the general controller *SimController* accessing the class *Configuration*, which encapsulates all instructions concerning the input/output of Parameters, configurations and the classes for Monte Carlo, Brownian Dynamics and Diffusion calculations. It also channels the grid of cells containing the polymer segments with the class *CellGrid*, which accesses (e. g. to calculate energies) the segment describing class *Bead* via iteration by the class *Iterator* and connects to the class *LoadBalancing* distributing the cells to a 2D processor grid such that equal processor loads are achieved. Communication between processors uses the Message Passing Interphase (MPI) standard organized in the class *Communicate*. Several 'helper' programmes are connected to the class *Configuration* from outside ***VirtNucSim***: ***ChromCreate*** creates starting configuration, ***CompNuc*** creates whole nuclei, ***Geometry*** analyses the simulations and VirtMic visualizes the simulations as a virtual microscope.

tent of the single bands (and therefore rosettes), the loop number of rosettes is also inversely proportional to the linker size (Tab. 2.1). For MLS models with small loops, the segment length was reduced to 200 or 100 nm as mentioned in 2.2.2 (Tab. 2.1). The base pair content of the linker was taken in equal parts from the neighbouring rosettes. Rounding effects which could have resulted in a loss of base pairs were avoided by adding loops to rosettes with small loop numbers.

2.3.3 Starting Configurations

As a starting configuration, the rosettes of the MLS model were stacked on top of each other. This approximated the natural shape and size of metaphase chromosomes (Pienta & Coffey, 1984; Fig. 2.1D). The linker between the rosettes was first put into the rosettes as an additional loop. Opening of its base led to decondensation into interphase. For most simulations of the RW/GL model, the big loops were also folded into rosettes, and opening of the loop basepoint allowed the decondensation into giant loops and interphase. The linker between loops was set as in the MLS model. Other starting configurations like prefolding of the giant loops into random

walks or along the axis of the metaphase chromosome were tested extensively (Knoch, 1998). They only influenced the relaxation speed into interphase by Monte Carlo simulations, but had no effect on results.

2.3.4 Excluded Volume and Nuclear Volume Properties

To test the influence of the excluded volume interaction on chromosome properties all simulations of Tab. 2.1 were performed with a low (U_0=0.1kT) and high (U_0=1.0kT) probability for chain crossing. These probabilities could reflect the activity of chain crossing mediated by Topoisomerase-IIα and β (Gasser *et al.*, 1986; Sikorav & Jannink, 1994; Duplantier *et al.*, 1995; Jannink *et al.*,1996; Nitiss, 1998; Berger, 1998; Knopf & Waldeck, 2001). A high excluded volume potential also influences the interaction between chromosome territories in a whole cell nucleus: how much volume there is available, how sharp the territories are separated or how much they can intermingle/invaginate into each other. Therefore, the single chromosomes were put into a spherical potential representing the surrounding chromosomes. The volume fraction of a chromosome in the nucleus equals the fraction of the chromosomal DNA content of the genome (Monier et al. 2000). In a spherical nucleus with a radius of 5μm, as typically used in experiments, this results for chromosome XV in a spherical potential with 1250nm radius, i. e. a volume of 8.2μm^3. The height of the potential was directly related to the height of the excluded volume potential (Fig. 5.2). Consequently, chromosomes with low excluded volume interaction could fold almost freely into chromosome territories with an average extension depending on the linker size and not on the embedding volume. In contrast, chromosomes with a high excluded volume potential were strictly confined to a spherical shape with a corresponding average extension. The influence of nuclear volume changes was also tested for the B-MLS model (Tab. 2.1) within spherical potentials of 970 and 1400nm radius (4.1 and 12.0μm^3) as well as a high spherical potential and a low excluded volume potential between the chain segments, which was used to speed up the simulations.

2.4 Morphology of Simulated Chromosomes

The folding of the 30nm chromatin fiber proposed by the RW/GL and the MLS models leads to chromosome territories with different morphologies (Fig. 2.1).

Only the MLS model forms distinct chromosome territories with a ‘sharp’ edge in agreement with experiments (Zirbel *et al.*, 1993; Eils *et al.*, 1995). The rosettes form visible compartments within a territory (Fig. 2.1B), and despite being sometimes very close (especially for small loops or low loop numbers), loops from different rosettes do not stretch significantly into other subcompartments. The rosette

Tab. 2.1 Simulated Chromosome Models with Their Physical Properties
The band number are the number of subcompartments or loops per chromosome for the corresponding chromosome model. The mean band distance is $\langle R_B \rangle = \sqrt{(300nm)^2 \cdot LI_L/(300nm)}$ and the mean territory size is $\langle R_{Ltotal} \rangle = \sqrt{(300nm)^2 \cdot (NB-1) \cdot LI_L/(300nm)}$ (see Equ. 2.4).

Model	Loop properties		Linker properties		Band properties		Mean band distance		Mean territory size	
	Size L_S [Mbp]	Length L_L [µm]	Size LI_S [Mbp]	Length LI_L [µm]	# of bands NB	# of loops per rosette	Theoretic $< R_B >$ [nm]	Simulated $< R_B >$ [nm]	Theoretic $< R_{Ltotal} >$ [µm]	Simulated $< R_{Ltotal} >$ [µm]
A-MLS	0.126	1.2	0.063	0.6	96	7.9±4.3	424	520	3.13	4.0
B-MLS	0.126	1.2	0.126	1.2	96	7.5±4.3	600	630	5.85	6.1
C-MLS	0.126	1.2	0.189	1.8	96	7.1±4.3	734	720	7.16	7.3
D-MLS	0.126	1.2	0.252	2.4	96	6.5±4.3	848	870	8.27	8.4
E-MLS	0.063	0.6	0.126	1.2	96	16.1±9.2	600	610	5.85	5.8
F-MLS	0.084	0.8	0.126	1.2	96	11.8±6.8	600	615	5.85	5.9
G-MLS	0.105	1.0	0.126	1.2	96	9.2±5.2	600	622	5.85	6.0
H-MLS	0.158	1.5	0.126	1.2	96	6.1±3.6	600	625	5.85	5.9
I-MLS	0.252	2.4	0.126	1.2	96	3.8±2.0	600	610	5.85	5.8
a-RW/GL	0.126	1.2	0.063	0.6	561	-	424	450	10.0	9.7
a'RW/GL	0.126	1.2	0.013	0.12	561	-	193	200	5.67	5.9
b-RW/GL	0.252	2.4	0.063	0.6	338	-	424	440	7.79	7.3
c-RW/GL	0.504	4.8	0.063	0.6	187	-	424	439	5.79	5.6
d-RW/GL	1.0	10	0.126	1.2	94	-	600	620	5.82	6.3
e-RW/GL	2.0	20	0.189	1.8	48	-	734	740	5.09	5.8
f-RW/GL	3.0	30	0.252	2.4	33	-	848	840	4.80	5.6
g-RW/GL	4.0	40	0.315	3.0	25	-	948	960	4.64	6.4
h-RW/GL	5.0	50	0.378	3.6	20	-	1000	1010	4.53	7.0

overlap is under 10% for an MLS model with 126kbp loops and linkers (Münkel & Langowski, 1998; Münkel *et al.*, 1999). Qualitatively, the visible overlap is proportional to the loop size and inversely proportional to the number of rosette loops. Since the resolution limit of light microscopy is smaller than the diameter of subcompartments, it is possible to identify them experimentally as globular structures (2.6.1), despite the invisibility of the single chromatin fiber. Chromosome territories in the RW/GL model lack a sharp border: The bigger the loops, the more they intermingle freely, overlap and loop out of the territory (Fig. 2.1A). Size dependent, loops can span whole chromosome territories (2.7). Small 'giant' loops of 126kbp do not overlap like big ones, but the chain of loops as a whole intermingles. Loop size independent, there are no substructures within the territory (Fig. 2.1A&C) resulting in a smooth mass distribution contrasting the globular MLS model.

A very surprising feature visible in the images is the existence of big unoccupied spaces within the chromosome territories, which allow high accessability to the interior of the territory. Consequently, territories do not have a closed surface. On length scales <10nm the chromatin fiber itself is the "surface" of a chromosome. The big

subcompartment sized voids in the MLS (Fig. 2.1C) or an outlooping big loop or chain of loops in the RW/GL model (Fig. 2.1A&C) are naturally occupied by other chromosomes in a nucleus. This agrees with the simulation of whole cell nuclei (Chapter 3). Nevertheless, the mean spacing between chromatin fibers ranges obviously from 50 to 100 nm agreeing with theoretic estimates also predicting a chromatin volume occupation of 4.4 to 9% in a 5 µm nucleus and a mean fiber distance of 63 to 90 nm. The structural overlaps and these volume relationships depend, of cours, on the available nuclear volume (Chapter 3&5). In summary, this morphologic analysis suggests that small molecules can reach every chromosomal location and that the diffusion of particles is only moderately obstructed by chromatin. This agrees with experiments and simulation of particle diffusion (Chapter 5).

2.5 Radial Mass and Density Distribution of Chromosome Territories

For the analysis of average properties of chromosomes on the scale of a whole chromosome territory, the radial mass and density distribution were calculated. The radial mass distribution as function of the radius was calculated as the number of segments in spherical shells (width 5 nm) centered at the chromosomal center of mass. The radial density is the number of segments in a shell divided by the volume of this shell. The mean was taken over 100 to 150 configurations.

For MLS models A-D with the same loop size of 126 kbp the plateau in the radial density up to a radius of 1.0 µm and the peak height of the radial mass distribution are inversely proportional, and the mean extension of the radial mass distribution is proportional to the linker size between the subcompartments (Fig. 2.4A&B). Changes of the loop size did only affect this behaviour implicitly, due to the influence on the distance between succeeding subcompartments (2.6.2). Due to the low barrier of the surrounding spherical potential with radius 1.2 µm, the mean radius and extension of a territory is only restricted by the linker between the subcompartments. The mean radial extension of a territory was the extension at half the maximum peak height of the radial mass distribution. It can be calculated theoretically with Equ. 2.4 assuming that the linker between rosettes conducts a random walk. Simulation and theoretical values are in good agreement (Tab. 2.1). The existence of a plateau can therefore only be explained by the interplay between the existence of an excluded volume interaction (although it might be low), the repulsive entropy of fast fluctuating rosette loops and the linker pulling the rosettes together.

Surprisingly, the radial density of the RW/GL models a-h show only rudimentary if any plateau and a much smoother density decrease with increased distance to the center of mass (Fig. 4.2C). The radial density and the peak height of the radial mass distribution are again inversely proportional, and the mean extension of the radial mass distribution is proportional to the linker size and in addition the loop size

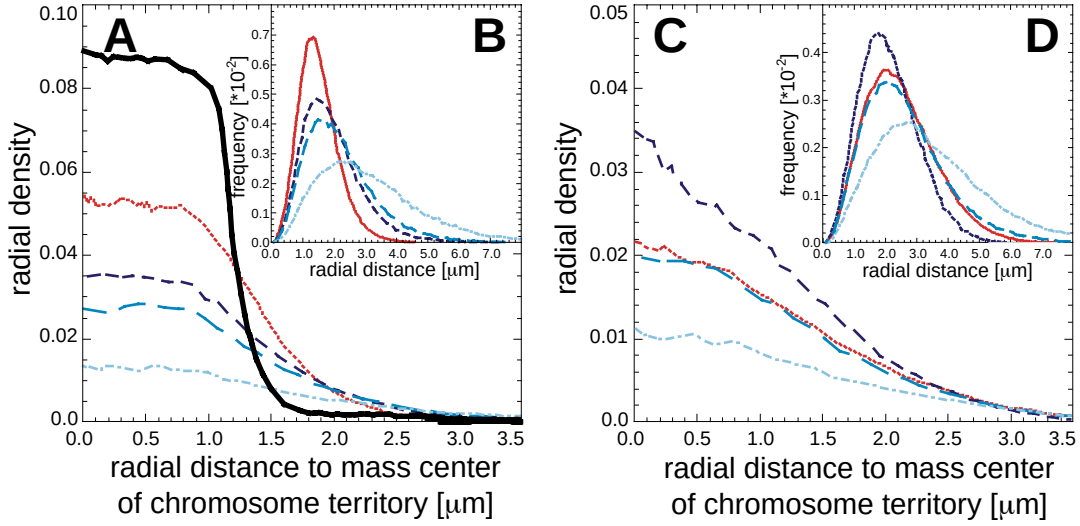


Fig. 2.4 Radial Density and Radial Mass Distribution of Chromosomes
The radial density (**A**) and mass (**B**) distributions of A-D MLS models (Tab. 2.1) with 126 kbp loop size are proportional to the linker size (A-MLS: dotted; B-MLS: dashed; C-MLS: long dashed; D-MLS: dash dotted; B-MLS with high excluded volume and high embedding volume potential: solid). The radial density (**C**) and mass (**D**) distributions of a-, b-, e-, h-RW/GL models (Tab. 2.1) are proportional to the linker length and proportional to the loop size but show no density plateau as in (**A**) (a-RW/GL: dash dotted, b-RW/GL: long dashed, e-RW/GL: dashed, h-RW/GL: dotted).

(Fig. 4.2C&D). Thus, the territory extension depends on both the linker and loop size in contrast to the MLS model (see above). As total linker length was nearly kept constant (2.3.1), variations from the theoretical calculation of the mean territory extension are smaller than expected. The slower decrease of the radial density due to the big loops, however, results in higher average extensions than predicted theoretically (Tab. 2.1). The reason for the slow radial density decrease and the rudimentary plateau is caused by the big intermingling loops connected with linkers relatively small compared to their loop size. Thus, the loops are more densely packed near the linkers while stretching out further than the linkers extent. This is in agreement with the explanation of the plateau for A-D MLS models.

2.6 Properties of MLS - Subcompartments

2.6.1 Radial Mass and Density Distribution

The rosettes of the MLS model form distinct subcompartments with a diameter larger than the resolution limit of light microscopy (Fig. 2.1B). The radial mass and density distribution were computed as for the whole chromosomes (2.5) with a shell width of 1 nm for higher resolution. The mean was taken over all subcompartments

of one chromosome and averaged over 100 to 150 configurations. It should be noted that the varying size of ideogram bands leads to different number of loops in the rosettes, since the loop size was kept constant in all rosettes (2.3.2).

For MLS models E-H with linker sizes of 126 kbp the peak height of the radial mass distribution is inversely proportional and the radial extension of the radial mass distribution is proportional to the loop size (Fig. 2.5 A). The mean radial extension of a subcompartment is half the position of the maximum of the radial mass distribution. Loop sizes of 63, 84, 105, 126, 158 and 252 kbp result in mean rosette radii of 195, 210, 250, 300, 350 and 430 nm, respectively. This leads to mean diameters of 390, 420, 500, 600, 700 and 860 nm. This agrees with the extension of single loops measured by simulated position dependent spatial distance measurements between genomic markers as function of their genomic distance (2.7.2). The values determined her reflect the average of the loop size distribution in real chromosomes. Only minor but systematic effects resulting from the inverse proportionality between loop numbers and linkers were observed (Fig 2.5 B). For comparison with microscopic experiments, a total possible extension of a rosette was estimated for the radial distance for which the histogram frequency and therefore mass probability drops below 10 %. This resulted in radii of 255, 285, 340, 380, 450 and 560 nm and diameters of 510, 570, 680, 760, 900 and 1120 nm. These values might reflect an upper limit for the comparison with experimental diameters which depend on preparation, image reconstruction and the threshold dependencies for volume determinations. Experimental data from BrdU incorporation reveal a diameter distribution of so called foci with from 400 to 700 nm (Berezney *et al.*, 1995; Zink *et al.*, 1998; Bornfleth, 1999). An MLS model with loop sizes of 63 to 126 kbp is favoured considering the image reconstruction methods used.

2.6.2 Spatial Distance between Succeeding Subcompartments

The distance between succeeding subcompartments depends on the interaction between the length of the linker, the diameter and loop size of the rosettes. The distance between the center of mass of each pair of succeeding subcompartments was calculated for a chromosome and averaged over 100 to 150 chromosome configurations. (The results were obtained with high excluded volume interaction 2.3.4).

In all MLS models the mean distance between succeeding subcompartments is proportional to the length of the linker (Fig. 2.5 C). Linker length of 63, 126, 189 and 252 kbp result for models A-D with loop sizes of 126 kbp in a mean distance between succeeding subcompartments of 520, 630, 720 and 870 nm. With Equ. 2.4 the theoretical values are 424, 600, 734 and 848 nm (Tab. 2.1). These differ from the observed values by 22.5, 5.0, 3.1 and 2.5 %. The effect has with a maximum for loops of around 126 kbp and vanishes the smaller and the bigger the loops are (Tab. 2.1). The discrepancy between theoretic and observed distance using a linker of 63 kbp and loops of 126 kbp, can be explained with an entropic repulsion of the rosettes due to their mean extension of 300 nm or maximum diameter of 760 nm

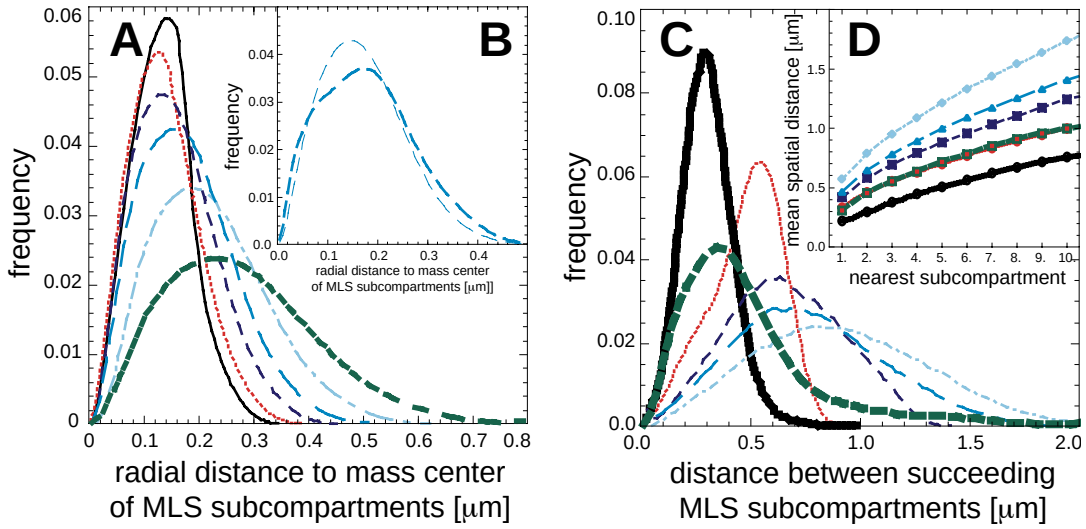


Fig. 2.5 Radial Mass Distribution and Succeeding/Nearest Distance of Subcompartments
(**A**) The extension of the radial mass distribution as function of the radial distance to the mass center of MLS subcompartments is proportional to the loop size in B and E-I MLS models with the same linker size of 126 kbp and high excluded volume interaction between the chain segments (E-MLS solid, F-MLS dotted, G-MLS long dashed, B-MLS short dashed, H-MLS long dash dotted, I-MLS short dash dotted). (**B**) Comparison of the radial mass distribution as in (**A**) of the B-MLS model reveals a wider distribution for low (thin dashed) than for high (thick dashed) excluded volume interaction. The distribution also depends on the loop numbers in rosettes due to the loop size (red region of variation). (**C**) The spatial distance distribution between the mass center of succeeding MLS subcompartments of A-D MLS models is proportional to the linker length for relaxed configurations (A-MLS dotted, B-MLS dashed, C-MLS long dashed, D-MLS dash dotted). After sudden density increase and insufficient relaxation the mean spatial distance decreases temporarily (A-MLS thick dotted, B-MLS thick dashed). (**D**) The mean spatial distance to the first, second etc. spatially nearest subcompartment of a subcompartment shows the same proportionalities as in **C**.

(2.6.1). This is supported by the asymmetric distance distribution shifted to higher distances (Fig. 2.5A). The mean distance of 520 nm also suggests, that the repulsion does not lead to really distinct subcompartments and that the loops arrange themselves like teeth of gear wheels (see also the morphology of very small nuclei in Chapter 3). This entropic repulsion supports also the increase of the segment length to 50 nm from an initial Kuhn length L_K of 300 nm (2.3.2). Additionally, it shows the shortcomings of reducing the subcompartments to mere spheres in simulations of whole cell nuclei. The distance of succeeding RW/GL loops models agree very well with the theoretic prediction, although the 13 kbp linkers are slightly stretched (a' RW/GL model, Tab. 2.1).

2.6.3 Spatial Distance between Nearest Subcompartments

The distance between the nearest neighbours of subcompartments depends not only on the interaction between the length of the linker, the diameter and loops size of the rosettes, but also on the embedding volume and therefore the nuclear density. The

distance to the spatially nearest, the second nearest etc. subcompartment, was calculated from the distances between the center of mass of one subcompartment to all the other subcompartments and sorting of these distances. The average was taken of 100 to 150 chromosome configurations.

Similar to the distance between succeeding subcompartments the nearest neighbour distances for the A-D MLS models with loop sizes of 126 kbp are proportional to the linker length (Fig. 2.5D). Linkers of 63, 126, 189 and 252 kbp resulted in a mean nearest neighbour distances of 355, 424, 467 and 573 nm. Compared to the mean separation of succeeding subcompartments these values are 185, 206, 276 and 275 nm smaller. For the third nearest subcompartment the values are much the same as for succeeding subcompartments. Since the nearest subcompartments are mostly not connected directly through a linker but by an arbitrary genomic separation, the entropic repulsion has here a much smaller effect than for succeeding subcompartments. The latter is due to the constraining linker, which forces the succeeding subcompartments to interact intensively with each other, while arbitrary subcompartments might arrange due to favourable energetic minima. Therefore, the nearest neighbour distance could well be smaller than the mean separation of succeeding subcompartments.

2.7 Spatial Distances between Genomic Markers

Measuring spatial distances between genomic markers as function of their genomic separation corresponds to different assumptions of the structure, stability and dynamics of chromosome organizations.

2.7.1 Position Independent Spatial Distances

Position independent measurements of spatial distances reflect chromosome models where loops can form at sequence independent positions. For variable loops in a fixed rosette (i. e. a rosette with constant base pair content and defined genomic position) this could either mean a chromatin fiber slithering through the “bases” of the loops. These distances reflect also possible variabilities of loop arrangements or whole of chromosome organizations between different cells, before or after cell division of one cell, cell cycle dependencies and even random experimental preparation artefacts disturbing or destroying the *in vivo* organization. Consequently, a marker pair could reside in any possible position relative to a loop base or rosette. Thus, the spatial distance is position independent.

For calculation, pairs of markers were placed randomly (e. g. regardless of any folding structures) on the chromosome (Fig. 2.6A) with genomic separations from 5.2 kbp (the base pair content of one chain segment) up to the whole chromosome.

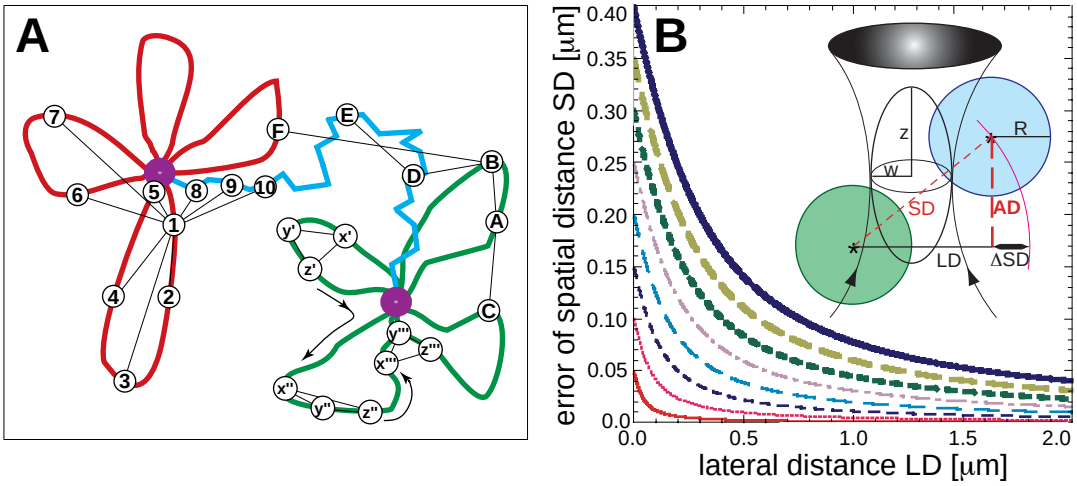


Fig. 2.6 Measuring Procedures for Spatial Distances and Error in Epifluorescence Microscopy
Various measuring procedures for spatial distances reflect structural properties of chromosomes (**A**): For position independent spatial distances randomly positioned markers could reside both on the same loop (A-B), on different loops (A-C), on a loop and in a linker (B-D), both in the linker (D-E) or on loops belonging to different rosettes (B-F). For position dependent measurements one marker is positioned (1) and the spatial distance measured successively to markers in the rosette (1to 2 through 7) and then to the linker (1-8 through 10). Dependencies between position dependent spatial distances occur if a set of genomic markers is shifted through a rosette (x'y'z' to x"y"z" etc.). A spatial distance between a marker pair in cis-position of loops is (A-C), a trans-position is (B-x"). (**B**) Error of the spatial distance ΔSD as function of the lateral distance LD, the axial separation AD of FISH signals with radius R, an objective with lateral resolution W and axial resolution Z (Equ. 2.17). In an epifluorescence microscope the signal would appear in the same focal plane, thus for lateral distances below 500nm the real three-dimensional distance is underestimated significantly by at least 50 to 100nm (AD = 50 nm: solid, 100nm: dotted, 150nm: dashed, 200nm: long dashed, 250nm: dash dotted, 300nm: thick dashed, 350nm: thick long dashed, 400nm: thick solid).

Since chromosome XV contains 21,000 chain segments, it was impossible to calculate all distances between all marker pairs. Thus, for genomic separations >25 Mbp, 5000 pairs were taken. Below 25 Mbp it was possible to calculate the distances between all marker pairs. The mean was taken over the number of pairs for one genomic separation and over 100 to 150 chromosome configurations.

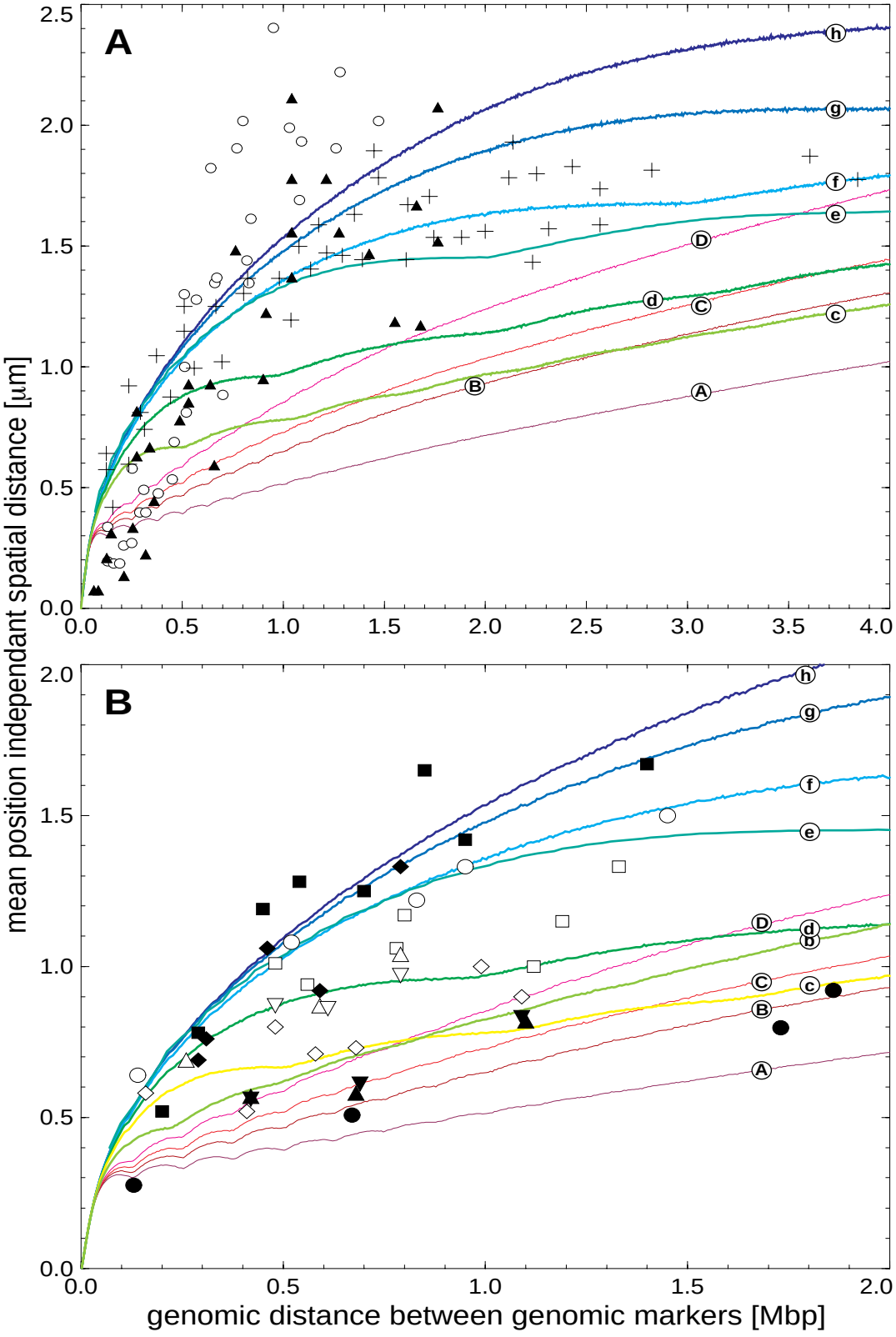
For RW/GL as well as for MLS models the general properties of the spatial distances between position independently placed markers as function of their genomic separation are equal (Fig. 2.7, Fig. 2.8): The distance first increases monotonously as expected from a random walk. At genomic separations of half the loop size the increase stagnates in a plateau and shows a local minimum at a genomic separation of one loop size. The spatial distance of the plateau is proportional to the loop size (Fig. 2.7, Fig. 2.8). Due to the random positioning of the marker pair the distance need not decrease to zero: E. g. the distance of a marker pair with a genomic separation of 126kbp in an MLS rosette with loop sizes of 126kbp is zero for markers positioned on a loop base. For markers located at the tip of two successive loops, which point in opposite directions, the distance could be twice the extension of the

loops. The spatial distance of the plateau is somewhat larger than the mean extension of a loop (Fig. 2.9) since also spatial distances of linkers contribute to the average. Therefore, the top of the plateau also shifts slightly to higher genomic separations. Inspection of the distance distributions (Fig. 3.9C), reveals further details due to a fixed MLS model and similar to the analyses of marker sets (2.7.3).

For genomic separations greater than one loop size the increase of the spatial distance is proportional to the linker size between loops of the RW/GL model (Fig. 2.4D) and rosettes of the MLS models A-D (Fig. 2.4B, Fig. 2.5C). Since the linker is mainly responsible for the mean extension of territories, MLS models E-H with different loop sizes but the same linker size, show the same spatial distance behaviour for genomic separations exceeding the mean MLS subcompartment size of ~1 Mbp (Fig. 2.7). For RW/GL models the same principle holds, although the genomic separation for which the spatial distance is not proportional to the loop size anymore but only proportional to the linker size depends itself on the loop size. For the chosen extension of a chromosome territory of 5µm this transition varies from ~4Mbp for the c-RW/GL model with 0.5Mbp loops (Fig. 2.7A) to ~15Mbp for the h-RW/GL model with 5.0Mbp loops.

Consequently, the differentiation between the RW/GL and the MLS model in the experimental spatial distance measurements depend on the loop and linker size between loops or rosettes. Assuming the current available resolution of experimental measurements (2.9), a C-MLS model with loops of 126kbp and large linkers of 256kbp can be distinguished very well from a f-RW/GL model with 3.0Mbp loops up to a genomic separation of 3Mbp (Fig. 2.7A). With a 126kbp linker, a B-MLS model is well distinguishable from a c-RWGL model with 0.5Mbp loops up to a genomic separations of 1.3Mbp. Experiments with good preparation conditions and statistics should even distinguish between a b-RW/GL model with loops of 250kbp and A-D MLS models with 126kbp loops below genomic separations of 0.5Mbp. An I-MLS model with 256kbp loops and 126kbp linkers can still be distinguished well from a c-RW/GL model with 0.5Mbp loops up to a genomic separation of 1.5Mbp (Fig. 2.7A, Fig. 2.8). In summary, RW/GL and MLS models can be distinguished with position independent spatial distance measurements for a wide range of model parameters under conditions assuming no fixed loop and/or rosette sizes.

Fig. 2.7 Simulated Position Independent Spatial Distances Compared to Experiment - I
Spatial distances between position independent placed genomic markers as function of their genomic distance and comparison to experiments. Lines are simulated and are the same in Fig. 2.7A&B and Fig. 2.8A&B, symbols are experimental values. For details of model name and properties see Tab. 2.1 and for interpretation of experimental values see Tab. 2.2. Thin lines: From bottom to top A-D MLS-models with 126 kbp loops size and varying linker sizes. Thick lines: RW/GL-models with different loop sizes a-RW/GL to h-RW/GL. (**A**) open circles: van den Engh ’92; full triangles: Trask ’93; pluses: Fig. 2B from Yokota ’95. (**B**) full circles: Fig. 3B from Yokota ’95; data from Yokota ’97: open circles: Fig. 2A 4p16.3, full squares: Fig. 2B 6p21.3, open squares: Fig. 2C 21q22.2, full rhombi: Fig. 2D Xq28, open rhombi: Fig. 2D Xp21.3, full up-triangles Fig. 4B MAA-Xp21.3, open up-triangles: Fig. 4B MAA-Xq28, full down-triangles: Fig. 4A PFA-Xp21.3, open down-triangles: Fig. 4A PFA-Xq28.


A
B
mean position independant spatial distance [µm]
genomic distance between genomic markers [Mbp]
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
h
g
f
e
D
d
C
c
B
A

2.7.2 Position Dependent Spatial Distances

Position dependent spatial distance measurements assume that loops and rosettes are fixed structures at least for most of the time during cell cycle and that there are only minor differences between different cells or cells after cell division etc.. Consequently, spatial distances may vary strongly due to the relative position of marker pairs in respect to loop bases or a certain position in a rosette.

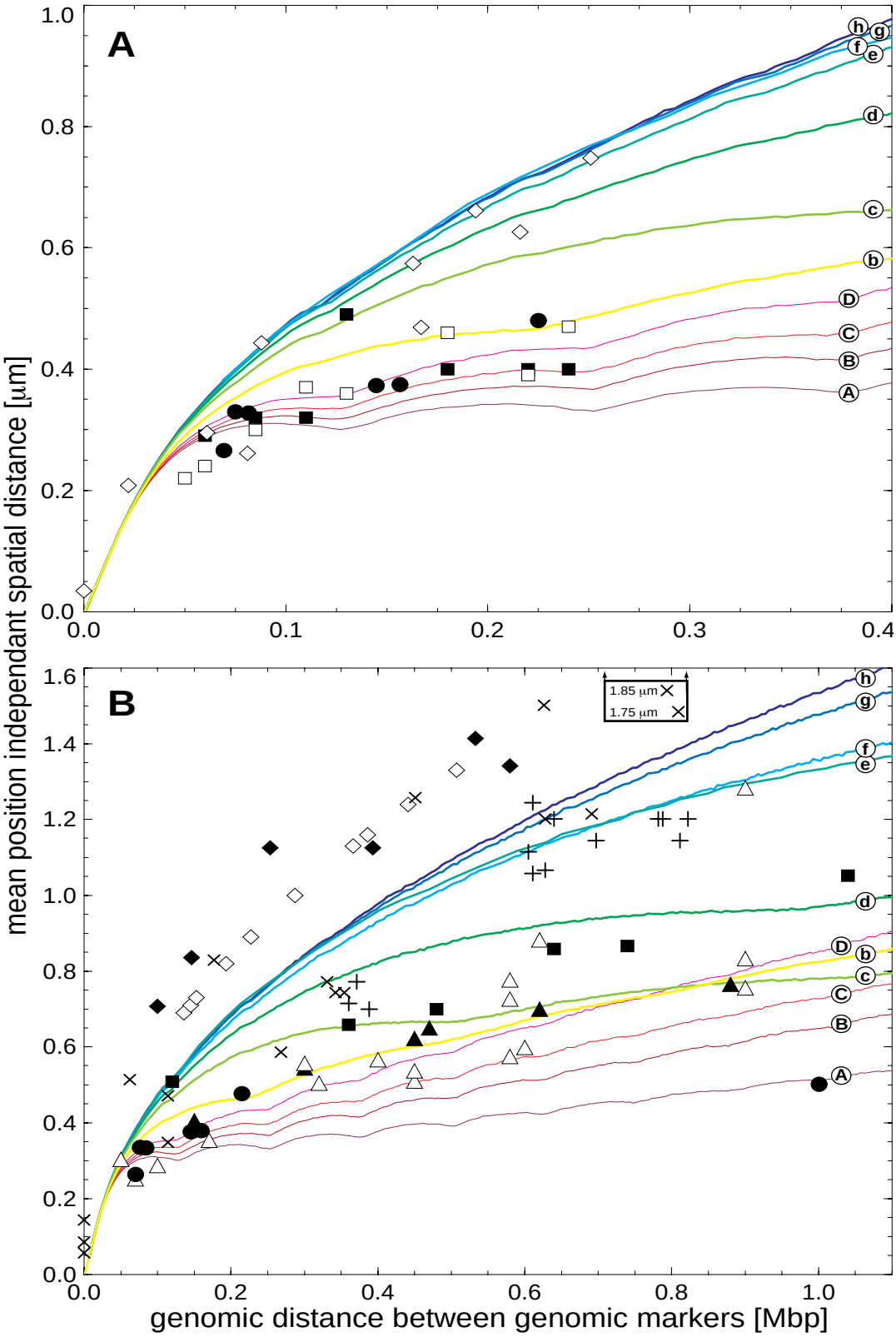
Therefore, the distances between the base of a rosette loop and the following chromatin chain were computed (Fig. 2.6A). The minimum genomic separation used was 5.2kbp (one chain segment). In contrast to the position independent distances only one marker pair existed for one genomic separation within one configuration. Therefore, the average was only taken over 100 to 150 configurations.

For RW/GL and for MLS models the spatial distances increase monotonously for small genomic separations as expected from a random walk (Fig. 2.9A). This is similar to the case of position independent measurements. At genomic separations of exactly half the loop size the spatial distance has a maximum followed by a minimum of or near to zero at a genomic separation of one loop size. The maximum is proportional to the loop size (Fig. 2.8). In an MLS rosette now a another loop follows. For the case of an RW/GL model or if the loop was the last one in an MLS rosette, the spatial distance increases rapidly within the succeeding linker. The increase is proportional to the linker size (Fig. 2.9B), in agreement with the mean distance between rosettes (Fig. 2.5C). The linker is then followed by another loop or series of loops whose next minimum of the spatial distance is theoretically equal to the mean length of the linker. This minimum is, however, underpronounced due to the movements of the next loop or rosette mediated by the linker. For bigger genomic separations the RW/GL model behaves like a staircase (Fig. 2.9A) and the MLS model like a stretched staircase(Fig.2.9B). The mean behaviour of these staircases are described by the position independent spatial distances (Fig. 2.9A&B).

Consequently, the exact local folding morphology could be determined and the RW/GL and the MLS models could distinguished, even if they had the same loop size. The prerequisite for experiments is that the genomic markers cover a genomic region with a resolution of 5 to 10kbp and that spatial distances can be measured

---

Fig. 2.8 Simulated Position Independent Spatial Distances Compared to Experiment - II
Spatial distances between position independent placed genomic markers as function of their genomic distance and comparison to experiments. Lines are simulated values and are the same through Fig. 2.7A&B and Fig. 2.8A&B, symbols are experimental values. For details of model name and properties see Tab. 2.1 and for interpretation of experimental values see Tab. 2.2. Thin lines: From bottom to top A-D MLS-models with 126 kbp loops size and varying linker sizes. Thick lines: RW/GL-models with different loop sizes a-RW/GL to h-RW/GL (a-RW/GL is not shown in Fig. 2.8A and 2.8B). (**A**) full circles: Knoch '98/Rauch '99; data from Monier '99: full squares: fibroblasts 11q13, open squares: lymphocytes 11q13; open rhombi: Trask '89. (**B**) full circles: Knoch '98/Rauch '99; full squares: Lawrence '90; data from Senger '93: full triangles: one colour, open triangles: two colour, data from Warrington '94: full rhombi: Fig. 1, open rhombi: Tab. 2.1; data from Trask '91: crosses: Fig. 4A, pluses: Fig. 4B.


A
B
h
g
f
e
d
c
b
D
C
B
A
1.85 µm
1.75 µm
mean position independant spatial distance [µm]
genomic distance between genomic markers [Mbp]
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
0.1
0.3

with a high enough accuracy that the minima for genomic separations of the order of a loop size can be measured. Should the described staircases not exist in experiments, this could be due to massive destructions by preparation artefacts, or the inexistence of fixed loop and/or rosette sizes as discussed for position independent spatial distances or due to a total different folding of the chromatin fiber.

2.7.3 Dependencies of Spatial Distances in a Set of Genomic Markers

Spatial distances in a set of marker pairs with a set of corresponding genomic separations are correlated to each other in a characteristic manner, if and only if fixed loop and rosette structures as for position dependant distance measurements exist. I. e. the spatial distances between a marker, and its neighbour markers, is defined not only strictly by the genetic separation to these markers but also by their specific location relative to the fixed structure.

To investigate these distance dependencies in a whole marker set, the same measuring process as for the position dependent measurements was used. Additionally the starting point and thus the whole marker set was moved through a rosette revealing the distance dependencies as function of the relative position to a loop or rosette (Fig. 2.6A). For illustration markers at position 0, 21, 42, 63, 126 and 252 kbp were chosen. For easier interpretation, only the distances to the zero position at 0 kbp were considered. The marker ensemble was shifted segment wise through the rosette of a B-MLS model with 126 kbp loops starting at a loop base.

The resulting spatial distances are characterized by coupled oscillations related to the position of the marker set (Fig. 2.10): For a genomic separation of 126 kbp the spatial distance has a minimum if the 0 kbp and the 126 kbp marker both reside near the loop base. A maximum is reached if the 0 kbp and the 126 kbp marker reside both at the loop tips: The spatial distance then is the mean of the succeeding loops being in a neighbouring cis and in a opposite trans position (definition in Fig. 2.6A), which so far is in agreement with single position dependent measurements. Consequently, the spatial distances in the marker set are coupled, thus the maximum of the 126 kbp genomic separation a priori defined the spatial distances of the other markers. This reflects the prerequest that loops and rosettes are fixed structures. The experimental inexistence of such a coupling would suggest a dynamic chromosome organization.

2.8 Excluded Volume and Embedding Nuclear Volume Dependencies

The excluded volume interaction between the chain segments affects the measured parameters of chromosomes depending on the used model and the embedding

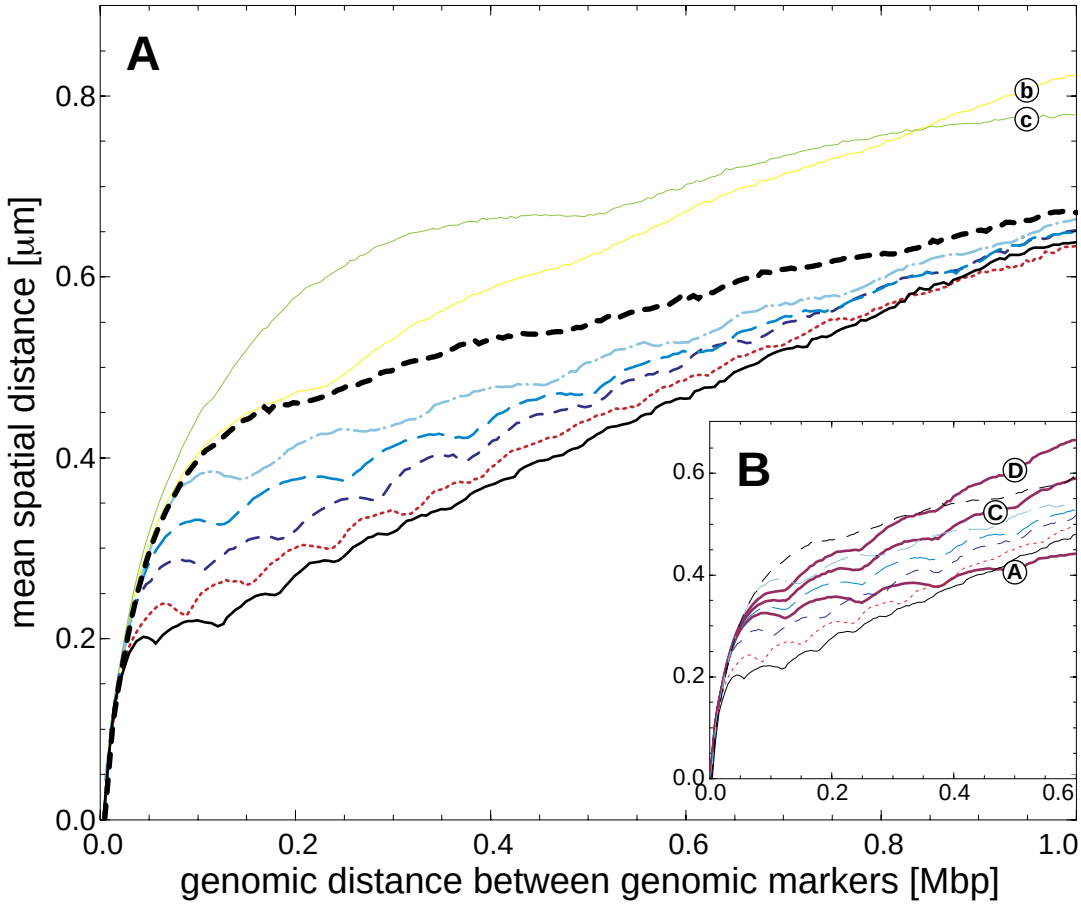


Fig. 2.9 Position Independent Spatial Distances as Function of the Loop Size
Spatial distances between position independent placed genomic markers as function of their genomic distance for MLS models with 126kbp linkers are proportional to the loop size for genomic separations <1Mbp (**A**; for comparison b- and c-RW/GL models are also shown; E-MLS: solid; F-MLS dotted; G-MLS: dashed; B-MLS: long dashed; H-MLS: long dashdotted; I-MLS: thick dashed). Comparison to position independent spatial distances for A-, C-, D-MLS models with the same loop but different linker size reveals the complex interplay between loops and linkers (**B**; legend as in **A**).

spherical potential. The latter influence depends on the relation between chromosome size, mean radial extension and spherical volume available. Thus, confinement to a spherical volume with 1250nm radius has a smaller effect in an A-MLS model with a mean extension of 4μm than for a B-MLS model with 5.4μm mean extension. The same holds for halving or doubling the spherical volume (970 and 1400nm radius). The properties of chromosome territories (2.4 throughout 2.7), resulted from simulations using a low excluded volume interaction and a low barrier of the spherical potential with 1250nm radius (U_0=0.1kT). An increase of both potentials to U_0=1.0kT confined the territory more rigidly to the embedding volume. It also led to a ~10-fold decrease in the relaxation time into thermodynamic equilibrium (a low excluded volume interaction but a high spherical potential yielded similar relaxation times as if both were low).

The basic morphological properties of the chromosome models did not change with nuclear radius, although the mean nearest neighbour distances between arbitrary subcompartments (Fig. 2.5D) models decrease in agreement with the density increase stated below. Consequently, the distinction between subcompartments decreases as well (denser parts of Fig. 2.1B). In the RW/GL models the large loops cannot reach out of the chromosome territory anymore, despite their free intermingling within. The latter suggests that large RW/GL loops reach out of the territories into other territories which agrees with the simulation of whole nuclei (Chapter 3).

The strict confinement to the embedding volume resulted in an increase of the radial density plateau with a sharp drop in the density near the edge for both the MLS (Fig. 2.4A) and the RW/GL model. The height of the plateau and the mean radial extension is therefore also proportional to the embedding volume. The shape of the density curve of the RW/GL model in connection with the free intermingling of giant loops could therefore also be interpreted as artefact possibly only vanishing for even higher excluded volume interactions. A low excluded volume interaction led to a wider radial mass distribution of subcompartments with two peaks, while a high excluded volume interaction resulted in a distribution with only one peak (Fig. 2.4B). This results from a quite high segment density near the center of a rosette, thus a high excluded volume interaction pushes these segments further outwards. This explains the shift from two to one peak. This shift demonstrates also the interplay of a high density near the center of a rosette and the lower density in the outer region of the rosette.

Despite the decreased distance between arbitrary subcompartments in the MLS model due to the density increase (Fig. 2.5C) for high excluded volume interaction and high spherical potential (Fig. 2.5D), the distance between succeeding subcompartments remained unchanged. This holds not for incompletely relaxed chromosome territories or for a sudden increase in the barrier of the embedding potential leading to a rapid decrease in the available volume (Fig. 2.5C). Thus, during a rapid volume compression subcompartments are pressed together and the available space is filled up. Consequently, the distance between succeeding subcompartments is first decreased before the relative position of succeeding subcompartments increases again to their average relaxed value. This is mainly constrained by the linker length, while keeping the reduced distance to arbitrary subcompartments nearby.

Increasing the height of both potentials did not change the distances for genomic separations <10 Mbp for the MLS or the RW/GL models. This agrees with unchanged subcompartment diameters and distances between subcompartments. Thus, the changes in the embedding volume seem still very moderate considering the nuclear volume relationships (Chapter 5). For genomic separations >10 Mbp, however, the increase in the spatial distance was slower and reached the diameter of the spherical volume as maximum, in agreement with density changes and subcompartment distances. This is also in agreement with simulation of whole nuclei with different diameters (Chapter 3).

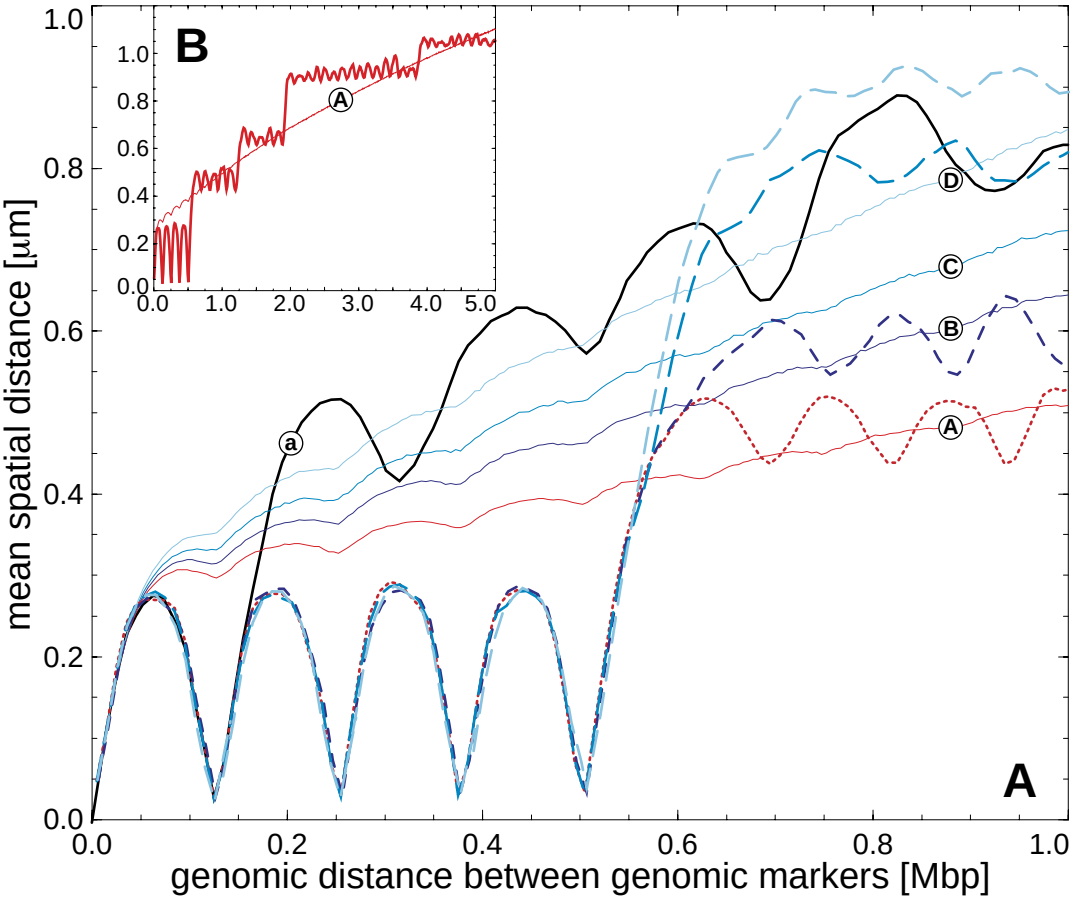


Fig. 2.10 Position Dependent versus Position Independent Spatial Distances
Comparison between spatial distances of position dependent and position independent placed genomic markers as function of their genomic distance are both proportional to the loop and linker size (**A**). Position dependent spatial distances show the detailed loop structure and connection to the next loop or rosette, whereas position independent spatial distances show the loop structure only in a smeared out manner. The loop size is 126 kbp and the linker varies for all shown models (A-MLS dotted, B-MLS dashed, C-MLS long dashed, D-MLS dash dotted, a-RW/GL model, position independent spatial distances for A-D MLS models are thin lines; Tab. 2.1). Comparison of spatial distances of position dependent and independent placed genomic markers for genomic separations up to 5 Mbp for 5 subcompartments of the A-MLS model (**B**). Position independent spatial distances represent the mean or smeared out structure of position dependent spatial distances.

2.9 Comparison between Simulated and Experimental Spatial Distances

2.9.1 General Properties of Methods Used in Different Studies

For the evaluation of experimental spatial distance measurements between small genomic markers labelled by fluorescence *in situ* hybridization (FISH) as function

of their genomic separation three characteristics and their interplay are important: the purpose, the used preparation conditions and the applied imaging techniques followed by distance extraction procedures (for a detailed overview see Tab. 2.2).

Purpose: First spatial distance measurements were used to find the sequential order of clones in gene mapping assays (Lawrence *et al.*, 1988; Lawrence *et al.*, 1989; Lawrence *et al.*, 1990; Trask *et al.*, 1989; Trask *et al.*, 1991; Senger *et al.*, 1993; Warrington & Bengtsson, 1994). Lawrence *et al.* (1988, 1989, 1990) and Trask *et al.* (1989, 1991) showed that the spatial distance increased monotoneously with the genomic separation. This was interpreted as a random walk behaviour of the chromatin fiber (van den Engh *et al.*, 1992). Consequently, the reverse assumption to find the genomic separation for a known spatial distance, seemed reasonable. However, assuming the chromatin fiber to conduct a random walk is incompatible with the formation of defined chromosome territories. Thus, due to the few chromosomal regions spatially mapped and the development of better sequencing techniques, the focus shifted to the detailed determination of metaphase and interphase chromosome packing (Trask *et al.*, 1993; Yokota *et al.*, 1995). This resulted in the Random-Walk/Giant-Loop model (Sachs *et al.*, 1995). Further studies thereafter were to our knowledge only focused on the three-dimensional structure of interphase chromosomes (Yokota *et al.*, 1997; Monier, 1997; Knoch *et al.*, 1998; Knoch *et al.*, 1998; Rauch, 1999; Knoch *et al.*, 2000; Rauch *et al.*, 2002).

Preparation: The preparation method is adapted to the experimental goal: To find the sequential order of clones the markers should be as far apart as possible and bright. Therefore, the first studies were done on interphase nuclei with decondensed chromosomes. To enhance decondensation, the cells were hypotonically swollen with 75 mM KCl, resulting in an increased nuclear volume and flattening of nuclei a smaller height of $<5\,\mu m$. The cells were even dropped on the cover slips to enhance the flattening (Trask *et al.*, 1989). For hybridization of the probe to the target, both have to be denatured in the nucleus to single strand DNA, usually being performed with 50% (before 1995) to 70% formamide in 2*SSC to decrease the denaturation temperature to 72 °C. For the chromosomes to survive this treatment (despite the preservation of the detailed chromatin folding), the cells have to be fixed before denaturation. The higher the degree of fixation the more of the native chromosomal structure is preserved and the more difficult the denaturation as well as hybridization process. For sequencing assays structure preservation was, however, of minor importance (Fiber-FISH techniques represent the total loss of structure). Most studies used a 3:1 methanol acetic acid solution for fixation and air dried the cells onto the coverslips (dropping of cells on the coverslips required fixation before). Then the cells were rehydrated for denaturation and hybridization. For a bright FISH signal for microscopic detection, the probe was labelled with biotin or digoxigenin and amplified with fluorescently labelled avidin or antibodies. Unfortunately, this results in a higher background and size of the FISH signal (Knoch, 1998). In summary, the

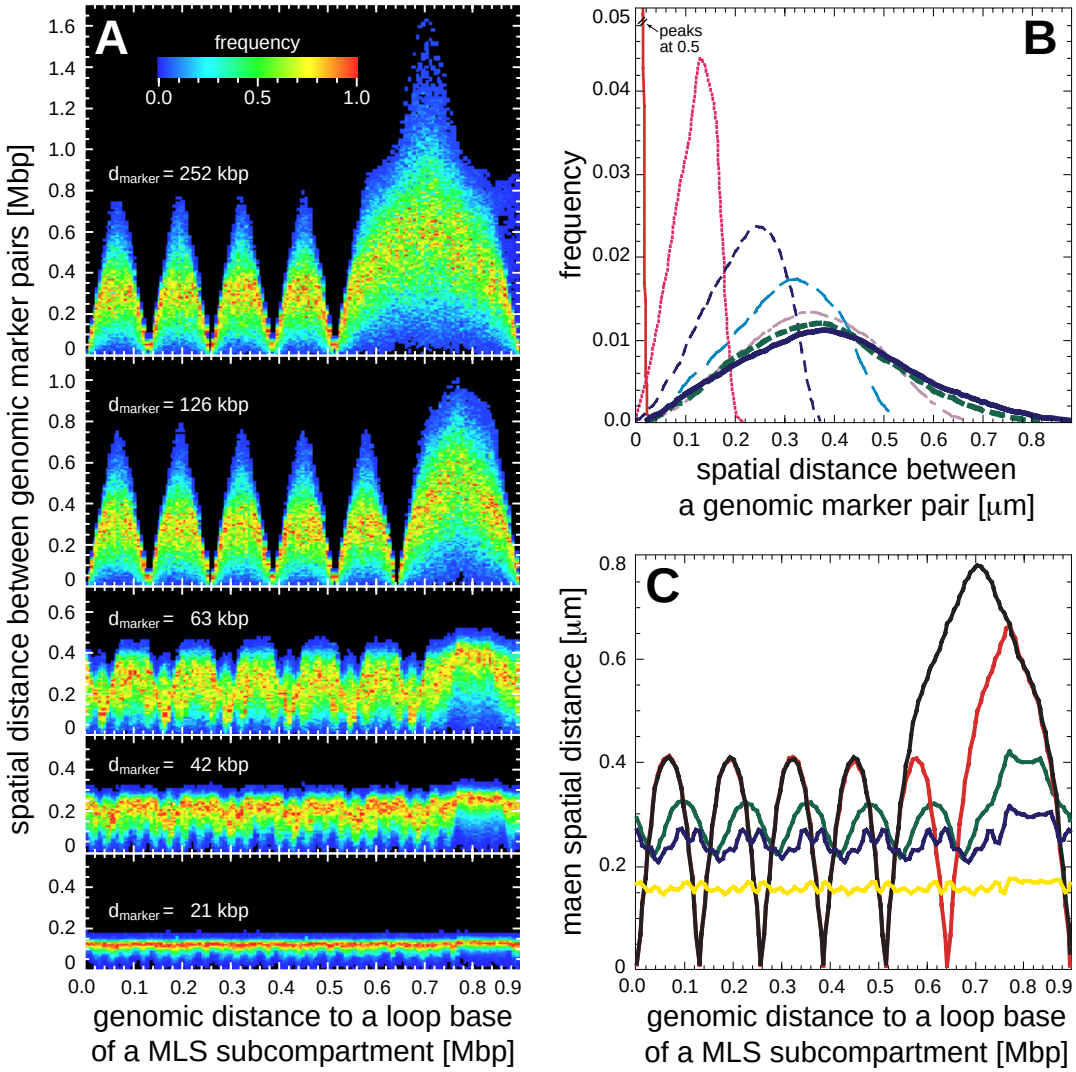


Fig. 2.11 Spatial Distances in an Ensemble of Genetic Markers
Behaviour of the spatial distances of a set of markers for various positions of the marker set relative to a rosette of the B-MLS model with loop and linker sizes of 126 kbp: (**A**) Distance distributions for different genomic separations as function of their shift in steps of 5.2 kbp relative to a loop basepoint of a rosette show characteristic oscillations and correlations as expected for fixed loop and rosette structures. The shift into a linker is clearly visible as well as the increased standard deviation due to the higher positional variability for markers near loop tips. For clarity and comparison with (**C**) the maximum value of the colour coded frequency was calibrated to 1.0. (**B**) Distance distributions for a genomic separation of 126 kbp as function of their shift in steps of 10.5 kbp relative to a loop basepoint of a rosette. For small shifts and therefore location of the markers near the loop base i. e. the rosette center the distribution is asymmetric with a bias to bigger distances. Shift of 0 kbp: solid red, 10.5 kbp: red dotted, 21 kbp: dashed blue, 31.5 kbp: long dashed blue, 42 kbp: dash dotted purple, 52.5 kbp: thick dashed green, 63 kbp: thick solid blue. (**C**) Mean spatial distance of the distance distributions as function their shift from (**A**) show clearly correlations between the oscillations. Genomic separation 21 kbp: yellow, 42 kbp: blue, 63 kbp: green, 126 kbp: red, 252 kbp: black.

chromosomes are exposed to methanol/acetic acid, flattening, drying, rehydration, heat, formamide and high salt concentrations. Thus, the chromosomal structure is distorted.

For the detailed determination of the chromosome structure more structure preserving preparations were used (first described by Popp *et al.*, 1990): Here, cells are grown directly on coverslips and fixed in 4% paraformaldehyde. This left the cells intact, that supplying the cell nucleus with the hybridization probes, requires the digestion of holes into the cell membranes mostly with 0.1% Triton X-100 and 0.5% Saponin. This is enhanced by dipping the coverslip into liquid nitrogen after incubation in 10 to 20% glycerol. This procedure was used with some minor protocol differences by Yokota *et al.* (1995, 1997), which lead to the first ambiguities about the RW/GL model, and by Monier (1997), Knoch *et al.* (1998), Knoch *et al.* (1998), Rauch (1999), Knoch *et al.* (2000) and Rauch *et al.* (2002).

Imaging: The measurement of spatial distances is characterized by the two classes of imaging methods and distance extraction procedures used: Epifluorescence microscopy with photographic image acquisition is especially suited for fast and frequent measurements with low spatial resolution. Confocal laser scanning microscopy with digital imaging is time consuming, but results in high spatial precision. Thus, in sequential mapping epifluorescence microscopy was used. Here only lateral distances between FISH signals in the same focal plane can be determed accurately without special image reconstruction procedures. Due to the axial elongation of the focus (up to 600 nm) and the mean FISH signal diameter of 1000 to 1500 nm, the FISH signal's center of mass can be axially shifted, while still appearing in the same focal plane. This effect is the greater the more the FISH signals are elongated or aspherical themselves. Consequently, the real spatial distance between the mass centres of signals is projected into the two-dimensional focal plane (Fig. 2.6B). The underestimation error ΔSD of the real spatial distance depends on the axial distance AD of the FISH signals and the measured projected lateral distance LD. The positive deviation ΔSD does not average out to zero and consequently adds to the statistical measurement error:

$$\Delta SD = \sqrt{LD^2 + SD^2} - LD. \tag{2.17}$$

The error is significant regarding typical ADs of 50 to 400 nm and LDs of 0 to 2.0 µm (Fig. 2.4B). For a mean AD = 250 nm and LD = 500 nm, SD is underestimated by ΔSD = 55 nm. AD = 200 nm and LD = 300 nm (the extension of a 126 kbp loop) results in a significant ΔSD = 85 nm and 120 nm for LD = 200 nm.

Fast acquisition of the statistical necessary amount of spatial distances in the same focal plane, required also preparation by hypotonic swelling and flattening in the first studies. The spatial distances were measured by projecting a photographic film to a screen. The distances were then measured with a ruler between the mass centres of the FISH signals, and multiplying with the magnification. In later works focusing on the three-dimensional structure, the images were projected onto a digi-

tizing board measuring the distances automatically (Yokota *et al.*, 1997). Although Warrington & Bengtsson (1994) had used confocal laser scanning microscopy for the sequential ordering of clones, this method thereafter was only used to determine the three-dimensional structure of chromosomes with high precision (Monier, 1997; Knoch *et al.*, 1998; Knoch *et al.*, 1998; Rauch, 1999; Knoch *et al.*, 2000; Rauch *et al.*, 2002). Here whole three-dimensional image stacks of confocal planes with separations of 200 to 250 nm were acquired. The images were filtered and deconvoluted with the point spread function for higher resolution. Finally the FISH signals, their center of mass and volume, were extracted automatically and the spatial distance was determined as the real three-dimensional distance between the FISH signal. Rauch (1999, 2002) even corrected for the chromatic shift in two colour experiments. Here signals could be shifted laterally by -23 to+45 nm ±40 nm and axially by -106 to +107 nm ±66 nm. The determination of the point spread function and the chromatic shift with fluorescent spherical beads depends on the special microscope setup and requires often calibration. This technique allows the measurement of spatial distances between the mass centres of FISH signals down to spatial separations of 50 nm with 70 nm accuracy (Bornfleth, 1999).

2.9.2 Comparison of Simulated Position Independent Spatial Distances to Experiments

The experimental spatial distances were compared to the simulated position independent spatial distances, since these make the least assumptions about the chromosomal structure for the different proposed models and are the most resistant against preparation artefacts (Fig. 2.7 A&B, Fig. 2.8 A&B; for a general comparison overview Tab. 2.2). Generally, the experimental data have statistical errors of ±70 nm.

Trask *et al.* (1989) used one colour FISH to determine spatial distances around the hamster dihydrofolate dereductase gene. The data fit to an MLS or RW/GL model with a loop size of ~80 kbp for a genomic separation <80 kbp. Here the spatial distance shows a minimum followed by a steep jump for a genomic separation of 90 kbp (Fig. 2.8 A). For larger genomic separations the data fit better to an RW/GL model with >1.0 Mbp loops, despite another significant jump. Taking into account the one colour labelling of the FISH signals and the underestimation of the spatial distance due to the use of an epifluorescence microscope, an RW/GL model with loops of ~100 kbp or > 1 Mbp could also be supported.

The spatial distances in the one colour study of Lawrence *et al.* (1989) for the human dystrophin gene agree with RW/GL models of 0.5 to 1.0 Mbp (Fig. 2.8 B).

Trask *et al.* (1991) used one and two colour FISH to map a Xq28 region. The one colour data fit best to an RW/GL model with 0.7 to 1.0 Mbp loops for genomic separations <0.4 Mbp, including frequent jumps (Fig. 2.8 B). For bigger genomic separations the data fit to RW/GL models with 1.0 to >5.0 Mbp loops. The two colour distances are significantly smaller than the one colour data for genomic separations

>0.6Mbp and suggesting RW/GL models with loops <3.0Mbp (Fig. 2.8B). This interpretation is supported by the good statistics for genomic separations ~ 0.8 Mbp.

Mapping of the genomic regions 4p16.3 (first published in van den Engh *et al.*, 1992; Trask *et al.*, 1993) and 6p21 (Trask *et al.*, 1993) with two colour FISH both fit RW/GL or MLS models with ~100kbp loops for genomic separations <300kbp, again taking into account the error and the underestimation of the spatial distance (Fig. 2.7A). The distances for larger genomic separations fit to RW/GL models with >5Mbp loops for 4p16.3 and 1.0 to >5.0Mbp loops for 6p21. The spreading of the values is, however, so big that the interpretation by a pure random walk behaviour of van den Engh (1992) seems most reasonable.

Senger *et al.* (1993) used one and two colour FISH to fine map the human MHC class II region (6p21). As methods "standard procedures" were used, perhaps similar to the various cited Lawrence and Trask publications. The one colour data fit to MLS models with 0.12 to 0.25Mbp loops and linkers or RW/GL models with 0.1 to 0.5Mbp loops. The two colour distances are smaller and show a significant jump for a separation of ~0.6Mbp. Therefore, the distances can be fitted to an MLS model with 0.1 to 0.25Mbp loops and linkers, or an RW/GL model with 0.1 to 0.5Mbp loops.

Warrinton & Bengtsson (1994) used two colour FISH as one method to derive a high resolution sequence map of 5q31-q33. A confocal laser scanning microscope was used but the preparation details were not given (due to citations of Trask and van den Engh presumably similar). To determine the genomic separation between clones, distances from 4p16.3 mapped by Trask *et al.* (1993) were taken for calibration. The calibration data fit an RW/GL model with >5.0Mbp loops (Fig. 2.8B) and consequently the spatial distances for the 5q31-33 region fit the same model. Taking into account the various models supported by different data (see above and below), show clearly the unreliability of this approach for sequential mapping. Sequential mapping works only if the chromosome structure shows a strong monotonous relation between genomic and spatial distance or if structural features proposed e. g. in the MLS model are deliberately destroyed to receive such a monotonous relation.

The first study solely dedicated to the fine structure of chromosomes was conducted by Yokota (1995) on chromosome IV for genomic separations of 0.15 to 190Mbp using two different preparation conditions. The data describe different spatial distance relationships for genomic separations below and above 4Mbp. This suggested the RW/GL model by Sachs *et al.* (1995). The spatial distances >4Mbp are twice as big for the hypotonic methanol/acetic acid (MAA) than for the paraformaldehyde (PFA) fixation reflecting the swelling and decondensation effect of the methanol/acetic acid fixation. Since the RW/GL and MLS models were simulated such that they behave similarly for separations of >10Mbp, only distances <4Mbp are compared. MAA fixation distances fit to an RW/GL model with 2.0 to 4.0Mbp loops (Fig. 2.7A). The PFA fixation suggests, however, an MLS model with a loop and linker size of 0.1 to 0.15Mbp (Fig. 2.7B) in disagreement

with the interpretation of Yokota *et al.* (1995). This discrepancy for the same genomic region, but different preparation conditions lead to the MLS model.

In a second study by Yokota *et al.* (1997) G- and R-band related genomic regions were further investigated with MAA and PFA preparations in fibroblasts and HeLa cells (Fig. 2.7A). The MAA fixed R-band regions suggest various RW/GL models: the 4p16.3 region suggests an RW/GL model with 2.0 to 3.0 Mbp loops, 6p21.3 suggests an RW/GL model with 4.0 to 5.0 Mbp loops and Xq28 suggests an RW/GL model with 1.0 to 5.0 Mbp loops including a jump ~0.7 Mbp. The MAA fixed G-band regions result in similar RW/GL models: 21q22.2 fits an RW/GL model with 1.0 to 2.0 Mbp loops and Xp21.3 fits an RW/GL model with 0.5 to 1.0 Mbp loops with a minimum in the spatial distance ~0.6 to 0.7 Mbp, possibly indicating a single loop. The PFA fixation in fibroblasts reveal, however, smaller distances: the G-band Xp21.3 region suggests an RW/GL model with 0.25 Mbp loops or an MLS model with 0.126 Mbp loops and ~0.200 Mbp linkers. The similar fixed R-band Xq28 region fits best to an RW/GL model with 1.0 Mbp loops. The same result is obtained for the same genomic separations in HeLa cells for an MAA fixation. On the one hand this clearly suggests no difference in the organisation between fibroblasts and HeLa cells (despite PFA and MAA fixation) in agreement with the Yokota interpretation. The results for the PFA fixed fibroblasts and MAA fixed HeLa cells are, however, significantly different from the former obtained in MAA fixed fibroblasts favouring models with higher compaction, that means smaller RW/GL loops. This favours also the MLS model and suggests a difference between fibroblast and HeLa cells although both fixed with the MAA preparation (undiscussed in Yokota *et al.*, 1997). Within the group of the MAA fixed fibroblast the G-bands show also a higher compaction than the R-bands, in agreement with Yokota.

The first high precision study to determine the detailed structure of chromosomes using a confocal microscope and computerized image analysis was done by Monier (1997): For a genomic region in 11q13 in fibroblasts the data suggest an MLS model with 0.126 kbp loops and 0.180 Mbp linkers (Fig. 2.8A). The mapping of the same region in lymphocytes favours also an MLS model with 0.100 Mbp loops and a 0.180 to 0.240 Mbp linkers (Fig. 2.9A). The results support the view that the local structure remains the same, although fibroblasts and lymphocytes have different volumes. This view is also supported by the simulations involving different embedding volumes of chromosomes (2.6) and the simulations of whole nuclei (Chapter 3).

The study with the highest precision and resolution conducted by Knoch (1998, 1998, 2000) and Rauch (1999, 2002) mapped the PraderWilli/Angelmann region on chromosome 15q11-21 with 0.069 to 0.213 Mbp clone separations (in publications before 2002, the smallest separation was 0.019 Mbp, due to a missplaced marker; the results, remain the same). The mapping included also a genomic separation of 1.0 Mbp (≈10 µm chromatin contour length) distal from the center of the mentioned set of clones to a another genomic marker. The spatial distances fit to an MLS model

Tab. 2.2 Properties of Experimental Spatial Distance Measurements
The more sophisticated methods were used, the more the spatial distance measurements agree with MLS-models. (DHFR: dihydrofolate reductase gene, Dystrophin: human dystrophin gene, MHC: major histocompatibility complex, F: fibroblasts, WI38F: WI-38 fibroblasts, HFF: human foreskin fibroblasts, L: lymphocytes, cf: arrest through confluency, ar: arrest in G1, MAA: methanol acetic acid, PFA: paraformaldehyde, FM: formamide, Dig: digoxigenin, photo: image acquisition with film

Study	Location	Preparation of Cells			FISH			Microscopy		Fit to model
		Cell cycle	KCl [nM]	Fix-ative	Melt-ing	Label	Col-ours	# of nuclei	Image acquisi-tion	
Fig. 3B, Trask '89	DHFR	UA41 G1-cf	75 dropped	MAA 3:1	FM 50 %	Biotin	1	20-37	photo, wall	RWGL 0.08-J RWGL 1.0
Fig. 3B, Lawrence'90	Dystro-phin	WI38F G1	75 dropped	MAA 3:1	FM 50 %	Biotin	1	20-60	photo, wall	RWGL 0.5-1
Fig. 3A, Trask '91	Xq28	F G1-cf	75 dropped	MAA 3:1	FM 50 %	Biotin	1	30-60	photo, wall	RWGL 0.7 J RWGL 2.0->5.0
Fig. 3B, Trask '91	Xq28	F G1-cf	75 dropped	MAA 3:1	FM 50 %	Biotin Dig	2	30-60	photo, wall	RWGL 1.0-3.0
Fig. 3, v.d. Engh '92 or Fig. 5A, Trask '93	4p16.3	F G1-cf	75 dropped	MAA 3:1	FM 50 %	Biotin Dig	2	?	photo, d-board	L_S <=0.1 for GS < 0.5 < RWGL >5.0
Fig. 5B, Trask '93	6p21	F G1-cf	75	MAA 3:1	FM 50-70 %	Biotin Dig	2	?	photo, d-board	L_S <=0.1 for GS < 1.0 < RWGL 1.0-5.0
Fig. 5, Senger '93	MHC 6p21.31	HFF G1-cf	?	?	FM 50 %	Biotin	1	> 30	photo, wall	MLS L_S&LI_S= 0.12-0.25 RWGL 0.1-0.5
Fig. 5, Senger '93	MHC 6p21.31	HFF G1-cf	?	?	FM 50 %	Biotin Dig	2	> 30	photo, wall	MLS L_S=0.1 LI_S=0.18 RWGL 0.1-0.5
Fig. 1, Warrington '94	4p16.3	F G1-cf	75	MAA 3:1	FM 50 %	Biotin Dig	2	?	?	RWGL > 5.0
Tab. 1, Warrington '94	5q31-33	L	?	?	?	?	?	?	CLSM BioRad	RWGL > 5.0
Fig. 2B, Yokota '95	4p16.3	F G1-cf	40 dropped	MAA 3:1	FM 70 %	Biotin Dig	2	40-360	photo, d-board	RWGL 2.0-4.0
Fig. 3B, Yokota '95	4p16.3	F G1-cf	-	PFA 4 %	FM 70 %	Biotin Dig	2	40-350	photo, d-board	MLS L_S&LI_S= 0.1-0.125

with a loop size ~0.126 Mbp and a linker size of 0.060 to 0.100 Mbp due to the spatial distance of only 0.500 µm for the genomic separation of 1 Mbp (Fig. 2.8 A&B).

2.10 Discussion of the Simulation of Single Chromosomes

The folding of the 30 nm chromatin fiber into chromosome territories is still a largely unresolved problem (Chapter 1). The current knowledge assumes that the fiber is compacted in different stages: chromatin loops, aggregates of loops, and arrangement of these subcompartments into chromosome territories. The details of

at a normal epifluorescence microscope, wall: distance measurements with ruler between center of mass of signals by projection of slide to wall, d-board: distance measurements between center of mass of signals by projection slide to a digitizing board, CLSM: confocal laser scanning microscope, BioRad: Bio-Rad CM software for spatial distance measurements, GS: genomic separation, J: jump of spatial distance which could indicate a linker between a loop, L_S: loop size, LI_S: linker size).

Study	Location	Preparation of Cells			FISH			Microscopy		Fit to model
		Cell cycle	KCl [nM]	Fix-ative	Melt-ing	Label	Col-ours	# of nuclei	Image aquisi-tion	
Fig. 2A, Yokota '97	4p16.3 R-band	F G1-cf	40 dropped	MAA 3:1	FM 70 %	Biotin Dig	2	37-178	photo, d-board	RWGL 2.0-3.0
Fig. 2B, Yokota '97	6p21.3 R-band	F G1-o	40 dropped	MAA 3:1	FM 70 %	Biotin Dig	2	37-178	photo, d-board	RWGL 4.0-5.0
Fig. 2C, Yokota '97	21q22.2 G-band	F G1-o	40 dropped	MAA 3:1	FM 70 %	Biotin Dig	2	37-178	photo, d-board	RWGL 1.0-2.0
Fig. 2D, Yokota '97	Xp21.3 G-band	F G1-o	40 dropped	MAA 3:1	FM 70 %	Biotin Dig	2	37-178	photo, d-board	RWGL 0.5-0.9
Fig. 2D, Yokota '97	Xq28 R-band	F G1-of	40 dropped	MAA 3:1	FM 70 %	Biotin Dig	2	37-178	photo, d-board	RWGL 1.0-5.0j
Fig. 2A, Yokota '97	Xp21.3 G-band	F	-	PFA 4 %	FM 70 %	Biotin Dig	2	37-178	photo, d-board	RWGL 0.25 MLS L_S=0.126 LI_S=0.200
Fig. 2A, Yokota '97	Xq28 R-band	F	-	PFA 4 %	FM 70 %	Biotin Dig	2	37-178	photo, d-board	RWGL 1.0
Fig. 4B, Yokota '97	Xp21.3 G-band	HeLa	40 dropped	MAA 3:1	FM 70 %	Biotin Dig	2	37-178	photo, d-board	RWGL 0.25 MLS L_S=0.126 LI_S=0.200
Fig. 4B, Yokota '97	Xq28 R-band	HeLa	40 dropped	MAA 3:1	FM 70 %	Biotin Dig	2	37-178	photo, d-board	RWGL 1.0
Monier '97	11q13	F	-	PFA 4%	FM 70 %	Biotin Dig	1	22-69	CLSM	MLS L_S=0.126 LI_S=180
Monier '97	11q13	L	-	PFA 4%	FM 70 %	Biotin Dig	1	22-69	CLSM	MLS L_S=0.1 LI_S=0.18-0.24
Knoch '98/ Rauch '99	15q11-21	F	-	PFA 4%	FM 70 %	Biotin Dig	1 & 2	60-120	CLSM	MLS L_S=0.1 LI_S=0.06-0.125

these structures and their dynamics, changes during the cell cycle and the transport properties in nuclei, are still debated. On the chromatin fiber level various models were proposed: In the Multi-Loop-Subcompartment (MLS) model (1.3.8, Fig. 1.14), chromatin loops form rosettes which are connected by a linker. In the Random-Walk/Giant-Loop (RW/GL) model, big loops are attached to a flexible backbone (1.3.7, Fig. 1.14).On the level of the whole nucleus the Inter-Chromosomal Domain (ICD) model has been proposed: Here transport in the nucleus occurs through a network of channels between dense chromosome territories (1.3.5, Fig. 1.13).

Whether the MLS and the RW/GL models can form chromosome territories in interphase, whether they lead to different morphologies on the level of whole nuclei and whether they can be distinguished experimentally on the fiber level has remained unclear. So far it is not understood whether the different fiber topologies are in agreement with the ICD model.

To investigate these hypothesis the MLS and the RW/GL models were simulated approximating the 30 nm chromatin fiber as a polymer chain. The properties of this chain were defined by a stretching and bending potential. An excluded volume interaction kept the chain from crossing and a spherical boundary potential simulated the confinement of the nucleus. Monte Carlo and Brownian Dynamics methods were applied to generate chain configurations at thermodynamical equilibrium from a startconfiguration resembling a metaphase chromosome.

Both the MLS and the RW/GL model form chromosome territories with different morphology: The MLS model reveals territories with a sharp "edge" and the rosettes result in distinct subcompartments visible with light microscopy. In contrast, the big RW/GL loops lead to a homogeneous chromatin distribution. Only the MLS model led to a low overlap of chromosomes, arms and subcompartments in agreement with experiments. The size of these subcompartments is also in agreement with experiments based on fluorescence *in situ* hybridization (FISH), replication labelling *in vivo* by BrdU, and recently developed *in vivo* labelling of chromatin with histone-autofluorescent proteins (Chapter 7). Comparison to experiments of the radial mass and density distributions of chromosomes and subcompartments as well as the spatial distances between the nearest and arbitrary subcompartments revealed best agreement for an MLS model with loop and linker sizes of 63 to 126 kbp.

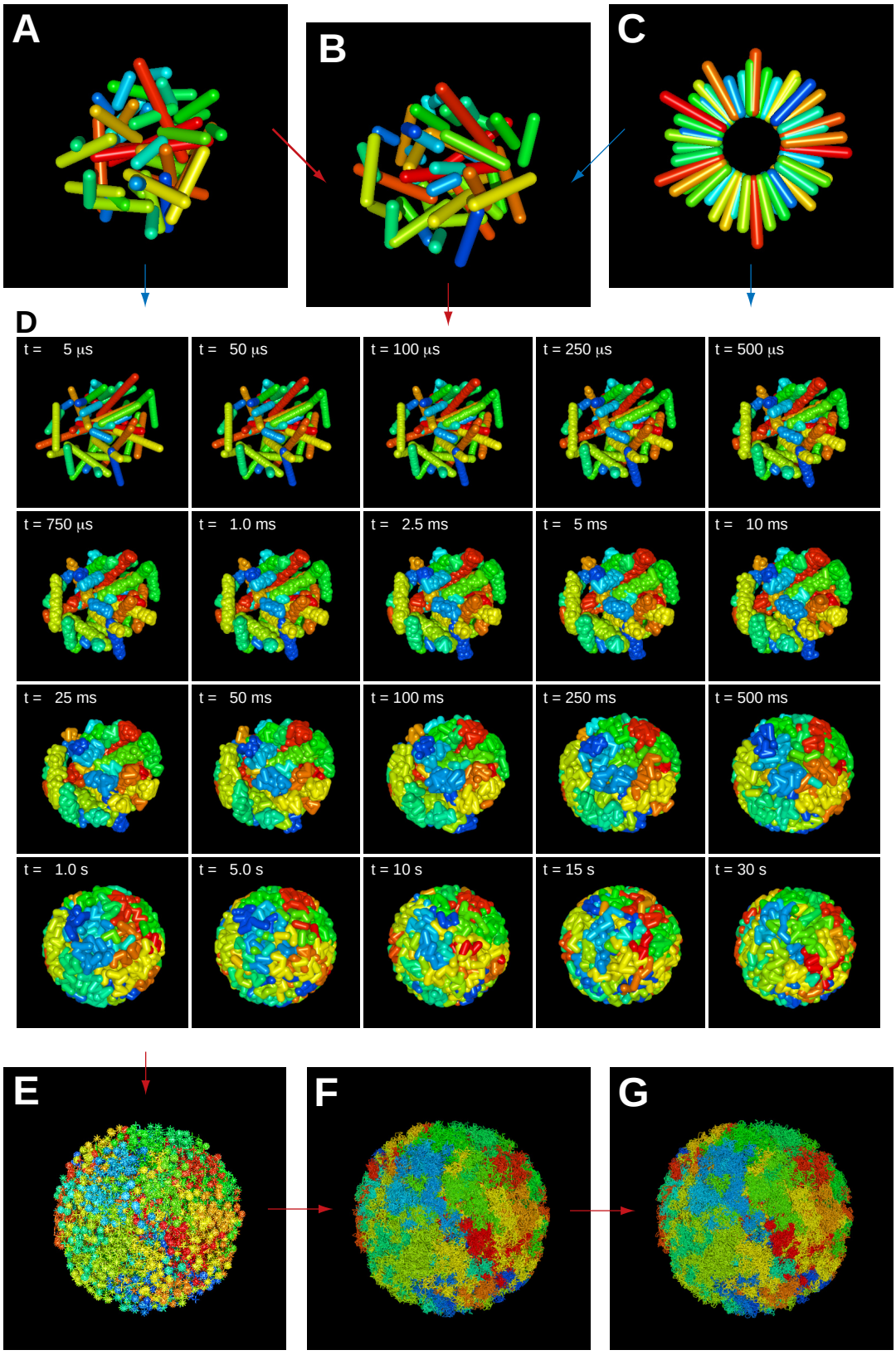
To investigate the MLS and RW/GL models on the level of the chromatin fiber, different methods for spatial distance measurements between genetic markers as function of their genomic separation were introduced and analysed. These methods correspond to different assumptions of the structure, stability and dynamics of chromosome organizations. The results reveal the characteristics of the MLS and the RW/GL model in detail and predict that the two models can be distinguished experimentally with current technologies.

For comparison with experimental distance measurement based on FISH, the 11 published studies were reviewed according to their purpose, preparation and used imaging technique: The trend to investigate the detailed three-dimensional organization of the nucleus led to better structure preserving preparation methods and imaging techniques with ever higher resolution. Therefore, the early studies result in better agreement with the RW/GL model. The latter favour again an MLS model with loop and linker sizes of 63 to 126 kbp (Monier, 1997; Knoch *et al.*, 1998; Knoch *et al.*, 1998; Rauch, 1999; Knoch *et al.*, 2000; Rauch *et al.*, 2002). These studies together with simulations also show that it is possible to determine the three-dimensional organization of interphase cell nuclei in detail.

The morphology of the models reveals also large spaces between the chromatin fibers. Although some of these spaces are occupied by other chromosomes in whole nuclei, the distances between the chromatin fibers is large enough to allow access to nearly every chromosomal location for tracers < 10 nm. Therefore, the diffusion of small molecules and typical proteins is only moderately obstructed. This is in agreement with recent single molecule experiments measuring the diffusion of particles

in nuclei of living cells. Therefore, the assumptions of the Inter-Chromosomal Domain (ICD) model seem simplified, since hyper freeways are not necessary if already the village streets are arenas.

Consequently, comparison of the simulations to experiments favour an MLS model with loop and linker sizes of 63 to 126 kbp and disagree with the hypothesis of the RW/GL and the ICD models. Additionally, the local and global characteristics of cell nuclei are tightly inter-connected. Therefore, the changes described qualitatively during mitosis, apoptosis or in the pathological classification of cancer, can be related to structural changes on the chromatin level. The simulations also propose the intuitive view of a nucleus as evolutionary optimized bioreactor: The genetic information in the nucleus is packaged, to fulfil all the requirements of pure storage, and on the other hand by its structural distribution guarantees an easy transport to every target site in the nucleus by Brownian diffusion. This guaranties random mixing, which leads to the most efficient reaction probability possible in a fluidic system. Structural and chemical modifications can then lead to the subtle regional regulation of processes, allowing possibilities of control well beyond the pure DNA sequence.


A
B
C
D
t = 5 μs
t = 50 μs
t = 100 μs
t = 250 μs
t = 500 μs
t = 750 μs
t = 1.0 ms
t = 2.5 ms
t = 5 ms
t = 10 ms
t = 25 ms
t = 50 ms
t = 100 ms
t = 250 ms
t = 500 ms
t = 1.0 s
t = 5.0 s
t = 10 s
t = 15 s
t = 30 s
E
F
G

3 Simulation of Interphase Nuclei

3.1 Introduction

The simulations of single chromosomes were extended to nuclei containing all 46 chromosomes, to investigate in addition to the folding of the 30 nm into chromosome territories also the chromosome arrangement and the related microscopic morphology. Again different Multi-Loop-Subcompartment (MLS) models, in which small loops form rosettes, connected by a linker, and Random-Walk/Giant-Loop (RW/GL) models, in which large loops are attached to a flexible backbone, were simulated. The 30 nm chromatin fiber was modelled as a polymer chain with stretching, bending and excluded volume interactions. A spherical boundary potential simulated the confinement of the nucleus. Simulated annealing and Brownian Dynamics methods as well as a four step decondensation procedure from metaphase were applied to generate interphase configurations at thermodynamical equilibrium. Both the MLS and the RW/GL model form chromosome territories with different morphologies: The MLS rosettes result in distinct subcompartments visible in electron and confocal laser scanning microscopic images. The RW/GL model leads to a homogeneous chromatin distribution. Even small changes of the model parameters induced significant rearrangements of the chromatin morphology. The low overlap of chromosomes, arms and subcompartments observed in experiments could only be reproduced with the MLS model. The chromatin density distribution in CLSM

Fig. 3.1 Startconfigurations and Decondensation into Interphase of Simulated Cell Nuclei
Metaphase chromosomes were placed as cylinders randomly (**A**) or in a metaphase plate (**B**) into a spherically constrained potential before optional relaxing by simulated annealing avoiding unphysical configurations (**C**). The cylinders were split into spheres and decondensated into interphase by $<6x10^6$ Brownian Dynamic (BD) steps of 5 µs or 30 s (**D**). Then the 30 nm chromatin fiber was placed into the spheres with scaled down chain segments avoiding concatenation (**E**) and softly relaxed with 10^3 BD steps linearly increasing from 0.01 to 0.5 µs (**F**). Segment sizes were 300 nm or smaller that chromatin loops consisted of ≥4 segments. Finally the resolution was increased to 50 nm segments, and relaxed further by $5x10^3$ BD steps of 0.5 µs to ensure proper morphologic and quantitative analysis (**G**). Shown for an MLS model with 126 kbp loops/linkers and 5 µm nuclear radius.

image stacks reveals a bimodal behaviour in agreement with recent experiments. As for the single chromosomes, the MLS model with loop and linker sizes of 63 to 126 kbp yielded the best agreement with experimental parameters. Visual inspection of the nuclear morphology revealed big spaces allowing high accessibility to nearly every nuclear location. Thus, the diffusion of particles is only moderately obstructed in agreement with experiments. A channel like network for molecular transport between chromosome territories, as postulated by the ICD model, was not apparent in the simulations.

3.2 Starting Configuration, Decondensation into Interphase and Properties of Simulated Models

The simulation of whole nuclei with all 46 human chromosomes is more complex than the simulation of single chromosomes described in Chapter 2: In addition to the organization of the 30 nm chromatin fiber within the chromosome, constraints such as nuclear size, metaphase starting configuration or adjectant chromosomes have to be taken into account. The computer power also increases by a factor of 46. Thus, whole nuclei contain ~1.2×10^6 chromatin segments at the highest resolution (in Chapter 2 the mean sized chromosome XV existed of only 2×10^4 segments). This necessitated a sophisticated four step strategy with growing resolution to receive reasonable relaxed interphase configurations at thermodynamic equilibrium. Since the variety of possible Monte Carlo moves is reduced (2.2.2) and their acceptance rates decrease dramatically due to the nuclear density, only Monte Carlo based simulated annealing and the slower Brownian Dynamic methods were applied.

3.2.1 Metaphase Starting Configurations and Simulated Annealing

The startingconfigurations for the metaphase-interphase decondensation, consisted of cylindrical metaphase chromosomes. These were placed either randomly or within a randomly arranged metaphase plate into a spherical boundary potential (Fig. 3.1 A&C). Some experimental data exist for the precise chromosome positioning (Fletcher, 1994; Leitch, 1994; Allison; 1999) and suggest a random arrangement in the metaphase plate. The spherical boundary potential equaled the exclued volume interaction of Chapter 2 (Equ. 2.6) and was adjusted to generate nuclei with diameters of 6, 8, 10 and 12 µm. Since neither the length, the diameter nor the condensation degree as function of the nuclear size were proficiently known, the cylinder length was set to 3.2, 4.3, 5.3, and 6.0 µm for chromosome I, and the cylinder radius was chosen to be 300, 400, 500, and 600 nm, for the respective nuclear sizes.

The startingconfigurations were relaxed by Monte Carlo like simulated annealing to avoid unbiological overlaps or unphysical configurations and to speed up the

decondensation process: The cylinders were randomly shifted by a maximum of 5% of the corresponding nuclear radius and randomly rotated by a maximum of 5°. Then the potential energy between the different cylinders as well as between the cylinders and the boundary potential was approximated: The cylinders were splitted into spheres and the excluded volume potential (Equ. 2.6) and the boundary potential were calculated. The resolution was 200 spheres for chromosome 1. A move was accepted if either the internal energy difference $\Delta H = H_n - H_m$ between the original H_n and the new configuration H_m decreased and if the transition probability

$$P_{transition} = \frac{p_n}{p_m} = e^{\left(\frac{-\Delta H_m}{k_B T}\right)} \tag{3.1}$$

between the two states with the single probabilities p_n and p_m, was larger than a random number from the interval [0,1]. If the energy decreased, a cooling factor reduced the temperature, the maximum shift distance and the rotation angle. The procedure was stopped if the internal energy difference ΔH fell below 10^{-5} in 20 successive steps. Depending on the initial conditions 100 to 500 simulated annealing steps were needed to relax the metaphase startingconfiguration (Fig. 3.1C).

3.2.2 Decondensation from Metaphase into Interphase

To guaranty highly relaxed interphase configurations with the available computer power, the decondensation from metaphase consisted of three steps at different resolutions. All of these steps included the chain properties and potentials described in 2.2.1 and the Brownian Dynamics algorithm described in 2.2.3:

First the cylinders of the startingconfiguration were split into spheres each corresponding to an interphase ideogram band obtained from the 850 metaphase ideogram banding pattern of Francke (1994), by division with three, since 2550 to 3000 interphase bands were reported (Yunis, 1981). The spheres were connected by a chromatin linker using the chain properties of chromatin (2.2.1). The mean length of the relaxed linker was set to 425, 600, 735 and 850 nm according to the desired interphase model. (3.2.3; see Tab. 2.1 and the results of Chapter 2). To allow numerically stable and fast decondensation, the sphere radius was set to 175, 234, 292, 350 nm for nuclear diameters of 6, 8, 10 and 12 µm, respectively. Extensive tests revealed that only the speed of decondensation, was influenced by these values, but none of the other properties of chromosome models. It should be noted that the use of these different settings cannot be avoided by simple rescaling of the chromosome configurations for different nuclei, since the linker connecting the spheres is a fixed parameter depending on the chromosome model. The Brownian Dynamics algorithm was applied until complete decondensation from metaphase into interphase was achieved. This was checked by visual (Fig. 3.1D) and internal energy inspection and was supported by the later analysis of chromosome properties. Using

Brownian Dynamics steps of 5 µs, usually 0.5, 1.0, 2.5 and 5.0 s for the different nuclei were sufficient for decondensation. Controls were conducted up to 30 s.

After this initial decondensation, the spheres were replaced by the 30 nm chromatin fiber topology according to the chromatin model. To avoid concatenation of chromatin loops belonging e. g. to different chromosomes, the chain segments of these loops were scaled down (Fig. 3.1 E). Then the segments and loops were softly decondensed/relaxed with 10^3 Brownian Dynamics steps linearly increasing from 0.01 to 0.5 µs. The adjustment of this step times was neccessary to avoid explosive detensioning due to the high density of scaled down loops. Segment sizes were 300 nm or such that a chromatin loop consisted at least of four segments. Finally, the resolution was increased to 50 nm (40 nm for the 126-RW/GL^X-models) and relaxed by $5x10^3$ Brownian Dynamics steps of 0.5 µs. Thus, relaxation on the highest level of resolution took 2.5 ms (Fig. 3.1 F).

The simulation used the improved simulation and parallelization code of Chapter 2, with a parallelization efficiency increased from 85 to 95 % using 16 processors. The simulations of whole nuclei including their analysis presented here have totalled ~260.000 h (~30 years) on a single R6000 processor with 120 MHz (note: the 96.000 CPUh for single chromosomes were simulated at 60 MHz).

3.2.3 Multi-Loop-Subcompartment (MLS) Model

In the MLS model small loops are forming rosettes connected by a chromatin linker. The rosettes form substructures of a chromosome such as those found in interphase studies on transcription and replication (Berezney *et al.*, 1995; Zink & Cremer, 1998; Zink *et al.* 1998). In metaphase chromosomes these rosettes could be related to the ideogram banding pattern (Pienta & Coffey, 1984; Laemmli, 1994). As loop sizes 63, 84, and 126 kbp and as linker sizes 63, 126, 252 and 504 kbp, corresponding to mean interphase distances of 425, 600, 735 and 850 nm, were chosen. The results of Chapter 2 revealed best agreement between simulation and experiment for 63 to 126 kbp linkers. The loop number in one rosette is proportional to the total DNA content of a rosette divided by the loop size. To simulate different DNA contents of the rosettes, which is the case in real chromosomes, the 850 band metaphase ideogram banding pattern of Francke (1994, Fig.1 10) was used. Each metaphase band was divided by three, since during decondensation into interphase the bands split up into ~2550 to 3000 (Yunis, 1981). Since the DNA in a linker is taken from the DNA content of the single bands and therefore rosettes, the loop number of rosettes is also inversely proportional to the linker size (Tab. 2.1). For MLS models with small loops, the segment length was reduced to 200 or 100 nm as mentioned in 2.2.2 (Tab. 2.1). The base pair content of the linker was taken in equal parts from the neighbouring rosettes. Rounding effects which could have resulted in a loss of base pairs were avoided by adding loops to rosettes with low loop numbers. Throughout Chapter 3 the following nomenclature L_S-LI_S-$MLS^{r\text{-}nucleus}$ (L_S: loop size; LI_S: linker size; r-nucleus: nuclear radius) is used for MLS models.

Tab. 3.1 Simulated Nuclei and Physical Properties of the Used MLS Chromosome Model
The number of bands are the number of subcompartments or loops per chromosome for the corresponding chromosome model. The mean band distance is $\langle R_B \rangle = \sqrt{(300nm)^2 \cdot LI_L/(300nm)}$ and the mean territory size is $\langle R_{Ltotal} \rangle = \sqrt{(300nm)^2 \cdot (NB-1) \cdot LI_L/(300nm)}$.

Model	Loop properties		Linker properties		Mean band properties		Mean band distance		Mean territory size	
	Size L_S [Mbp]	Length L_L [µm]	Size LI_S [Mbp]	Length LI_L [µm]	# of bands NB	# of loops per rosette	Theoretic $< R_B >$ [nm]	Simulated $< R_B >$ [nm]	Theoretic $< R_{Ltotal} >$ [µm]	Simulated $< R_{Ltotal} >$ [µm]
63-63-MLS^3 63-63-MLS^4 63-63-MLS^5 63-63-MLS^6	0.063	0.6	0.063	0.6	103±48	20.1±9.6	424	430 429 430 438	4.26	4.0 4.2 4.5 4.6
84-63-MLS^3 84-63-MLS^4 84-63-MLS^5 84-63-MLS^6	0.084	0.8	0.063	0.6	103±48	15.1±7.1	424	457 465 470 478	4.26	4.1 4.3 4.6 4.7
126-63-MLS^3 126-63-MLS^4 126-63-MLS^5 126-63-MLS^6	0.126	1.2	0.063	0.6	103±48	10.0±4.7	424	503 511 520 527	4.26	4.1 4.5 4.8 4.9
63-126-MLS^3 63-126-MLS^4 63-126-MLS^5 63-126-MLS^6	0.063	0.6	0.126	1.2	103±48	19.2±9.0	600	604 605 605 609	6.02	4.3 5.6 6.2 6.3
84-126-MLS^3 84-126-MLS^4 84-126-MLS^5 84-126-MLS^6	0.084	0.8	0.126	1.2	103±48	14.3±6.7	600	605 607 610 613	6.02	4.4 5.7 6.3 6.4
126-126-MLS^3 126-126-MLS^4 126-126-MLS^5 126-126-MLS^6	0.126	1.2	0.126	1.2	103±48	9.6±4.5	600	615 620 630 638	6.02	4.6 5.8 6.3 6.5
63-189-MLS^3 63-189-MLS^4 63-189-MLS^5 63-189-MLS^6	0.063	0.6	0.189	1.8	103±48	18.1±8.5	734	729 735 738 744	7.35	4.9 6.3 7.0 7.2
84-189-MLS^3 84-189-MLS^4 84-189-MLS^5 84-189-MLS^6	0.084	0.8	0.189	1.8	103±48	13.6±6.4	734	730 735 740 745	7.35	4.8 6.5 7.2 7.3
126-189-MLS^3 126-189-MLS^4 126-189-MLS^5 126-189-MLS^6	0.126	1.2	0.189	1.8	103±48	9.0±4.7	734	730 739 743 748	7.35	4.8 6.5 7.3 7.5
63-253-MLS^3 63-252-MLS^4 63-252-MLS^5 63-252-MLS^6	0.063	0.6	0.252	2.4	103±48	17.1±8.0	848	841 845 850 853	8.53	5.0 6.8 7.5 8.4
84-253-MLS^3 84-252-MLS^4 84-252-MLS^5 84-252-MLS^6	0.084	0.8	0.252	2.4	103±48	12.8±6.0	848	840 845 850 855	8.53	5.0 6.9 7.7 8.5
126-252-MLS^3 126-252-MLS^4 126-252-MLS^5 126-252-MLS^6	0.126	1.2	0.252	2.4	103±48	8.6±4.0	848	855 859 863 865	8.53	5.2 7.0 7.8 8.7

3.2.4 Random-Walk/Giant-Loop (RW/GL) Model

In the RW/GL model big loops are assumed to be attached to a non-DNA backbone with no closed loop base (Sachs *et al.*, 1995; (1.3.7), Fig. 1.14). Here the RW/GL model was simulated with loops attached at basepoints, which are connected by a linker. The loop size was 126, 252, 504 kbp or an average of 1.32±0.6 Mbp. The latter correspond to the mean size of an ideogram band. The loops were either attached evenly spaced along the backbone of the decondensed chain of spheres (3.2.2) or in the case of 1.32±0.6 Mbp loops, was obtained by opening all but one loop base in the corresponding MLS rosette (3.2.3). The second method is also closer to a real decondensation process from metaphase into interphase. For loops smaller than 500 kbp the RW/GL term 'giant' seems, inappropriate and is more similar to a Pienta & Coffey (1984) like model of interphase organization (1.3.4, Fig. 1.11 A&B). In the first case. the number of loops was inversely proportional to the loop size and resulted in 984, 492 and 246 loops and linkers sizes of 0.013, 0.026 and 0.052 kbp for a mean sized chromosomes with 137±61 Mbp. In the second case, the number of loops and their spacing equalled that of the number of interphase ideogram bands or rosettes (3.2.3). Therefore, the mean extension of the chromosome territories was defined in the decondensation setting. Rounding effects which could have resulted in a loss of base pairs were avoided by subtraction or addition of loops to chromosomes as needed. Throughout Chapter 3 the following nomenclature L_S-RW/GL$^{r\text{-nucleus}}$ (L_S: loop size; r-nucleus: nuclear radius) is used.

3.2.5 Excluded Volume and Nuclear Volume Properties

For the simulated annealing and the decondensation from metaphase into interphase a high ($U_0 = 2.0kT$), and for the decondensation/relaxation of the detailed folding of the 30 nm chromatin fiber a low ($U_0 = 1.0kT$) excluded volume interaction was used. Thus, the probability of chain crossing was increasing, which could reflect the activity of chain crossing mediated by Topoisomerase-IIα and Topoisomerase-IIβ (Gasser *et al.*, 1986; Sikorav & Jannink, 1994; Duplantier *et al.*, 1995; Jannink *et al.*,1996; Nitiss, 1998; Berger, 1998; Knopf & Waldeck, 2001). The height of the boundary potential was always high with $U_0 = 2.0kT$. The even lower $U_0 = 0.1kT$ potential used in Chapter 2 resulting in a much faster relaxation process was also tested in all steps of relaxation: The initial decondensation process was slowed down due to lower entropic repulsion. However, the relaxation/decondensation of rosettes into giant loops of the RW/GL model and their intrusion into neighbouring chromosome territories was enhanced. A variety of other excluded volume setups were tested having no influence on the final results. The might, however, lead to a different behaviour concerning the dynamics of the chromatin fiber, which was not investigated here.

Tab. 3.2 Simulated Nuclei and Physical Properties of the Used RW/GL Chromosome Model
The number of bands are the number of subcompartments or loops per chromosome for the corresponding chromosome model. The mean band distance is $\langle R_B \rangle = \sqrt{(300nm)^2 \cdot LI_L/(300nm)}$ and the mean territory size is $\langle R_{Ltotal} \rangle = \sqrt{(300nm)^2 \cdot (NB-1) \cdot LI_L/(300nm)}$.

Model	Loop properties		Linker properties		Mean band properties		Mean band distance		Mean territory size	
	Size L_S [Mbp]	Length L_L [µm]	Size LI_S [Mbp]	Length LI_L [µm]	# of bands NB	# of loops per rosette	Theoretic $< R_B >$ [nm]	Simulated $< R_B >$ [nm]	Theoretic $< R_{Ltotal} >$ [µm]	Simulated $< R_{Ltotal} >$ [µm]
126-RW/GL^3	0.126	1.2	0.013	0.120	984±436	-	197	199	6.02	4.0
126-RW/GL^4								205		5.5
126-RW/GL^5								210		6.0
126-RW/GL^6								215		6.1
252-RW/GL^3	0.252	2.4	0.026	0.250	492±218	-	273	268	6.02	4.2
252-RW/GL^4								270		5.5
252-RW/GL^5								280		6.1
252-RW/GL^6								283		6.2
504-RW/GL^3	0.504	4.8	0.052	0.500	246±59	-	387	394	6.02	4.1
504-RW/GL^4								395		5.7
504-RW/GL^5								400		6.2
504-RW/GL^6								400		6.2
1320-RW/GL^3	1.32 ±0.62	12.6 ±5.9	0.126	1.2	103±48	-	600	600	6.02	4.9
1320-RW/GL^4								608		6.6
1320-RW/GL^5								610		6.9
1320-RW/GL^6								613		7.6

3.3 Morphology and General Properties of Simulated Nuclei

To investigate the morphology and general properties on the scale of whole nuclei, which result from the folding of the 30nm chromatin fiber according to the different RW/GL and MLS models, the fiber configuration was rendered, electron microscopic (EM) and confocal laser scanning microscopic (CLSM) images were simulated, and the radial mass distribution and the intensity mass distribution in CLSM images were determined:

3.3.1 Rendering and Simulation of EM and CLSM Images

The morphology at the highest possible resolution was portrayed by rendering the chain segments as cylinders with spherical ends and a 30nm diameter using the program POVRay-V3.0. Due to the 1.2x10^6 segments in a nucleus, only visible segments in an outer spherical shell, whose width depended on the nuclear radius, were used. Single chromosomes were painted with one colour for each chromosome species, not differentiating between homologous chromosomes.

To visualize the chromatin distribution in nuclei with high resolution, electron microscopic images were calculated. The nuclei were placed in a three-dimensional

grid, whose grid spaceing corresponded to the electron microscopic resolution. To map the chain segments into the grid, the cylindrical segments were split using a cylindrical parametrization with 1nm resolution. Then the mass was plotted into each grid position, summarized and finally normalized to a 0 to 256 intensity range. For two-dimensional images the central section of a nucleus was visualized. Chromosomes were painted with one colour for each species and the colour saturation was changed to differentiate homologous chromosomes.

For comparison with experimental chromatin distributions (Chapter 7) three-dimensional confocal laser scanning microscopic (CLSM) image stacks were calculated: The nuclei were placed in a three-dimensional grid with a resolution of 70x70x70nm or 80x80x80nm. This corresponds to an oversampling of 2-3 in lateral and 3-9 in axial direction in agreement with standard experimental CLSM setups. To map the cylindrical segments to this grid of pixels, they were split using a cylindrical parametrization of 5nm resolution. For convolution with the point spread function (PSF), the cylindrical parametrization was replaced by a three-dimensional Gaussian distribution: According to the theoretic resolution of a 100x 1.4 oil immersion PL APO objective (7.2.5) and the more realistic resolution of a 60x 1.2 water immersion PL APO objective, the lateral focal width at half the maximum $FWHM_{x,y}$ was set to 139 or 240nm and axially $FWHM_z$ to 236 or 720nm, respectively. The Gaussian distribution was parameterized with a Cartesian one of 10nm resolution and its values put into the pixel grid. The total pixel intensity was summarized and normalized to a 0 to 256 intensity range. For a two-dimensional image the central section of a nucleus was visualized. The same colour coding for the chromosomes as for electron microscopic images was applied.

3.3.2 Morphology of Simulated Nuclei

The semi-quantitative morphologies by rendered, simulated electron microscopic and confocal laser scanning imaging (Fig. 3.2, Fig. 4.1, Fig. 5.1, Fig. 5.2) were consistent with the folding of the 30nm chromatin fiber or the RW/GL and MLS models. The morphologies are in agreement with the parameters discussed in Chapter 3 and with experiments:

In the MLS model chromosome territories form. Their distinctiveness depends on the interplay between the size of the loops, linkers and nuclei. This is shown by rendering, electron and confocal laser scanning microscopic maps of chromosomes (Fig. 3.2A-Cαγφ, Fig. 4.1B, Fig. 5.1, Fig. 5.2): Usually chromosome territories are compact (Fig. 3.2AαγφI&Bαγφ&Cαγφ, Fig. 4.1B, Fig. 5.1, Fig. 5.2AI&B &C). The larger the linker in respect to the nuclear radius the larger the fingering of territories. This is most apparent in nuclei with 3μm radius (Fig. 3.2AαγφII&III, Fig. 5.2AII&III). Compact territories exist most likely for 63 and 126kbp linkers even in small nuclei. This agrees with the predictions from Chapter 2, with FISH experiments (Zirbel *et al.*, 1993; Monier *et al.*, 1997; Dietzel *et al.*; 1997; Knoch *et al.*, 1998; Knoch *et al.*, 1998; Rauch, 1999; Knoch *et al.*, 2000; v. Hase, 2000;

Kreth, 2001; Habermann, 2001; Cremer, 2001; Tanabe *et al.*, 2002; Rauch *et al.*, 2002) and *in vivo* experiments by BrdU replication labelling (Berezney *et al.*, 1995; Zink *et al.*, 1998; Zink *et al.*, 1998; Zink *et al.*, 1999; Bornfleth, 1999). The MLS rosettes create distinct subcompartments within the chromosome territories and are visible by rendering, in electron microscopic and confocal laser scanning microscopic images (Fig. 3.2 A-Cαβδε, Fig. 4.1 A). In the latter they appear as globules. The chromatin morphology depends again on the interplay between the size of the loops, the linker and the nucleus. Even small changes in these parameters are clearly visible: e. g. for loop sizes of 63 to 126 kbp (Fig. 3.2 A&Bαβδε, Fig. 5.2 A&B, or the linker sizes of 63 to 126 kbp (Fig. 3.2 Aαβδε I&II, Fig. 5.2 AI&II). Remarkably, the loop base points and rosette centres appear also in EM images as dark spots. The overlap of chromosome territories and subcompartments is low at least for nuclear diameters ≥ 8 µm in agreement with the upper morphology and experiments (Münkel & Langowski, 1998; Münkel *et al.*, 1999).

In the RW/GL model also chromosome territories with a very homogeneous morphology form (Fig. 3.2 D, Fig. 5.2 D. The distinctiveness of the territories is lost the bigger the loops, since 1.32±0.6 Mbp loops intermingle freely. The morphology does not show distinct features, despite the speckle like loop bases and statistical imbalances (Fig. 3.2 D, Fig. 5.2 D; here an unrelaxed nucleus is shown to demonstrate the effects better). The large intermingling loops lead to high territory overlap, in contrast to the MLS model and experiments. However, 126 kbp small loops do not overlap like big ones (although the chain of loops as a whole intermingles) and form distinct territories with the same overlap dependencies as the MLS model.

In summary, it is very well possible to investigate with light microscopy even small changes in the three-dimensional organization of nuclei, despite the invisibility of the single chromatin fibers due to the resolution limit (Chapter 7). Of course, the more general the organizational changes in real nuclei, the easier they are apparent qualitatively. Thus, pathological diagnoses of e. g. cancer, based on the nuclear morphology are due to structural changes on the chromatin level.

Beyond the morphology of chromatin, also the morphology of the nucleoplasma, i. e. the space between the chromatin fibers, depends on the linker, loop and nuclear size. The chromatin fiber spacing qualitatively is at least 50 to 100 nm (Fig. 3.2, Fig. 4.1, Fig. 5.1, Fig. 5.2; see also Fig. 2.1) and agrees with estimates of the chromatin volume fraction of 4.4 to 9% in a 5 µm nucleus and a mean fiber distance of 63 to 90 nm (Chapter 5). These voids allow high accessability to the interior of these territories for tracers of corresponding size. Therefore, the definition of the surface of a chromosome territory depends on the scale of the probing tracers or observation. While, for a large particle with 500 nm diameter the chromosome territory is inpenetratable, for particle diameters < 10 nm the chromatin fiber itself is the surface of chromosomes. Nevertheless, it is possible to imagine the embedding hull around a chromosome territory possibly defined by chemical markers. This pure morphologic considerations reveal that small molecules, mRNA or typical proteins could reach nearly every location in the nucleus by only moderately obstructed diffusion.

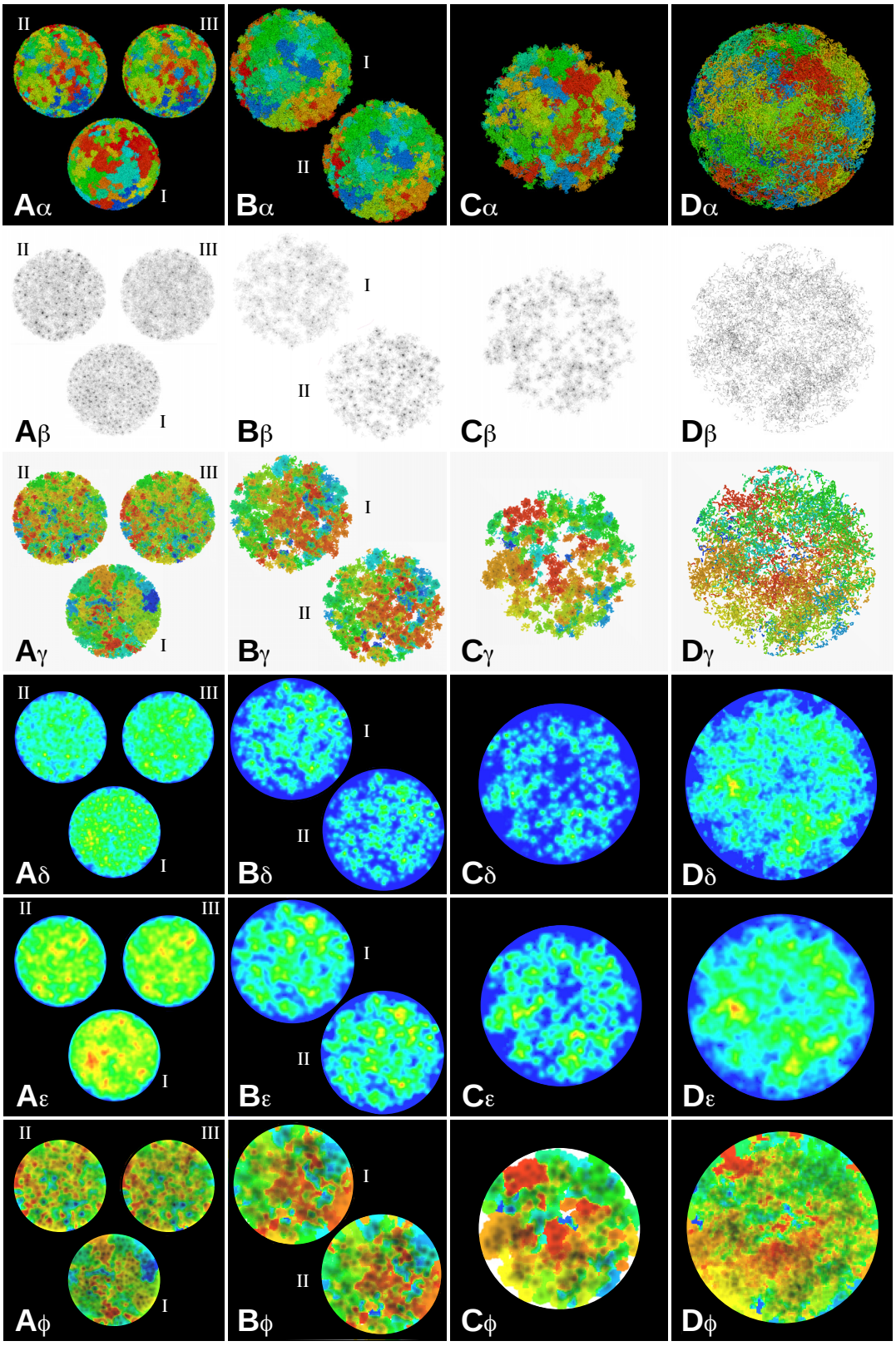
This is in agreement with the simulation of the diffusion of particles in nuclei and experiments (Chapter 5). The Inter-Chromosomal Domain (ICD) model (1.3.5, Fig. 1.13) proposes a channel like network in which the transport of particles takes place. Since a channel like network does not exist between the chromosome territories of simulated nuclei and due to the upper results, this hypothesis seems oversimplified and could possibly be refuted. In summary, the nucleus seems be optimized for low energy consumption necessary for transport. The chromatin fiber distribution presumably leads here to a suble guiding of diffusing particles.

3.3.3 Radial Mass Distribution of Nuclei

To analyse average nuclear properties, the nuclear radial mass and density distributions were calculated as the number of segments in spherical shells (width 50 nm) centered at the nuclear center of mass as function of the shell radius and for the density divided by the shell volume. The average was taken of ten nuclei.

The height of the plateau as well as the extension of the radial mass and density distribution is mainly proportional to the nuclear radius and agrees with the theoretic expectation (Fig. 3.3 A&B). The slope of the decay from the plateau of the density and from the maximum of the mass distribution is also proportional to the nuclear radius. This indicates that the internal pressure generated by the excluded volume interaction between the chromatin fiber is inversely proportional to the

Fig. 3.2 Morphology of Cell Nuclei by Rendering, EM Images and CLSM Images
Distinct differences in the morphology of chromosome models 63-63-MLS^3 (**AI**), 63-252-MLS^3 (**AII**), 126-252-MLS^3 (**AIII**), 126-126-MLS^4 (**BI**), 84-126-MLS^4 (**BII**), 126-126-MLS^5 (**C**), 1320-RW/GL^6 (**D**), are visible in the three-dimensional rendering of the actual 30 nm chromatin fiber (**α**), the electron microscopic (EM) image (**β**), the EM image chromosome map (**γ**; homologous chromosome painting legend below), the confocal laser scanning microscopic (CLSM) image with a high resolution 100x 1.4 oil immersion PL APO objective (**δ**), the CLSM image with a lower resolution 60x 1.2 water immersion PL APO objective (**ε**; colour intensity coding below), and the CLSM image chromosome map (**ϕ**): The rosettes of the MLS model form subcompartments being visible as separated organizational and dynamic entities (**A**, **B**, **C**). Their constituative loop size and number properties are visible and differ between 63 and126 kbp loops even in nuclei with very small radii of 3 µm (**AII**, **AIII**), as well as for the smaller difference between a 84 and 126 kbp loops (**BII**, **BI**; is also be visible in 3 µm nuclei). 63 and 252 kbp linkers could change the distinction between subcompartments using a constant loop size (**AI**, **AII**), being most effective in small nuclei. Distinct chromosome territories form the better, the smaller the linker between subcompartments and depending on the nuclear radius (**AI**, **B**, **C**; Tab 3.1). In contrast, the large loops of the RW/GL model intermingle freely (even more in 3 µm small nuclei), neither forming distinct features like in the MLS model (**D**), nor forming clearly separated chromosome territories due to high overlap (Fig. 3.6&8). Using small loops of 126 kbp would neither intermingle freely, nor form distinct subcompartments, although low overlapping chromosome territories form as expected from the MLS models (Fig. 3.6&8) and from the simulation of single chromosomes (Fig. 2.1 C). It should be noted that the EM and CLSM images were normalized to highest intensity in each image, thus not representing absolute intensities.


II
III
I
Aα
I
II
Bα
Cα
Dα
II
III
I
Aβ
I
II
Bβ
Cβ
Dβ
II
III
I
Aγ
I
II
Bγ
Cγ
Dγ
II
III
I
Aδ
I
II
Bδ
Cδ
Dδ
II
III
I
Aε
I
II
Bε
Cε
Dε
II
III
I
Aϕ
I
II
Bϕ
Cϕ
Dϕ

nuclear radius. Therefore, in big nuclei the chromatin is pressed less against the nuclear boundary. The behaviour is also proportional to the linker size between subcompartments of the MLS (Fig. 3.3A&B) and the loop size of the RW/GL model, since larger linkers or loops lead to a more isotropic packaging of chromosomes. This contrasts the viewpoint that the higher compactness of chromosome territories, due to small linkers, leads to higher exclusion of territories pushing them more to the nuclear border.

The nuclear radial mass and density distributions were, of course, independent of the starting configuration. Calculation of the distributions for specific chromosomes, i. e. measuring their location in respect to the nuclear center (e. g. for the big chromosome I or the small chromosome Y) yielded the same results as above for starting-configurations with randomly arranged chromosomes. However, the mean radial mass distribution for chromosomes arranged randomly in a metaphase startingconfiguration is proportional to the size of the chromosomes (Fig. 3.3C). This is due to the interiour location of small chromosomes compared to large chromosomes in the metaphase plate and the corresponding nuclear localization (Fig. 3.1C). Thus, the bigger the chromosomes the more they are localized at the nuclear membrane in interphase. Qualitative analyses of the decondensation process (Fig. 3.1D) reveals slow diffusion of decondensating chromosomes. This not only supports the upper result but also shows, that chromosomes keep their relation from the startingconfiguration to interphase. The results agree with the unrandom localization of chromosomes found in recent experiments mapping interphase chromosomes by fluorescence *in situ* hybridization (FISH; v. Hase, 2000; Kreth, 2001; Habermann, 2001; Cremer, 2001; Tanabe *et al.*, 2002). Beyond, the simulations show that already a biased positioning in the metaphase plate could explain this effect. This does not exclude special constraints holding or dragging the chromosomes in place during interphase. Such constraints might be necessary, to account for diffusion of the chromosomes destroying this order over long times scales (e. g. hours or days). Whether this is the case, can not be answered from the simulated decondensation process, since here the chromosomes are represented by spheres with a smooth gliding surface in contrast to e. g. MLS rosettes. Therefore, the slow diffusion of spheres could hardly be extrapolated to rosettes. Nevertheless, the results reveal already that not many additional features were needed to hold chromosomes effectively in place.

3.3.4 Intensity and Mass Distribution in CLSM-Stacks of Nuclei

To analyse the nuclear morphology on the level of the simulated confocal laser scanning microscopic images the absolute intensity and mass distribution as function of the nucleosome concentration were calculated. The intensity distribution provides also the pixel or nuclear volume fraction at the given chromatin concentration and thus the total number of nucleosomes at a particular concentration or within a pixel. The unnormality of the mass distribution was tested by fitting to a bimodal Gaussian distribution. The distributions were normalized and averaged over ten nuclei.

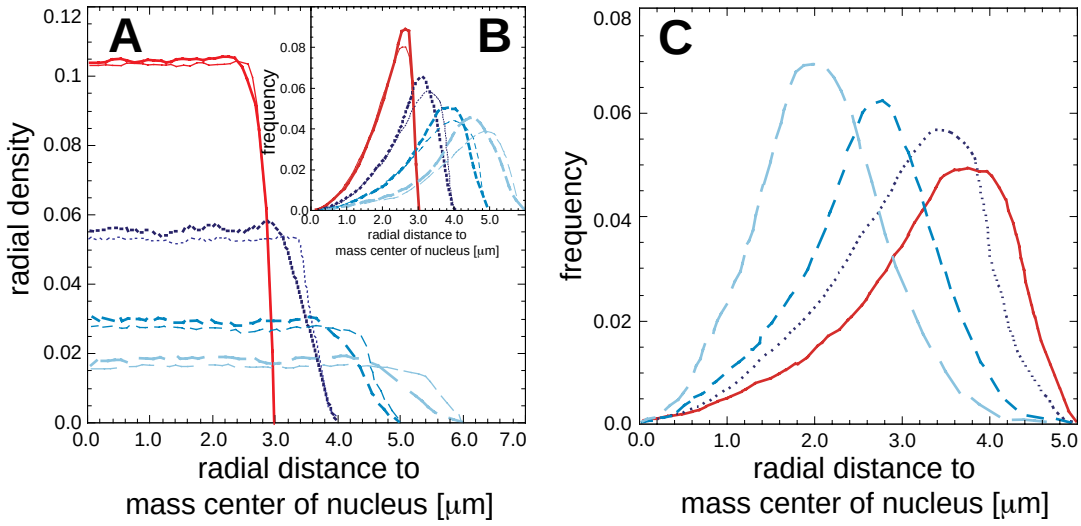


Fig. 3.3 Radial Mass and Radial Density Distribution of Nuclei and Chromosome Location
The plateau and extension of the radial density distribution of nuclei (**A**) and the extension of the radial mass distribution (**B**) are proportional to the nuclear radius (126-63-MLS^X, thick lines; radii of 3, 4, 5, 6 µm are, solid, dotted, dashed, long dashed). The radial density distribution is also proportional to the linker length (**A**, **B**, 126-252-MLS^X, thin lines). Randomly arranged chromosomes in the startconfiguration resulted in the same radial mass distribution for the single chromosomes in respect to the nuclear center of mass (**A**, **B**; Fig. 3.1A&B). The radial mass distribution of chromosomes arranged in a metaphase plate in the startconfiguration (Fig. 3.1C), however, is proportional to the size of the chromosome (**C**): small chromosomes locate preferentially in the inner part of the nucleus as in the metaphase plate and big ones locate in the outer part (126-126-MLS^5, chromosomes I, VIII, XV and Y are solid, dotted, dashed, long dashed).

In the MLS model the absolute intensity distribution, the position and deviation of its peak from the average nuclear chromatin concentration are proportional to the nuclear radius and the loop and linker sizes (Fig 3.4A&B). The height of the saddle leading to the peak of the distribution is inversely proportional to these parameters. The saddle indicates the low and border intensity regime of the nucleus. For 3, 4, 5 and 6 µm nuclear radii, the mean nucleosome concentrations are 251, 107, 55.3 and 31.5 µM in agreement with the distribution averages and the peak positions at 40, 70, 160 and 420 µM. Thus, the denser a nucleus, the more volume is occupied by the high density fraction, since the MLS rosettes are packed closer. The highest densities are populated equally in all nuclei, depending only on the rosette structure. The center density of rosettes is loop size proportional. Thus, in a nucleus with 6µm radius the intensity distribution approximated that of free subcompartments. Since the RW/GL model has a very homogeneous morphology due to the large intermingling loops, the intensity distribution exhibits always a clear peak around the average nucleosome concentration, with a smaller width at e. g. half the maximum. Since the RW/GL model lacks structures like rosettes the highest concentrations are less populated. The intensity distribution is also proportional to the microscopic resolution: the distribution is shifted to higher concentrations and shows more pro-

nounced peaks in bigger nuclei due to the smaller smearing out of structures for a high 100x than for a low 60x resolution objective. The above described proportionalities were also found for the mass distribution (Fig. 3.4C) and confirm the nuclear volume dependencies (3.3.2, Chapter 5). The mass distribution is only bimodal for the MLS model (Fig. 3.5D). The fraction of the mass described by the Gausian nearer to the saddle is proportional to the saddle behaviour and quantifies the relation between chromatin, the low and border intensity regime, and the nucleoplasma.

The comparison of the absolute distributions to standard CLSM experiments, is compromised by the objective resolution, image reconstruction and threshold settings and most importantly the relationship between intensity and nucleosome concentration. A comparison to a recent experimental quantification of the absolute intensity and mass distributions by fluorescent fluctuation microscopy (Weidemann *et al.*, 2002) suggest good agreement for nuclei with 4 to 5 µm radius and an MLS model. Whether the found bimodality of the mass distribution could be attributed to the mass/nucleoplasma or to two states of chromatin organization remains unclear.

3.4 Properties of Chromosome Territories

3.4.1 Radial Mass and Density Distribution of Territories

To analyse of average properties of chromosomes the radial mass distribution and the radial density was calculated from the number chain segments in spherical shells (width 5 nm) centered at the chromosomal center of mass. The radial density is the number of segments in a shell divided by the volume of this shell. For single chromosome species the average was taken over the two homologous chromosomes in ten nuclei. For an overall investigation the average was taken over all chromosomes.

Averaging over all different sized chromosomes, revealed in all MLS models a plateau in the radial density up to radi of ~1 to 2 µm (Fig. 3.5A). Its height and the peak height of the radial mass distribution are inversely proportional (Fig. 3.5B), and the average extension is proportional for both distributions to the linker length between the subcompartments. Changes of the loop size did only effect this behaviour implicitly due to the influence of loops on the distance between succeeding subcompartments (2.6.2, 3.5.2). Since the loop size change is relatively small, it is largely averaged out in comparison to the territory extension. Of course, the radial mass and density distribution is also proportional to the chromosome size (Fig. 3.5C). Dueto their since clustering into groups (Fig. 1.9) the decay of the radial density reveals a steplike behaviour. The territory extension is also proportional to the nuclear radius with the degree of influence being inversely proportional to the nuclear radius and proportional to the linker length (Fig. 3.5D). Thus, for small linker length of 63 kbp the territory extension is not influenced much for all nuclear radii in contrast to linker length of 850 kbp, where the nuclear radius plays a

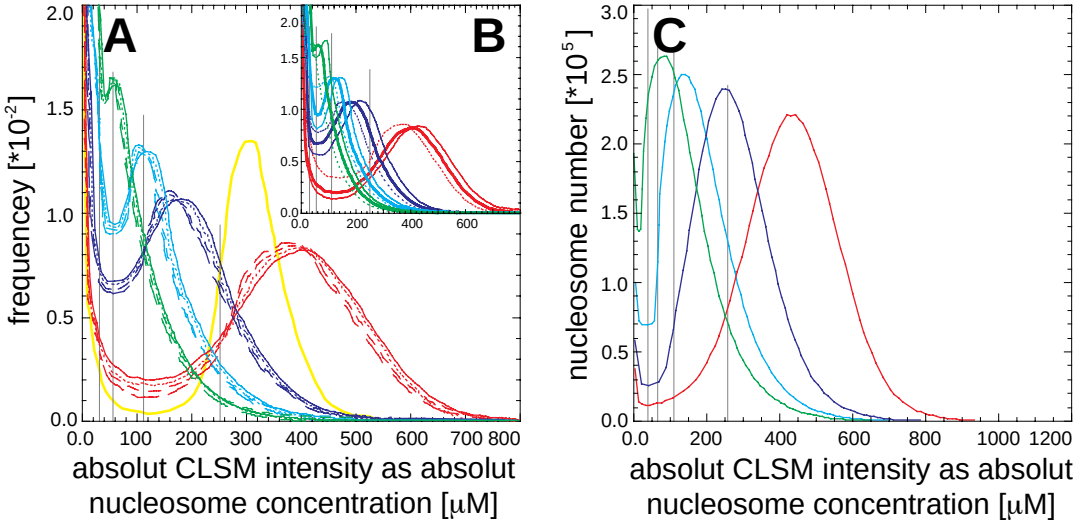


Fig. 3.4 Absolute Intensity and Mass Distributions of 3D-CLSM Image Stacks
The absolute intensity distribution, its peak, the peak deviation from the mean nuclear density (upright lines) are proportional. The saddle height indicating the low and border intensity regime is inversely proportional to the nuclear radius (**A**, 126-X-MLS^X, linkers of 63, 126,189 and 252 kbp are solid, dotted, dashed, and long dashed, nuclear radii of 3, 4, 5 and 6 µm are red, blue, light blue and green, 1320-RW/GL^3 is thick yellow). The intensity distribution is also proportional to the linker (**A**) and loop size (**B,** X-126-MLS^X, loops of 63 and 126 kbp are dotted and thick line) and inversely proportional to the objective resolution (**B**, for 126-126-MLS^X 100x and 60x objective are thick and thin lines). The nucleosome number distribution as function of the absolute chromatin density follows the same proportionalities and is bimodal (**C**, 126-126-MLS^X as in **A**, lower fraction dotted, upper fraction dashed). The bimodality is inversely proportionality to the saddle behaviour. (**D**, for MLS models, linkers marked as radii in **A**, loop sizes 63, 84 and 126 kbp are thick, normal and thin).

major role up to 5 µm. It should be noted, that in the MLS and the RW/GL model the territory extension is only defined by the linker and that realistic linker length should have mean sizes of 63 to 126 kbp in comparison with experiments. The existence of a plateau in all chromosomes can therefore be explained by the interplay between the excluded volume interaction, the repulsive entropy of fast fluctuating rosette loops and the linker pulling the rosettes together with a minor influence of the nuclear radius. For a comparison between theoretic and simulated territory extensions see Tab. 3.1.

The radial density of the RW/GL models shows only a rudimentary plateau and much slower density decrease depending on the loops size (Fig. 3.5 A), although the effect is not as big as in the case of single chromosomes (2.5). The radial density and the peak height of the radial mass distribution as a whole are again inversely proportional, and the mean extension of the radial mass distribution is proportional to the length of the linker between this time the loops. In addition a proportionality to the loop size exists mainly for the 1.32±0.6 Mbp loop size. Thus, the territory extension is created by the interplay between the total linker length and loop size as in the case for the simulation of single chromosomes (Chapter 2). As the sum of the linker length was kept constant (3.2.4) the reason for the slow radial density

decrease and the rudimentary plateau is the connection between the intermingling loops of 1.32±0.6Mbp and the with small linkers compared to big loop size. Thus the loops are more densely packed near the linkers while stretching out further than the linkers extent. This agrees with the explanation of the plateau for MLS models. For a comparison between theoretic and simulated territory extensions see Tab. 3.2.

3.4.2 Roundness of Territories

To analyse the shape of the territories with respect to the deviation from a sphere, the roundness R was calculated using the square roots of the eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$ of the tensor of inertia with

$$R = \frac{\lambda_2 \lambda_3}{\lambda_1^{\,2}}. \tag{3.2}$$

For single chromosome species the average was taken over the two homologous chromosomes in ten nuclei.

The roundness of chromosomes was mainly inversely proportional to their size and the nuclear radius, and proportional to the linker size (Fig. 3.5E&F). Thus, the roundness not only behaves as the theoretic prediction permitting more three-dimensional states for longer or more unconstrained configurations, but also shows similar proportionalities as the radial mass and density distributions (3.4.1). Loop size changes effected this behaviour like the distance of succeeding subcompartments (2.6.2, 3.5.2). Only the big 1.32±0.6Mbp loops in the RW/GL models lead to rounder territories than expected, since these loops could distribute isotropically and covering the unisotropy of their connecting backbone.

3.4.3 Spatial Distance between Arbitrary and Nearest Chromosome Territories

To investigate the relationship between chromosome parameters and nuclear size, the spatial distance distribution of chromosomes and the distance to the nearest neighbouring chromosome territories were calculated. The distance to the spatially nearest, the second nearest etc. subcompartment, was caluculated from the distances between the center of mass of one subcompartment to all the other subcompartments and sorting of these distances. The mean was taken over ten nuclei.

In all simulations the mean distance between arbitrary chromosome territories is mainly proportional to the nuclear volume (Fig. 3.3G). For nuclear radii of 3, 4, 5 and 6µm the means were 2.4, 3.3, 4.2, 5.0µm and their deviation from the radius was inversely proportional to the radius. This effect is due to the distance of the center of mass of the territory to the nuclear border and therefore proportional to the territory extension, which depends in turn on linker and loop sizes of the model (3.4.1). The mean distance between arbitrary territories is consequently also proportional to the linker size between subcompartments (Fig. 3.3G), since the linker leads

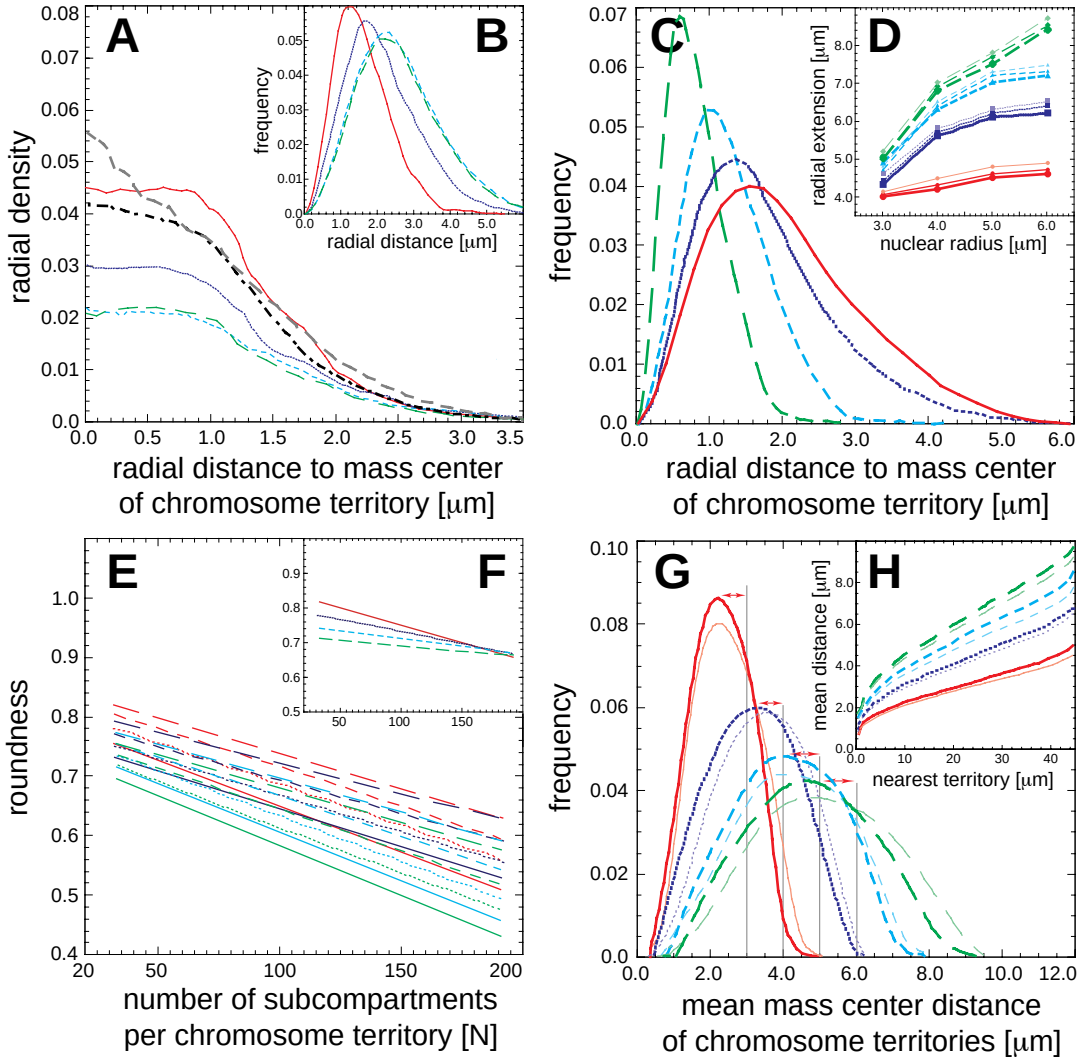


Fig. 3.5 Properties of Chromosome Territories in Cell Nuclei
Radial density (**A**) and mass (**B**) distributions are proportional to the linker size (X-126-MLS^5, linkers of 63, 126,189 and 252 kbp are solid, dotted, dashed, and long dashed) and implicitly also to the loop size extending the linker (**D**). The same proportionality holds for 126-RW/GL5 and 1320-RW/GL5 models (thick dash dotted and thick long dashed), although the extension is explicitly depending to the loop size. The radial density shows a plateau for the MLS in contrast to RW/GL models connected to loop arrangement and size. The extension of territories is proportional to the chromosome size (**C**, 126-126-MLS^5, chromosomes I, VIII, XV and Y are solid, dotted, dashed and long dashed) and proportional to the restricting nuclear radius (**D**, for MLS models, linkers marked as in **A**, loop sizes 63, 84 and 126 kbp are thick, normal and thin). The roundness of chromosome territories is inversely proportional to the chromosome and nuclear size and proportional to the linker size (**E**, 126-X-MLS^X, linkers as in **A**, **B**, 3, 4, 5, 6 μm radius are red, blue, light blue and green) and in the RW/GL model inversely proportional to the loop size (**F**, 1320-RW/GL^X) The distances distribution between arbitrary chromosomes (**G**) is proportional to the nuclear radius and the deviation between the mean and the radius is inversely proportional to the nuclear radius. Its mean is also slightly proportional to the linker length even in small nuclei (126-63-MLS^X thick, 126-252-MLS^X thin lines; 3, 4, 5, 6 μm radius are solid, dotted, dashed, long dashed). The nearest neighbour distances show the same proportionalities (**H**, legend as in **G**).

to larger and more evenly packed chromosomes in interphase nuclei. For the RW/GL model the mean distance is additionally proportional to loop size and is especially pronounced for big loops of 1.32±0.6 Mbp, with mean values similar to those for large linker sizes of 252 kbp (data not shown). Thus, the results agree not only with the theoretical expectation but also with the findings of subcompartment properties (3.5). The sorted nearest neighbour distances behaved exactly as the distance distribution described above (Fig. 3.3H). The inset of the arbitrary distance distribution as well as the first nearest neighbour distances clarify also, that the center of mass of chromosomes could be very close, as the chain of subcompartments or loops of the MLS or the RW/GL model could wound around each other. This is closer than the mean extension of the territories would suggest. This is in qualitative agreement with territory painting by fluorescence *in situ* hybridization.

The distance distribution and the nearest neighbour distances showed, of course, the same proportionalities independent of the starting configuration. Calculation of the detailed values for specific chromosomes (say e. g. between the homologous chromosomes) showed, however, only the same results for startconfigurations with randomly arranged chromosomes. Results for specific chromosomes arranged randomly in a metaphase plate starting configuration agreed with the position of these chromosomes in the nucleus. The mean distance between the homologous chromosomes I, VIII, XV and Y was found to be 4.0, 5.2, 6.2, 6.6 µm for a nucleus with 5 µm radius (3.3.3, Fig. 3.3C).

3.4.4 "Volume" and Overlap of Territories Based on CLSM Images

The volume of a chromosome territory depends on the used measure and equals the volume of the 30 nm chromatin fiber itself for a measure of similar scale. Using the resolution of CLSM images is a possibility for defining the occupied region of a chromosome territory, then being defined as "volume" and calculated as the number of boxes in the three-dimensional grid belonging to one chromosome during the calculation of CLSM image stacks (3.3.1). This is comparable to defining an enveloping surface and calculating the volume inside. Since the exponential decaying Gaussian parameterization led to infinite chromosomal volumes, only pixels >1% of the Gaussian maximum intensity were counted. The chromosome overlap as nuclear fraction then is the sum of boxes belonging to at least two different chromosomes and dividing by the total nuclear volume. Thus, the volume dependence is also a function of the nucleosome concentration distribution (3.3.4). The analyses were averaged over all chromosomes in ten nuclei.

The mean chromosomal volume is mainly proportional to the linker size and the nuclear radius. The degree of influence is inversely proportional to the nuclear radius and proportional to the linker length (Fig. 3.6A), in agreement with the radial mass and density. Loop size changes did mostly affect this behaviour implicitly to the extent that the loop size influenced the distance between succeeding subcompartments (2.6.2, 3.5.2). Of course, the territory volumes represented by their

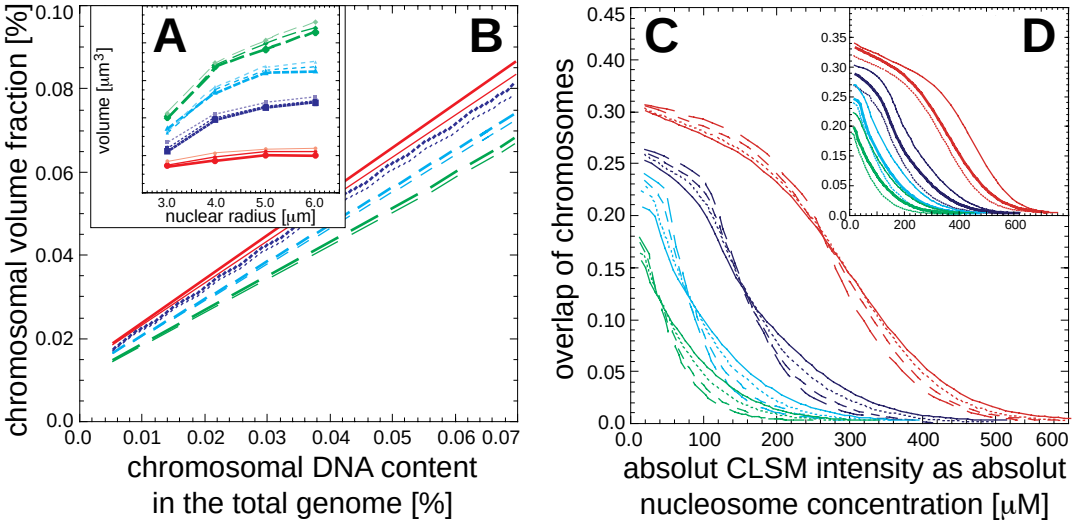


Fig. 3.6 Volume and Overlap of Chromosome Territories
The nuclear volume fraction of chromosome territories is proportional in the first place to the chromosome size (**B**, 126-X-MLS^X models, linkers of 63, 126,189 and 252 kbp are solid, dotted, dashed, and long dashed, nuclear radii of 3, 4, 5 and 6 µm are red, blue, light blue and green) and inversely proportional to the nuclear radius as more space is available and is not too much influenced by the resolution of the objective, itself being proportional to the chromosome size (100x objective thick lines, 60x objective thin lines). The obvious volume dependency on the intensity/density, e. g. by thresholding is in agreement with the intensity/distribution and could be calculated from that relationship (Fig. 3.4). The mean chromosomal volume is proportional to the linker size and the nuclear radius and little influenced by the loop size in the MLS model (**A**, shown for MLS models, linkers of 63, 126,189 and 252 kbp are solid, dotted, dashed, and long dashed, loop sizes 63, 84 and 126 kbp are thick, normal and thin). The overlap of chromosome territories follows the same proportionalities as the volume and decreases for increased intensity/density thresholding (**C**, same legend as in **A** and **B**, for 100x). Since the intensity/density distribution for a high resolution objective is thinner than for a low resolution objective smearing out or broadening given structures, the decay sets in later for the low resolution objective (**D**, 100x thick and 60x thin lines). The overlap is not dramatically influenced by the loop size as in the case for the volume (**D**, 63-126-MLS^X is dotted).

nuclear volume fraction are proportional to the chromosome size, being in that representation also inversely proportional to the nuclear radius due to the increased space available (Fig. 3.6B). Thus, the volume or fraction are also proportional to the objective resolution with value shifts consistent with the other proportionalities (Fig. 3.6B). In the RW/GL model the volume is also proportional to the loop size, being especially pronounced for big loops of 1.32±0.6 Mbp, with mean values similar to those for large linker sizes of 252 kbp. Consequently, the overlap of chromosome territories is consistent with the volume behaviour: the overlap is proportional to the linker and loop size and inversely proportional to the nuclear radius as then more space is available and the chromosome territories are not pressed into each other (Fig. 3.6C&D). Different loop sizes did affect this behaviour implicitly to the extent that the loop size influenced the distance between succeeding subcompartments (2.6.2, 3.5.2). The longer the linker and the more fingered the territory the bigger is this slight effect. In the RW/GL model the overlap is again proportional to

the loop size and especially pronounced for the 1.32±0.6 Mbp big loops. Since for a high resolution objective the intensity/density distribution is thinner than for a low resolution objective, smearing out structures, the overlap is also inversely proportional to the objective resolution and to the chromatin density (Fig. 3.6 C&D).

Despite these proportionalities the evaluation and comparison of the absolute volumes and overlaps to experiments is compromised not only by their obvious dependence on the objective resolution and chromatin density but mainly by the various image reconstruction and threshold settings. Nevertheless, simulations with a high resolution 100x objective and an estimated threshold of ~10 to 15% (normalized to the intensity distribution maximum) of the total intensity range gave reliable comparison results: The volume of an MLS model or a small looped RW/GL model in contrast to a large looped RW/GL model agree with experimental measurements of the volume (Monier 1997). Concerning the overlap a comparison results in best agreement for MLS models in nuclei with 4, but better 5 to 6 µm radius (Dietzel *et al.*, 1997; Münkel & Langowski, 1998; Münkel *et al.*, 1999) and does not agree with RW/GL models with loops bigger than 252 kbp.

3.5 Properties of MLS - Subcompartments

3.5.1 Radial Mass Distribution of Subcompartments

The rosettes of the MLS model form distinct subcompartments having a diameter larger than the resolution limit of light microscopy (Fig. 3.7). Therefore they are visible in the simulation of confocal laser scanning images (Fig. 3.7). The radial mass distribution and radial density were computed as for the whole chromosomes with a shell width of 1 nm for higher resolution. The mean was taken over all subcompartments (~5000) in one nucleus. It should be noted that the varying size of ideogram bands led to different number of loops in the rosettes, although the loop size was kept constant in all rosettes (Tab. 3.1).

For all MLS models the peak height of the radial mass distribution is inversely proportional and the radial extension of the radial mass distribution is proportional to the loop size (Fig. 3.7 A). The mean radial extension of a subcompartment is half the maximum peak height of the radial mass distribution. For 126 kbp linkers and 4 µm nuclear radius, loop sizes of 63, 84 and 126 kbp result in mean rosette radii of 195, 215, and 300 nm leading to mean diameters of 390, 430 and 860 nm, respectively. This agrees with the extension of single loops measured by simulated position dependent and independent spatial distance measurements between genomic markers as function of their genomic separation in the simulation of single chromosomes (2.7.1 and 2.7.2) and whole nuclei (3.6). Only minor (in total ~25 nm) but systematic effects resulting from the inverse proportionality between loop numbers and linkers were observed (Fig. 3.7 A) as already shown in Chapter 2 (2.6.1,

Fig. 2.5A). The mean radial extension is also slightly proportional to the nuclear radius (Fig. 3.7B) and is due to the confinement of chromosomes into a nucleus in agreement with the findings for the mean territory extension. This effect was not seen in the simulation of single chromosomes in Chapter 2, because the confinement was not as restricting as in nuclei.

For comparison with microscopic experiments, a total possible extension of a rosette was estimated for the radial distance for which the histogram frequency and therefore mass probability drops below 10%. This resulted in radii of 255, 290 and 380nm and diameters of 560, 580and 720nm. These values might reflect an upper limit for the comparison with experimental diameters which depend on preparation, image reconstruction and the threshold dependencies for volume determinations. Experimental data from BrdU incorporation reveal a diameter distribution of so called foci with from 400 to 700nm (Berezney *et al.*, 1995; Zink *et al.*, 1998; Bornfleth, 1999). Consequently, an MLS model with loop sizes of 63 to 126kbp is favoured considering the image reconstruction methods used.

3.5.2 Spatial Distance between Succeeding Subcompartments

The distance between succeeding subcompartments depends on the interaction between the linker length, the diameter and the loop size of rosettes. Therefore, the distance between the mass center of succeeding subcompartments was calculated and the values put in a histogram with a 1.0 nm resolution. The mean was taken over all succeeding pairs of subcompartments of all chromosomes in a nucleus.

In all MLS models the mean distance between succeeding subcompartments is proportional to the length of the linker and the loop size (Fig. 3.7C). In nuclei with 5µm radius, linker length of 63, 126, 189 and 252kbp resulted for loop sizes of 63, 84 and 126kbp in a mean distance between succeeding subcompartments of 430, 605, 738 and 850nm, 470, 610, 740 and 850nm, as well as 520, 630, 743 and 863nm, respectively. The theoretical values for these linker length calculated by Equ. 2.4 would be 425, 600, 735 and 850nm and were introduced into the simulation during the decondensation into interphase using spheres (3.2.2). Therefore the observed values differed by 1.1, 0.8, 0.4 and 0.0%, 10.5, 1.6, 0.6 and 0.0%, as well as 22.3, 5.0, 1.7 and 1.5%. Determination of the mean distance between spheres before introducing the detailed folding of the chromatin fiber (3.2.3) and further relaxation revealed in all simulations the theoretic distance, thus the found variance was caused by entropic repulsion of rosettes. The biggest discrepancy is found for a 63kbp linker and a 126kbp loop in agreement with the mean subcompartment extensions and the simulations of single chromosomes. This entropic repulsion supports also the increase of the segment length to 50nm from an initial Kuhn length L_K of 300nm (3.2.2). Additionally, it shows the incapabilities of reducing the subcompartments to mere spheres in simulations of whole cell nuclei during decondensation. The mean distance between succeeding subcompartments depended only to a minor (in total ~20nm) extend on the nuclear radius with variances in the range of

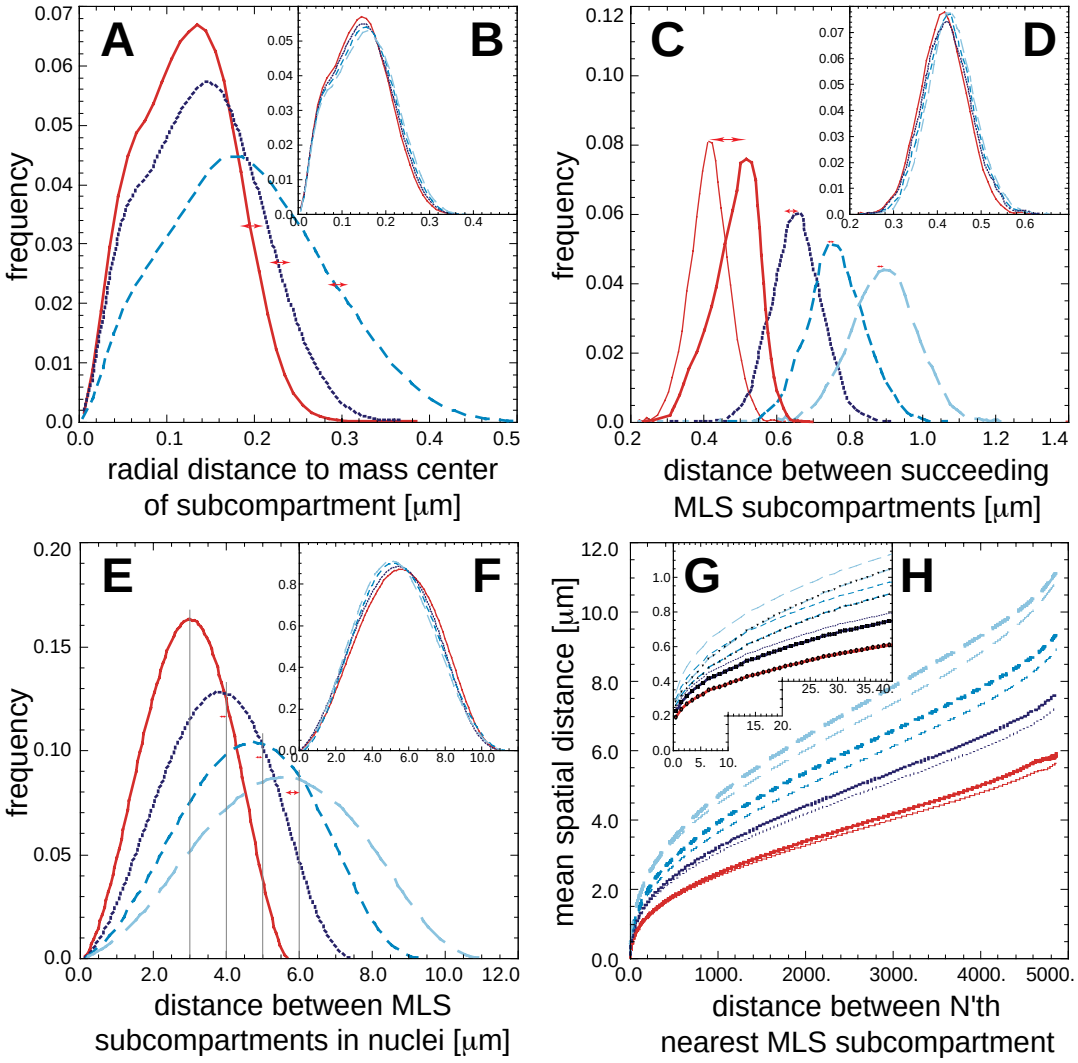


Fig. 3.7 Properties of MLS Subcompartments in Cell Nuclei
The radial extension is proportional to the loops size (**A**, X-126-MLS^4, loops of 63, 84,126 kbp are solid, dotted, dashed), slightly to the linker length (variance: arrows in **A**) and also slightly to the nuclear radius (**B**, 84-126-MLS^X, 3, 4, 5, 6 μm radius are, solid, dotted, dashed, long dashed). The spatial distance distribution between the mass center of succeeding MLS subcompartments is proportional mainly to the linker length (**C**, 126-X-MLS^4, 63, 126, 189, 252 kbp as in **B**), the loop size (the difference from 63 to 126 kbp loops is indicated by arrows and explicitly shown by the thin 63-63-MLS^4) and also little to the nuclear radius (**D**, legend as in **B**, but 63-63-MLS^X). The distance distribution between arbitrary subcompartments is proportional to the nuclear radius (**E**, same models as in **B**) and the deviation between the mean and the radius is inversely proportional to the nuclear radius (arrows in **E**) and proportional to the loop size. The mean is also slightly inversely proportional to the linker length even in big nuclei (**F**, legend as in **C**, but with X-126-MLS^6). The nearest neighbour distances show for distant neighbours the same proportionalities (**G**, shown for all nuclear radii, legend in **B**, and 126-63-MLS^X thick, 126-252-MLS^X thin lines) as the arbitrary distances (**E**, **F**). For close subcompartments the nearest distance is, however, proportional to the linker length and value for the first nearest neighbour distance is smaller than the distance between succeeding subcompartments, depending on parameters (**H**).

the mean chromosome extension variation for the same proportionality (Fig. 3.7D) and thus are not influenced strongly by other neighbouring subcompartments.

3.5.3 Spatial Distance between Arbitrary and Nearest Subcompartments

In nuclei the distance between arbitrary or the nearest neighbours of subcompartments depends mainly on the available nuclear volume with an average density. The linker length, the loop size or the subcompartment diameter are only of minor importance. This is different to the simulation of single chromosomes (2.6.3) in which the neighbouring subcompartment belong all to the same chromosome. Thus, the linker length is also of major importance, if the chromosome could extend freely without a volume restricting constrain. To calculate the distance to the spatially nearest, the second nearest etc. subcompartment, the distances between the center of mass of one subcompartment to all the other subcompartments were computed and then sorted according to the distance. The general distance distribution resulted in the distance distribution between arbitrary subcompartments. The average was taken over all subcompartments in a nucleus in ten nuclei.

In all MLS models the mean distance between arbitrary subcompartments is mainly proportional to the nuclear volume (Fig. 3.7E). For nuclear radii of 3, 4, 5 and 6 µm the means were 3.0, 3.8, 4.7 and 5.6 µm and their deviation from the radius was inversely proportional to the radius. This effect is caused by the distance of the center of mass of a subcompartment to the nuclear border and therefore quite exactly proportional to the loop size and subcompartment radius. The mean distance between arbitrary subcompartments is also slightly inversely proportional to the linker size (Fig. 3.7F). Thus longer linkers lead to better packaging and more evenly distributed subcompartments. The sorted nearest neighbour distances behaved as the arbitrary neighbour distance distribution (Fig. 3.7G). For close nearest neighbours the distance is proportional to the linker length due to the influence of subcompartments lying in the same chromosome (Fig. 3.7H). The first nearest neighbour distances matched those expected from evenly distributed subcompartments and is smaller than the mean distance to the succeeding subcompartment. Since the nearest subcompartments are mostly not connected directly through a linker but by an arbitrary genomic separation, the entropic repulsion of rosettes has here a much smaller effect than for succeeding subcompartments. Since the nearest subcompartments are mostly not connected directly through a linker but by an arbitrary genomic separation, the entropic repulsion has here a much smaller effect than for succeeding subcompartments. The latter is due to the constraining linker, which forces the succeeding subcompartments to interact intensively with each other, while arbitrary subcom-partments might arrange due to favourable energetic minima. Therefore, the nearest neighbour distance could well be smaller than the mean separation of succeeding subcompartments. The analysis of the nearest neighbour distances of the single chromosomes within a nucleus resulted in the same proportionalities and agreed with those found in Chapter 2.

3.5.4 "Volume" and Overlap of Subcompartments Based on CLSM Images

As in the volume and overlap calculation of chromosome territories (3.4.4) the subcompartment volume depends on the used measure and equals the volume of the 30 nm chromatin fiber itself for a measure of similar scale. For calculation the same procedure was used as in 3.4.4, despite averaging over all subcompartmens for subcompartment volumes in ten nuclei.

The volume of subcompartments is like the mean radial mass and density distribution mainly proportional to the loop size (Fig. 3.8 A) and slightly dependent to the linker size and nuclear radius (Fig. 3.8 B). Only minor (in total ~25 nm) but systematic effects resulting from the inverse proportionality between loop numbers and linkers were observed (Fig. 3.7 A, 2.6.2, 3.4, 3.5.2). The overlap is proportional to the loop and linker size (Fig. 3.8 C&D) and in all other respects consistent with the overlap behaviour of whole chromosome territories regarding the higher values due to the additional overlap contribution from within a territory. Since RW/GL loops intermingle the larger their size and as small loops are not forming distinct subcompartments the overlap reaches values >70 %. Consequently, again the comparison of the absolute volume and overlaps to experiments is compromised by the objective resolution, chromatin density, image reconstruction and threshold settings. Nevertheless, with the approximation with a 100x objective and an estimated normalized threshold of ~10 to 15% (normalized to the intensity distribution maximum) gives reasonable results: For MLS models in nuclei with 4, but better 5 to 6 µm radius the subcompartment overlap is in agreement with experimental values (Münkel & Langowski, 1998; Münkel *et al.*, 1999). The experiments disagree with the RW/GL model, even if small loops are considered, since the loops intermingle freely and due not form distinct subcompartments.

3.6 Position Independent Spatial Distances between Genomic Markers

The measuring process of spatial distances between genomic markers as function of their genomic separation reflects different assumptions of the structure, stability and dynamics of chromosome organizations. Position independent measurements of spatial distances reflect chromosome models in which there are no fixed loop and/or rosette sizes. (i. e. a rosette with constant base pair content and defined genomic position) this suggests a chromatin fiber slithering through the "bases" of the loops. These distances reflect also possible variabilities of chromosome organizations between different cells, before or after cell division of one cell, cell cycle dependencies and even random experimental preparation artefacts disturbing or destroying the *in vivo* organization. Consequently, a marker pair could reside in any possible position relative to a loop base or rosette. Thus, the spatial distance is position inde-

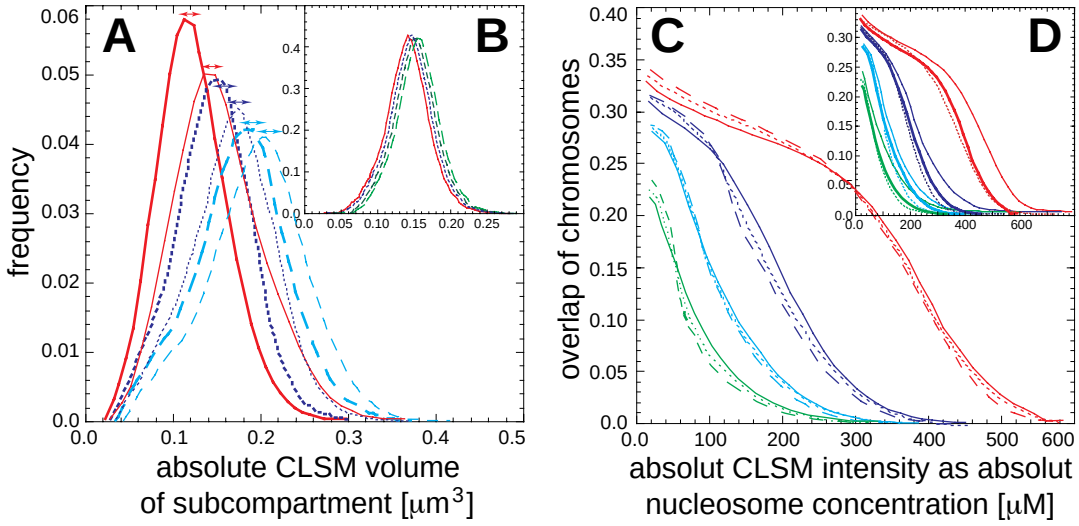


Fig. 3.8 Volume and Overlap of MLS Subcompartments
The subcompartment volume is proportional to the loop size (**A**; X-126-MLS^4, 63, 84,126 kbp loops are solid, dotted, dashed), slightly to the linker size (variance: arrows in **A**) and also slightly to the nuclear radius (**B**; 84-126-MLS^X, 3, 4, 5, 6 μm radius are, solid, dotted, dashed, long dashed). The volume dependency on the chromatin density, e. g. by thresholding is in agreement with the intensity/density distribution and could be calculated from that relationship (Fig. 3.4). The subcompartment overlap is mainly proportional to the nuclear radius (**C**; 3, 4, 5, 6 μm radius are, red, blue, light blue, green) and the loop size (legend as in **A**). It is higher than the chromosome territory overlap (Fig. 3.6). Since the intensity/density distribution for a high resolution objective is thinner than for a low resolution objective smearing out or broadening given structures, the decay sets in later for the low resolution objective (**D**, 100x thick and 60x thin lines). The overlap is hardly influenced by the linker size as in the case for the volume (**D**, 63-252-MLS^X is dotted).

pendent. For calculation, pairs of markers were placed randomly - e. g. with no respect to any folding structures - on the chromosome (Fig. 2.6 A) with genomic separations from 5.2 kbp (the basepair content of one chain segment) up to the whole chromosome. Due to the ~26.000 segments of a mean chromosomes it was impossible to calculate all distances between all marker pairs. Therefore, below 25 Mbp all distances were calculated but above 5000 marker pairs were randomly chosen. The mean was taken over all chromosomes in ten nuclei.

For RW/GL as well as for MLS models the general properties of the position independent spatial distances between markers as function of the genomic separation are equal (Fig. 3.9 A&B) and agree with those from the simulation of single chromosomes (2.7.1): The spatial distance first increases monotonous as expected from a random walk. At genomic separations of half the loop size the increase stagnates in a plateau and shows a local minimum at a genomic separation of one loop size. The spatial distance of the plateau is proportional to the loop size. Due to the random positioning of the marker pair the distance need not decrease to zero: E. g. the distance of a marker pair with a genomic separation of 126 kbp in an MLS rosette with loop sizes of 126 kbp is zero for markers positioned on a loop base. For markers located at the tip of two successive loops, which point in opposite direc-

tions, the distance could be twice the extension of the loops. The spatial distance of the plateau is a little bigger than the mean extension of a loop (Fig. 2.9 in Chapter 2) since also spatial distances of linkers contribute to the average. Therefore, the top of the plateau also shifts slightly to higher genomic separations.

The complex interplay between loops, linkers and rosettes is even more reflected in the distance distribution (Fig. 3.9 C): the oscillations introduced by the loops is much clearer and the not-to-zero-decrease for loop size multiples is shown by distance "spikes" due to the higher probability of markers located in the linker, when the genomic separation surpasses the total rosette size. The shifts into new rosettes combined with the decrease of the loop based oscillations is accessible and suggests higher flexibility of rosettes than for loops. Thus, although position independent spatial distances reflect chromosome models with flexible loop and rosette sizes better than position dependent spatial distances, the analysis of a fixed MLS model still reveals at least part of its structure. The implicit linker dependence due to rosette repulsion is also again present (3.4, 3.5). For genomic separations greater than one loop size the increase of the position independent spatial distance is proportional to the linker size between loops of the MLS models (Fig. 3.9A) and of the RW/GL model (Fig. 3.9B). Therefore, for small genomic separations the position independent spatial distances are not related to the nuclear radius. For large linkers and genomic separations the spatial distances reflect the same nuclear radius dependencies, and reach distance values as found for the radial mass distribution of chromosome territories (3.4.1).

3.7 Discussion of the Simulation of Interphase Nuclei

The nuclear arrangement of chromosome territories, the folding of the 30 nm chromatin fiber into chromosome territories and the relation between the fiber and the microscopic morphology of nuclei are subject of current research (Chapter 1). It is assumed that the fiber is compacted in different stages: chromatin loops, aggregates of loops, and arrangement of these subcompartments into chromosome territories. The dynamics of these structures, changes during the cell cycle and the transport properties in nuclei, are also still debated. On the level of the chromatin fiber various models were proposed: In the Multi-Loop-Subcompartment (MLS) model (1.3.8, Fig. 1.14), chromatin loops form rosettes which are connected by a linker. In the Random-Walk/Giant-Loop (RW/GL) model, big loops are attached to a flexible backbone (1.3.7, Fig. 1.14). On the level of the whole nucleus the Inter-Chromosomal Domain (ICD) model has been proposed: Here nuclear transport occurs in a network of channels between dense chromosome territories (1.3.5, Fig. 1.13).

Whether the MLS and the RW/GL models can form chromosome territories in interphase nuclei, whether they lead to different morphologies on the level of whole nuclei and whether they can be distinguished experimentally on the fiber level has

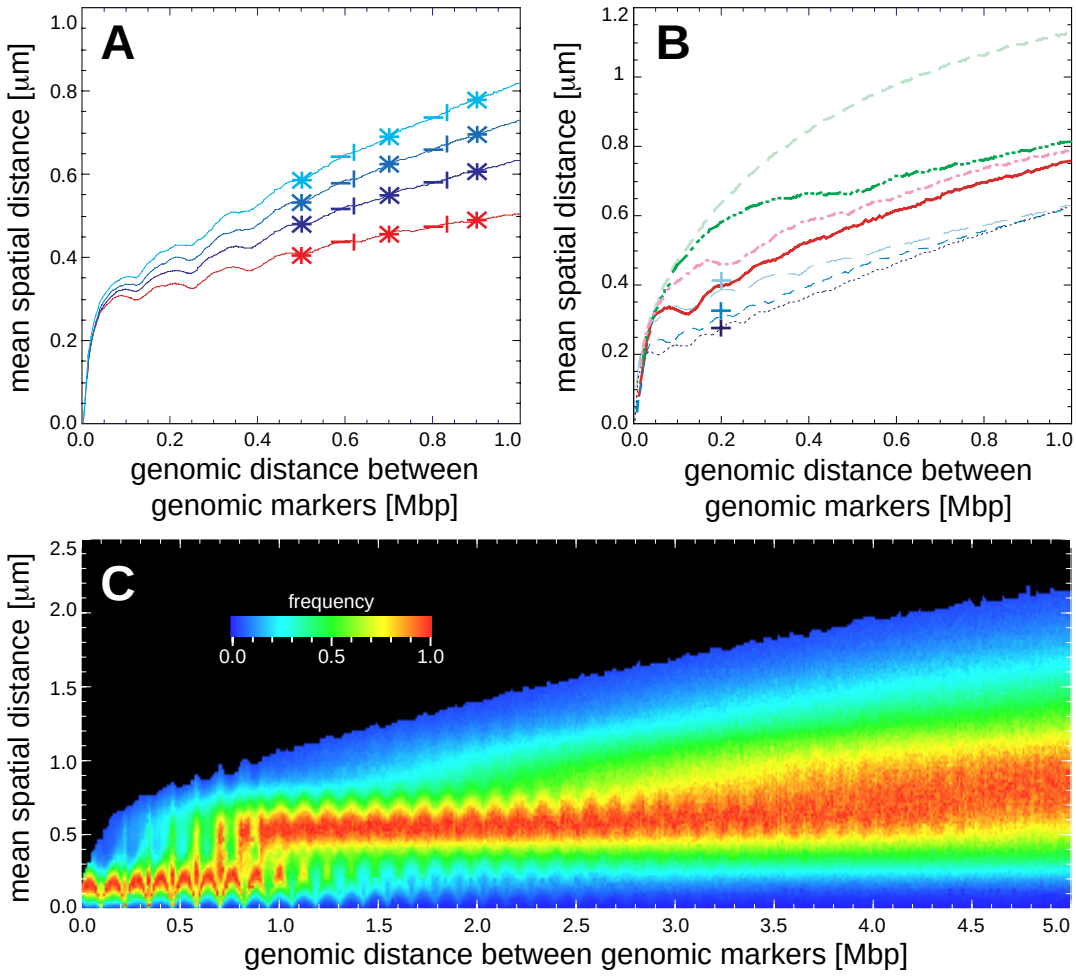


Fig. 3.9 Position Independent Spatial Distances between Genomic Markers in Cell Nuclei
The spatial distance between position independent placed markers is proportional to the linker size (**A**, 126-X-MLS[5], 63, 126, 189 and 252 kbp linkers are red, dark blue, blue and light blue) and to the loop size in the MLS and the RW/GL model (**B**, X-126-MLS[5], 63, 84 and 126 kbp loops are dotted, dashed and long dashed; X-RW/GL[5], 126, 252, 504 and 1320 kbp loops are all thick and solid, dash dotted, long dash dotted and long dashed). For small genomic separations no nuclear radius dependency exists (**A**, 3, 4, 5 and 6 μm radii are crosses, diagonal crosses, dash and upright dash), although for bigger separations the radial mass dependencies hold. The implicit linker extension by rosette repulsion is also present (**B**, 63-126-MLS[5], crosses). The position independent spatial distance distributions (**C**, 126-126-MLS[5]) reveal loop introduced oscillations up to ~3 Mbp and distance spikes for loop size multiples due to the higher probability of markers located in the linker while the genomic separation surpasses the total rosette size. The shifts into new rosettes at 0.75, 1.50, 2.25, 3.0 and 3.75 Mbp combined with the oscillation decrease suggests higher rosette than loop flexibility.

remained unclear. The nuclear arrangement of chromosomes and its relation to these folding topologies is also unknown. So far it is also not understood whether the different fiber topologies are in agreement with the ICD model.

To investigate these hypothesis the simulations of single chromosomes were extended to nuclei containing all 46 chromosomes, corresponding to diploid human

cells. The MLS and the RW/GL models were simulated assuming the 30 nm chromatin fiber as a polymer chain. The properties of this chain were defined by a stretching and bending potential. An excluded volume interaction kept the chain from crossing and a spherical boundary potential simulated the confinement of the nucleus. Simulated annealing and Brownian Dynamics methods as well as a four step decondensation procedure from metaphase were applied to generate interphase configurations at thermodynamical equilibrium.

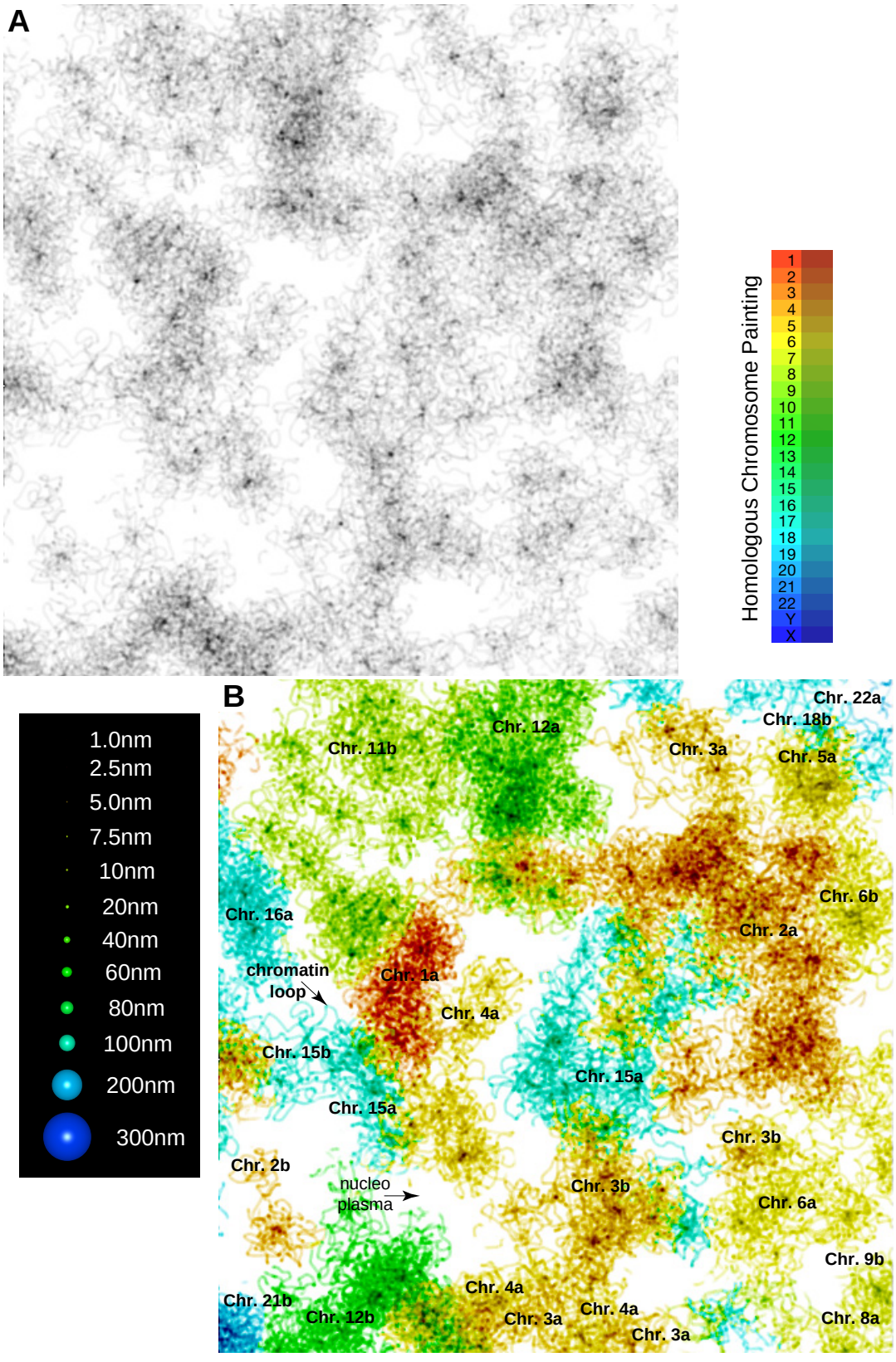
Both the MLS and the RW/GL model form chromosome territories that were visualized by rendering and simulation of electron (EM) and confocal laser scanning microscopic (CLSM) images. The morphology of both models is different: The MLS model reveals territories with a sharp "edge" and the rosettes result in distinct subcompartments visible with light microscopy. In contrast, the large RW/GL loops lead to a homogeneous chromatin distribution. Only the MLS model led to a low overlap of chromosomes, arms and subcompartments in agreement with experiments. The size of these subcompartments and their spatial distance are also in agreement with experiments based on fluorescence *in situ* hybridization (FISH), replication labelling *in vivo* by BrdU, and recently developed *in vivo* labelling of chromatin with histone-autofluorescent proteins (Chapter 7). Even small changes of the model parameters induced significant rearrangements of the chromatin morphology. Thus, pathological diagnoses of e. g. cancer, based on the nuclear morphology are due to structural changes on the chromatin level. The chromatin density distribution as seen in CLSM image stacks reveals a bimodal behaviour in agreement with current experiments. The comparison of simulated spatial distances between genetic markers as function of their genomic separation to experiments favours again an MLS model with loop and linker sizes of 63 to 126 kbp.

The nuclear localization of chromosome territories depended on the initial metaphase position: Small chromosomes are located more in the inner and large chromosomes more in the outer regions of the nucleus. This agrees with recent labelling of territories with fluorescence *in situ* hybridization (FISH). The results could indicate that only a special position in metaphase and no chromosomal attachement/localisation mechanisms in interphase are necessary to explain these results.

Additionally, the morphologies of the simulated cell nuclei show large spaces in rendered or simulated electron microscopic images. These voids allow high accessibility to most nuclear locations and also to the interior of chromosome territories by tracers of corresponding size. Therefore, the diffusion of small molecules and typical proteins is only moderately obstructed. This is in agreement with recent single molecule experiments measuring the diffusion of particles in nuclei of living cells. A channel like network for molecular transport between chromosome territories, as postulated by the Inter-Chromosomal Domain (ICD) model, was not apparent in the simulations. Therefore, the assumptions of the Inter-Chromosomal Domain (ICD) model seem at least simplified.

In summary, the results from the simulation of whole nuclei agree with the results from the simulation of single chromosomes. Again an MLS model with loop

and linker sizes of 63 to 126 kbp is favoured and the the hypothesis of the RW/GL and the ICD models seem unlikely. Additionally, the local and global characteristics of cell nuclei are tightly inter-connected. Therefore, the changes described qualitatively during mitosis, apoptosis or in the pathological classification of cancer, can be related to structural changes on the chromatin level. The simulations again propose the intuitive view of a nucleus as evolutionary optimized bioreactor: The genetic information in the nucleus is packaged, to fulfil all the requirements of pure storage, and on the other hand by its structural distribution guarantees an easy transport to every target site in the nucleus by Brownian diffusion. This guaranties random mixing, which leads to the most efficient reaction probability possible in a fluidic system. Structural and chemical modifications can then lead to the subtle regional regulation of processes, allowing possibilities of control well beyond the pure DNA sequence.


A
Homologous Chromosome Painting
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Y
X
1.0nm
2.5nm
5.0nm
7.5nm
10nm
20nm
40nm
60nm
80nm
100nm
200nm
300nm
B
Chr. 22a
Chr. 18b
Chr. 12a
Chr. 11b
Chr. 3a
Chr. 5a
Chr. 6b
Chr. 16a
Chr. 2a
chromatin loop
Chr. 1a
Chr. 4a
Chr. 15b
Chr. 15a
Chr. 15a
Chr. 3b
Chr. 2b
nucleo plasma
Chr. 3b
Chr. 6a
Chr. 9b
Chr. 4a
Chr. 21b
Chr. 12b
Chr. 3a
Chr. 4a
Chr. 8a
Chr. 3a

4 Scaling Properties of the Nuclear Organization

4.1 Introduction

The human genome consists of several stages of packaging, which cover huge length and time scales. Therefore, the scaling behaviour of the 30 nm chromatin fiber topology and the morphology of confocal laser scanning microscopic (CLSM) image stacks was determined. Both were obtained from simulations of single chromosomes and whole nuclei based on the chromatin fiber topology of the Random-Walk/Giant-Loop (RW/GL) model, in which large loops are attached to a flexible backbone, and the Multi-Loop-Subcompartment (MLS) models, in which small loops form rosettes, connected by a linker. For the analysis, a variety of scaling/fractal dimensions were calculated. The scaling of the chromatin fiber revealed different power-law behaviours on different scales. This multi-scaling is related to the random walk behaviour of the fiber, the globular nature of loops or rosettes, and the arrangement of loops or rosettes. Within the multi-scaling a fine-structure was present for the MLS model due to the rosette loops. This fine-structured multi-scaling behaviour agrees with the correlation behaviour in the DNA sequence of human chromosomes. Thus, the sequential and three-dimensional organization of genomes are closely interconnected. The scaling of CLSM image stacks reflected the model and imaging properties in detail. Thus, the chromatin fiber topology is closely connected to nuclear morphology. Therefore, scaling analyses of the nuclear morphology are a suitable approach to differentiate between different cell states, e. g. during the cell cycle, due to malignancy, in apoptosis or in response to drugs.

Fig. 4.1 Multi-Scaling Morphology of Simulated Electron Microscopic Images
The multi-scaling of the 3D nuclear organization resulting from the multi-packing of DNA is revealed in the electron microscopic image (**A**) and chromosome map (**B**, homologous chromosome painting legend, right) of a 126-126-MLS^5 nucleus deliberately unrelaxed, thus the channel like voids between chromosome territories are an artefact. Consequently, their invisibility in reality refutes the Inter Chromosomal Domain (ICD) hypothesis. The volume properties make clear that the dynamics and diffusion are obstructed according regarding the scale (spheric size legend, left).

4.2 Scaling Analyses Methods

In order to describe quantitatively the different chromatin topologies and microscopic morphologies resulting from the Multi-Loop-Subcompartment (MLS) or the Random-Walk/Giant-Loop (RW/GL) models of simulated chromosomes and simulated nuclei, their scaling behaviour was investigated. The scaling behaviour describes how a parameter, e. g. the length of the coast of Britain, depends e. g. on the scale of observation, i. e. the length of the ruler used to measure the coast line. Since the coast-line is self-similar it follows a power-law. The exponent of the power-law is the scaling or fractal dimension of the coast line. For Britain it values ~1.24. Mathematical lines, surfaces and volumes have fractal dimensions of 1.0, 2.0 and 3.0, which agrees with their Euklidian dimension. Thus, the British coast scales like a folded and not a straight line. (For another example see also 4.5)

To determine the scaling behaviour of different chromatin topologies and microscopic morphologies the scaling behaviour of several parameters was determined:

4.2.1 Exact Spatial-Distance and Exact Yard-Stick Dimension

The scaling behaviour of the one-dimensional axis of the 30 nm chromatin fiber was investigated by the exact spatial-distance dimension D_{SD} and the exact yard-stick dimension D_Y. Neglecting the different measuring process $D_{SD} \Leftrightarrow D_Y$ in the limit (Mandelbrot, 1977).

D_{SD} is the inverse exponent in the scaling relation between the distance $R_{SD}(c_{SD})$ connecting two chain positions and the contour length c_{SD} e. g. in bp

$$R_{SD}(c_{SD}) \sim c_{SD}^{(1/D_{SD})}. \tag{4.1}$$

This corresponds to the measurement of spatial distances between genomic markers as function of their genomic separation c_{SD} (Chapter 2, 2.7). Since R_{SD} depends for large c_{SD} on the marker position, the average over several randomly placed marker pairs need to be taken (see also position independent spatial distances, 2.7.1, Fig. 2.4A, Fig. 4.2A). Genomic separations ≥5.2 kbp (the base pair content of one chain segment) up to the whole chromosome were used. All possible maker pairs for genomic separations <25 Mbp and 5000 pairs >25 Mbp were taken. The mean was taken over 100 to 150 chromosome configurations.

The exact yard-stick dimension D_Y is the negative exponent in the scaling relation between the curve length $C_Y(l_Y)$ and the yard-stick length l_Y

$$C_Y(l_Y) = N_Y \cdot l_Y \sim l_Y^{1-D_Y} \text{ and } N(l_Y) \sim l_Y^{-D_Y} \tag{4.2}$$

with the number of yard-sticks N_Y, depending itself on l_Y. The beginning coordinates of a randomly chosen chain segment defined the initial startpoint from which the necessary N_Y to reach both chain ends was calculated by walking along the chain with l_Y sized steps. First the distance to the next segments was determined

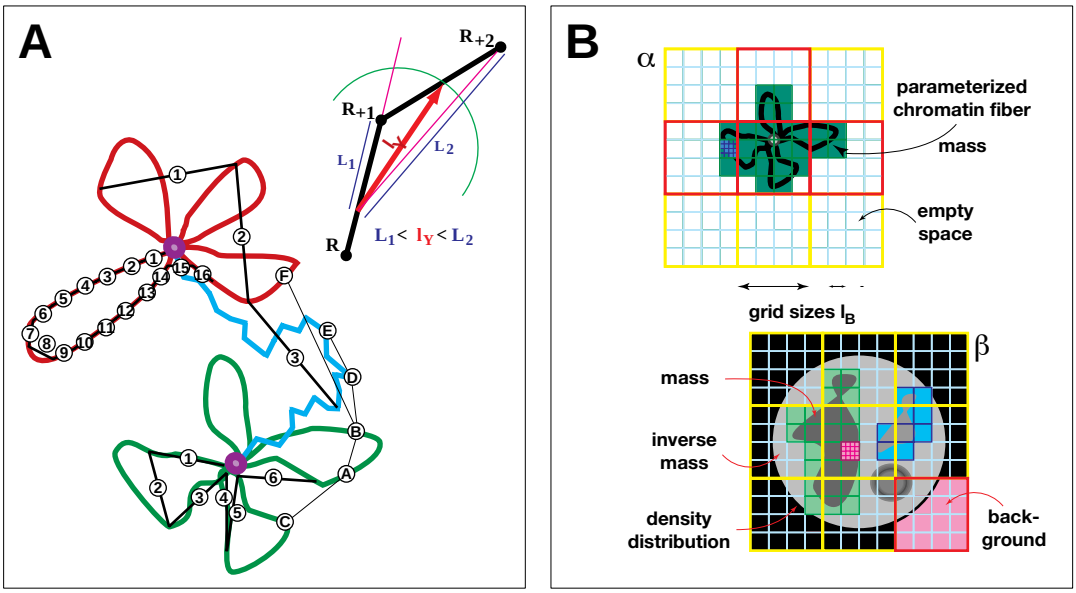


Fig. 4.2 Analyses Used for the Scaling Analyses
(**A**) The spatial-distance dimension D_{SD} was determined from position independent spatial distances as function of randomly positioned genetic markers with a genomic/curvature separation c_{SD}. Thus, markers could reside both on the same loop (**A-B**), on different loops (**A-C**), on a loop and in a linker (**B-D**), both in the linker (**D-E**) or on loops belonging to different rosettes (**B-F**). The exact yard-stick dimension D_Y was calculated by walking along the fiber using a yard-stick l_Y. Thus, the start and end of a small l_Y mostly reside in the same loop (**1-16**) in contrast to large l_Y often lying in different loops (**1-6**) or rosettes (**1-3**).The end-point E of l_Y was determined exactly by finding the chain segment, where $L_1 < l_Y < L_2$, before solving the corresponding vector equation. (**Bα**) The scaling behaviour of the voluminous chromatin fiber was determined by mapping the fiber to a 3D-grid of boxes, before the number of occupied boxes N_B of side length l_B in a measuring grid was counted and the pure box-counting dimension D_B was calculated. The smaller l_B, the better the volume of the fiber is approximated, thus the bigger l_B the more the chromosome is point-like approximated. (**Bβ**) In contrast to D_B, the weighted box-counting dimension D_{Bw} was determined as function of the intensity threshold of the chromatin density distribution of CLSM images and was additionally corrected for boxes part of the nucleus and the background (pink). Therefore, also D_{Bw} inverse-mass could be determined which is not possible for D_B as the empty space between the chromatin fiber is not restricted and thus infinite. For the weighted lacunarity dimension D_{Lw} the number of pixels n_i within one box was determined. For the local dimension $D_{Local-w}$, the diffuseness d_{local}, the skewness s_{local} and the kurtosis k_{local} the box did not exceed or was $l_B = 9 pixels$.

until it surpassed l_Y (Fig. 4.2A). Hence, the exact chain position and the startpoint of the next step where the distance equals l_Y is a coordinate point within the previous segment. This is given by the corresponding trigonometric vector equation. Near the chain end and/or for big l_Y, the distance could be $> l_Y$, thus N_Y was determined by the fraction of l_Y to reach the chain end. The resolution of l_Y was 5, 25 and 50 nm up to scaling length of 10^2, 10^3 and 10^4, respectively. Since, N_Y depends on the initial startpoint the bigger l_Y, $S(l_Y) = abs(0.05 \cdot l_Y)$ (l_Y in nm) startpoints were averaged, thus the standard error of D_Y was always <0.01. The mean was taken over 100 to 150 statistical chromosome configurations.

4.2.2 Weighted Box-Counting and Lacunarity Dimension

To determine the scaling behaviour of the real three-dimensional mass distribution of the 30 nm chromatin fiber and to determine the scaling of the mass, the inverse-mass (the mass complement, the chromatin "free" volume or nucleoplasma, Viswanathan & Heaney, 1995) and the iso-mass ("surfaces"; Pfeiffer, 1985; Cox & Wang, 1993) from simulated confocal laser scanning microscopic image stacks, the pure box-counting dimension D_B, the weighted box-counting dimension D_{Bw} and the weighted lacunarity dimension D_{Lw} were determined (Mandelbrot, 1977, Einstein *et al.*, 1998a; Einstein *et al.*, 1998b; MacLellan & Endler, 1998). On large scales $D_{SD} \Leftrightarrow D_Y \Leftrightarrow D_B$ (Mandelbrot, 1977; Avenir, 1992; Peitgen *et al.*, 1992).

D_B and D_{Bw} are the negative exponents in the scaling relation between the occupied volume $V_B(l_B)$ and the measuring volume with sidelength l_B

$$V_B(l_B) = N_B \cdot l_B^3 \sim l_B^{3-D_{B,Bw}} \text{ and } N_B \sim l_B^{-D_{B,Bw}} \tag{4.3}$$

with the number of box-volumes N_B containing at least one mass occupied pixel and depending itself on l_B. Thus, D_B and D_{Bw} are, in principle, a generalization of the yard-stick dimension to three-dimensional space (4.2.1). They, however, only approximate e. g. the length of a one-dimensional line due to the infinite positional relations of the line to the measuring box (Falconer, 1993).

To obtain D_B for the chromatin fiber of single chromosomes the chain of segments characterized by the start and end coordinates of each segment had to be transformed into a volume representation: The chromosome configurations were placed in a three-dimensional grid of boxes with a resolution of 5x5x5 nm. Therefore, the cylindrical segments were split using a cylindrical paramtrization with 1 nm resolution. Grid positions occupied by the chromatin fiber were set to a value of 1, else 0. A grid of boxes with sidelength l_B was placed on top of the mass containing grid and the number of boxes N_B containing mass was determined. All possible sidelength were used. Since N_B depends on the initial grid startposition the bigger l_B, $S(l_B) = abs(0.1 \cdot l_B)$ (l_B in grid positions) equally distributed startpositions were averaged, thus the standard error for D_B was <0.01. The average was taken over 100 to 150 statistical independent chromosome configurations.

To obtain D_{Bw} for the (inverse-, iso,-) mass of simulated confocal laser scanning microscopic (CLSM) image stacks, nuclei were placed in a three-dimensional grid with a resolution of 70x70x70 nm or 80x80x80 nm. This corresponds to an oversampling of 2-3 in lateral and 3-9 in axial direction in agreement with standard experimental CLSM setups. To map the cylindrical segments to this grid of pixels, they were split using a cylindrical parametrization of 5 nm resolution. For convolution with the point spread function (PSF), the cylindrical parametrization was replaced by a three-dimensional Gaussian distribution: According to the theoretic resolution of a 100x 1.4 oil immersion PL APO objective (7.2.5) and the more realistic resolution of a 60x 1.2 water immersion PL APO objective, the lateral focal width at half the maximum $FWHM_{x,y}$ was set to 139 or 240 nm and axially $FWHM_z$ to 236 or 720 nm, respectively. The Gaussian distribution was parametrized with a Car-

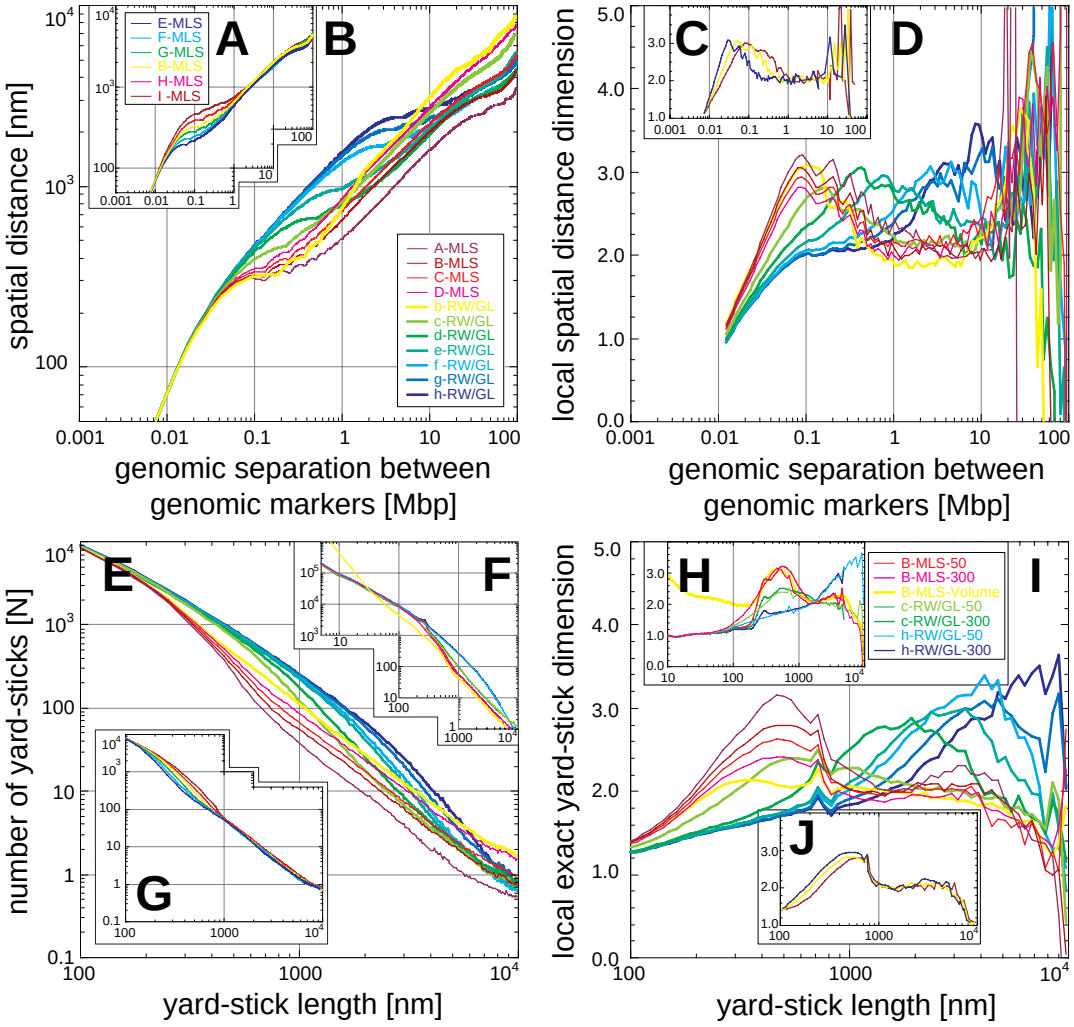


Fig. 4.3 Spatial-Distance and Yard-Stick Dimensions of Simulated Single Chromosomes
The spatial-distance function $R_{SD}(c_{SD})$ (**A**,**B**) and exact yard-stick curve length function $C_Y(l_Y)$ (**E**, **F**, **G**) shows power-law behaviour as expected for fractal self-similar polymer foldings. The slopes are the spatial-distance dimension D_{SD} and the exact yard-stick dimension D_Y. The finite size of chromosomes generates a cut-off >~80 Mbp or >~8 µm after which the pow-law behaviour breaks down. Beyond non-trivial power law behaviour due to the deviation of D_{SD} and D_Y from 1.0 (a stiff linear segment) or ~2.0 (a random walk), four major scaling regions exist. The detailed dimension behaviour is given by the local dimensions $D_{SD}(c_{SD})$ and $D_Y(l_Y)$ (**C**, **D**, **H**, **I**, **J**) with fluctuations the bigger the closer c_{SD} and l_Y are to the cut-off. The general multi-scaling behaviour of D_{SD} and D_Y is characterized by an increase from an initial 1.0 for small c_{SD} and l_Y characterizing the stiff chromatin segments, over values ~2.0 as for the random walk of the segments to a maximum of 3.0 stating the ring-shaped loops of both the MLS and the RW/GL model and globular state of the rosettes of the MLS models according to the c_{SD} and l_Y. In the MLS model thereafter again local dimensions ~2.0 are reached describing the random organization of the rosettes relative to each other. Within the general behaviour a fine structure attributable to the loops aggregated in rosettes is present for MLS models, better measured with D_{SD}. The maximum position and height is proportional to the loop and linker size in the MLS and the RW/GL model, and all the inherent model properties described in Chapter 2 are visible. Computation of D_Y for 300 nm segment resolutions shows at l_Y = 300 nm a peak smeared out for 50 nm (**F**,**H**). Despite the higher starting dimension, D_Y equals D_B measuring the voluminous chromatin fiber (**F**, **H**) in agreement with theoretic prediction.

tesian one of 10 nm resolution and its values put into the pixel grid. The total pixel intensity was summarized and normalized to a 0 to 256 intensity range.

The weighted box-counting dimension D_{Bw} was determined similar to D_B, but as function of intensity thresholds (Haidekker *et al.*, 1997). Additionally, the weighting factor $w_i = n/l_b^2$ with the number of pixels n belonging to the nucleus was applied to each box. Thus, boxes covering only part of a nucleus do not distort the result. This approach equals the use of the yard-stick fraction for big l_Y in D_Y.

Since different mass distribution could lead to the same box-counting dimension, the weighted lacunarity dimension D_{Lw}, which is hole sensitive, was calculated for the (inverse-, iso-) mass (Einstein *et al.*, 1998a; Einstein *et al.*, 1998b): D_{Lw} is the exponent in the scaling relation between the lacunarity function $\Lambda(l_B)$ and the measuring volume with sidelength l_B. $\Lambda(l_B)$ is obtained from the distribution $Q_n(M, l_B)$ of the number of mass containing pixels n_i in each box with sidelength l_B, in terms of the moments $Z_{Q_n}^{(q)}(l_B) = \sum_{M=1}^{l_B^2} M^q Q_n(M, l_B)$, thus

$$\Lambda_{Lw}(l_B) = \frac{Z_{Q_n}^{(2)}(l_B)}{[Z_{Q_n}^{(1)}(l_B)]^2} = \sum_{i=1}^{S} w_i \sum_{i=1}^{S} \left(\frac{n_i^2}{w_i}\right) \Big/ \left(\sum_{i=1}^{S} n_i\right)^2 \quad \text{and} \quad \Lambda_{Lw}(l_B) \sim l_B^{D_{Lw}} \tag{4.4}$$

with the weighting factor w_i from above and the number of boxes covering the nucleus S. In general, D_{Lw} was calculated as D_{Bw}.

4.2.3 Calculation of the Scaling Exponents/Dimensions

To determine the local spatial-distance dimension $D_{Sp}(l_{SP})$, the local yard-stick dimension $D_Y(l_Y)$, and the local box-counting dimension $D_B(l_B)$, the logarithm of the corresponding scaling relation was taken and the unsymmetric finite difference quotient of second order was applied (6.2.1). For the scaling of the (inverse-, iso-) mass of confocal images as function of intensity thresholds, D_{Bw} and D_{Lw} were determined from a linear regression in the power law-regime excluding the cut-off region (Peitgen, 1992).

4.2.4 Weighted Local Dimension, Diffuseness, Skewness and Kurtosis

The local/regional and thus spatially resolved scaling behaviour of the mass distribution in confocal laser scanning images was investigated by the weighted local dimension $D_{local-w}$, which is the scaling exponent in the relation between the total mass $M_{local}(l_{local})$ and the measuring volume of sidelength l_{local}

$$M_{local}(l_{local}) \sim l_{local}^{D_{local,w}}. \tag{4.5}$$

The mass was calculated for boxes around all image pixels belonging to the nucleus with l_{local} of 1, 3, and 5 pixels only (tests with 7 and 9 pixels were also performed; Kaye, 1989; Rodríguez-Iturbe & Rinaldo, 1997). Since the mass in microscopic images ranges from 0 to 255, the local scaling dimensions $D_{local-w}$ could be greater

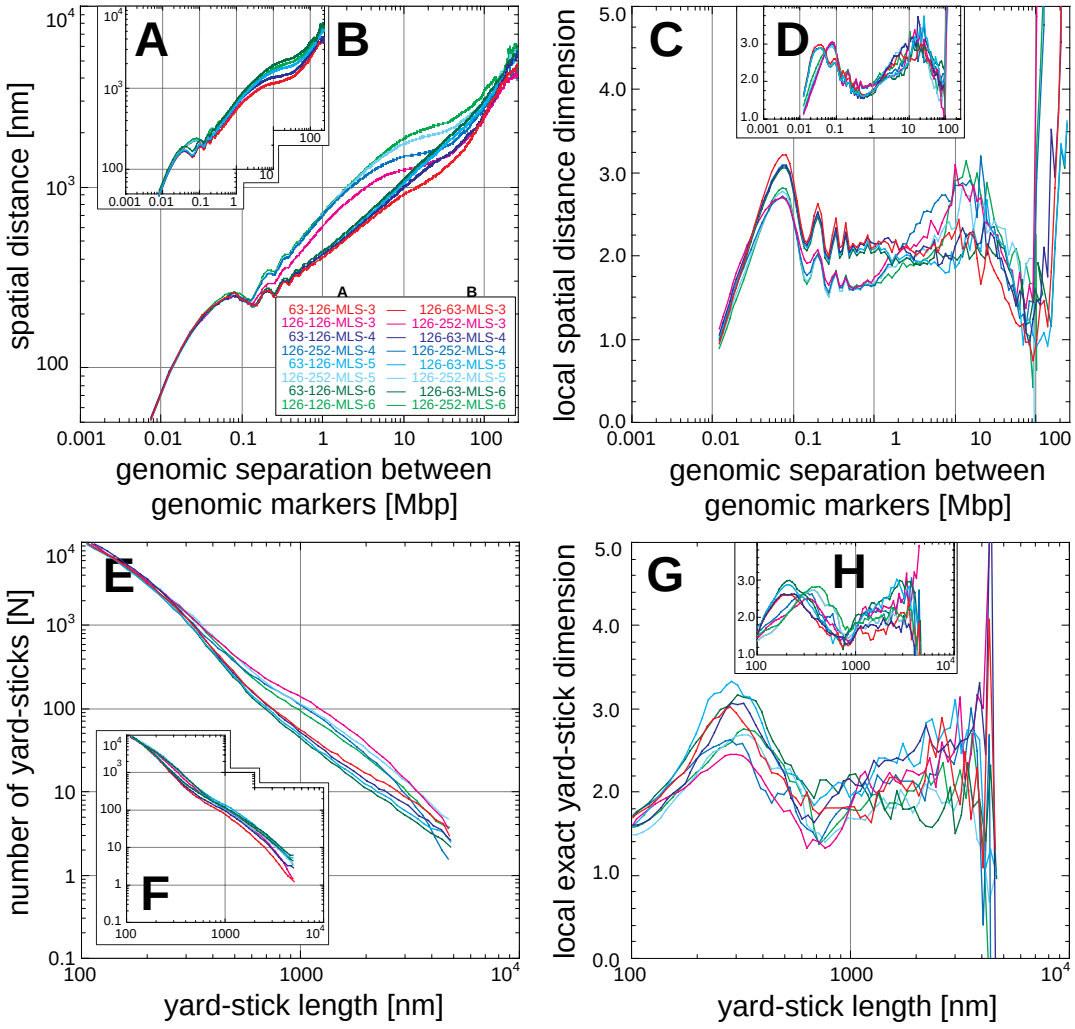


Fig. 4.4 Spatial-Distance and Yard-Stick Dimensions of Simulated Nuclei
The spatial-distance function $R_{SD}(c_{SD})$ (**A**, **B**) and the exact yard-stick curve length function $C_Y(l_Y)$ (**E**, **F**) behave like those of single chromosomes: Despite the general power-law behaviour with a cut-off >~150 Mbp or >~8 µm, there are again four different major scaling regimes and the local dimensions D_{SD} (**C**, **D**) and D_Y (**G**, **H**) reveal multi-scaling behaviour with a fine structure attributable to the loop structure according to the model topology: The local dimensions increase from 1.0 characterizing the one-dimensional chromatin segment, over values ~2.0 as for random walks to the maximum of 3.0 due to the ring-shaped loops of both the MLS and the RW/GL model and globular state of the rosettes of the MLS models. The model topologies, however, are characterized by he maximum position and height being proportional to the loop and linker size in the MLS and the RW/GL model and by the nuclear radius influencing the fourth scaling region or the local dimensions for >~90 Mbp or >~5 µm. All the model properties described in Chapter 3 are present.

than 3.0. Therefore, the total intensity for one box was normalized to the real mass distribution where a pixel could either hold mass or not. To account for pixels in the box not belonging to the nucleus, the mass in a box was weighted with the fraction of nuclear pixels.

Additionally, the local diffuseness d_{local} (Einstein *et al.*, 1994), the local skewness s_{local} and the local kurtosis k_{local} (Mongelard *et al.*, 1999) were calculated to further interpret the results of $D_{local-w}$. They are frequently used measures in the quantitative analysis of pathological specimens (Haralick *et al.*, 1973; Irinopoulou *et al.*, 1993; Cross *et al.*, 1997; Cross, 1997). The skewness and kurtosis are independent from the local diffuseness and describe how the shape of the intensity/mass distribution diverges from a Gaussian distribution: whereas the skewness measures the imbalance ($s_{local} = 0$ for an ideal Gaussian) between the upper and lower part of the distribution in relation to the average, the kurtosis judges the flattening ($k_{local} < 0$) or sharpening ($k_{local} > 0$) of the distribution.

The local diffuseness d_{local} was calculated as the local standard deviation in the intensity within one box of $l_{local} = 5 pixels$ with the number of pixels in the box $N = 125$), the intensity of the N's pixel I and the average box intensity $\bar{I}$:

$$d_{local} = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(I_n - \bar{I})^2}. \tag{4.6}$$

The local skewness s_{local} was calculated according to

$$s_{local} = \frac{1}{N}\sum_{n=1}^{N}\left[\frac{I_n - \bar{I}}{d_{local}}\right]^3 \tag{4.7}$$

and the local kurtosis k_{local} was calculated according to

$$k_{local} = \left(\frac{1}{N}\sum_{n=1}^{N}\left[\frac{I_n - \bar{I}}{d_{local}}\right]^4\right) - 3 \tag{4.8}$$

For Equ. 4.6, Equ. 4.7 and Equ. 4.8 pixels belonged to the nuclear volume and were corrected for background pixels (Fig. 4.2Bβ).

4.3 Scaling Behaviour of the 30 nm Chromatin Fiber

The scaling of the 30 nm chromatin fiber shows a complex behaviour: Beyond the simple appearance of a scaling behaviour, the scaling varies on different scales and reveals also a fine structure:

4.3.1 Appearance of Scaling

The exact spatial-distance function $R_{SD}(c_{SD})$ (Equ. 4.1, 4.2.1) and the exact yardstick function $C_Y(l_Y)$ (Equ. 4.2, 4.2.1), and their exponents, the local dimensions $D_{Sp}(l_{SP})$ and $D_Y(l_Y)$ (4.2.3) were calculated for the one-dimensional axis of the 30 nm chromatin fiber of simulated single chromosomes (Chapter 2) and single chromosomes in whole nuclei (Chapter 3). The calculation of $R_{SD}(c_{SD})$ and

$C_Y(l_Y)$ were exact, since the coordinates of the markers separated by c_{SD} for the spatial distance and the l_Y long yard-sticks on the chromatin backbone were exact. The summation and averaging procedure excluded numeric instabilities. The calculation of the local dimensions were also exact concerning the resolution of c_{SD} and l_Y, smoothing out structures for increased c_{SD} and l_Y. Additionally, the box-volume function $V_B(l_B)$ (Equ. 4.3) and its exponent, the local box-counting dimension $D_B(l_B)$ (4.2.3) were determined for single chromosomes (Chapter 2). Thus, the results of $D_Y(l_Y)$ are extended to the real three-dimensional mass distribution of the chromatin fiber. $D_B(l_B)$ bridges also the gap to the scaling analyses of simulated confocal laser scanning microscopic (CLSM) image stacks. The volume parametrization of the chromatin fiber, the resolution of the mass containing grid and the placement of measuring boxes of sidelength l_B were set that the calculation of $V_B(l_B)$ was hardly compromised $<<l_B$. The determination of $D_B(l_B)$ was exact reminding these constraints and the chosen resolution of l_B. Considering the simulation properties of the model, for scales c_{SD}, l_Y and l_B >4 to 6 chain segments the one-dimensional backbone representation is comparable to the volume representation of the 30 nm chromatin fiber, thus $D_{SD} \Leftrightarrow D_Y \Leftrightarrow D_B$ (Mandelbrot, 1977).

In all analyses of MLS and RW/GL chromosome topologies $R_{SD}(c_{SD})$, $C_Y(l_Y)$ and $V_B(l_B)$ show power-law behaviour with varying slopes indicating non-trivial scaling exponents. This is corroborated by the spatial-distance, yard-stick and box-counting dimensions $D_{Sp}(l_{SP})$, $D_Y(l_Y)$ and $D_B(l_B)$ with varying values of 0.0, 1.0, 2.0 or 3.0. These values are characteristic for a point, a line, a plane or random walk and a volume (Fig. 4.3AB&E-G, Fig. 4.4AB&EF). $D_{Sp}(l_{SP})$ and $D_Y(l_Y)$, differ from $D_B(l_B)$ for small c_{SD}, l_Y and l_B. This is due to the fiber representation as a one-dimensional axis and as voluminous cylinders. The scaling behaviour existed nearly up to the entire scale of the chromosome, but at least up to ~80 Mbp or ~8 µm for single chromosomes and ~150 Mbp or ~8 µm in whole nuclei. Therefore, the scaling spans ~4 or ~3 orders of magnitude concerning genetic separation or curve length and subtracting the minimal resolution of the simulations being 5000 bp or 50 nm per chain segment. For c_{SD} and l_Y approaching the entire chromosome size a cut-off exists at which the scaling behaviour breaks down. The cut-off is characterized by growing fluctuations. The $D_Y(l_Y)$ thereby tends to 0.0 which suggests that the chromosome is represented as a point-like object.

4.3.2 Multi-Scaling and Fine-Structured Multi-Scaling

Beyond the appearance of simple power-law scaling with a single slope covering the entire scale, $R_{SD}(c_{SD})$, $C_Y(l_Y)$ and $V_B(l_B)$ show a more complex behaviour: For all MLS and RW/GL fiber topologies the slopes vary considerably within four or three major scaling regimes (Fig. 4.3AB&E-G, 4.4AB&EF). This indicates a varying degree of scaling, which is called multi-scaling. Therefore, the local dimensions $D_{Sp}(l_{SP})$, $D_Y(l_Y)$ and $D_B(l_B)$ limited only by the c_{SD}, l_Y and l_B resolution, were determined, since they are the most detailed measure to investigate this behaviour.

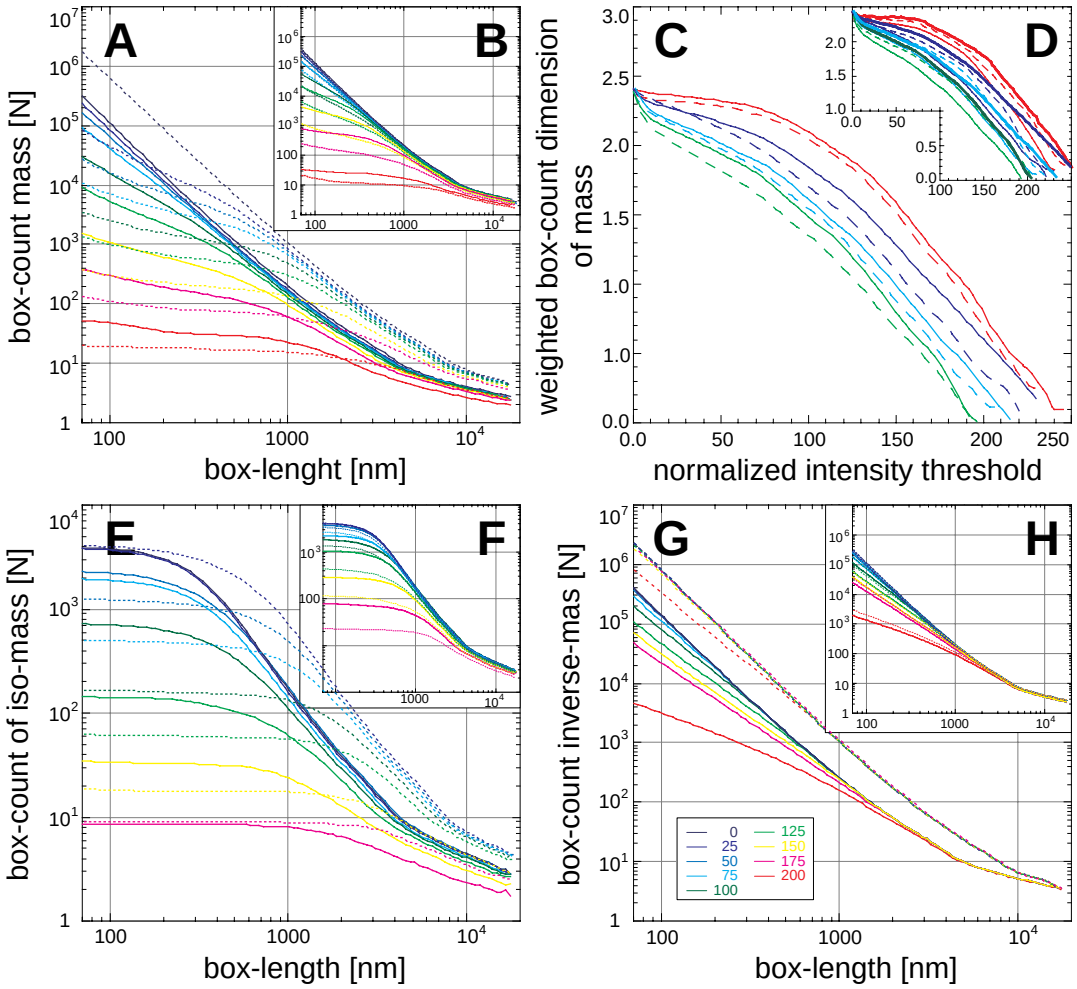


Fig. 4.5 Weighted Box-Counting Dimension of Simulated Nuclear CLSM Image Stacks
The weighted box-volume function $V_B(l_B)$ of the mass (**A**, **B**), inverse-mass (**E**, **F**) and iso-mass (**G**, **H**) shows distinct power-law behaviour for different (inverse-, iso-) mass threshold in the analyses of simulated confocal image stacks (126-252-MLS^3: solid, 126-252-MLS^6: dotted in **A**, **E**, **G**; 63-63-MLS^3: solid, 126-63-MLS^6: dotted in **A**, **E**, **G**). The box volume function show an upper cut-off due to and proportional to the finite size of the nuclei (4.5; Fig. 4.9). The scaling below the cut-off shows two scaling regions with slopes the more different the higher the threshold describing the different structural patterns (**E**). The determination of the weighted box-counting dimensions D_B as function of the mass threshold (**C**, **D**) shows a decreasing behaviour. D_B is inversely proportional to the nuclear radius, the linker and loop size and proportional to the objective resolution (**C**, 126-X-MLS-^{X}, linkers of 63 and 252 kbp are solid and dashed, nuclear radii of 3, 4, 5 and 6 µm are red, blue, light blue and green; **D**: X-126-MLS^X, loops of 63 and 126 kbp are dotted and thin, 100x, 60x objectives are thin and thick lines).

$D_{Sp}(l_{SP})$ and $D_Y(l_Y)$ are characterized by starting values of 1.0 for small c_{SD} and l_Y, an increase to a general maximum approaching 3.0 and a decay to ~2.0 for mid sized c_{SD} and l_Y, before a final decrease approaching the cut-off (Fig. 4.3CD&H-J, 4.4CD&GH). The first regime is attributable to the stiff one-dimensional chain seg-

ments, used in the simulations. The increase describes the increasing random walk behaviour with the growing chain, until the maximum is reached for looped fiber foldings. The subsequent decay is steeper for the single RW/GL than for the MLS loops, which here form globular rosettes. This indicates the transition to a random walk above the MLS rosettes or the RW/GL loops. Thus, dimensions are again ~2.0, before the final decrease at which chromosomes are represented as point like object. Consequently, the general scaling behaviour is indeed multi-scaling representing the detailed 30nm chromatin fiber topologies of the MLS and RW/GL models. The local box-counting dimension $D_B(l_B)$ deviates from this general behaviour only in the start value for small l_B and thus the increase to the general maximum. Therefore, indeed the approximation of the dimensions $D_{SD} \Leftrightarrow D_Y \Leftrightarrow D_B$ is found.

Within the general multi-scaling behaviour a fine structure is present in the decrease of the general maximum of the MLS model. From its periodicity it is attributable to the loops of MLS rosettes. The fine structure is better visible in $D_{Sp}(l_{SP})$ than in $D_Y(l_Y)$ or $D_B(l_B)$. This is due to the measuring process averaging out fine structures in the latter two cases. Thus, the fluctuations present in the $D_{Sp}(l_{SP})$ at large c_{SD}, are suppressed in the $D_Y(l_Y)$ at large l_Y. However, the latter reveals more clearly the general multi-scaling behaviour of the large RW/GL loops. Consequently, the MLS model presents fine structured multi-scaling behaviour in contrast to the RW/GL model. Notably, also the finite length of the chain segment is present in $D_Y(l_Y)$ for simulations using 300nm long segments, in contrast to 50nm segments, where the effect is mostly averaged out (Fig. 4.3F&I).

The detailed model dependencies are, also reflected by the fine-structured multi-scaling behaviour: The position and height of the general maximum are proportional to the loop and linker size of the MLS and RW/GL models for single chromosomes and also slightly proportional to the nuclear radius for single chromosomes in nuclei. The latter influences the scaling behaviour at large scales (>90Mbp or >5µm), since the nuclear radius influences the mean chromosome extension. Thus, the extension of the power-law scaling regime, the inset of the final general decay of D_{SD} and D_Y as well as the cut-off are proportional to the chromosome size. The fine-structure, the fine-structure spacing and visibility is loop size proportional.

Consequently, the topology of the 30nm chromatin fiber does not only exhibit a simple power-law, but moreover a fine-structured multi-scaling behaviour. This characterises in detail the Multi-Loop-Subcompartment (MLS) and Random-Walk/Giant-Loop (RW/GL) models, especially the loops and globular state of the MLS rosettes. On scales of 10^4 to 10^6 bp this MLS scaling is in excellent agreement with the existence of fine-structured multi-scaling long-range correlations in human DNA sequences (Chapter 6). Thus, the sequential and the structural scaling behaviour are causally connected as in the case of the correlation behaviour of the creating function (correlation coefficients of 0.5) and the three-dimensional structural behaviour (exact yard-stick dimension of 2.0) of random walks (Stanley & Ostrowsky, 1986; Stanley & Ostrowsky, 1988).

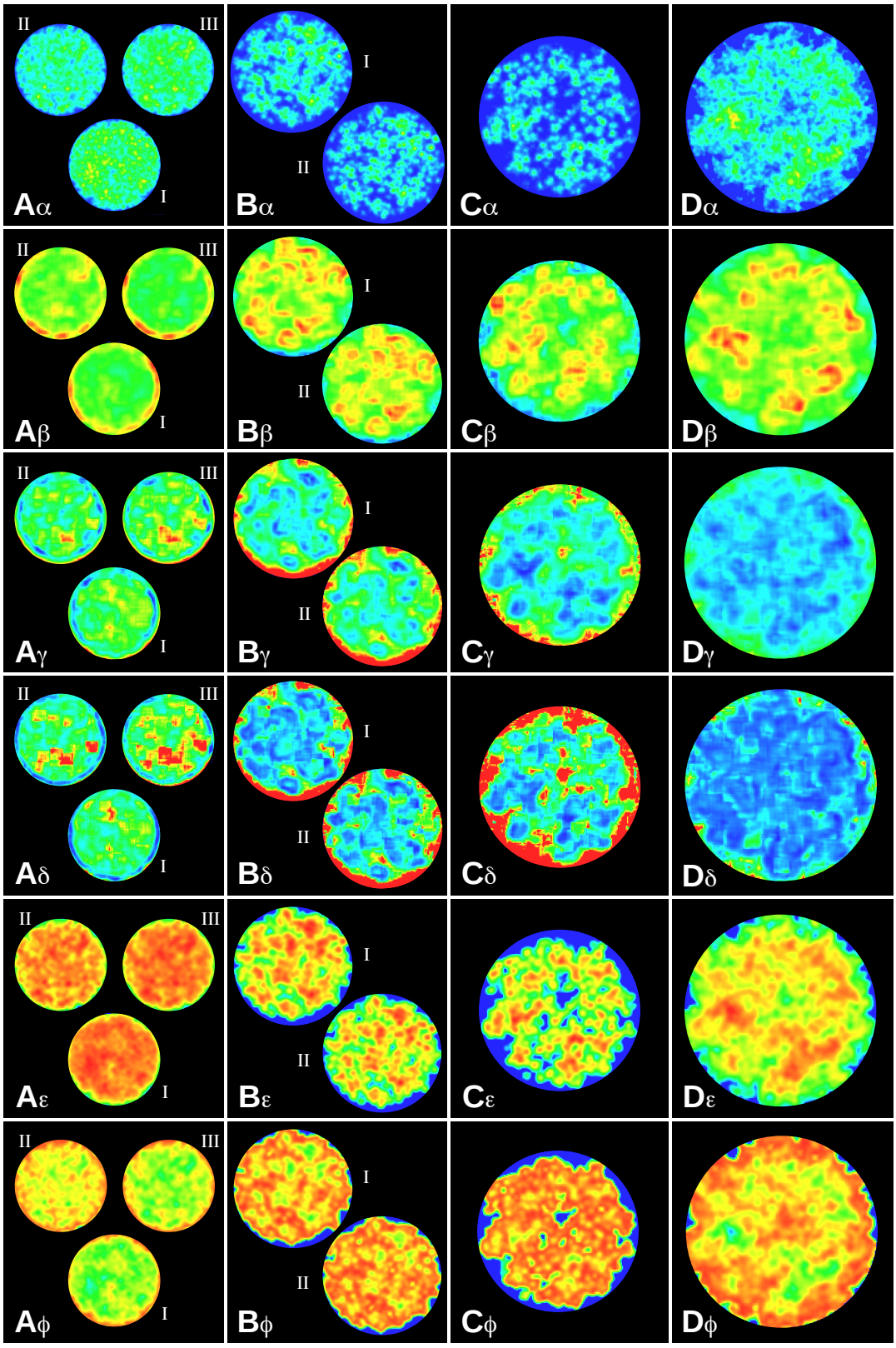
4.4 Scaling of Simulated Confocal Images of Nuclei

For analyses of the scaling behaviour on the scale of whole nuclei the weighted box-counting and weighted lacunarity dimensions D_{Bw} and D_{Lw} were calculated as function of (inverse-, iso-) mass thresholds in simulated confocal laser scanning microscopic (CLSM) image stacks (Chapter 3). The mass is given directly by the intensity and the inverse-mass is the complement of the mass. The iso-mass is the mass at a threshold (equivalent to iso-mass "surfaces"). $D_{local-w}$, and the statistical parameters diffuseness d_{local}, skewness s_{local} and kurtosis k_{local} were calculated, to analyse scaling behaviour with spatial resolution and their contributions to D_{Bw} and D_{Lw} in respect to the morphology of the local (inverse-, iso-) mass dimension.

4.4.1 Weighted Box-Counting and Weighted Lacunarity Dimension

The weighted box-volume function $V_B(l_B)$ and the weighted lacunarity function $\Lambda(l_B)$ were calculated as function of (inverse-, iso-) mass thresholds. Both $V_B(l_B)$ and $\Lambda(l_B)$ show for all masses nearly the same power law behaviour (Fig. 4.5A&B&E-H; due to their similarity only $V_B(l_B)$ is shown). $V_B(l_B)$ is proportional to the mass threshold at every position of l_B since at higher thresholds there is less contributing mass. Since, in nuclei there is less iso-mass, than mass, than inverse-mass, $V_B(l_B)$ fans apart differently for different thresholds. The degree of fanning is proportional to the amount of mass. In general, $V_B(l_B)$ shows power-law behaviour for all masses over two orders of magnitude before reaching a cut-off

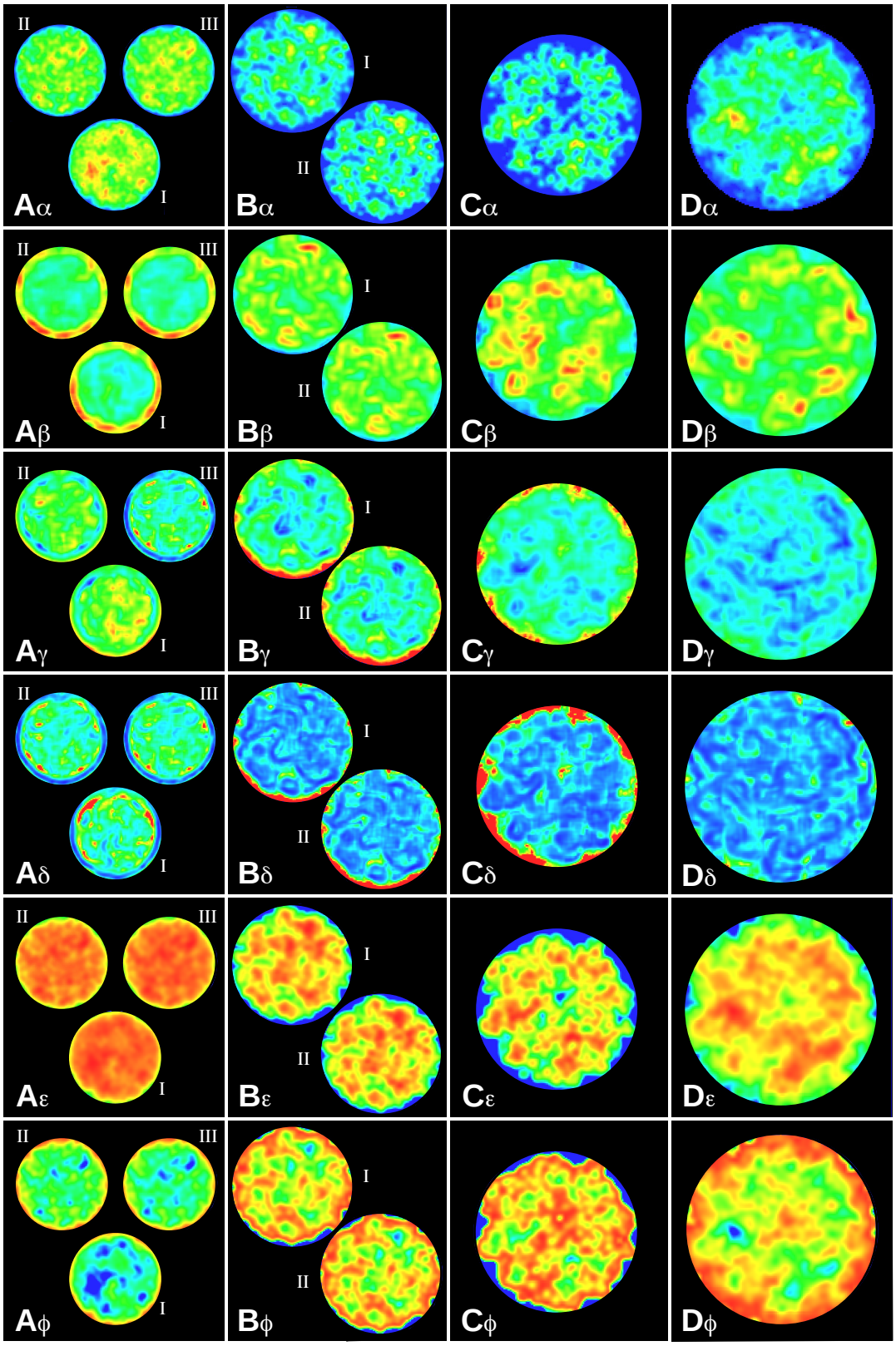
Fig. 4.6 Morphology-I of Weighted Local Mass Dimension, Diffuseness, Skewness and Kurtosis
The distinct differences in the morphology of chromosome topologies (Fig. 3.2) visible in the confocal laser scanning microscopic (CLSM) images with a high resolution **100x 1.4 oil** immersion PL APO objective (**α**) for the models 63-63-MLS^3 (**AI**), 63-252-MLS^3 (**AII**), 126-252-MLS^3 (**AIII**), 126-126-MLS^4 (**BI**), 84-126-MLS^4 (**BII**), 126-126-MLS^5 (**C**), 1320-RW/GL^6 (**D**), are visible and directly connected to the morphology in the images of the local diffuseness indicating and proportional to structural edges (**β**), the local skewness proportional to the imbalance of the intensity distribution (**γ**), the local kurtosis proportional to the sharpness of the intensity distribution (**δ**), and the weighted local dimension of the mass (**ε**), and the inverse-mass (**ϕ**) indicating the mass scaling of the local intensity distribution: The subcompartment forming rosettes of the MLS model are visible as separated entities (**A**, **B**, **C**), distinguished by their constituting loop size (**AII**, **AIII**, **BII**, **BI**) and linker size (**AI**, **AII**), despite their washing out by the measurement procedures (4.2.4). In contrast, the large RW/GL loops create much smoother/homogener morphologies with less edges, imbalances, more sharpness as well as scaling nearer to $D_{Local-w}$ of 3.0. Small RW/GL loops perform similarly. The determined values are in agreement and spatial resolve the distribution origin of the corresponding values (Fig. 4.8, 4.4). The images were calibrated according to the colour legend below, for the intensity/diffuseness/skewness/kurtosis the frequency needed to be 0.5% of the maximum.


II
III
I
Aα
I
II
Bα
Cα
Dα
Aβ
Bβ
Cβ
Dβ
Aγ
Bγ
Cγ
Dγ
Aδ
Bδ
Cδ
Dδ
Aε
Bε
Cε
Dε
Aϕ
Bϕ
Cϕ
Dϕ

with similar power-law behaviour for all thresholds. The position of this cut-off is proportional to the nuclear radius and indicates the transition at which nuclei are represented as point like object. This is in agreement with the scaling of the nuclear lamina and the cut-off found there (4.5). Before the cut-off the scaling splits in two scaling regions with slope difference and the point of transition between both slope regions being proportional to the nuclear radius, the loop and linker size, the objective resolution and the threshold. This behaviour describes the different structures left above the threshold and is therefore different for the mass (Fig. 4.5A&B), the inverse-mass (Fig. 4.5G&H) and the iso-mass (Fig. 4.5E&F).

For the determination of the weighted box-counting dimension D_{Bw} and the weighted lacunarity dimension D_{Lw} as function of the threshold, the linear regression from the starting l_B to the upper cut-off was calculated, averaging over both scaling regions. Again D_{Bw} and D_{Lw} show nearly the same proportionalities as described above. From an initial value of ~2.4 the mass D_{Bw} decreases to zero as function of the threshold. The amplitude of the curve (despite the starting value) is inversely proportional to the nuclear radius, the loop and linker size, and the objective resolution (Fig. 4.5C&D). For the iso-mass the starting value is ~2.0 and is faster reduced to zero with the stated proportionalities for the mass. The inverse mass exhibits a starting value of ~2.8 and shows the inversed proportionalities. The absolute values and proportionalities are in agreement with the average of the weighted local mass dimension distributions (4.4.2, Fig. 4.8A-C). Consequently, the weighted box-counting dimension D_{Bw} and the weighted lacunarity dimension D_{Lw} distinguish between different three-dimensional organizations of chromatin and result in absolute values for the scaling behaviour.

Fig. 4.7 Morphology-II of Weighted Local Mass Dimension, Diffuseness, Skewness and Kurtosis
The distinct differences in the morphology of chromosome topologies (Fig. 3.2) visible in the confocal laser scanning microscopic (CLSM) images with a low resolution **60x 1.2 water** immersion PL APO objective (**α**) for the models 63-63-MLS^3 (**AI**), 63-252-MLS^3 (**AII**), 126-252-MLS^3 (**AIII**), 126-126-MLS^4 (**BI**), 84-126-MLS^4 (**BII**), 126-126-MLS^5 (**C**), 1320-RW/GL^6 (**D**), are visible and directly connected to the morphology in the images of the local diffuseness indicating and proportional to structural edges (**β**), the local skewness proportional to the imbalance of the intensity distribution (**γ**), the local kurtosis proportional to the sharpness of the intensity distribution (**δ**), and the weighted local dimension of the mass (**ε**), and the inverse-mass (**ϕ**) indicating the mass scaling of the local intensity distribution: The subcompartment forming rosettes of the MLS model are visible as separated entities (**A**, **B**, **C**), distinguished by their constituting loop size (**AII**, **AIII**, **BII**, **BI**) and linker size (**AI**, **AII**), despite their washing out by the measurement procedures (4.2.4). In contrast, the large RW/GL loops create much smoother/homogener morphologies with less edges, imbalances, more sharpness as well as scaling nearer to $D_{Local-w}$ of 3.0. Small RW/GL loops perform similarly. The determined values are in agreement and spatial resolve the distribution origin of the corresponding values (Fig. 4.8, 4.4). The images were calibrated according to the colour legend below, for the intensity/diffuseness/skewness/kurtosis the frequency needed to be 0.5% of the maximum


II
III
I
Aα
I
II
Bα
Cα
Dα
Aβ
Bβ
Cβ
Dβ
Aγ
Bγ
Cγ
Dγ
Aδ
Bδ
Cδ
Dδ
Aε
Bε
Cε
Dε
Aϕ
Bϕ
Cϕ
Dϕ

4.4.2 Weighted Local Mass Dimension

The weighted local mass function $M_{local}(l_{local})$ and its exponent, the weighted local mass dimension $D_{local-w}$, were calculated for the (inverse-) mass distribution in a box of side length $l_{local} = 5 pixels$ around a pixel. The mass was weighted to account for pixels not belonging to the nucleus (4.2.2, 4.2.3). Already the spatial description of $D_{local-w}$ represents correctly the rosettes of the MLS model and the interplay between the linker and loop size as well as the nuclear radius (Fig. 4.6A-Cαεϕ, Fig. 4.7A-Cαεϕ). The homogeneity of the RW/GL model is also represented correctly (Fig. 4.6Dαεϕ, Fig. 4.7Dαεϕ). Even small changes of model parameters are visible: e. g. the change from 63 to 126kbp loops (Fig. 4.6A&Bαεϕ, Fig. 4.7A&Bαεϕ) or linkers (Fig. 4.6A&BαεϕI&II, Fig. 4.7A&BαεϕI&II). High mass $D_{local-w}$ colocalizes with the MLS subcompartments or statistical clumps the in RW/GL model. The inverse holds for the inverse-mass. Thus, the origin of features in the $D_{local-w}$ distributions of the (inverse-) mass are causally clarified.

The $D_{local-w}$ distributions of the mass (Fig. 4.8A&B) and the inverse-mass (Fig. 4.8C&D) are characterized by one major maximum located between 2.0 and 2.5 for the mass and between 2.5 and 2.9 for the inverse-mass. This agrees with the fact that there is much more space between the mass than mass itself (5.3). For small nuclei of 3µm radius, the inverse-mass shows two peaks due to the different scaling in the nucleus and at the nuclear membrane. The RW/GL model leads to much higher local dimensions, showing the homogener mass distribution of this fiber topology. The distribution, its peak height and width are proportional to the nuclear radius for the (inverse-) mass, inversely proportional to the linker size, loop size and objective resolution for the mass and proportional for the inverse mass. This agrees with the morphology. Values <2.0 for the mass and <2.5 for the inverse-mass are due to edge effects. Therefore, the weighted local mass dimension differentiates not only between the chromatin and nucleoplasma distribution but also between the different chromatin organisations introduced by the MLS and RW/GL models.

4.4.3 Weighted Local Diffuseness

The local standard deviation, the so called local diffuseness d_{local} (4.2.4) was calculated as for $D_{local-w}$. d_{local} is the same for mass and inverse-mass (Equ. 4.6). As for $D_{local-w}$, the visualization of d_{local} reveals its spatial origin: high d_{local} colocalizes with borders, e. g. of subcompartments, showing again the features of the MLS model (Fig. 4.6A-Cαβ, Fig. 4.7A-Cαβ) and the homogeneity of the RW/GL model (Fig. 4.6Dαβ, Fig. 4.7Dαβ). The change from 63 to 126kbp loops or linkers is again present (Fig. 4.6A&Bαβ, Fig. 4.7A&Bαβ (Fig. 4.6A&BαβI&II, Fig. 4.7-A&BαβI&II), although d_{local} smoothes out distinct structures more than $D_{local-w}$.

The normalized and absolute local diffuseness d_{local} distributions from images with normalized intensities (Fig. 4.8E&F) or absolute nucleosome concentration (Fig. 4.8G&H) exhibits one or two peaks depending on the nuclear radius and den-

sity: For a 3µm radius two peaks, for a 4µm radius only one peak although asymmetric to lower intensities/nucleosome concentrations exists. For 5 and 6µm radii again two peaks exist of which the first has as saddle due to many values near 0.0. The general distribution average is inversely and the distribution width is proportional to the linker (Fig. 4.8E&G) and loop (Fig. 4.8F&H) size. Thus larger loops and linkers lead to a more homogeneous chromatin distribution in agreement with the analysed parameters of simulated nuclei (Chapter 3). The average and its width are also proportional to the resolution due to the smoothing effect of a low 60x objective in contrast to a 100x objective (Fig. 4.8F&H). The comparison between the analyses of the normalized and absolute distributions (Fig. 4.8E&F versus Fig. 4.8G&H) show that the normalized d_{local} distributions distinguish well between the different morphologies. Consequently, the diffuseness distribution is also a reasonable measure to analyse different chromatin organisations in nuclei.

4.4.4 Weighted Local Skewness and Weighted Local Kurtosis

The local skewness s_{local} (4.2.4, Equ. 4.7) qualifies the asymmetry ($s_{local} = 0$ for an ideal Gaussian) and the local kurtosis k_{local} (4.2.4, Equ. 4.8), judges the flattening ($k_{local} < 0$) or sharpening ($k_{local} > 0$) of the local intensity or nucleosome concentration distribution. They were calculated for the (inverse-) mass in a box of side length $l_{Local} = 5 pixels$ around a pixel, excluding background pixels. In general the skewness distribution indicates a strong local asymmetry of the local mass distribution to the right (Fig. 4.8I&J; for the inverse mass, of course to the left according to Equ. 4.8) and the kurtosis distribution reveals a massively sharp local (inverse-) mass distribution (Fig. 4.8K&L). Both the skewness and the kurtosis are proportional to the nuclear radius, the linker and loop size and inversely proportional to the objective resolution in agreement with the expectations from the nuclear morphology. Their spatial mapping is again in agreement with nuclear morphology as for the local mass dimensions (4.4.2) and the diffuseness (4.4.3). Thus, also the skewness and the kurtosis are independent measures for the analyses distinguishing between different nuclear morphologies, although they might differ only slightly.

4.5 Scaling of the Nuclear Membrane

To analyse the shape of the nuclear membrane or nuclei, additionally the pure box-volume function $V_B(l_B)$ and its exponent the pure box-counting dimension $D_B(l_B)$ were calculated for the surface of spherical nuclei with 3, 4, 5 und 6µm radius. Although the real shape of nuclei is more oblath and could be oddly shaped, the analysis set a standard and clarify scaling effects due to the nuclear size and shape.

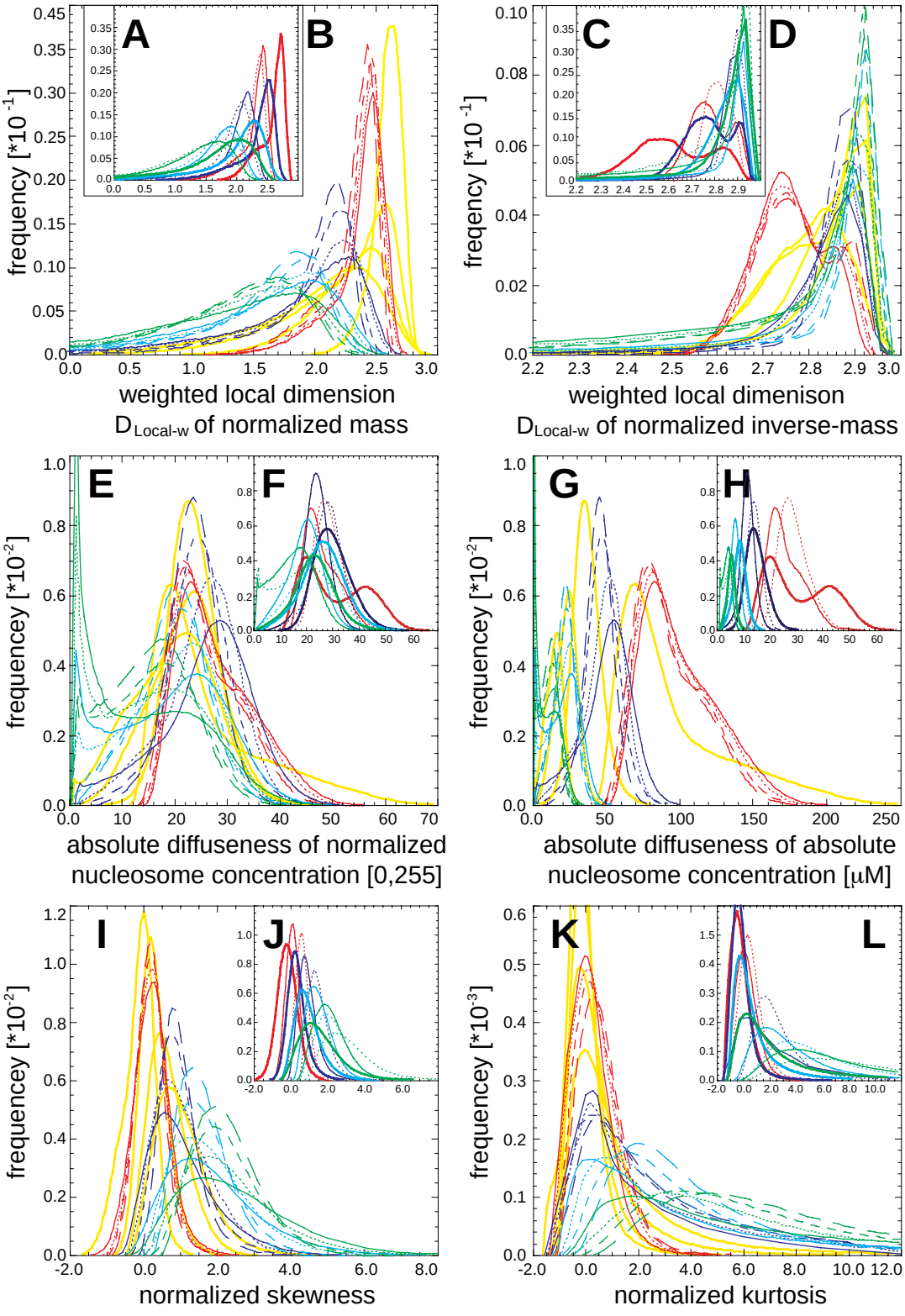
The box-volume function $V_B(l_B)$ shows power-law behaviour with a general slope indicating a surface like scaling over ~2 orders of magnitude (Fig. 4.9A,

4.4.1). This is in agreement with the (inverse-, iso-) mass scaling (Fig. 4.5, 4.4.1). The number of boxes N_B is proportional to the nuclear radius, but not the general slope. The limited pixel resolution leads to a lower cut-off at ~100 nm and the finite size of the nuclei leads to an upper cut-off at 5, 6, 8 and 9 µm. This is ~25% less than the nuclear diameters and in agreement with scaling theory of finite objects. The pure local box-counting dimension $D_B(l_B)$ initiates at 1.8, before increasing to 2.0 and a little beyond shortly before the cut-off. Thereafter, $D_B(l_B)$ decreases to ~0.5. While approaching the cut-off fluctuations set in due to lower statistics for growing l_B. The constant underestimate of $D_B(l_B)$ being $\leq$2.0 is due to the notorious underestimate by the local box-counting dimension in agreement with theoretic prediction. Thus, for l_B smaller than ~1000 to 2000 nm depending on the nuclear radius, the spherical membrane is near to a smooth surface. Thereafter its curvature is determined, before the membrane is measured as a sphere and finally reduced to a point object. Different pixelation with pixelsizes of 70 and 80 nm for the high resolution 100x 1.4 and the low resolution 63x 1.2 objective leads only to a shift of the scaling behaviour (Fig. 4.9 A&B). Consequently, the general and local box-counting dimensions D_B and $D_B(l_B)$ are also an accurate measure to analyse the shape of the nuclear membrane and thus the whole nucleus.

4.6 Discussion of Scaling Properties of the Nuclear Higher Order Structure

The three-dimensional organization of the human genome consists of several stages of packaging spanning huge length and time scales with unknown scaling behaviour. Since the development of scaling or fractal analyses a huge number of meth-

Fig. 4.8 Weighted Local Dimension, Diffuseness, Skewness and Kurtosis
The unnormalized weighted local dimension $D_{local-w}$ distributions of the density mass (**A**, **B**) and inverse density mass (**C**, **D**) distribution are characterized by one major peak located between 2.0 and 2.5 for the mass and between 2.5 and 2.9 for the inverse mass. For small nuclei of 3µm radius the inverse mass shows two peaks due to the different scaling within and at the membrane of the nucleus. The RW/GL model leads to much higher values showing their higher homogeneity. The peak height and position are mainly proportional to the nuclear radius for the (inverse-) mass, inversely proportional to linker size, loop size and objective resolution for the mass and proportional in the case of the inverse mass (**A**-**D**). The normalized (on a scale from 0 to 256; **E**,**F**) and absolute nucleosome concentration (**G**, **H**) local diffuseness d_{local} distribution consists of two peaks (merging for 4 and 5 µm radii) which are proportional to mainly the nuclear volume, the loop and linker length as well as the objective resolution. The skewness distribution indicates a local imbalance of the mass distribution to the right (the inverse-mass distribution is unbalanced to the left) and the kurtosis reveals a sharp (inverse-) mass distribution. The skewness s_{local} and kurtosis k_{local} are proportional to the nuclear radius, the linker and loop size and inversely proportional to the objective resolution. (**A**, **C**, **E**:126-X-MLS-^{X}, linkers of 63, 126,189 and 252 kbp are solid, dotted, dashed, and long dashed, nuclear radii of 3, 4, 5 and 6 µm are red, blue, light blue and green, 1320-RW/GL^3 are thick yellow; **B**, **D**, **F**: X-126-MLS^X, loops of 63, 126 kbp are dotted and thin, 100x, 60x objective are thin and thick lines.)


A
B
C
D
E
F
G
H
I
J
K
L
frequency [*10 -1]
weighted local dimension
D Local-w of normalized mass
freqency [*10 -1]
weighted local dimenison
D Local-w of normalized inverse-mass
frequencey [*10-2]
absolute diffuseness of normalized
nucleosome concentration [0,255]
absolute diffuseness of absolute
nucleosome concentration [μM]
normalized skewness
frequencey [*10-3]
normalized kurtosis

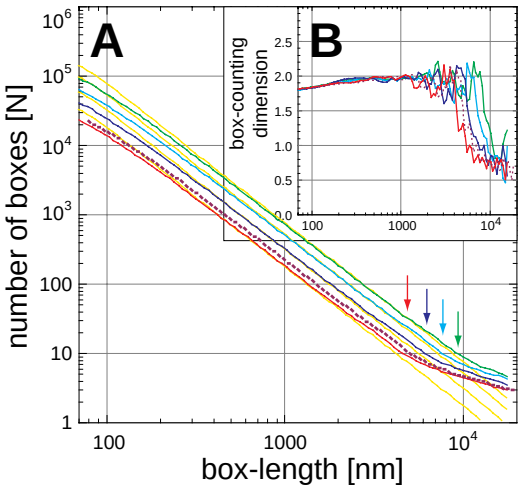


Fig. 4.9 Nuclear Membrane/Surface Dimension of Simulated Nuclei
Calculation of the pure box-volume function for the pixels belonging to the nuclear membrane by counting the occupied boxes N_B results in power-law behaviour as function of the boxlength l_B over ~2 orders of magnitude (**A**, radii of 3, 4, 5, and 6 µm are red, blue, light blue and green, for an initial pixel size of 70 nm used for high resolution 100x 1.4 oil immersion objective). For different sized nuclei, N_B is, proportional to the nuclear radius. A lower cut-off at ~100 nm due to the pixel approximation of the membrane surface and an upper cut-off due to the finite nuclear size at ~ 5, 6, 8 and 9 µm (arrows) exist, where the general pure box-counting dimension D_B deviates from the theoretic surface value of 2.0 (yellow lines). Thus, the determination of the pure local box-counting dimension $D_B(l_B)$ is feasible (**B**): from an initial value 1.8, $D_B(l_B)$ increases to values of ~2.0 and little beyond before the upper cut-off is reached and $D_B(l_B)$ decreases with fluctuations to ~0.5. Thus, for l_B smaller ~1000 to 2000 nm the spherical membrane is near to smooth surface, before its curvature is determined and before the surface is measured as a sphere reduced to a point object. The difference using a pixel size of 80 nm for low resolution 63x 1.2 water immersion objective leads only to a shift but not to a general difference in the scaling behaviour (**A**, **B**; dotted dark red).

ods were developed and applied to near to every existing subject (Mandelbrot, 1977; Brickmann & Bär, 1986; Stanley & Ostrowsky, 1986; Stanley & Ostrowsky, 1988; Kaye, 1989; Avnir, 1992; Peitgen *et al.*, 1992; Falconer, 1993; Nonnenmacher *et al.*, 1993; Novak, 1994; Harrison, 1995; Hastings & Sugihara, 1996; Rodríguez-Iturbe & Rinaldo, 1997), ranging e. g. from the art of Pollock's drip paintings (Taylor, 1999), over music (Hsü & Hsü, 1990; Hsü & Hsü, 1991), to the shape of fern leaves (Tayloer, 1999). From the succeeding principles Takahashi (1989) even proposed a fractal model of chromosomes and chromosomal DNA replication. The success of scaling analyses has also led to their use in pathological specimens to characterize benign from malign behaviour (Rigaut, 1983; Irinopoulou *et al.*, 1993; Landini & Rippin, 1993; Nonnenmacher *et al.*, 1993; Kriete, 1996; Sandau & Kurz, 1996; Byng *et al.*, 1997; Cross *et al.*, 1997; Haidekker *et al.*, 1997; Einstein *et al.*, 1998a; Einstein *et al.*, 1998a).

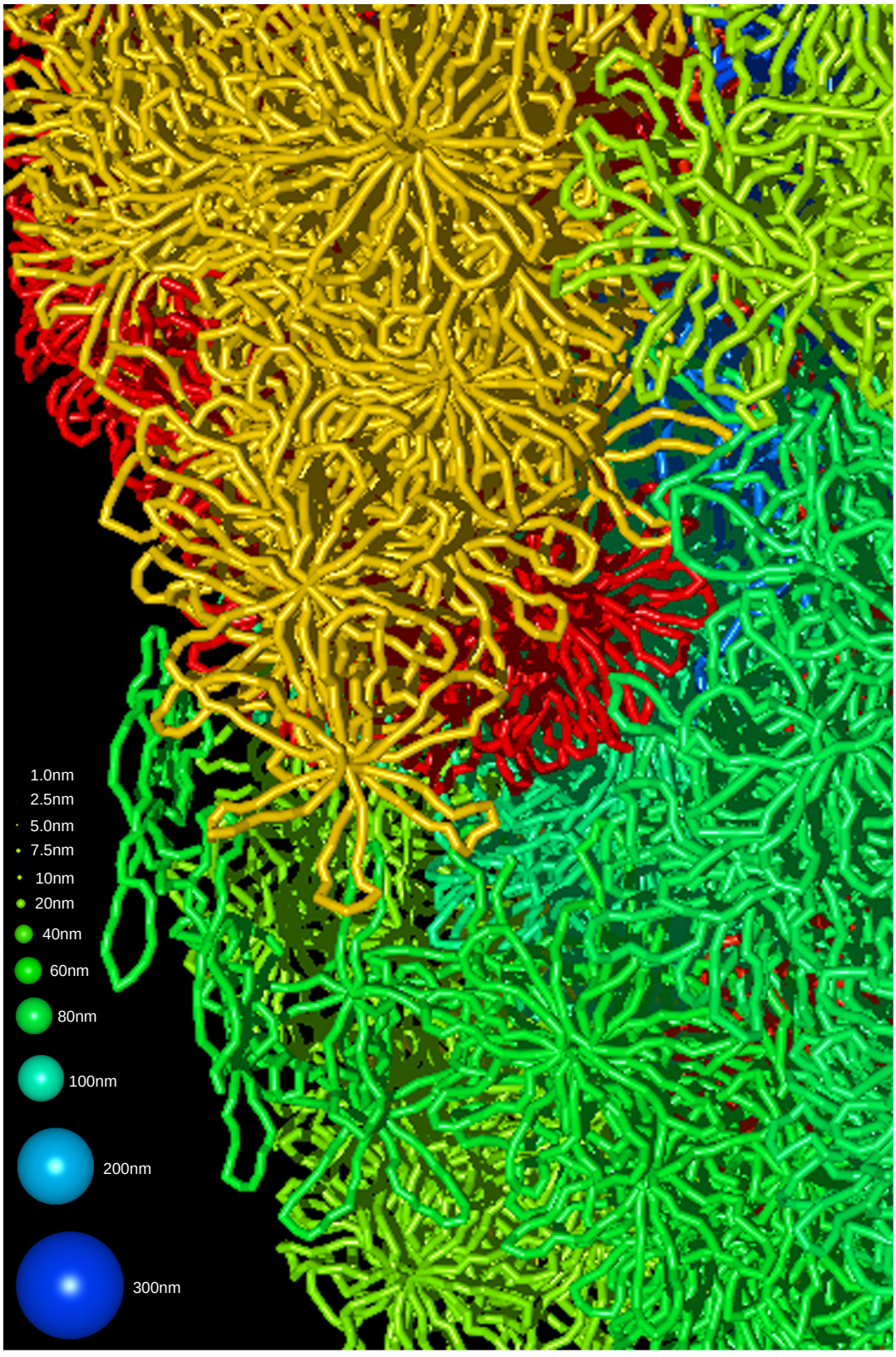
In order to describe the topology of the chromatin fiber and the morphology of confocal laser scanning microscopic (CLSM) image stacks, their scaling behaviour was investigated. The topology and the morphology were obtained from simulations

of single chromosomes (Chapter 2) and whole nuclei (Chapter 3). The simulations were based on the Random-Walk/Giant-Loop (RW/GL) model, in which large loops are attached to a flexible backbone, and the Multi-Loop-Subcompartment (MLS) models, in which small loops form rosettes, connected by a linker. For the analysis, a variety of scaling/fractal dimensions were calculated: The scaling of the one-dimensional axis of the 30 nm chromatin fiber was analysed with the exact spatial-distance and yard-stick dimensions and its voluminous scaling was investigated by the pure box-counting dimension. The scaling behaviour of the (inverse-, iso-) mass distribution of CLSM image stacks was investigated with the weighted box-counting, lacunarity and local mass dimensions. To support the latter, additionally, the local diffuseness, skewness and kurtosis distributions were calculated.

The scaling of the chromatin fiber revealed different power-law behaviours on different scales: in the MLS model four and in the RW/GL model three major scaling regimes existed. In the MLS model these are attributable to the stiff one-dimensional chain segments, the random walk attributes of the chain, the globular nature of small loops and rosettes, and the random walk arrangements of the rosettes. In the RW/GL model, the second and third region dominate and are shifted to higher length scale. Within the multi-scaling a fine-structure was present in the MLS model due to the rosette loops. This fine-structured multi-scaling agrees with the correlation behaviour in the DNA sequence of human chromosomes. Thus, the sequential and three-dimensional organization of genomes are closely interconnected.

The general scaling behaviour of CLSM image stacks showed also power law behaviour as function of (iso-, inverse-) mass thresholds. The scaling dimensions reflected the model and imaging properties in detail. Determination of the local diffuseness, skewness and kurtosis as well as their spatial visualization supported these results. Thus, the topology of the chromatin fiber and the morphology of nuclei are closely connected. Therefore, scaling analyses of the nuclear morphology are a suitable approach to differentiate between different cell states, e. g. during the cell cycle, due to malignancy, in apoptosis or in response to drugs.

Consequently, the analyses of the scaling behaviour links on the one hand the sequential with the three-dimensional organization of nuclei and on the other hand connects changes of this three-dimensional organization to morphological changes on the level of the whole nucleus.


1.0nm
2.5nm
5.0nm
7.5nm
10nm
20nm
40nm
60nm
80nm
100nm
200nm
300nm

5 Simulation of the Dynamics in Interphase Nuclei

5.1 Introduction

The impact of the three-dimensional genome organization on molecular mobility is closely related to many nuclear processes. Here the diffusion of spheres was simulated by Brownian Dynamics in computer generated nuclei with a Multi-Loop-Subcompartment (MLS) chromatin fiber topology. The tracers interacted with the static fiber by an excluded volume potential. Visual inspection of the nuclear morphology revealed big spaces allowing high accessibility to nearly every spatial location. This is supported by estimations of the nuclear volume occupied by chromatin <30%, leaving >70% of space for diffusion with an average mesh spacing of 29 to 82 nm for nuclei of 6 to 12 µm diameter. This agrees with the simulated displacement for 10 nm sized particles of ~1 to 2 µm within 10 ms. Therefore, the diffusion of biologically relevant tracers is only moderately obstructed. The anomaly parameter D_W characterizing the degree of obstruction ranged from 2.0 (obstacle free diffusion) to 4.0, in agreement with experiments. The degree of obstruction was proportional to the nuclear density, the fiber diameter, the interaction hardness and the tracer size. Different fiber topologies had no effect on the average particle displacement. Consequently, molecules and proteins might reach every nuclear location by energy independent diffusion without a special channel like network.

Fig. 5.1 Morphology of Obstructed Dynamics and Diffusion in the Organization of Nuclei
The detailed few from the outside onto the rendered three-dimensional organization resulting from a 126-126-MLS^6 nucleus (Chapter 3), demonstrates the structure and low overlap of chromosome territories and the rosette like subcompartments and that the mean spacing between chromatin fibers ranges at least from 50 to 100 nm for this nuclear radius (Tab. 5.1). Hence, the obstruction of diffusing particles (spherical legend in image) is proportional to their size. Thus, small chemical substances as nucleotides or ATP molecules reach every location in the nucleus and most relevant proteins or protein subunits should only be obstructed moderately. Consequently, active transport of molecules should be restricted to few exceptions as well as a channel like network for transportation. Additionally, the lack of a planelike space between chromosome territories and the size needed to be distinct from the visible voids refute the Inter Chromosomal Domain (ICD) hypothesis.

5.2 Simulation Methods

Particles free in solution experience due to statistical interactions with the molecules of the solvent Brownian motion.

5.2.1 Analytical Description of Obstructed Diffusion

Without external forces, the particle concentration $c(\boldsymbol{r}, t)$ is connected to the particle flux $\boldsymbol{j}(\boldsymbol{r}, t)$ by Fick's first law

$$\boldsymbol{j}(\boldsymbol{r}, t) = -D(\boldsymbol{r}, t)\nabla c(\boldsymbol{r}, t) \tag{5.1}$$

where $\boldsymbol{D}(\boldsymbol{r}, t)$ is in the most general case a tensor being reduced in an infinite and isotropic solution to the constant diffusion coefficient D_0. The diffusion coefficient is inversely proportional to the friction coefficient f and thus the viscosity η and to the Stoke's or hydrodynamic radius R_h, thus

$$D_0 = \frac{k_B T}{f} = \frac{k_B T}{6\pi\eta R_h} \tag{5.2}$$

with the Boltzmann's constant k_B and the temperature T. Equ. 5.1 and the continuity equation

$$\frac{\partial}{\partial t} c(\boldsymbol{r}, t) = -\nabla \boldsymbol{j}(\boldsymbol{r}, t) \tag{5.3}$$

result in Fick's second law

$$\frac{\partial}{\partial t} c(\boldsymbol{r}, t) = D_0 \nabla^2 c(\boldsymbol{r}, t). \tag{5.4}$$

The steady-state concentration for a finite sample with N particles is $P_{D_0}(\boldsymbol{r}, t) = N/V$. Fluctuations of the concentration due to particle motions following a Markovian process are described by the transition probability $P_{trans-D_0}$ for a particle from a location $\boldsymbol{r}_1$ at time t to a location $\boldsymbol{r}_2$ at time $t+\tau$ is

$$P_{trans-D_0}(\boldsymbol{r}_2, t+\tau|\boldsymbol{r}_1, t) d^3 r_2 = P_{trans-D_0}(\boldsymbol{r}_2, \tau|\boldsymbol{r}_1, 0) d^3 r_2. \tag{5.5}$$

With the boundary condition $c(\boldsymbol{r}, 0) = \delta(\boldsymbol{r})$ the transition probability is

$$P(\boldsymbol{r}_2, \tau|\boldsymbol{r}_1, 0) = c(\boldsymbol{r}_2 - \boldsymbol{r}_1, \tau) = \frac{1}{(4\pi D_0 \tau)^{D_T/2}} \cdot \exp\left[-\frac{|\boldsymbol{r}_2 - \boldsymbol{r}_1|^2}{4 D_0 \tau}\right] \tag{5.6}$$

using the number of translational degrees of freedom D_T. The mean square displacement of the particles is then given by

$$\left\langle |\boldsymbol{r}_2 - \boldsymbol{r}_1|^2 \right\rangle = 2 D_T D_0 \tau. \tag{5.7}$$

In a cell nucleus the diffusion of particles is constrained by the chromatin fiber. The structural properties and the space between the fibers can be described by scaling laws (Chapter 4). The diffusion or transport of particles in such a statistical sys-

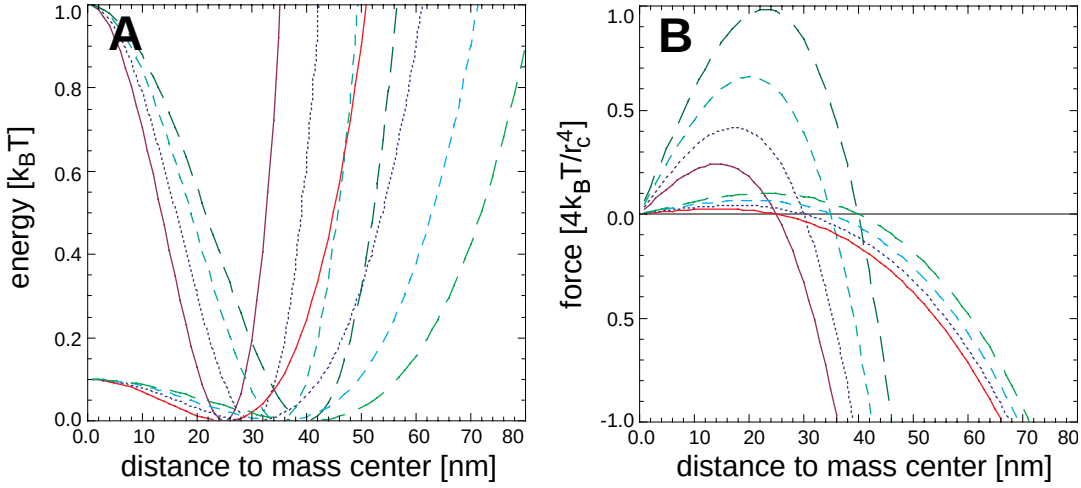


Fig. 5.2 Excluded Volume Potential and Forces
The Excluded Volume potential (**A**, Equ. 2.6) and its force (**B**) avoids crossing of the chromatin fiber or spheres for a high prefactor U_0 of 1.0 (dark colours), in contrast to a low prefactor of 0.1 (light colours). This is proportional to the chromatin fiber diameters(diameters of 25, 30, 35 and 40 nm are solid, dotted, dashed and long dashed). The Excluded Volume potential and force was, only relevant below the critical distance for which it was repulsive, else it is a repulsive-binding potential.

tem is called percolation (de Gennes, 1976; Gefen & Aharony, 1981; Herrmann *et al.*, 1984; Stauffer, 1995) and depends critically on the distribution and concentration of the mass or obstacles (Halperin *et al.*, 1985; Witten & Cantor, 1984). For concentrations c_0 above the critical percolation threshold/limit p_c the diffusion is blocked (percolation freeze-out). For two- and three-dimensional grids the threshold is reached at concentrations c_0 of 0.41 and 0.69, respectively, considering particles and obstacles of similar size (Bunde & Havlin, 1996). For statistical systems with unhomogeneous or multi-scaling behaviour (Chapter 4), the percolation limit varies as function of the tracer size (Stanley, 1984; Orbach, 1986; Sznitman, 1991; Wiggins, 1991; Voss, 1992). Therefore, the mean square displacement does not scale linearly as a function of time (Saxton, 1987-1995), but follows

$$\left\langle \left| \boldsymbol{r}_2 - \boldsymbol{r}_1 \right|^2 \right\rangle \propto \tau^{2/D_W} = 2D_T \frac{\omega_0^2}{4\tau_{diff}} \left(\frac{\tau}{\tau_{diff}} \right)^{2/D_W - 1} \tau = 2D_T D(\tau)\tau \qquad (5.8)$$

with the scaling dimension of the particle random walk D_W, also called the anomaly parameter (Alexander & Orbach, 1982; Aharony, 1984; Lobb & Frank, 1984; Nagle, 1992; Qian *et al.*, 1999), on the scale of observation ω_0. This anomalous diffusion behaviour can be described with a time-dependent diffusion coefficient

$$D(\tau) = D_0 \left(\frac{\tau}{\tau_{diff}} \right)^{2/D_W - 1} \text{ with } \tau_{diff} = \frac{\omega_0^2}{4D_0}. \qquad (5.9)$$

A free random walk corresponds to $D_W = 2.0$ (Rammal & Toulouse, 1983; Rudnick & Caspari, 1987), whereas D_W increases in the presence of obstacles. For sta-

tistical systems near the percolation limit D_W equals ~2,87 in two and ~3.80 in three dimensions (Bunde & Havlin, 1996). This holds for a pure exclusion interactions between particle and obstacles, and is biased for other, e. g. binding interactions (Havlin & Weiss, 1984; Harder & Havlin, 1985; Wachsmuth, 2001).

5.2.2 Simulation Method

The diffusion of 100 to 1000 spheres with 1 to 300 nm diameters was simulated by Brownian dynamics in nuclei with diameters of 6 to 12 µm. The nuclei were simulated according to Chapter 3. The Multi-Loop-Subcompartment (MLS) model with loops and linkers of 63 to 126 kbp were used for the folding of the 30 nm chromatin fiber. This fibrous network was kept static while the movement of the particles was calculated by the Brownian Dynamics algorithm (2.2.3) also applied for the simulation of single chromosomes and whole nuclei (Chapter 2, Chapter 3). The diffusing spheres interacted with the chromatin fiber by the excluded volume potential and force (Equ. 2.6, Fig. 5.2) used also for keeping the chromatin fibers from crossing (Chapter 2, Chapter 3). Beyond the nuclear radius, the fiber diameter was variated between 25, 30, 35 and 40 µm. The hardness of the excluded volume potential was set to U_0 of 1.0 and 2.0. The change in U_0 leads to a broader potential equivalent to changing the fiber diameter. Throughout Chapter 5 the following nomenclature L_S-LI_S-$MLS^{r\text{-}nucleus}$ (L_S: loop size; LI_S: linker size; r-nucleus: nuclear radius) is used.

5.3 Morphologic Volume and Diffusion Relationships

A morphologic feature visible in the rendered images of simulated single chromosomes (Fig. 2.1) and whole nuclei (Fig. 3.1, Fig. 3.2A-Cαγϕ, Fig. 4.1, Fig. 5.1, Fig. 5.3, Fig. 9.1) is the existence of big unoccupied spaces within chromosome territories. These voids allow high accessability to the interior of these territories for tracers of corresponding size. Therefore, the definition of the surface of a chromosome territory depends on the scale of the probing tracers or observation. While, for a large particle with 500 nm diameter the chromosome territory is inpenetratable, for particle diameters <10 nm the chromatin fiber itself is the surface of chromosomes. Nevertheless, it is possible to imagine the embedding hull around a chromosome territory possibly defined by chemical markers. The big subcompartment sized voids in the MLS model (Fig. 1D) or an outlooping big loop or chain of loops in the RW/GL model (Fig. 1C&D) are of course occupied by other chromosomes in whole nuclei. (Fig. 3.1, Fig. 3.2A-Cαγϕ, Fig. 4.1, Fig. 5.1, Fig. 5.3, Fig. 9.1). However, the space between the chromatin fiber stays obviously in a range from 40 to 120 nm. The volume fraction taken by the 30 nm chromatin fiber and the mean spacing of a nuclear isotropic distribution of the fiber can be estimated quantitatively.

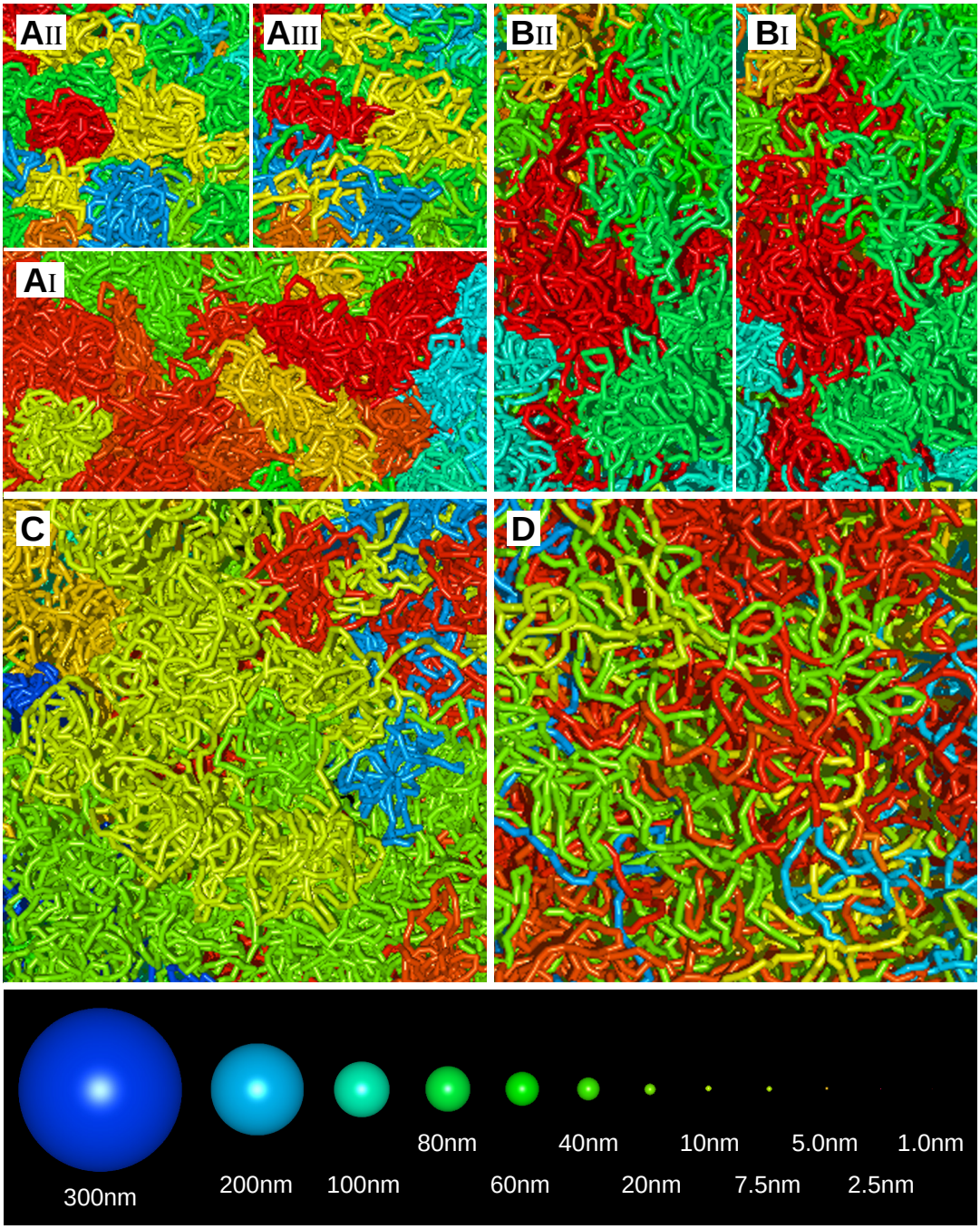


Fig. 5.3 Visualization of the Volume Relationships in Rendered Nuclei
The different volume occupying properties and morphologies of chromosome models 63-63-MLS[3] (**AI**), 63-252-MLS[3] (**AII**), 126-252-MLS[3] (**AIII**), 84-126-MLS[4] (**BI**), 84-126-MLS[4] (**BII**), 126-126-MLS[5] (**C**), 1320-RW/GL[6] (**D**), are visible in close-ups of the three-dimensional rendering of the actual 30 nm fiber and agree with other morphologic aspects (Fig. 3.2). The mean spacing behaves like the theoretic prediction (Tab. 5.1) and is at least 50 to 100 nm for nuclei with 6 µm radius. Thus again the diffusion of small molecules and proteins is only obstructed moderately and thus no active transportation processes and a one- or two-dimensional distinct channel network seem necessary.

Tab. 5.1 General Volume, Concentration and Spacing Properties of Nuclei
The mean nuclear nucleosome concentration was calculated using the nucleosome number from 5.3.2, the volume fractions using a cylindrical and a nucleosomal approximation according to 5.3.1 and 5.3.2. The mean isotropic mesh spacing was estimated according to 5.3.3.

Nuclear Diameter [µm]	Nuclear Volume [µm^3]	Mean Nuclear Nucleosome Concentration [µM]	Chromatin Volume Fraction [% of Nuclear Volume]				Mean Isotropic Mesh Spacing [nm]	
			Cylindrical 30nm Fiber	Cylindrical 40nm Fiber	Nucleosome + DNA	Nucleosome +DNA+RNA +protein	1x Chain	2x Chain
6	115	251	41.0	73.0	20.1	60.0	41	29
8	268	107	17.5	31.3	8.6	26.0	64	45
10	523	55	9.0	16.0	4.4	13.3	90	63
12	904	32	5.2	9.3	2.6	7.7	117	82

5.3.1 Volume Fraction of the Chromatin Fiber in Cell Nuclei

The volume fraction occupied by the chromatin fiber could be estimated by two approximations leading to a lower and an upper limit:

Cylindrical Approximation of the 30nm Chromatin Fiber: In the first approximation the 30nm chromatin fiber is considered as a homogeneously filled cylinder with a diameter of 30nm and a basepair density of 105bp per nanometer fiber length. The human genome consists of around 7x10^9 bp which divided by the basepair density yields a total fiber length of 6.66x10^{4}µm = 6.66cm. The volume of the 30nm chromatin fiber of the human genome then is

$$V_{chromatin,\,zylinder} = \pi \cdot (0.015\mu m)^2 \cdot 6.66 \cdot 10^4 \mu m \approx 47\mu m^3 . \qquad (5.10)$$

Including entropic interaction diameter of the fiber and a contribution of associated proteins results perhaps in a 40nm diameter with a volume of 84µm^3. Consequently, for a typical nucleus of 5µm radius (Monier, 2000) the upper limit of chromatin volume fraction ranges from 9.0 to 16% (Tab. 5.1). For very small nuclei of 3µm radius the latter approximation crosses even the percolation limit (Tab. 5.1).

Nucleosomal Approximation of the 30nm Chromatin Fiber: A finer approach considers the cylindrical nucleosomes within the chromatin fiber (1.2.2) with volume

$$V_{nucleosome} = \pi \cdot (5.5nm)^2 \cdot 6nm \approx 568nm^3 \qquad (5.11)$$

based on the atomic nucleosome structure (Luger *et al.*, 1997). A nucleosome consists of the histone octamer and one H1 histone, totaling 128kDa and of 146bp DNA per nucleosome with 96kDa. Thus, a nucleosome weighs 224kDa with a dry mass density of $\delta_{nucleosome} \approx 0.39$kDa/nm^3 using $V_{nuclesome}$. There are 7x10^{9}bp/200bp≈3.5x10^{7}nucleosomes in a human cell nucleus (including the linker DNA between the nucleosomes). The total dry mass in a cell nucleus of histones therefore

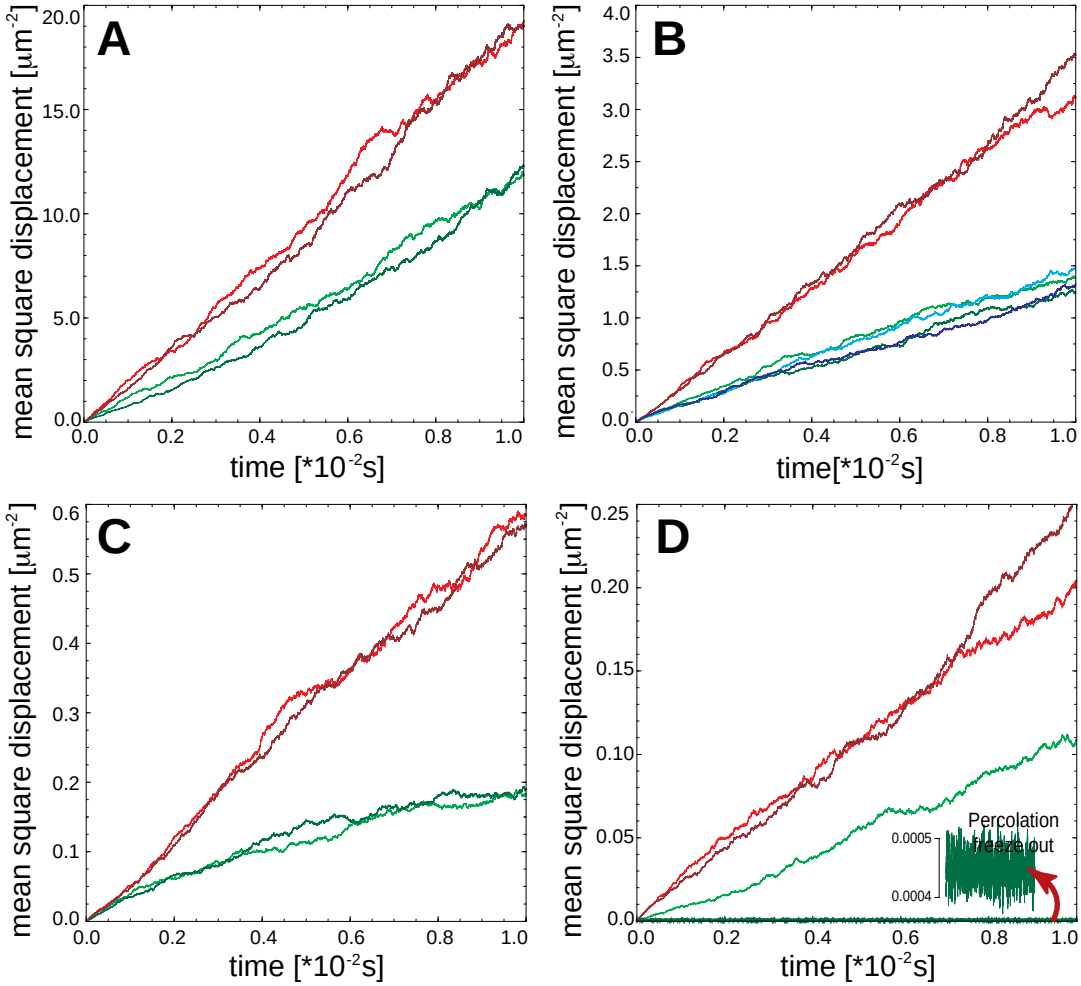


Fig. 5.4 Mean Square Displacement of Spheres in a Static Nuclear Chromatin Mesh
The mean square displacement of spheres with hydrodynamic diameters of R_h 1 (**A**), 5 (**B**), 25 (**C**) and 50nm (**D**) is mainly proportional to the size of the nucleus (3µm red and 6µm blue) and thus its mean density as well as to the size of the spheres. Due to the same mean density also no effects of different chromosomal morphologies were found (all with 126-126-MLS^X, and 63-126-MLS^3 blue **B**). For small spherical radius and low density a change in the excluded volume interaction resulted in no apparent differences (high U_0 is dark, low U_0 is light).In small nuclei with a high density the increase to a high U_0 leads to the slight increase necessary to increase the density just over the percolation threshold/limit resulting in percolation freeze-out, despite small spatial fluctuations around the particles mean position (**D**). Apparently, the dependency of the mean square displacement is not a linear function of time. This indicates anomalous diffusion.

is $m_{histones} \approx 4.5\text{x}10^9\text{kDa} \approx 7.2\text{pg}$ and a total dry mass of DNA $m_{DNA} \approx 4.6\text{x}10^9\text{kDa} \approx 7.3\text{pg}$. This is in very good agreement with experiment (Cold Spring Harbour, 1978). Therefore, the volume of the chromatin fiber here results in

$$V_{chromatin,\,nucleosome} \approx \frac{m_{histones}}{\delta_{nucleosome}} + \frac{m_{DNA}}{\delta_{nucleosome}} \approx 23.2\mu m^3 \qquad (5.12)$$

The corresponding volume fractions of the entire chromatin fiber are 4.4% for a spherical cell nucleus with 10 µm diameter. In a nucleus non-histone proteins total another ≈14 pg and substances like RNA another ≈14 pg. With the reasonable estimate of a similar density of these components the volume fraction increases to 8.9%. For nuclei of 3 µm radius the latter stays just below the percolation limit.

In summary, the cylindrical and the nucleosomal approximation of the volume fractions of the chromatin fibers stays for standard nuclei of 525 μm^3 between 2.6 and 16.0%. Although in very small nuclei of 115 μm^3 the standard estimate ranges from 20.1 to 41.0% and is near or crosses the percolation limit in the extended approximation, the latter needs to be carefully interpreted since i*n vivo* cells with such small nuclei, are very inactive and hardly any nuclear processes take place. Taking also the results for nuclei with 8 µm diameter and 268 μm^3 (Tab. 5.1) into account, <30% of the nucleus might be occupied volume, thus more than 70% of the volume in typical cell nuclei are available space for diffusion. This is far from the percolation limit and suggests for small particles <10 nm only a moderate obstruction of their diffusion behaviour.

5.3.2 Approximation of the Mean Isotropic Mesh Spacing

Although the results of 5.3.1 suggest a rather "empty" nucleus, the mean spacing of chromatin fiber is of equal importance: Neglecting a specific chromosome model, a theoretic scaling estimate of the mean spacing L_{mesh} in a isotropic mesh of chromatin fibers can be obtained by assuming a spherical volume filled with a three-dimensional grid of cubes. The edges of the cubes are occupied by the chromatin fiber. Then the nuclear volume $V_{nucleus}$ and the total edge length of the mesh which is the total length of the chromatin fiber l_{fiber}, are

$$V_{nucleus} = NL_{mesh}^3 \text{ and } l_{fiber} = N3L_{mesh} \tag{5.13}$$

with the number of cubes N and the edge length of the cubes, which is the mesh spacing L_{mesh}. Thus, the mean mesh spacing can be expressed by

$$L_{mesh} \propto \sqrt{\frac{V_{nucleus}}{l_{fiber}}}. \tag{5.14}$$

For a spherical nucleus with 10 µm diameter and a chromatin fiber of 6.6x10^4 µm length, L_{mesh} is around 90 nm (Tab. 5.1). A hypothetical doubling of the fiber length, e. g. due to a partial decondensation of the chromatin conformation results in $L_{mesh} \approx 60$ nm. Despite the unknown distribution of non-histone proteins, many of these associate to the chromatin fiber reducing the mesh spacing. RNA and decondensed 30 nm chromatin fibers or nucleosome free DNA contribute also to a smaller mesh spacing. Nevertheless, the mesh spacing is >29 to 82 nm for nuclei of 6 to 12 µm diameter (Tab. 5.1), suggesting again that particles <10 nm show only moderately obstructed diffusion.

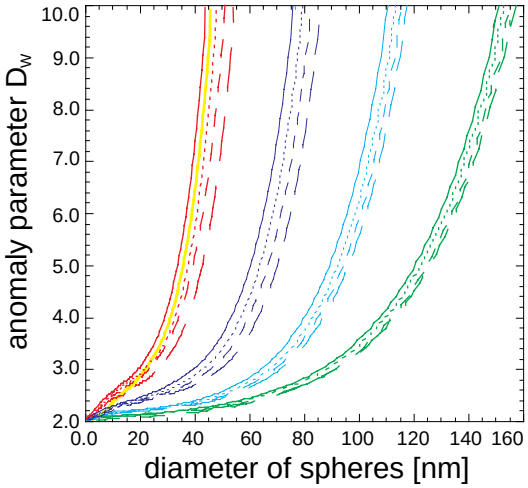


Fig. 5.5 Obstructed Diffusion Coefficient of Spheres in a Static Nuclear Chromatin Mesh
The anomaly parameter D_W indicating the degree of diffusion obstruction for spheres as function of their diameter is in the first place proportional to the nuclear radius and therefore the nuclear density (nuclear radii of 3, 4, 5, and 6 µm are red, blue, light blue and green). Changes of the chromatin fiber diameter leads to a smaller proportionality (diameters of 25, 30, 35 and 40 nm are solid, dotted, dashed, and long dashed). Increasing the excluded volume interaction between spheres and the chromatin fiber from a U_0 of 1.0 to 2.0 shifts D_W to higher values equivalent to changing the fiber diameter from 25 to 30 nm (as an example: yellow line). All simulations were done in nuclei with a 126-126-MLS^X chromosome topology.

5.4 Particle Diffusion in a Static Nuclear Chromatin Mesh

The morphologic relationships in the nucleus suggested only moderately obstructed diffusion of typical sized molecules and proteins in the nucleus, since the volume occupancies are well below the three-dimensional percolation limit assuming standard cellular conditions. To test this assumption the diffusion of particles with diameters from 1 to 300 nm was simulated with the Brownian Dynamics algorithm introduced in Chapter 2 (2.2.3). The spheres interacted with the kept static chromatin fiber through the same excluded volume interaction keeping also the fibers from crossing during the simulation of single chromosomes and whole nuclei (Equ. 2.6, Fig. 5.2). The nuclei used were chosen among those simulated in Chapter 3. 100 to 1000 spheres were randomly placed in the nucleus and diffused for 1.0 s.

A mean square displacement of spheres exists for particles of up to ~50 nm in nuclei of 3 µm diameter and up to ~120 nm in 6 µm (Fig. 5.4). A particle of 10 nm diameter moves around 1 to 2 µm within 10 ms and therefore could rapidly cross a whole nucleus. Thus, the displacement is proportional to the nuclear size and density as well as to the size of the spheres. This agrees with the theoretic prediction. No differences in the mean square displacement were found for different chromosome topologies (Fig. 5.4B) since the average nuclear density and mean isotropic

mesh spacing are not influenced by different chromatin topologies. However, locally an influence might be present. Changes of the interaction potential by its prefactor U_0 from 1.0 to 2.0, did not result in big shifts of the displacement in the case of small spheres. For big spheres, however, this small change could even lead to a crossing of the percolation limit (Fig. 5.4D). In this so called percolation freeze-out the particle is trapped and fluctuates only thermally around one position.

Apparently the mean square displacement is not a linear function of time since the chromatin fiber acts as obstacle. This could either be described by a time dependent diffusion coefficient or better with the anomaly parameter D_W (Equ. 5.8, Equ. 5.9). For the simulated diffusion of particles D_W is proportional to the nuclear radius and density, the diameter of the chromatin fiber and the hardness of the excluded volume interaction (Fig. 5.5). For small spheres and large nuclei, D_W ranges typically from 2.0 (the value indicating obstacle free diffusion) to 3.0 or 4.0. For the case in which spheres and nuclear density approach the percolation limit, D_W tends to infinity. Thus, a mean square displacement of 0.0 is connected to a $D_W = \infty$.

Qualitatively the results of the mean square displacement and the calculation of the anomalous parameter D_W are in agreement with recent experiments (Misteli, 2001; Wachsmuth *et al.*, 2000, Wachsmuth, 2001). Quantitatively, the simulated D_W is smaller than in experiments, since in the simulations only the chromatin fiber was simulated. Therefore, assuming a chromatin fiber of 40nm diameter which is equivalent to adding the amount of proteins and RNA present in real nuclei (Tab. 5.1, 5.3.3), leads to much greater agreement with experiments.

5.5 Discussion of the Simulation of Dynamics of Interphase Nuclei

The dynamics of large cellular structures, especially those of the three-dimensional organization of the cell nucleus, and the diffusion of small molecules and proteins within this structural framework are subject of current research. The investigation of these processes are mainly based on fluorescence methods like fluorescence recovery after photobleaching (FRAP) or fluorescence correlation spectroscopy (FCS). Whereas FRAP measurements led to many quantitative analyses, the single molecule FCS technique is still under development (Brock *et al.*, 1998; Brock *et al.*, 1999; Schwille *et al.*, 1999; Gennerich & Schild, 2000).

Using photobleaching methods the hydrodynamic investigation of the cytoplasm revealed for macromolecules an increased viscosity of 2,5 to 10 times that of water (Luby-Phelps *et al.*, 1986; Luby-Phelps *et al.*, 1987; Luby-Phelps *et al.*, 1993; Luby-Phelps, 1994, Kao *et al.*, 1993; Swaminathan *et al.*, 1996; Swaminathan *et al.*, 1997; Seksek *et al.*, 1997; Partikian *et al.*, 1998). Investigation of particles in the denser network of the cell nucleus revealed a moderately obstructed diffusion

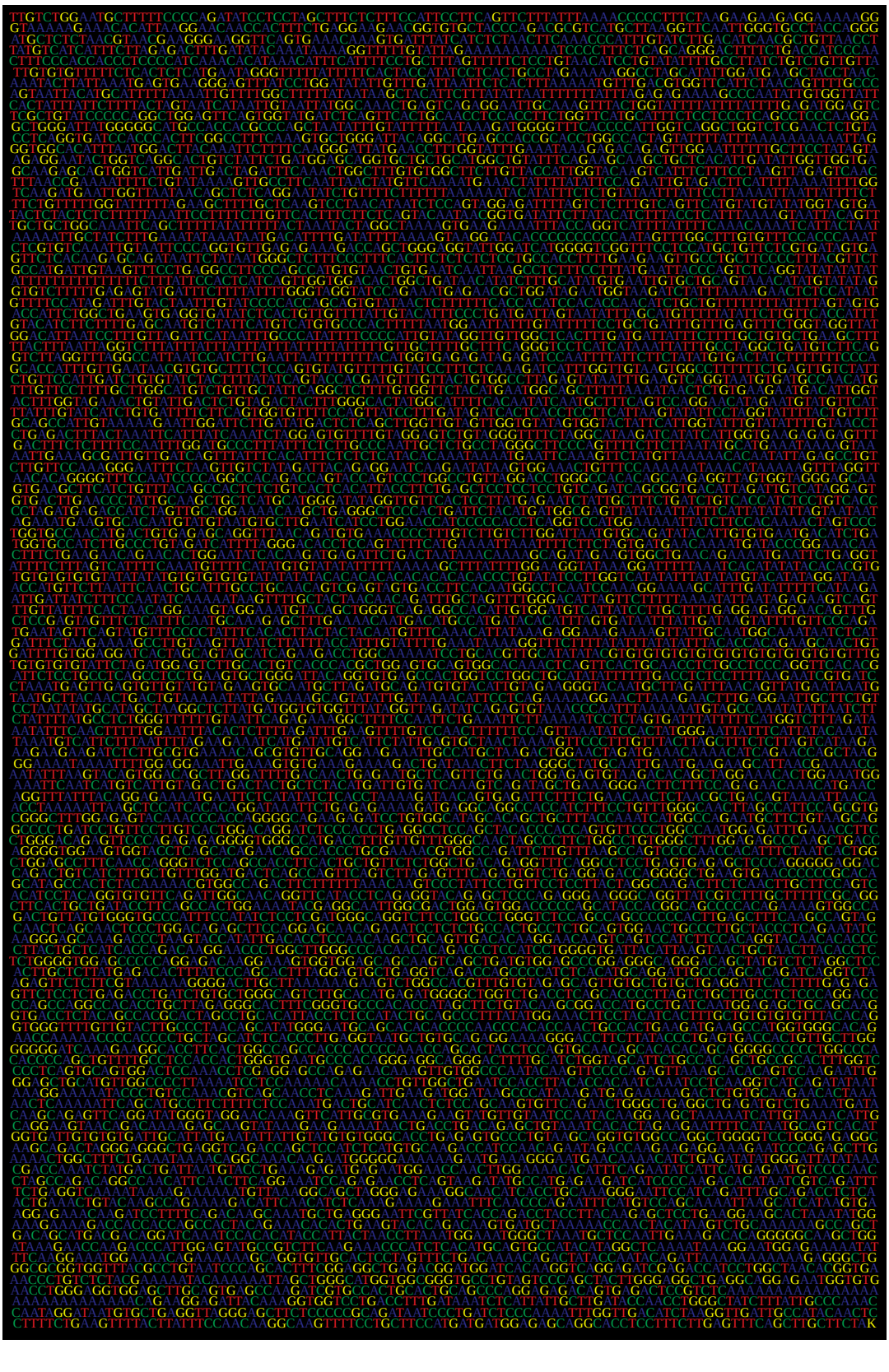
behaviour (Wedekind et al., 1996; Seksek *et al.*, 1997, Politz *et al.*, 1998; Politz, 1998; Wachsmuth *et al.*, 1998; Lukacs *et al.*, 2000; Wachsmuth, 2000). Many proteins are also influenced by their binding to the framework due to their function (Abney *et al.*, 1997; Phair & Misteli, 2000; Misteli *et al.*, 2000; Misteli, 2001).

To investigate the nuclear motion of particles, their diffusion was simulated by Brownian Dynamics in computer generated cell nuclei using the Multi-Loop-Subcompartment topology for the three-dimensional organization of the 30 nm chromatin fiber. The tracers interacted with the static fiber by an excluded volume potential.

The morphology of these cell nuclei shows big spaces in rendered or simulated electron microscopic images. These voids allow high accessibility to most nuclear locations and also to the interior of chromosome territories by tracers of corresponding size. Estimation of the nuclear volume fraction taken by the chromatin fiber exhibited that $<30\%$ of the nucleus might be occupied volume. This leaves more than 70% of the volume available for diffusion in typical cell nuclei. This is far from the percolation limit and suggests for small particles <10 nm only a moderate obstruction of their diffusion behaviour. This agrees with estimations of the mean mesh spacing of >29 to 82 nm for nuclei of 6 to 12 µm diameter.

Simulation of diffusion revealed that the particle displacement is proportional to the nuclear density and the tracer size. A particle of 10 nm diameter moves around 1 to 2 µm within 10 ms and therefore could rapidly cross a whole nucleus. Changes of the interaction potential and the diameter of the chromatin fiber were similar to small nuclear density changes. Different fiber topologies had no effect on the mean particle displacement, since the mean density and isotropic mesh spacing was unchanged. However, locally an influence, might be present. Big particles are trapped because the relation between their size and the available space exceeds the percolation limit. Changes of simulation parameters resulting in an increased mean density, also led to this percolation freeze-out. The anomaly parameter D_W characterizing the degree of obstruction ranged from 2.0 (obstacle free diffusion) to 3.0 or 4.0. These results are not only in agreement with the estimates but also with the experiments from above. Although the latter show a bit higher obstruction this could be explained by the simulated nuclei excluding e. g. mRNA. *In vivo* the chromatin organization itself is also dynamic, leading to less obstruction of bigger particles. Morphologically and dynamically the Inter-Chromosomal Domain (ICD) hypothesis suggesting dense inaccessible chromosome territories requiring a channel like network for molecular transport disagrees with these results.

Consequently, the simulations support the intuitive view of a nucleus as evolutionary optimized bioreactor: The genetic information containing DNA on the one hand is packaged fulfilling all requirements of pure storage and on the other hand guarantees by its nucleoplasmic suspension the easiest transport to every target site in the nucleus by diffusion. The latter is energy independent and guarantees by random mixing the most efficient reaction probability possible in fluidic systems. The detailed structural as well as its chemical modification of the genome could lead to a subtle regional regulation of processes.



6 Long-Range Correlations in DNA Sequences

6.1 Introduction

The sequential organization, i. e. the relations within sequences, and its connection to the three-dimensional organization of genomes is still a largely unresolved problem. Here, long-range power-law correlations were found by correlation analysis on almost the entire observable scale of 113 completely sequenced chromosomes of $0.5x10^6$ to $3.0x10^7$ bp from Archaea, Bacteria, *Arabidopsis thaliana*, *Saccharomyces cerevisae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens*. The local correlation coefficient shows close to random correlations on the scale of a few base pairs, a first maximum from 40 to 3400 bp (for *Arabidopsis thaliana* and *Drosophila melanogaster* divided in two submaxima), and often a region of one or more second maxima from 10^5 to $3x10^5$ bp. This multi-scaling behaviour was species specific. Within this multi-scaling behaviour an additional fine-structure is present and attributable to the codon usage in all except the human sequences. Here it is connected to nucleosomal binding. Computer generated random sequences assuming a block organization of genomes, the codon usage and nucleosomal binding explain all these results. Mutation by sequence reshuffling destroyed all correlations, thus their stability seems evolutionary tightly controlled and connected to the spatial genome organization on large scales. The correlation behaviour was used to construct trees, which were similar to the corresponding phylogenetic trees for β-Tubulin genes of Oomycetes and Eukarya genomes. For Archaea and Bacteria tree construction led to a new classification system with four major tree branches/ classes. In summary, these findings suggest a complex sequential organization of genomes closely connected to their three-dimensional organization.

Fig. 6.1 Correlations in a Piece of Sequence from *Homo sapiens* Chromosome XXI
An arbitrary piece with 12380 bp of the latter analysed human chromosome XXI supports already the qualitative impression that correlations are present on various length scales, although the prejudiced human ability of pattern recognition needs to be reminded. A GC rich region and a stretch of A's in the bottom part should especially be recognized (A: blue; C: green; G: yellow; T: red).

6.2 Correlation Analysis, Random Sequence Design and Tree Construction

6.2.1 Correlation Analysis of DNA Sequences and Genomes

The analysis of correlations in genetic sequences and especially long-range power-law correlations attempted here, is based on the concentration profile of single bases along the DNA sequence: The square root of the mean-square deviation between the concentration of bases c_l in a window of length l and the concentration $\bar{c}_L$ of bases in the entire DNA sequence with length L was calculated:

$$C(l) = \sqrt{\langle (c_l - \bar{c}_L)^2 \rangle_s}. \tag{6.1}$$

The average was taken over all $s = L - l + 1$ possible positions of the window within the whole sequence. Bases used were adenine (A), thymine (T), guanine (G), and cytosine (C) as well as their reduction to purines (A+G) and pyrimidines (T+C). Due to the complementarity of purines and pyrimidines the corresponding results are equal. Analysing the correlations as base/non-base or purine/pyrimidine, is equivalent to mapping the DNA sequence to the trajectory of a one-dimensional random walk. In the following only the results of purines/pyrimidines are considered.

For a fractal self-similar sequence like a random walk the concentration fluctuation function $C(l)$ shows power-law behaviour:

$$C(l) \sim l^{\delta} \text{ with } -1.0 \le \delta \le 0.0, \tag{6.2}$$

where -1.0 characterizes a negatively, -0.5 a randomly and 0.0 a positively correlated sequence. The power-law behaviour of $C(l)$ is connected to the power-law behaviour of the min- and maximum deviation function $F(l) \sim l^{a}$ by Peng *et al.* (1992), the common autocorrelation function $A(l) \sim l^{\gamma}$, and the power spectrum $S(f) \sim (1/f)^{\beta}$ via

$$\delta = \alpha - 1 = \frac{\beta - 1}{2} = \frac{-\gamma}{2} \tag{6.3}$$

(for details see: Borovik *et al.*, 1994; Stanley *et al.*, 1994). $C(l)$ are $F(l)$ related to the common autocorrelation function $A(l) \sim l^{\gamma}$ by double summation, e. g.

$$C^2(l) = \sum_{i=1}^{L} \sum_{j=1}^{L} A(j - i). \tag{6.4}$$

Using this definition, random fluctuations are substantially reduced compared to the fluctuations in $A(l)$ and the analysis leads to a more reliable characterization of the DNA sequence (Peng *et al.*, 1992).

Numerical calculation of $C(l)$ by using Equ. 6.1 in this sequence of operations

$$C(l) = \sqrt{\frac{1}{L - l + 1} \sum_{s=1}^{L-l} \left(\frac{1}{l} \sum_{k=1}^{l} n - \frac{1}{L} \sum_{k=1}^{L} N \right)^2} \tag{6.5}$$

by means of the probabilities for a base at a certain position $n = P(s+k)$, $N = P(k)$ and e. g. $P = 1$ for purines and $P = 0$ elsewhere, leads to extreme numerical instabilities (Fig. 6.2 A). These instabilities were avoided by expansion of Equ. 6.5 to

$$C(l) = \frac{1}{Ll}\sqrt{\frac{1}{L-l}\sum_{s=1}^{L-l}\left[\left(\sum_{k=1}^{l} Ln\right) - \left(\sum_{k=1}^{L} lN\right)\right]^2} \tag{6.6}$$

and by utilizing infinitely exact calculation tools provided by the GNU multiple precision package GMP. The latter is due to the start of deviations from the exact result (Fig. 6.2 A) and gets especially important for sequences longer than 10^5 base pairs. To save computer power, the program adjusted the precision automatically depending on the sequence length, guaranteeing a precision >8 digits.

To determine the local correlation coefficient $\delta(l)$ for the analysis of the general behaviour and fine-structural features of long-range correlations as a function of the window size l, the following asymmetric finite difference quotient of second order was applied to $\tilde{C}(\tilde{l}) = \log C(l) \sim \delta \log l$ with $\tilde{l} = \log l$:

$$\delta(l_i) = \frac{k}{h(h+k)}\tilde{C}(\tilde{l}_i + h) - \frac{k-h}{hk}\tilde{C}(\tilde{l}_i) - \frac{h}{k(h+k)}\tilde{C}(\tilde{l}_i - k) \tag{6.7}$$

with

$$k = \tilde{l}_i - \tilde{l}_{i-1} = \log l_i - \log l_{i-1}, \tag{6.8}$$

$$h = \tilde{l}_{i+1} - \tilde{l}_i = \log l_{i+1} - \log l_i, \tag{6.9}$$

$$\tilde{C}(\tilde{l}_i - k) = (\log C)(l_{i-i}) = C_{i-1}, \tag{6.10}$$

$$\tilde{C}(\tilde{l}_i) = \log C_i = \log C_i, \tag{6.11}$$

$$\tilde{C}(\tilde{l}_i + h) = \log C(l_{i+1}) = \log C_{i+1}. \tag{6.12}$$

To reduce the enormous computer power needed to calculate $C(l)$ and $\delta(l)$ for every possible l, every window from 1 to 10^4bp and 250 logarithmicaly distributed windows for every order of magnitude were chosen. Calculations were performed on PCs and an IBM SP2, using in total ~5000 h CPU time. On the latter the analyses were split into jobs of a few minutes computing single or few windows, thus being an extremely efficient filler for the unavoidable gaps in batch mode of big parallel machines. The analyses would also be predestined for grid computing, e. g. a screen saver application.

6.2.2 Design of Arteficial Random DNA Sequences/Genomes

To investigate the error behaviour and to determine the origin of various correlation properties, artificial sequences based on different assumptions about their composition were constructed:

Random sequences were constructed using a uniform distribution from a R250 random number generator which is based on 16 parallel copies of a linear shift register with a period of $2^{250} - 1$ (Kirkpatrick & Stoll, 1981). This is a far longer period in comparison to the usually used linear congruent generator and is as well computationally faster (Maier, 1991). Composition of base pairs was either uniform or biased by the human base pair distribution (A: 30%, C: 20%, G: 20%, T: 30%).

Random block sequences were assembled from blocks of random length with a randomly biased base pair composition. The block length B was chosen uniformly either from the interval [0, B] or [B-10%, B+10%]. The degree of bias in the base pair composition, defining the difference between blocks, was chosen independently for each block. The concentration of purines per block varied uniformly in [0.5-D, 0.5+D] with D being 0.050, 0.075, 0.100, 0.150, 0.200, 0.250, 0.300, 0.350, 0.400, 0.450, or 0.500.

Random codon sequences were composed by random arrangement of codons. Composition of codons using a uniform distribution equals the construction of totally random sequences (see above), thus random codon sequences were based on the biased codon usage tables provided by the Kazusa DNA Research Institute, Kisarazu, Japan (http://www.kazusa.jp). Used were the tables available on 13th October 2001.

Random gene sequences were designed as hybrids between totally unbiased random sequences and random codon sequences assuming the following: Codons with a distribution biased by codon usage tables were distributed randomly within connected blocks. These blocks correspond to genes and were placed in a totally unbiased random sequence. These genes were chosen to have a length of 999bp and were equally distributed within the random sequence. Therefore, variation of the fraction of genes in the sequence led to a change not only in the number of genes/blocks but also the length of the random sequence separating them.

Random nucleosome sequences were either based on a 230bp consensus sequence or two special sequence motives of nucleosomal binding sites. These were arranged in 2750bp long genes/blocks, as in the case of random gene sequences. For the consensus sequence the three nucleosomal binding sequences 602nvp_rev, 605nvp and 618nvp_rev found by SELEX experiments (Lowary & Widom, 1998) were compared: Base pairs present in at least two of the sequences were kept constant, while the other base pairs were chosen in an unbiased random manner: nnnGnnTGnT TCnnTnAnACC GAnnnnATCn nTTnnGnnAT GGACTACGnn GnGnCCnnGA GnnnnCnGGT GCCnnnnnCG CnCAATnnnG TnnAGACnnT CTAGnnCCGC TTAAACGCnn nTACnnCTnT CCCCCnCnTA nCGCCAAGGGG nnTnCnnnCT AGTCnCnAnn CACnTGTnnGn AnnCnTAAnC TGCAnnnnnT nACAnnGnCC TTGCC. Genes/blocks, consequently, are not a mere concatenation of the same consensus sequence, and thus reduce irrelevant correlations. The special sequence motives GCTCTAGAGC GCTCTAGAGC GCTCTAGAGC and CGTTTAAGCG TATCTAGAGC were suggested by Lowary & Widom (1998) to be the

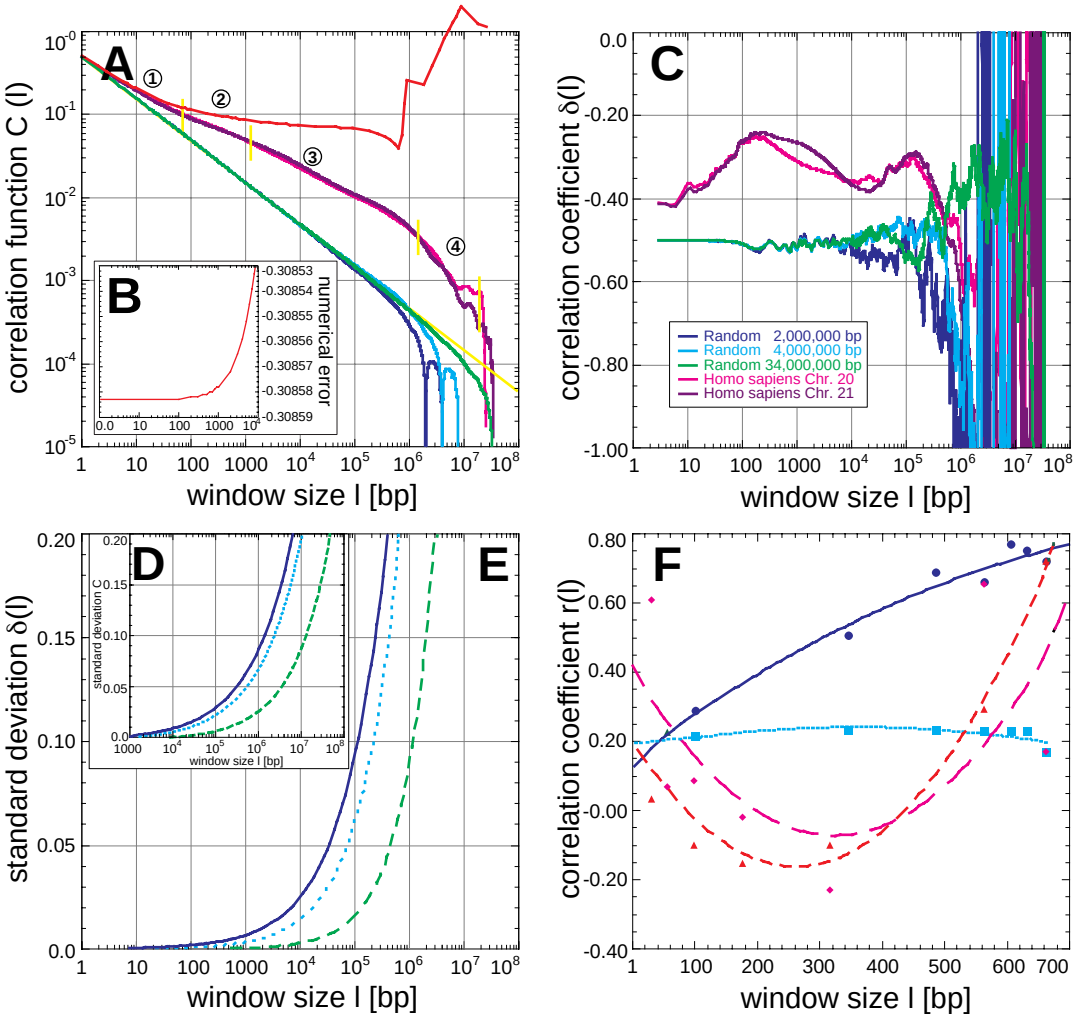


Fig. 6.2 Introduction to the Correlation Function $C(l)$ and Correlation Coefficient $\delta(l)$ (**A**) The correlation function $C(l)$ of random sequences shows power-law behaviour as expected for a fractal self-similar sequence (legend in **C**). The error made by not using exact numerics is shown by $C(l)$ of the *Homo sapiens* chromosome 21 (red line) and the absolute numerical error (**B**). The slope is the correlation coefficient δ whose value in the linear region is -0.5 (yellow line) indicating random correlations. The finite length of sequences generates a cut-off after which the power-law behaviour breaks down, thus concatenation of two sequences creates a double cut-off. Sequences of *Homo sapiens* exhibit not only a positively correlated power law behaviour due to a δ bigger than -0.5, but also four regions (numbers 1-4) with different degree of correlation. The detailed correlation behaviour is given by the local correlation coefficient $\delta(l)$ (**C**) which fluctuates around -0.5 for random sequences. The fluctuations became the bigger as the window size approaches the cut-off. *Homo sapiens* reveals a distinct positive correlated pattern with less fluctuations. To distinguish real from statistical correlations the standard deviation computed from 20 random sequences with similar base pair distribution as for *Homo sapiens* were calculated for $C(l)$ (**D**) and $\delta(l)$ (**E**). The standard deviation of $\delta(l)$ shifts only to higher window sizes depending on the sequence length but the behaviour is equal (colours same as in **C**). Comparison of correlation trees based on different window regions of the $C(l)$ (**F**; lower region: solid; upper region: dashed) and $\delta(l)$ (lower region: dotted; upper region: long dashed) to the phylogenetic tree reveals the distributed information content which can be fitted best with a second order polynomial. The information content for tree construction is the highest for using the entire sequence.

underlying motives for nucleosomal binding. Genes/blocks used a random mixture of both sequences with a distribution of 60%:40% according to their length.

6.2.3 Correlation Based Tree Construction and Classification of Genomes

To investigate the relationship between the different correlation behaviour of genetic sequences and genomes as well as for the comparison with classic phylogenetic trees, $C(l)$ and $\delta(l)$ were used to create correlation matrices r_{ij} by pairwise correlation between each single data set, applying the Pearson product moment correlation coefficients (PPMCC) s_{ij}:

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}^2 s_{jj}^2}} \text{ with } s_{ij} = \frac{1}{N-1} \sum_{k,\,l\,=\,1}^{N-1} (y_k - \bar{y}_k)(y_l - \bar{y}_l) \tag{6.13}$$

where k, l represent the values y of two data sets, with the average of the values $\bar{y}$, N the number of values in each data set and with $-1.0 \le r_{ij} \le 1.0$ (-1.0 is negatively, 0.0 is randomly and 1.0 is positively correlated). The r_{ij} calculated from $C(l)$ were always greater 0.0, whereas r_{ij} for $\delta(l)$ were also negative. Hence the former could be interpreted directly as measure of similarity S_{ij} which per definition ranges from 0.0 to 1.0 (Lefkovitch, 1993, p. 173), whereas the latter had to be transformed with

$$S_{ij} = \frac{(r_{ij} + 1)}{2}. \tag{6.14}$$

To construct trees for correlation based genome classification, it was necessary to calculate distance measures D_{ij} from the similarity measure S_{ij} in the form

$$D_{ij} = -\ln S_{ij} \tag{6.15}$$

according to Lefkovitch (1993, p. 173).
The D_{ij} matrices were directly imported into PAUP a standard platform for tree construction (Swofford, 2000), using the standard algorithms neighbour joining (NJ; Saitou & Nei, 1987) and unweighted pair group method by average (UPGMA; Swofford *et al.*, 1996, pp. 446) to find the relationships of the distance measures.

For the analysis of β-Tubulin genes from Oomycetes, PAUP was also used to calculate matrices of the genetic distances using their DNA sequence according to the algorithm of Kimura (Swofford *et al.*, 1996, p. 456). These were compared to the distance matrices D_{ij} based on the fractal analysis by determining the PPMCC of these total matrices.

Genetic distances were also used to construct trees by NJ for optical comparison. To obtain a phylogenetic tree for the Archaea, the gene for the 16S rRNA (Madigan *et al.*, 1997, pp. 621) was selected from the completely sequenced genomes. In each case, bootstrap analyses indicating the reliability of a tree branch were performed with 1000 replications (Swofford *et al.*, 1996, pp. 507).

6.3 Appearance of Long-Range Correlations

The concentration fluctuation function $C(l)$ (Equ. 6.1) and its exponent (Equ. 6.7), the local correlation coefficient $\delta(l)$, were calculated for the 6 longest available sequences of *Homo sapiens*, the 3 longest available sequences of the fruitfly *Drosophila melanogaster*, all 16 completely sequenced chromosomes of the yeast *Saccharomyces cerevisae*, the preliminary sequences of the 3 chromosomes of the yeast *Schizosaccharomyces pombe*, 3 completely sequenced chromosomes of the plant *Arabidopsis thaliana* (Tab. 6.1), as well as for the completely sequenced genomes of 15 Archaea (Tab. 6.2) and 64 sequences of 60 Bacteria four of which are bi-chromosomal (Tab. 6.2). The sequence length varied from $3x10^5$bp for the yeast chromosome III to $2.8x10^7$bp for a sequence piece of the human chromosome XIV. Longer stretches of undefined base pairs were not present within the sequences, except for negligibly unknown single bases (especially in the human sequences). Since most Archaea and Bacteria genomes are circular, the linear data base sequences were overlapfree concatenated to itself to cover the entire range of possible sequence correlations. Only *Agrobacterium tumefaciens* (AE007870) has a linear bacterial chromosome.

The exact calculation of $C(l)$, in principale being only a simple counting problem, required the use of a computationally stable algorithm (Equ. 6.6) and the use of the multiple precision package GMP for longest sequences. Otherwise fast growing numerical errors and function breakdowns for large l were unavoidable (shown for human chromosome XXI in Fig. 6.2A&B). Such effects have not been addressed for in any of the corresponding literature. The calculation of $\delta(l)$ was also exact, considering the chosen resolution of l to save computer power: from 1 to 10^4bp every l, and for $>10^4$bp 250 logarithmically distributed l were selected. Thus, for l $>10^4$bp local correlations $\delta(l)$ with high frequencies were smoothed out.

In all analysed sequences the concentration fluctuation function $C(l)$ shows power-law behaviour with varying slopes indicating a non-trivial degree of correlation (Fig. 6.2A). This is corroborated by the local correlation coefficient $\delta(l)$ with significantly varying values >-0.5, the characteristic value for random sequences (Fig. 6.2C). Thus, positive long-range correlations of non-random origin were found almost on the entire sequence scale, but certainly $<10^5$ to 10^6bp, in all of the analysed sequences (Fig. 6.3A&B, 6.4A-C, 6.5A-D, 6.6A-C, 6.7A-L, 6.14A-D).

Naturally, the finite lengths of the sequences generate a cut-off for the local concentration c_l approaching the mean concentration $\bar{c}_L$ for large l (Fig. 6.2A), resulting in break down of the power-law behaviour. The concatenation of sequences leads to a double cut-off. Since for cut-off-approaching l, the independent number of sequence windows $s = L - l$, over which the average is taken (Equ. 6.1), decreases rapidly, random deviations do not average out anymore. Thus, with growing l, fluctuations with increasing frequency and amplitude showed up in $C(l)$ and more apparently in $\delta(l)$.

Tab. 6.1 Attributes and Correlation Properties of Analysed Eukarya Genomes
To simulate the whole chromosome I of *Arabidopsis thaliana* the sequences of the top and bottom arm being separated by an unsequenced centromeric region were concatenated. Accession numbers of *Saccharomyces cerevisae* are annotated with the version in brackets. The sequences of *Schizosaccharomyces pombe* are the preliminary from 10.12.2001. The sequences of *Drosophilia melangoster* are the three largest available sequences with "Gold Standard Quality" downloaded on 10.12.2001 from http://www.fruitfly.org. The human sequence of chromosome XXI is the one by Hattori *et al.* (2000) with no apparent accession number. The sequence of chromosome XXII was downloaded

Genome	Accession Number	Category	Length [bp]	Correlation Properties				
				Start [N,R,P] [bp]	First Maximum [bp]	Transition [bp]	Second Maximum [F, R] [bp]	Fine Structure [C,F]
Arabidopsis thaliana Chr. I top+bottom	-	P	28890626	P	60/600	171	580	C
Arabidopsis thaliana Chr. I top	AE00517	P	14221746	P	60/550	160	550	C
Arabidopsis thaliana Chr. I bottom	AE005173	P	14668880	P	60/650	185	620	C
Arabidopsis thaliana Chr. II	AE002093	P	19646744	P	60/680	180	660	C
Arabidopsis thaliana Chr. IV	NC001268	P	17549956	P	60/650	160	680	C
Saccharomyces cerevisae Chr. I	NC001133(1)	Y	230203	P	500	-	-	C
Saccharomyces cerevisae Chr. II	NC001133(1)	Y	813139	P	435	-	-	C
Saccharomyces cerevisae Chr. III	NC001133(2)	Y	316613	P	450	-	-	C
Saccharomyces cerevisae Chr. IV	NC001133(2)	Y	1531929	P	410	-	-	C
Saccharomyces cerevisae Chr. V	NC001133(2)	Y	576870	P	640	-	-	C
Saccharomyces cerevisae Chr. VI	NC001133(1)	Y	270148	P	640	-	-	C
Saccharomyces cerevisae Chr. VII	NC001133(1)	Y	1090936	P	540	-	-	C
Saccharomyces cerevisae Chr. VIII	NC001133(2)	Y	562638	P	620	-	-	C
Saccharomyces cerevisae Chr. IX	NC001133(1)	Y	439885	P	460	-	-	C
Saccharomyces cerevisae Chr. X	NC001133(1)	Y	745440	P	460	-	-	C
Saccharomyces cerevisae Chr. XI	NC001133(1)	Y	666445	P	580	-	-	C
Saccharomyces cerevisae Chr. XII	NC001133(1)	Y	1078173	P	560	-	-	C

To distinguish real from these statistical correlations, random sequences with an initial length of 2, 4, and 34 Mbp (Tab. 6.1, 6.2, and 6.3) as well as their concatenation were created, using either equal or biased human base pair distributions. Both types of random sequences show the same behaviour, because $C(l)$ is based on the concentration deviation from the mean concentration. Only the onsets of fluctuations and cut-offs differ according to the length of the sequence. Therefore, the calculated standard deviation based on 20 such sequences for each length is could be fitted with the same but shifted exponential function (Fig. 6.2 D&E). The standard deviations for $C(l)$ and $\delta(l)$ remain small, e. g. $SD_{\delta(l)}$ is <0.1 up to ~1.3 and <0.05 up to ~1.6 orders of magnitude below the maximum sequence length. Consequently, positive long-range correlations are indeed present almost up to the entire scale of the analysed sequences, taking the standard deviation as function of the sequence length into account.

from The Institute for Genome Research (TIGR) website at http://www.tigr.org. Specie categories are plant (P), yeast (Y), insect (I) and primate (Pr). Properties of correlation are classified with N for negative (crossing the random regime with a value of 0.5 in bp), R for random and P for positive correlation coefficients for window sizes of a few base pairs. The transition to the second maximum could be a minimum (M). Second maxima are dividable into those with a fine structure not attributable to statistics (F) and those not clearly separable from fluctuations based on the cut-off length of sequences (R). The fine structure is categorized by codon usage (C) and nucleosomal binding (N).

Genome	Accession Number	Category	Length [bp]	Correlation Properties				
				Start [N,R,P] [bp]	First Maximum [bp]	Transition [M] [bp]	Second Maximum [F, R] [bp]	Fine Structure [C,F]
Saccharomyces cerevisae Chr. XIII	NC001133(1)	Y	924430	P	560	-	-	C
Saccharomyces cerevisae Chr. XIV	NC001133(1)	Y	784330	P	450	-	-	C
Saccharomyces cerevisae Chr. XV	NC001133(1)	Y	1091284	P	550	-	-	C
Saccharomyces cerevisae Chr. XVI	NC001133(1)	Y	875709	P	420	-	-	C
Schizosaccharomyces pombe Chr. I	V-011213	Y	5602103	P	900	1.2*^4	R1.0*^5	C
Schizosaccharomyces pombe Chr. II	V-011213	Y	4430733	P	850	1.4*^4	R1.0*^5	C
Schizosaccharomyces pombe Chr. III	V-011213	Y	2467649	P	610	2.0*^4	R1.0*^5	C
Drosophila melanogaster Chr. 2L	2L-1011210	I	22651956	P	40/3100	-	-	C
Drosophila melanogaster Chr. 2R	2R-2-011210	I	14631223	P	40/3800	-	-	C
Drosophila melanogaster Chr. 3R	3R-1-011210	I	28460979	P	40/3400	-	-	C
Homo sapiens sapiens Chr. XI	NT009151	Pr	19322668	P	200	1.0*^5	R3.5*^5	N
Homo sapiens sapiens Chr. XIV	NT026437	Pr	28334988	P	200	1.7*^4	F1.4*^5	N
Homo sapiens sapiens Chr. XV	NT010321	Pr	9197381	P	200	2.0*^4	1.0*^5	N
Homo sapiens sapiens Chr. XX	NT011362	Pr	24982240	P	200	1.2*^4	1.3*^5	N
Homo sapiens sapiens Chr. XXI	Nature	Pr	33820172	P	200	2.0*^4	1.3*^5	N
Homo sapiens sapiens Chr. XXII	TIGR WLC010213	Pr	33705278	P	200	1.2*^4	2.0*^4 1.9*^5	N

6.4 Multi-Scaling of Long-Range Correlations

Beyond the appearance of simple long-range correlations with a single slope covering the whole length scale, the concentration fluctuation function $C(l)$ conducts a far more complex behaviour: in all the analysed sequences the slopes vary considerably within different scaling regions (Fig. 6.2, 6.14). This is called multi-scaling. The local coefficient of correlation $\delta(l)$ is the more sensitive measure to investigate these general patterns in the limit of the chosen resolution of l. On scales with minor fluctuations and small standard deviation (6.3, Fig. 6.2C&D), $\delta(l)$ generally shows a global maximum between 40 and 3400 bp. This is sometimes followed by a region of one or several significant maxima around $6x10^4$ to $3x10^5$ bp (Fig. 6.3 A&B, 6.4 A-C, 6.5 A-D, 6.6 A-D, 6.7 A-L, 6.14 A-D). Both regions are con-

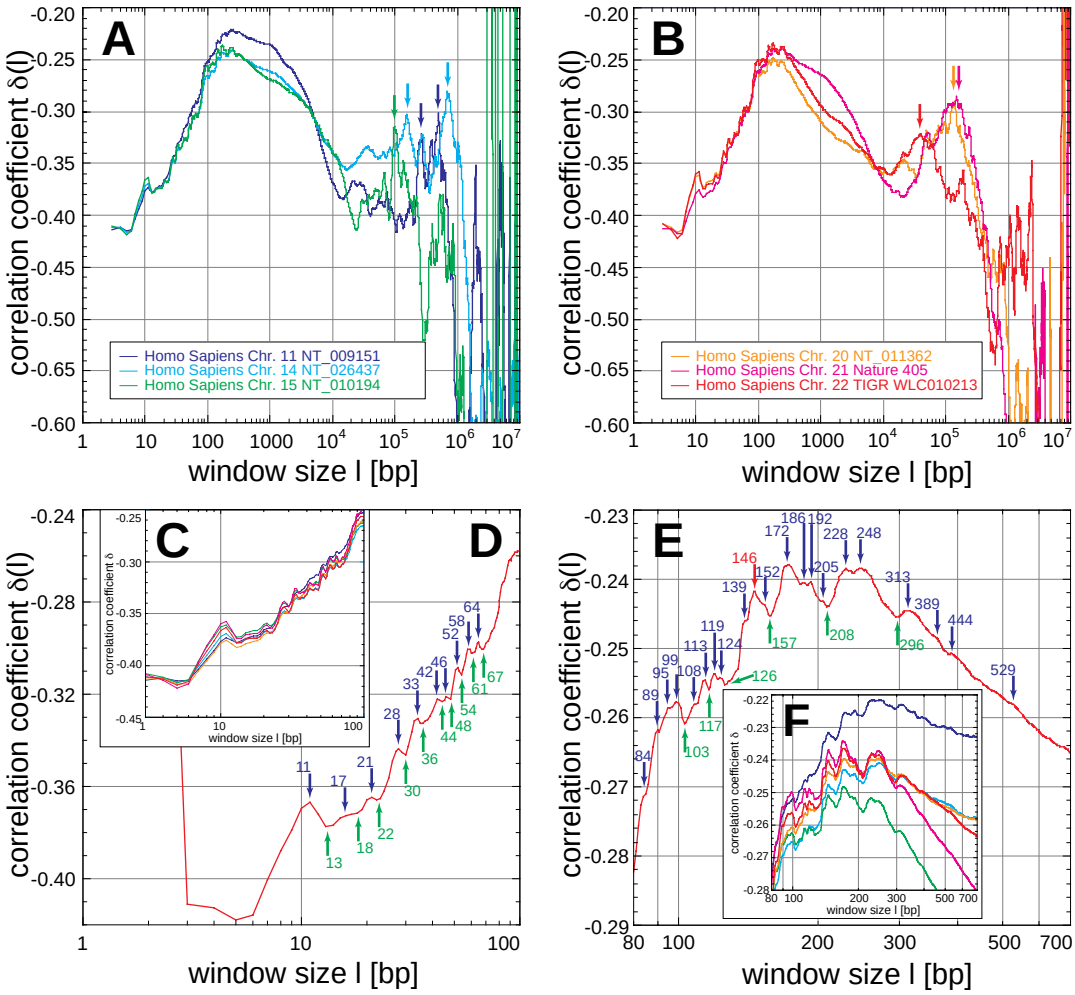


Fig. 6.3 Correlations in Chromosomes of *Homo sapiens* and their Fine Structural Features
The correlation coefficient $\delta(l)$ shows strong positive correlations for the chromosomes (**A**, **B**). In general, δ increases from a starting value until a plateaued maximum, before a decrease and a second statistical significant maximum for chromosomes 20, 21 and 22. Finally, δ decreases to values characteristic for random sequences and enters the fluctuation regime. Within this general behaviour a distinct fine structure is visible in all chromosomes (**C**, **F**) which survives averaging (**D**, **E**; see also Fig. 6.10). The very pronounced local maximum at 11 bp might be related to the double helical pitch, whereas the local minima and maxima are related to the nucleosome which is obvious for 146 bp and less obvious for 172 bp, 205 bp, 228 bp and 248 bp (**D**, **E**). The second maximum around 10^5 might be associated to chromatin loops and thus the three-dimensional organization of the human genome.

nected either directly or via a transition zone characterized by one or several minima. Consequently, in all the analysed sequences positive multi-scaling long-range correlations up to almost the entire length were found, beyond the already described simple power-law behaviours. The specific characteristics of these multi-scaling properties leading to different morphologic classes as well as their possible origin and interpretation are demonstrated in the following sections (Tab. 6.1, 6.2):

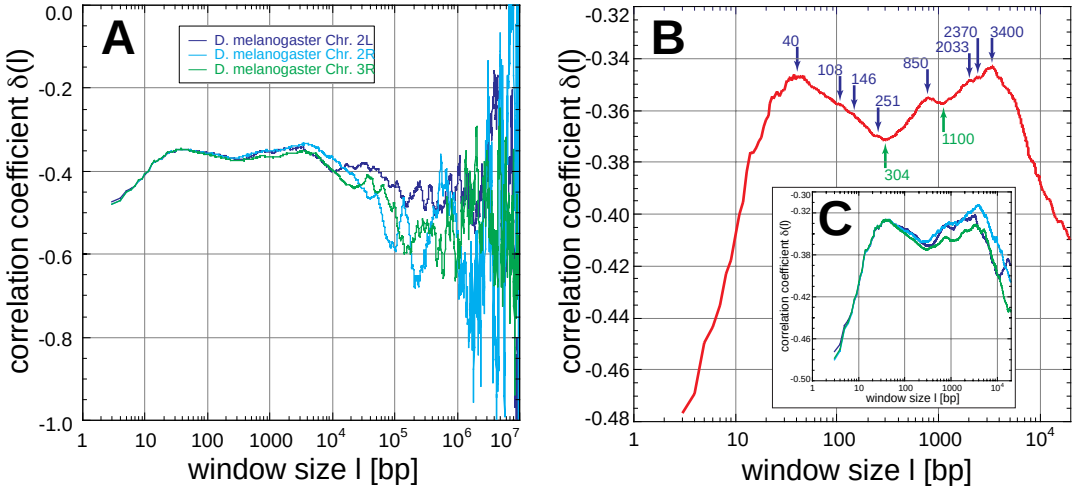


Fig. 6.4 Correlations in Chromosomes of *Drosophila melanogaster*
The analysed sequences of *Drosophila melangaster* show positive correlations analysing $\delta(l)$ (**A**, **B**, **C**). The averaged δ (**B**), has two main maxima (40bp, 3400bp) with several local maxima inbetween (108bp, 146bp, 251bp, 850bp, 2033bp, and 2311bp) and two major minima (302bp, 1100bp). These features appear in all chromosomes (**C**) and are similar to those of *Arabidopsis thaliana* (Fig. 6.6).

6.4.1 General Behaviour of the Multi-Scaling in Eukarya

Homo sapiens: The 6 largest available sequences from chromosomes XI, XIV, XV, XX, XXI, XXII with length of $9x10^6$ to $3.8x10^7$bp were analysed (Tab. 6.3). Sequences of chromosomes XX, XXI and XXII cover huge chromosomal parts with many ideogram bands, in contrast to those of chromosomes XI, XIV and XV. δ increases in all human sequences from an initial value around -0.42 to a maximum between -0.26 and-0.22, located at ~200bp (Fig. 6.3A&B). Despite the very similar ascent, the descent to the minimum between -0.40 and -0.35 at $2x10^4$ to $3x10^4$bp diverges: a slower, before a transition to a faster descent is characteristic for chromosome XI, XIV, XV and XXI, in contrast to an initial steeper descent for chromosome XX and XXII. The transition locates between 2000 and 4000bp in all 6 sequences. Thereafter, a clear second maximum was found for chromosome XXII at ~$4x10^4$bp and for chromosomes XX and XXI at $1.3x10^5$bp. The significance of these maxima is not only highlighted with respect to the standard deviation (Fig. 6.2E) but also in their steadyness compared to the spiked fluctuations of random sequences (Fig. 6.2C). Chromosomes XI, XIV and XV exhibit also significant peaks in the region between 10^5 and $5x10^5$bp, although their appearance is accompanied by a high degree of fluctuations. Whether these fluctuations or the substructure of the clear maxima of chromosomes XX, XXI and XXII feature real regularity, might remain unclear until the really complete (i. e. gap lacking) sequencing of all 24 human chromosomes.

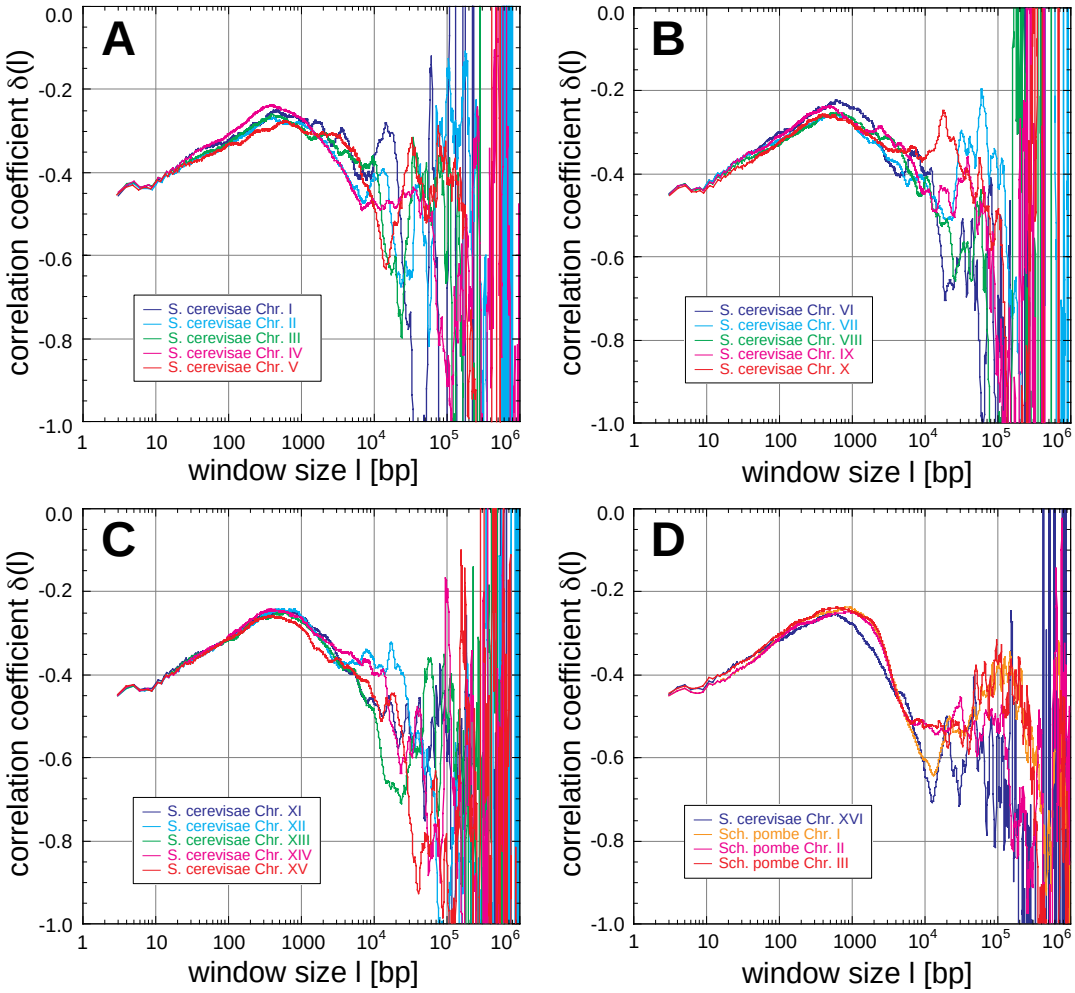


Fig. 6.5 Correlations in the *Saccharomyces cerevisae* and *Schizosaccharomyces pombe* Genomes Analysis of the chromosomes reveal correlations up to window sizes of 10^4 to 10^5bp for *Saccharomyces cerevisae* and up to $10^{5.5}$bp for *Schizosaccharomyces pombe*. The general behaviour of $\delta(l)$ is characterized by an increase of δ to maxima around 500 and 900bp, respectively. Thereafter, δ decreases until random correlation is reached for *Saccharomyces cerevisae*, or for *Schizosacchormyces pombe* minimum between 1.2 to $2.0\text{x}10^4$bp followed by a second maximum around 10^5bp. The zig-zag fine structure resulting from the codon usage is also present up to large window sizes.

Drosophila melanogaster: The 3 longest available sequences contain in contrast to the human, yeast, Archaea and Bacteria two flat maxima below 10^4bp like *Arabidopsis thaliana* with -0.347 and -0.345 at 40 and 3000bp, separated by a major minimum with -0.37 at ~304bp (Fig. 6.4A&C). Inbetween several smaller local maxima at 108bp, 146bp, 251bp, 850bp, 2033bp and 2311bp and one local minimum at 1100bp are present and survive averaging (Fig. 6.4B&C). Beyond scales of 3000bp, δ decreases to values characteristic for random correlations.

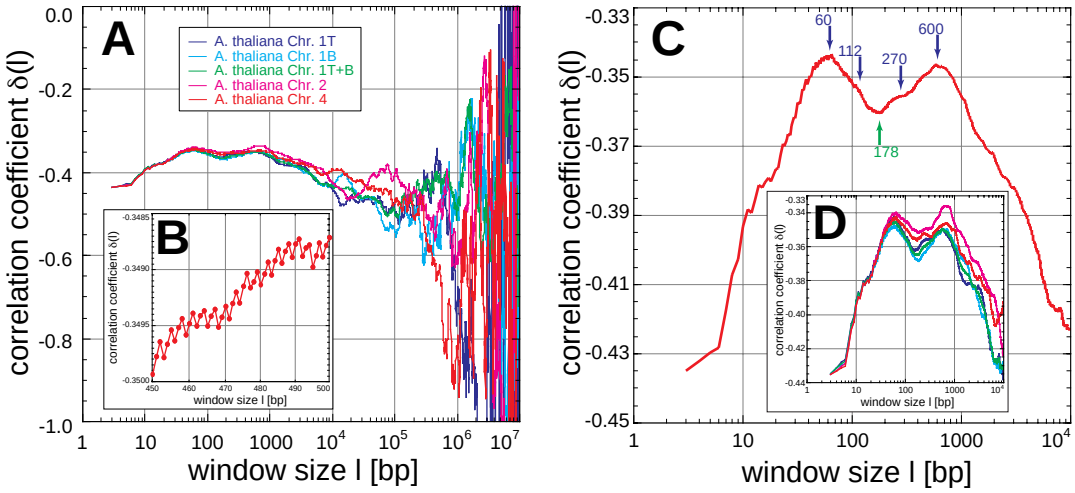


Fig. 6.6 Correlations in Chromosomes of *Arabidopsis thaliana*
The analysed chromosomes of Arabidopsis thaliana reveal positive correlations(**A**, **C**, **D**). The averaged δ (**C**) increases to two main maxima (60bp, 600bp), two small local maxima in between (112bp, 270bp) and one major minimum (178bp). These features appear in all chromosomes (**D**) and are similar to those of *Drosophila melanogaster* (Fig. 6.4). The zig-zag fine structure visible in the curves is due to correlations based on the codon usage (**B**) and is still present for large window sizes.

Saccharomyces cerevisae: The genome with its 16 chromosomes and length of $3x10^5$ to $1.5x10^6$bp (Tab. 6.1), is completely sequenced in contrast to all the other large genomes. δ increases from -0.45 linearly to a maximum around -0.25 between 400 and 650bp, and thereafter decreases until the region characteristic for random correlation and fluctuations are reached (Fig. 6.5A-D). The significance of the peaks and fluctuations on scales $>10^4$bp is unclear. Below 10^4bp, however, the behaviour of δ is astonishingly similar.

Schizosaccharomyces pombe: In the case of the 3 preliminarily sequenced chromosomes with length of $2.4x10^6$ to $5.6x10^6$bp (Tab. 6.1) δ increases from -0.45 linearly to a maximum around -0.23 between 600 and 900 bp, thereafter decreases to a minimum between $1.2x10^4$ and $2.0x10^4$bp before reaching a second significant maximum region around 10^5bp containing many fluctuations (Fig. 6.5D). Despite the much longer sequences the behaviour is remarkably similar to *Saccharomyces cerevisae* below the first maximum.

Arabidopsis thaliana: 3 of the five chromosomes were available, from which chromosome I was unfortunately split in pieces of the top and bottom arm due to the unsequenced centromer (Tab. 6.1). These two arms were concatenated to test changes in the analysis from single arms to a total chromosome. While the human, yeast, Archaea and Bacteria possesses one maximum below 10^4bp, *Arabidopsis thaliana* like *Drosophilia melangoster* demonstrates a different behaviour with two flat maxima with -0.342 and -0.345 at 60 and 600bp, separated by a major minimum

Tab. 6.2 Attributes and Correlation Properties of Analysed Archaea (A) and Bacteria (B) Genomes
Properties of correlation are classified with N for negative (crossing the random regime is given in bp), R for random and P for positive correlation coefficients for window sizes of a few base pairs. The transition between maxima is characterized by a stetic linear increase (L), a plateau with more

Archaea and Bacteria	Accession Number	Category	Length [bp]	Correlation Properties					
				Start [N,R,P] [bp]	First Maximum [bp]	Transition [L,PL] [M] [bp]	Second Maximum [K, F] [R,T] [bp]	Fine Structure [C,F]	Class [A] [A'] [A''] [B]
Aeropyrum pernix K1	BA000002	A	1669695	P	460	M7.0*^4	F6.0*^4	C	A
Archaeoglobus fulgidus	AE000782	A	2178400	P	420	M1.8*^4	R	C	A
Halobacterium sp. NRC-1	AE004437	A	2014239	N<21	900	L	F1.0*^5	C	B
Methanobacterium thermoautotrophicum delta-H	AE000666	A	1751377	P	1100	M3.7*^4	R2.0*^5	C	A''
Methanococcus jannaschii L77117	L77117	A	1664957	P	500	M2.0*^4	R2.3*^5	C	A
Methanopyrus kandleri AV19	AE094390	A	1694969	P	630	M1.0*^4	F1.0*^5	C	n. d.
Methanosarcina acetivorans C2A	AE010299	A	5751492	P	540	L	R3.0*^5	C	n. d.
Pyrobaculum aerophilum	AE009441	A	2222430	P	385	Ms	R	C	n. d.
Pyrococcus abyssi	AL096836	A	1765118	P	400	M2.0*^4	R	C	A
Pyrococcus furiosus DSM3638	AE009950	A	1908256	P	400	M1.5*^4	R3.0*^5	C	n. d.
Pyrococcus horikoshii	BA000001	A	1738505	P	400	M2.0*^4	R	C	A
Sulfolobus solfataricus	AL596259	A	2992245	P	370	M8.6*^4	R	C	A
Sulfolobus tokodaii	BA000023	A	2694756	P	385	M6.6*^4	R	C	n. d.
Thermoplasma acidophilum	AL139299	A	1564906	P	440	M3.0*^4	R	C	A
Thermoplasma volcanium	BA000011	A	1584854	P	440	M6.0*^4	R	C	A
Agrobacterium tumefaciens C58 circular chromosome	AE007869	B	2841581	N<4	800	M	T3.2*^5	C	n. d.
Agrobacterium tumefaciens C58 linear chromosome	AE007870	B	2074782	N<4	700	M	T3.2*^5	C	n. d.
Aquifex aeolicus	AE000657	B	1551335	P	370	M1,6*^4	-	C	A
Bacillus halodurans	BA000004	B	4202353	P	1000	T5.0*^3	K1.0*^5	C	B
Bacillus subtilis	AL009126	B	4214814	P	850	T3.5*^3	F1.0*^5	C	B
Borrelia burgdorferi	AE000783	B	910681	P	600	M5.0*^4	R1.5*^5	C	A
Brucella melitensis 16M Chr. I	AE008917	B	2117144	N	720	M1.0*^4	T2.5*^6	C	n. d.
Brucella melitensis 16M Chr. II	AE008918	B	1177787	N	830	M2.0*^4	T1.0*^5	C	n. d.
Buchnera sp. APS	BA000003	B	640681	P	850	M6.6*^4	R1.5*^5	C	A
Campylobacter jejuni	AL111168	B	1641481	P	660	M2.0*^4	T1.2*^5	C	A'
Caulobacter crescentus	AE005673	B	4016947	N<65	790	M7.0*^3	T4.0*^5	C	A'
Chlamydia muridarum	AE002160	B	1069393	P	650	L	K6.0*^4	C	B
Chlamydia pneumoniae CWL029	AE001363	B	1230230	P	500	PL	F1.0^5	C	B
Chlamydia pneumoniae AR39	AE002161	B	1229784	P	500	PL	F1.0^5	C	B
Chlamydia pneumoniae J138	BA000008	B	1228266	P	500	PL	F1.0^5	C	B
Chlamydia trachomatis	AE001273	B	1042519	P	650	L	F6.0*^4	C	B
Clostridium acetobutylicum ATCC824	AE001437	B	3940880	P	630	PL	K1.0*^5	C	B

less fast increase to the second maximum (P) or by a distinct minimum (M). Second maxima are dividable into those in the form close to a cap without much structure (K), those with a plateau including a fine structure not attributable to statistics (F), those not clearly separatable from fluctuations based

Archaea and Bacteria	Accession Number	Category	Length [bp]	Correlation Properties					
				Start [N,R,P] [bp]	First Maximum [bp]	Transition [L,PL] [M] [bp]	Second Maximum [K, F] [R,T] [bp]	Fine Structure [C,F]	Class [A] [A'] [A''] [B]
Clostridium perfringens 13	BA000016	B	3031430	P	615	PL	K.10*^5	C	n. d.
Corynebacterium glutamicum	AX114121	B	3309400	N<12	1000	PL	F1.4*^5	C	B
Deinococcus radiodurans Chr. I	AE000513	B	2648577	N<5	-	-	-	C	A'
Deinococcus radiodurans Chr. II	AE001825	B	412344	N<5	-	-	-	C	n. d.
Escherichia coli K12	U00096	B	4639221	N<5	860	M8*^3	F2.2*^4	C	B
Escherichia coli O157:H7-EDL933	AE005174	B	5468733	N<5	1000	PL	F2.2*^4	C	B
Escherichia coli O157:H7-RIMD0509952	BA000007	B	5498450	N<5	1000	PL	F2.2*^5	C	B
Fusobacterium nucleatum ATCC 25586	AL731704	B	2174500	P	1400	PL	R3.0*^5	C	n. d.
Haemophilus influenzae	L42023	B	1830023	P	720	M7.5^4	R1.8*^5	C	A
Helibacter pylori J99	AE001439	B	1643831	P	860	M1.8^4	T2.0*^5	C	A
Helibacter pylori 26695	AE000511	B	1667825	P	860	M6.0*^4	T2.0*^5	P	A
Lactococcus lactis IL1403	AE005176	B	2365589	P	950	L	K1.0*^5	C	B
Listeria innocua Clip 11262	AL592020	B	3011208	P	600	L	K1.0*^5	C	n. d.
Listeria monocytogenes EGD	AL591824	B	2944528	P	600	L	K1.0*^5	C	n. d.
Mesorhizobium loti MAFF303099	BA000012	B	7036071	N<10	660	M5.0^4	T8*^5	C	A'
Mycobacterium leprae TN	AL450380	B	3268203	N<25	700	L	F1.7*^5	C	B
Mycobacterium tuberculosis H37Rv	AL123456	B	4411529	N<22	1000	L	F3.6*^5	C	B
Mycobacterium tuberculosis CDC1551	AE000516	B	4403661	N<22	1000	L	F3.6*^5	C	B
Mycoplasma genitalium G37	L43967	B	580074	P	900	PL	6.0*^4	C	B
Mycoplasma pneumoniae M129	U00089	B	816394	P	1000	PL	F8*^4	C	B
Mycoplasma pulmonis UAB-CTIP	AL445566	B	963879	P	630	M8.3*^4	R1.3*^5 R2.6*^5	C	A
Neisseria meningitidis Sero Group A, Strain Z2491	AL157959	B	2184406	R	1100	PL	F3.5*^5	C	B
Neisseria meningitidis MC58	AE002098	B	2272351	R	1300	PL	F3.5*^5	C	B
Nostoc PCC7120	BA000019	B	6413771	R	690	M3.5*^4	F1.6*^5	C	n. d.
Pasteurella multocida PM70	AE004439	B	2257487	R	650	PL	F1.0*^5	C	B
Pseudomonas aeruginosa PA01	AE004091	B	6264403	N<11	950	L	F3.8*^5	C F(13)	B
Ralsteria solancearum GMI1000	AL646052	B	3716413	N	1900	PL	K2.0*^5	C	n. d.
Rickettsia conorii Malish 7	AE006914	B	1268755	P	690	ML	R2.5*^5	C	n. d.
Rickettsia prowazekii Madrid-E	AJ235269	B	1111523	P	690	M1.2*^5	R2.5*^5	C	A

on the cut-off length of sequences (R) and those being a mixture of F and R (T). The general fine structure is categorized into codon usage (C) or another distinct fine structure (F). General classification is based on the Archaea and Bacteria tree and consists of the classes A, A', A'' and B (Fig. 6.16).

Archaea and Bacteria	Accession Number	Category	Length [bp]	Correlation Properties					
				Start [N,R,P] [bp]	First Maximum [bp]	Transition [L,PL] [M] [bp]	Second Maximum [K, F] [R,T] [bp]	Fine Structure [C,F]	Class [A] [A'] [A''] [B]
Salmonella entericia servovar Typhi CT18	AL513382	B	4809037	N<4	950	L	T2.5*^5	C	n. d.
Salmonella typhimurium LT2	AE006468	B	4857432	N<4	950	L	T2.5*^5	C	n. d.
Sinorhizobium meliloti 1021	AL591688	B	2160837	P	750	M1.0*^4	F3.0*^5	C	B
Staphylococcus aureus Mu50	BA000017	B	2878134	R	1000	L	K8.6*^4	C	B
Staphylococcus aureus N315	BA000018	B	2813641	R	1000	L	K1.2*^5	C	B
Streptococcus pneumoniae	AE005672	B	2160837	P	860	L	F8.3*^5	C	B
Streptococcus pneumoniae R36	AE007317	B	2038615	P	860	L	F8.3*^5	C	n. d.
Streptococcus pyogenes SF370	AE004092	B	1852441	P	860	L	F8.3*^4	C	B
Streptomyces coelicolor A3(2)	AL644882	B	8667507	N<5	720	M2.0*^4	F1.0*^5	C	n. d.
Synechocystis sp. PCC6803	AB001339	B	3573470	P	500	M8.0*^4	R1.0*^5	C	A
Thermotoga maritima	AE000512	B	1860725	P	690	M6.0*^5	R1.8*^5	C	A
Treponema pallidum	2275888	B	1137944	P	1000	L	K	C	B
Ureaplasma urealyticum	1503438	B	751719	R	900	M2.5*^4	F1.0*^5	C	A
Vibrio cholerae Chr. I	AE003852	B	2961116	N<5	630	L	F2.0*^5	C	B
Vibrio cholerae Chr. II	AE003853	B	1072311	N<5	630	L	F1.0*^5	C	B
Wigglesworthia brevipalpis	BA000021	B	697721	P	562	M2.0*^4	F	C	n. d.
Xylella fastidiosa	AE003849	B	2679306	N<10	2200	M4.0*^4	R	C	A''
Yersinia pestis CO92	AL590842	B	4653728	N<5	900	L	F1.7*^5	C	n. d.

with -0.36 at ~178bp (Fig. 6.6A&D). Inbetween, two smaller local maxima are present at 112 and 270bp. Averaging all sequences leaves these structures unchanged (Fig. 6.6C&D). Beyond scales of 600bp, δ decreases to values characteristic for random correlations. The growing fluctuations are statistical insignificant, despite the length of the sequences between $1.5x10^7$ and $2.8x10^7$bp. Concatenation of the top and bottom arm led to no changes below 10^4bp, but structures present in the above discussed separated arms were averaged out.

6.4.2 General Behaviour of the Multi-Scaling in Archaea and Bacteria

Archaea and Bacteria (Tab. 6.2) revealed a more diverse behaviour than expected from the similarity between the chromosomes within the single Eukarya species. Nevertheless, the appearance of distinct general behaviours allows qualitative and later quantitative classification by tree construction (6.6) into four major classes referred to as A, A', A'' and B with distinct general behaviour (for a quantitative

classification see 6.6). In class A consisting of some Bacteria (e. g. Aquifex aeolicus) and most of the Archaea (e. g. *Eropyrum pernix* and except *Halobacterium sp. NRC1*), δ increases up to a general maximum around -0.14 at ~550bp and thereafter decreases with growing fluctuations (Fig. 6.7A&B). Separate analyses of the Archaea and Bacteria within the class A reveal a shift of the maximum position with -0.15 at ~450bp and -0.13 at 650bp, rescpectively. The region of second local maxima at around 10^5bp within the present fluctuations seems due to the limited number of available sequences statistical insignificant, although the second maxima are gaining more significance between $5x10^4$ and 10^5bp for the Bacteria. Class A', containing e. g. *Campylobacter jejuni*, possesses a lower first maximum around -0.27 at ~850bp, followed by a minimum of around -0.35 between 5000 and $2.5x10^4$bp. Succeeding with a linear increase a statistical significant plateaued maximum between $6x10^4$ and $3x10^5$bp, in which small fluctuations are present, is reached. Finally, the plateau decreases sharply without much fluctuation. The latter quantitatively found class A'' consisting only of *Methanobacterium thermoautotrophicum delta-H* and *Xylella fastidiosa* seems to be a mixture of class A and A'. Yet another behaviour is shown by the biggest class B encountering 30 Bacteria (e. g. *Bacillus halodurans* or *Clostridium acetobutylicum*): Here the first maximum is only hinted after the usual increase and reaches out into plateaued saddle points at ~2000bp. Therafter, δ rises to presumably a second maximum at ~10^5bp with extreme degree of correlation sometimes even above -0.1. For window sizes >10^5bp δ decreases sharply with hardly any fluctuation supporting again the high correlation degree suppressing the fluctuations, common else and for random sequences.

In summary, the general correlation behaviour of Archaea and Bacteria is characterized by a first maximum below 10^3bp with decreased height and increased position, the more a second maximum appears. The transition between these maxima exhibits a minimum or a saddle point, depending on the growing presence of the second maximum. The sometimes found extreme correlation degree is unlike that found in any Eukarya. It needs to be noted that sequences from the same Archaea or Bacteria but from different strains, behave very similar, and thus suggest evolutive constancy of correlations in connection with their phylogenetic relationship as for the correlation behaviour within the single species of Eukarya.

6.4.3 Origin and Interpretation of Multi-Scaling

The distinct morphologic classes found qualitatively within the general correlation behaviour (6.4.2), imply a higher degree of sequential organization than proposed by a mere multi-scaling behaviour. To determine quantitatively a possible origin of these multi-scaling behaviours, random sequences were designed assuming a block organization of genomes (6.2.2). For Eukarya, and especially for *Homo sapiens*, such a block organization is already proposed by e. g. the formation of stainable and in their AT/GC and gene content differing ideogram bands of metaphase chromo-

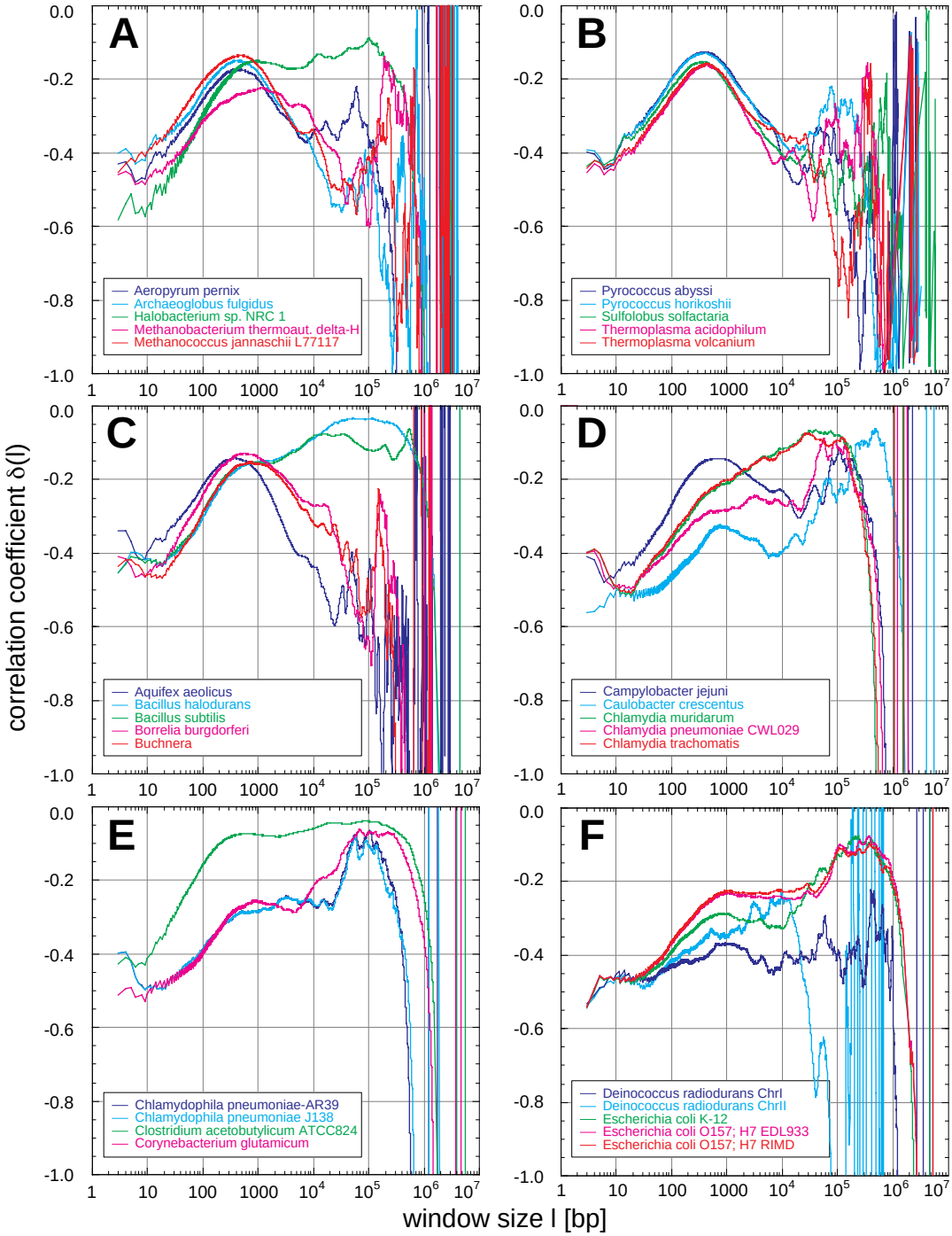
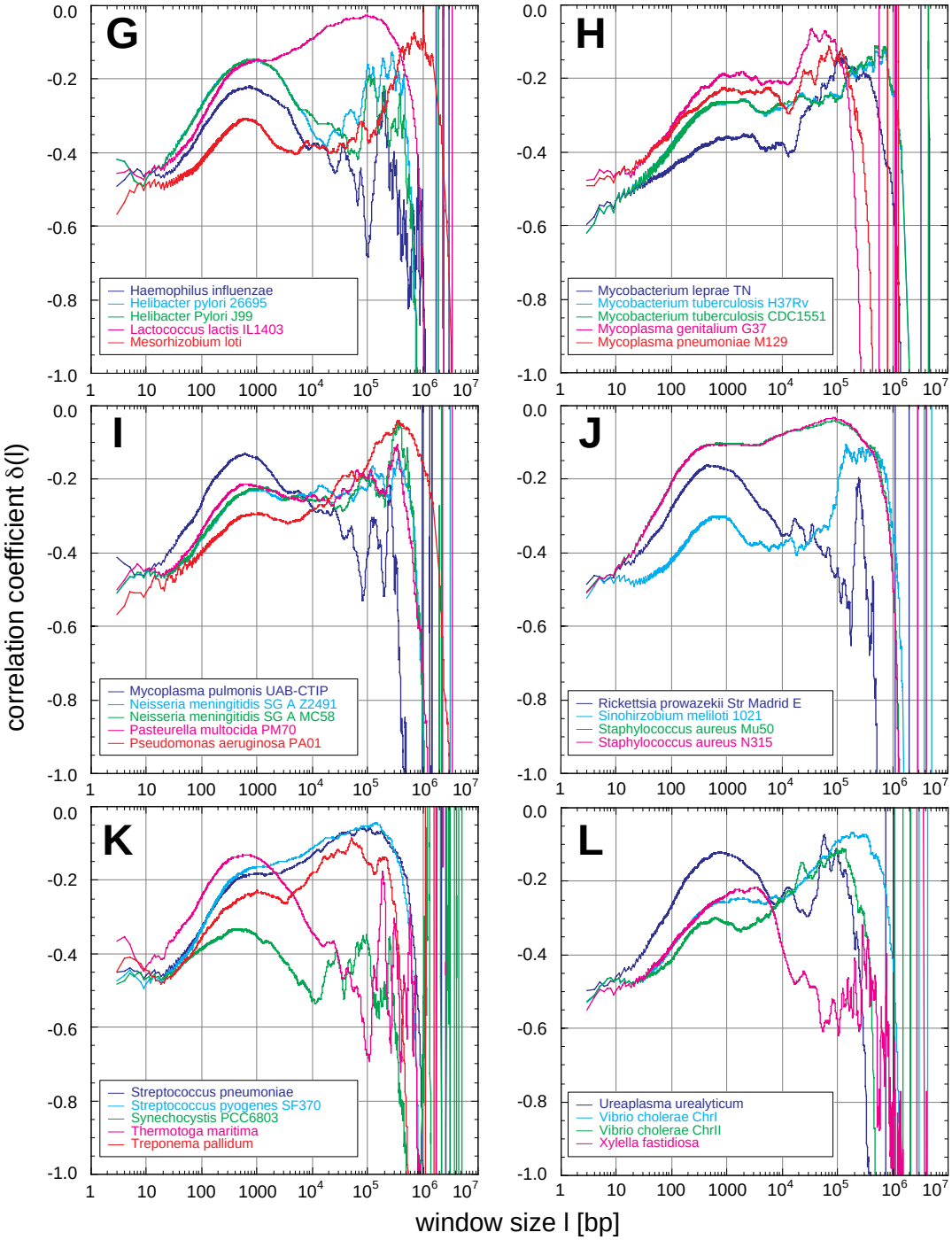


Fig. 6.7 Correlations in Archaea and Bacteria Genomes and Their Classification
The analysis of the correlation coefficient $\delta(l)$ of Archaea (**A**, **B**) and Bacteria (**C**-**L**) reveals behaviours separable into four major classes referred to as A, A', A'' and B as represented by tree construction (Fig. 6.15) and averaged (Fig. 6.16). In general Archaea and Bacteria are characterized by a first maximum below 10^3bp with decreased height and increased position, the more a second maximum is appearing. The transition exhibits a minimum or a saddle point also connected to the growing presence of the second maximum. The often found extreme degree of correlation is unlike that found in



any of the Eukarya. Prime example for the Archaea is *Archeoglobus fulgidus*, for class A *Aquifex aeolicus* and for class A' *Campylobacter jejuni*. Class A'' is a mixture of class A and A' consisting only of *Methanobacterium thermoautotrophicum delta-H* and *Xylella fastidiosa*. The biggest class B of 30 Bacteria like e. g. *Bacillus halodurans* or *Clostridium acetobutylicum* and is charactericed by the extreme degree of correlation and sharp descent without fluctuations. Sequences from the same Archaea or Bacteria but different strains show almost identical behaviour.

somes (Francke, 1994; see also Chapter 1). A block structure could also be suggested by the three-dimensional organization of genomes (Chapters 1-4 and 7).

The random block sequences with a total length of 10 Mbp, were composed from blocks with a random length B chosen either from [0, B] or [B-10%, B+10%]. This avoids artificial correlations from a fixed block length. While [0, B] approximates a primitive fractal block pattern with a certain degree of self similarity due to the broad distributed block length, [B-10%, B+10%] modells a softened periodicity. The difference between blocks was created by changing the uniform purin/pyrimidin compositions to concentrations chosen uniformly from [0.5-D, 0.5+D] with D varying from 0.00 to 0.50. The overall composition remained therefore unchanged.

All created block sequences own one global maximum, whose position, width and descent is proportional to the block length and whose ascent as well as and initial values are also proportional to the concentration deviation D (Fig. 6.8A&B). In contrast, the height of the maximum is inversely proportional to the deviation D. The behaviours are in agreement with the measuring process leading to the concentration fluctuation function $C(l)$ and the local coefficient of correlation $\delta(l)$. Both kinds of the used block length distributions yielded similar results with somewhat smaller values for the block length distribution from [0, B] (Fig. 6.8A). Remarkably only after the descent, fluctuations common for random sequences with uniform or biased base pair composition set in (Fig. 6.2C). These fluctuations are suppressed by correlations induced by the blocks. This suppression is proportional to the block length. In detail, the maximum height changes from -0.42 to nearly -0.005 and its position shifts from 35 to 1.5×10^4 bp for block length from 50 to 10^6 bp and a deviation D of 0.100 (Fig. 6.8A). For changes of D from 0.050 to 0.500 the height of the maximum changes from -0.27 to -0.03 and from -0.04 to -0.005 for a block length of 10^3 and 10^5 bp. Thus, blocks of large length and/or large concentration deviations create correlations of extremely high degree. The correlation degree for $\delta(l = 3)$ as function of the deviation D, follows $\delta(l = 3, D) = -0.5 + 0.113D + 0.855D^2$, a quadratic fit with $R = 0.99$, in contrast to the linear dependence found in the simulation of the fine structural pattern of the codon usage (6.5.1).

To understand the obvious evolutive persistence of the multi-scaling long-range behaviour, simple random rearrangements of blocks with the same properties as those used to create the random block sequences were applied to these sequences: Already after 10^5 rearrangements the multi-scaling properties disappeared completely. Determination of a real time scale was not an aim of this investigation. Applying the more complex and yet partly known possibilities of mutation and rearrangement processes, could lead to the determination of a precise time scaling for the loss of correlations. Thus, evolutive persistence is only guaranteed by defined and not totally random rearrangements in real genomes. At least for correlations on scales $>10^3$ bp this seems to require the involvement of the three-dimensional organization of genomes and vice versa. Here the nucleosomal and the local 30 nm chromatin fiber conformation might also play an important role. Consequently, the general sequential and the three-dimensional organization seem to be closely inter-

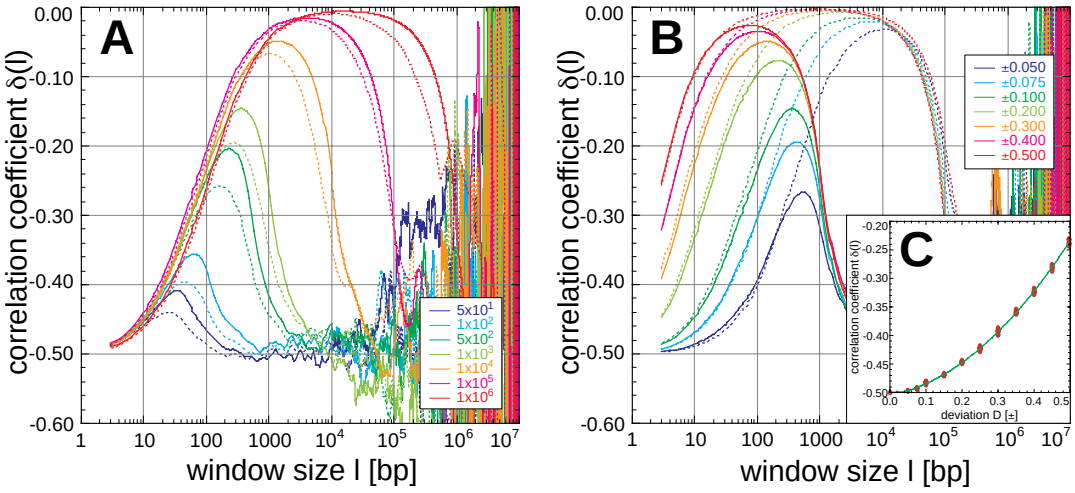


Fig. 6.8 Simulation of the Block Structure of Genomes
Simulation of random sequences using blocks of random length B either from the interval $B \pm 10\%$ or 0 to B, and with a deviation from the uniform purin/pyrimidine concentration, leads to a global maximum in the correlation coefficient (**A**,**B**). Its position, height and descent are proportional to the block length (**A**; $B \pm 10\%$: solid line, 0-B: dotted line, **B**: see legend there, for a deviation of 0.100) and the ascent to the maximum and its height are proportional, whereas its position is inversely proportional to the concentration deviation (**B**; $B \pm 10\%$ with $B = 10^3$, solid line, $B = 10^5$: dotted line, deviation see legend there). Notably, the descent is remarkably smooth although fluctuations increase exponentially as function of window size l (Fig. 6.2). The degree of correlation follows a quadratic dependence $\delta(l = 3, D) = -0.5 + 0.113D + 0.855D^2$ with $R = 0.99$ (**C**), in contrast to the linear dependence found for simulations of the codon usage.

woven. An intuitive understanding might be achieved from card games and how cheats avoid destroying a favourable card sequence during the mixing process.

In summary, the general morphology of the multi-scaling correlation behaviour (6.4.1) is at least partly explained in all analysed sequences by a relatively simple block organization with evolutive persistence (6.4.2). In reality, of course, the mixture of block lengths and deviations is more complex than assumed here. Especially integration of blocks within blocks could fine tune the general behaviour. Nevertheless, the detailed description of the general morphology in 6.4.2 can already be quantified reasonably well: In the case of *Homo sapiens* the first maximum could be due to block length of ~500 bp and concentration deviations of 0.050 to 0.075. The second maximum present in the sequences of chromosomes 20, 21, and 22, cannot be explained by simple block structure on the order of 10^5 bp, although its smooth and fluctuationless behaviour is similar to those of large blocks. A more pronounced periodicity, consisting of evenly spaced blocks with a deviation in base pair composition and a length of around 10^5 bp, could be the origin for these second maxima. Such periodicities were found in the simulation of the codon usage and nucleosomal binding sites (see below, 6.5). The behaviour of chromosomes from *Saccharomyces cerevisae* is best described by a block length of 5000 bp and deviations of 0.05. Sequences of *Arabidopsis thaliana* can be a mixture of two block sizes of 50 to

100bp and 5000bp, with deviations below 0.05. Concerning Archaea and Bacteria, the first maximum in the morphologic classes of Archaea, A, A' and A'' is best described by 5000 to 10^4bp blocks with deviations from 0.30 to 0.075. The second maxima appearing to increase from groups A' to A'' can be explained by increasing presence of blocks with large length or by more pronounced periodicities as proposed for *Homo sapiens*. In class B this kind of interpretation gets clearer: Here the behaviour can be explained by merging block lengths of 5000bp and 10^5 to 10^6bp with deviations in the base pair concentration >0.075.

6.5 Fine Structure of Long-Range Correlations

Within the multi-scaling long-range correlations further fine structures were found which are attributable to the codon usage and nucleosome associated sequences after comparison to arteficially designed random sequences (6.2.2):

6.5.1 Codon Usage Associated Fine Structure

A fine structure with a periodicity of 3bp up to window lengths of several hundred (Fig. 6.6B) or even a few thousand base pairs is present in all but the human sequences (Fig. 6.3A&B, 6.9A). In the bacterium *Pseudomonas aeruginosa* PA01 this 3bp periodicity is dominated by another periodicity of 12bp (Fig. 6.7I, 6.9B). The sequences of *Homo sapiens* show yet another finestructure (6.5.2). To relate this fine structure to the codon usage and to distinguish it from the fine structure found in the human (Fig. 6.3C-F) and *Pseudomonas aeruginos*a PA01 sequences, 10Mbp long random sequences were generated: These sequences consisted completely of codons with a distribution of codon usage tables (6.2.2). As expected, uniformly distributed codons, the simplest codon usage table, totally lack a fine structure (Fig. 6.9C). However, a uniform distribution of amino acids using the human codon usage tables, already introduces enough imbalance to create the requested fine structure. Use of the real human codon distribution only increased the behaviour. Random codon sequences based on the corresponding codon usage table displayed the fine structure for all analysed sequences as shown e. g. for *Drosophila melanogaster*, *Saccharomyces cerevisae*, *Schizosaccharomyces pombe*, *Arabidopsis thaliana*, *Chlamydia muridarum*, *Mycobacterium tuberculosis* and *Pseudomonas aeruginosa* PA01. Thus, neither the fine structure present in *Homo sapiens* nor the 12bp periodicity in *Pseudomonas aeruginosa* PA01 are based on the codon usage. The latter possibly is due to an uncommon but distinct succession of codons. The simulations also correctly reproduce the correlation degree for $\delta(l = 3)$ and whether this starting value is > or < -0.5. The fine structure also approximates rapidly -0.5, thereafter fluctuating around it. Thus, no increase of δ is created as in the real sequences, thus finally attributed mainly to the block structure of genomes (6.4.3).

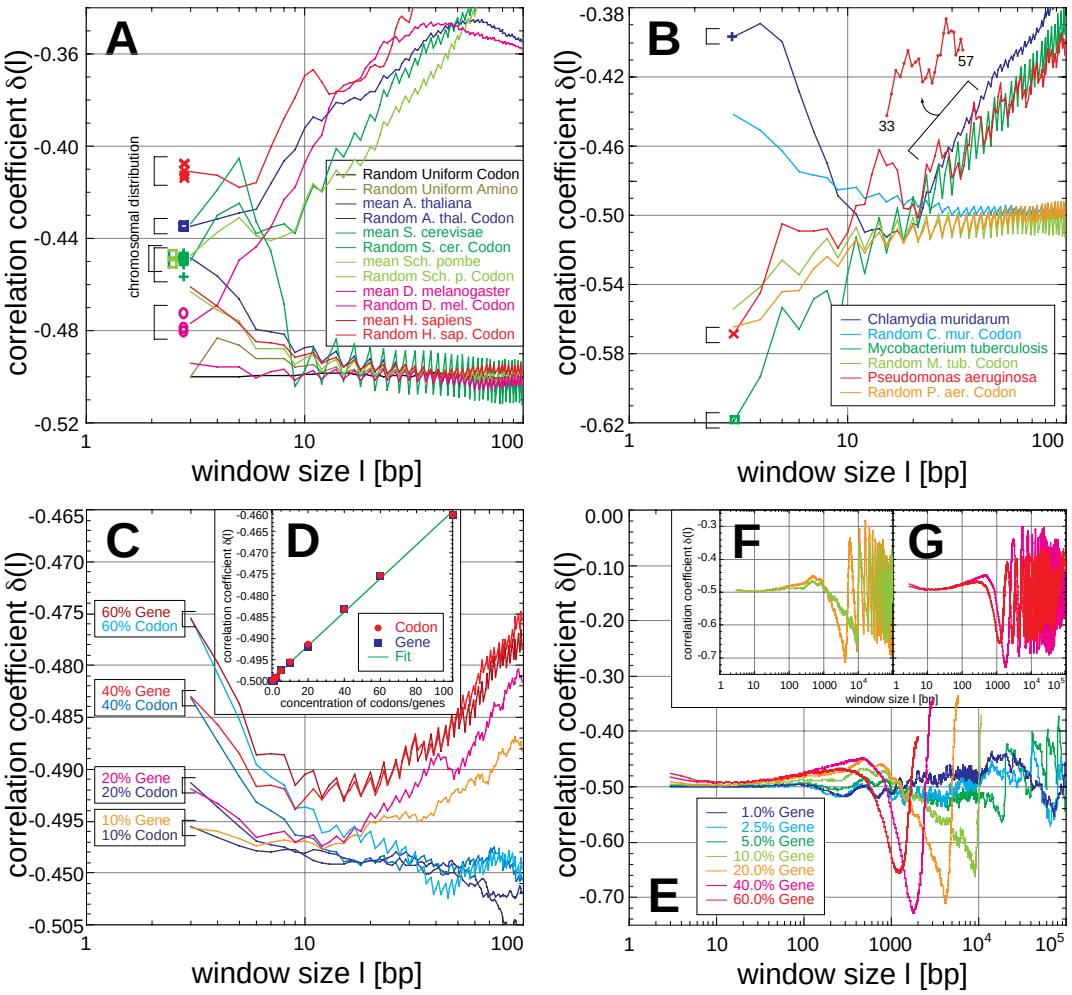


Fig. 6.9 Appearance and Simulation of the Codon Usage
In all but the human sequences a fine structure with a periodicity of three up to window length of several hundred base pairs (Fig. 6.6) is present which is associated to the codon usage (**A**, **B**). Already a uniform distribution of the 20 amino acids in artificial random sequences causes this feature. Specie specific codon usage is responsible for the starting behaviour $\delta(3) < -0.5$ or $\delta(3) > -0.5$. *Pseudomonas aeruginosa* PA01 has an additional periodicity of 12 bp dominating the codon usage pattern and is unexplainable by a simple codon usage (**B**). The appearance, visibility of the codon usage as well as the degree of correlation at $\delta(3)$ is proportional to the concentration $c_{codon/gene}$ of human distributed codons within a random sequence and is more apparent for codons organized in genes/blocks (**C**, see 100% value in **A**). The degree of correlation follows a linear dependence with $\delta(l = 3, c_{codon,gene}) = -0.5 + 0.039 c_{codon,gene}$ and $R = 0.99$ (**D**). Organization of codons in genes/blocks led to a maximum of $\delta(l)$ and oscillations due to the gene/block length and separation (**C**, **E**, **F**, **G**; see also Fig 6.8).

To further investigate the codon amount needed within the sequence to produce the fine structure, random sequences were set up with a concentration $c_{codon/gene}$ of codons. The codons chosen from a variety of codon usage tables were either randomly mixed under a random sequence (random codon sequence) or organized in

blocks of 333 codons or 999bp which were distributed equally in the sequence (random gene sequence; 6.2.2). Whereas the former simulate mutated, distorted or free for deletion genes, the latter come close to functional genes. The appearance of the codon fine structure is proportional to the codon concentration and sets in for concentrations of ~10% for gene and >50% for codon sequences (Fig. 6.9C). Thus, the earlier inset of the fine structure for gene sequences is caused by the uninterupted succession of codons within a gene. This proximity enhancement is not present in random codon sequences. The degree of correlation for the human codon distribution at $\delta(l = 3)$ follows $\delta(l = 3, c_{codon,gene}) = -0.5 + 0.046c_{codon,gene}$ with $\delta(l = 3, c_{codon,gene}) = -0.5 + 0.046c_{codon,gene}$ and $R = 0.99$ for random codon as well as gene sequences (Fig. 6.9D). For *Drosophila melanogaster*, *Saccharomyces cerevisae*, *Schizosaccharomyces pombe*, *Arabidopsis thaliana*, *Chlamydia muridarum*, *Mycobacterium tuberculosis* and *Pseudomonas aeruginosa* PA01, similar linear laws were found with slopes of 0.047, 0.043, 0.043, 0.045, 0.044, -0.055 and -0.056, respectively. Consequently, the dependence is based on the degree of correlation within the codon.

Beyond the fine structure, the random gene, in obvious contrast to the codon sequences, demonstrates also a general multi-scaling behaviour as found in the design of random block sequences: a first maximum before 10^3bp is followed by periodicities proportional to the different separations between genes for different $c_{codon/gene}$ (Fig. 6.9D-F). The height and position of the first maximum is more pronounced the greater the deviations between the genes and the rest of the sequence are, and thus is maximal for $c_{codon/gene} = 50\%$ with a δ of -0.44 at 480bp. Consequently, the multi-scaling created by genes has a minor influence in comparison to the block organization discussed above (6.4.3). Nevertheless, small sequence regions with a strong deviating base pair concentration in connection with a periodic spacing could explain the second maxima found around 10^5bp in the human sequences and uninterpretable with the simple block approach (6.4.3). A straightforward calculation based on the total length of the haploid human genome of ~3.5×10^9bp and the ~35,000 genes so far found, also results in a mean gene spacing of 10^5bp. Thus, the second maxima might originate in the gene spacing or gene density within these sequences.

6.5.2 Nucleosomal Binding Associated Fine Structure

The fine structure present in all human sequences (Fig. 6.3C-F), is far more complex than that of the codon usage (6.5.1): The very pronounced local maximum at 11bp might be associated to the double helical pitch, whereas the local minima and maxima thereafter seem related to the nucleosome, being obvious for a maximum at 146bp and less pronounced at 172bp, 205bp, 228bp and 248bp.

To confirm this possible connection, again 10Mbp long random nucleosome sequences were created in which nucleosome binding sequences were organized in blocks (6.2.2). The blocks were equally distributed within a total random sequence.

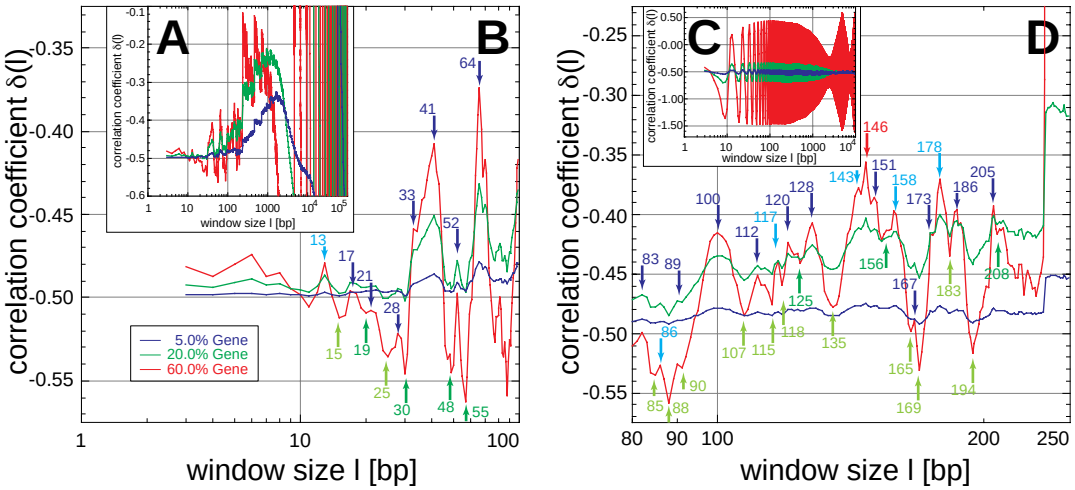


Fig. 6.10 Simulation of the Fine Feature Attributable to the Nucleosome
The fine feature present in all human sequences (Fig. 6.1) is in agreement with the pattern found in simulations using a consensus nucleosomal binding sequence (**A**,**B**,**D**) organized in a block/gene fashion (see also Fig. 6.8). The positions of the local maxima are in many cases the same as in the human genome (dark numbers/arrows are in agreement within ±1bp), whereas the similarity of the position of the local minima is difficult to compare to their smearing out in the human sequence due to the block structure of genomes (Fig. 6.3). Use of a mixture of two special sequence motifs results in highly ordered periodicities of 10bp, attributable to the helical pitch and the base pairs bound to the nucleosomal core (**C**). The appearance, visibility as well as the degree of correlation is again proportional to the concentration of the blocks/genes in the random sequence (legend in **B**), leading also to a general maximum and oscillations of $\delta(l)$ (**A**, embedding hull in **C**).

The gene size of 2750bp was either designed from a consensus sequence of 230bp or a mixture of two special sequence motives of 30 and 20bp. All three motives were based on nucleosomal binding studies (6.2.2). The consensus sequence which contains fixed as well as variable bases is somewhat more resistant against periodicities than the exact mixture of the motives. The fine-structure of the consensus sequence exhibits a very similar pattern with 75% of maxima found within ±1bp also in the real human sequences e. g. 146bp (Fig. 6.10B&D). The low similarity of ~33% for local minima is, however, difficult to compare due to their smearing out by the general multi-scaling behaviour of the human sequences. A correlation between 2000 and 4000bp attributable to the transition of the multi scaling behaviour (6.4.1) was not found. It could, however, be associated to short-range correlations between entire nucleosomes. The general multi-scaling behaviour as found in *Arabidopsis thaliana* stays also unsupported. The appearance, visibility as well as the degree of correlation is once more proportional to the concentration of genes within the random sequence. This carefully predicts a concentration of nucleosomal binding sequences of 5 to 10% in the human sequences. The use of the mixture of two sequence motives results in a first maximum at 13bp as for the consensus sequence and in a highly ordered periodicity of 10bp (Fig. 6.10C), being strongly proportional to the concentration. This periodicity is attributable to the double heli-

cal pitch and not to the short motif length. Both kinds of random nucleosome sequences also produce the multi-scaling behaviour suggested by the block/gene organizatin as in the investigation of the general block organization (6.4.3) or of the codon usage (6.5.1). Into this the fine-structure is embedded (Fig. 6.10A&C). Especially for the mixture of the sequence motives these fine structure periodicities propose an embedding hull defining the block/gene based periodicity (Fig. 6.10C).

6.6 Classification of Correlations by Tree Construction

The fine-structured multi-scaling correlation behaviour of single chromosomes show a higher similarity within different Eukariotic species than between major regna. Especially Archaea and Bacteria sequences can be visually subdivided into different groups with similar behaviour. To investigate quantitatively the relationship between the different correlation behaviours, their classification as well as the comparison with classic phylogeny, $C(l)$ and $\delta(l)$ were used to construct trees (6.2.3). Here the whole or an intervall $[0 \leq l_i, l_j \leq L]$ (with the maximum sequence length L) of the single data sets were used to create correlation matrices by pairwise correlation. These were first transformed into similarity matrices and then into distance matrices for the use in the common program for tree construction PAUP. Trees herein were either constructed by neighbour joining (NJ) or the unweighted pair group method by average (UPGMA). Although UPGMA is said to be older and less sophisticated than NJ, it produced in many cases numerically more stable results. To validate this approach in respect to classic phylogeny, first sequences of β-Tubulin genes of Oomycetes were subject to correlation analysis followed by tree construction, before the Eukarya, Archaea and Bacteria analyses were further enquired.

6.6.1 Tree Construction of β-Tubulin Genes of Oomycetes

$C(l)$ and $\delta(l)$ were calculated for 9 equal sized sequences with ~665bp of β-Tubulin genes from *Peronospora* (Tab. 6.3). Although, the sequences of the β-Tubulin genes on first sight seem only to reveal random correlations (Fig. 6.11A&C) with high degree of insignificant fluctuations considering the shifted standard deviation for such small sequences (Fig. 6.2D&E). However, the graphs show similarities for various scaling regions (Fig. 6.11B&D), which demonstrate that information is still present even though it is visually not apparent. The best tree constructed from $C(l)$ using the scaling region of $l_i=0$ to $l_j=600$bp by neighbour joining (NJ, 6.2.3), resulted in a very good agreement with the assembled phylogenetic tree (Fig. 6.11): *Phytium ultimum* and *Achlya klebsiana* are separated correctly, the subgroups of *Peronospora brassicae* and *Peronospora thlaspeos* as well as *Peronospora trifolii hybridi*, *Peronospora pulveracea* and *Peronospora lamii* are correctly identified.

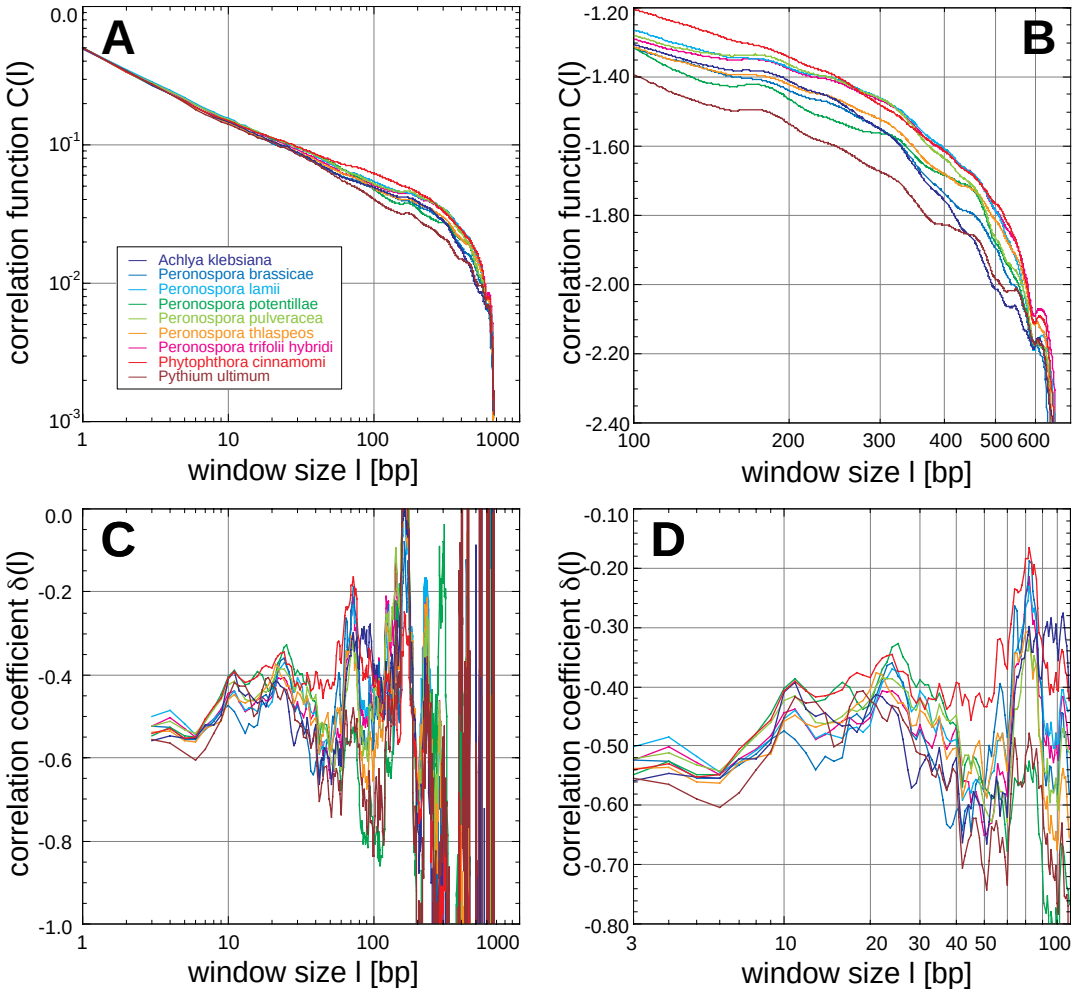


Fig. 6.11 Correlation Analysis of β-Tubulin Genes
The analysis of the DNA sequence of β-Tubulin genes reveals only random correlations in the correlation function $C(l)$ (**A**) and correlation coefficient $\delta(l)$ (**C**). The details of the graphs, however, show obvious similarities between certain sequences which is clearer visible in **B** and **D**. These similarities show the information still contained in the random correlation and could be used for tree creation as shown in Fig. 6.11.

The small differences found are also insecure in the phylogenetic tree as the bootstrap values below 80 indicate (bootstrap values describe the significance of a tree branch and range from [0, 100]). The position of *Phytophthora cinnamomi* would be correct if *Peronospora potentillae* was left out of the analyses. Thus, only *Peronospora potentillae* is misslocated which also has a relatively insecure position in the phylogenetic tree with a bootstrap value of 63.

To locate the regions $[0 \leq l_i, l_j \leq L]$ for which the distance matrices created from $C(l)$ or $\delta(l)$ result in the best agreement with the phylogenetic distance matrices, the regions, i. e. l_i and l_j, were variated. The degree of correlation between both matrices and thus tree types is fittable to second order polynomials and is the highest

Tab. 6.3 Attributes and Correlation Properties of Analysed ß-Tubulin Genes
Due to the unavailability of accession numbers for the unpublished sequences of M. Göker, herbarium numbers are used (MG#). All sequences belong to Oomycetes (O). Properties of correlation are classified with N for negative (crossing the random regime is given in bp), R for random and P for positive correlation coefficients for window sizes of a few base pairs. The fine structure is categorized into codon usage (C) or another distinct fine structure like codon usage (F).

Oomycetes	Accession Number	Category	Length [bp]	Correlation Properties	
				Correlated [N/R/P] [bp]	Fine Structure [C,F]
Achlya klebsiana	J05597	O	661	N<8	-
Peronospora brassicae GÄUM.	(MG-1866)	O	660	N<9	-
Peronospora lamii A. BRAUN	(MG-1867)	O	665	R	-
Peronospora potentillae GÄUM.	(MG-1833)	O	664	N<8	-
Peronospora pulveracea FUCKEL	(MG-1763)	O	664	N<8	-
Peronospora thlaspeos arvensis GÄUM.	(MG-1852)	O	662	N<10	-
Peronospora trifolii hybridi GÄUM.	(MG-1802)	O	665	N<9	-
Phytophthora cinnamomi	U22050	O	664	N<8	-
Pythium ultimum	AF115397	O	664	N<10	-

using $C(l)$ or $\delta(l)$ from 0 bp to the entire sequence length ([0, L], Fig. 6.2 F). A reasonable degree of agreement is also found using [0, l_j<<L] for $\delta(l)$. Thus, the information content important for tree construction is located at [0, $l_j \approx$ L] using $C(l)$ or $\delta(l)$ and [0, l_j<<L] using $\delta(l)$.

In summary, these results suggest that tree construction based on the correlation behaviour can be employed succesfully for the investigation of the relationship between DNA sequences. This method could be especially usefull in cases where the sequences show little similarity, thus no standard construction procedures for phylogenetic trees could be applied anymore.

6.6.2 Tree Construction of Eukarya, Archaea and Bacteria Genomes

Already, the qualitative and complex partitioning into different general behaviours of the fine-structured multi-scaling behaviour (6.4), suggests a more complex tree construction for the Eukarya, Archaea and Bacteria than for β-Tubulin genes. Here different correlation behaviours interfere and fluctuations near the cut-off (6.3) as well as the substantial increase in the standard deviation for $C(l)$ and $\delta(l)$, set-in at different length scales, since the sequences are all of different length. Thus, the region [0, $l_j \approx$ L] for tree construction as in the case of β-Tubulin genes was unusable. Nevertheless l_j was chosen as large as possible.

Tree construction from $C(l)$ and $\delta(l)$ using various scaling regions $l_j < 10^5$ bp for the Eukarya, resulted in complete separation of *Homo sapiens*, *Saccharomyces cerevisae* and *Arabidopsis thaliana*, in agreement with classic phylogeny, where yeasts are related closer to humans than plants. Reducing l_j led to even better separation with higher bootstrap values and therefore higher liability of tree branches,

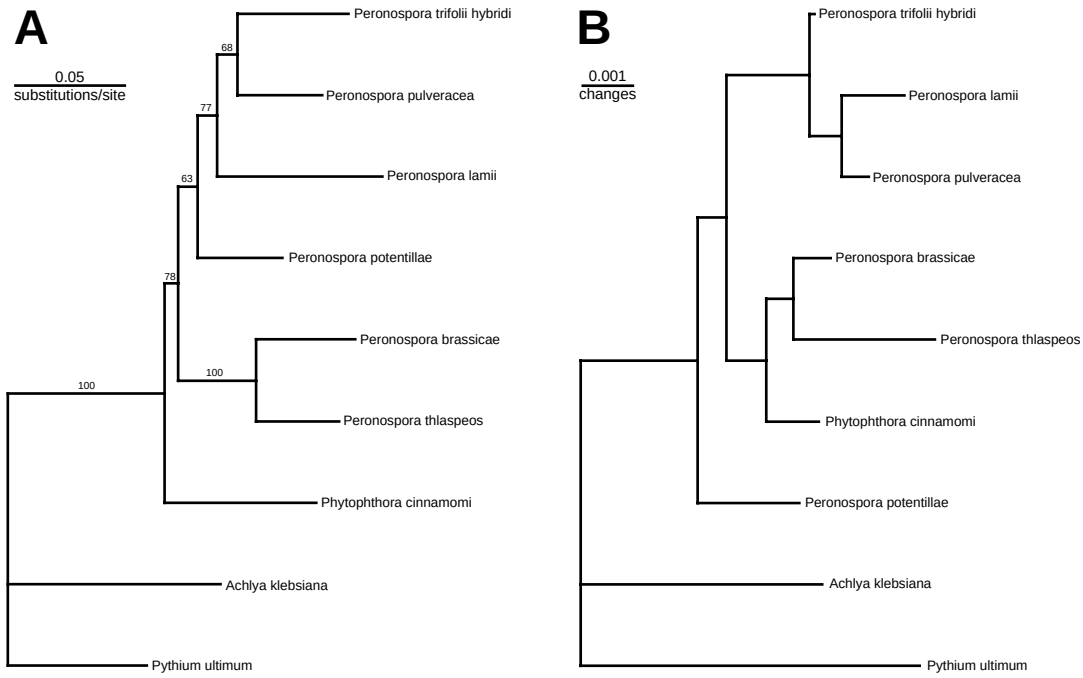


Fig. 6.12 Comparison of Trees Based on Phylogeny and $C(l)$ for β-Tubulin Genes
The best tree based on the function $C(l)$ from 0 to 660 bp (**B**) is very similar to the phylogenetic neighbour joining (NJ) tree (**A**): The subgroups of Peronospora brassicae and Peronospora thlaspeos as well as Peronospora trifolii hybridi, Peronospora pulveracea and Peronospora lamii are correctly identified. The small differences in the latter subgroup is also insecure in the phylogenetic tree as the bootstrap value below 80 indicates. The position of Phytophthora cinnamomi would be correct if Peronospora potentillae was left out of the analyses. Thus, only Peronospora potentillae dislocated which has a very insecure position in the phylogenetic tree with a bootstrap value of 63. Pythium ultimum and Achlia klebsiana share the same positions as the tree was rooted with these two species.

despite eventual single swaps of branches. Thus, the hypothesized information distribution could also hold in the case of largely or completely sequenced genomes.Although, a comparison with $\delta(l)$ supports the sequence order within species, as e. g. in the case for sequences of the bottom arm, bottom arm and concatenated chromosome I of *Arabidopsis thaliana*, the general origin of this order remains unclear: Correlation to sequence length, base pair composition, or any other simple measure of hereditary connections (as far as they are by now available) has remained unsuccessful.

Tree construction mainly based on $\delta(l)$ for the Archaea and comparison to phylogenetic trees constructed from 16S rRNA resulted in correct identification of subgroups containing *Thermoplasma acidophilum* and *Thermoplasma volcanium* as well as *Pyrococcus abyssi* and *Pyrococcus horikoshii* (Fig. 6.14A&B). This is probably due to the very low genetic difference. Tree construction mainly based on $\delta(l)$ 100 and 10^5bp of the Archaea together with the Bacteria leads to two main branches with several subbranches (Fig. 6.15). The Archaea were relatively well separated from the Bacteria. Close related Bacteria and subgrouping of different strains from

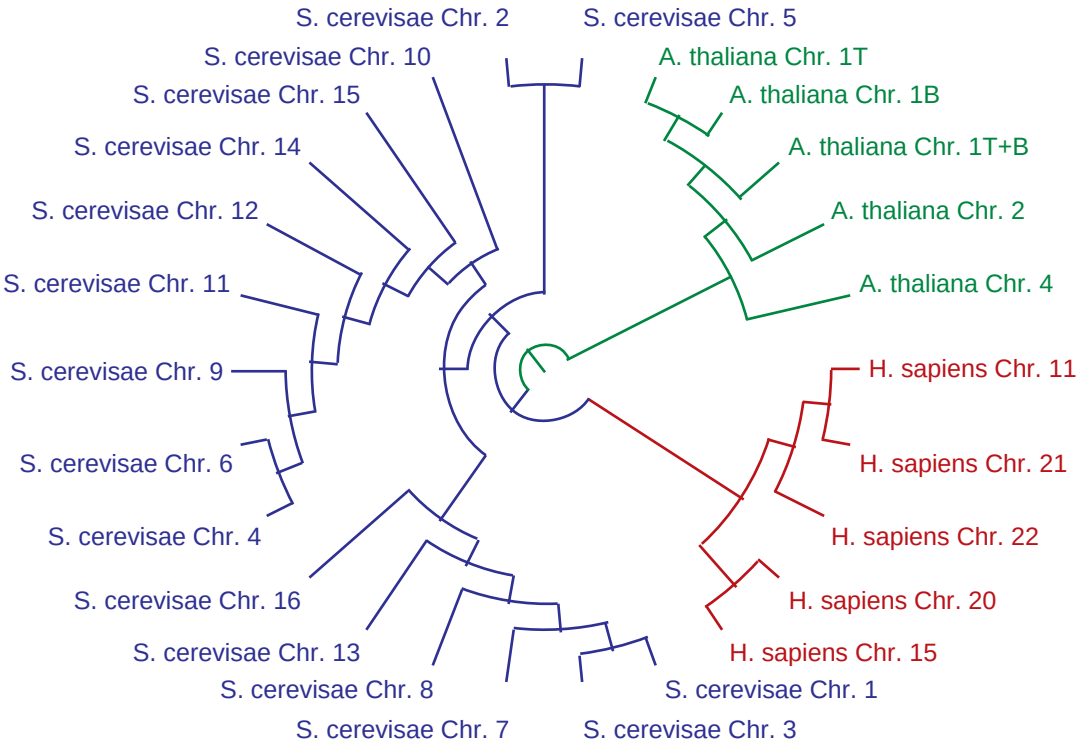


Fig. 6.13 UPGMA Tree of Eukarya Chromosomes Using $C(l)$
Tree construction using the correlation function $C(l)$ for window sizes of 1 to 200 bp shows separation into species for chromosomes of *Homo sapiens*, *Saccharomyces cerevisae*, and *Arabidopsis thaliana* in agreement with classic phylogeny. The order of chromosomes within the species remains unclear and is neither due to chromosome length nor base pair composition.

an Archaea or Bacteria were also correctly grouped. However, a comparison of the correlation based tree to the classic phylogenetic trees remains unsatisfactory. Again no statistically significant connection was found to any other apparent parameter like sequence length, base pair composition, gene size, gene separation, gene density or evolutive development (see also Mira *et al.*, 2001). Detailed comparison of $\delta(l)$ between species within the same subbranches not only supported the already described four correlation behaviours in Archaea and Bacteria (6.4.2), but also lead to a quantitative basis with clear separated borders between the classes A, A', A'' and B. Thus, the qualitative description is supported by tree construction and vice versa.

This classification scheme could also be used to average $C(l)$ and $\delta(l)$ of one class. The resulting average has a little standard deviation. Thus, the distinct classes of the multi-scaling behaviour appears is much clearer revealed (Fig. 6.16 A&B).

Consequently, tree assembly based on the fine-structure and multi-scaling of long-range correlations for Eukarya and especially for Archaea and Bacteria might lead to a new classification system integrating different properties of the general genome organization.

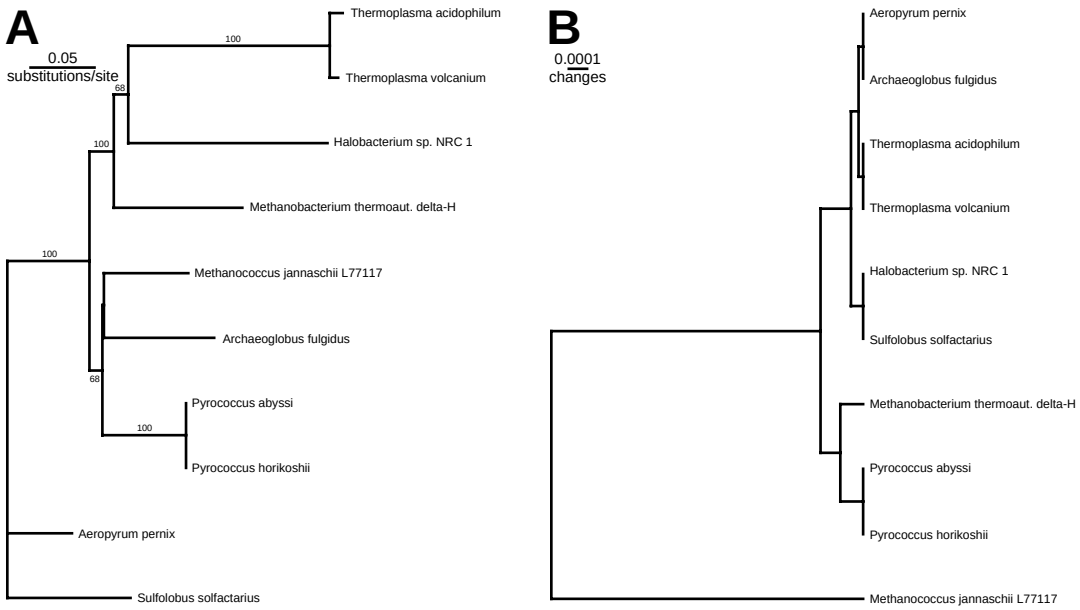


Fig. 6.14 Comparison of Trees based on Phylogeny and $\delta(l)$ for Archaea
Comparison of the UPGMA tree based on $\delta(l)$ using windowsizes from 1 to 200 bp (**B**) to a phylogenetic NJ tree constructed by with the 16S rRNA gene (**A**). The subgroups of *Thermoplasma acidophilum* and *Thermoplasma volcanium* as well as of *Pyrococcus abyssi* and *Pyrococcus horikoshii* were found correctly in the $\delta(l)$ tree. Other relationships were not found correctly.

6.7 Discussion of Long Range Correlations in DNA Sequences

The general sequential organization of genomes and their evolution has been of major interest since the discovery of DNA, its double helical structure (Watson, Crick & Franklin, 1953) and the discovery that it is the primary carrier of information and inheritance (Delbrück, Hershey & Luria, Nobelprizes 1969). First investigations determining the chemical properties, sequential order and self-reproduction of transfer-ribonucleicacids (tRNA) showed not only an organization into codons of 3 bp, but also a maximum stability of self-replicated tRNA at ~75 bp (Eigen & Winkler-Oswatitsch, 1981a; Eigen & Winkler-Oswatitsch, 1981b; Eigen *et al.*, 1981). Thus, the information on the sequence level of genomes evolved in a very defined and delicate interaction with its underlying material carrier - the DNA and other involved molecular agents. However, larger (i. e. $>10^3$ bp) and more sequences were unanalysable until the development of sequencing techniques based on the discovery of the polymerase chain reaction (PCR, Mullis, Nobelprize 1993) and theoretic advances of correlation analyses for texts, time courses, languages and music (Mandelbrot, 1983; Hsü & Hsü, 1990; Hsü & Hsü, 1991; Rabinovitch *et al.*, 1992):

Long-range correlations at least up to 800 bp were found in the mostly noncoding (76% introns) gene of the human-blood coagulation factor VII by fitting the

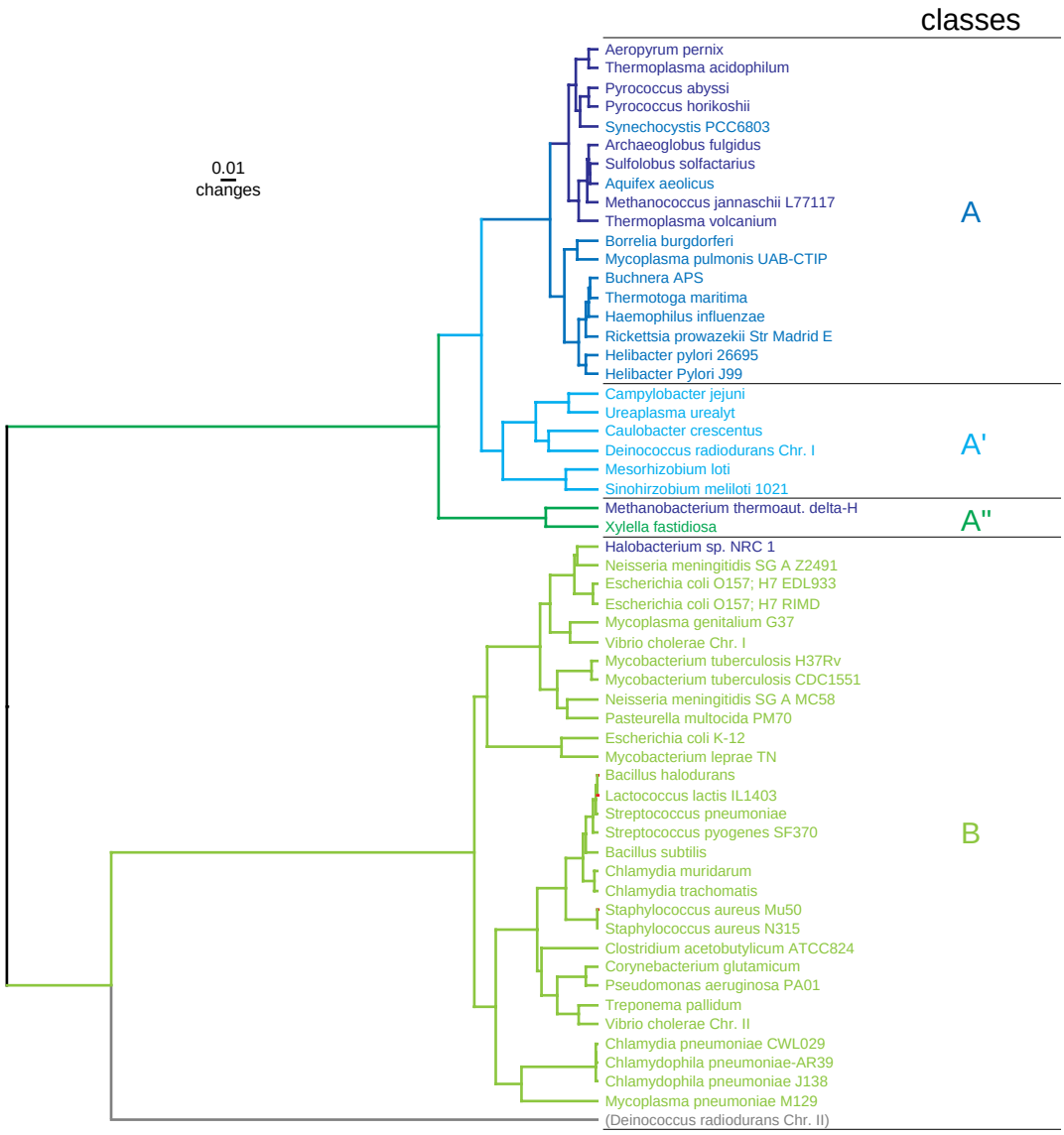


Fig. 6.15 Classification Tree of the Correlation Coefficient $\delta(l)$ for Archaea and Bacteria
The construction of the UPGMA tree using $\delta(l)$ for window sizes from 100 to 10^5bp leads to two main branches. The upper branch can be subdivided in three subbranches A (in which most Archaea form a subgroup; dark blue colour), A' and A'' due to the different and distinct behaviours of $\delta(l)$ (see also Fig. 6.16). Although the lower branch B shows also distinct subbranches the only difference lies in the values of $\delta(l)$ but not in its behaviour. Due to the small sequence size of *Deinococcus radiodurans* Chr. II its separated branch needs to be neglected.

power spectrum $P(f)$ of the mutual information function to a power law $1/f^{\beta}$ (Li & Kaneko, 1992). Despite limited statistics, the correlation coefficient β was different for intron and exon containing regions. This was explained by repetitive subsequences whose generation should be comparable to the copy-and-error mechanism of modern music composition. Mapping several sequences to a two state random

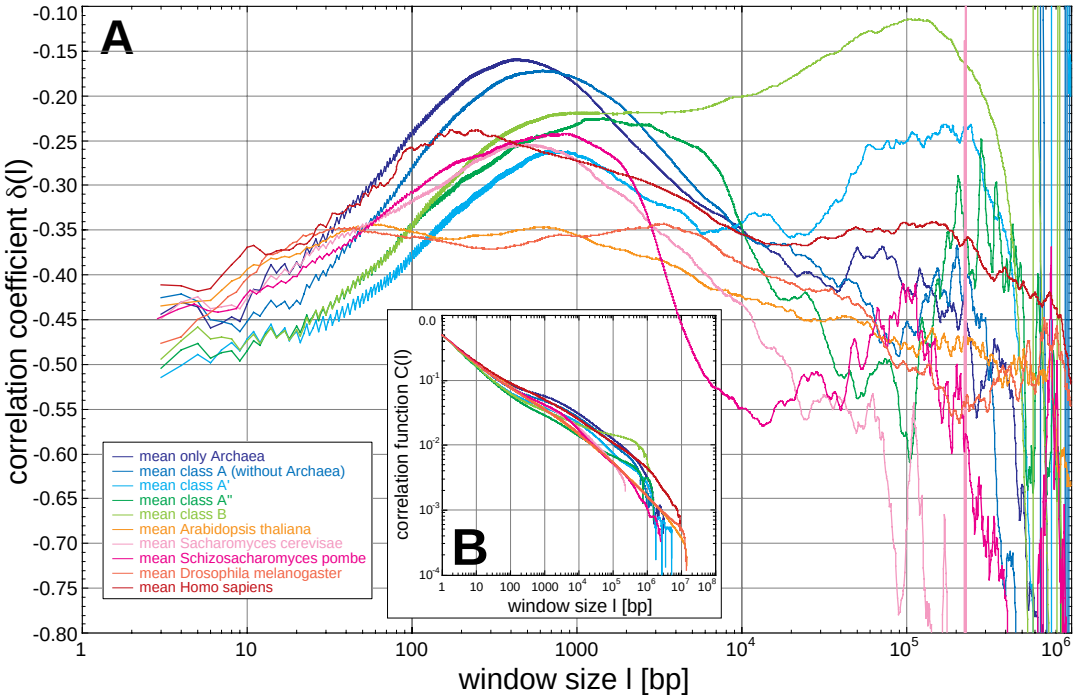


Fig. 6.16 Comparison of Means of Correlation Coefficients $\delta(l)$ for All Analysed Genomes (**A**) Shown are the means over the single analysed chromosomes for each of the Eukarya genomes, the Archaea and the classes A (without the Archaea), A', A'' and B. Comparison reveals that only *Homo sapiens* does not show the zig-zag pattern due to the codon usage, although it shows a fine structure not present in any other genome or class. All genomes show a maximum between window sizes of 100 to 1000 bp of which only the maxima *Homo sapiens* seems to be connected to the nucleosome. The classes A' and B show a second maximum after a decrease of δ with very high correlations for window lengths of ~10^5 bp in contrast to the other genomes. Only *Homo sapiens* shows also a second maximum although in the mean it is washed out and is not statistical significant in contrast to analysis of some of the single analysed human chromosomes. (**B**) For comparison the means of the concentration fluctuation function $C(l)$ for the same averages are shown.

walk and using the so called Peng-function (Peng *et al.*, 1992) for analyses, extended the long-range correlations to 10^3 bp in intronrich, in contrast to intronless genes. Here only random correlations were present. These behaviours were interpreted as non-equilibrium and equilibrium states, being of general fractal nature. Simultaneously, long-range correlations with similar extend and a '$1/f^{\beta}$-noise' character were found by Voss (1992), in 25.000 sequences (the total GeneBank Release 68) in 10 different organism groups (primate, rodent, mammal, vertebrate, invertebrate, plant, virus, organelle, bacteria and phage). The use of the (equal-symbol) spectral density function Reif, 1965; Robinson, 1974) also revealed a periodicity of 3 bp caused by the codon usage and a periodicity of 9 bp of unknown origin, but characteristic for primates, vertebrates and invertebrates.

Besides the widespread astonishment on how such correlations could have persisted and evolved over thousands of base pairs (Amato, 1992; Madox, 1992), the reports induced a broad discussion about the validity of the results: On the one hand,

the originality of correlations was questioned and attributed to the mere presence of regions with biased base pair composition (Nee, 1992). Computer generation of such patchy sequences seemed to support these results. Random mutation and reshuffling of such sequences as well as the bacteriophage lambda destroyed any correlation (Carlin & Brandel, 1993). On the other hand, the existence of long-range correlations was rejected at all since the Peng function shows a not exactly linear power-law behaviour (Prabhu & Claverie, 1992; Chatzidimitriou-Dreismann & Larhammar, 1993). Proposition of a Levy-walk model for the sequences solved these inconsistencies (Buldyrev *et al.*, 1993). It possibly could also account better for the evolution of long-range correlations than their interpretation as stationary fractional Brownian motion (Allegrini *et al.*, 1998). Long-range correlations could finally be established by Peng *et al.* (1994) by development of detrended fluctuation analysis (DFA), which is an alternative method differentiating local patchiness from long-range correlations. The existence between intron and exon containing sequences was also proven by DFA (Buldyrev *et al.*, 1995). Concerning the evolution and persistence, copy and deletion models were discussed, as well as close connections to the globular three-dimensional organization of genomes (Takahashi, 1989; Grosberg *et al.*, 1993; Stanley *et al.*, 1994; Borovik *et al.*, 1994; Mira *et al.*, 2001).

Additionally, methods and results were further validated by comparing different methods (Borovik *et al.*, 1994, Luo *et al.*, 1998) and extended to fractal cantor pattern recognition (Provata & Almirantis, 2000), factorial moments analysis (Mohanty & Narayana-Rao, 2000), rescaled range transition matrix analysis (Yu & Chen, 2000), as well as two-dimensional visualizations (Yu *et al.*, 2000; Hao *et al.*, 2000a; Hao *et al.*, 2000b). Sequence evolution mechanisms inspired by language evolution were also proposed (de Oliveira, 1999; Mackiewicz *et al.*, 1999; Hao *et al.*, 2000a).

Regarding periodicities or correlations connected to the codon usage (Voss, 1992) or nucleosomal binding sequences (Ambrose *et al.*, 1990) only sequences known to contain these features were analysed: A variety of periodicities were found (Blank & Becker, 1996; Liu & Stein, 1997; Lowary & Widom, 1998; Baily *et al.*, 2000).

Despite all the above described efforts the general sequential organization of completely sequenced genomes and especially the extend to which long-range correlations exist, whether the degree of correlation itself depended on the scale of observation (multi-scaling), whether such correlations are linked to the three-dimensional genome organization and whether they are specie specific has not been resolved yet.

To investigate these properties the concentration fluctuation function $C(l)$ and its exponent $\delta(l)$, the local correlation coefficient, were calculated using numerically exact algorithms for in total 113 complete sequences of $0.5\text{x}10^6$ to $3.0\text{x}10^7$ bp from *Homo sapiens*, *Drosophila melanogaster*, *Saccharomyces cerevisae*, *Schizosaccharomyces pombe*, *Arabidopsis thaliana*, Archaea and Bacteria. The results revealed long-range correlations in all analysed sequences almost up to the

entire length scale of sequences, but at least up to 10^5 to 10^6bp. This is an increase of at least 2 to 3 orders of magnitude compared to the cited literature.

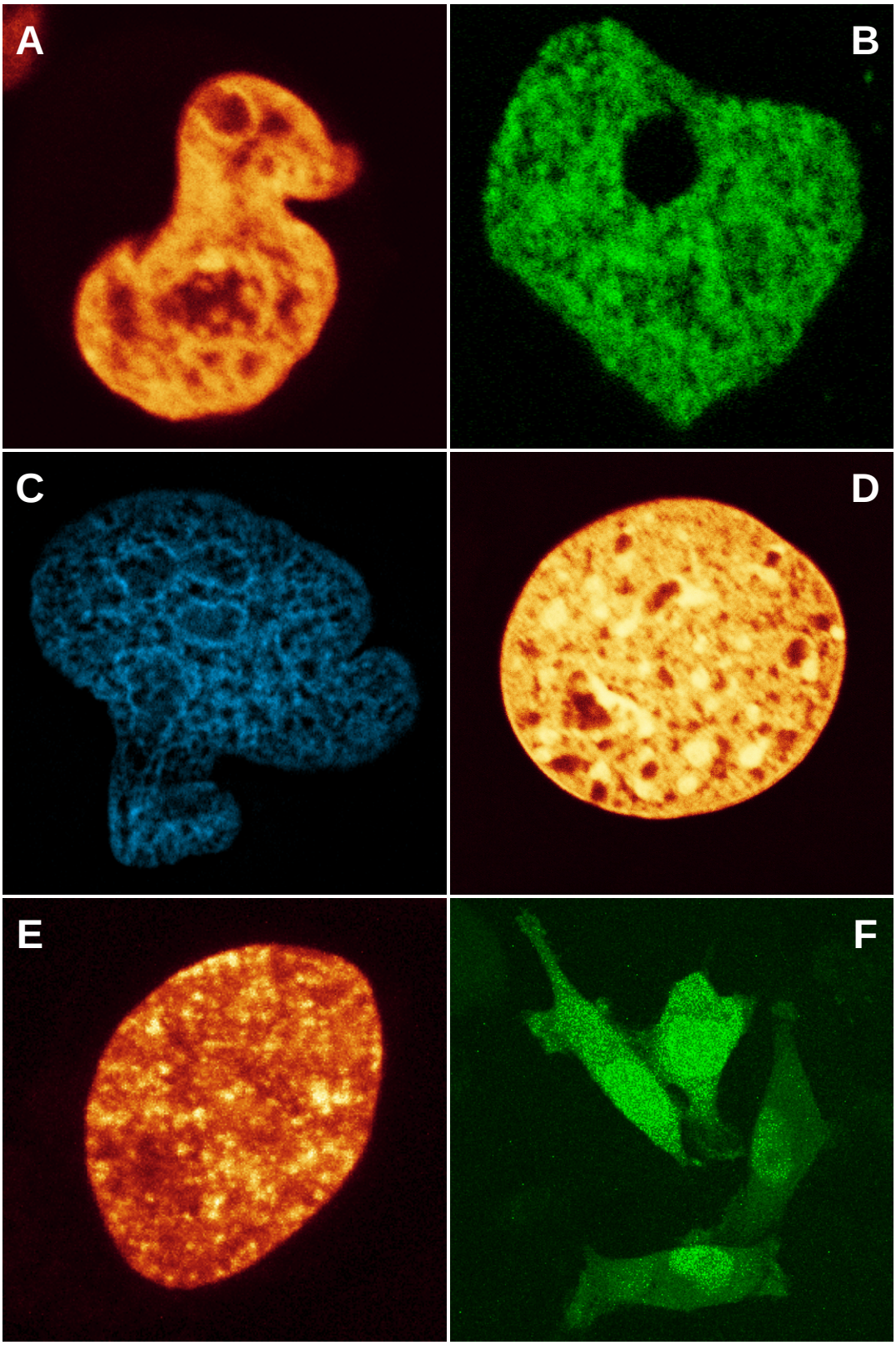
Beyond the appearance of a simple power-law behaviour of the long-range correlations showed more complex behaviour. The calculation of the local correlation coefficient $\delta(l)$ showed a maximum between 50 and 2000bp and sometimes a region of one or more second maxima at ~10^5bp. This so called multi-scaling behaviour was species specific. Especially the human sequences display very pronounced second maxima. This might be connected to the globular organization in chromosome subcompartments. Also many Bacteria show a remarkable degree of correlation at this scale, whose origin is unknown. Analysis of computer-generated random sequences suggested that this type of correlation might originate from a block wise sequence organization. The existence of a multi-scaling behaviour not only disclosses the above described literature discussion about the validity of correlation results, but also suggest a block organization for their origin.

Investigation of the evolutive persistence of the multi-scaling by simulation of mutations by sequence reshuffling within the sequences resulted in total loss of (multi-scaling) correlations. Thus, the general sequential and the three-dimensional organization seem to be closely connected as otherwise precise rearrangement on large scales cannot take place without loosing the multi-scaling behaviour.

Within the multi-scaling correlation behaviour, a additional species specific fine-structures were found which are attributable to the codon usage except for the human sequences. Here the fine-structure is connected to nucleosome binding. The connection was again proven by arteficial random sequence design.

To investigate quantitatively the relationship between the different correlation behaviours, their classification as well as the comparison with classic phylogeny, trees were constructed from the multi-scaling correlations. For β-Tubulin genes of Oomycetes and Eukarya the results are in agreement with the phyologenetic trees. For Archaea and Bacteria four new major tree branches/classes were found. Since these classes seem unconnected to any apparent parameter like, e. g. base pair composition or gene content, tree construction by correlation analysis might lead to a new classification system of genomes integrating the different properties of the general organization of whole genomes

In summary, these findings suggest a complex sequential organization of genomes closely connected their three-dimensional organization.


A
B
C
D
E
F

7 “Chromatin Alive“: *In vivo* Analysis of the Chromatin Distribution in Cell Nuclei

7.1 Introduction

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

Fig. 7.1 Interphase Morphology of Various Histone-AFP Constructs in Different Cell Lines
The morphology differs between the human HeLa-H2A-YFP (**A**, image sidelength (IS):20 µm), the primate Cos7-H1-GFP (**B**, IS: 20 µm), the human LCLC103H-H2A-CFP (**C**, IS: 30 µm) and the mouse ID13-H2A-YFP (**D**, IS: 25 µm) cell lines, thus they have a different three-dimension organization of their genome although expressing the same histone. The morphology of the human HeLa-mH2A1.2 (**E**, IS: 20 µm) cell line differs from the usual HeLa-H2A-YFP (**A**), whereas expression H2A by the natural promoter did no differ (HeLa-H2A-NP, **F**, IS: 100 µm).

7.2 Vector Construction, Cell Culture and Microscopy

7.2.1 Vectors

Basis: The histone genes were inserted into the pSV-HIII-CFP and pSV-HIII-YFP plasmids, both being based on the promotorless pECFP-1 plasmid (Clontech). For eukaryotic expression the Simian Virus (SV40) HindIIIc promoter region (base pair 1046 to 5171, 1118bp) was inserted in reverse direction into the HindIII site of the multiple cloning site (MCS) of pECFP-1 for higher expression efficiency. For the pSV-HIII-YFP plasmid the eCFP was exchanged with the eYFP coding sequence of pEYFP-NUC plasmid (Clontech) using AgeI-BsrgI restriction sites (Tab. 7.1&2).

PCR of Histone Genes with Corresponding Sites for Ligation: The histone genes were inserted into the MCS between the HindIIIc-regulatory-region and the start codon of the AFP gene. To avoid histone AFP protein interactions a spacer of 18 to 33bp (6 to 11 amino acids) was inserted between both sequences (Tab. 7.1). Histone genes were amplified with polymerase chain reaction (PCR) from either existing plasmids (H1.0, H2A.i, H2B.a, all three from A. Alonso, DKFZ, Heidelberg; H2B.d-NPII-H2A.d, from D. Doenecke, University of Göttingen), Caski genomic DNA (H3.1, H4.a) or a cDNA clone (mH2A1.2, IMAGp998A141538, I.M.A.G.E. Consortium, RZPD, Berlin) (accession numbers in Tab. 7.1). The PCR was carried out in 100µl reaction volume in small siliconized 600µl polypropylen Eppendorf tubes (Biozym Diagnostik GmbH). The reaction used 10µl of GeneAmp® 10x PCR Buffer II (Perkin Elmer), a 2µl mix of dNTP with a 25mM concentration (Peqlab Biotech GmbH), 8µl of 25mM of $MgCl_2$ ($MgCl_2$ Solution, Perkin Elmer), and 2µl of 2U/µl DNA Polymerase AmpliTaq® (Perkin Elmer). As PCR template, 4ng of plasmid DNA or 1µg of genomic DNA at a maximum of 3µl were used. 0.3µg of forward and reverse primers containing also newly introduced restriction sites for ligation (see Tab. 7.2) were used at a maximum of 3µl. For thermal cycling a 3min preheating to 96°C was followed by 30 (plasmid DNA) or 33 (genomic DNA) cycles of 30sec at 96°C denaturation, 30 sec at 50°C annealing, 30 sec at 72°C elongation and a final extension of 10min at 72°C. The PCR product was purified from enzymes, salt, unused primers and dNTP with the QIAquickTM PCR Purification Kit (Quiagen). Products were analysed on a 8% polyacrylamide gel (6cm, TBE buffer [90mM Tris, 90mM Boracid, 3mM EDTA, pH 8,3], 90V for 90min, Lambda HindIII and 100bp ladder as markers) before and after the purification.

Restriction of the PCR Products and Vectors: Various restriction enzymes for cleavage of the PCR product before insertion into pSV-HIII-CFP and pSV-HIII-YFP were used (Tab. 7.1&7.2). Restriction used a maximum of 10µg of PCR product or 20µg of acceptor-plasmid. 10 to 20U (1 to 2µl) of enzymes (Boehringer, Gibco BRL, New England Bio Labs, Promega) and 3µl of the 10x Restriction Buffer (from

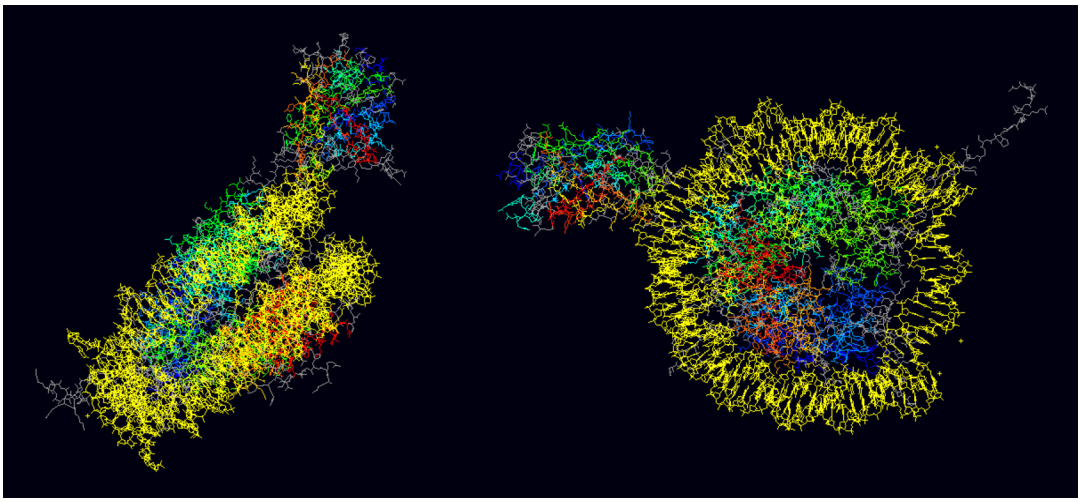


Fig. 7.2 Structure Proposal for a Nucleosome Containing a Histone-AFP Fusionprotein
The nucleosome containing the histones H2A, H2B, H3, H4 including the wound around DNA (yellow) with a directly but randomly attached GFP molecule (colours indicate the different folding features of the proteins) demonstrate the scaling dimensions of the two structures. As the position of the C-terminus (white), that means the position of the highly variable histone tails, is unknown the GFP was not attached their as in the constructed histone-AFP fusionproteins. Consequently, one needs to imagine that the GFP could be as far extended as the visible histone tail (upper right) demonstrates.

the corresponding companies) were added in 30 ml reaction volume and 2 h incubation. To cleave the EcoR1-BamH1 and the EcoR1-Sal1 sites double digests were performed. The restriction of the PCR product was either directly purified with the QIAquickTM PCR Purification Kit (Qiagen) or cut out of a 1% Agarose Gel (20 cm gel, TAE buffer [4.8 g/l Tris, 1.15 ml/l acetic acid, 1 mM EDTA, pH 8.0], 150 V, 4 °C) and then extracted with the QIAquickTM Gel Extraction Kit (Qiagen) or the Ultrafree®-DNA Gel Extraction Kit (Millipore). For further purification an ethanol precipitation and a wash step with 70% ethanol followed.

Ligation and Transformation into Bacteria: The acceptor plasmid (max. 3 µg) was mixed with the restricted PCR products (max. 300 ng) and 1 µl T4 DNA ligase with 1 U/µl (Boehringer) and 3 µl of 10x T4 Ligase Buffer (Boehringer) yielding a reaction volume of 30 µl. The reaction was conducted at 4 °C overnight.

For transformation into Epicurian Coli® XL10-Gold Ultra Competent Cells (Stratagene) a 10 µl aliquot of bacteria stock from -80 °C was thawed on ice for 10 min. Then 1.7 µl of ß-mercaptoethanol were added and kept on ice for 10 min under gentle mixing every 2 min. 10 µl of the ligation mixture (at maximum 1 µg) and 10 µl Luria Bertani (LB) medium was added. After incubation for 30 min a 45 sec heat treatment at 42 °C followed. Subsequently the cells were cooled on ice for 2 min, then 450 µl with 37 °C SOC medium (20 g/l Bactotrypton, 5 g/l Yeast Extract, 0.5 g/l NaCl, 10 mM MgCl, 10 mM $MgSO_4$, 2.5 mM KCl, 20 mM glucose, ph 7.5, sterile filtrated) were added and the bacteria were incubated at 37 °C for

Tab. 7.1 Constructed Histone-AFP Plasmids and Their Properties
The promoter used is given in the name of the plasmid and is either the early and late promoter contained in the SV40 HindIII restriction fragment, the CMV promoter from Clontech or the natural Promoter NPII. If the same plasmid was constructed with different colours the plasmid name was abbreviated with XFP (see column eight for details). The plasmid with the bidirectional promoter NPII contains H2B.d and H2A.d, of which only the H2A.d is fused to YFP. A new MCS was inserted between the H2B.d and its stop codon for later use with a non-homologous AFP according to the found construct conversions during simultaneous co-transfection (Chapter 8; Fig. 7.1).

Plasmid	Histone				Spacer Size [bp] ([AA])	Fluorescent Protein		Length of Fusion Protein [bp] ([AA])	Plasmid Size [kbp]
	Name	Accession Number	Length [bp] ([AA])	Insertation MCS Site		Terminality	eCFP eGFP eYFP DsRed1/2		
pSVHIII-H1.0-XFP	H1.0	M87841	582 (194)	SalI BamHI	21 (7)	N	C/G/Y R1/R2	1320 (440)	5.9
pSVHIII-H1.0-YFP-His	H1.0	M87841	582 (194)	SalI BamHI	21 (7)	N	Y	1320+18 (442)	6.0
pSVHIII-H2A.i-XFP	H2A.i	X83549	390 (130)	EcoRI BamHI	21 (7)	N	C/G/Y R1/R2	1128 (376)	5.7
pSVHIII-H2B.a-XFP	H2B.i	X57127	378 (126)	EcoRI BamHI	21 (7)	N	C/G/Y R1/R2	1116 (372)	5.7
pSVHIII-H3.1-XFP	H3.1	X57128	408 (136)	EcoRI BamHI	21 (7)	N	C/G/Y R1/R2	1146 (382)	5.7
pSVHIII-H4.a-XFP	H4.a	X60481	309 (103)	EcoRI BamHI	21 (7)	N	C/G/Y R1/R2	1047 (349)	5.7
pSVHIII-mH2A1.2-XFP	mH2A1.2	IMAGp998A141538	1116 (372)	EcoRI SalI	42 (14)	N	C/Y R1/R2	1875 (625)	6.4
pCMV-GFP-mH2A1.2	mH2A1.2	IMAGp998A141538	1116 (372)	EcoRI SalI	33 (11)	C	G	1866 (622)	5.8
pH2B.d-NPII-H2A.d-YFP	H2B.d NPII H2A.d	Z83336 - Z83739	378 (126) 322 (-) 390 (130)	HindIII BamHI	18 (6) - 21 (21)	N - N	- - Y	396 (132) 1128 (376)	5.6

60 min, gently mixed every 10 min. Finally the bacteria were put on agar plates under Kanamycin (30 µg/ml agar) selection. Existing clones were picked after 24 h and transferred into 20 ml LB medium containing Kanamycin (30 µg/ml) and incubated on a shaker at 37 °C overnight. Plasmid DNA was extracted from the bacteria with the Nucleo Bond Plasmid Kit (Clontech). Successful ligation was tested by a restriction of 300 ng plasmid DNA using the corresponding restriction enzymes and a 1% agarose gel (25 cm, TAE buffer [0.4 M Tris, Acetic Acid, 0.01 M EDTA, pH 7.5], 150 V, for 3 h at 4 °C (Fig. 7.4). DNA of successful clones was sequenced (Andreas Hunziker, German Cancer Research Center (DKFZ), Heidelberg, Germany) with the reverse EGFP-N1 or DsRed1-N1 primer (Clontech) and the forward primer of the histone PCR (Tab. 7.2). The reverse primer is located within the AFP gene, thus the purity of the histone insertation could be determined. Thereafter one clone was chosen for a DNA maxiprep, again followed by a DNA product control as described.

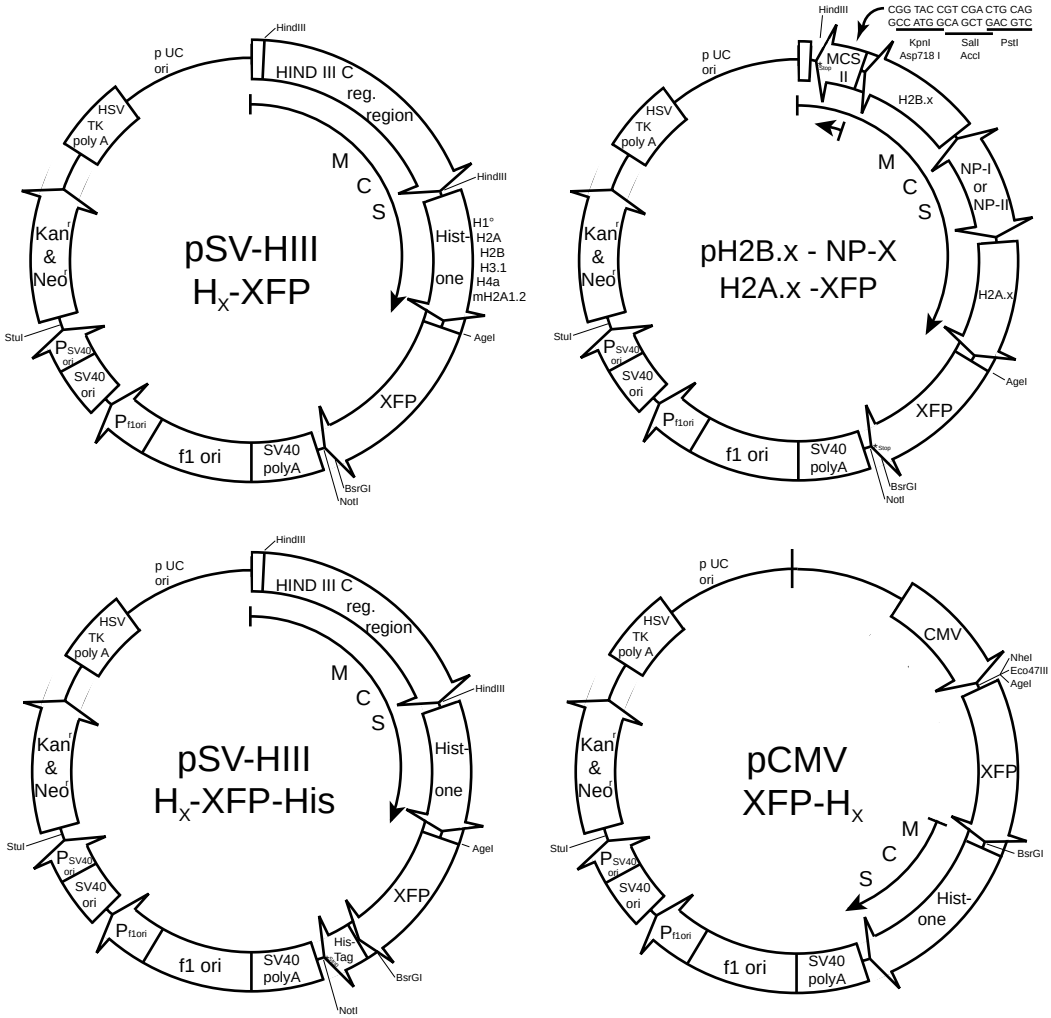


Fig. 7.3 Basic Maps of Constructed Histone-AFP Plasmids
In the most frequently used vector the histone-AFP fusion protein is expressed by the early SV40 promoter located in the inversely located HindIIIC fragment of the Simian Virus 40 (**pSV-HIII-H_x-XFP**). For investigation of the natural histone regulation the bidirectional H2B and H2A gene complexe with the inbetween located bidirectional promoter (NP) was inserted C-terminally to the AFP (**pH2B.x-NP-X-H2A.x-XFP**). To allow the labelling with a second AFP C-terminally to the H2B a second multiple cloning site (MCS) was added. For purification of histones on adsorption columns a so called His-tag consisting of the coding sequence for six histidin amino acids was introduced at the C-terminus of the XFP (**pSV-HIII-H_x-XFP-His**). For determination of terminality effects of the AFP the histones were also inserted N-terminally of the AFP (**pCMV-XFP-H_x**). For vector construction see 7.2.1 and for constructed vectors Tab. 7.1.

7.2.2 Transfection in Eucariotic Cells and Cell Culture

0.9ml of RPMI 1640 medium (Gibco BRL) and ~4µg of the pSV-HIII-H_x-XFP, pSV-HIII-H_x-XFP were mixed and sterile filtrated with a Millex-GV4-22µm filter (Millipore). Simultaneously 0.9 ml of RPMI 1640 were mixed with 10 µl Lipo-

Tab. 7.2 Primers Used for Plasmid Construction
Primers were produced with a DNA synthesizer of Applied Biosystems, as 0.2 µmol synthesis to a final concentration of 1-2 µg/µl and purified by reversed phase high performance liquid chromatography (R-HPLC) by W. Weinig (German Cancer Research Center (DKFZ), Heidelberg). (F: forward; R: reverse; sequences are from 5' to 3'; * newly inserted reverse Multiple Cloning Site (MCS).)

Histone	Primer Direction [F/R]	Restriction Site	Sequence of Restric-tionsite	Sequence in Histone	Primer Length [bp]
H1.0	F	SalI	CAGTCGACG	ATGACCGAGAATTCCACG	27
H1.0	R	BamHI	TGGATCCCG	CTTCTTCTTGCCGGCCCT	27
H2A	F	EcoRI	CGAATTCTG	ATGTCGGGACGCGGCAAG	27
H2A	R	BamHI	TGGATCCCG	TTTGCCTTTGGCCTTGTG	27
H2B	F	EcoRI	CTTCGAATTCTG	ATGCCTGAACCAGCT	27
H2B	R	BamHI	CGGTGGATCCCG	CTTGGAGCTTGTATACTTGG	32
H3	F	EcoRI	CTTCGAATTCTG	ATGGCTCGTACGAAGCAAACAGCT	36
H3	R	BamHI	CGGTGGATCCCG	TGCCCTTTCCCCACGGATGCG	33
H4	F	EcoRI	CTTCGAATTCTG	ATGTCTGGACGTGGTAAGGGC	33
H4	R	BamHI	CGGTGGATCCCG	ACCGCCAAAGCCATAAAGGG	32
mH2A1.2	F	EcoRI	CTTCGAATTCTG	ATGTCGAGCCGCGGTGGG	30
mH2A1.2	R	SalI	TACCGTCGACTG	GTTGGCGTCCAGCTTGGC	30
mH2A1.2	$F_{sequencing}$	-	-	GGAAGCCATCATCACACCACC	21
mH2A1.2	$R_{sequencing}$	-	-	CCAGCTACTTCCAAGGGCCCG	21
H2B.d-NPII-H2A.d	F	HindIII	GCTCAAGCTTCGA	TCA - *CGGTACCGTCGACTGCAG** - CTT-GGAACTGGTGTACTTGGTGACG	59
H2B.d-NPII-H2A.d	R	BamHI	CGGTGGATCCCG	CTTGCCCTTGGCCTTGTGGTGGC	35

fectamin (Transfection Reagent Kit, Gibco BRL). Both solutions were combined and incubated for 10 min at room temperature and slightly mixed 5 times. Meanwhile small cell culture flasks with ~10^6 cells were washed twice with RPMI 1640 and the transfection mixture was added. After 6 h of incubation the transfection mixture was discarded and 10 ml of RPMI 1640 containing 10% fetal calf serum (FCS) were added. At least 60% of the cells showed positive AFP fluorescence after 24 h (transient transfection). For stable transfection the cells were selected under G418 pressure (500 µg/ml) resulting in >80% positive cells after 2 weeks. To receive monoclonal populations concerning the expression of the fusionprotein, cells were dry-trypsinized (~1 ml of a 0.125% trypsin solution was added and discarded leaving only a trypsin film on the cells) with a 0.125% trypsin solution (8 g/l NaCl, 0.2 g/l KCl, 0.02 g/l KH_2PO4, 1.15 g/l Na_2HPO_4-$7H_2O$, 0.01 g/l $CaCl_2$, 0.1 g/l $MgSO_4$-$7H_2O$, 1,25 g/l trypsin), diluted and put into Limbo plates. Positive clones were picked and grown up resulting in 99% positive cells with similar fluorescence.

HeLa (ATCC CCL 2), ID13 (kindly provided by P. Howley, National Institute of Health, Bethesda, USA), COS–7 (ATCC CRL-1651) and LCLC–103H, (DSMZ #ACC 384) cells were grown in phenol-red free (due to its fluorescence emission at the same spectral position as the used AFPs) RPMI 1640 supplemented with 10%

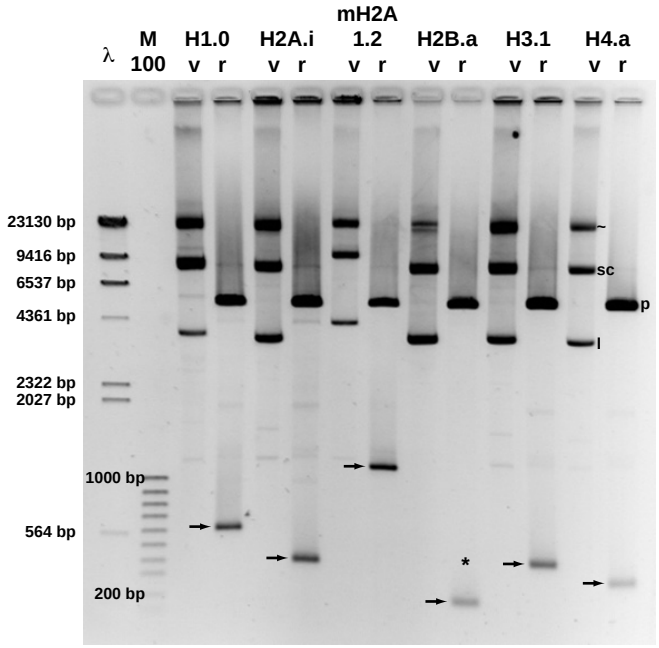


Fig. 7.4 Control Gel of Successful Histone-AFP Vector Construction after Ligation
Agarose gel (1%, 20 cm, at 4°C overnight, λ is the λ-HindIII DNA Marker and M100 is the 100 bp ladder marker) showing the vector (v) in its usual conditions (sc: super coiled; l: linearized) and the restricted vector (r; p: rest of vector) with its insert (arrows). The H2B histone is restricted twice by EcoRI due to restriction site in the middle of its coding sequence and runs at 180 bp and not at the 378 bp (*). ~ marks bacterial DNA due to improper DNA extraction of vectors.

FCS in a humidified atmosphere under 5% CO_2 and at 37°C. The cells were usually trypsinized twice a week with no apparent effect on the expression or fluorescence of the fusionprotein for more than 4 years (e. g. ~300 to 350 passages!) in the case of HeLa and Cos-7 cells. ID13 cells are resistant against G418 selection thus a proper cloning was only possible through dilution cloning.

For microscopy, cells were seeded in an appropriate density on 45 mm diameter round coverslips with 170±10 µm thickness (Langenbrink) using phenol-red free RPMI 1640 with 10% FCS. The ethanol washed and autoclaved coverslips were placed in 10 cm Petri dishes, thus enough medium buffered cells excretions and debris. After 24 h the medium was exchanged and the coverslip transferred to the measurement chamber with 4 ml medium. The chamber was then equilibrated for 1 h under the usual culture conditions prior to uncooled transfer to the microscope.

7.2.3 Proliferation and Cell Cycle Analysis of Cell Clones

Proliferation: $2x10^5$ cells were seeded in 2.5 cm diameter Petri dishes. After 6, 24, 48, 72 and 96 h the cells were dry-trypsinized, stopped and suspended in 2 ml of phenol-red free RPMI 1640 with 10% FCS. To determine the cell number, 100 µl of a

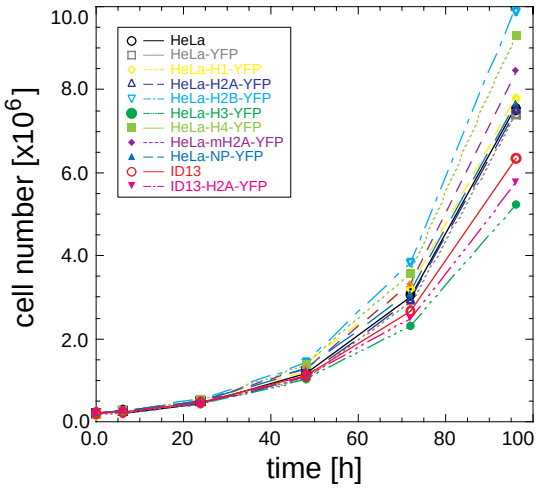


Fig. 7.5 Proliferation of Histone-YFP Expressing Cell Lines
The cell number increase as function of time is comparable in the control HeLa or ID13 cells and YFP or histone-AFP expressing stable cell lines being more obvious by calculating the proliferation rate (Tab. 7.3). Notably, after 60h or ~3 cell divisions proliferation depends on the medium conditions, cell culture flask as well as the number of apoptotic cells, thus results might not represent the initial cellular growth properties adequately anymore.

good mixed suspension was put into a Neubauer counting chamber and manually counted. To assure statistical significance, 3 to 5 Petri dishes were counted at each time point and averaged. The proliferation rate was calculated as the exponent in an exponential fit of the cell number as function of time. Since not all cell lines were measured simultaneously, the proliferation rates (Fig. 7.5, Tab. 7.3) were calibrated to the mean value of various measurements of pure HeLa cells used as standard.

Cell Cycle Analysis: Cells were seeded in 25cm^2 cell culture flasks. After 24h the medium was put into a centrifuge tube, and the cells were dry-trypsinized for ~2min, adding the discarded trypsin to the tube, stopped and suspended with 3ml phenol-red free RPMI 1640 with 10% FCS. The trypsination was carried out to assure the cell being single and no agglomerates were present. The cell suspension was added to the tube and centrifuged twice at 700rpm in a Sorval centrifuge, the supernant was discarded and the pellet was dissolved in Hank's solution avoiding agglomerates. Then cells were fixed by harshly squirting 1ml ethanol at -20°C through the pipette on the cell pellet. To dissolve the pellet and to destroy agglomerates the solution was pipetted up and down ~5 to 10 times. To complete fixation 1ml of ethanol was added and the tube stored at -20°C for a minimum of 24h.

For cell cycle analysis by flow cytometry (Fig. 7.6, Tab. 7.3) cells were again centrifuged and the pellet resuspended in a DNA staining solution containing 5µM DAPI (4,6-Diamidino-2-phenylindole-2HCl, SERVA, Heidelberg, Germany) and 5µM SR101 (Sulforhodamine 101, Eastman Kodak, Rochester, USA) as a protein counter stain (Stoehr *et al.*, 1978). The analyses was carried out on a

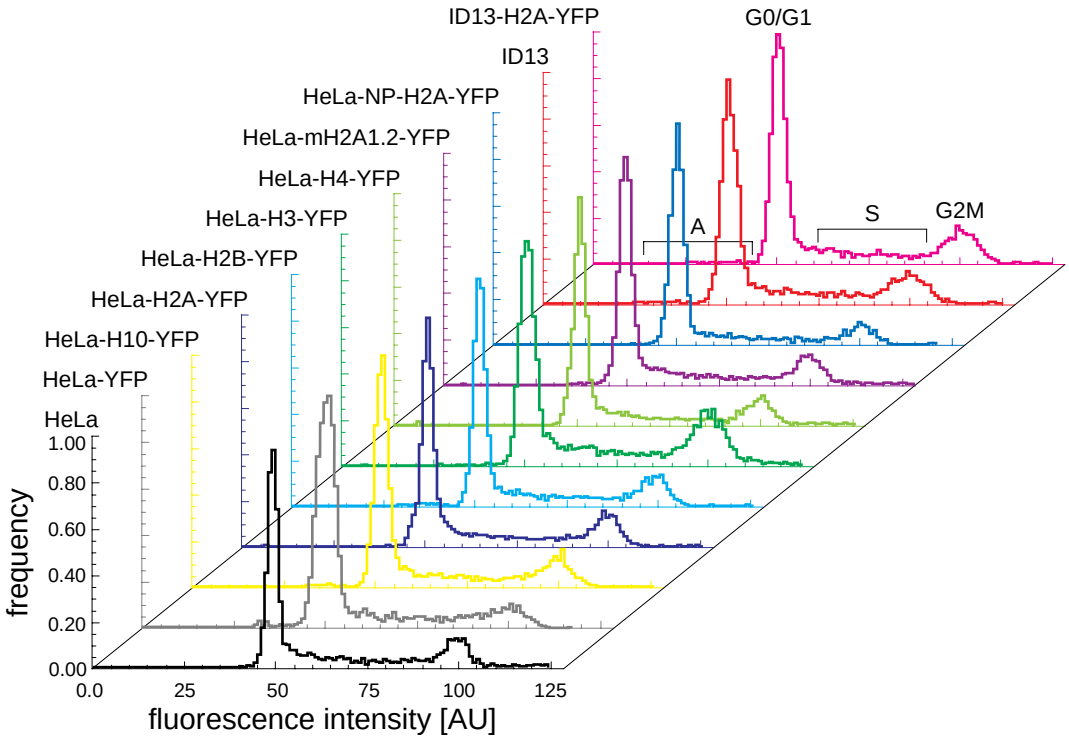


Fig. 7.6 Cell Cycle Analysis of Histone-YFP Expressing Cell Lines
Quantification by fitting the G1/G0, S and G2M peaks of the shown cell cycle distributions obtained by fluorescence activated cell sorting (FACS) to Gaussian distributions, reveals no difference between control HeLa or ID13 cells and YFP or histone-YFP expressing stable cell lines (Tab. 7.3). The minor appearance of apoptosis is hardly visible in the presented distributions.

CYTOFLUOROGRAPH System 30-L (Ortho Diagnostics Systems Inc., Westwood, MA, USA) using the UV lines (351-364 nm) of an argon ion laser for DAPI excitation. Small amounts of unspecific non-DNA fluorescence of DAPI was quenched by energy transfer mechanisms between DAPI and SR101. The DAPI emission was collected above 450 nm. Processing and cell cycle analysis of flow cytometric data was performed according to Dean & Jett (1974) and Stoehr *et al.* (1976) on a PC-based computer system Stoehr *et al.* (1991).

7.2.4 Fluorescence Properties of Cell Clones

Cells were seeded in 25 cm^2 cell culture flasks. After 24 h the cells were washed twice with Hank's solution (0.14 g/l $CaCl_2$, 0.4 g/l KCl, 0.06 g/l KH_2PO4, 0.1 g/l $MgCl_2$-$6H_2O$, 0.1 g/l $MgSO_4$-$7H_2O$, 8 g/l NaCl, 0.09 g/l Na_2HPO_4-$7H_2O$, 1 g/l D-Glucose), carefully dry-trypsinized for ~2 min and stopped with 3 ml phenol-red free RPMI 1640 with 10% FCS in an final concentration of 10^6 cells/ml. Agglomerates of even two cells were carefully avoided.

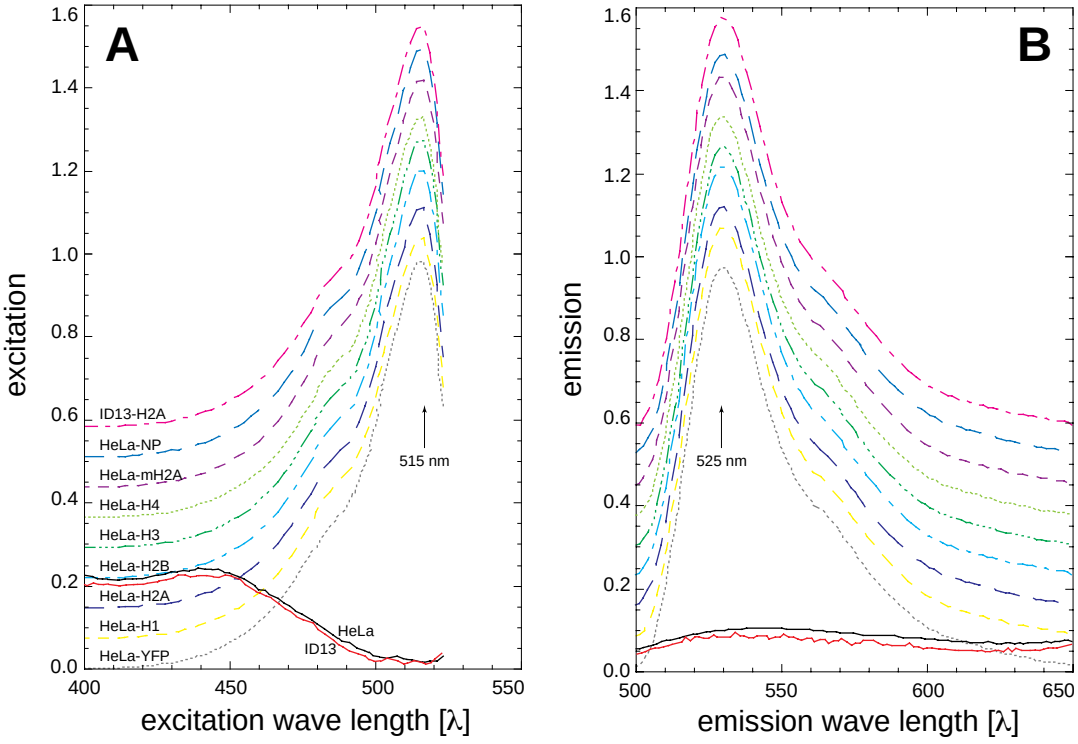


Fig. 7.7 Spectral Properties of YFPs in Histone-YFP Expressing Cell Lines
The normalized excitation (**A**) and emission (**B**) spectra of stable cell lines (HeLa or ID13) expressing either the pure YFP or the histone-YFP fusionprotein agree very closely with spectra obtained for YFP in solution (Clontech manual). The subtracted background was determined by HeLa and ID13 cell lines not expressing histone-YFP fusionproteins (shown magnified ten-fold). This implies that hardly any disturbance of fluorescence properties by interactions of the YFP exist, thus the functional organization of the cells seems uninfluenced. The mean fluorescence intensity of the cell lines provided by the amplitude of the unnormalized spectra are equivalent to FACS measurements (Fig. 7.8).

Spectral Properties of AFP in Cell Clones: Excitation and emission spectra were measured with an SLM-AMINCO 8100 fluorescence spectrometer (SLM, Urbana, IL, USA) using a 150W Xenon lamp. 500µl of the above prepared cell suspension were immediately measured in quartz cuvettes with 3mm path length at 20°C. Sedimentation was low enough to insure no influence on the spectra. Excitation spectra for eYFP were collected from 300 to 525nm at an emission wavelength of 530nm with a 4mm monochromator slit width for excitation and emission. Emission spectra were excited at 488nm and recorded from 500 to 650nm using the same monochromator setting. Finally, the spectra were background corrected, instrument corrected and normalized (Fig. 7.7, Tab. 7.3). From the absolute amplitude the approximate the nucleosomes fraction containing a histone-AFP per nucleus was calculated, calibrating with the extinction coefficient obtained by absorption spectra, the cell number determined (7.2.3), the interphase nucleosome number of ~1.75×10^7, regarding the cell cycle statistics obtained by FACS (7.2.3) and averaging over the cell cycle stages.

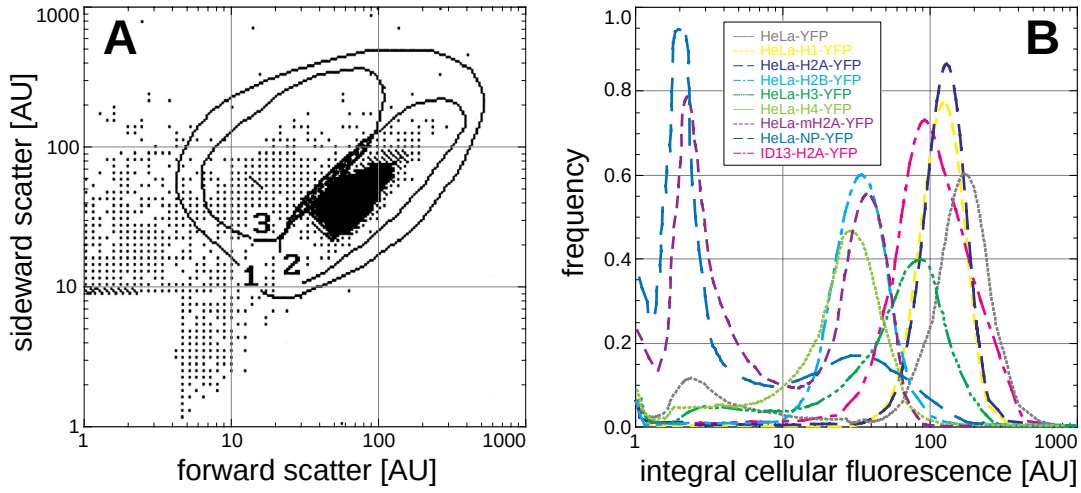


Fig. 7.8 Fluorescence Intensity Distribution of Histone-AFP Expressing Cell Lines
The forward and sideward scatter measured by fluorescence activated cell sorting (FACS) reveals cell populations very similar for all tested cell lines (**A**, HeLa control cell line). The fluorescence distribution of the integral cellular YFP fluorescence was determined from the total population 1 in (**A**) and varies between the different cell lines (**B**). The mean of the main maximum (> and integral of 20) is quantified in Tab. 7.3. HeLa-H3 and HeLa-H4 show also an additional distribution of low fluorescent cells. HeLa-mH2A and HeLa-NP lines were investigated biclonal as well as monoclonal achieved by further subcloning and characterised only by the second maximum of the fluorescence distribution (results of proliferation, cell cycle and spectral properties are based on the latter).

Fluorescence Intensity Distribution of Cell Clones: The fluorescence distribution of the cell clones were measured with the conventional flow cytometer Profile I (1983, Coulter Corporation, EPICS Division, Hialeah, FL, USA) exciting with the 488 nm line of an Argon ion laser at ~15 mW. The above prepared cell suspension was cooled on ice to avoid aggregation and softly mixed resulting in a homogeneous suspension before 100 µl was sucked into the cytometer. To restrict the fluorescence measurements to intact cells, the forward and sideward laser light scattering discriminated the intact cells from debris and other artefacts (Fig. 7.8, Tab. 7.3).

7.2.5 Confocal Laser Scanning Microscopy

Three-dimensional images were collected with an inverse Leica TCS SP (Leica) confocal laser scanning microscope (Fig. 7.9) equipped with 63x1.32 NA (theoretic $FWHM_{xy}$ = 148 nm, $FWHM_z$ = 286 nm) and 100x1.4 NA oil immersion PL APO objectives (theoretic $FWHM_{xy}$ = 139 nm, $FWHM_z$ = 236 nm), an Argon Krypton laser with at 458 nm, 476 nm and 488 nm, and Helium Neon laser with 543 nm, 633 nm excitation lines. eYFP constructs were excited with the 488 nm and/or the 514 nm line using a DD-458/514 or TD 488/543/633 beamsplitter. The emitted fluorescence was integrally collected with the spectrometer from 495 or 520 to 650 nm, respectively. The photomultiplier responding linearly from 450 V to 800 V, was

Tab. 7.3 Properties of Clones Expressing Histone-XFP Fusionproteins
For the quantitative characterization only fusionproteins with YFP were used: The proliferation was investigated by the growth exponent and the doubling time from the growth curve (Fig. 7.5) revealed only minor differences in comparison to the control cells. The same hold for the cell cycle phases and apoptosis fraction determined from the cell cycle phase distributions (Fig. 7.6). No spectral shifts of the autoflurescent fusionproteins were measurable in the excitation and emission spectra (Fig. 7.7). The clonality and the mean relative autofluorescent intensity (FL) of the integral cellular fluorescence distribution (main peak; *: autofluorescence of the control cells; Fig. 7.8) differed according to the selection purpose. The mean fraction of nucleosomes containing a fusionprotein determined by the absolute spectra agreed with the fluorescence distribution (mean of total cell line; brackets indicate the estimated mean fluorescence of the mean peak according to the FACS measurements).

Cell Line	Proliferation Rate		Cell Cycle Phase				Spectral Shift	FACS		Incorporation Rate ±2.0 [%]
	Exponent ±0.002	doubling time ±1.2 [h]	G1/G0 ±3 [%]	S ±2 [%]	G2/M ±2 [%]	Apoptosis ±0.5 [%]		Clonality	Mean FL ±6 [AU]	
HeLa	0.038	18.2	55.5	31.0	13.5	1.9	-	-	1.0*	-
HeLa-YFP (free)	0.038	18.2	58.0	24.9	15.2	2.3	no	poly	153.3	-
HeLa H1.0-YFP	0.038	18.2	53.9	31.2	13.9	2.4	no	mono	125.6	22.1
HeLa H2A-YFP	0.038	18.2	55.6	33.0	11.4	1.7	no	mono	129.2	23.4
HeLa H2B-YFP	0.041	16.9	54.2	28.7	17.1	2.5	no	mono	34.1	6.3
HeLa H3-YFP	0.034	20.3	49.4	28.5	22.0	2.0	no	bi	58.1	9.8
HeLa H4-YFP	0.040	17.3	57.7	26.2	16.1	2.1	no	bi	27.3	5.2
HeLa mH2A1.2	0.039	17.7	57.9	29.7	12.4	2.8	no	bi	41.4	7.3
HeLa NP	0.038	18.2	56.2	30.6	13.2	2.0	no	bi	33.5	5.7
ID13	0.036	19.2	57.3	29.9	12.8	2.1	-	-	1.0*	-
ID13 H2A-YFP	0.035	19.8	54.7	28.6	16.7	1.8	no	mono	118.2	19.2

operated at 700 V. The intensities were mapped to a 8 or 12 Bit range. The photomultiplier off-set was calibrated with the dark-current using a glow over/under lookuptable and set to -1. Images had a size of 512x512xZ pixels (Z: number of axial planes according to the cell thickness), with pixel sizes of 70x70x210 nm for images taken for the quantitative investigation of the chromatin distribution. The scan speed was 400 Hz and each line was averaged two to four times.

The coverslip with the cells (7.2.2) were placed in the ROC measurement chamber (LaCon Gbr, distributed through Leica) being closed with a glass-top and allowing air (especially gH_2O and CO_2) exchange and was put into a heating device (Heating Insert P, PeCon Gbr, distributed through Leica) which was placed on top of the microscope stage. The heating device was hold on 39 °C assuring the coverslip having a temperature ≥34 °C despite heat dissipation. On the heating device a gas incubator (Incubator S, PeCon Gbr) containing two water sinks was placed providing a humidified atmosphere with 5% CO_2 at 37 °C. The atmosphere was pumped through the incubator, being CO_2 enriched and heated to as well 39 °C to account for heat dissipation in an atmosphere controller (CTI-Controller 3700, PeCon Gbr) and humidified by the water sinks. To account for the evaporation of the water and to allow constant humidification the water was filled up automatically with a piston-pump (Labotron, Messtechnik GmbH, Gelting, Germany) with a flow of 300 µl/h.

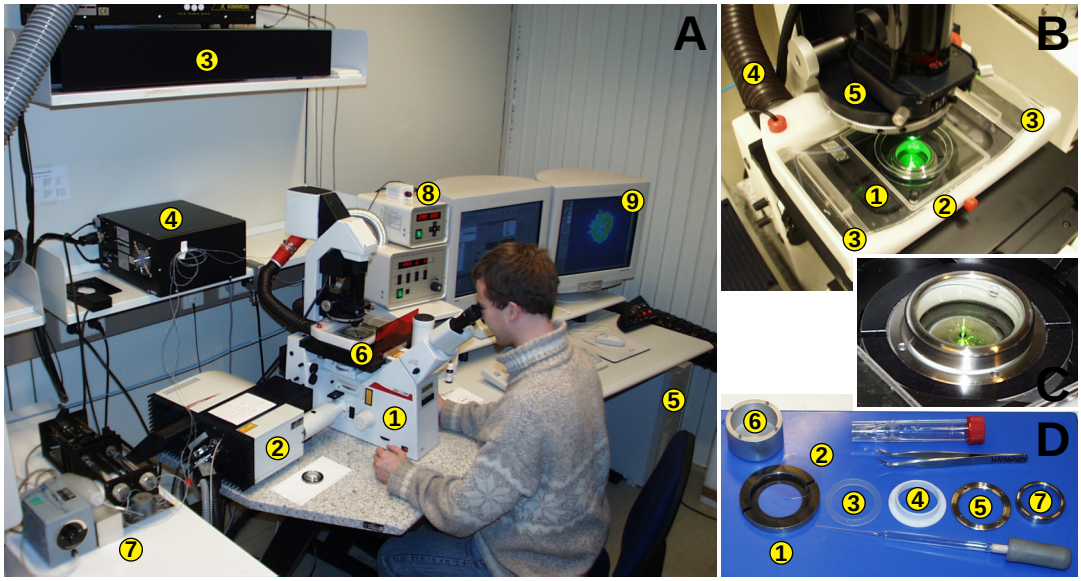


Fig. 7.9 Confocal Laser Scanning Microscope, Incubator and Measuring Chamber Set-Ups
The measurements described in Chapter 7 were conducted on a Leica TCS SP confocal scanning laser microscope consisting of an inverted microscope (**A.1**) with a condenser (**B.5**) to which the confocal scanning module (**A.2**) is attached containing the accustooptical modulators, the confocal optics, the grid spectrometer and the photomultipliers. The Argon Krypton laser (**A.3**; power supply: **A.4**) and the Helium Neon laser (under the table, invisible) are coupled into the scanning module via glass fiber optics. A Hg-fluorescence lamp is attached to the microscope from the back of the microscope (invisible, see green fluorescence in **B**, **D**). The whole microscope is controlled by hard ware modules (under the table) and by a PC (**A.5**). The measuring chamber (**C**) sits in a heating stage hold on 39˚C (**B.1**) enclosed by an incubator (**A.6**, **B.2**), consisting of two sinks (**B.3**), constantly supplied with water through a piston pump (**A.7**). Through the incubator a constant air of 37˚C flows (**B.4**) containing 5% CO_2 and saturated with gH_2O by the sinks. The temperature and CO_2 concentration is supervised by two controllers (**A.8**). The measuring chamber (**C**, **D**) consists of a chamber ring (**D.1**) into which the coverslip is placed (**D.2**) unto which a silicon seal (**D.3**) and a medium ring (**D.4**) is placed and secured by a screw ring (**D.5**) screwed by into the chamber ring (**D.1**) with a four point screwer (**D.6**). After filling 4ml of medium into the medium ring the measuring chamber is closed by a silicate glass top (**D.7**). For measurement the room is not lighted, leaving only the monitors as light source (**A.7**) on which currently a near to ready condensated metaphase is just imaged.

7.3 Quantitative Analysis Methods

To analyse the three-dimensional nuclear chromatin morphology imaged by confocal laser scanning microscopy first the nuclear volume and surface has to be segmented to allow accurate and artefactfree analysis. Then the (absolute) intensity and nucleosome distribution, the scaling behaviour by calculation of the (inverse-) mass using weighted box-counting-, lacunarity-, local fractal dimensions and the local diffuseness skewness and kurtosis as well as the scaling behaviour of the nuclear border were determined. For the latter, the detailed analysis description is given in Chapter 4 (4.2.2, 4.2.3, 4.2.4). These analyses are still in progress.

7.3.1 Nuclear Volume and Surface Segmentation, and Nuclear Roundness

To quantify the three-dimensional morphology correctly, the nuclear volume and the surface were segmented with a script within the image analysis software Heurisco. Therefore, a threshold dependent two-dimensional diffusion filter was applied separately to each axial image plane. The threshold was set to distinguish nuclear from random background intensities and to detect simultaneously the nuclear volume homogeneously. This fills up low intensity voids as the nucleolus, but leaves out surface invaginations. After determination of the nuclear pixels, an edge detection algorithm implemented within Heurisco determined the nuclear surface pixels. Finally, the information about pixels was stored in binary matrix with the same dimensions as the three-dimensional CLSM image stack. Consequently, the nuclear volume V_{Nuc} was obtained by $V_{Nuc} = N_{Nuc} \cdot v_{Pixel}$, with the number of pixels belonging to the nuclear volume N_{Nuc} and the volume of the pixel v_{Nuc} given by the pixel dimension. Accordingly, the nuclear surface area A_{Nuc} can be approximated by $A_{Nuc} = N_{Surf} \cdot r_{Pixel}$ with the number of pixels belonging to the surface N_{Surf} and the pixel dimension r_{Pixel}.

The most general and dimensionally reduced measure to characterise nuclei at the pixel resolution is given by the nuclear roundness R_{VS}

$$R_{VS} = 36\pi \frac{V_{Nuc}^2}{A_{Nuc}^3} \tag{7.1}$$

which is minimal for a sphere (similar to the chromosomal roundness (Equ. 3.2).

7.3.2 (Absolute) Intensity and Nucleosome Distribution of Nuclei

The absolute intensity and nucleosome distribution as function of the nucleosome concentration is the most straight forward morphologic property to analyse the three-dimensional organization of nuclei. It is also well suited for the confocal microscopy level of resolution. Due to the linearity of the photomultipliers and the CLSM image intensities ranging from 8 or 12 Bit, the nucleosome concentration can be calculated by relating the mean intensity to the mean concentration of the nucleosomes $\sim 1.74 \cdot 10^7 / V_{Nuc}$ for interphase nuclei. This assumes a constant relation between nucleosome number and volume and an equal incorporation rate of histone-AFPs into the nucleosomes. Variation of the histone-AFP production is of minor importance due to the calibration of each nucleus to the 8 or 12 Bit range. Consequently, the structure or structural changes in nuclei are mapped comparable even if the cell cycle phases and thus the volume and production rate change. Therefore, the intensity could also be transferred to a nucleosome concentration distribution. Its frequency provides the nuclear volume fraction at the given chromatin concentration and also the total number of nucleosomes at a particular concentration. The unnormality of this mass distribution could be tested by fitting to a bimodal Gauss function.

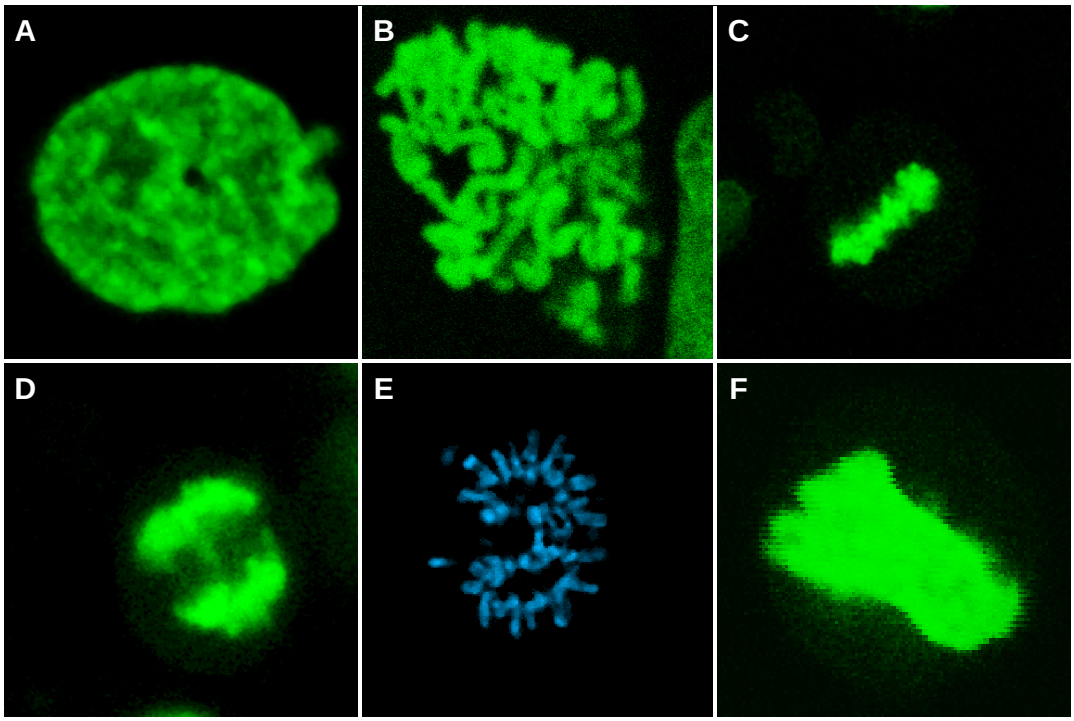


Fig. 7.10 Stages of Mitosis
The images demonstrate the stages of mitosis found in populations of HeLa and LCLC103H (only **E**) cells expressing H2A-YFP and H2A-CFP (only **E**). At the beginning of prophase the chromosomes start to condensate with apparent morphologic clustering (**A**) until full condensation into snakelike cylindrical chromosomes at the end of prophase (**B**), continuing with pairing of mother/daughter chromosomes usually indicated as prometaphase (not shown, because it seems not to be a separate state in the continuous flow from prophase to metaphase) until the metaphase plate is formed in metaphase (sideview, **C**, for a frontview, although in anaphase, **E**). Then the chromosomes are dragged apart from the metaphase plate in anaphase (sideview **D**, near the end of anaphase, **E**) into the daughter cells by spindle fibers already formed on the way from prophase to metaphase, and also organizing the chromosome pairing and the structure of the metaphase plate. After total separation of the mother into daughter cells the chromatin decondenses in telophase (**F**) showing more and more the morphology typical of interphase nuclei (Fig. 7.1). (image side length: **A**, **B**, **F** 25 µm, else 35 µm; **E** with courtesy of F. Bestvater, German Cancer Research Center (DKFZ), Heidelberg, Germany.)

7.4 *In vivo* Properties of Histone-AFP Expressing Cell Lines

To qualitatively and quantitatively investigate the three-dimensional organization and dynamics *in vivo* on the level of whole nuclei, fusion proteins of all histone classes with autofluorescent proteins (AFP) were created and expressed in various cell types (7.2.1): Since the spatial, volume and the mass of the nucleosome and the autofluorescent proteins indicate the possibility of steric hindrance or other influences on the chromatin fiber and gene regulation (Fig. 7.2), the autofluorescent protein was sequentially attached to the C-terminus of the histone, separated by a linker

(Tab. 7.1). The amino acid sequence of the linker was checked for free flexibility and adjusted such that the sequence of the multiple cloning site of the used basic vector could be used (Tab. 7.1&7.2). Both the highly flexible C-terminal histone tail and the linker allowed an estimated maximum distance of the autofluorescent protein to the nucleosome of 12nm with a mean of ~7nm. This still admits access of regulatory proteins to the histone tail. For expression of the protein the early SV40 promoter located in the HindIIIC fragment of the Simian Virus 40 (SV40) was used. Its expression pattern in relation to that of the natural histone genes within the cell cycle was the closest available and its overexpression properties are moderate. This resulted in the standard vector pSV-HIII-H_x-XFP built with most combinations of the histones H1, H2A, H2B, H3, H4 and mH2A1.2 and the autofluoresecent proteins eCFP, eGFP, eYFP, DsRed-1 and DsRed-2 (Tab. 7.1, Fig. 7.3A). The resulting fusionproteins could not only be expressed in the case of eCFP, eGFP, eYFP unproblematically but also resulted in an unusual high number of expressing cells of ~60% 24h after transfection. Fusionproteins with DsRed-1 and DsRed-2, however, led to a high rate of dead cells presumably due to DsRed tetramerization and consequent chromatin clumping. Therefore, these fusion proteins could only be recommended for transient use. For comparison of the histone expression by the SV40 promoter to the natural histone promoter, the bidirectional H2B and H2A gene complex with its bidirectional promoter (NP-I or NP-II) was inserted N-terminally to the autofluorescent protein. This resulted in the vector pH2B.x-NP-X-H2A.x-XFP (Fig. 7.3B). To allow later attachment of a XFP to the H2B, a multiple cloning site was integrated C-terminally to the H2B. Additionally, to the common vector pSV-HIII-H_x-XFP a so called His-tag, consisting of six histidin amino acids was introduced at the C-terminus of the autofluorescent protein, to allow purification of the histones by adsorbtion chromatography (Fig. 7.3C). For determination of effects due to the terminality of the autofluorescent protein also the general vector pCMV-XFP-H_x was used for mH2A1.2 (Fig 7.3D). Unfortunately, the promoter of the Cytomegaly Virus (CMV) leads to a much higher expression than the SV40 promoter region. The vector containing the natural promoter and the His-tag resulted in the same expression efficiencies, whereas the vector containing the CMV promoter led to only ~40% of expressing cells after 24h. According to the measurement and experimental purposes the cell lines were subcloned.

To characterise the general cell line properties quantitatively and to investigate influences of the histone-AFP fusionproteins or the pure AFP in comparison to the control cell lines, several parameters were determined for histone-YFPs: The proliferation rates characterized by the exponent of the growth curve and the doubling time resulted in values of ~0.038±0.002 and ~18.2±1.3h, respectively. Thus no significant difference between the YFP or histone-YFP expressing and the HeLa or ID13 controls seems to exist (Fig. 7.5, Tab. 7.3). The same held for the quantification of the cell cycle phase and apoptosis fraction resulting in 55.5±3, 31.0±2, 13.5±2 and 1.9±0.5% of cells being in G1/G0 phase, S phase, G2M phase and apoptosis, respectively (Fig. 7.6, Tab. 7.3). The normalized excitation and emission spec-

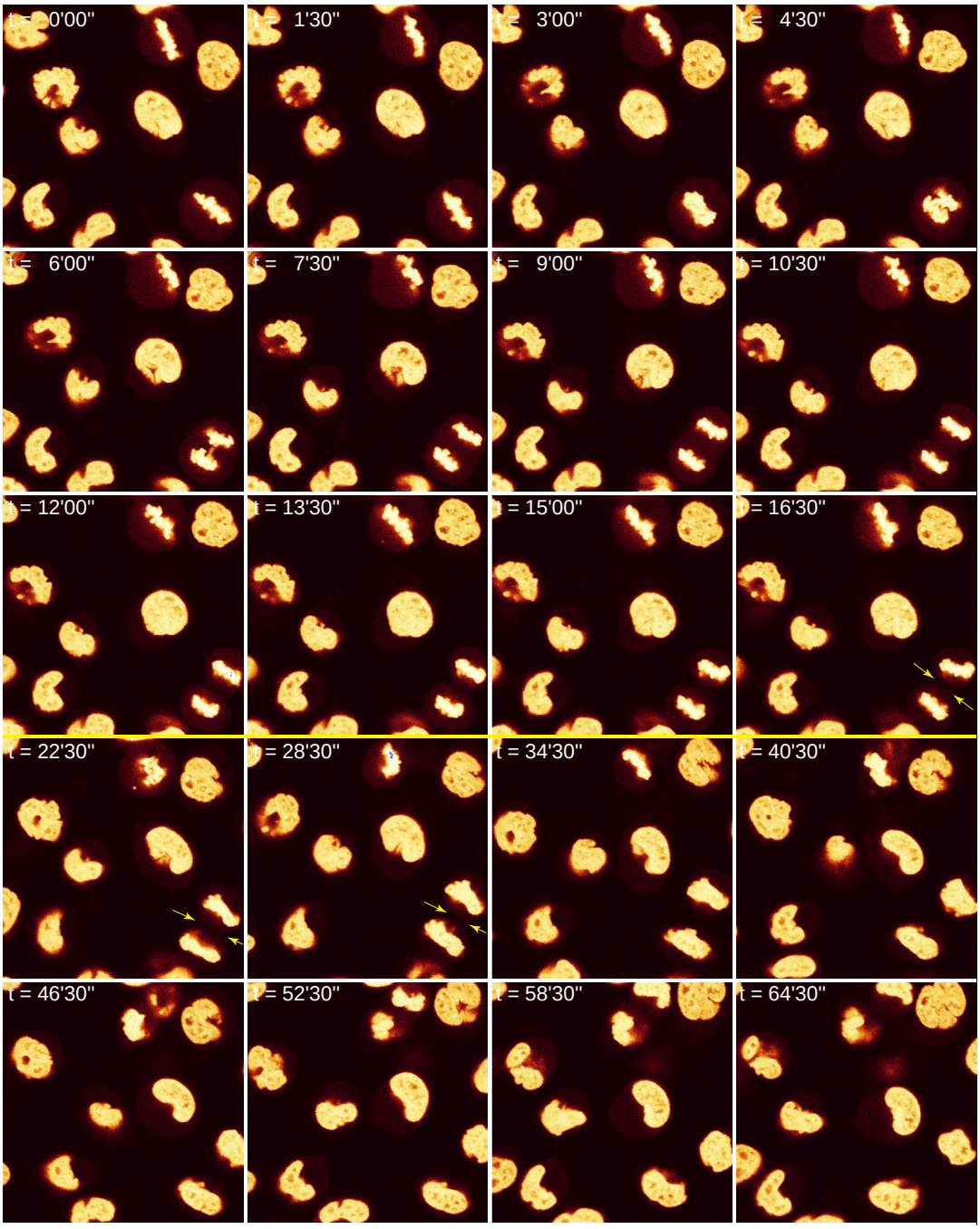


Fig. 7.11 Time Course of a Mitosis
The separation of metaphase plates takes only 6 to 9 min, whereas a whole mitosis from first condensation in prophase to full decondensation of chromosomes telophase within the daughter cells takes around 2 h (con- and decondensation usually need 40 to 50 min each). Images were slightly intensity oversaturated to show that a significant amount of histone-YFP is present in the cytoplasm, thus after separation of the metaphase plate, division of cells by cell membrane separation can be observed (arrows), taking usually 20 to 30 min. The strong impression always persists that daughter cells look alike concerning their decondensation morphology. (image side length: 100 µm)

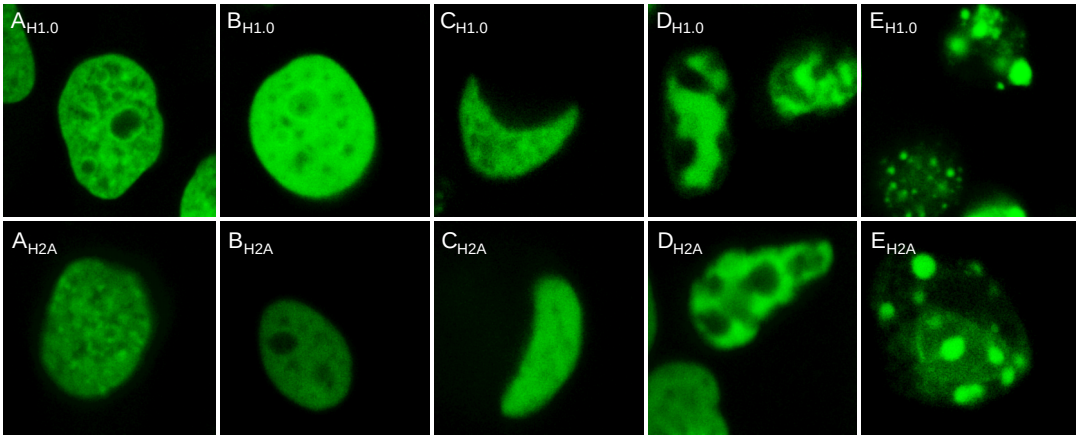


Fig. 7.12 Stages of Apoptosis Induced by Sodiumbutyrate
Images demonstrating the stages of apoptosis found at the same time in populations of Hela cells expressing H1.0-YFP or H2A-YFP due to incubation with deacetylation inhibitor Sodiumbutyrat. (**A**) Control nuclei without Sodiumbutyrat. (**B**) Nuclei showing lower granularity of the three-dimensional organization of the chromatin distribution than in (**A**) induced by Sodiumbutyrat (no signs of apoptosis are visible, see also Fig. 7.13). In the H2A image, the laser intensity was lowered to 400, the contrast to 2500 and the nuclei were bleached intensively to demonstrate that changes are due to structural changes and not only overproduction of H1.0-YFP. (**C**)"apoptotic half moon" shaped nuclei with starting agglomeration of chromatin. (**D**) Start of apoptotic condensation of chromatin in the nucleus. In the H1.0 image the upper right apoptotic nucleus in transition from apoptotic condensation to apoptotic fragmentation. (**E**) Apoptotic fragmentation of nuclei and final stage of apoptosis.

Unlike the cell culture (7.2.2) and microscope procedure (7.2.5) cells, were grown in Lab Tek 8-well chamber coverslips (Nalge Nunc International, Naperville, IL, USA). Cells were incubated with 6 mM Sodiumbutyrate for 12 or 36 h. Prior to microscopy the chamber slides were filled to the brim with $C0_2$ saturated medium also containing 6 mM Sodiumbutyrat. The brim was sealed with silicone paste and covered with a top, avoiding air leftovers and assuring that turning the chamber slide upside down no leackage appeared. Images were taken with an upright Zeiss 410 confocal microscope equipped with an oil immersion 63x1.32 NA PL APO objective and an Argon Krypton laser using the 488 nm line for excitation and a corresponding beamsplitter. The laser intensity was 450±10 V, the contrast 4500±100 and the pinhole was 17 (internal Zeiss units). Images had a size of 512x512xZ pixels (Z: number of axial planes), with a pixel size of 66.1x66.1x200 nm (A,B) and 49.6x49.6x 200 nm (C,D,E). The images were filtered with a 3*3 median filter.

tra of the histone-YFP cell lines match with cell lines expressing pure YFP or even isolated YFP in solution within ~1.5 nm (Fig. 7.7, Tab. 7.3). Therefore, major interactions of the autofluorescent protein with the nucleosome, the chromatin fiber or others could be excluded. Consequently, these general properties suggest no major influence on the cells of the histone-YFPs fusions, despite they might exist leading to more subtle effects. The single cell integral fluorescence distribution of the cell lines was measured with fluorescence activated cell sorting (FACS) and differed by a factor of ~4 between cell lines and ranged from mono, over bi, to poly clonality (according to the selection requirements for cell clones; Fig. 7.8, Tab. 7.3). The mean fluorescence intensity of clones is uncorrelated with any of the other properties, thus the histone-AFP expression although differing by a factor of ~4 seems not to influence the general cellular properties (Tab. 7.3). The absolute average fraction

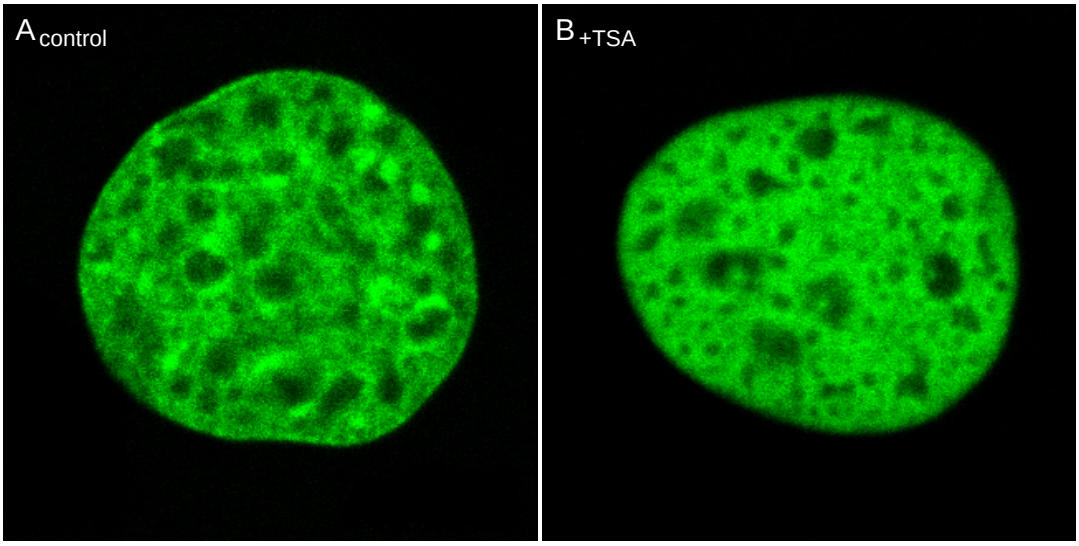


Fig. 7.13 Structural Chromatin Changes Induced by the Deacetylase Inhibitor Trichostatin-A (TSA) The three-dimensional structure of the chromatin distribution changes from a course grained (**A**) to a more isotrope distribution (**B**) after adding TSA in nuclei of ID13 cells expressing H2A-YFP. The intensity distribution was normalized, thus the change is not due to higher expression in agreement to experiments using Sodiumbutyrat (Fig 7.12). (sidelength 20 µm; with courtesy of K. Rippe, Division Physics of Molecular Biological Processes, Kirchhof Institute for Physics, Heidelberg, Germany).

of labelled nucleosomes at least containing one histone-AFP molecule was calculated from the absolute amplitude of the spectra calibrated by the extinction coefficient in an absorption spectrum. This resulted in incorporation rates of 5 to 23% differing by a factor of ~4 in agreement with the FACS measurements. Meanwhile, for the HeLa-H2B-YFP cell line the incorporation rate into nucleosomes was also determined by biochemical methods combined with fluorescence fluctuation microscopy (FFM) and resulted in 5.0±1% (Weidemann *et al.*). Here also a prolongation of the nucleosomal repeatlength from 185±10 to 204±3bp was found. This might lead to a more open chromatin conformation. In summary, the quantitative characterization of the cell lines revealed that the expression of the histone-AFP fusionproteins has no major adverse effects on the cells in comparison to the control cell lines.

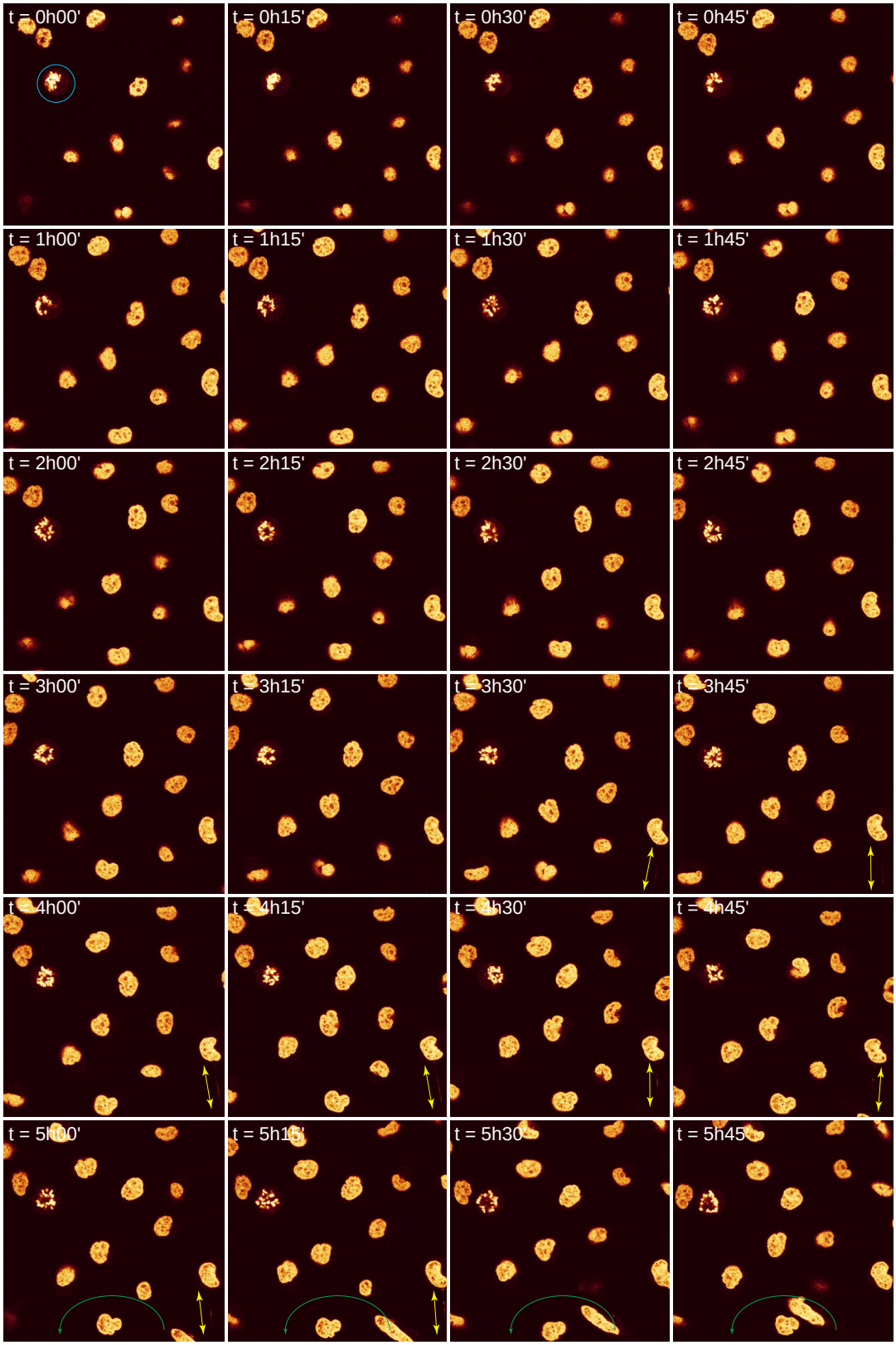
7.5 Qualitative Chromatin Morphology of Cell Nuclei *in vivo*

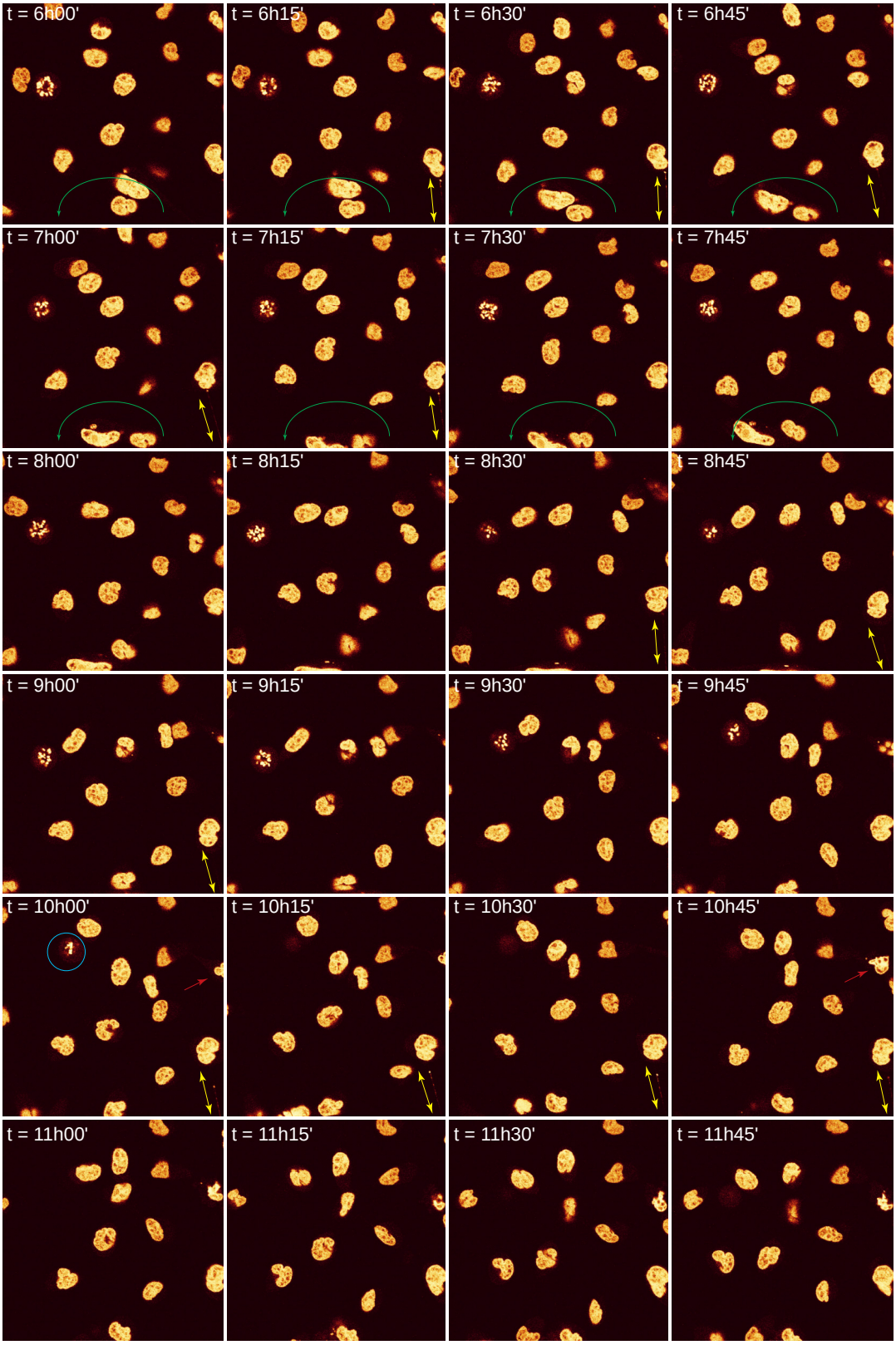
The nuclear morphology of cells expressing histone-AFP fusionproteins was assessed during interphase, mitosis and apoptosis to investigate the overall possibilities of this method of chromatin labelling. To judge the cell cycle and health of cells is also important during microscopy and for quantification of the chromatin distribution. For microscopy always a confocal laser scanning microscope (CLSM) was used with conditions guaranteeing the health of the cells (7.2.5; Fig. 7.12).

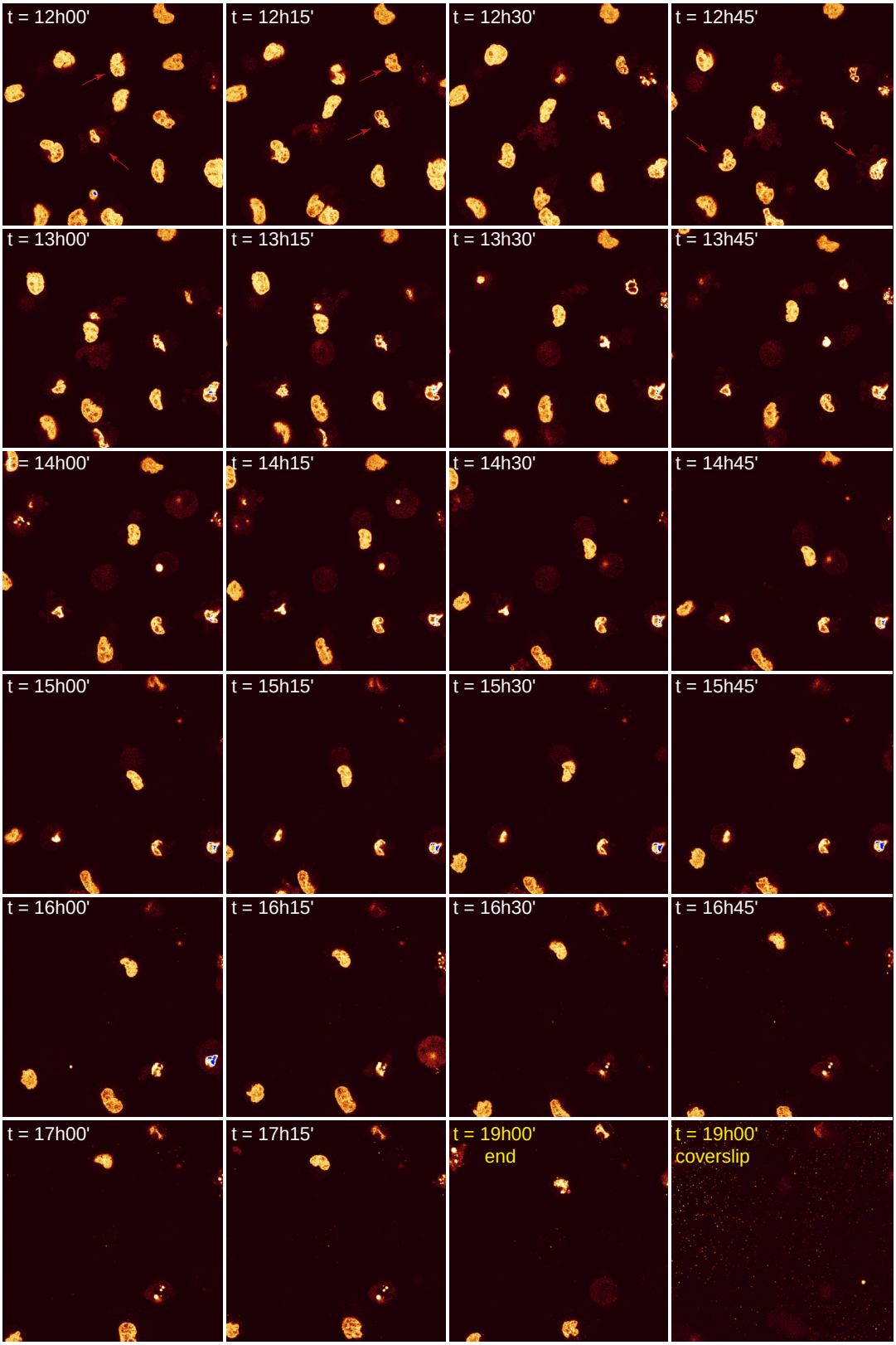
7.5.1 Interphase

The histone-AFP fusionproteins are localized mainly in the cell nucleus (Fig. 7.1, Fig. 7.10-14, Fig. 7.16 and Fig. 8.1) with hardly any fusionproteins in the cytoplasm. This is due to their rapid and active transport into the nucleus after syntheses mediated by the nuclear localization sequence (NLS) of the histones and the low production rate (Baake *et al.*, 2001). Thus, the nuclear membrane separating the nucleo- from the cytoplasm is visible in ~80 to 85% of the cells adding G1/G0 and S phase. Within the nucleus the chromatin is not distributed homogeneously in contrast to the near to total isotropic distribution of free YFP: Globular spots/blobs/foci/granules and lacunas of various sizes dominate. The nucleoli are surrounded by a band of structured dense chromatin as near the nuclear membrane. While the histones H1, H2A, H2B, H3 and H4 all produce the same pattern, there are differences in the chromatin distribution between different cell types: Within the human HeLa cells the blobs have sizes of ~250 to 700 nm with similar sized lacunas. Both have a big size variance and a tendency to clump. The mouse ID13 cells show a much more isotropic morphology with a few highly dens spots and circular voids. This morphology is often denoted as 'pizza-nuclei'. The primate Cos7 cells show a very isotropic morphology of separated globular structures with sizes similar to those of HeLa cells. These morphologies are in agreement with fluorescence *in situ* hybridization (Fig. 1.12, Fig. 1.13B), the *in vivo* replication labelling or staining with DAPI or Hoechst33342. However, the first is based on harsh preparations, the second is only temporary and the last is based on intercalation of fluorophores into the DNA leading to rapid cell death. Not only are these techniques highly artefact prone, but also show the morphology less clear than the histone-AFP fusionproteins. The histone-YFP interphase morphology agrees also with the prediction of a Multi-Loop-Subcompartment (MLS) like organization of the 30 nm chromatin fiber (Fig. 1.14, 1.3.7, Chapter 2&3, Fig. 3.2, Fig. 4.6&7). Additionally, the nuclear

Fig. 7.14 Time Course of Apoptosis Induced by Deoxyglucose
4 mg/ml Deoxyglucose was added to the medium of the already calibrated cells in the microscope chamber at 0 h. Image stacks were taken every 15 min totalling 19 h. Shown are image planes 8 µm above the coverslip for better vision as cells get rounder and raise away from the coverslip during apoptosis (image sidelength 200 µm). During the experiment no expression increase of H2A-YFP was observed, thus changes were only due to apoptosis. First signs of apoptosis with form changes of the nuclear membrane starts at around 7 to 9 h. "Open" apoptosis sets in after 9 to 11 h (red arrows) in agreement with the literature and leaves hardly any cell in the end stage of apoptosis after 19 h. The last image plane (directly above the coverslip) shows the massive appearance of the chromatin/cell debris after apoptotic fragmentation. A comparison to non fluorescent HeLa cells directly after the time course revealed that the fluorescence is not due to autofluorescence of proteins etc., but due to YFP. Consequently the YFP is not or only partially degraded by proteases etc.. Before the onset of apoptosis a HeLa-H2A-YFP cell marching over several micrometers in about 3 h was observed (green pathway). This is astonishing because Deoxyglucose blocks the glucose and thus energy turnover. In contrast a cell in metaphase did not proceed with mitosis for 10 h which could be due to the activation of mitosis checkpoints before turning on G2-apoptosis. Sometimes dividing cells are still connected by a channel connecting the nuclei which holds for about 11 h (yellow double arrow).


t = 0h00'
t = 0h15'
t = 0h30'
t = 0h45'
t = 1h00'
t = 1h15'
t = 1h30'
t = 1h45'
t = 2h00'
t = 2h15'
t = 2h30'
t = 2h45'
t = 3h00'
t = 3h15'
t = 3h30'
t = 3h45'
t = 4h00'
t = 4h15'
t = 4h30'
t = 4h45'
t = 5h00'
t = 5h15'
t = 5h30'
t = 5h45'


t = 6h00'
t = 6h15'
t = 6h30'
t = 6h45'
t = 7h00'
t = 7h15'
t = 7h30'
t = 7h45'
t = 8h00'
t = 8h15'
t = 8h30'
t = 8h45'
t = 9h00'
t = 9h15'
t = 9h30'
t = 9h45'
t = 10h00'
t = 10h15'
t = 10h30'
t = 10h45'
t = 11h00'
t = 11h15'
t = 11h30'
t = 11h45'


t = 12h00'
t = 12h15'
t = 12h30'
t = 12h45'
t = 13h00'
t = 13h15'
t = 13h30'
t = 13h45'
t = 14h00'
t = 14h15'
t = 14h30'
t = 14h45'
t = 15h00'
t = 15h15'
t = 15h30'
t = 15h45'
t = 16h00'
t = 16h15'
t = 16h30'
t = 16h45'
t = 17h00'
t = 17h15'
t = 19h00'
end
t = 19h00'
coverslip

shape differs greatly: HeLa cells often have very odd shaped nuclei with many invaginations, in contrast to the round ID13 cells or Cos7 cells, although the latter are not totally round. The expression of the histones by the natural promoter led to the same intra-nuclear morphologies. Here a fraction of cells comparable to the number of cells in S phase showed considerable more fusionproteins in the cytoplasm than by expression with the SV40 promoter. Time courses through the whole cell cycle indicated indeed a link to S phase and replication.

The histone macroH2A1.2 (mH2A1.2) should be located in the inactive X chromosome (Pehrson *et al.*, 1992; Vijay-Kumar *et al.*, 1995; Pehrson *et al.*, 1997; Constanzi & Pehrson, 1998; Pehrson *et al.*, 1998; Lee *et al.*, 1998; Csankovszki *et al.*, 1999; Mermoud *et al.*, 1999; Rasmussen *et al.*, 1999; Rasmussen *et al.*, 2000), thus the mH2A1.2-AFP would be a possible *in vivo* label for a single chromosome. The 42kDa heavy mH2A1.2 consists of three parts containing a region with 50% sequential but higher structural similarity to the usual H2A, a region to the C-terminal end with 57% similar to H1 and a region common for DNA binding zink-finger proteins. The latter two might play a role in the inactivation of the X chromosome. mH2A1.2-YFP shows a more granular morphology but localizes as H2A-YFP. Unfortunately, no apparent incorporation into the X chromosome of the female HeLa cells exists. To exclude steric influences of the C-terminal AFP location, a GFP-mH2A1.2 fusionprotein with N-terminal GFP was constructed whose expression led always to rapid cell death. Consequently, the function of mH2A1.2-YFP might be blocked by the AFP, while GFP-mH2A1.2 seems functional but presumably inactivated the total genome due to its overexpression.

7.5.2 Mitosis

To allow cell division the chromosomes are condensed from their territory like distribution into well transportable cylinders before destruction of the nuclear membrane. Already the first stages of condensation at the beginning of prophase are visible as a more condensed chromatin morphology (Fig. 7.10A), before the full condensation into snakelike cylinders at the end of prophase (Fig. 7.10B). Thereafter, the pairing of chromosomes and formation of the metaphase plate, in which the chromosomes are radially arranged, were followed in detail (Fig. 7.10C). Most of the fusionprotein is located in the chromosomes suggesting incorporation into the nucleosomes in agreement with bleaching experiments in metaphase (Weidemann, 2002). The chromosomes are separated and dragged apart in anaphase (Fig. 7.10D&E), thus the chromosome arrangement can be very well followed and single chromosomes could be identified by their length and/or volume. The total separation into the two daughter cells is visible by the cytoplasmic histone-AFP distribution increased by the release of unincorporated fusionproteins during nuclear membrane degradation and due to higher expression in S-phase (Fig. 7.11, arrows). Thereafter, the chromatin decondenses again, while the nuclear membrane is build around the chromosomes in telophase (Fig. 7.10F). Over time the nuclei show more

and more the typical interphase morphology. Time courses (Fig. 7.11) revealed that a whole mitosis takes ~2h with the separation of the metaphase plate taking only 6 to 9min. This leaves ~40 to 50min for condensation and decondensation, excluding the resting time needed for checkpoint inquiry by the cells before mitosis continuation. Decreasing the temperature or CO_2 concentration etc. led to slower path through mitosis. This indicates the adequacy of the measuring conditions and is in turn a standard.

7.5.3 Apoptosis

Apoptosis was induced in the cells with the deacetylase inhibitor Sodiumbutyrat (Ng & Bird, 2000; Siavoshian *et al.*, 2000; Marks *et al.*, 2000) or the glucose cycle inhibitor Deoxyglucose starving the cells. Incubation with Sodiumbutyrat leads to neutralization of the positive charged histone tails resulting in decondensation of the chromatin fiber. Accordingly, in Hela cells first the histone-AFP expression was increased due to upregulation of gene activity. Then the interphase chromatin distribution was homogenised, which was proven by normalization of the intensity distributions or extensive bleaching (Fig. 7.12A&B). Thus, Sodiumbutyrate leads to a more open chromatin conformation and to visible changes in the three-dimensional organization of cell nuclei. The same behaviour was found for the deacetylase inhibitor Trichostatin-A (TSA) in HeLa and ID13 cells (Fig. 7.13). After ~10h most nuclei change from a round to a half-moon shape, the so called “apoptotic half moon” (Fig. 7.12C), before the chromatin is digested and agglomerated into so called apoptotic bodies (Fig. 7.12D). Thereafter, the nuclei and the whole cells are fragmented and totally destructed (Fig. 7.12E). To investigate apoptosis without an initial chromatin change the cells were treated with Deoxyglucose resulting in the same apoptosis course as above, despite lower number of cells showing apoptotic half moons (Fig. 7.14). Apoptosis could also be distinguished from pure salt or pH death. Here usually the nuclear membrane folded and invaginated, before the cell implodes or bursts depending of the direction of the condition changes.

7.6 Quantitative Chromatin Morphology of Cell Nuclei *in vivo*

For the quantitative analyses of the three-dimensional organization of interphase nuclei, scaling analyses as shown in Chapter 4 for simulated confocal image stacks are especially suited. Prior to the scaling analyses, the nuclei in experimental image stacks have to be segmented to avoid artefacts from the background (7.3.1) in contrast to the simulated image stacks where the nuclear volume and surface are a priori known. Then, the absolute and normalized nucleosome density distributions (7.3.2), the weighted scaling analysis (assuring the equal treatment of regions in the middle

or at the membrane of nuclei, Chapter 4) and the scaling of the nuclear membrane and thus the general nuclear shape (Chapter 4) could be investigated. Since in a cell population a variety of cell phases according to the cell cycle and cells with different health status are present, ~100 interphase nuclei need to be measured to achieve results of statistical significance. Interphase nuclei were chosen by the general adherence and stretching behaviour of the whole cell and the nuclear morphology. The cells are currently measured for each of the cell lines expressing the histone-YFP fusionproteins H1, H2A, H2B, H3, H4 and mH2A1.2 in female human HeLa cells and H2A-YFP fusionproteins in human LCLC102, primate Cos7, and mouse ID13 cells. To show that the ongoing measurements are analysable in agreement with simulated nuclei in Chapter 3&4 by the (absolute) and normalized nucleosome distribution, the scaling behaviour by calculation of the (inverse-) mass using weighted box-counting-, lacunarity-, local fractal dimensions and the local diffuseness, skewness or kurtosis as well as the scaling behaviour of the nuclear border, a first test-result is presented here:

Two nuclei of Cos7 cells expressing H1-YFP with a usual (Fig. 7.15Bβ) and more condensed (Fig. 7.15Bα) morphology and thus apparently different, were segmented. The mean chromatin densities were 123µm and 167µm. The mass distributions were bimodal with fractions of 18:82% and 3:97%. Both results agree with absolute chromatin density measurements (Weidemann *et al.*, in submittance). The roundness R_{VS} was 0.74 and 0.62. The weighted volume function $V_B(l_B)$ as well as the weighted box-counting dimensions D_{Bw} were calculated as function of an intensity threshold according to Chapter 4. The weighted box-volume function $V_B(l_B)$ of the mass showed distinct power-law behaviour for each intensity threshold before reaching a cut-off with similar behaviour for all thresholds (Fig. 7.15A&B). The cut-off indicates the transition where whole nuclei are seen more and more as a point like object for large side length of the measuring box. Consequently, the determination of the weighted box-counting dimension D_{Bw} by a linear regression is not only justified but also shows as function of the threshold a behaviour distinguishing between both nuclei (Fig. 7.15C): the more homogeneous nucleus has a higher initial plateau with a D_{Bw} of ~2.6, with an early and steep decent the higher the threshold in contrast to the more condensed nucleus with an initial value of ~2.4 slightly decrease before a less steep decent. In summary, the mean chromatin density, the bimodality of this distribution, the roundness and the weighted box-counting dimension could not only be determined but also distinguish between nuclei with a different three-dimensional chromatin organization.

7.7 Discussion of “Chromatin Alive“ and Future Aspects

The morphology and dynamics of chromatin in the cell nucleus has been debated since the first microscopic studies (Chapter 1). The qualitative and quantitative anal-

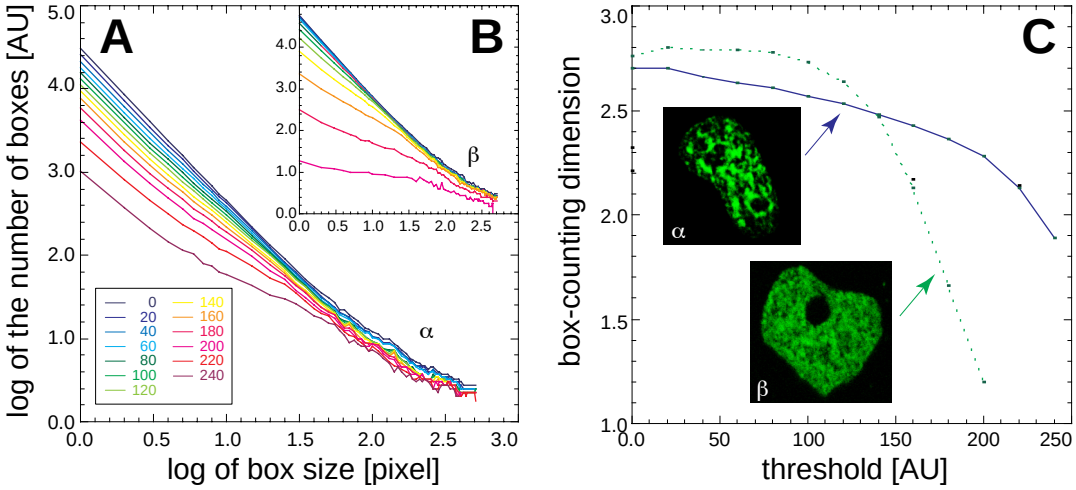


Fig. 7.15 Volume and Surface Distributions of H2A-YFP Expressing Cell Lines
The weighted box-volume function $V_B(l_B)$ shows distinct power-law behaviour for two nuclei for each intensity threshold before a cut-off is reached with similar behaviour for all thresholds (**A**, **B**). The box-counting of more granular and condensed nuclei (**α**) falls-off more rapidly than the more homogeneous nuclei (**β**) as function of the threshold. Thus, calculation of the box-counting dimension D_{Bw} as function of the intensity threshold differs also between the two nuclei (**C**): the more homogeneous nucleus (**β**) shows a higher initial plateau with a D_{Bw} of ~2.6, with an earlier and steeper decent in contrast to the more condensed nuclei (**α**) showing from the beginning a slight before the steep decent in agreement with the general theoretic expectation and Chapter 4.

yses were either compromised by the used fixation protocols in electron microscopy and fluorescence *in situ* hybridization (FISH) or the artefacts by *in vivo* staining with DNA intercalating fluorophores (e. g. Feulgen, ethidium bromide, 4,6-Diamidino-2-phenylindole-2HCl (DAPI) or the bisbenzimide Hoechst33342 ($C_{27}H_{29}N_6O$); Belmont *et al.*, 1984). These methods preclude the investigation of the dynamics and often result in distortion of the nuclear morphology and rapid cell death. The recently developed *in vivo* labelling during replication by incorporation of e. g. BrdU leads only to a temporal label due to its rapid degradation (Zink *et al.*, 1998).

To overcome these limitations for an *in vivo* investigation of the three-dimensional structure and dynamics of the chromatin distribution a novel technique was established: DNA vectors encoding the fusion proteins between all histones H1.0, H2A, mH2A1.2, H2B, H3, H4 and the autofluorescent proteins CFP, GFP, YFP, DsRed1 DsRed2 were constructed. Their expression in HeLa, LCLC103H, Cos7 and ID13 cells and subsequent cloning have led to stable cell lines with unchanged properties for over 4 years. In these cell lines between 2.6 and ~20% of the nucleosomes have incorporated a fluorescent fusion protein. Due to active histone transport >>80% of the labelled histones are located in the nucleus. No apparent influence of fusionproteins on the cell cycle status, the proliferation rate or the AFP fluorescent excitation/emission spectra but an increase in the nucleosomal repeat length (Weidemann *et al.*, in submittance) was found. With this approach the structure and dynamics of histones, nucleosomes, chromatin, chromosomes and whole nuclei to cell

cycle, differentiation, apoptosis or general gene regulation as well as drug influences could be investigated *in vivo*. The interphase morphology showed globular structures as predicted by a Multi-Loop-Subcompartment (MLS) like model for the three-dimensional organization of the 30 nm chromatin fiber. All stages of mitosis as well as the single chromosomes are clearly distinguishable. Induction of apoptosis by the glucose cycle blocker Deoxyglucose or the deacetylase inhibitors Sodiumbutyrat and Trichostatin A revealed also all stages of apoptosis. Additionally, the deacetylase inhibitors first led to a smoothing of the interphase morphology, due to a change of the chromatin conformation, before the apoptotic nuclear breakdown.

In Chapter 3 and Chapter 4 it was theoretically shown that such apparent differences in the chromatin morphology could be investigated by scaling analysis of the (inverse-, iso-) mass distribution by the weighted box-counting, lacunarity and local fractal dimensions as well as by statistical properties as the nucleosome density, the local diffuseness, local skewness and local kurtosis distributions. Here it was shown that with the *in vivo* chromatin label the interphase morphology and changes thereof could be investigated succesfully with these analyses methods.

Beyond the above described histone-AFP fusionproteins for the labelling of the chromatin distribution *in vivo* were already applied for various purposes (Kanda *et al.*, 1998; Lever *et al.*, 2000; Misteli *et al.*, 2000; Phair & Misteli, 2000; Dey *et al.*, 2000; Monier *et al.*, 2000; Perche *et al.*, 2000; Sadoni *et al.*, 2001; Kimura &Cook, 2001; Bestvater *et al.*, 2002; Weidemann *et al.*, in submittance) and provide a lot of opportunities whose principle feasibility seems obvious combining the results shown with state of the art techniques:

7.7.1 *In vivo* Method for Co-Localization and Dynamics

The *in vivo* labelling of chromatin could also be used for colocalization experiments combined with other fusionproteins involving autofluorescent proteins (AFP) or other fluorescent tags. The clear visibility of the border of the nuclear membrane recommends the histone-AFP fusion proteins also as a general nuclear marker (Chapter 8; Fig. 8.1). The method is superior over conterstaining of DNA with e. g. DAPI or Hoechst33342, due to their fuzzy signal distribution and most important their toxicity. The latter does not allow continuous experiments on the same defined cells. It should be noted that simultaneous double transfection of two or more fusion proteins using AFPs need to be avoided due to conversion of their fluorescence properties as described in chapter 8 “GFP-Walking”. Of course, histone-AFP expressing cells could also be used in all experiments involving isotonic and non-dehydrative fixation of cells being common in immunolabelling or in fluorescence in situ hybridization (FISH) experiments. Here the chromatin counterstain needs not to be applied separately avoiding various staining artefacts. Histone-AFPs also allow the investigation of the histone or of the chromatin dynamics in the nucleus as described for mitosis and apoptosis. Production or transportation of histones could be blocked and the influences could be analysed (A. Alonso, German Cancer

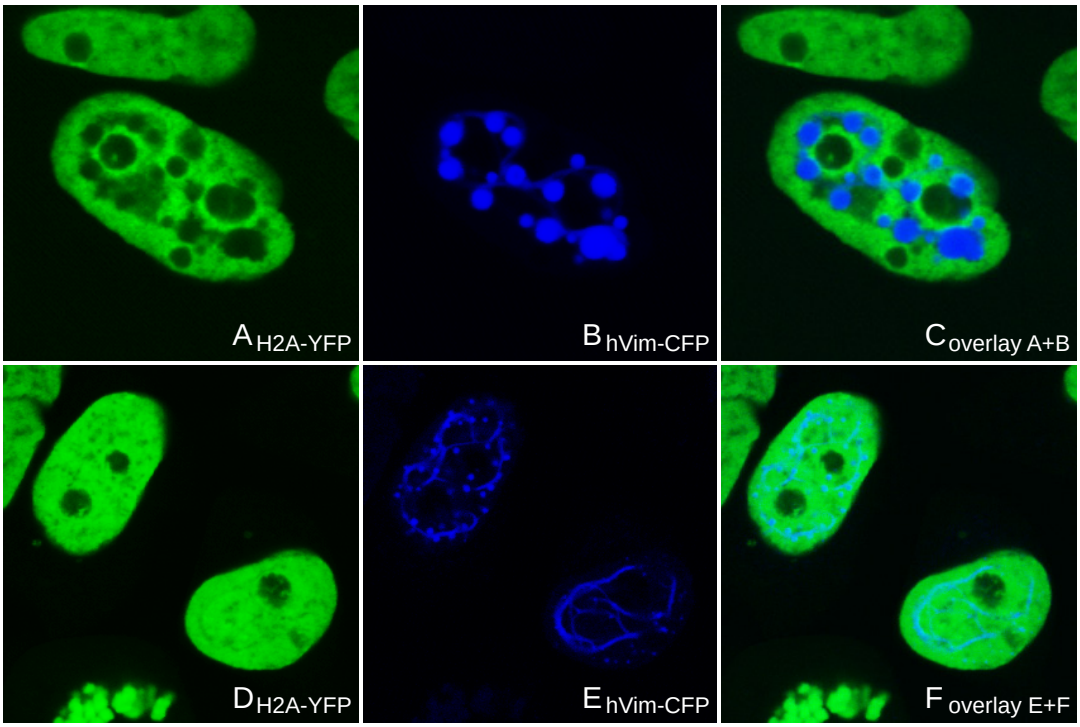


Fig. 7.16 Colocalization of Human Vimentin hV-CFP against Chromatin H2A-YFP
Human Vimentin (hV) does exclude and does not colocalize with chromatin in living HeLa cells stably expressing H2A-YFP and overtransfected with pure hV and hV-CFP (relation 10:1). The voids left by the hV-CFP (B, E) are clearly visible in the images of the H2A-YFP (A, D). A close investigation of the overlay of both CLSM channels reveals an empty space between hV and chromatin which is due to entropic exclusion by the faster movement of hV against chromatin. The mostly globular distribution of hV (B) can be transferred to the fibrous state with small aggregation globules (E) by changing the temperature. (The CLSM images were taken according to 7.2.2 and 7.2.5; with courtesy of M. Reichenzeller, Division Cellular Biology, German Cancer Research Center (DKFZ), Heidelberg, Germany).

Research Center (DKFZ), unpublished). The movement of the whole nucleus or substructures of the chromatin organization as well as in relation to colocalized proteins could also be determined *in vivo*: E. g. the colocalization and dynamics of vimentin-YFP in relation to the chromatin distribution and dynamics in nuclei (Fig. 7.16; Reichenzeller, unpublished; Reichenzeller *et al.*, 2000) or the localization and dynamics of proteases during apoptosis. Another example is the investigation of the relationship between the three-dimensional organization of the nucleus and its dynamics by measuring with fluorescence fluctuation microscopy (FFM) the local obstruction of diffusing particles as function of the local chromatin distribution (Wachsmuth *et al.*, 2000; Wachsmuth, 2001).

7.7.2 Improvement of Apoptosis Analysis

Beyond the pure description of the nuclear appearance, the use in the control of cell culture and easy screening during experiments of apoptosis by visual inspection, histone-AFP fusion proteins could also facilitate classical apoptosis tests based on the release of histones, measured e. g. by fluorescence activated cell sorting (FACS), enzyme-linked immunosorbant assays (ELISA) or similar methods:

Apoptosis analysis by FACS usually measures the DNA content in nuclei left after fixation, marking the DNA with DAPI, permeabelization and washing out the small and labelled DNA fragments degraded by nucleases during apoptosis (7.2.3). Appropriate fixation methods allow now the measurement of the histone or chromatin amount left in the nuclei measurable e. g. by fluorescence correlation spectroscopy (FCS) or spatially and intensity resolved planeometric microscopy (SIRPM, Chapter 8). The free histone amount could possibly be measured with fluorescence correlation spectroscopy in the cell culture medium or directly in nuclei, although no structural changes by apoptosis are yet clearly visible.

ELISA tests on apoptosis measure the amount of free histones by binding them immunochemically onto the bottom of well-plates. The bound histones are detected by immunolabelling with antibodies carrying the enzymatic ability to catalyse a reaction changing the fluorescence properties of a fluorophore and thus amplifying the initial signal. The detected final fluorescence is proportional to the amount to free histones, which is proportional to the degree of apoptosis. Using histone-AFP fusion proteins only needs eventual binding to a matrix, whereas the signal amplification can be avoided by the corresponding optical detection systems, e. g. microscopy or fluorescence correlation spectroscopy (FCS).

7.7.3 Specific Labeling and Specific Isolation of Histones

The specific labelling and isolation is another advantage of the histone-AFP fusion proteins. The labelling specificity is reached by the defined position of the AFP as part of the amino acid sequence at the histone termini. Isolation can be performed with standard protocols of protein purification and could be facilitated by insertation of a sequence of 6 histidin amino acids (so called HIS-Tag or with variation in the sequence HAT-Tag) allowing purification of proteins by affinity chromatography. Current labelling techniques of isolated histones use e. g. NH_2 sidegroups of the amino acid lysin for tagging with succinimidylester conjugated fluorescence dyes followed by purification procedures. Therefore, not only the labelling process could be omitted but also the purification process is reduced. The specific labelled and isolated histones could be used in all applications wasting conventionally labelled histones. Such experiments include *in vitro* and *in vivo* studies about the production, transport, localization and degradation of histones as well as their modifications by enzymes and their involvement in gene regulation. Of special interest here are commercially available kits based on the use and detection of histones.

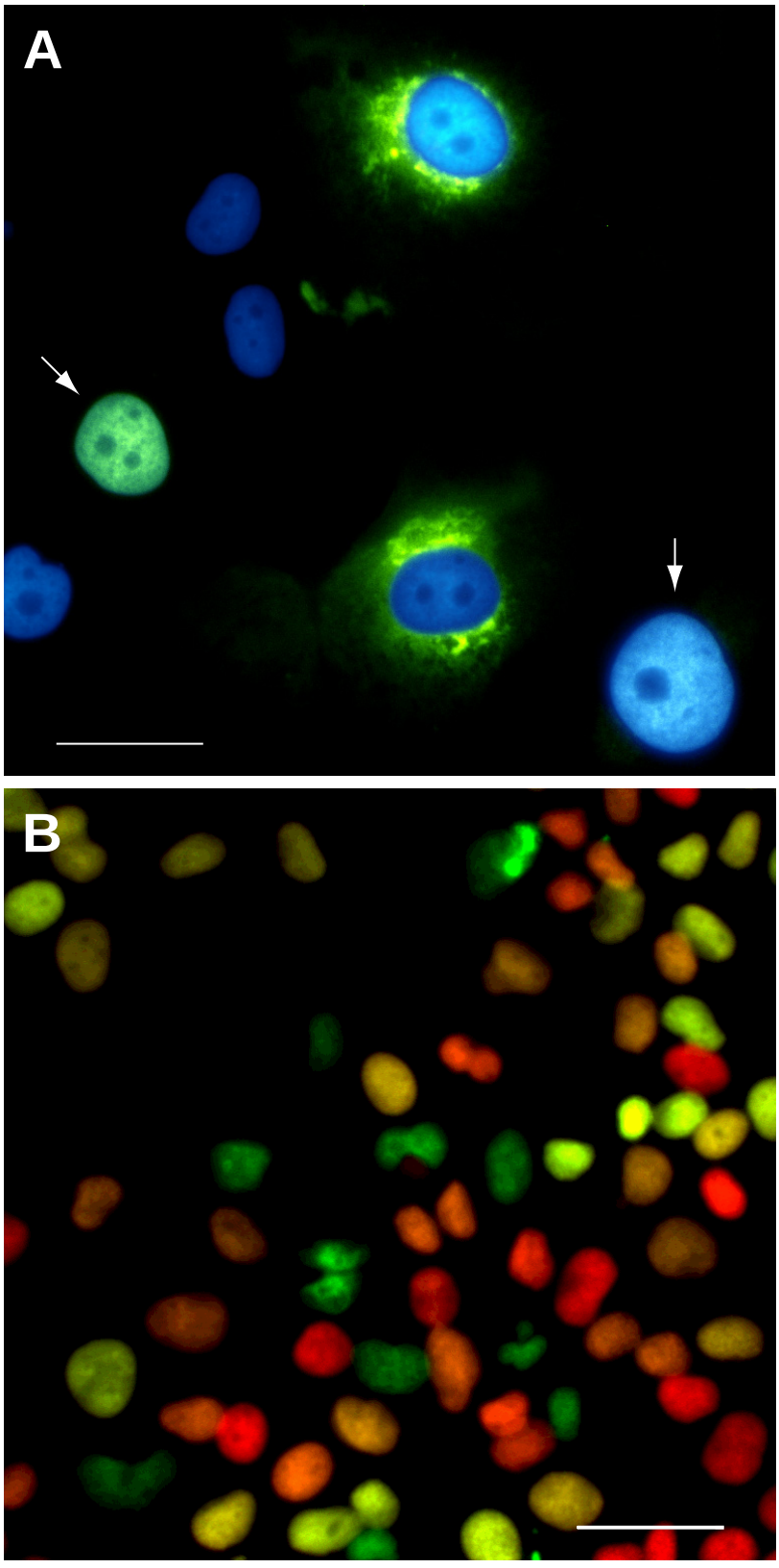
7.7.4 *In Organo* and *In Organismo* Labelling of Chromatin

Cell lines and even organotypic cell cultures are mainly compromised by immortalization and/or transformation as well as by lacking the natural cell interaction and general environment of living tissues, organs and whole organisms. The importance of the *in organo* and/or *in organismo* investigation of complex cellular properties is obviously stressed by the long process of substance development for medical applications ranging from the cellular level, over animal tests to phase I-IV clinical trials in humans. The use of histone-AFP fusionproteins *in organo* and *in organismo* might not only be necessary but also seems possible considering the inexistence of at least obvious influences of histone-AFPs on various properties of cells, the long unchanged stabilities of cloned cell lines and the broad usage possibilities. *In organo* and/or *in organismo* chromatin labelling could be achieved in various ways:

The histone-AFP fusion proteins could be produced, isolated and subsequently transported into the corresponding tissue. Transportation could either be achieved by microinjection or by connection of the histone-AFP to transportation agents used and developed for drug delivery (Lindgren *et al.*, 2000). These methods are, however, compromised by artefacts due to the isolation and transportation process.

Transplantation of cells or tissues already expressing histone-AFP fusion proteins avoids transportation problems. Transplantation of tumors derived from tumor cell lines into animals is a widely used approach in cancer research (Viney, 1995; Sharma & Schreiber, 1999; Siegel *et al.*, 2000; Compagni & Christofori, 2000). The inability to construct whole organs from single cells *ex organismo* (despite first successes in skin and cartilage tissue engineering; Pohamac *et al.*, 1998; La France & Armstrong, 1999; Germain *et al.*, 2000; Holy *et al.*, 2000; Yanas, 2000), could be overcome by transplantation of histone-AFP expressing (embryonic) stem-cells into organs *in organismo*. Such stem cells could differentiate into the corresponding tissue (Camper *et al.*, 1995; Halene & Kohn, 2000; Ourednik *et al.*, 2001).

The most convenient way would be the construction of a transgenic animal (considering now only mammals) expressing the histone-AFP fusion proteins in specific or most tissues (Rosenberg, 1997; Roth *et al.*, 1999). Usually a vector containing the genetic information is transferred in isolated embryonic stem cells, which are cultivated and then analysed for specific parameters like expression patterns or genomic insertation region. The insertation of the AFP directly to a natural histone encoding sequence would, of course be, most favourable (Hurstling, 1997; Willis *et al.*, 1998). Finally, the transgenic animal is created by transferring the stem cell into blastocytes, then being transferred in pseudo-pregnant female animals. The genetic information could also be transferred into already differentiated embryonic cells within the blastocyte. Thus, cells and tissues could be harvested from such animals and used for research. For free GFP such a strategy was already used successfully in mice (Okabe *et al.*, 1997; see also Jacks, 1996), with expression in all tissues despite erythrocytes and hairs. No effects on the animal were described.


A
B

8 “GFP-Walking”: Artificial Construct Conversions Caused by Simultaneous Co-Transfection

8.1 Introduction

Several variants of the green fluorescent protein with distinct spectral characteristics have been developed for multicolour labelling experiments *in vivo* as already described in Chapter 7 “Chromatin Alive“. In Chapter 8 “GFP-Walking“ it is shown that simultaneous co-transfection of fluorescent protein chimeras, a convenient and widely used approach, can cause false positive results due to conversion of their spectral properties. Standard transfection result in ~8%, depending on the treatment of the DNA up to 26%, of the cells expressing altered fusion proteins. This could lead to severe misinterpretation of the results. The conversion is independent of the transfection method and the cell type. The results show that conversion is based on homologous recombination/repair/replication (RRR) events occurring between the nucleotide sequences of the fluorescent proteins. Conversion can be avoided by consecutive transfection or by fluorescent constructs with low sequence similarities. The appearance of conversion makes it possible to easily exchange spectral properties in fusion proteins, to create libraries or to assemble DNA fusion constructs *in vivo*. The detailed quantification of the conversion rate could be used to investigate RRR processes in general.

Fig. 8.1 Fluorescence Exchange by Simultaneous Co-Transfection
(A) Fluorescence image of LCLC–103H cells stably transfected by H2A–eCFP and CATB–eYFP. Cells expressing converted H2A–XFP (arrows) appear beside the expected expression patterns (cyan nuclei and yellow ER/Golgi). Scale bar, 50 µm. **(B)** Cell population enriched for converted cells after G418-selection. This false colour image reveals a variety of recombined H2A–XFP expressions in nuclei (see also Fig. 8.2) and demonstrates clearly the apparent dominance of the effect (green = CFP, red = YFP, yellow = CFP + YFP). Scale bar, 100 µm.

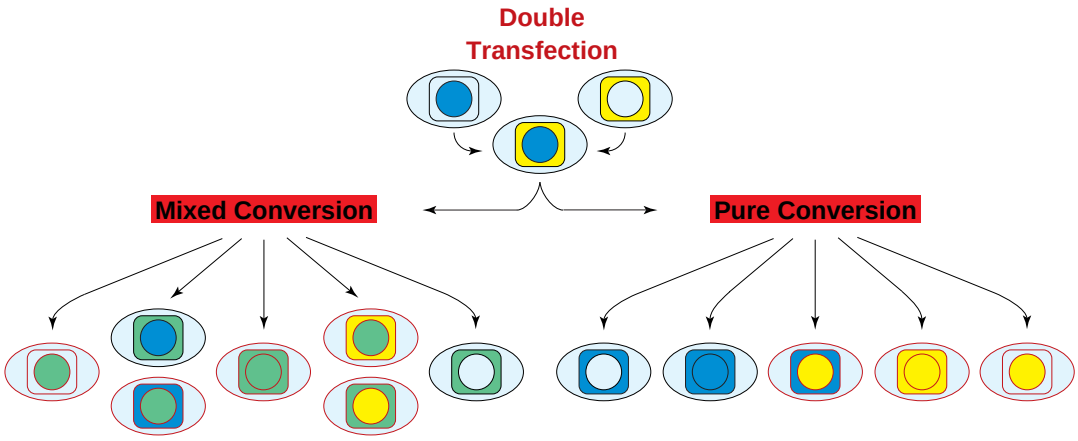


Fig. 8.2 Complexity of Possible False Positive Phenotypes
All cases were observed experimentally. Mixed conversion (green) requires the expression of both the correct as well as the converted fusion protein. The red encircled cases were used to determine the conversion rate R_{H2A}.

8.2 Materials and Methods

8.2.1 Cell Culture

Large cell lung carcinoma cells LCLC–103H, (DSMZ #ACC 384), HeLa (ATCC CCL 2) and COS–7 (ATCC CRL-1651) cells were cultivated in RPMI–1640 medium supplemented with 10% FCS and 4 mM L–Glutamine at 37°C in a 5% CO_2 atmosphere. Stably transfected cell populations were obtained by G418 selection (800 µg/ml). Cells were passaged twice a week. For quantification experiments the cells were seeded after 9 days of G418 selection in Petri dishes at an appropriate density.

8.2.2 Vectors

The pSV-HIII-H2A-CFP vector (described in 7.2.1) used, was obtained by amplifying the coding region of the human histone H2A.i gene (Acc# X83549) with PCR and insertation into the promoterless plasmid pECFP–1 (Clontech) via the EcoRI and BamHI restriction sites. The HindIIIc fragment of the SV40 was inserted into the HindIII site of pECFP–1 in reverse direction, thus the fusion protein H2A–eCFP is expressed through the early SV40 promoter. H2A–DsRed1 was obtained by exchanging the eCFP with the DsRed1 sequence (see also 7.2.1). The sequence encoding the human cathepsin B protein lacking the C-terminal pro-peptide was PCR amplified from the IMAGE clone ID 380482 (Ressourcen Zentrum und Primär Datenbank GmbH, Berlin) with the restriction sites KpnI and SalI. The eYFP nucle-

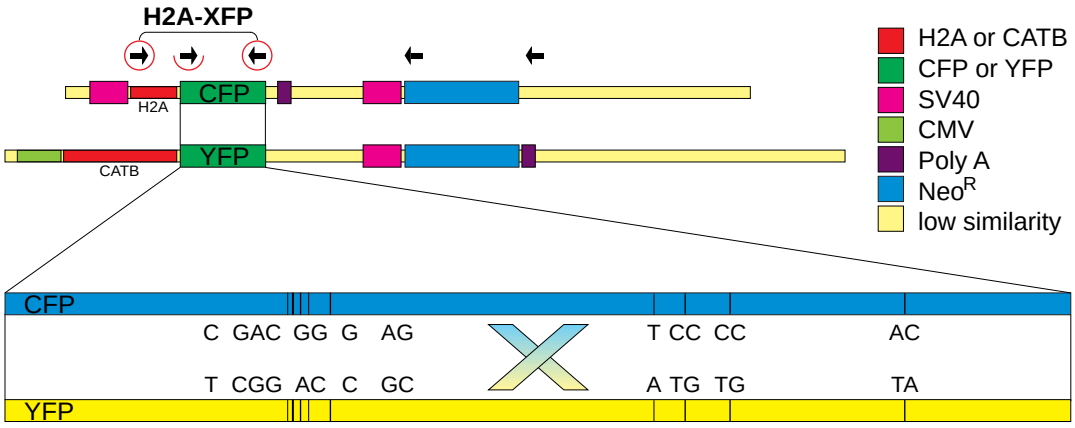


Fig. 8.3 Sequence Comparison of pSV–H2A–eCFP and pcDNA3–CATB–eYFP
Not only the 16 nucleotide difference among eCFP and eYFP over 718 nucleotides are revealed, but also the homologies of the Neo resistance genes and the SV40 promoters. Arrows indicate the direction of some of the PCR primers used to investigate the course and to prove conversion on the DNA level (half and full encircled arrows).

otide sequence was obtained by PCR from a pEYFP–1 (Clontech) derived plasmid with SalI and NotI. Both PCR products were cloned into the pcDNA3 vector (Invitrogen) containing a CMV promoter. In both plasmids the fluorescent protein is attached to the C-terminus of the functional protein (Fig. 8.3). The promoterless pEGFP–1 and the pDsRed1–N1 vectors were obtained from Clontech.

8.2.3 Transfection Procedures

Transfection reagents: Cells were seeded at a density of $10^4/cm^2$ and transfected ~16 h later with equal amounts of both plasmids using either FuGENE 6™ (Roche Molecular Biochemicals), DMRIE–C, CellFECTIN, Lipofectin, or GibcoPlus (Transfection Reagent Kit, Gibco BRL) according to manufacturer's protocols.

Ca_2PO_4 precipitation: Cells were grown in 10 cm Petri dishes to 50% confluence and the medium was changed 2 h prior to transfection. A 2x HBS solution (50 mM Hepes, 280 mM NaCl, 1.5 mM Na_2HPO_4, pH = 6.95 exact!) was diluted drop by drop in the same volume of a 2.5 M $CaCl_2$ solution containing the DNA (1 µg DNA per plasmid filled up to 20 µg DNA with inert pUC17 DNA for better precipitation) while vortexing continuously. After 20 min of incubation the transfection mixture containing the precipitates was added to the culture; 16 h later the cells were washed and the culture medium was renewed.

Electroporation: 10^7 cells were harvested and suspended in 1 ml of cell culture medium. 5 µg of each plasmid were electroporated into the cells with 250 V, 3.9 A,

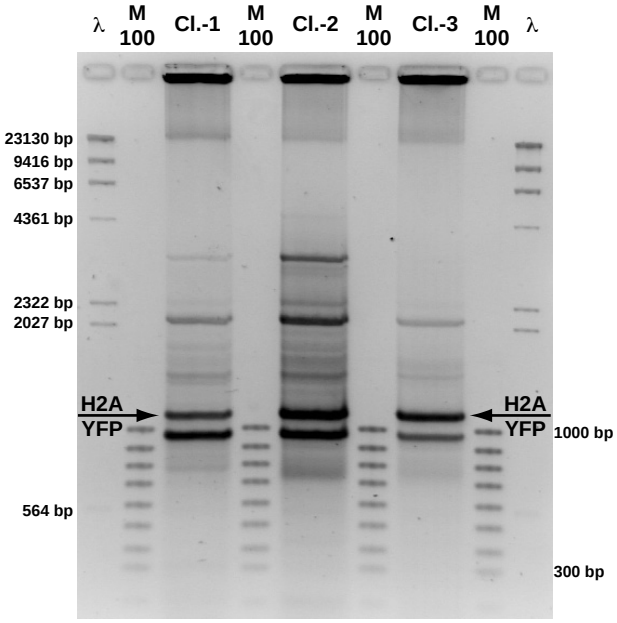


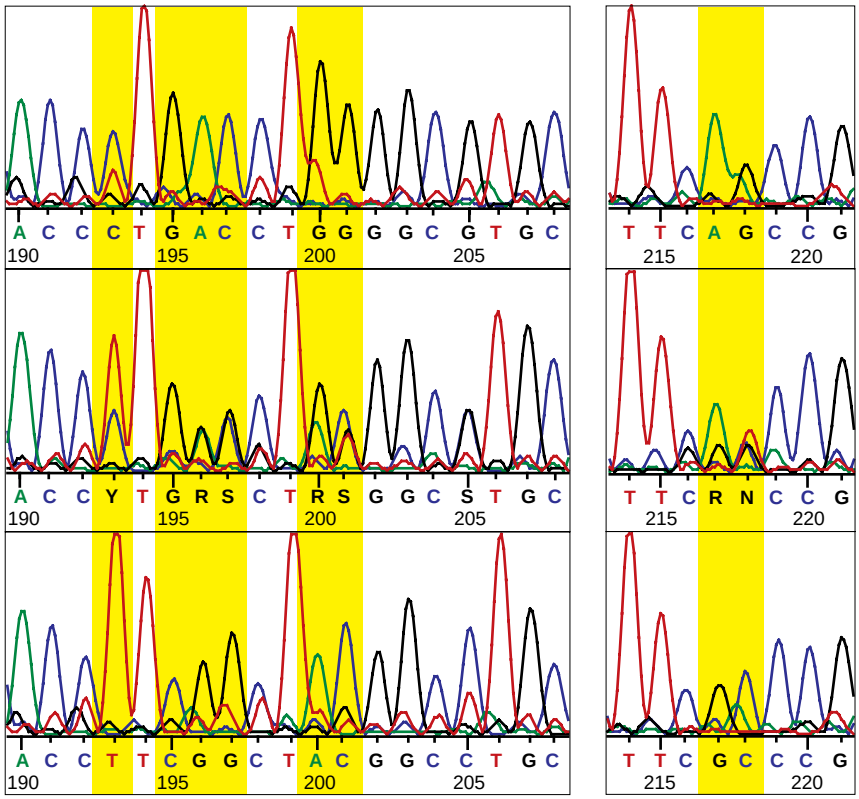
Fig. 8.4 Gel Separation of the PCR Amplification of H2A-YFP from Three Converted Clones
Agarose gel (1.5%, 20cm, 4°C, overnight) showing the successful amplified converted constructs (arrows, λ is the λ-DNA Marker and M100 is the 100bp ladder marker, see also Fig. 7.3). In the pockets the template genomic DNA is visible beside the other bands being PCR byproducts

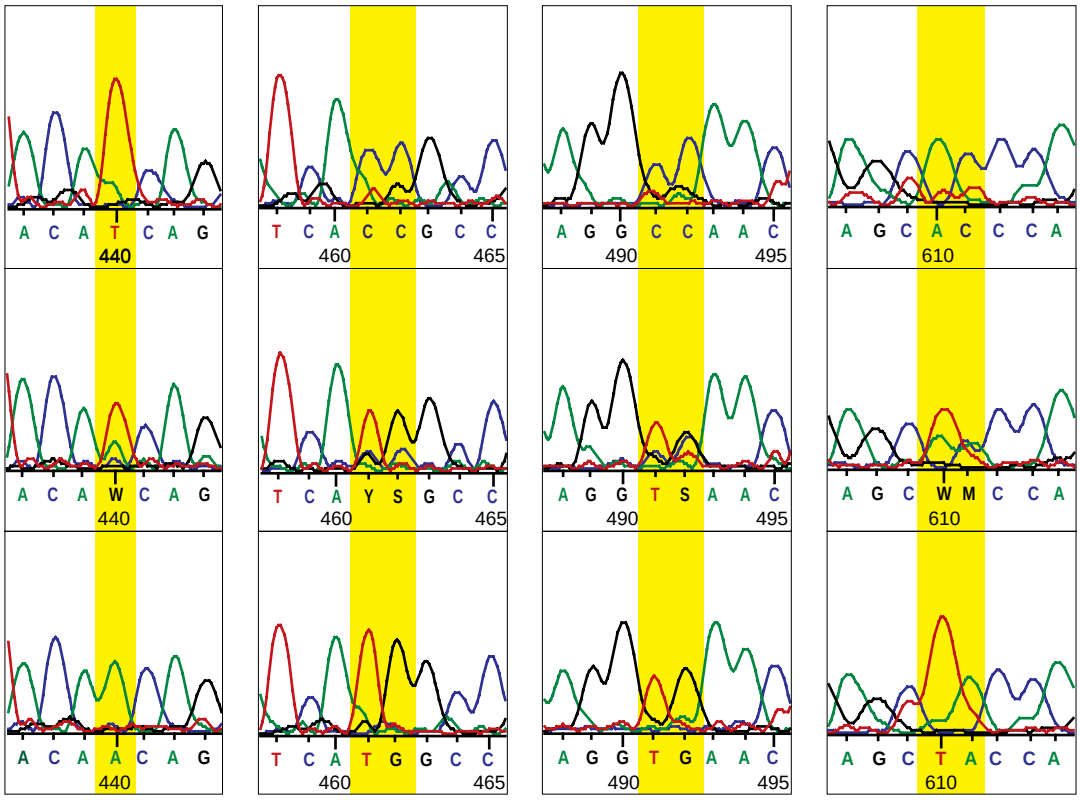
and 125µF at room temperature. Subsequently, the cells were reseeded in Petri dishes; the cells were washed and the medium was renewed after 24h.

8.2.4 Determination of Converted Sequences

Converted H2A cell clones were isolated and grown to confluence in 25cm^2 culture flasks (~10^7 cells). Genomic DNA was extracted by adding 5ml of lysate buffer (1% SDS, 10mM EDTA, 50mM Tris) and 0.1mg/ml (end concentration) proteinase K and incubation at 37°C overnight. The lysate was purified twice by adding 5 ml of phenol/chloroform (1:1v/v), followed by 1 to 3min of whirling and centrifugation at 3,000rpm for 5min. The DNA was precipitated by adding NaCl to 0.1M and 70% ethanol end concentration. The genomic DNA was removed from the solution with a pasteur pipette, resuspended in TE buffer (10mM Tris, 1mM EDTA). This genomic DNA was used as template for PCR amplification of XFP

Fig. 8.5 Sequencing Results Proving Conversion
Sequencing spectra of the XFP part of a PCR amplified H2A-CFP (pure H2A-CFP clone, 1st row), XFP (H2A-CFP+CATB-YFP clone, 2nd row), and H2A-YFP (converted clone, 3rd row) proving conversion. Numbers indicate the base pair position in respect to the beginning of XFP; base pairs are: A=Adenin (green), T=Thymidin (red), G=Guanin (black), C=Cystein (blue), W=A+T, R=A+G, M=A+C, N=T+G, V=T+C, S=G+C; the amplitude is the signal size in the sequencing gel.


A C C C T G A C C T G G G G C G T G C
190 195 200 205
T T C A G C C G
215 220
A C C Y T G R S C T R S G G C S T G C
190 195 200 205
T T C R N C C G
215 220
A C C T T C G G C T A C G G C C T G C
190 195 200 205
T T C G C C C G
215 220


A C A T C A G
440
T C A C C G C C
460 465
A G G C C A A C
490 495
A G C A C C C A
610
A C A W C A G
440
T C A Y S G C C
460 465
A G G T S A A C
490 495
A G C W M C C A
610
A C A A C A G
440
T C A T G G C C
460 465
A G G T G A A C
490 495
A G C T A C C A
610

and H2A–XFP and sequencing using the primers(CGA-ATT-CTG)-ATG-TCG-GGA-CGC-GGC-AAG (H2A–EcoRI–fd) (see also Tab. 7.2), (GGGT-ACC)-ATG-GTG-AGC-AAG-GGC-GAG-GAG-CT (XFP–KpnI–fd), and (GGGT-ACC)-CTT-GTA-CAG-CTC-GTC-CAT-GCC-GA (XFP–KpnI–rv) by the following PCR protocol (Advantage genomic PCR Kit, Clontech): 1' preheating at 94°C, 35x 30'' denaturation cycles at 94°C and 3'30'' elongation cycles at 65°C, and a final 1' annealing step at 65°C. The PCR product was subjected to agarose gel electrophoresis, the respective bands were recovered from the gel (QIAEX II Agarose Gel Extraction, Qiagen) and sequenced (sequencer model: 373A; Big Dye Terminator Cycle Sequencing Kit; Applied Biosystems). Sequence comparisons were done with the sequence align programs AlignPlus (Scientific and Educational Software) and MACAW (Greg Schuler, Version 2.0.5). The original plasmids and DNA from a correct cell clone were used as controls.

8.2.5 Fluorescence Microscopy

Images were collected with an Axiovert S100 TV (Zeiss) equipped with 10x, 20x, and 40x objectives for phase contrast and fluorescence modes, excitation filters for CFP (436/10, Omega) and YFP (515/10, Omega) mounted on a Ludl filterwheel, a dualband emission filter (470/30–555/40, Omega) and a dualband beamsplitter (475/565, Omega) mounted in a filter slider, and a charge coupled device (CCD) camera (C4742–95, Hamamatsu). The image acquisition and processing was performed by the OpenLab software (Improvision). For quantification analysis, the fluorescence images were calibrated to the brightest intensities and about 75 images were captured from each sample containing about $5x10^3$ cells in total for each experiment. For background subtraction defocused images for each channel were taken under the same conditions. The crosstalk was $<1\%$ and $<3\%$ for the eCFP excitation in the eYFP channel and the eYFP excitation in the eCFP channel respectively and was determined both from separate control experiments and during the quantification analyses.

8.3 Qualitative Description and Proof of Conversion

8.3.1 Observation of the Conversion Effect

Mammalian transfection vectors encoding the human histone (H2A.i) tagged with the enhanced cyan fluorescent protein (eCFP) and the human cysteine protease cathepsin B (CATB) tagged with the enhanced yellow fluorescent protein (eYFP), were cotransfected into human tumour cells LCLC–103H by the use of the transfection agent FuGENE 6. Both constructs resulted in comparable expression rates. Observation by fluorescence microscopy revealed, besides the expected cyan emis-

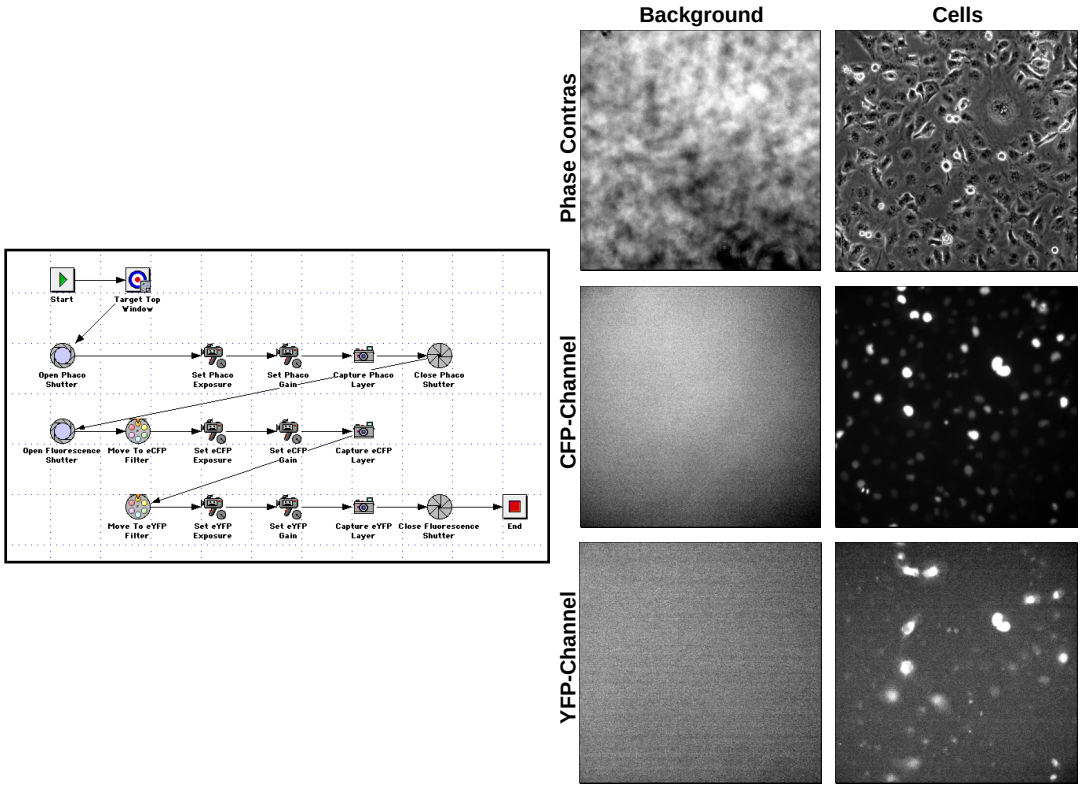


Fig. 8.6 Imaging Procedure
The macro used to regulate the image taking once the area of observation within the Petri dish was chosen by hand **(left)**. Background and images in the phase contrast, the CFP- and the YFP channel show clearly that the observation area is not uniformly illuminated **(right)**.

sion in the nucleus (H2A) and the yellow signals in endoplasmatic reticulum/Golgi/lysosomes (CATB), also signals with the reversed localization. The effect is highly evident in stably transfected populations (Fig. 8.1A, 8.1B; Tab. 8.1(1)).

8.3.2 Variation of Protocols and Generality of Conversion

To verify the effect and to explore its nature and generality, the transfection protocol was varied, the plasmids were manipulated, different transfection mediators and various cell lines were used. A basic control experiment was to transfect the constructs individually which in both cases did not change the fluorescence properties (Tab. 8.1 (2,3)), and thus excluding contaminated plasmid preparations or spontaneous mutations. Supertransfection of a stable H2A–eCFP clone with CATB–eYFP did not result in any conversion (Tab. 8.1 (4)), thus the conversion can not be attributed to an association effect of the fluorescent proteins coexpressed in the same cell.

Extracellular preannealing of the DNA was minimized by mixing the plasmids separately with the transfection mediator and applying them in a consecutive man-

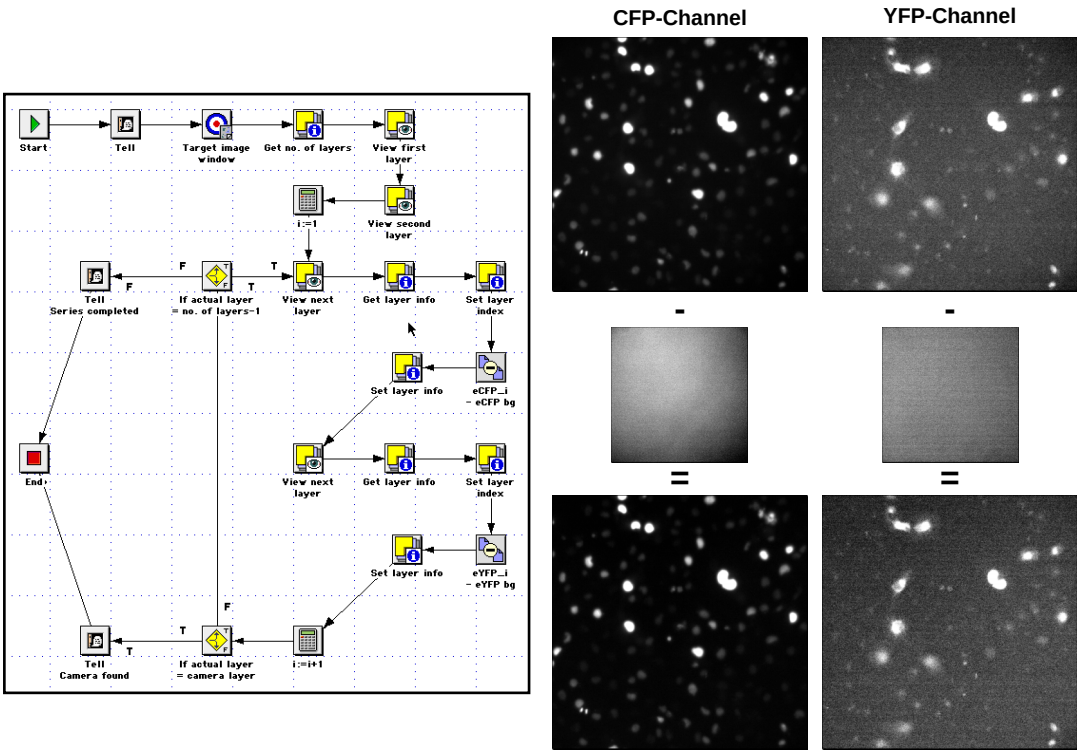


Fig. 8.7 Image Analysis - Background Subtraction
The macro used to subtract the background from the CFP- and YFP-channel **(left)**. The positive effect and more homogenous illumination after background subtraction **(right)**.

ner but immediately one after the other. This treatment considerably reduced the conversion (Tab. 8.1 (5)). Expanding the time interval between the application of the two plasmids from one to several hours, also resulted in a reduction but not in the complete disappearance of the phenomenon. The kinetics revealed no substantial difference between one and eight hours delay (Tab. 8.1 (6–11)). Obviously, one hour delay is already sufficient for a substantial reduction of conversion (Tab. 8.1 (6)). Washing the culture before the second transfection markedly increased the degree of expression of the second construct, but did not affect the conversion rate in comparison to the non-washed sample (Tab. 8.1 (9)).

A promoterless vector encoding only the enhanced green fluorescent protein (eGFP) and the H2A–eCFP construct were cotransfected. The advantage of a promoterless XFP vector over an XFP-chimera containing vector substantially reduces the background and thus facilitates the detection of the conversion events (the signal cannot be suppressed completely due to genomic integration of the XFP sequence downstream to other intrinsic promoters).

This experiment also showed a considerable conversion rate (Tab. 8.1 (14)), which illustrates that conversion is neither affected by the associated protein, the XFP, the promoter nor the vector. It is well known that linear plasmids and strand

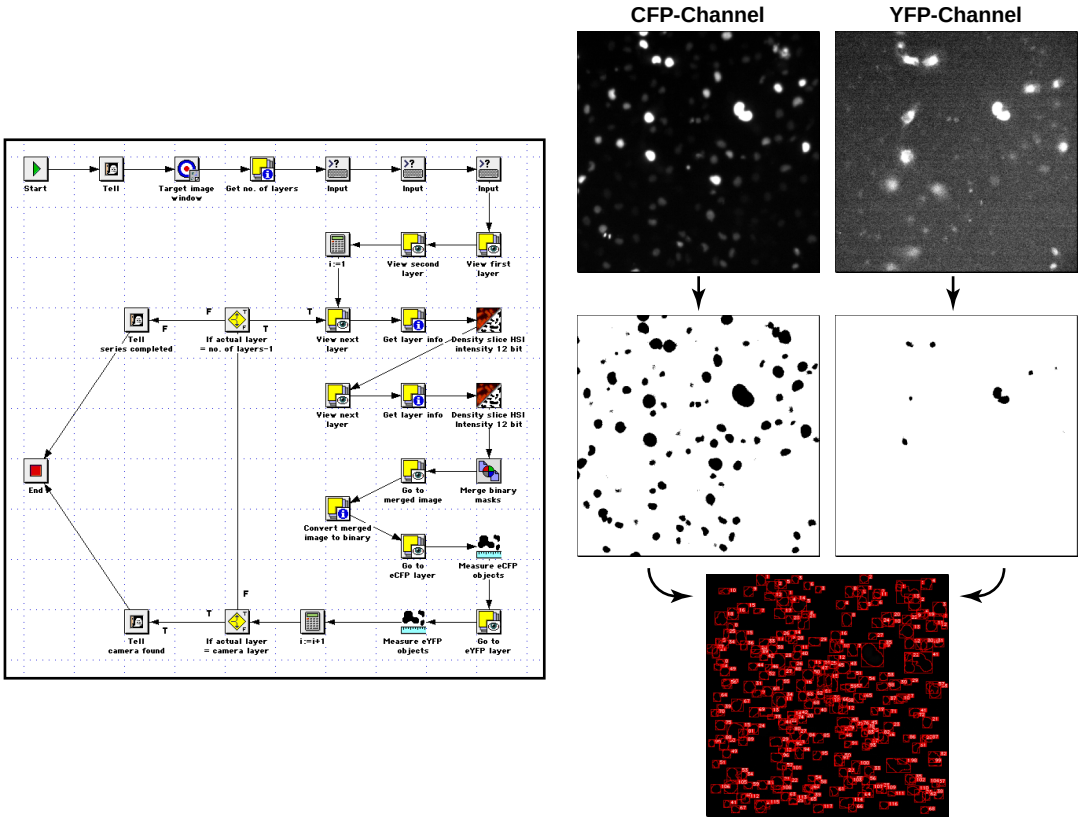


Fig. 8.8 Separation of the H2A from the CATB Signals and Analysis of Signals
Macro used for the separation and analysis of signals **(left)**. From each channel a binary mask after intensity thresholding is created, merged and the signals analysed according to their area **(right)**.

breaks considerably enhance the effect of homologous recombination and integration of DNA into a genome (Anderson & Eliason 1986, Elliott *et al.* 1998, Liang *et al.* 1998). Thus, the vectors were linearized, denatured at 96°C and reannealed by slow cooling to facilitate heteroduplex formation. Particularly the latter approach (which might be the most repair dominated) led to a considerable increase of conversion (Tab. 8.1 (12,13)). All these results strongly suggest that an intimate contact or extracellular preannealing supports the process, while the conversion itself must take place inside the cell, presumably due to RRR activities. To further address the possibility of homologous recombination, simultaneous transfection with the combination of H2A–DsRed1 and CATB–eCFP chimeras was performed. DsRed1 shares only low sequence similarity with eCFP and recombination is therefore unlikely. In fact, this approach revealed no convertants (Tab. 8.1 (15)).

To control the influence of transfection method, the mediators DMRIE–C, CellFECTIN, Lipofectin, GibcoPlus or calcium phosphate precipitation were used and electroporation was performed. Conversions appeared in all cases but at different rates (Tab. 8.1 (16–21)). HeLa and COS–7 cells, frequently used in transfection experiments, exhibited the same conversion phenomenon (Tab. 8.1 (22–25)).

8.4 Proof of Conversion by DNA Sequence Analysis

To prove the exchange of the XFP sequences at the molecular level, cell clones with converted fluorescence properties were isolated. Different primer sets were applied to the extracted genomic DNA to fish the recombined XFP sequences by PCR analysis (Fig. 8.4). Surprisingly, only primers for the XFP sequence worked but none starting further up- or downstream within the vector sequences. This suggests that the integration of the DNA terminates with the XFP-sequence. In the inspected clones the amplified H2A–XFP sequence revealed a complete transition from the eCFP to the eYFP sequence in all 16 variant nucleotides (compare 1st and 3 rd row in Fig. 8.5), in contrast to the corresponding controls (1st and 2nd row in Fig. 8.5).

8.5 Quantification of the Conversion Rates

Once the generality of the effect was established, the conversion rates were quantified to further substantiate the results. Trials to determine these by fluorescence activated cell sorting (FACS) analysis failed because the signals could neither be discriminated by their size nor by the intracellular location of the fluorescent objects (data not shown). However, digital microscopy and image processing offer a unique alternative: (i) microscopy allows direct control of transitions; (ii) microscopic images allow the setting of a signal–area threshold, necessary for unambiguous detection of objects. Moreover, for double transfection the theory predicts in total 15 expression patterns, one of which is the desired transfection, two others express the single constructs and 12 are conversions. The complexity of possible false positive phenotypes is illustrated in Fig. 8.2. All of these were experimentally verified. To reach statistical relevant results close to those obtained in FACS analysis an algorithm which facilitates the interactive objective evaluation of the specimen according to these parameters was developed:

8.5.1 Space and Intensity Resolved Planeometric Microscopy (SIRPM)

After background subtraction which was necessary due to the unisotropic illumination of the field of observation (Fig. 8.6), the stronger H2A signals (C_i, Y_i) were separated from the weaker CATB (c_i, y_i) signals by an intensity threshold (Fig. 8.7). However, a pure intensity threshold does not completely separate the H2A from the CATB signals (see also Fig. 8.7). Therefore, from the remaining objects binary masks were created to calculate the mean grey value (i) and the area (a). To eliminate redundancies, both masks were merged and applied to each channel (Fig. 8.8). By setting an area threshold only the mean grey values of the nuclear histone signals (C_a, Y_a) were selected, the smaller CATB signals (c_a, y_a) and arte-

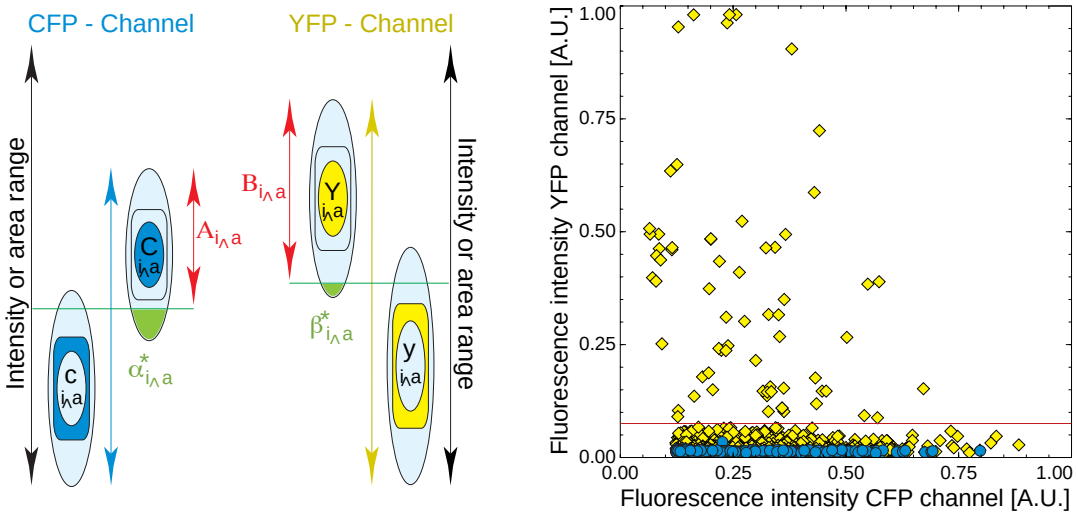


Fig. 8.9 Quantification of the Total Conversion Rate and Typical Signal Analysis Plot of a Standard Co-Transfection Experiment

(Left) To quantify the conversion the H2A signals were distinguished from the CATB signals by the different range of their area and intensity distribution ($c_{i \wedge a}$ and $C_{i \wedge a}$, $y_{i \wedge a}$ and $Y_{i \wedge a}$) separately in both fluorescent channels. **(Right)** Standard conversion experiment. As expected, most nuclei show cyan fluorescence and a small amount of yellow signals due to the near spatial localization of the CATB. Converted nuclei are separated by an intensity threshold (all yellow rhombi above the red line). Control cells expressing only H2A–eCFP illustrate that cross talk between the CFP and the YFP channels is not responsible for the conversion effect (blue circles).

facts were ignored. The mean grey values were then used to optimize the intensity threshold. To determine the conversion rate correctly, the errors $\alpha_{i \wedge a}$* and $\beta_{i \wedge a}$* (which occur while distinguishing the H2A signal distributions $C_{i \wedge a}$ and $Y_{i \wedge a}$ from the CATB signal distributions $c_{i \wedge a}$ and $y_{i \wedge a}$, resulting in a loss of positive H2A signals) needed to be equal in both the CFP and the YFP channel. Thus, the conversion rate from H2A–eCFP to H2A–YFP is

$$R_{H2A} = \frac{B_{i \wedge a}}{A_{i \wedge a} + B_{i \wedge a} - (A_{i \wedge a} \wedge B_{i \wedge a})} \tag{8.1}$$

where $A_{i \wedge a}$ is the number of H2A–eCFP expressing nuclei, $B_{i \wedge a}$ the number of H2A–YFP, and $(A_{i \wedge a} \wedge B_{i \wedge a})$ the number of the nuclei containing both constructs (Fig. 8.9, left). The results of the automated procedure were manually controlled by a hand recount in an adequate sample; comparable results were obtained. Presumably, conversion affects H2A and CATB equally, but it is far easier to determine for the H2A, since the CATB signal distribution is difficult to distinguish from artefacts. Thus, the apparent total conversion rate, which describes the cell population expressing wrong labelled constructs, is

$$R_{total} = 2R_{H2A} - {R_{H2A}}^2 \tag{8.2}$$

where ${R_{H2A}}^2$ is the number of cells showing conversion of both constructs.

8.5.2 Detailed Quantification of the Conversion Rates

For quantification, transfected populations were enriched for expressing cells by G418-selection and a statistical relevant number of cells (~5x10^3) was recorded and evaluated. The constructs used, the cell type, the transfection method (in respect to transfectant and protocol) and the quantification results are summarized in Tab. 8.1 and correspond to the qualitative description and proof of conversion described already in Chapter 8.3. Focusing on the need for the scientist to know the total number of false positive results of cells within the cell culture R_{total} calculated with Equ. 8.2 is discussed here:

Conversion in a standard double transfection experiment appears in up to ~8% of stably transfected cells (mean of 4 independent experiments). The variance between the different transfectants could either be real or due to the not optimized protocols used (in contrast to the FuGENE 6 protocol). The corresponding control experiments using individual transfection of both constructs, supertransfection or using DsRed1 revealed 0.0% conversion. The exact agreement of the calculated values to zero, stress the quality of the SIRPM algorithm. Applying the two constructs separately or with different time delays reduced conversion to ~2%. Interestingly, the delay kinetics shows no continuously decreasing behaviour but seems to stay at a low but constant level starting with time delay as small as 1h. Thus the conversion effect could be to a greater part be influenced by the transfectant and only to a minor part by the intracellular interaction of both constructs. This hypothesis could be supported by the protocol using calcium phosphate and electroporation as transfection method with total conversion rates of also ~2.0%.

Linearized DNA shows a similar total conversion rate (6.7%) as in the standard transfection experiment which suggests that the artificial linearization is similar to the natural condition needed for RRR processes. Manipulations of DNA, which favour the annealing, increase conversion to ~26%, which is close to the theoretical value of 25%. This number is based on the following assumptions: (i) the four possibilities of reannealing after double-strand melting of two linearized sequences are equally distributed, (ii) the RRR processes are inable to discriminate between the mismatched strands. In this case base pair mismatch repair might dominate the conglomerate of RRR processes.

8.6 Discussion of “GFP-Walking“ and Future Aspects

Since its discovery by Morin & Hastings (1971) the green fluorescent protein (GFP) has become an important *in vivo* marker in the life sciences (Chalfie *et al.* 1994). The importance of fluorescent proteins could be estimated from the ~3,500 articles, thereof ~1,500 in the year 2000, having been published since their discovery in the 1970s. The poor fluorescence properties, proteolytic and photochemical instability

Tab. 8.1 Quantification of the Conversion Rate for Different Experimental Conditions
n.q. = not quantified.

#	Construct	Cell type	Transfection method		Conversion	
			Transfectant	Protocol	+/-	R_{total} [%]
1	H2A–eCFP + CATB–eYFP	LCLC–103H	FuGENE 6	simultaneous	++	8.2
2	H2A–eCFP	LCLC–103H	FuGENE 6		-	0.0
3	CATB–eYFP	LCLC–103H	FuGENE 6		-	0.0
4	H2A–eCFP + CATB–eYFP	LCLC–103H	FuGENE 6	secondary transfection of stable line	-	0.0
5	H2A–eCFP + CATB–eYFP	LCLC–103H	FuGENE 6	separate mix + simultaneous	+	≥2.0
6	H2A–eCFP + CATB–eYFP	LCLC–103H	FuGENE 6	1h delay	+	~2.0
7	H2A–eCFP + CATB–eYFP	LCLC–103H	FuGENE 6	2h delay	+	~2.0
8	H2A–eCFP + CATB–eYFP	LCLC–103H	FuGENE 6	4h delay	+	2.0
9	H2A–eCFP + CATB–eYFP	LCLC–103H	FuGENE 6	4h delay (no wash)	+	~2.0
10	H2A–eCFP + CATB–eYFP	LCLC–103H	FuGENE 6	8h delay	+	~2.0
11	H2A–eCFP + CATB–eYFP	LCLC–103H	FuGENE 6	simultaneous; linearized	++	6.7
12	H2A–eCFP + CATB–eYFP	LCLC–103H	FuGENE 6	simultaneous; linearized + 96°C	+++	26
13	H2A–eCFP + eGFP (promoterless)	LCLC–103H	FuGENE 6	simultaneous	+	≥2.0
14	H2A–DsRed1 + CATB–eCFP	LCLC–103H	FuGENE 6	simultaneous	-	0.0
15	H2A–eCFP + CATB–eYFP	LCLC–103H	DMRIE–C	simultaneous	++	7.5
16	H2A–eCFP + CATB–eYFP	LCLC–103H	CellFECTIN	simultaneous	+	≥2.0
17	H2A–eCFP + CATB–eYFP	LCLC–103H	Lipofectin	simultaneous	++	≥4.7
18	H2A–eCFP + CATB–eYFP	LCLC–103H	GibcoPlus	simultaneous	-	≥2.0
19	H2A–eCFP + CATB–eYFP	LCLC–103H	calcium phosphate	simultaneous	+	~2.0
20	H2A–eCFP + CATB–eYFP	LCLC–103H	electroporation	simultaneous	+	~2.0
21	H2A–eCFP	HeLa	FuGENE 6		-	0.0
22	H2A–eCFP + CATB–eYFP	HeLa	FuGENE 6	simultaneous	++	n.q.
23	H2A–eCFP	COS–7	FuGENE 6		-	0.0
24	H2A–eCFP + CATB–eYFP	COS–7	FuGENE 6	simultaneous	++	n.q.

as well as extended posttranslational folding time of the wild type GFP (wtGFP), resulted in the development of a variety of improved fluorescent proteins (XFPs) with different spectral properties suitable for multicolour labelling experiments (Heim *et al.* 1995, Cubitt *et al.* 1995, Cormack *et al.* 1996). The enhanced variants derived from wtGFP share about 99% sequence identity; others, like the recently discovered DsRed1 and its improved variant DsRed2 show only 22% sequence identity with wtGFP (Matz *et al.* 1999, Fradkov *et al.* 2000, Wall *et al.* 2000).

Several methods have been developed for the transfer of DNA into cells. The most widely used transfection techniques include calcium phosphate coprecipitation (Graham & van der Eb 1973), electroporation (Andreason & Evans 1988, Shigekawa & Dower 1988), use of viral vectors (Piccini *et al*. 1987), cationic liposome-mediated transfection (Felgner *et al.* 1987), lipopolyamines (Remy *et al.* 1994), dendrimers (Haensler & Szoka 1993), and non-liposomal lipid formulations (Uyttersprot *et al.* 1998). Double transfection of DNA can be achieved in a successive or a simultaneous way. Successive transfection is often less efficient; supertransfection of clones already stably expressing one construct is time consuming due to the selection and the propagation process. In contrast, simultaneous co-transfection is not only the faster, more convenient and efficient method (e.g. in FRET experiments (Pollok & Heim 1999)), but also the only possibility in time critical experiments involving e.g. rapid cell death. However, in this case, DNA containing regions with a high degree of similarity could undergo extrachromosomal rearrangements. Although concatemer formation and homologous recombination (Anderson & Eliason 1986, Stark *et al.* 1992, Haber 1999, Thacker 1999, Flores-Rozas & Kolodner 2000) are known to occur during transfection procedures and might promote insertion into chromosomes (Bishop 1996), the details of the processes involved are not clear yet.

The results of Chapter 8 prove that the conversion of fluorescence properties in co-transfection is a significant event and is caused by recombination/repair/replication processes. Experiments were evaluated by quantitative microscopy using an image analysing algorithm which substitutes and surpasses fluorescence activated cell sorting (FACS) because it discriminates not only for signal intensity but also for area/location. The observations are relevant for the interpretation of all transfection experiments in which constructs with similar sequences are used, irrespective of whether the tags or the genes of interest bear the similarity. It should be emphasized that one should be aware of those misleading results. They can be overcome by (i) the delayed protocols or (ii) low sequence similarities of the markers. Since DsRed1, the only alternative at the time of experiments, is compromised for other reasons such as tetramer formation, higher toxicity or deficient addressing (unpublished data by Felix Bestvater), the need for the development of better markers is further emphasized. At the same time this illustrates possible difficulties in already available systems containing multiple XFP sequences with regards to intra– or intermolecular interactions and thus potential conversion.

It is known that recombination depends on strand breaks which are more frequent in transformed cells (Johnson et al. 1999, Richardson & Jasin 2000) or cells damaged by irradiation or chemicals. One of the essential steps in the conversion mechanism is the preannealing of the DNA. It is likely that transfection conditions facilitate such preannealing but that the final conversion is accomplished by intracellular recombination. These activities imply an active DNA RRR machinery well known for mammalian cells (Haber 1999, Thaker 1999). The probability for chromosomal insertion of converted constructs after RRR processes is presumably

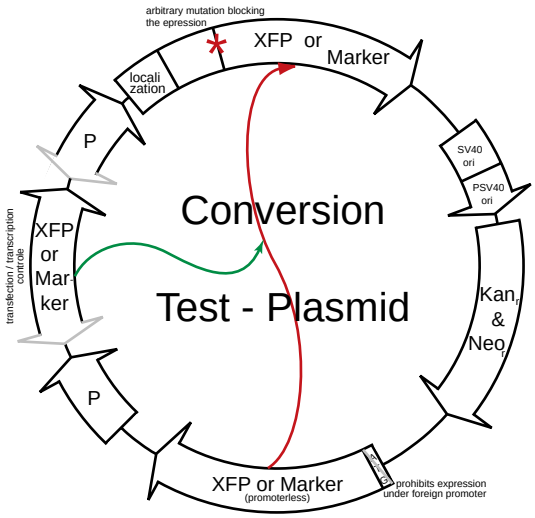


Fig. 8.10 Test Plasmid for Conversion Optimized for FACS Usage
It consists of an expressed (promoter P) but non-functional marker which can be functionalyzed through conversion using the correct marker sequence which lacks the start codon ATG to prevent expression through a natural promoter. A standard for quantification is introduced through another marker which might be linked through a bidirectional promoter to the non-functional marker.

higher for plasmids with double-strand breaks than for intact (and thus not converted) plasmids.

Beyond the above described methods and results, the simultaneous transfection protocol and the high rate of conversion also provides new opportunities whose principle feasibility was already shown above:

8.6.1 *In vivo* and *in vitro* Method for the Integral Investigation of RRR Processes

As conversion is based on RRR processes it could be utilized to investigate RRR properties of cells in general and particularly in the unstable genomes of transformed cells. The appearance and quantification of conversion leads to an integral analysis of RRR processes as their effect is measured directly on the DNA level. In contrast, current research most of the time investigates expression levels of known factors involved and rarely their activity. Conversion has the advantage that such investigations could now be performed *in vivo* with fewer restrictions and artefacts. Various experiments could also be conducted on the same cell population or even on single cells without their unsustainable consumption. Cells with special RRR properties could be separated and cultivated for further use. Of course, conversion therefore saves time and money by avoiding complex preparation methods.

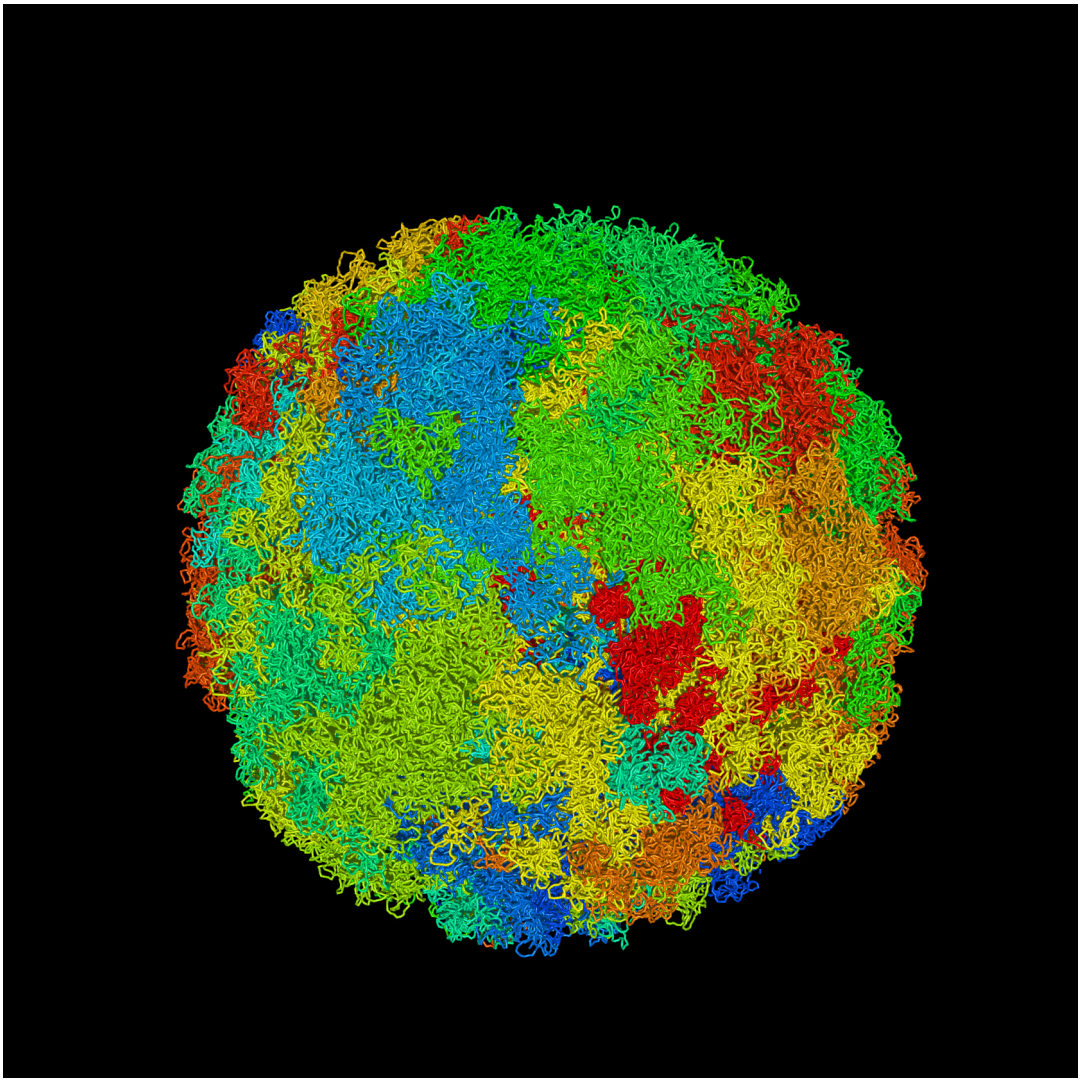
8.6.2 An Optimized Plasmid for FACS Analysis

The advantages of conversion for the integral investigation of RRR processes described in 8.6.1 can be further improved by the use of optimized vectors suitable for FACS analysis with multiple XFP sequences and integrated non-homologeous calibration markers. Such vectors are not only a direct, reproducible and quantifiable approach but also combine the positives of conversion with those of mass screenings.

The makeup of such an optimized vector is proposed in Fig. 8.10: A mutated and therefore non-functional XFP or other marker is expressed through some promoter P. In contrast the correct and functional marker (avoiding concentrationand stoichometric matters by being on the same plasmid) is not expressed. This is secured by omitting the start codon ATG to prevent a natural promoter near to the genomic integration site having an expression effect. Due to conversion the non-functional marker is functionalized and expressed through the promoter P. Additionaly the expressed marker can carry a localization signal for easier use and detection. For quantification another marker can be used as transfection control. This standard is optimized if the promoter P is bidirectional, thus the marker and standard are expressed to the same degree. The sequence of the marker or sequences at its starting or ending flanks could be adjusted such that special sequence dependent properties of RRR processes could be investigated.

8.6.3 *In vivo* and *in vitro* Method for the Creation of DNA Constructs

The basis of conversion - that is the manipulation of a DNA sequence by RRR processes - could be used to easily exchange fluorescence or other properties of (fusion) proteins, to create DNA libraries or in general to assemble DNA constructs *in vivo*. For fluorescence exchange this strategy was already utilized successfully by simultaneous double transfection of H2A-eCFP with the 718 bp sequence of eGFP (Tab. 8.1 (13)). This works much faster than cloning the gene into a vector system and expressing it in cells, which requires various enzymatic digests, sequence purifications, ligations and control gels. Such an approach could be especially useful in cases where a construct cannot be build or transported into cells or organism due to special sequence properties like a functional structure or the mere size of the final construct. In such cases the final construct could be mutated or separated in parts which by-pass these obstacles and assembled *in vivo*.



9 Summary and Synthesis

9.1 Summary

The cell nucleus is organized in a complex manner to store, transcribe, and replicate the genomic information necessary for most processes from the cellular level, over embryogenesis to cognitive ability. This organization was discussed since the discovery of the nucleus. According to current knowledge the nucleus is organized in seven levels of packaging to function efficiently: the DNA double helix (i), winds around a protein complex forming the nucleosome (ii), which condenses irregularly to the 30 nm chromatin fiber (iii), which is folded into chromatin loops (iv), which aggregate to chromosomal subdomains (v), which constitute a chromosome (vi), which are nonrandomly arranged in the nucleus (vii). With increasing level of packaging the knowledge gets sparser and many hypothesis are still contrasting or inconclusive. This thesis approaches the sequential and three-dimensional organization of the human genome by integrating different aspects from all nuclear scales:

To investigate the folding of the 30 nm chromatin fiber into chromosome territories, their morphology and experimental distinguishability, single chromosomes based on the Multi-Loop-Subcompartment (MLS) model, in which small loops form rosettes, connected by a linker, and the Random-Walk/Giant-Loop (RW/GL) topologie, in which large loops are attached to a flexible bachbone, were simulated for various loop and linker sizes. The 30 nm chromatin fiber was modelled as a polymer chain with stretching, bending and excluded volume interactions. A spherical boundary potential simulated the confinement of other chromosomes and the nucleus. Monte Carlo and Brownian Dynamics methods were applied to generate chain configurations at thermodynamic equilibrium. These simulations of single chromosomes were extended to nuclei of diploid human cells containing all 46

Fig. 9.1 Rendered Image of Nucleus
Shown is the rendered image of a nucleus with 5 µm radius in which the 30 nm chromatin fiber is folded according to the Multi-Loop-Subcompartment (MLS) model, with 126 kbp loops and linkers. The formation of chromosome territories and subcompartments is the characteristics of this model.

chromosomes, to determine the chromosome arrangement and the related microscopic morphology, besides the validation of the results of the simulation of single chromosomes. The chromatin fiber was simulated as in the case of single chromosomes. Since the computer power increased by a factor of 46, this time simulated annealing and Brownian Dynamics methods as well as a four step decondensation procedure from metaphase were applied to generate interphase configurations again at thermodynamic equilibrium. Both the MLS and the RW/GL model form chromosome territories with different morphologies: The MLS rosettes result in distinct subcompartments visible with light microscopy. This morphology and the size of these subcompartments agree with the morphology found by expression of histone autofluorescent protein fusions (see below) and FISH experiments. In contrast, the big RW/GL loops lead to a homogeneous chromatin distribution. Even small changes of the model parameters induced significant rearrangements of the chromatin morphology. Thus, in turn pathological diagnoses of e. g. cancer based on the nuclear morphology, might be related to structural changes on the chromatin level. The position of chromosome territories in interphase depends on their metaphase location, and suggests a possible origin of current experimental findings. Only the MLS model leads to a low overlap of chromosomes, arms and subcompartments, again in agreement with experiments. The chromatin density distribution in CLSM image stacks of the MLS model but not the RW/GL model reveals a bimodal behaviour in agreement with recent experiments. 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. (For an overview of all analysed parameters see Tab. 9.1.)

In order to characterize the levels of packaging of the genome, the scaling behaviour of the 30 nm chromatin fiber topology and scaling behaviour the morphology of simulated confocal laser scanning microscopic (CLSM) image stacks was determined. Both were obtained from simulations of single chromosomes and whole nuclei using the RW/GL and MLS chromatin fiber topologies. For the analysis, various scaling/fractal dimensions were calculated. The scaling of the chromatin fiber revealed different power-law behaviours on different scales. This multi-scaling is created by the random walk behaviour of the fiber, the globular nature and the arrangement of loops or rosettes. Within the multi-scaling regime a fine-structure was present for the MLS model arising from the rosette loops. A similar fine-structured multi-scaling behaviour was also found in the correlation behaviour on the level of the DNA sequence of human chromosomes (see below). Thus, the sequential and three-dimensional organization of genomes are closely interconnected. The scaling of CLSM image stacks also reflected the model and imaging properties in detail. Thus, the chromatin fiber topology is also closely connected to nuclear morphology. Therefore, scaling analyses of the nuclear morphology are a suitable approach to differentiate between different cell states, e. g. during the cell cycle, due to malignancy, in apoptosis or in response to drugs. Consequently, the scaling behaviour shows that all nuclear organization levels are connected.

To determine the impact of the three-dimensional genome organization on molecular mobility, the accessability of nuclear loci and the hypothesis of the Inter-Chromosomal Domain (ICD) model, the diffusion of spheres was simulated by Brownian Dynamics in computer generated nuclei with an MLS chromatin fiber topology. The tracers interacted with the static fiber by an excluded volume potential. Visual inspection of the morphology of simulated chromosomes or nuclei revealed big spaces allowing high accessibility to nearly every spatial location. A channel like network for molecular transport between chromosome territories, as postulated by the ICD model, was not apparent in the simulations. The big spaces are supported by estimating the nuclear volume occupied by chromatin of <30%, leaving >70% of space for diffusion with an average mesh spacing of 29 to 82 nm for nuclei of 6 to 12 µm diameter. This agrees with the simulated mean displacement for 10 nm sized particles of ~1 to 2 µm within 10 ms. Therefore, the diffusion of biological relevant tracers is only moderately obstructed. The anomaly parameter D_W characterizing the degree of obstruction ranged from 2.0 (obstacle free diffusion) to 4.0, in agreement with experiments. The degree of obstruction was proportional to the nuclear density, the fiber diameter, the interaction hardness and the tracer size. Different fiber topologies had no effect on the average particle displacement. Consequently, molecules and proteins might reach every nuclear location by energy independent diffusion without a special channel like network.

The sequential organization, i. e. the relations within DNA sequences, and its connection to the three-dimensional organization of genomes was investigated by correlation analyses of 113 completely sequenced chromosomes of 0.5×10^6 to 3.0×10^7 bp from Archaea, Bacteria, *Arabidopsis thaliana*, *Saccharomyces cerevisae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens*. All sequences revealed long-range power-law correlations almost on the entire observable scale. The local correlation coefficient shows close to random correlations on the scale of a few base pairs, a first maximum from 40 to 3400 bp (for *Arabidopsis thaliana* and *Drosophila melanogaster* divided in two submaxima), and often a region of one or more second maxima from 10^5 to 3×10^5 bp. This multi-scaling behaviour was species specific. Computer generated random sequences assuming a block organization of genomes reproduced such multi-scaling. Within this multi-scaling behaviour an additional fine-structure is present and attributable to the codon usage in all except the human sequences. Here it is connected to nucleosomal binding. Computer generated random sequences assuming the codon usage and nucleosomal binding agree with these results. Mutation by sequence reshuffling destroyed all correlations, thus their stability seems evolutionary tightly controlled and connected to the spatial genome organization on large scales. This is supported by the scaling behaviour of the chromatin topology (see above).The correlation behaviour was used to construct trees, which were similar to the corresponding phylogenetic trees for β-Tubulin genes of Oomycetes and Eukarya genomes. For Archaea and Bacteria tree construction led to a new classification system with four major tree branches/classes. In summary, these findings suggest a complex sequen-

Tab. 9.1 Comparison between Simulated and Experimental Chromosome Parameters
The general comparison distinguishes between measurement category and the defined experimental parameter. Then the predictive value from simulations of single chromosomes and whole nuclei is judged. The experimental data are characterized by their availability for this comparison and by their principle availability, i. e. the measurements were done, but not analysed suitable for comparison (bracket). Additionally, the experimental method is judged by its *in vivo* applicability (v). Finally, the simulation and experimental results are compared (Multi-Loop-Subcompartment model: MLS; Ran-

Measurement Category	Parameter	Simulation		Experimental Data Available		Comparison Result
		Chromosomes [-+]	Nuclei [-+]	[N/Y]	Method (v)	
Qualitative Morphology	General appearance on the fiber level	++	+++	N	-	-
	Electron microscopy (EM)	+	+++	Y	EM, cryo EM	non-ICD
	Confocal Laser Scanning Microscopy (CLSM)	+	+++	Y	FISH, BrdU (v), His-AFP (vv)	MLS non-ICD
Quantitative Morphology	Intensity distribution of CLSM images	+	+++	Y	BrdU (v), His-AFP (vv)	MLS
	Diffuseness distribution of CLSM images	+	+++	N (Y)	BrdU (v), His-AFP (vv)	n. d.
	Skewness distribution of CLSM images	+	+++	N (Y)	BrdU (v), His-AFP (vv)	n. d.
	Kurtosis distribution of CLSM images	+	+++	N (Y)	BrdU (v), His-AFP (vv)	n. d.
Shape/Form	Roundness of chromosomes	+	+++	N (Y)	FISH, BrdU (v)	n. d.
Colocalization	Overlap of chromosomes	-	+++	Y	FISH	MLS
	Overlap MLS subcompartments	+	+++	Y	FISH, BrdU (v)	MLS
Extension	Radial mass and density distribution of nuclei	-	+++	N (Y)	FISH, BrdU (v), His-AFP (vv)	n. d. (MLS)
	Radial mass/density distribution of chromosomes	++	+++	N (Y)	FISH, BrdU (v)	n. d. (MLS)
	Radial mass/density distribution of MLS subcompartments	++	+++	N (Y)	FISH, BrdU (v)	n. d. (MLS)
Distances	Distances between arbitrary chromosomes	-	+++	N (Y)	FISH, BrdU (v)	(MLS RW/GL)
	Distances between nearest chromosomes	-	+++	N (Y)	FISH, BrdU (v)	(MLS RW/GL)
	Distances between arbitrary MLS subcompartments	+	+++	Y	FISH, BrdU (v)	MLS

tial organization of genomes closely connected to their three-dimensional organization.

The *in vivo* morphology and dynamics of chromatin is difficult to assess by electron microscopy, fluorescence *in situ* hybridization (FISH) and *in vivo* stains since these methods require fixation or produce artefacts. To overcome these limitations a novel *in vivo* technique for chromatin labelling was established: DNA vectors encoding the fusion proteins of all histones H1.0, H2A, mH2A1.2, H2B, H3, H4 and the autofluorescent proteins CFP, GFP, YFP, DsRed1 DsRed2 were developed and expressed stably in HeLa, LCLC103H, Cos7 and ID13 cells. 2.6 to ~20% of the nucleosomes carry a label. No apparent influence of the cell cycle status, the proliferation rate or the AFP fluorescent excitation/emission spectra, but recently a somewhat increased nucleosomal repeat length was detected. With this approach the

dom-Walk/Giant-Loop model: RW/GL; Inter-Chromosomal Domain model: ICD; n. d. not determined; result in brackets: qualitatively the data indicate the corresponding model). The results from structural destruction of chromatin by ion-irradiation leading to DNA fragment distributions were done by P. Quicken, Institute for Ray-Biology, Ludwig-Maximilian University,Munich, using the simulated configurations of single chromosomes.

Measurement Category	Parameter	Simulation		Experimental Data Available		Comparison Result
		Chromosomes [-,+]	Nuclei [-,+]	[N/Y]	Method (v)	
Distances	Distances between nearest MLS subcompartments	++	+++	Y	FISH, BrdU (v)	MLS
	Distance between genetic markers	++	+++	Y	FISH	MLS
	Distances between genetic markers in ensemble	++	+++	Y	FISH	MLS
Scaling Properties	Exact spatial-distance dimension of chromatin backbone	++	+++	Y	FISH	MLS
	Exact yard-stick dimension of chromatin fiber backbone	++	+++	N (N)	-	n. d.
	Box-counting dimension single chromosomes	++	+++	N (Y)	(FISH, BrdU (v))	n. d. (MLS)
	Weighted box-counting dimension in CLSM Images	-	+++	Y	FISH, BrdU (v), His-AFP (vv)	n. d. (MLS)
	Weighted lacunarity dimension in CLSM Images	-	+++	Y	FISH, BrdU (v), His-AFP (vv)	n. d. (MLS)
	Weighted local dimension in CLSM Images	-	+++	Y	FISH, BrdU (v), His-AFP (vv)	n. d. (MLS)
Structural Destruction	DNA fragment distribution	+++	+++	Y	carbon-ion irradiation	MLS
Dynamics	Diffusion of particles	-	+++	Y	fluorescent dyes, Dextranes, AFP	MLS non-ICD
	Diffusion of chromosomes	+	+++	N(Y)	BrdU (v), His-AFP (vv)	n. d.
	Diffusion of MLS subcompartments	+	+++	Y	BrdU (v), His-AFP (vv)	n. d.
	Diffusion of chromatin loops	++	+++	N (Y)	BrdU (v), His-AFP (vv)	n. d.

structure and dynamics of histones, nucleosomes, chromatin, chromosomes and whole nuclei during cell cycle, differentiation, and apoptosis could be investigated *in vivo*. The interphase morphology showed globular structures as predicted by the Multi-Loop-Subcompartment model. All stages of mitosis as well as apoptosis were clearly distinguishable. Deacetylase inhibitors led to a smoothing of the interphase morphology. With this *in vivo* chromatin label the interphase morphology and changes thereof could be investigated by quantitative scaling and statistical analyses. The technique could also be applied for cell culture control and counterstaining, or *in organo* and *in organismo* by creation of transgenic animals.

This now widely used technique of chromatin labelling by histon-autofluorescent protein fusions led to the discovery of construct conversions in simultaneous co-transfections, a convenient and widely used approach in multicolour labelling

experiments *in vivo*, using green fluorescent proteins with distinct spectral characteristics. These co-transfections can cause false positive results due to conversion of their spectral properties. Standard transfection result in ~8%, depending on the treatment of the DNA up to 26%, of the cells expressing altered fusion proteins. This could lead to severe misinterpretation of the results. The conversion is independent of the transfection method and the cell type. The results show that conversion is based on homologous recombination/repair/replication (RRR) events occurring between the nucleotide sequences of the fluorescent proteins. Conversion can be avoided by consecutive transfection or by fluorescent constructs with low sequence similarities. The appearance of conversion makes it possible to easily exchange spectral properties in fusion proteins, to create libraries or to assemble DNA fusion constructs *in vivo*. The detailed quantification of the conversion rate could be used to investigate RRR processes in general.

9.2 Synthesis

The cell nucleus can be viewed as an optimized bioreactor in which the sequential and three-dimensional organization coevolved:

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

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

Three - Dimensional Organization of the Human Interphase Nucleus
Experiments compared to Simulations

Tobias A. Knoch, Christian Münkel, Waldemar Waldeck and Jörg Langowski [1]

in collaboration with J. Rauch, H. Bornfleth and C. Cremer [2] I. Solovei and T. Cremer [3] P. Quicken, A. Friedel and A. Kellerer [4]

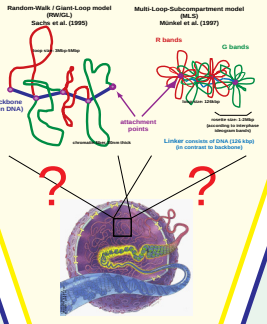
[1] Division Biophysics of Macromolecules, German Cancer Research Center, Im Neuenheimer Feld 280
D - 69120 Heidelberg, Federal Republic of Germany
http://www.DKFZ-Heidelberg.de/Macromol/Welcome.html

[2] Institute for Applied Physics, Albert Ueberle Str. 3-5, 69120 Heidelberg, FRG
[3] Institute for Anthropology and Human Genetics, Richard Wagner Str. 10, 8033 Munich, FRG
[4] Strahlenbiologisches Institut, Schillerstr. 42, 80336 München, FRG and GSF, Ingolstädter Landstr. 1, 85764 Neuherberg, FRG

German Human Genome Project

INTRODUCTION

Despite the succesful linear sequencing of the genome its three dimensional structure is widely unknown although it is important for gene regulation and replication. With a comparison between experiments and simulations we show here an interdisciplinary approach leading to the determination of the three- dimensional organization of the human genome.



FISH

Fluorescence in situ hybridization (FISH) is used for the specific marking of chromosome arms (Fig. 1A) and pairs of small chromosomal DNA regions (Fig. 1B). The labeling is visualized with confocal laser scanning microscopy followed by image reconstruction procedures. Chromosome arms show only small overlap and globular substructures as predicted by the MLS-model (Fig. 1A & 7A). A comparison between simulated and measured spatial distances between genomic regions as function of their genomic distances results in a good agreement with the MLS-model having loop sizes of arround 126 kbp and linker sizes between 63 kbp and 126 kbp (Fig. 2).

Fig. 1A & 1B: FISH-images of a territory painting of chromosome 15 (left, 1A) and genomic markers YAC-48 and YAC60 (right 1B) with a genomic separation of 1.0 Mbp in interphase of fibroblast cells.

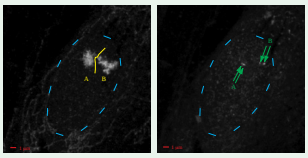
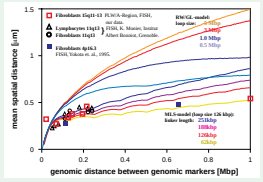


Fig. 2: Comparison of the RW/GL- and the MLS-model with experimentally determined interphase distances.



SIMULATION

For the prediction of experiments we simulated various models of human interphase chromosome 15 with Monte Carlo and Brownian Dynamics methods. The chromatin fiber was modelled as a flexible polymer fiber. Only stretching, bending and excluded volume interactions are considered. Chromosomes are further confined by a spherical potential representing the surrounding chromosomes or the nuclear membrane. Only the MLS model leads to clearly distinct functional and dynamic subcompartments in agreement with experiments (Fig. 7B & 1A) in contrast to the RW/GL models where big loops are intermingling freely and featureless (Fig. 7C & 7D).

Fig. 7A: Starting configuration with the form and size of a metaphase chromosome.

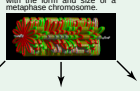


Fig. 7B: MLS model with 126 kbp loops and linkers.

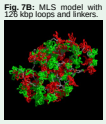


Fig. 7C: RW/GL model with 126 kbp loops.

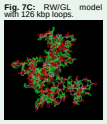


Fig. 7D: RW/GL model with 5 Mbp loops.

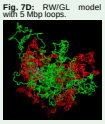
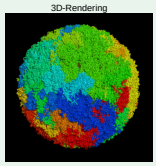
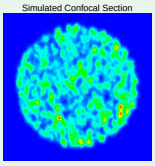


Fig. 8A & 8B: Simulation of a human interphase nucleus containing all 46 chromosomes with 1,200,000 polymer segments. The MLS-model leads to the formation of distinct and non-overlaping chromosome territories.

3D-Rendering



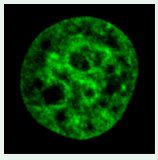
Simulated Confocal Section



CONCLUSION

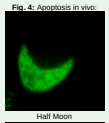
Simulations of chromsomes and the whole cell nucleus show that only the MLS-model leads to the formation of non-overlapping chromosome territories and distinct functional and dynamic sub-compartments. Spatial distances between FISH labeled pairs of genomic markers as function of their genomic distance result in a MLS-model with loop sizes of 120 kbp and linker sizes of 63 to 126 kbp. With the developement of GFP-fusionproteins it is possible to study the chromatin distribution and dynamics resulting from cell cycle, treatment by chemicals or radiation in vivo. The chromatin distribution is similar to those found in the simulation of whole cell nuclei. Fractal analysis of the simulations reveal the multifractality of chromosomes in agreement with porous network research. It is possible to quantify the in vivo chromatin distribution with fractal analysis and to relate the result to differences in morphology. The simulation of fragment distributions based on double strand breakage after carbon-ion irriadiation differs in different models. Here again a comparison to experiments favours a MLS-model.

CHROMATIN LABELING in vivo

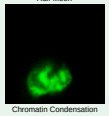


For the investigation of the structure and dynamics of the chromatin distribution in vivo a fusion protein between a chromatin associated protein and an auto-fluorescent protein (AFP) was developed and stably expressed in cells. Changes in the chromatin distribution due to the cell cycle (Fig. 5), differentiation, chemicals (Fig. 4) or radiation are now possible without fixation and staining and therefore artefactfree and time saving. The structures visible in confocal images of normal interphase nuclei (Fig. 3: nucleus left with an 20 µm image sidelength) are similar to those found in simulated confocal images of the MLS-model (Fig. 8B). The structures can be analyzed quantitatively by global and local fractal analysis (Fig. 10) and hence can be linked to the detailed folding of the chromatin fiber.

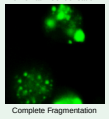
Fig. 4: Apoptosis in vivo:



Half Moon

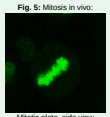


Chromatin Condensation

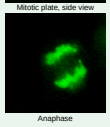


Complete Fragmentation

Fig. 5: Mitosis in vivo:



Mitotic plate, side view



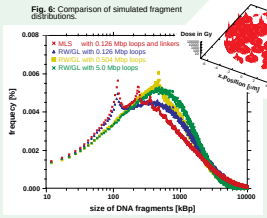
Anaphase

Sidelength of all images: 40 µm

CARBON - ION IRRADIATION

The length distribution of DNA fragments after irradiation with carbon ions and the sites of double strand breakage depend on the spatial arrangement of the 30 nm chromatin fiber in the nucleus. Simulated configurations of different chromosome models were taken as the basis for the detailed simulation of these fragment distributions. The RW/GL-model and the MLS-model lead to clearly distinct fragment distributions (Fig. 6). A comparison with experiments favours an MLS-model. The specifity of breakage sites is currently analyzed. Data by P. Quicken.

Fig. 6: Comparison of simulated fragment distributions.



Radiation dose distribution in a cell nucleus.

FRACTAL ANALYSIS

Fractal analysis is especially suited to quantify the unordered and non-euclidean chromatin distribution of the nucleus. The dynamic behaviour of the chromatin structure and the diffusion of particles in the nucleus are also closely connected to the fractal dimension. The fractal analysis of the simulation of chromosome 15 lead to multifractal behaviour in agreement with porous network research (Fig. 9). Therefore chromosome territories show a higher degree of determinism than previously thought. First tests of fractal analysis of chromatin distributions in vivo result in significant differences for different morphologies (Fig. 10) and might favour an MLS-model like chromatin distribution.

Fig. 9: Comparison of RW/GL- and MLS- model with fractal dimension of the chromatin fiber from simulations.

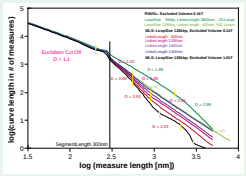
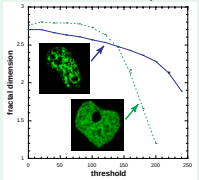


Fig. 10: Fractal Dimension as function of the intensity threshold.



Grants and Prizes

Grants

Dissertation grant of the German Cancer Research Center (DKFZ) to conduct this thesis, Heidelberg, Germany, 1998-2001.

Grant for computer resources: Structure and Dynamics of Chromosomes in the Human Cell Nucleus. Machine IBM-SP2, Supercomputing Center (SCC) Karlsruhe, University of Karlsruhe, Karlsruhe, Germany, 1999-2001.

Prizes

Student Travel Award of the Molecular Graphics and Modelling Society for the talk: Three-dimensional organization of chromosome territories and the human interphase cell nucleus - simulations and experiments. *Molecular Modelling in the LARGE - Bridging scales in space, time and complexity*, Molecular Graphics and Modelling Society, 17th International Meeting, San Diego Paradise Point Resort, San Diego, California, USA, 6. - 10. December, 1998.

Student Travel Award of the German Cancer Research Center (DKFZ) for the poster: Three-dimensional organization of the human interphase nucleus - experiments compared to simulations. *Posterpresentation of Scientific Studies from Diploma- and PhD- Students*. German Cancer Research Center (DKFZ), Heidelberg, 10. - 14. January, 1999.

Poster 1 Student Travel Award Winning Poster, DKFZ, 1999.

Three - Dimensional Organization of Chromosome Territories in the Human Interphase Cell Nucleus

Tobias A. Knoch, Christian Münkel and Jörg Langowski [1)]

Joachim Rauch, Harald Bornfleth and Christoph Cremer [2)] **with Irina Solovei and Thomas Cremer** [3)]

[1)] Division Biophysics of Macromolecules, German Cancer Research Center, Im Neuenheimer Feld 280
D - 69120 Heidelberg, Federal Republic of Germany
http://www.DKFZ-Heidelberg.de/Macromol/Welcome.html

[2)] Instiute for Applied Physics, Albert Ueberle Str. 3-5, 69120 Heidelberg, FRG

[3)] Institute for Anthropology and Human Genetics, Richard Wagner Str. 10, 8033Munich, FRG

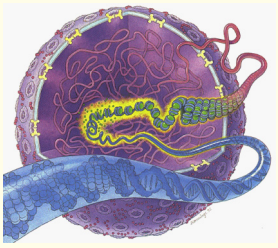
German Human Genome Project

PURPOSE

The eukaryotic cell is a prime example of a functioning nano machinery. The synthesis of proteins, maintenance of structure and duplication of the eukaryotic cell itself are all fine-tuned biochemical processes that depend on the precise structural arrangement of the cellular components. The regulation of genes – their transcription and replication - has been shown to be connected closely to the three-dimensional organization of the genome in the cell nucleus. Despite the successful linear sequencing of the human genome its three-dimensional structure is widely unknown.

With the simulation of chromosomes and cell nuclei in comparison with fluorescence in situ hybridization we show here an approach leading to the detailed determination of the three-dimensional organization of the human genome:

Best agreement between simulations and experiment is reached for a Multi-Loop-Subcompartment model, thus the human genome shows a higher degree of determinism than previously thought.

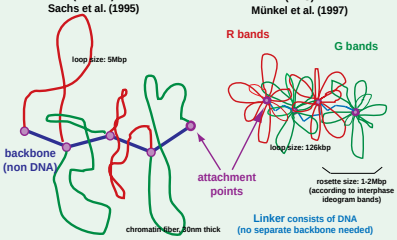


Typical textbook illustration of the human cell nucleus:
1) human cell nuclei differ from spherical shape,
2) the DNA is not a closed pipe,
3) nucleosomes are not regularly organized into chromatin,
4) chromatin does not float around randomly in the nucleus.
V. Hennings (illustrator) in Molecular and Cellular Biology by Stephen L. Wolfe, 1993.

SIMULATIONS

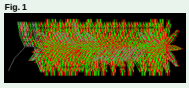
Random Walk / Giant Loop model (RW/GL)
Sachs et al. (1995)

Multi-Loop-Subcompartment model (MLS)
Münkel et al. (1997)



With **Monte Carlo** and **Brownian Dynamics** methods we simulated various models (see sketch left) of **human interphase chromosome 15** assuming a flexible polymer chain. To save computer power we start with ~3,500 300nm=31kbp and later we relax with ~21,000 50nm=5,2 kbp long segments. For simulation of a single chromosome it is placed in a potential well whose height is related to the excluded volume interaction (EVI). The EVI keeps the chain from self crossing. Starting configurations have the approximate form and size of a metaphase chromosome (Fig. 1) from which the following decondensation into interphase resembles the natural process.

Fig. 1



Rosettes in the Multi-Loop-Subcompartment model correspond to the size of chromosomal interphase band domains.
For **simulation of a whole interphase nucleus** 46 methaphase chromosomes are placed randomly in a nucleus confined by an EVI. The simulations are made on two IBM SP2 parallel computers with 80 and 512 nodes.

Fig. 2
Ray traced image of the Random-Walk/Giant-Loop model, loop size 5Mbp, after ~80.000 Monte-Carlo and 1000 relaxing Brownian Dynamics steps. Large loops intermingle freely thus forming no distinct features like in the MLS model.

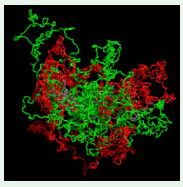


Fig. 3
Ray traced image of the Multi-Loop Subcompartment model, loop size 126kbp, linker size 126 kbp, after ~50.000 Monte-Carlo and 1000 relaxing Brownian- Dynamics steps. Here rosettes form subcompartments as separated organizational and dynamic entities.

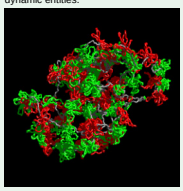
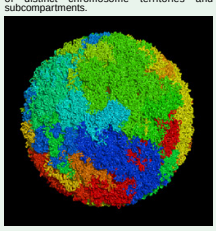
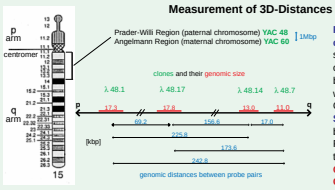


Fig. 4
Simulation of a **human interphase cell nucleus** with all 46 chromosomes with 1,200,000 polymer segments after 0.5s Brownian Dynamics simulation with 10µs steps. The MLS-model leads to the formation of distinct chromosome territories and subcompartments.

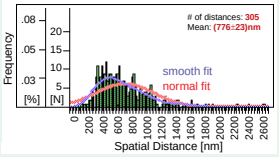


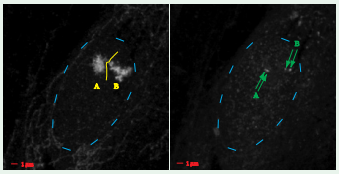
EXPERIMENTS

Measurement of 3D-Distances between Genomic Markers



Distance Distribution **Fig. 5**





Fluorescence in situ hybridization (FISH) in connection with **confocal light microscopy** is used for the specific marking of small chromosomal DNA regions. Despite the low spatial resolution of FISH, it is possible to interprete the results (f. e. the 3D distance between genetic markers as a function of their genomic distance) with our simulations.
Chromosome 15 and the **Prader-Labhard-Willi/Angelmann Syndrom** region was chosen, because the genomic distance between markers is well known (see sketch left) and because the PLW/A-syndrom is a candidate for structure mutation (in contrast to the common base pair mutation).
Collaboration with B. Horsthemke, Institute for Human Genetics, Essen, FRG.

Methods: Human fibroblast cells grown on coverslips to confluent layers and being assumed to rest now in the same cell cycle phase are fixed in isotonic environment with paraformaldehyde.
For Hybridization we use digoxigenin labeled DNA probes. The probes are detected with fluorescent dyes.

Confocal image series were taken with a Leica TCS NT confocal microscope with an axial displacement of Δz = 200nm. The images are median and background filtered. After manual threshold determination from an extended focus view (Fig. 6) for spot finding we proceed with image reconstruction specially adjusted to the microscope. Finally the **3D-spatial distances** are determined between the centers of mass of the spots (Fig. 5). The experimental distance distributions are then compared to the computed ones. We use a workstation cluster of 10 Silicon Graphics Indigo and Indigo II for computation.

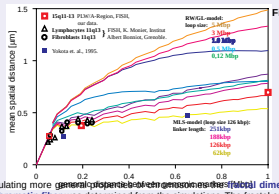
The fibroblast nuclei are found to have their in vivo size (~20µm * 10µm * 6µm) so that we conclude that at least on the micrometer length scale we preserved the nuclear structure. With two colour FISH it is possible to detect 3D-distances below the optical resolution.

Fig. 6
Chromosomes form distinct territories in interphase and genomic markers lay clearly separable within the territories.
Left: Territory painting by FISH of chromosome 15; by chance the

CONCLUSION

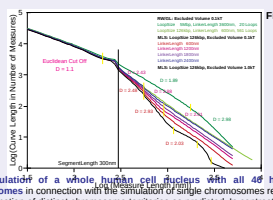
Best agreement between simulations and experiments is reached for a **Multi-Loop-Subcompartment model** with a loop size of roughly126kbp and a linker length of 1,200nm. Supposed that defined loop bases exist it might be possible to determine the mean positioning of genes relative to each other.

Fig. 7



For calculating more general shape descriptors the fractal dimension of the chromatin fiber was determined from the simulations. The fractal analysis resulted in multifractal behaviour (data not shown here) in good agreement with predictions drawn from porous network research (Avnir, 1989; Mandelbrot, private communications).

Fig. 8



The simulation of a whole human cell nucleus with all 46 human chromosomes in connection with the simulation of single chromosomes resulted in the formation of distinct chromosome territories as predicted. In contrast to the RW/GL-model the MLS-model leads to low overlap between chromosome territories as well as chromosome arms, in agreement with overlap analysis of confocal image series (data not shown here).

Patents and Publications Related to this Thesis

Patents

Knoch, T. A., Waldeck, W., Müller, G., Alonso, A. & Langowski, J. DNA-Sequenz und Verfahren zur *in vivo* Markierung und Analyse von DNA/Chromatin in Zellen. *German Patent Application 10013204.9-44* and *International Patent Application PCT/DE01/01044*.

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

Publications

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*

Poster 2 Poster Shown on Various Conferences, 1998.

in Science and Engineering 1999, editors Egon Krause and Willi Jäger, 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.

Westphal, G., van den Berg-Stein, S., Braun, K., **Knoch, T. A.**, Dümmerling, M., Langowski, J., Debus, J. & Friedrich, E. Detection of the NGF receptors TrkaA and p75NTR and effect of NGF on the growth characteristics of human tumor cell lines. *J. Exp. Clin. Canc. Res.* 21(2), 255-267, 2002.

Westphal, G., Niederberger, E., Blum, C., Wollman, Y., **Knoch, T. A.**, Dümmerling, M., Rebel, W., Debus, J. & Friedrich, E. Erythropoietin Receptor in Human Tumor Cells: Expression and Aspects Regarding Functionality. *Tumori* 88(2), 150-159, 2002.

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

***:** Shared first authorship due to equal contribution.

Diffusion and transport in the human interphase cell nucleus
FCS experiments compared to simulations

Malte Wachsmuth, Tobias A. Knoch, Waldemar Waldeck, and J rg Langowski

Division Biophysics of Macromolecules, German Cancer Research Center,
Im Neuenheimer Feld 280, D - 69120 Heidelberg, Germany
http://www.dkfz-heidelberg.de/Macromol/Welcome.html

FCS experiments

Basics

Mean square displacement (MSD) of a free Brownian particle:

$$\left\langle \mathbf{r}(t)^2 \right\rangle = 6D_0 t, \qquad D_0 = \frac{kT}{6\pi\eta r_h} \qquad (1), (2)$$

(D_0: diffusion coefficient, r_h: hydrodynamic radius, η: viscosity).
Mean square displacement in the presence of obstacles:

$$\left\langle \mathbf{r}(t)^2 \right\rangle = 6D(t)\, t \propto t^{1/d_w} \qquad (3)$$

This behaviour is called obstructed diffusion. The anomaly parameter d_w characterizes the time-dependent diffusion coefficient $D(t)$ and equals 2 for free diffusion. It increases with increasing obstacle concentration and depends strongly on geometric properties like the obstacle size or the fractal dimensions of the distribution.
If the obstacles form cages, dead-ends, or cavities, molecules can be trapped, resulting in an apparently slowly and a freely diffusing fraction of molecules.

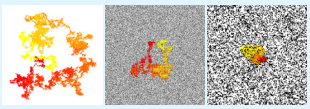
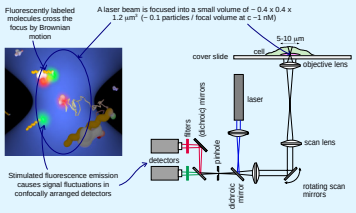


Fig. 1: Simulated random walk of 10^6 steps on an empty rectangular 1500x1500 site lattice (left); the path colour is changed from red to yellow with time. In the presence of statistically distributed obstacles with a density of 35% and a size of 2x2 (middle) or 8x8 sites (right), respectively, the area covered by the random walk gets smaller and shows different "compactnesses" for different obstacle geometries.

Fluorescence correlation spectroscopy



A computer calculates the autocorrelation function (ACF) of the detector signals. The concentration and the diffusion coefficient of the molecules can be derived.
The excitation and detection path are coupled via a scanning unit into a conventional inverted microscope (Olympus IX70), providing a diffraction limited focus and a corresponding spatial resolution. The compact FCS/scanning module can be easily attached to the video port of the microscope and shows a high optical and mechanical stability.

Diffusion scans

		two components	obstructed diffusion	number of cells
AT-1:	$D_{\text{monomer, wt}}/D_{\text{monomer, cell}}$	5.3 – 0.6	5.5 – 0.6	9
	$D_{\text{monomer, cell}}/D_{\text{flu-prot., cell}}$	1.4 – 0.2	1.5 – 0.4	6
COS-7:	$D_{\text{monomer, wt}}/D_{\text{monomer, cell}}$	4.7 – 0.5	4.5 – 0.7	8
	$D_{\text{monomer, cell}}/D_{\text{flu-prot., cell}}$	1.2 – 0.1	1.3 – 0.3	4

Table 1: The diffusion coefficient and therefore the viscosity sensed by eGFP and the fusion protein is about 5 times higher than in water in AT-1 cells and in COS-7 cells. This holds for nuclei as well as for the cytoplasm, implying that the nuclear "solvent" is similar to the cytosol.

From FCS data along straight lines in AT-1 and COS-7 cells expressing eGFP or an eGFP-b-galactosidase fusion protein (Fig. 2) we obtain a deviation from ideal diffusion especially in the nucleus.
Interpretation as a fast moving fraction everywhere in the cell and a slower one mainly in the nucleus: only an inhomogeneous chromatin distribution with high local densities leads to trapping and subsequent observation of two distinct fractions.
Applying the obstructed diffusion model: diffusion obstruction is found mainly in the nucleus. Even low obstacle densities lead to a remarkable deviation from free diffusion.

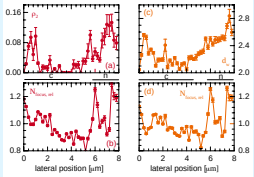
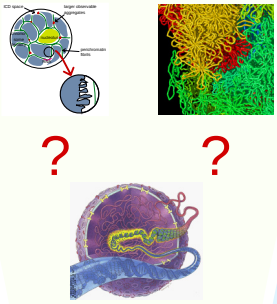


Fig. 2: FCS scan through a COS-7 cell expressing the fusion protein: (a) fraction of a slow component and (b) relative number of molecules, found with the two component free diffusion model; (c) anomaly parameter and (d) relative molecule number from the obstructed diffusion model, as a function of the position in the cell ("c" - cytoplasm, "n" - nucleus).

Introduction

Despite the succesful linear sequencing of the human genome the three-dimensional arrangement of chromatin, functional, and structural components is still largely unknown. Molecular transport and diffusion are important for processes like gene regulation, replication, or repair and are vitally influenced by the structure.
With a comparison between fluorescence correlation spectroscopy (FCS) experiments and simulations we show here an interdisciplinary approach for the understanding of transport and diffusion properties in the human interphase cell nucleus.



Conclusion

FCS in combination with a scanning device is a suitable tool to study the diffusion characteristics of fluorescent proteins in living cell nuclei with high spatial resolution. Computer simulations of the three-dimensional organization of the human interphase nucleus allows a detailed test of theoretical models in comparison to experiments. Diffusion and transport in the nucleus are most appropriately described with the concept of obstructed diffusion. A large volume fraction of the nucleus seems to contain a cytosol-like liquid with an apparent viscosity 5 times higher than in water. The geometry of particles and structure as well as their interactions influence the mobilities in terms of speed and spatial coverage. A considerable amount of genomic sites is accessible for not too large particles. FCS experiments and simulations based on the polymer model are in a good agreement. Using recently developed in vivo chromatin markers, a detailed study of mobility vs. structure is subject of current work.

Diffusion vs. structure

The diffusion of particles in living interphase nuclei depends on the local structure. The development of in vivo chromatin markers allows to investigate this relation using FCS. The correlation between diffusion obstruction and structure vanishes for small particles and probably increases with increasing particle size (Fig. 2, 3, 7).

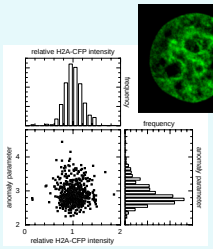


Fig. 3: The degree of diffusion obstruction plotted against the chromatin density, represented by the H2A-CFP fluorescence intensity. Data from FCS of Alexa568 dye in LCLS103H cell nuclei stably expressing a H2A-CFP fusion protein.

Simulations

For the prediction of experiments we simulated various models of human interphase chromosome 15 with Monte Carlo and Brownian Dynamics methods. The chromatin fiber was modelled as a flexible polymer. Only stretching, bending and excluded volume interactions are considered. Chromosomes are further confined by a spherical potential representing the surrounding chromosomes or the nuclear membrane. Only the rosette-like MLS model leads to clearly distinct functional and dynamic subcompartments in agreement with experiments (Fig. 4B) in contrast to the RW/GL models where big loops are intermingling freely and featureless (Fig. 4C & 4D).

Fig. 4A: Starting configuration with the form and size of a metaphase chromosome.

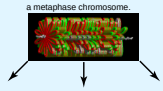


Fig. 4B: MLS model with 126 kbp loops and

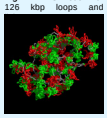


Fig. 4C: RW/GL model with 126 kbp loops.

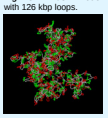


Fig. 4D: RW/GL model with 5 Mbp loops.

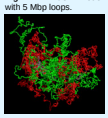
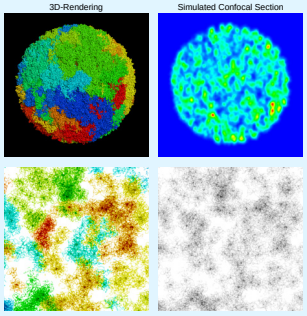


Fig. 5A - 5D: Simulation of a human interphase nucleus containing all 46 chromosomes with 1,200,000 polymer segments. The MLS-model leads to the formation of distinct and non-overlaping chromosome territories.



The diffusion of spherical particles with radius r_s in a nucleus is simulated using Brownian Dynamics methods. The mean square displacement of the particles depends on r_s, the radius of the nucleus, i.e. the obstacle concentration, and also critically on the interaction between particles and structure (Fig. 6 & 7). The results agree with theoretical expectations as well as with FCS experiments (Table 1).

Fig. 6: Comparison of mean square displacements of particles with different radius r_s in nuclei with 3 m (g, y) and 6 m (r,b) radius and low (r,g) and high (b, y) interaction.

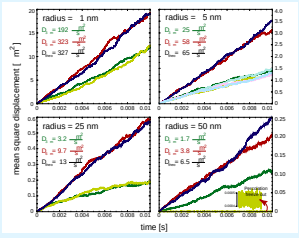
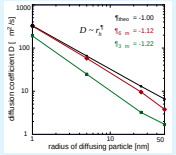


Fig. 7: Apparent diffusion coefficient D as a function of the particle radius r_s for different nuclear radii.



Literature

Abney, J. R., Cutler, B., Fillbach, M. L., Axelrod, D. & Scalettar, B. A. Chromatin dynamics in interphase nuclei and its implications for nuclear structure. *J. Cell. Biol.* 137(7), 1459, 1997.

Adolph, K. W. & Kreisman, L. R. Surface structure and isolated metaphase chromosomes. *Exp. Cell Res.* 147, 155-166, 1983.

Aharony, A. Percolation, fractals and anomalous diffusion. *J. Stat. Phys.* 34(5), 931-939, 1984.

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

Alexander, S. & Orbach, R. Density states on fractals: "fractons". *J. Physique - Lettres* 43, L625-L631, 1982.

Allegrini, P., Buiatti, M., Grigolini, P. & West, B.J. Fractional Brownian motion as a nonstationary process: an alternative pradigm for DNA sequences. *Phys. Rev. E* 57(4), 4558-4567, 1998.

Allen, M. P. & Tildesley, D. J. Computer simulations of liquids. *Oxford University Press, Oxford*, 1989.

Allison, D. C. & Nestor, A. L. Evidence for a relatively random array of human chromosomes on the mitotic ring. *J. Cell Biol.* 145, 1-14, 1999.

Amato, I. DNA shows unexpected patterns writ large. *Science* 257, 747, 1992.

Ambrose, C. Lowman, H., Rajadhyaksha, A., Blasquez, V. & Bina, M. Location of nucleosomes in Simian Virus 40 chromatin. *J. Mol. Biol.* 214, 875-884, 1990.

Anderson, R. A., & Eliason, S. L. Recombination of homologous DNA fragments transfected into mammalian cells occurs predominantly by terminal pairing. *Mol. Cell. Biol.* 6, 3246–3252, 1986.

Andreason, G. L. & Evans G. A. Introduction and expression of DNA molecules in eukaryotic cells by electroporation. *Biotechniques* 6, 650–660, 1988.

Audit, B., Thermes, C., Vaillant, C., d'Aubenton-Carafa, Y., Muzy, J. F. & Arneodo, A. Long-range correlations in genomic DNA: A signature of the nucleosomal structure. *Phys. Rev. Lett.* 86 (11), 2471-2474, 2001.

Audit, B., Vaillant, C., Arneodo, A., d'Aubenton-Carafa, Y. & Thermes, C. Long-range correlations between DNA bending sites: relation to the structure and dynamics of nucleosomes. *J. Mol. Biol.* 316, 903-918, 2002.

Avramova, Z., Petrov, P. & Tsanev, R. DNA-folding by a stably DNA-linked protein in eukaryotic chromatin. *Nuclear Structure and Function. by Harris, J. R. & Zbarsky, I. B. (Editors)*, Plenum Press, New York - London, 337-339, 1990.

Poster 3 Poster together with Malte Wachsmuth, 2001.

Avnir, D. (Ed.) The fractal approach to heterogeneous chemistry - surfaces, colloids, polymers. *John Wiley & Sons (Chichester, New York, Brisbane, Toronto, Singapore)*, ISBN 0-471-91723-0, 1992.

Baake, M., Doenecke, D. & Albig, W. Characterisation of nuclear localisation signals of the four human core histones. *J. Cell. Biochem.* 81, 333-346, 2001.

Bak, A. L. Zeuthen, J. & Crick, F. H. C. Higher-order structure of human mitotic chromosomes. *Proc. Natl. Acad. Sci. USA* 79(4), 2541-2545, 1979.

Bailey, K.A., Pereira, S. L., Widom, J. & Reeve, J. Archael histone selection of nucleosome positioning sequences and the procaryotic origin of histone-dependent genome evolution. *J. Mol. Biol.* 303, 25-34, 2000.

Baudy, P. & Bram, S. Chromatin fiber dimensions and nucleosome orientation: a neutron scattering investigation. *Nucleic. Acids. Res.* 5 (10), 3697-3714, 1978.

Baudy, P. & Bram, S. Neutron scattering on nuclei. *Nucleic. Acids. Res.* 6 (4): 1721-1729, 1979.

Baumgärtner, A. & Binder, K. Monte carlo studies on the freely jointed polymer chain with excluded volume interaction. *J. Chem. Phys.* 71(6), 2541-2545, 1979.

Beaucage, G. Approximations leading to a unified exponential/power-law approach to small angle scattering. *J. Appl. Cryst.* 28, 717-728, 1995.

Beaucage, G. Small-angle scattering from polymeric mass fractals of arbitrary mass-fractal dimension. *J. Appl. Cryst.* 29, 134-146, 1996.

Bednar, J. & Woodcock, C. L. Cryoelectron microscopic analysis of nucleosomes and chromatin. *Methods. Enzymol.* 304, 191-213, 1999.

Belmont, A. S. , Kendall, F. M. & Nicolini, C. Three-dimensional intranuclear DNA organization in situ: three states of condensation and their redistribution as a function of nuclear size near the G1-S border in HeLa S-3 cells. *J. Cell. Sci.* 65, 123-138, 1984.

Belmont, A. S., Braunfeld, M. B., Sedat, J. W. & Agard, D. A. Large-scale chromatin structural domains within mitotic and interphase chromosomes *in vivo* and *in vitro*. *Chromosoma* 98, 129-143, 1989.

Belmont, A. S. & Bruce, K. Visualization of G1 Chromosomes: A Folded, Twisted, Supercoiled Chromonema Model of Interphase Chromatid Structure. *J. Cell Biol.* 127 (2), 287-302, 1994.

Benyajati, C. and Worcel, A. Isolation, characterization and strucutre of the folded interphase genome of Drosophilia melangoster. *Cell* 8, 393-407, 1976.

Benyajati, C., Spoerel, N., Haymerle, H. & Ashburner, M. The messenger RNA for alcohol dehydrogenase in Drosophila melanogaster differs in its 5' end in different developmental stages. *Cell* 33, 125-133, 1983.

Berezney, R., Mortillaro, M. J., Ma, H., Wei, X. & Samarabandu, J. The nuclear matrix: a structural milieu for genomic function. *Int. Rev. Cytol.*, 1-65, 1995.

Berger, J. M. Structure of DNA topoisomerases. *Biochim. Biophys. Acta.* 1400 (1-3), 3-18, 1998.

Bishop, J. O. Chromosomal insertion of foreign DNA. *Reprod. Nutr. Dev.* 36, 607–618, 1996.

Blank, A. & Becker, P. B. The effect of nucleosome phasing sequences and DNA topology on nucleosome spacing. *J. Mol. Biol.* 260, 1-8, 1996.

Bloom & Fawcett. A textbook of hisotology. *Chapman and Hall*, 12th edition.

Bornfleth, H., Sätzler, K., Eils, R. & Cremer, C. High-precision distance measurements and volume-conserving segmentation of objects near and below the resolution limit in three-dimensional confocal fluorescence microscopy. *J. Microscopy* 189(2), 2871-2886, 1998.

Bornfleth, H. Precision distance microscopy and model-based segmentation for the analysis of the functional topology of the cell nucleus. *PhD thesis*, Faculty for Physics and Astronomy, Ruprecht Karls University of Heidelberg, 1999a.

Bornfleth, H., Edelmann, P., Zink, D., Cremer, C. & Cremer, T. Quantitative motion analysis of subchromosomal foci in living cells using four-dimensional microscopy. Biophys. J. 77, 2771-2886, 1999b.

Borovik, A. S., Grosberg, A. Y. & Frank-Kamenetskii, M. D. Fractality of DNA Texts. *J. Biomol. Structure* 12 (3), 655-669, 1994.

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

Brickmann, J. & Bär H.-J. Chaos und fractale dimension. *Nachr. Chem. Tech. Lab.* 34, 566-572, 1986.

Brigder, J. M., Herrmann, H., Münkel, C. & Lichter, P. Identification of an interchromosomal compartment by polymerization of nuclear-targeted vimentin. *J. Cell Sci.* 111 (9), 1241-1253, 1998.

Brock, R., Hink, M. A. & Jovin, T. M. Fluorescence correlation microscopy of cells in the presence of autofluorescence. *Biophys. J.* 75(5), 2547-, 1998.

Brock, R., Vamosi, G., Vereb, G. & Jovin, T. M. Rapid characterization of green fluorescent protein fusion proteins on the molecular and cellular level by fluorescence correlation microscopy. *Proc. Nat. Acad. Sci. USA* 96(18), 10123-, 1999.

Buldyrev, S. V., Goldberger, A. L., Havlin, S., Mantegna, R. N., Matsa, M.E., Peng, C.-K., Simons, M. & Stanley, H. E. Long-range correlation properties of coding and noncoding DNA sequences: GenBank analysis. *Phys. Rev. Lett.* 51 (5), 5084-5091, 1993.

Buldyrev, S. V., Goldbergerger, A. L., Havlin, S., Mantegna, R. N., Matsa, M. E., Peng, C.-K., Simons, M. & Stanley, H.E. Long-Range correlation properties of coding and noncoding DNA sequences: GenBand analysis. *Phys. Rev.* 51(5), 5084-5091, 1995.

Bunde, A. & Havlin, S. (Eds) Fractals and Disordered Systems. 2nd ed, *Springer-Verlag (Berlin, Heidelberg, New York)*, ISBN 3-540-56219-2, 1996.

Byng, J. W., Yaffe, M. J., Lockwood, G. A., Little, L. E., Tritchler, D. L. & Boyd, N. F. Automamographic analysis of mammographic densities and breast carcinoma risk. *Cancer* 80(1), 66-74, 1997.

Camper, S. A., Saunders, C. E., Kendall, S. K., Keri, R. A., Seasholtz, A. F., Gordon, D. F., Birkmeier, T. S., Keegan, C. E., Karolyi, I. J. Roller, M. L. *et al.* Implementing transgenic and embryonic stem cell technology to study gene expression , cell-cell interactions and gene function. *Biol. Reprod.* 52 (2), 246-257, 1995.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. & Prasher, D. Green fluorescent protein as a marker for gene expression. *Science* 263, 802–805, 1994.

Camargo, M. & Cervenka, J. Patterns of DNA replication of human chromosomes. II. Replication map and replication model. *Am. J. Hum. Genet.* 34, 757-780, 1982.

Cambell, R. D. Describing the shapes of fern leaves: a fractal geometrical approach. *Acta Biotheoretica* 44, 119-142, 1996.

Chatzidimitriou-Dreismann, C. A. & Larhammar, D. Long-range correlations in DNA. *Nature* 361, 212-213, 1993.

Chirico, G. & Langowski, J. Calculating the hydrodynamic properties of DNA through a second-order Brownian dynamics algorithm. *Macromolecules* 25, 769-775, 1994.

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

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

Compagni, A. & Christofori, G. Recent advances in research on multistage tumorgenesis. *Br. J. Cancer* 83 (1), 1-5, 2000.

Constanzi, C. & Pehrson, J. R. Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature* 393, 599-601, 1998.

Cook, P. R., Brazell, I. A. & Jost, E. Characterization of nuclear structures containing superhelical DNA. *J. Cell Sci.* 22, 303-324, 1976.

Cormack, B. P., Valdivia, R. H. & Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* 173, 33–38, 1996.

Cox, B. L. & Wang, J. S. Y. Fractal surfaces measurement and applications in the earth sciences. *Fractals* 1(1), 87-115, 1993.

Craig, J. M. Bickmore, W. A. Chromosome bands - flavours to savour. *BioEssays* 15, 349-354, 1993.

Cremer, C., Zorn, C. & Cremer, T. An ultraviolet microbeam for 257nm. *Microskopy Acta* 75, 331-337, 1974.

Cremer, T., Cremer, C., Baumann, H., Luedtke, E. K., Sperling, K., Teubner, V. & Zorn, C. Rabl's model of the interphase chromosome arrangement, tested in chinese hamster cells by premature chromosome condensation and Laser-UV-Microbeam experiments. *Human Genetics* 60, 46-46, 1982a.

Cremer, T., Cremer, C., Schneider, T., Baumann, H., Hens, L. & Kirsch-Volders, M. Analysis of chromosome positions in the interphase nucleus of Chinese hamster cells by laser-UV-microirradiation experiments. *Hum. Genet.* 62, 201-209, 1982b.

Cremer, T. Von der Zellenlehre zur Chromosomentheorie, Naturwissenschaftliche Erkenntnis und Theorienwechsel in der frühen Zell- und Vererbungsforschung. *Springer Berlin-Heidelberg-New York*, 1985.

Cremer, T., Lichter, P., Borden, J., Ward, D. C. & Manuelidis, L. Detection of chromosome aberrations in metaphase and interphase tumor cells by in situ hybridization using chromosome specific library probes. *Hum. Genet.* 80, 235-246, 1988.

Cremer, T., Kurz, A., Zirbel, R., Dietzel, S., Rinke, B., Schrack, E., Speicher, M. R., Mathie, U., Jauch, A., Emmerich, P., Scherhan, H., Ried, T., Cremer C. & Lichter, P. Role of Chromosome Territories in the Functional Compartmentalization of the Cell Nucleus. *Cold Spring Harbor Symp. Quant. Biol.* 58, 777-792, 1993.

Cremer, C., Münkel, C., Granzow, M., Jauch, A., Dietzel, S., Eils, R., Guan, X. Y., Meltzer, P. S., Trent, J. M., Langowski, J. & Cremer, T. Nuclear architecture and the induction of chromosomal aberrations. *Mutat. Res.* 366, 97-116, 1996.

Cremer, T., Kreth, G., Koester, H., Fink, R. H. A., Heintzmann, R., Cremer, M., Solovei, I., Zink, D. & Cremer, C. Chromosome territories, interchromatin domain compartment and nuclear matrix: an integrated view of the functional nuclear architecture. *Critical Reviews in Eucariotic Gene Expression*, 2000.

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

Cremer, M., von Hase, J., Volm, T., Brero, A., Kreth, G., Walter, J., Fischer, C., Solovei, I., Cremer, C. & Cremer, T. Non-random radial higer-order chromatin arrangements in nuclei of diploid human cells. *Chrom. Res.* 9, 541-567, 2001.

Cross, S. S. Fractals in phathology. *J. of Pathol.* 182, 1-8, 1997.

Cross, S. S., Bury, J. P., Stephenson, T. J. & Harrison, R. F. Image analysis of low magnification images of fine needle aspirates of the breast produces useful discrimination between benign and malignant cases. *Cytopathol.* 8, 265-273, 1997.

Csankovszki, G. Panning, B., Bates, B. Pehrson, J. R. & Jaenisch, R. Conditionals deleteion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation. *Nature Gen.* 22, 323-324, 1999.

Cubitt, A. B., Heim, R., Adams, S.R, Boyd, A.E, Gross, L.A. and Tsien, R.Y. Understanding, improving and using green fluorescent proteins. *Trends Biochem. Sci.* 20, 448–455, 1995.

de Gennes P. G. La percolation: un concept unificateur. La Recherche 7, 919-927, 1976.

de Oliveira, P. M. C. Studying DNA evolution through successive file editions. *Physica A* 273, 70-74, 1999.

Dean, P. N. & Jett, J.H. Mathematical Analysis of DNA Distributions Derived from Flow Microfluorometry. *Journal of Cell Biology* 60, 523, 1974.

de la Tour, E. B. & Laemmli, U. K. The metaphase scaffold is helically folded: Sister chromatids have predominantly opposite helical handedness. *Cell* 55, 937-944, 1988.

Dey, A., Ellenberg, J., Farina, A., Coleman, A. E., Maruyama, T., Sciortino, S., Lippincott-Schwartz, J. & Ozato, K. Abromodomain protein, MCAP, associates with mitotic chromosomes and affects G(2)-to-M transition. *Mol. Cell. Biol.* 20, 6537-6549, 2000.

Dietzel, S., Eils, R., Sätzler, K., Bornfleth, H., Jauch, A., Cremer, C. & Cremer, T. Evidence against a looped structure of the inactive human X chromosome territory. *Exp. Cell. Res.* 240, 187-196, 1998a.

Dietzel, S., Jauch, A., Kienle, D., Qu, G., Holtgreve-Grez, H., Eils, R., Münkel, C., Bittner, M., Meltzer, P. S., Trent, J. M. & Cremer, T. Separate and variably shaped chromosome arm domains are disclosed by chromosome arm painting in human cell nuclei. *Chromosome Res.* 6, 25-33, 1998b.

Dietzel, S., Schiebel, K., Little, G., Edelmann, P., Rappold, G. A., Eils, R., Cremer, C. & Cremer, T. the 3D positioning of ANT2 and ANT3 genes within female X chromosome territories correlates with gene activity. *Exp. Cell Res.* 252, 363-375, 1999.

Doi, M. & Edwards, S. F. The theory of polymer dynamics, *Oxford University Press*, Oxford, 1986.

Duplantier, B., Jannink, G. & Sikorav, J. L. Anaphase chromatid motion: involvement of type II DNA topoisomerases. *Biophys. J.* 69 (4), 1596-1605, 1995.

Dupraw, E. J. DNA and Chromosomes. New York, Holt, Rinehart, Winston. 1970.

Earnshaw, W. C. & Laemmli, U. K. Architecture of metaphase chromosomes and chromosome scaffolds. *J. Cell Biol.* 96, 84-93, 1983.

Earnshaw, W. C. & Heck, M. M. S. Localization of topoisomerase II in mitotic chromosomes. *J. Cell Biol.* 96, 84-93, 1985.

Ehrlich, L., Münkel, C., Chirico, G. & Langowski, J. (1997). A Brownian dynamics model for the chromatin fiber. *Comput. Appl. Biosci.* 13 (3), 271-9, 1997.

Eigen, M. &Winkler-Oswatitsch, R. Transfer-RNA: The early adaptor. *Naturwissenschaften* 68, 217-228, 1981a.

Eigen, M. & Winkler-Oswatitsch, R. Transfer-RNA, an early gene? *Naturwissenschaften* 68, 282-292, 1981b.

Eigen, M, Gardiner, W., Schuster, P. & Winkler-Oswatitsch, R. Ursprung der genetischen Information. *Spektrum der Wissenschaft* 6, 36-56, 1981.

Eils, R., Bertin, E., Saracoglu, K., Rinke, B., Schröck, E., Parazza, F., Usson, Y., Robert-Nicoud, M., Stelzer, E. H. K., Chassery, J.-M., Cremer, T. & Cremer, C. Application of laser confocal microscopy and 3D-Voronoi diagrams for volume and surface estimates of interphase chromosomes. *J. Microsc.* 177, 150-161, 1995.

Eils, R., Dietzel, S., Bertin, E., Schröck, E., Speicher, M. R., Ried, T., Robert-Nicoud, M., Cremer, C. & Cremer, T. Three-dimensional reconstruction of painted human interphase chromosomes: active and inactive X-chromosome territories have similar volumes but differ in shape and surface structure. *J. Cell Biol.* 135 (6), 1427-1440, 1996.

Einstein, A. J., Barba, J., Unger, P. D. & Gil, J. Nuclear diffuseness as a measure of texture: definition and application to the computer-assisted diagnosis of parathyroid adenoma and carcinoma. *J. of Microscopy* 176(PT2), 158-166, 1994.

Einstein, A. J., Wu, H.-S., Sanchez, M. & Gil, J. Fractal characterization of chromatin appearance for diagnosis in breast cytology. *J. of Pathol.* 185, 366-381, 1998.

Einstein, A. J., Wu, H.-S. & Gil, J. Self-affinity and lacunarity of chromatin texture in benign and malignant breast epithelial cell nuclei. *Phys. Rev. Let.* 80(2), 397-400, 1998.

Elliott, B., Richardson, C., Winderbaum, J., Nickoloff, J. A. & Jasin M. Gene conversion tracts from double-strand break repair in mammalian cells. *Mol. Cell. Biol.* 18, 93–101, 1998.

Ermak, D. L. & McCammon, J. A. Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys.* 69 (4), 1352-1359, 1978.

Falconer, K. J. Fraktale Geometrie - Mathematische Grundlagen und Anwendungen. *Spektrum Akademischer Verlag (Heidelberg, Berlin, Oxford)*, ISBN 3-86025-075-2, 1993.

Felgner, P. L. et al. Lipofection: a highly efficient, lipid-mediated DNA-transfection procedure. *Proc. Natl. Acad. Sci. USA* 84, 7413–7417, 1987.

Finch, J. T. & Klug, A. Solenoidal model for the superstructure in chromatin. *Proc. Natl. Acad. Sci. USA* 73, 1897-1901, 1976.

Fletcher, H. L. The radial pos itions of metaphase chromosomes may be a consequence of the relative strength of their interaction with the spindle and their size. *Chrom. Res.* 2, 21-24, 1994.

Flick, K. E., Jurica, M. S. Monnat, R. J. & Stoddard, B. L. DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI. *Nature* 394, 96-101, 1998.

Flores-Rozas, H. & Kolodner, R. D. Links between replication, recombination and genome instability in eukaryotes. *Trends Biochem. Sci.* 25, 196–200, 2000.

Fradkov, A. F., Chen, Y., Ding, L., Barsova, E.V., Matz, M.V., and Lukyanov, S.A. Novel fluorescent protein from Discosoma coral and its mutants possesses a unique far-red fluorescence. *FEBS Lett.* 479, 127–130, 2000.

Francke, U. Digitized and differentially shaded human chromosome ideograms for genomic applications. *Cytogenet. Cell. Genet.* 65, 206-219, 1994.

Fransz, P. *et al.* Organization of interphase chromosomes in Arabidopsis thaliana: Euchromatic loops around heterochromatic chromocentres. (submitted).

Freire, J. P. & Horta, A. Mean reciprocal distances of short polymethylene chains. Calculation of the translational diffusion coefficient of n-alkanes. *J. Chem. Phys.* 65, 4049, 1976.

Gasser, S. M., Laroche, T., Falquet, J., Boy de la Tour, E. & Laemmli, U. K.Metaphase chromosome structure. Involvement of topoisomerase II. *J. Mol. Biol.* 188 (4), 613-629, 1986.

Gefen, Y. & Aharony A. Anaomalous diffusion on percolating clusters. *Phys. Rev. Let.* 50(1), 77-80, 1981.

Gennerich, A. & Schild, D. Fluorescence correlation spectroscopy in small cytosolic compartments depends critically on the diffusion model used. *Biophys. J.* 79(6), 3294, 2000.

Germain, L., Carrier, P., Auger, F. A., Salesse, C. &Guerin, S. L. Can we produce a human corneal equivalent by tissue engineering? *Prog. Retin. Eye. Res.* 19 (5), 497-527, 2000.

Goldman, M. A., Holmquist, G. P., Gray, M. C. Caston L. A. & Nag, A. Replication timing of genes and middle repetitive sequences. *Science* 224, 686-692, 1984.

Graham, F. L. & van der Eb, A. J. A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology* 52, 456–467, 1973.

Grossberg, A., Rabin, Y., Havlin, S. & Neer, A. Crumpled globule model of the three-dimensional structure of DNA. *Europhys. Lett.* 23 (5),373-378, 1993.

Haber, J. E. DNA recombination: the replication connection. *Trends Biochem. Sci.* 24, 271–275, 1999.

Habermann, F. A., Cremer, M., Walter, J., Kreth, G., von Hase, J., Bauer, K., Wienberg, J., Cremer, C., Cremer, T., & Solovei, I. Arrangements of macro- and microchromosomes in chicken cells. *Chrom. Res.* 9, 569-584, 2001.

Haensler, J. & Szoka, F. C. Jr. Polyamidoamine cascade polymers mediate efficient transfection of cells in culture. *Bioconjug. Chem.* 4, 372–379, 1993.

Haidekker, M. A., Andresen, R. Evertsz, C. J. G., Banzer, D. & Peitgen, H.-O. Assessing the degree of osteoporosis in the axial skeleton using the depencence of the fractal dimension on the grey level threshold. *The Brit. J. of Radiol.* 70, 586-593, 1997.

Halene, S. & Kohn, D. B. Gene therapy using hematopoietic stem cells: Sisyphus approaches the crest. *Hum. Gene. Ther.* 11 (9), 1259-1267, 2000.

Halperin, B. I., Feng, S. & Sen, P. N. Differences between Lattice and continuum percolation transport exponents. *Phys. Rev. Let.* 54(22), 2391-2110, 1985.

Hammermann, M., Toth, K., Rodemer, C., Waldeck, W., May, R.P. & Langowski, J. Salt-dependent compaction of di- and trinucleosomes studied by small-angle neutron scattering. *Biophys. J.* 79 (1): 584-594, 2000.

Hao, B., Xie, H., Yu, Z. & Chen, G. Factorizable language: from dynamics to bacterial complete genomes. *Physica A* 288, 10-20, 2000a.

Hao, B., Lee, H. C. & Zhang, S. Fractals related to long DNA sequences and complete genomes. *Chaos, Solit. & Fractals* 11, 825-836, 2000b.

Hastings, H. M. & Sugihara, G. Fraktale - ein Leitfaden für Anweder, *Spektrum Akademischer Verlag (Heidelberg, Berlin, Oxford)*, ISBN 3-86025-337-9, 1996.

Haralick, R. M., Shanmugam, K. & Dinstein, H. Textural features for image classification. *IEEE Trans. Syst., Man, Cybern.* SMC-3(6), 610-621,1973.

Harder, H. & Havlin, S. Diffusion on fractals with singular waiting-time distribution. *Phys. Rev. B* 36(7), 3874-3879.

Harrison, A. Fractals in Chemistry. Oxford Chemistry Primers, *Oxford University Press*, ISBN 0-19-855-768, 1995.

Hattori, M., Fujiyama, A., Taylor, T. D., Watanabe, H. Yada, T., Park, H.-S., Toyoda, A., Ishii, K., Totoki, Y., Choi, D.-K., Soeda, E., Ohki, M., Takagi, T., Sakaki, Y., Taudien, S., Blechschmidt, K., Polley, A., Menzel, U., Delabar, J., Kumpf, K., Lehmann, R., Patterson, K., Reichwald, A., Rump, M., Schillhabel, M., Schudy, A., Zimmermann, W., Rosenthal, A., Kudoh, J., Shibuya, K., Kawasaki, K., Asakawa, S., Shintani, A., Sasaki, T., Nagamine, K., Mitsuyama, S., Antonarakis, S. E., Minoshima, S., Shimuzu, N., Nordsiek, G., Hornischer, K., Brandt, P., Scharfe, M., Schön, O., Desario, A., Reichelt, J., Kauer, G., Blöcker, H., Ramser, J., Beck, A., Klages, S., Hennig, S., Riesselmann, L., Dagand, E., Haaf, T., Wehrmeyer, S., Borzym, K., Gardiner, K., Nizetic, D., Francis, F., Lehrach, H., Reinhard, R. & Yaspo, M.-L. The DNA sequence of human chromosome 21. *Nature* 405, 311-319, 2000.

Havlin, S. & Weiss, G. H. Relation between dynamic transport properties and static topological structure for the lattice-animal model of branched polymers. *Phys. Rev. Let.* 53(2), 178-181, 1984.

Heim, R., Cubitt, A. B. & Tsien, R. Y. Improved green fluorescence. *Nature* 373, 663–664, 1995.

Herrmann, H. J., Derrida, B. & Vannimenus, J. Superconductivity exponents in two- and three-dimensional percolation. *Phys. Rev. B* 30(7), 4080-4082,1984.

Holmquist, G. P. Review article: chromosome bands, their chromatin flavors, and their functional features. *Am. J. Hum. Genet.* 51, 17-37, 1992.

Holy, C. E., Shoichet, M. S. &Davies, J. E. Engineering three-dimensional bone tissue in vitro using biodegradable scaffolds: investigating initial cell-seeding density and culture period. *J. Biomed. Mater. Res.* 51 (3), 376-382, 2000.

Horowitz, R. A., Agard, D. A., Sedat, J. W. & Woodcock, C. L. The three-dimensional architecture of chromatin in situ: electron tomography reveals fibers composed of a continuously variable zig-zag nucleosomal ribbon. *J. Cell Biol.* 125 (1), 1-10, 1994.

Horowitz, R. A., Koster, A. J., Walz, J. & Woodcock, C. L. Automated electron microscope tomography of frozen-hydrated chromatin: the irregular three-dimensional zigzag architecture persists in compact, isolated fibers. *J. Struct. Biol.* 120 (3), 353-362, 1997.

Houchmandzadeh, B. & Dimitrov, S. Elasticity measurements show the existance of thin rigid cores inside mitotic chromosomes. *J. Cell Biol.* 145 (2), 215-223, 1999.

Houchmandzadeh, B., Marko, J. F., Chatenay, D. & Libchaber, A. Elasticity and structure of eukaryote chromosomes studied by micromanipulation and micropipette aspiration. *J. Cell. Biol.* 139 (1), 1-12, 1997.

Hsü, K. J. & Hsü, A. J. Fractal geometry of music. *Proc. Natl. Acad. Sci. USA* 87, 938-941, 1990.

Hsü, K. J. & Hsü, A. J. Self-similiarity of the “1/f noise“ called music. *Proc. Natl. Acad. Sci. USA* 88, 3507-3509, 1991.

Hursting, S. D. Experimental models of gene-environment interaction for cancer chemoprevention studies. *Curr. Opin. Oncol.* 9 (5), 487-491, 1997.

Ibel, K. Neutron diffraction of interphase nuclei. *J. Mol. Biol.* 160 (1), 77-85, 1982.

Iniesta, A. & Garcia de la Torre, J. A second-order algorithm for the simulation of the Brownian dynamics of macromolecular models. *J. Chem. Phys.* 92, 2015-2019, 1990.

Irinopoulou, T., Rigaut, J. P. & Benson M. C. Toward objective prognostic grading of prostatic carcinoma using image analysis. *Analyt. and Quantit. Cytol. and Histol.* 341-344, 1993.

Jacks, T. Lessons from the p53 mutant mouse. *J. Cancer Res. Clin. Oncol.* 122 (6), 319-327, 1996.

Jackson, D. A. & Pombo, A. Replication clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells. *J. Cell Biol.* 140, 12850-1295, 1998.

Jannink, G., Duplantier, B. & Sikorav, J. L. Forces on chromosomal DNA during anaphase. *Biophys. J.* 71 (1), 451-65, 1996.

Johnson, R. D., Liu, N. & Jasin, M. Mammalian XRCC2 promotes the repair of DNA double-strand breaks by homologous recombination. *Nature* 401, 397–399, 1999.

Kanda, T., Sullivan, K. F. & Wahl, G. M. Histone-GFP fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cells. *Curr. Biol.* 8, 277-385, 1998.

Kao, H. P., Abney, J. R. & Verkman, A. S. Determinants of the translational mobility of a small solute in the cytoplasm. *J. Cell. Biol.* 120(1), 175, 1993.

Kapitulnik, A. & Deutscher, G. Percolation characteristics in discontinous thin films of Pb. *Phys. Rev. B* 49(19), 1444-1448, 1982.

Karlin, S. & Brendel, V. Patchiness and correlations in DNA sequences. *Science* 259 677-680, 1993.

Kaye, B. H. A random walk through fractal dimensions. *VCH (Weinheim, New York, Basel, Cambridge, Tokyo)*, ISBN 3-527-26468-x, 1989.

Kimura, H. & Cook, P. R. Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B. *J. Cell. Biol.* 153, 1341-1353, 2001.

Kirkpatrick, S. & Stoll, E. Implementation the R250 random number generator. *J. Computational Physics* 40, 517, 1981.

Knoch, T. A. Three Dimensional Organisation of Chromosome Domains in Simulation and Experiment. *Diploma Thesis*, German Cancer Research Center (DKFZ) Heidelberg and Faculty for Physics and Astronomy of the Ruprecht Karls University of Heidelberg, Germany, 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 - Artickle Archive, Foresight Institute, Palo Alto, CA, USA*, http://www. foresight.org/, 1-6, 1998.

Knoch, T. A., Münkel, C., and Langowski „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 Egon Krause and Willi Jäger, High-Performance Computing Center (HLRS) Stuttgart, University of Stuttgart, Springer Berlin-Heidelberg-New York, ISBN 3-540-66504-8, 229-238, 2000.

Knoch, T. A. Approaching the Three-Dimensional Organization of the Human Cell Nucleus: 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. *PhD thesis*, Faculty for Physics and Astronomy, Ruprecht Karls University of Heidelberg, 2002.

Knopf, W. C. & Waldeck, W. DNA-binding enzymes, structural themes. *Enzyclopedia of Life Sciences*, artickle #2717, Macmillion Publishers Ltd., Nature Publishing Group, London, http://www.els.net, 2001.

Kornberg, R. D. & Klug, A. The nucleosome. *Scientific American* 2, 28-44, 1981.

Kowalczykowski, S. C. Initiation of genetic recombination and recombination-dependent replication. *Trends Biochem. Sci.* 25, 156–165, 2000.

Kreth, G. Simulation von Chromosomen und virtuelle Microscopie. *PhD thesis*, Faculty for Physics and Astronomy, Ruprecht Karls University of Heidelberg, 2001.

Kriete, A. Hierarchical data representation of lung to model morphology and function. in Höne, K. H. & Kikinis, R (Eds.) Visualization in Biomedical Computing. Proceedings of the 4th International Converence, VBC'96, Hamburg, Germany, Lecture Notes in Computer Science 1131, *Springer*, 1996.

La France, M. L. & Armstrong, D. W. Novel living skin replacement biotherapy approach for wounded skin tissues. *Tissue. Eng.* 5 (2), 153-170, 1999.

Laemmli, U. K. Levels of organization of the DNA eukaryotic chromosomes. *Pharmac. Rev.* 30, 469-476, 1976.

Landini G. & Rippin J. W. An "asymptotic fractal" approach to the morphology of malignant cell nuclei. *Fractals* 1(3), 326-335, 1993.

Lawrence, J. B. Villnave, C. A. & Singer, R. H. Sensitive, high-resolution chromatin and chromosome mapping *in situ*: Presence and Orientation of two closely integrated copies of EBV in a lymphoma line. *Cell* 52, 51-61, 1988.

Lawrence, J. B., Singer, R. H. & Marselle, L. M. Highly localized tracks of specific transcripts within interphase nuclei visualized by *in situ* hybridization. *Cell* 57, 493-502, 1989.

Lawrence, J. B., Singer, R. H. & McNeil, J. A. Interphase and Metaphase Resolution of Different Distances Within the Human Dystrophin Gene. *Science* 249, 928-932, 1990.

Lee, H, Hong, M., Kim, J. W., Hong, Y. M., Choe, Y.-K., Chang, S. Y., Lee, K. S. & Choe, I. S. Isolation of cDNA clones encoding human histone macroH2A1 subtypes. *Biochim. Biophys. Acta* 1399,73-77, 1998.

Lefkovitch, L. P. Optimal set covering for biological classification. *Ottawa*, 1993.

Leitch, A. R., Mosgöller, W., Schwarzacher, T., Bennett, M. D. & Heslop-Harrison, J. S.Genomic in situ hybridization to sectioned nuclei shows chromosome domains in grass hybrids. *J. Cell Sci.* 95, 335-341, 1990.

Leuba, S. H., Yang, G., Robert, C., Samori, B., van Holde, K., Zlatanova, J. & Bustamante, C. Three-dimensional structure of extended chromatin fibers as revealed by tapping-mode scanning force microscopy. *Proc. Natl. Acad. Sci. USA* 91, 11621-11625, 1994.

Lever, M. A., Th'ng, J. P., Sun, X. & Henzel, M. J. Rapid exchange of histone H1.1 on chromatin in living cells. *Nature* 408, 873-876, 2000.

Lewin, B. Genes VII.*Oxford University Press*, ISBN 0-19-879277-8, 2000.

Li, G., Sudlow, G. & Belmont, A. S. Interphase cell cycle dynamics of a late-replicating, heterochromatic homogeneously staining region: Precise choreography of condensation/decondensation and nuclear positioning. *J. Cell Biol.* 140(5), 975-989, 1998.

Li, W. & Kaneko, K. Longe-range correlation and partial $1/f^{\alpha}$ spectrum in a noncoding DNA Sequence. *Europhys. Lett.* 17 (7), 655-660, 1992.

Liang, F., Han, M., Romanienko, P. J. & Jasin, M. Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *Proc. Natl. Acad. Sci. USA* 95, 5172–5177, 1998.

Liu, K. & Stein, A. DNA sequence encodes information for nucleosome array formation. *J. Mol. Biol*. 270, 1997.

Lichter, P., Cremer, T., borden, J., Manuelidis, L. & Ward, D. C. Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. *Hum. Genet.* 80, 224-234, 1988.

Lindgren, M., Hällbrink, M., Prochiantz, A. and Langel, Ü. Cell-penetrating peptides. *TiPS* 21, 99-103, 2000.

Liu, B. & Sachs, R. K. A two-backbone model for interphase chromosome geometry. *Bulletin Math-Mathem.Biol.* 59(2), 325-337, 1997.

Lobb, C. J. & Frank, D. J. Percolative conduction and the Alexander-Orbach conjecture in two dimensions. *Phys. Rev. B* 30(7), 4090-4092, 1984.

Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* 276, 19-42, 1998.

Luby-Phelps, K., Taylor, D. L. & Lanni, F. Probing the structure of cytoplasm. *J. Cell Biol.* 102 (6), 2015, 1986.

Luby-Phelps, K., Castle, P. E., Taylor, D. L. & Lanni, F. Hindered diffusion of inert tracer particles in the cytoplasm of mouse 3T3 cells. *Proc. Natl. Acad. Sci. USA* 84(14), 4910, 1987.

Luby-Phelps, K., Mujumdar, S., Mujumdar, R. B., Ernst, L. A., Galbraith, W. & Waggoner, A. S. A novel fluorescence ratiometric method confirms the low solvent viscosity of the cytoplasm. *Biophys. J.* 65(1), 236, 1993.

Luby-Phelps, K. Physical properties of cytoplasm. *Curr. Opin. Cell Biol.* 6(1), 3, 1994.

Lukacs, G. L., Haggie, P., Seksek, O., Lechardeur, D., Freedman, N. & Verkman, A. S. Size-dependent DNA mobility in cytoplasm and nucleus. *J. Biol. Chem.* 275(3), 1625, 2000.

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

Luo, L., Lee, W., Jia, L., Ji, F. & Tsai, L. Statistical correlation of nucleotides in a DNA sequence. *Phys. Rev. E* 58 (1), 861-871, 1998.

Ma, H. Siegel, A. J. & Berezney, R. Association of chromosome territories with the nuclear matrix. Disruption of human chromosome territories correlates with the release of a subset of nuclear matrix proteins. *J. Cell Biol.* 146, 531-542, 1999.

Mackiewicz, P., Gierlik, A., Kowalczuk, M., Szczepanik, D., Dudek, M. R. & Cebrat, S. Mechanisms generating long-range correlation in nucleotide composition of the Borrelia burgdoferi genome. *Physica A* 273, 103-115, 1999.

Madigan, M. T., Martinko, J. M. & Parker, J. *Brock.* Biology of Microorganisms. 8th edition, *Prentice Hall International Inc.*, 1997.

Maddox, J. Long-range correlations within DNA. *Natur* 358, 103, 1992.

Madra, N. & Sokal, A. D. The Pivot algorithm: a highly efficient Monte Carlo method for the self avoiding random walk. J. *Statistical Physics* 50, 109-189, 1988.

Maier, W. L. A fast pseudo random number generator. *Dr. Dobb's Journal* 176, 1991.

Mandelbrot, B. B. The Fractal Geometry of Nature, *W. H. Freeman and Company,* New York, ISBN 0-7167-1186-9, 1983.

Marks, P. A., Richon, V. M. & Rifkind, R. A. Histone deacetylase inhibitors: inducers of differentiation or apoptosis of transformed cells. *J. Natl. Cancer Inst.* 92, 1210-1216, 2000.

Marsden, M. P. F. & Laemmli, U. K. Metaphase chromosome structure: evidence for a radial loop model. *Cell* 17, 849-858, 1979.

Matz, M. V., Fradkov, A. F., Labas, Y. A., Savitsky, A.P., Zaraisky, A.G., Markelov, M.L., and Lukyanov, S.A. Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat. Biotechnol.* 17, 969–973, 1999.

McLellan, T. & Endler, J. A. The relative success of some methods for measuring and describing the shape of complex objects. *Syst. Biol.* 47(2), 264-281, 1998.

Mermoud, J. E., Costanzi, C., Pehrson, J. R. & Brockdorff, N. Histone macroH2A1.2 relocates to the inactive X chromosome after initiation and propagation of X-inactivation. *J. Cell Biol.* 147(7), 1399-1408, 1999.

Metropolis, M. Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* 21, 1087-1092, 1953.

Mira, A. Ochman, H. & Moran, N. Deletional bias and the evolution of bacterial genomes. *Trends. Genet.* 17(19), 589-596, 2001.

Misteli, T., Gunjan, A. Hock, R., Bustin, M. & Brown, D. T. Dynamic binding of histone H1 to chromatin in living cells. *Nature* 408, 877-881, 2000.

Misteli, T. Protein dynamics: Implications for nuclear architecture and gene expression. *Science* 291(5505), 843, 2001.

Mohanty, A. K. & Narayana-Rao, A. V. S. S. Factorial moments analysis show characteristic length scale in DNA sequences. *Phys. Rev. E* 84 (8), 1832-1835, 2000.

Mongelard, F., Vourc'h, C., Robert-Nicoud, M. & Usson, Y. Quantitative assessment of the alteration of chromatin during the course of FISH procedures. *Cytometry* 36, 96-101, 1999.

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

Monier, K., Armas, J. C., Etteldorf, S., Ghazal, P. & Sullivan, K. F. Annexation of the interchromosomal space during viral infection. *Nat. Cell Biol.* 2, 661-665, 2000.

Morin, J. & Hastings, J. Energy transfer in a bioluminescent system. *J. Cell Physiol.* 77, 313–318, 1971.

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

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

Nagle, J. F. Long tail kinetics in biophysics. *Biophys. J.* 63, 366-370, 1992.

Nakayasu, H. & Berezney, R. Mapping replicational sites in the eucaryotic cell nucleus. *J. Cell. Biol.* 108 (1), 1-11, 1989.

Nee, S. Uncorrelated DNA walks. *Nature* 357, 450, 1992.

Ng, H. H. & Bird, A. Histone deacetylases: silencers for hire. *Trends Biochem. Sci.* 25, 121-126, 2000.

Nitiss, J. L. Investigating the biological functions of DNA topoisomerases in eukaryotic cells. *Biochim. Biophys. Acta.* 1400 (1-3), 63-81, 1998.

Nonnenmacher, T. F., Losa, G. A. & Weibel, E. R. (Eds.) Fractals in biology and medicine. *Birkhäuser (Basel, Boston, Berlin)*, ISBN 3-7643-2989-0, 1993.

Notbohm, H. Small angle scattering of cell nuclei. *Eur. Biophys. J.* 13 (6), 367-372, 1986.

Novak, M. M. (Ed) Fractals in the natural and applied sciences. Proceedings of the 2nd IFIP Working Conference on Fractals in the Natural and Applied Sciences, London, UK, 1993, IFIP Transactions A: Computer Science and Technology, *North-Holland* (Amsterdam, London, New York, Tokyo), ISBN 0-444-81628-3, 1994.

Okabe, M., Ikawa, M., Kominami, K., Nakanishi, T. & Nishimune, Y. "Green mice" as a source of ubiquitous green cells. *FEBS Lett.* 407 (3), 313-319, 1997.

Olins, A. L & Olins, D. E. Spheroid Chromatin Units (v Bodies). *Science* 183, 330-332, 1974.

Orbach, R. Dynamics of fractal networks. Science 231, 814-819, 1986.

Ostashevsky, J. Y. & Lange, C. S. The 30 nm chromatin fiber as a flexible polymer. *J. Biomol. Struct. & Dyn.* 11, 813-820, 1994.

Ourednik, V., Ourednik, J., Flax, J. D., Zawada, W. M., Hutt, C., Yang, C., Park, K. I., Kim, S. U., Sidman, R. L., Freed, C. R. & Snyder, E. Y. Segregation of human neural stem cells in the developing primate forebrain. *Science* 293, 1820-1824, 2001.

Pardoll, D. M., Vogelstein, B. & Coffey, D. S. Localization of SV40 genes within supercoiled loop domains. *Cell* 19, 527-536, 1980.

Parikian, A., Olveczky, B., Swaminathan, R., Li, Y. & Verkman, A. S. Rapid diffusion of green fluorescent protein in the mitochondrial matrix. *J. Cell Biol.* 140(4), 821, 1998.

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

Paulson, J. R. Scaffolding and radial loops: the structural organization of metaphase chromosomes. *Chromosomes and Chromatin 3*, CRC Press, Boca Raton, Florida, 3-36,1988.

Pehrson, J. R. & Fried, V. A. MacroH2A, a core histone containing a large nonhistone region. *Science* 257, 1398-1400, 1992.

Pehrson, J. R., Costanzi, C. & Dharia, C. Developmental and Tissue Expression Patterns of histone macroH2A1 subtypes. *J. of Cell. Biochem.* 65, 107-113, 1997.

Pehrson, J. R. & Fuji, R. N. Evolutionary conservation of histone macroH2A subtypes and domains. *Nucleic Acids Res.* 26(12), 2837-2842, 1998.

Peitgen, H.-O., Jürgens, H. & Saupe, D. Chaos and fractals - new frontiers of science. *Springer-Verlag Berlin Heidelberg (Berlin, Heidelberg, New York)*, ISBN 3-540-97903-4, 1992.

Peng, C.-K., Buldyrev, S. V., Goldberger, A. L., Havlin, S., Sciortino, F., Simons, M. & Stanley, H. E. Long-range correlations in nucleotide sequences. *Nature* 356, 168-170, 1992.

Peng, C.-K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H.E. & Goldberger, A. L. Mosaic organization of DNA nucleotides. *Phys. Rev. E* 49 (2), 1685-1689, 1994.

Perche, P. Y., Vourc'h, C., Konecny, L., Souchier, C., Robert-Nicoud, M., Dimitrov, S. & Khochbin, S. Higher concentrations of histone macroH2A in the Barr body are correlated with higher nucleosome density. *Curr. Biol.* 10, 1531-1534, 2000.

Pfeifer, P., Welz, U. & Wippermann, H. Fractal surface dimension of proteins: Lysozyme. *Chem. Phys. Let.* 113(6), 535-540, 1985.

Pfitzner, W. Über den feineren Bau der bei der Zellteilung auftretenden fadenförmigen Differenzierung des Zellkerns. *Morphologisches Jahrbuch* 7, 289-311. 1881.

Phair, R. D. & Misteli, T. High mobility of proteins in the mammalian cell nucleus. *Nature* 404, 604-609, 2000.

Piccini, A., Perkus, M. E. & Paoletti, E. Vaccinia virus as an expression vector. *Methods Enzymol.* 153, 545–563, 1987.

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

Pinkel, D., Landegent, J., Collins, C., Fuscoe, J., Segraves, R., Lucas, J. & Gray, J. W. Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and tanslocations of chromosomes 4. *Proc. Natl. Acad. Sci.* 85, 9138-9142, 1988.

Pomahac, B., Svensjo, T., Yao, F., Brown, H. & Eriksson, E. Tissue engineering of skin. *Crit. Rev. Oral. Biol. Med.* 9 (3), 333-344, 1998.

Poirier, M., Eroglu, S., Chatenay, D. & Marko, J. F. Reversible and irreversible unfolding of mitotic newt chromosomes by applied force. Mol. Biol. Cell. 11 (1), 269-276, 2000.

Politz, J. C., Browne, E. S., Wolf, D. E. & Pederson T. Intranuclear diffusion and hybridization state of oligonucleotides measured by fluorescence correlation spectroscopy in living cells. *Proc. Natl. Acad. Sci. USA* 95(11), 6043, 1998.

Politz, J. C. Use of caged fluorochromes to track macromolecular movement in living cells. *Trends Cell Biol.* 9(7), 284, 1999.

Pollok, B. A. & Heim, R. Using GFP in FRET-based applications. *Trends Cell Biol.* 9, 57–60, 1999.

Popp, S., Scholl, H. P., Loos, P., Jauch, A., Stelzer, E., Cremer, C. & Cremer, T. Distribution of chromosome 18 and X centric heterochromatin in the interphase nucleus of cultured human cells. *Exp. Cell. Res.* 189, 1-12, 1990.

Prabhu, V. V. & Claverie, J.-M. Correlations in intronless DNA. *Nature* 359, 782, 1992.

Provata, A. & Almirantis, Y. Fractal Cantor patterns in the sequence structure of DNA. *Fractals* 8 (1), 15-27, 2000.

Qian, H., Raymond, G. M. & Bassingthwaighte, J. B. Stochastic fractal behaviour in concentration fluctuation and fluorescence correlation spectroscopy. *Biophys. Chem.* 80, 1-5, 1999.

Rabinovich, M. I., Fabrikant, A. L. & Tsmiring, L. Sh. Finite-dimensional disorder. *Sov. Phys. Usp.* 35 (8), 629-649, 1992.

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

Rammal, R. & Toulouse, G. Random walks on fractal structures and percolation clusters. *J. Physique - Lettres* 44, L13-L22, 1983.

Rasmussen, T. P., Huang, T., Mastrangelo, M.-A., Loring, J., Panning, B. & Jaenisch, R. Messenger RNAs encoding mouse histone macroH2A1 isoforms are expressed at similar levels in male and female cells and result from alternative splicing. *Nucleic Acids Res.* 27(18), 3685-3689, 1999.

Rasmussen, T. P., Mastrangelo, M.-A., Eden, A., Pehrson, J. R. & Jaenisch, R. Dynamic relocalization of histone macroH2A1 from centrosomes to inactive X chromsomes during X inactivation. *J. Cell Biol.* 150(5), 1189-1198, 2000.

Rattner, J. B. & Lin, C. C. Radial loops and helical coils coexist in metaphase chromosomes. *Cell* 42, 291-296, 1985

Rauch, J. Spektrale Präzisionsdistanzmikroskopie zur Untersuchung der 3D-Topologie ausgewählter DNA-Punktmarker. *PhD thesis*, Faculty for Physics and Astronomy, Ruprecht Karls University of Heidelberg, 1999.

Reichenzeller, M., Burzlaff, A., Lichter, P. & Herrmann, H. *In vivo* observation of a nuclear channel-like system: Evidence for a distinct interchromosomal domain compartment in interphase cells. *J. Struct. Biol.* 129, 175-185, 2000.

Reif, F. Fundamentals of Statistical and Thermal Physics. *McGraw-Hill*, New York, 1965.

Remy, J. S., Sirlin, C., Vierling, P. & Behr, J. P. Gene transfer with a series of lipophilic DNA-binding molecules. *Bioconjug. Chem.* 5, 647–654, 1994.

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

Reznik, N.A., Yampol, G.P., Kiseleva, E. V., Khristolyubova, N. B., Gruzdev, A. D. Functional and structural units in the chromomere. *Genetica* 83 (3), 293-299, 1991.

Richardson, C. & Jasin, M. Frequent chromosomal translocations induced by DNA double-strand breaks. *Nature* 405, 697–700, 2000.

Rigaut, J. P. An emperical formulation relating boundary lengths to resolution in specimens showing 'non-ideally fractal' dimensions. *J. of Microscopy* 133(Pt1), 41-54, 1983.

Robinett, C. C., Straight, A., Li, G., Willhelm, C., Sudlow, G., Murray, A. & Belmont, A. S. In Vivo Localization of DNA Sequences and Visualization of Large-Scale Chromatin Organization Using Lac Operator/Repressor Recognition. *J. Cell Biol.* 135 (6), 1685-1700, 1996.

Robinson, F. N. H. Noise and Fluctuations, *Clarendon*, Oxford, 1974.

Rodríguez-Iturbe, I. & Rinaldo, A. Fractal river basins - chance and self-organization. *Cambridge University Press*, ISBN 0-521-47398-5, 1997.

Roth, J. B., Foxworth, W. B., McArthur, M. J., Montgomery, C. A. & Kier, A. B. Spontaneous and engineered mutant mice as models for experimental and comparative pathology: history, comparison and developemental technology. *Lab. Anim. Sci.* 49 (1), 12-34, 1999.

Rosenberg, M. P. Gene knockout and transgenic technologies in risk assessment: the next generation. *Mol. Carcinog.* 20 (3), 262-274, 1997.

Rudnick, J. & Gaspari, G. The shapes of random walks. *Science* 237, 384-389, 1987.

Sachs, R. K., van den Engh, G., Trask, B., Yokota, H. & Hearst, J. E. A random-walk/giant-loop model for interphase chromosomes. Proceedings of the *Nat. Acad. Sci. USA* 92, 2710-2714, 1995.

Sadoni, N., Sullivan, K. F., Weinzierl, P. Stelzer, E. H. & Zink, D. Large-scale chromatin fibers of living cells display a discontinuous functional organization. *Chromosoma* 110, 39-51, 2001.

Saitho, Y. & Laemmli, U. K. Metaphase chromosome structure: Bands arise from a differential folding path of the highly AT-rich scaffold. *Cell* 76, 609-622, 1994.

Saitou, N. & Nei, M. The Neighbour-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4 (4), 406-425, 1987.

Salganik, R. I., Dudareva, N. A., Popovsky, A. V., Kiseleva, E. V. & Rozov, S. M. Structure of the plant mitochondrial genome and light-regulated transcription of the mitochondrial genes. *Nuclear Structure and Function. by Harris, J. R. & Zbarsky, I. B. (Editors)*, Plenum Press, New York - London, 19-22, 1990.

Sandau, K. & Kurz, H. Measuring fractal dimension and complexity - an alternative approach with an application. *J. of Microscopy* 186(Pt2), 164-176, 1996.

Saxton, M. J. Lateral diffusion in an archipelago. *Biophys. J.* 52, 989-997, 1987.

Saxton, M. J. Lateral diffusion in an archipelago: Distance dependence of the diffusion coefficient. *Biophys. J.* 56, 615-622, 1989.

Saxton, M. J. Lateral diffusion in a mixture of mobile and immobile particles: AMonte Carlo study. *Biophys. J.* 58, 1303-1306, 1990.

Saxton, M. J. Lateral diffusion and aggregation: A Monte Carlo study. *Biophys. J.* 61, 119-128, 1992.

Saxton, M. J. Lateral diffusion in an archipelago: Dependence on tracer size. *Biophys. J.* 64, 1053-1062, 1993a.

Saxton, M. J. Lateral diffusion in an archipelago: Single-partickle diffusion. *Biophys. J.* 64, 1766-1780, 1993b.

Saxton, M. J. Anomalous diffusion due to obstacles: AMonte Carlo study. *Biophys. J.* 66, 394-401, 1994.

Saxton, M. J. Single-partickle tracking: effects on corrals. *Biophys. J.* 69, 389-398, 1995a.

Saxton, M. J. Anomalous diffusion due to binding: A Monte Carlo study. *Biophys. J.* 70, 1250-1262, 1995b.

Saxton, M. J. Anomalous diffusion due to binding: a Monte Carlo study. *Biophys. J.* 70 (3), 1250-62, 1996.

Schardin, M., Cremer, T., Hager, H. D. & Lang, M. Specific staining of human chromosomes in Chinese hamster x man hybrid cell lines demonstrates interphase chromosome territories. *Hum. Genet.* 71, 281-287, 1985.

Scher, H., Shlesinger, M. F. & Bendler, J. T. Time-scale invariance in transport and relaxation. *Physics Today*, 26-34, Jan 1991.

Schwille, P., Haupts, U., Maiti, S. & Webb, W. W. Molecular dynamics in living cells observed by fluorescence correlation spectroscopy with one- and two-photon excitation. *Biophys. J.* 77(4), 2251, 1999.

Sedat, J. & Manuelidis, L. A direct approach to the structure of eukaryotic chromosomes. *Cold Spring Harbor Symp. Quant. Biol.* 42, 331-350, 1977.

Seksek, O., Biwersi, J. & Verkman, A. S. Translational diffusion of macromolecule-sized solutes in cytoplasm and nucleus. *J. Cell. Biol.* 138(1), 131, 1997.

Senger, G., Ragoussis, J., Trowsdale, J. & Sheer, D. Fine mapping of the human MHC class II region within chromosome band 6p21 and evaluation of probe ordering using interphase fluorescence in situ hybridization. *Cytogenet. Cell Genet.* 64 (1), 49-53, 1993.

Sharma, P. & Schreiber-Agus, N. Mouse models of prostate cancer. *Oncogene* 18 (38), 5349-5355, 1999.

Shigekawa, K. & Dower, W. J. Electroporation of eukaryotes and prokaryotes: a general approach to the introduction of macromolecules into cells. *Biotechniques* 6, 742–751, 1988.

Shrout, M. K., Potter, B. J. & Hildebolt, C. F. The effect of image variations on fractal dimensions calculations. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod.* 84, 96-100, 1997.

Siavoshian, S., Segain, J. P., Kornprobst, M., Bonnet, C., Cherbut, C., Galmiche, J. P. & Blottiere, H. M. Butyrate and trichostatin A effects on the proliferation/differentiation of human intestinal epithelial cells: induction of cyclin D3 and p21 expression. *Gut.* 46, 507-514, 2000.

Siegel, P. M., Hardy, W. R. & Muller, W. J. Mammary gland neoplasia: insights from transgenic mouse models. *Bioessays* 22 (6), 554-563, 2000.

Sikorav, J. L. & Jannink, G. Kinetics of chromosome condensation in the presence of topoisomerases: a phantom chain model. *Biophys. J.* 66 (3 Pt 1), 827-37, 1994.

Sparvoli, F., Martin, C., Scienza, A., Gavazzi, G. & Tonelli, C. Cloning and molecular analysis of structural genes involved in flavonoid and stilbene biosynthesis in grape (Vitis vinifera L.). *Plant. Mol. Biol.* 24 (5), 743-55, 1994.

Stack, S. M., Brown, D. B. & Dewey, W. C. Visualization of interphase chromosomes. *J. Cell Sci.*, 281-299, 1977.

Stanley, H. E. Application of fractal concepts to polymer statistics and to anomalous transport in randomly porous media. *J. Stat. Phys.* 36(5/6), 843-860, 1984.

Stanley, H. E. & Ostrowsky, N. (Eds.) On growth and form - fractal and non-fractal patterns in physics. NATO ASI Series, Series E: Applied Sciences - No. 100, *Martinus Nijhoff Publishers (Dordrecht)*, ISBN 90-247-3234-4, 1986.

Stanley, H. E. & Ostrowsky, N. (Eds.) Random fluctuations and pattern growth - experiments and models. NATO ASI Series, Series E: Applied Sciences - No. 157, *Kluwer Academic Publishers (Dordrecht, Boston, London)*, ISBN 0-7923-0073-4, 1988.

Stanley, H. E., Buldyrev, S. V., Goldberger, A. L., Goldberger, Z. D., Havlin, S., Mantegna, R. N., Ossadnik, S. M., Peng, C.-K. and Simons, M. Statistical mechanics in biology: how ubiquitous are long-range correlations? *Physica A* 205, 214-253, 1994.

Stark, W. M., Boocock, M. R. & Sherratt, D. J. Catalysis by site-specific recombinases. *Trends Genet.* 8, 432–439, 1992. *Erratum: Trends Genet.* 9, 45, 1993.

Stauffer, D. & Aharony, A. Perkolationstheorie - eine Einführung. *VCH (Weinheim, New York, Basel, Cambridge, Tokyo)*, ISBN 3-527-29334-5, 1995.

Stoehr M., Gebhardt U., and Goerttler K. Computer assistance in multiparameter flow microfluorometry of mammalian cells. *Biotechnology and Bioengineering* 18 (8), 1057-1074, 1976.

Stoehr M., Vogt-Schaden M., Knobloch M., Vogel R. & Futterman G. Evaluation of eight fluorochrome combinations for simultaneous DNA-protein flow analyses. *Stain Technology* 55 (4), 205-215, 1978.

Stoehr, M. & Futterman G. A complete C-Program for 4-parameter flowcytometry using personal computer technology. *Heidelberger Zytometrie Symposium* Abstract, 17. - 19. October, 1991.

Stokke, B. T. & Brant, D. A. The reliability of wormlike polysaccharide chain dimensions estimated from elektrone micrographs. *Biopolymers* 30, 1161-1181, 1990.

Swaminathan, R., Bicknese, S., Periasamy, N. & Verkman, A. S. Cytoplasmatic viscosity near the cell plasma membrane: Translational diffusion of a small fluorescet solute measured by total internal reflection-fluorescence photobleaching recovery. *Biophys. J.* 71(2), 1140, 1996.

Swaminathan, R., Hoang, C. P. & Verkman, A. S. Photobleaching recovery and anisotropy decay of green fluorescent protein GFP-S65T in solution and cells: Cytoplasmic viscosity probed by green fluorescent protein translational and rotational diffusion. *Biophys. J.* 72(4), 1900, 1997.

Swofford, D. L., Olsen, G. J., Waddell, P. J. & Hillis, D. M. Phylogenetic inference. in *Molecular Systematics 2nd edition*, edited by Hills. D. M., Moritz, C. & Mable, B. K., Sunderland, 1996.

Swofford, D. L. Paup - Phylogenetic analyses using parsimony and other methods (Version 4b8). *Sinauer Associates*, Sunderland, MA, USA.

Sznitman, A.-S. Brownian motion, obstacles and random media. *Springer-Verlag Berlin Heidelberg (Berlin, Heidelberg, New York)*, ISBN 3-540-64544-3, 1991.

Tajbakhsh, J., Luz, H., Bornfleth, H., Lampel, S., Cremer, C. & Lichter, P. Spatial distribution of GC- and AT-rich DNA sequences within human chromosome territories. *Exp. Cell. Res.* 255, 229-237, 2000.

Takahashi, M. A fractal model of chromosomes and chromosomal DNA replication. 117-136, 1989a.

Takahashi, M. A fractal model of chromosomes and chromosomal DNA replication. *J. Theor. Biol.* 141, 117-136, 1989b.

Tanabe, H., Müller, S., Neusser, M., von Hase, J., Calcagno, E., Cremer, M., Solovei, I., Cremer, C. & Cremer, T. Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. Proc. Nat. Acad. Sci. USA, 4424-4429, 2002.

Taylor, R. P. Micolich, A. P. & Jonas, D. Fractal analysis of Pollock's drip paintings. *Nature* 399, 422, 1999.

Thacker, J. The role of homologous recombination processes in the repair of severe forms of DNA damage in mammalian cells. *Biochimie* 81, 77–85, 1999.

Trask, B. J., Massa, H., Kenwrick, S. & Gitschier, J. Mapping of human chromosome Xq28 by two-color fluorescence in situ hybridization of DNA sequences to interphase cell nuclei. Am. J. Hum. Genet. 48 (1), 1-15, 1991.

Trask, B., Pinkel, D. & van den Engh, G. The proximity of DNA sequences in interphase cell nuclei is correlated to genomic distance and permits ordering of cosmids spanning 250 kilobase pairs. *Genomics* 5, 710-717, 1989.

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

Traut, W. Chromosomen - Klassische und molekulare Cytogenetik, *Springer Berlin*, 1st ed., 1991.

Tsvetkov, A. G. & Parvenov V. N. Transcriptionally active amphibian oocyte nucleolar chromatin is organized in higher order structure. *Nuclear Structure and Function. by Harris, J. R. & Zbarsky, I. B. (Editors)*, Plenum Press, New York - London, 203-206, 1990.

Tsvetkov, A.G. & Parfenov, V. N. Organizatsiia dispergirovannogo nukleoliarnogo khromatina i ul'trastruktura iadryshek v oogeneze travianoi liagushki. (The organization of dispersed nucleolar chromatin and the ultrastructure of the nucleoli in the oogenesis of the common frog.) *Tsitologiia* 37 (7), 567-573, 1995.

Utsumi, K. R. Studies on the structure of chromosomes. II. Chromosome fibers as revealed by scanning electron microscopy. *Cell Struct. Funct.* 6, 395-401, 1981.

Uyttersprot, N., Costagliola, S. & Miot, F. A new tool for efficient transfection of dog and human thyrocytes in primary culture. *Mol. Cell. Endocrinol.* 142, 35–39, 1998.

van den Engh, G., Sachs, R. & Trask, B. J. Estimating genomic distance from DNA sequence location in cell nuclei by a random walk model. *Science* 257, 1410-1412, 1992.

van Holde, K. & Zlatanova, J. Chromatin higher order structure: Chasing a mirage? *J. Biol. Chem.* 270 (15), 8373-8376, 1995.

Verdier, P. H. & Stockmayer, W. H. Monte Carlo calculations on the dynamics of polymers in dilute solution. *J. Chem. Phys.* 36(1), 227-235, 1962.

Vijay-Kumar, S. Chandra, N. Dharia, C. & Pehrson, J. R. Crystallization and preliminary X-Ray crystallographic studies of nonhistone region of macroH2A1. *PROTEINS: Struct. Funct. and Genet.* 22, 290-292, 1995.

Viney, J. L. Transgenic and gene knockout mice in cancer research. *Cancer Metastasis Rev.* 14 (2), 77-90, 1995;

Visser A.E., E. R., Jauch A., Little G., Bakker P., Cremer T., Aten J.A. Spatial distribution of early and late replicating chromatin in interphase territories of active and inactive X-chromosomes. *Exp. Cell Res.* in press, 1998.

Viswanathan, R. & Heaney M. B. Direct Imaging of the percolation network in a three-dimensional disordered conductor-insulator composite. *Phys. Rev. Let.* 75(24), 4433-3661, 1995.

Voet, D & Voet, J. G. Biochemistry. second edition, *John Wiley & Sons Inc.*, 1995;

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

Vogelstein, B., Pardoll, D. M. & Coffey, D. S. Supercoiled loops and eucaryotic DNA replication. *Cell* 22, 79-85, 1980.

von Hase, J. Quantitative Bildanalyse der DNA-Verteilung spezifischer Chromosomenterritorien im Zellkern. *Diploma thesis*, Faculty for Physics and Astronomy, Ruprecht Karls University of Heidelberg, 2000.

Voss, R. F., Laiboitz, R. B. & Allessandrini, E. L. Fractal (scalin) clusters in thin gold films near the percolation threshold. *Phys. Rev. Let.* 49(19), 1441-1444,1982.

Voss, R. F. Evolution of long-range fractal correlations and 1/f noise in DNA base sequences. *Phys. Rev. Lett.* 68 (25), 3805-3808, 1992.

Vuppu, A. K., Garcia, A. A. & Vernia, C. Tapping mode atomic force microscopy of scleroglucan networks. *Biopolymers* 43, 89-100, 1997.

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

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

Waldeyer, W. Über Karyokinese und ihre Beziehung zu den Befruchtungsvorgängen. *Archiv für mikroskopische Anatomie* 32, 1-122, 1888.

Wall, M. A., Socolich, M. & Ranganathan, R. The structural basis for red fluorescence in the tetrameric GFP homolog DsRed. *Nat. Struct. Biol.* 7, 1133–1138, 2000.

Warrington, J. A. & Bengtsson, U. High-resolution physical mapping of human 5q31-q33 using three methods: radiation hybrid mapping, interphase fluorescence in situ hybridization, and pulsed-field gel electrophoresis. *Genomics* 24 (2), 395-398, 1994.

Wedekind, P., Kubitscheck, U., Heinrich, O. & Peters, R. Line-scanning microphotolysis for diffraction-limited measurements of lateral diffusion. *Biophys. J.* 71(3), 1621, 1996.

Weidemann, T. Quantitative investigation of distribution, mobility, and binding of fluorescently labeled histones *in vitro* and *in vivo* with fluorescence fluctuation microscopy. *PhD thesis*, Faculty for Biology, Ruprecht Karls University of Heidelberg, 2001.

Weidemann, T., Wachsmuth, M., Knoch, T. A., Müller, G., Waldeck, W. & Langowski, J. Nucleosome density distribution in interphase nuclei of living cells. (in submittance).

Widom, J. Physicochemical studies of the folding of the 100 Å nucleosome filament into the 300 Å filament. *J. Mol. Biol.* 190, 411-424, 1986.

Wiggins, S. Chaotic transport in dynamical systems. Interdisziplinary Applied Mathematics, *Springer-Verlag Berlin Heidelberg (Berlin, Heidelberg, New York)*, ISBN 3-540-97522-5, 1991.

Willis, R. A., Wei, C., Turner, M. J., Callahan, B. P., Pugh, A. E., Barth, R. K., Lord, E. M. & Frelinger, J. G. A transgenic strategy for analyzing the regulatory regions of the human prostate-specific antigen gene: potential applications for the treatment of prostate cancer. *Int. J. Mol. Med.* 1 (2), 379-386, 1998.

Wischnitzer, S. The submicroscopic morphology of the interphase nucleus. *Int. Rev. Cytol.* 34, 1-48, 1973.

Witten, T. A. & Kantor Y. Space-filling constraint on transport in random aggregates. *Phys. Rev. B* 30(7), 4093-4095, 1984.

Wolffe, A. Chromatin: Structure and Function. 2 edit, *Academic Press, London,* 1995.

Woodcock, C. L. & Horowitz, R. A. Chromatin organization reviewed. *Trends in Cell Biol.* 5, 272-277, 1995.

Woodcock, C. L. & Horowitz, R. A. Electron microscopic imaging of chromatin with nucleosome resolution. *Methods. Cell Biol.* 53, 167-186, 1998.

Woodcock, C. L. & Horowitz, R. A. Electron microscopy of chromatin. *Methods.* 12 (1), 84-95, 1997.

Woodcock, C. L. Chromatin fibers observed in situ in frozen hydrated sections. Native fiber diameter is not correlated with nucleosome repeat length. *J. Cell Biol.* 125 (1): 11-19, 1994

Woodcock, C. L., Grigoryev, S. A., Horowitz, R. A. & Whitaker, N. A chromatin folding model that incorporates linker variability generates fibers resembling the native structures. *Proc. Nat. Acad. Sci.* U-S-A. 90 (19), 9021-9025, 1993.

Yokota, H., Singer, M. J., van den Engh, G. J. & Trask, B. J. Regional differences in the compaction of chromatin in human G0/G1 interphase nuclei. *Chrom. Res.* 5 (3), 157-66, 1997.

Yannas, I. V. Synthesis of organs: in vitro or in vivo? *Proc. Natl. Acad. Sci. USA* 97 (17), 9354-9356, 2000.

Yokota, H., van den Engh, G., Hearst, J., Sachs, R. K. & Trask, B. J. Evidence for the organization of chromatin in megabase pair-sized loops arranged along a random walk path in the human G0/G1 interphase nucleus. *J. Cell Biol.* 130 (6), 1239-1249, 1995.

Yu, Z. G. & Chen, G. Y. Rescaled range and transition matrix analysis of DNA sequences. *Commun. Theor. Phys. (Beijing)* 33, 673-678, 2000.

Yu, Z., Anh, V. V. & Wang, B. Correlation property of length sequences based on global structure of the complete genome. *Phys. Rev. E* 63 (1), 011901-011908, 2000.

Yu, Z. G., Hao, B., Xie, H., Chen, G. Dimensions of fractals related to languages defined by tagged strings in complete genomes. *Chaos, Solit. & Fractals* 11, 2215-2222, 2000.

Yunis, J. J. (1981). Mid-prophase human chromosomes. The attainment of 2000 bands. *Hum. Genet.* 56, 293-298, 1981.

Zachar, Z., Kramer, J., Mims, I. P. & Bingham, P. M. Evidence for channeled diffusion of pre-mRNAs during nuclear RNA transport in metazoans. *J. Cell Biol.* 121 (4), 729-42, 1993.

Zink, D. & Cremer, T. Chromosome dynamics in nuclei of living cells. *Curr. Biol.* 8 (9), R321-4, 1998.

Zink, D., Cremer, T., Saffrich, R., Fischer, R., Trendelenburg, M. F., Ansorge, W. & Stelzer, E. H. Structure and dynamics of human interphase chromosome territories in vivo. *Hum. Genet.* 102 (2), 241-51, 1998.

Zink, D., Bornfleth, H., Visser, A., Cremer, C. & Cremer, T. Organization of early and late replicating DNA in human chromosome territories. *Exp. Cell Res.* 247, 176-188, 1999.

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

Zorn, C., Cremer, T., Cremer, C. & Zimmer, J. Laser UV microirradiation of interphase nuclei and post-treatment with caffeine. A new approach to establish the arrangement of interphase chromosomes. *Hum. Genet.* 35 (1), 83-9, 1976.

Zorn, C., Cremer, C., Cremer, T. & Zimmer, J. Unscheduled DNA synthesis after partial UV irradiation of the cell nucleus. Distribution in interphase and metaphase. *Exp. Cell. Res.* 124 (1), 111-9, 1979.