UNIVERSITY OF CALIFORNIA, SAN DIEGO

The 3-D Structure of the Immunoglobulin Heavy Chain Locus: Implications for Long-Range Genomic Interactions

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Biological Sciences

by

Suchit Jhunjhunwala

Committee in charge:

Professor Cornelis Murre, Chair
Professor Mark Ellisman
Professor Terry Hwa
Professor Maho Niwa
Professor Kit Pogliano

2009

The dissertation of Suchit Jhunjhunwala is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Chair

University of California, San Diego

2009

DEDICATION

I dedicate this work to my parents, for their priceless love and support. For all the undisclosed hardships they have endured for me.

EPIGRAPH

Let me anticipate what will be explained in much more detail later, namely, that the most essential part of a living cell -the chromosome fiber- may suitably be called an aperiodic crystal. In physics we have dealt hitherto only with periodic crystals. To a humble physicist's mind, these are very interesting and complicated objects; they constitute one of the most fascinating and complex material structures by which inanimate nature puzzles his wits. Yet, compared with the aperiodic crystal, they are rather plain and dull. The difference in structure is of the same kind as that between an ordinary wall paper in which the same pattern is repeated again and again in regular periodicity and a masterpiece of embroidery, say a Raphael tapestry, which shows no dull repetition, but an elaborate, coherent, meaningful design traced by the great master.

—E.W. Schrödinger, What is Life? 1944.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost I would like to thank Kees for being such a wonderful advisor and for his inductive enthusiasm. Its been a privilege to work under the guidance of such a bright scientist. All the members of the Murre Lab, past and current, have been very supportive and helpful.

The text of chapters 1 and 5 is partly reprinted from the paper *"Chromatin architecture and the generation of antigen receptor diversity"*, Cell, 138(3), 2009.

The text of chapter 2 is partly reprinted from the paper *"Visualization of Looping Involving the Immunoglobulin Heavy-chain Locus in Developing B Cells"*, Genes and Development, 19(3), 2005. I would like to thank Camil Sayegh to make me an integral part of the project when I joined the lab. Roy Riblet helped us with his expertise in the genetic organization of the genetic locus and provided us with some hybridization probes.

Parts of the other chapters - 3, 4, and 5, are reprinted from the paper *"The 3D-Structure of the Immunoglobulin Heavy Chain Locus: Implications for Long-Range Genomic Interactions"*, *Cell*, 133(2), 2008. I have spent years working in the dark along with Menno van Zelm and Mandy Peak to generate the data for this project. It would not have been possible to survive, as we call it, the 'dungeon' or the 'beach' without them. Roy Riblet once again helped us with the map and small probes for the locus. Tobias Knoch was the backbone of the computer

simulations for RW/GL and MLS models in this project, and it was his work that inspired us to compare the data to structural models. Steve Cutchin and Amit Chourasia, at the San Diego Supercomputer Center helped us in the trilateration study and made some wonderful 3D movies and figures. I thank Jacques van Dongen, Frank Grosveld and Douglas Forbes for their help with the manuscript. I thank Pete Carlton and John Sedat (UCSF) for their generous help with using their SIM system. Terry Hwa provided us his expert advice in comparing our data to random walk models and inferring compartmentalization.

Maho Niwa, Mark Ellisman, Terry Hwa, Kit Pogliano - members of my thesis committee have always been helpful with continually improving the study. While exploring different avenues to approach the structural organization of the locus, I conducted a number of electron microscopy experiments at Mark Ellisman's lab. I thank the members of his lab, especially Mason Mackey and Ben Giepmans for their generous help. Kit Pogliano helped us determine the resolution of the microscopy system.

VITA

1998-2003	B.Tech. & M.Tech., Biochemical Engineering and Biotechnology, Indian Institute of Technology, Delhi.
2003-2009	Ph.D., Biological Sciences, University of California San Diego.

PUBLICATIONS

Jhunjhunwala S., van Zelm M.C., Peak M.M., Murre C., "Chromatin architecture and the generation of antigen receptor diversity", *Cell*, 138(3), 2009.

Jhunjhunwala S., Van Zelm M.C., Peak M., S. Cutchin M.M., Riblet R., Van Dongen J.J., Grosveld F.G., Knoch T.A., Murre C., "The 3D-Structure of the Immunoglobulin Heavy Chain Locus: Implications for Long-Range Genomic Interactions", *Cell*, 133(2), 2008.

Sayegh C.E., Jhunjhunwala S., Riblet R., Murre C., "Visualization of Looping Involving the Immunoglobulin Heavy-chain Locus in Developing B Cells", *Genes Dev*, 19(3), 2005.

ABSTRACT OF THE DISSERTATION

The 3-D Structure of the Immunoglobulin Heavy Chain Locus: Implications for Long-Range Genomic Interactions

by

Suchit Jhunjhunwala

Doctor of Philosophy in Biological Sciences

University of California San Diego, 2009

Professor Cornelis Murre, Chair

The immunoglobulin heavy chain (Igh) locus is organized into distinct regions that encode multiple variable (V_H), diversity (D_H), joining (J_H) and constant (C_H) gene segments. DNA recombination takes place between the V_H, D_H and J_H segments at the Igh locus in developing B cells. The locus undergoes large-scale contraction to facilitate this recombination. However, its structural organization is unknown. It is likely that the structural organization plays a role in this process.

By simultaneously visualizing three subregions of the Igh locus using 3D fluorescence in-situ hybridization, we show that looping of the distal V_H segments to the C_H segments is observable in pro-B cells. This looping occurs at a

significantly higher frequency in pro-B cells compared to $CD8^+$ T cells. This indicated that there is a structural reorganization of the locus and not just a simple contraction of the chromatin fiber.

To decipher the topology of the locus, 12 genomic markers were used that spanned the entire locus. Spatial distance distributions between different combinations of these markers were determined and compared to computer simulations of different models of chromatin structure. These comparisons revealed that the data agreed with a topology that predicted higher order organization of the chromatin fiber into multiple subcompartments connected by linkers (Multi-Loop Subcompartment model). Compartmentalization of the locus was visualized by labeling the entire locus with hybridization markers. Relative locations of the different Igh sub-regions in 3D space were determined using a trilateration technique. Striking conformational changes can be seen between pre-pro-B and pro-B cells, when the locus transitions from a de-contracted to a contracted state.

The implications of the higher order organization of the locus on long-range genomic interactions are discussed. It is evident that the higher order organization is necessary for promoting long-range genomic interactions that would facilitate V(D)J recombination at the Igh locus. In absence of higher order organization, the expected frequency of interactions are much lesser.

Chapter 1

Introduction

The structure-function relation has become an increasingly important paradigm in the study of biological molecules. Immunoglobulins, or antibodies, are an integral part of the immune system that recognize potentially an infinite variety of foreign antigens, owing to their highly variable structure. The structural variability is possible due to the stochastic process of DNA recombination at the immunoglobulin gene. This study focuses on a hitherto unexplored structure-function relationship, that between the structure of a genetic locus and DNA recombination. We will highlight how the nuclear organization, chromatin topology and chromatin dynamics of the immunoglobulin heavy chain locus relate to its function.

1.1 Genetic Organization of the Antigen Receptor Loci of the Lymphoid System

The lymphocyte compartment consists of cells that express a diverse repertoire of antigen receptors, enabling organisms to generate a unique immune response to invading pathogens. B lymphocytes produce B-cell receptors (BCR), while the antigen receptors in T cells are called T-cell receptors (TCR). Dreyer and Bennett first proposed that antigen receptor diversity is generated by DNA recombination[1]. This original insight was confirmed in later studies, revealing that antigen receptor loci are organized into distinct genomic regions that contain variable (V), diversity (D) and/or joining (J) and constant (C) coding elements[2, 3, 4]. Since this early work, the understanding of the biochemical and molecular mechanisms that underpin the assembly of antigen receptor genes from their coding elements has blossomed. Each antigen receptor is built from two types of polypeptides, encoded by independent genetic loci in the genome.

The murine immunoglobulin heavy chain locus (Igh), which encodes the Ig heavy chain, consists of a single massive stretch of DNA (3 Mbp) in length, which is divided into distinct DNA elements encoding the variable (V), diversity (D), joining (J), and constant (C) regions (Figure 1.1A). Each subregion displays much complexity. For example, fifteen partially dispersed V region families encode approximately 195 V_H gene segments, depending on the genetic background, each

of which is approximately 500 bp in size. The density of the gene segments within the V-region cluster is relatively low, containing large intergenic regions up to 50 Kbp in size. Downstream of the V_H regions are located 10-13 D_H genes and four J_H genes as well as eight C regions that encode the various isotypes including C_μ, C_δ, $C_{\gamma 1}$ $C_{\gamma 2a}$, $C_{\gamma 2b}$, $C_{\gamma 3}$, C_ϵ and C_α.

The light chain of immunoglobulins is produced by one of two loci, Igκ or Igλ. The Igκ locus is comprised of approximately 120 V_κ gene segments that span almost 3 Mbp, a J_κ cluster, and a single constant region positioned within very close proximity (2.5 Kbp) to the J_κ cluster (Figure 1.1B). The organization of the Igλ locus is quite distinct from that of the Igh and Igκ loci. Rather than a common set of J gene segments located upstream of the constant region(s), the four constant regions of Igλ each contain their own unique J_λ gene segment. Moreover, only two V region gene segments, $V_\lambda 1$ and $V_\lambda 2$ are frequently utilized. $V_\lambda 2$ is located approximately 60 Kbp from $J_\lambda 2$. It will not recombine with other J_λ gene segments. On the other hand, $V_\lambda 1$, located 22 Kbp from the $J_\lambda 1$, will form joints with either $J_\lambda 1$ or $J_\lambda 3$ (Figure 1.1C). Thus, each chain of an antibody is produced using a similar theme of combining distinct gene segments.

The organization of genes encoding the T-cell receptor (TCR) gene segments is strikingly similar to that of the immunoglobulin loci (Figure 1.1, 1.2). Two distinct T-cell lineages, named $\alpha\beta$ and $\gamma\delta$ T cells, develop in the thymus from early T-lineage progenitor cells. These two cell types were originally identi-

fied based on the expression of either the $\alpha\beta$ or $\gamma\delta$ T-cell receptor. The $\alpha\beta$ TCR consists of two chains, the α chain linked to a β chain, while the $\gamma\delta$ TCR is composed of a heterodimeric complex containing the γ and δ chains. The TCRβ locus spreads over approximately 650 Kbp of genomic DNA (Figure 1.2B). It contains 31 V_β gene segments, of which twenty are functional and located upstream from two $D_\beta J_\beta$ clusters and two C_β regions. Each of the $D_\beta J_\beta$ clusters contains a single D_β and six J_β gene segments. Oddly, one of the V_β gene segments is located down-stream of the C_β regions.

The TCRα locus hews to a similar theme in that it is comprised of approximately 100 V gene segments located within a 1.5 Mbp region (Figure 1.2A). At least 200 Kbp separates the V_α regions from the J_α cluster. The TCR_α locus is unusual in that it contains many more J regions as compared to other antigen receptor loci, with 61 J_α gene segments that span 65 Kbp. Nested within the TCRα locus is the TCRδ locus, containing numerous V_δ, two D_δ and J_δ elements and one C_δ region. Unlike the TCRα, TCRβ and TCRδ loci, the TCRγ locus is small (less than 200 Kbp), containing few V_γ and J_γ gene segments (Figure 1.2C). Thus, the majority of antigen receptor loci are comprised of large numbers of V regions that span a vast genomic region and numerous clustered D or J gene segments.

It can be seen that the different antigen receptor loci, although expressed in different systems, have a common theme of organization. It is also likely

that a common structural organization underlies this theme, promoting DNA recombination at these loci in lymphoid cells during early stages of development.

1.2 Lymphocyte Development

Pluripotent hematopoietic stem cells, which have self-renewal capacity, develop into lymphoid-primed multipotent progenitors (LMPPs) that lack long-term self-renewal capacity, and have myeloid or lymphoid restricted differentiation potential. LMPPs develop into common lymphoid progenitors, which differentiate to become pre-pro-B cells. Pre-pro-B cells, in turn, develop into committed pro-B cells, that initiate and complete Igh V(D)J gene rearrangement. At the pro-B-cell stage, $D_H J_H$ joining precedes that of $V_H D_H J_H$ gene rearrangement. Once a productive $V_H D_H J_H$ gene rearrangement has been generated, a pre-B-Cell Receptor (pre-BCR) is assembled which acts, in turn, to inhibit the expression of the recombinase, RAG1 and RAG2, and promotes the survival and expansion of developing large pre-B cells. This proliferation phase is followed by exit from the cell cycle, during which RAG gene expression is re-induced to activate Igκ gene rearrangement. In the presence of auto-reactivity, continued Igκ VJ gene rearrangement will replace primary Igκ VJ joints, generating BCRs with novel specificities. If self-reactivity still continues, Igλ VJ gene rearrangement will be initiated to generate an Igλ chain. Once a BCR has been assembled that

lacks auto-reactivity, tonic BCR signaling will permanently suppress recombinase gene expression and promote developmental progression and migration into the peripheral lymphoid organs.

The development of $\alpha\beta$ T and $\gamma\delta$ T cells in the thymus is a process characterized by the sequential rearrangement of the variable gene segments of the T cell antigen receptor loci. Shortly after arriving in the thymus, T cell progenitors initiate rearrangement of the TCRβ chain. The rearrangement of the TCRβ locus is initiated and completed at a developmental stage that lacks the expression of the co-receptors for the T-cell receptor, CD4 and CD8, a stage commonly referred to as the double negative (DN) compartment. Upon rearrangement and expression of a productive TCRβ chain and its assembly into a pre-TCR complex, recombinase expression is suppressed and thymocytes undergo developmental progression that is characterized by rapid cellular expansion. During this phase thymocytes begin to express CD8, followed by CD4. Thymocytes that express CD4 and CD8, also named double positive (DP) cells, undergo cell cycle arrest and initiate TCRα VJ gene rearrangement. $V_\alpha J_\alpha$ rearrangements can be initiated multiple times, such that secondary TCRα rearrangements can replace progressively primary $V_\alpha J_\alpha$ joints. The primary rearrangements predominantly utilize the more 3' located V_α segments and the most 5' J_α elements. DP thymocytes then initiate the processes of positive and negative selection leading to the maturation of only those cells that express TCRs with moderate affinity for antigen-containing major

histocompatibility complexes presented by thymic epithelial antigen presenting cells. Positively selected thymocytes decrease expression of either CD4 or CD8 to develop into mature CD8 or CD4 single positive (SP) T cells. The TCRγ and TCRδ loci also undergo rearrangement in the DN compartment.

Antigen receptor rearrangements are mono-allelic in nature, and two distinct mechanisms may ensure mono-allelic rearrangement. First, antigen receptor rearrangement is mono-allelically activated, most well characterized for the Igκ locus[5]. Second, once a productive Igh or TCRβ V(D)J gene rearrangement has been generated, signaling mediated by the pre-BCR or pre-TCR antagonizes continued rearrangement by a feed-back mechanism[6, 7]. As a result, only one copy of a functional antigen receptor gene is produced per lymphocyte. Thus, the adaptive arm of the immune system is generated by distinct cell types, which undergo ordered gene rearrangement, to equip each lymphocyte with a single and unique BCR or TCR.

1.3 Nuclear Localization and Structural Reorganization of Antigen Receptors

The nuclear locations of genes vary during developmental progression[8]. Such repositioning of loci during differentiation is often conserved throughout evolution. A prominent example involves multiple genes in yeast that appear

to cluster at the nuclear membrane / nuclear lamina[9]. A subset of these genes is associated with nuclear pores that are undergoing active transcription whereas others that are associated with the inner nuclear membrane / nuclear lamina appear to be transcriptionally silenced. In mammalian cells, the expression of the β-globin locus is initiated at the nuclear membrane prior to its movement towards more centrally located domains[10]. A second nuclear compartment involves the pericentromeric heterochromatin. The association of genes with pericentromeric heterochromatin is often correlated with transcriptional silencing.

Evidence is accumulating indicating that the assembly of immunoglobulin loci is also regulated by nuclear location. The inner nuclear membrane / nuclear lamina associates with the Igh locus in hematopoietic progenitors[11]. The distal V_H region cluster is tethered to the nuclear membrane whereas the $D_H J_H$ elements are located away from the membrane (Figure 3). The spatial orientation of the Igh locus tethered to the inner nuclear membrane may allow the recombinase access to the $D_H J_H$ domain to promote $D_H J_H$ rearrangement while the V_H cluster remains closed. The presence of $D_H J_H$ rearrangements in hematopoietic progenitors, but the absence of $V_H - D_H J_H$ joints is consistent with such an orientation.

Upon commitment to the B-cell lineage, the Igh locus undergoes internalization and large-scale contraction, followed by V_H-$D_H J_H$ gene rearrangement[12, 11, 13, 14]. The observed contraction is not well understood, but it is believed to facilitate DNA recombination by bringing the V segments physically closer to the D/J seg-

ments. Once a productive V_H-$D_H J_H$ rearrangement has occurred, both Igh alleles undergo de-contraction[13] (Figure 3). The de-contraction of antigen receptor loci has been postulated to suppress further rearrangement and to enforce the allelic exclusion mechanism[13]. In addition to de-contraction, recruitment of the non-productive Igh allele to the repressive pericentromeric neighborhood further ensures allelic exclusion[13]. The Igκ locus also undergoes relocation and contraction / de-contraction. In summation, the antigen receptor loci undergo nuclear relocation and structural reorganization to regulate DNA recombination.

1.4 Ordered Assembly of Antigen Receptor Loci and Chromatin Territories

The rearrangement of antigen receptor loci is ordered. In pro-B cells, Igh $D_H J_H$ rearrangement is initiated prior to V_H-$D_H J_H$ joining[15]. TCRβ rearrangement is also sequential, i.e. D_β-J_β rearrangements occur prior to V_β-$D_\beta J_\beta$ joining. As described above, once a productive Igh or TCRβ V(D)J gene rearrangement has been generated, signaling mediated by the pre-BCR or pre-TCR antagonizes continued rearrangement by a feedback mechanism to ensure allelic exclusion[7] (Jung et al., 2002). Feedback signaling, however, does not suppress rearrangement of the entire Igh or TCR$_\beta$ V region repertoire. The four most proximal V_H regions as well as a V_β region, located downstream rather than 5' of the D_β-J_β cluster, escape the

allelic exclusion mechanism[16, 13, 17]. Thus we are faced with the question: what might explain these differences in V region usage?

Functional chromosomal domains may provide a means by which regions undergoing DNA recombination are distinguishable from regions not undergoing DNA rearrangement. Such functional domains have been described for the chicken β-globin cluster, which is marked by the presence of boundary elements, containing CTCF binding sites[18]. To determine whether such functional domains exist within antigen receptor loci, V_β and V_H regions, normally distally located and subject to the allelic exclusion mechanism, were inserted in a chromosomal position immediately 5' of the D domain or within the D-J cluster[19, 20]. Interestingly, the targeted V_H regions appeared to rearrange with substantially higher frequency than their endogenous counterparts. Furthermore, targeting of the V_H region to a location immediately 5' of the $D_H J_H$ region revealed that the ordered rearrangement process was perturbed as well[19]. As suggested previously, these data are consistent with a model in which the majority of V_H and D_H domains are located in functionally separate territories[19].

1.5 Chromatin Structure

The 3-dimensional organization of antigen receptor loci is not known. The structure is likely to permit gene segments to encounter each other frequently.

Resolving this question requires insight into long-range chromatin structure and dynamics through polymer physics.

1.5.1 The Chromatin Fiber

The unit of the chromatin fiber is the nucleosome. A nucleosome consists of a 146 bp DNA segment wrapped around an octamer consisting of two copies each of histones H2A, H2B, H3 and H4. The nucleosomes form a 10 nm fiber creating a structure resembling 'beads on a string'. This 10 nm fiber, in the presence of histone H1, condenses into a more compact 30 nm fiber, of which the precise structure is still not completely resolved[21]. How the chromatin fiber is folded into higher order structures beyond the 30 nm fiber remains largely unknown, but involves increasingly higher order folding. In the late 1970s and early 1980s, distinct folding patterns for chromosome structure were proposed, including topologies involving helical, radial or combined loop-helical folding[22, 23, 24]. Using electron microscopic analyses of chromosome spreads, Laemmli and collaborators showed that chromosomes appeared comprised of loops of ∼90 Kbp in size. It was postulated that such loops interact with a putative nuclear matrix during mitosis and cluster further into rosettes containing on average ∼18 loops, yielding ∼100 rosettes per average mitotic chromosome[23, 25]. More recently, serial thin-section electron microscopy has suggested a different topology for chromatin structure, namely a *chromonema* fiber, where a chain of diameter of 60-130 nm is interspersed

by more loosely folded segments that have a diameter of 30 nm[26].

1.5.2 Flexibility of the Chromatin Fiber

In essence, a chromatin fiber, because of its repetitive nature, resembles a polymer, and many of its physical properties can be analyzed in terms of polymer models established by Edwards, Flory, Kuhn and de Gennes (Morawetz, 1985). The most important parameter determining the flexibility of a polymer is the *persistence length*, which is defined as the minimum contour length of the polymer separating two segments of the chain needed for the segments to become decorrelated from each other, i.e., the orientation of one segment does not affect the orientation of the other. It is a statistical measure of the flexibility of a polymer. The persistence length for naked DNA is ∼50 nm. The persistence length for a chromatin fiber remains controversial but depending on the experimental strategy by which it was measured, has been found to vary between 30-200 nm[27].

1.5.3 Shape and Dynamics of a Simple Polymer Chain

The shape and dynamics of a polymer can be described in terms of a random walk behavior where the polymer is assumed to be made up of a series of rigid individual units, named Kuhn segments, linked by flexible hinges, allowing free rotation between consecutive Kuhn segments. Under this assumption, the flexibility is defined by the length of a Kuhn segment, which is twice the persistence

length. Several types of random walk models have been used to describe the structure and dynamics of polymer chains (Figure 1.4). In a freely jointed chain model, the hinges connecting adjacent Kuhn segments are completely free to rotate, and the polymer segments are allowed to overlap with each other, i.e., the orientation of one segment is independent of the orientation of its two adjacent segments (Figure 1.4A). A more realistic model frequently applied to describe polymer chain behavior is the self-avoiding chain (Figure 1.4B). A self-avoiding chain is similar to the freely jointed chain, except that the chain cannot cross its own path, i.e., the Kuhn segments cannot intersect with each other[28]. Yet another model is the worm-like or Kratky and Porod chain, which considers the polymer as a continuously flexible chain rather than freely jointed discrete segments (Kratky and Porod, 1949) (Figure 1.4C). The persistence length is the minimal repeating unit in this model and owes its conceptual origin to this model. Recent studies have described the dynamics of the yeast chromatin fiber in terms of the worm-like chain[29].

1.5.4 Shape and Dynamics of the Chromatin Fiber

Cytological analysis of chromosomes in mammalian cells have demonstrated a confined geometry for the chromatin fiber to chromosome arms as well as bands, arguing against the predictions of purely random walk behavior. Furthermore, spatial distance measurements as a function of genomic separation demonstrated

that the spatial distance scales as a function of genomic separation over 4 Mbp, with exponents that are incompatible with that of free random walk statistics[30, 31, 32, 33]. As a first approach to described the chromatin fiber in vertebrate nuclei, a chromatin topology named the Random Walk/Giant Loop model (RW/GL), has been proposed in which the chromatin fiber was described as undergoing random walk behavior but confined to large loops (2-5 Mbp)[30] (Figure 1.4D). However, measurements of spatial distances between genomic markers, that are spaced less than 4 Mbp apart, did not agree well with the RW/GL model[33]. Based on these observations, as well as on the original cytological observations made by Laemmli and collaborators in late 1970, the Multi-Loop-Subcompartment (MLS) model was proposed to describe long-range chromatin folding[34, 33]. The MLS model implies that the chromatin fiber is folded into rosettes, which are connected by flexible linkers (Figure 1.4E). Most recently, yet another topology, the Random-Loop (RL) model, has been proposed to describe the spectrum of chromatin configurations and dynamics (Figure 1.4F). Whereas loops in the RW/GL and MLS models are assumed uniform in size, the RL model, for genomic separation of 5-10 Mbp, permits both large and small loops to form, and loops can dynamically associate and dissociate from loop attachment points[35]. The development of novel computational methods to model new chromatin topologies should help to define experimental strategies that either support or reject model predictions.

The principal question addressed in this study concerns the role of chromatin topology in long-range genomic interactions at the immunoglobulin heavy chain locus. Comparison of experimental data to polymer-models of chromatin structure has permitted insights into our still-rudimentary knowledge of long-range genomic interactions and the ensemble of topologies that are adopted by antigen receptor loci, using the Igh locus as a model.

Acknowledgements

The text of this chapter is partly reprinted from the paper *"Chromatin architecture and the generation of antigen receptor diversity", Cell, 138(3), 2009.*

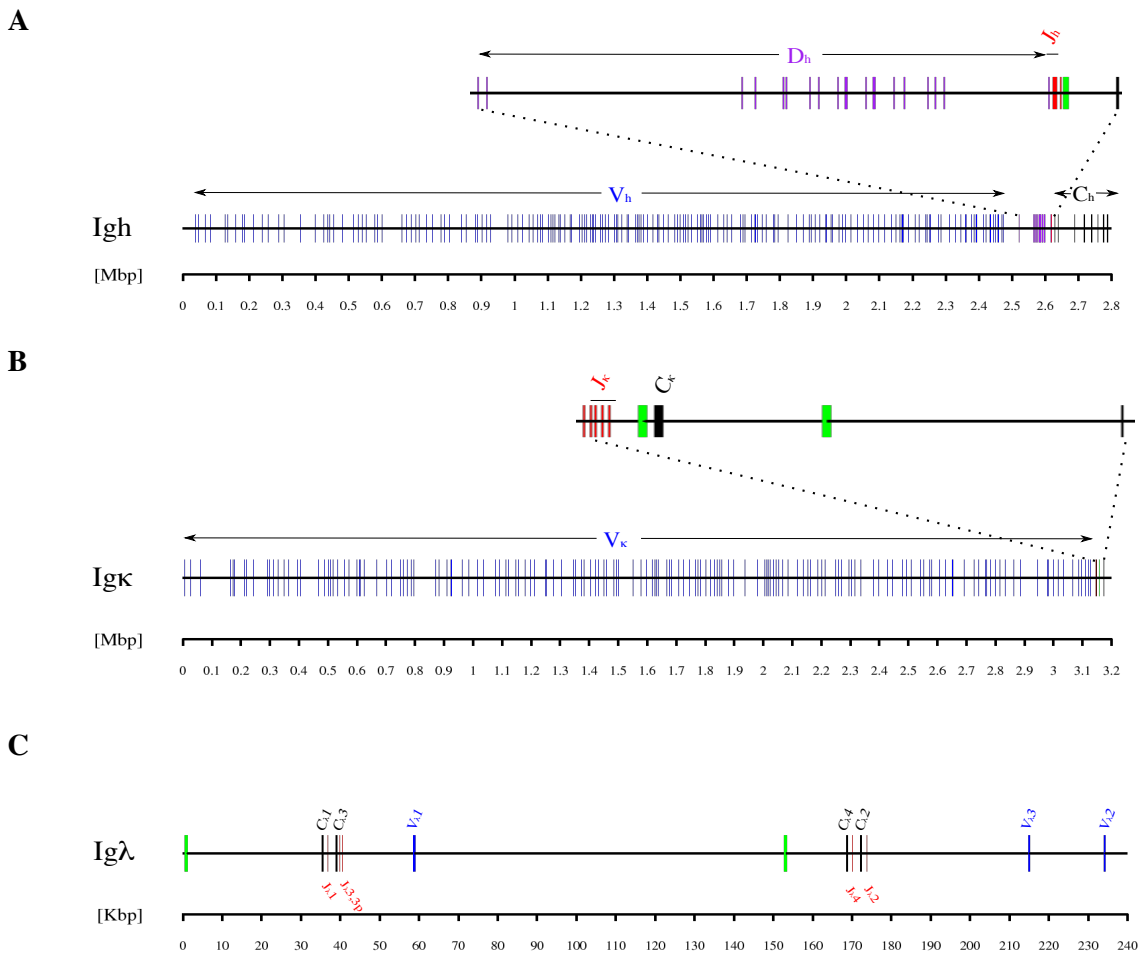


Figure 1.1: Genetic Organization of the Immunoglobulin Loci

The various coding elements for the immunoglobulin loci are organized as shown. The V segments are shown in blue. D segments (if present) are shown in purple. J segments are shown in red and the C segments are in black. Known enhancer elements in the loci are in green.

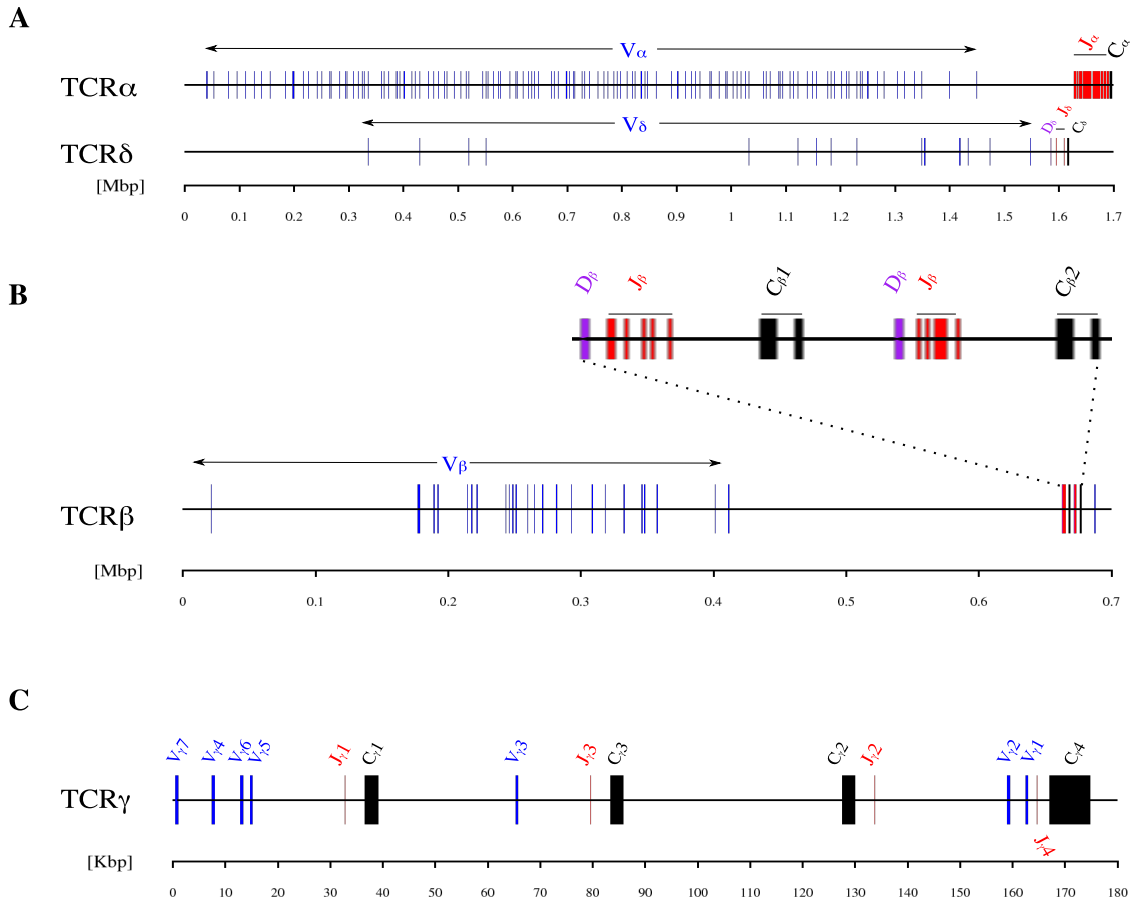


Figure 1.2: Genetic Organization of the TCR Genes

The various coding elements for the TCR loci are organized as shown. The V segments are shown in blue. D segments (if present) are shown in purple. J segments are shown in red and the C segments are in black. Known enhancer elements in the loci are in green.

Figure 1.3: Nuclear Location and Compaction of Antigen Receptor Loci during Lymphocyte Development
Nuclear locations of antigen receptor loci during developmental progression are indicated. (A) Indicated are the nuclear positions of the Igh and Igκ alleles at various stages of early B cell development. Pre-pro-B and pro-B cell stage are shown in blue. Large pre-B and small pre-B are indicated in green. Immature and mature-B cell compartments are shown in red. Blue cluster of loops indicate V regions. Red clusters of loops represent D/J/C coding elements. Dark dots represent the centromeric heterochromatin. Relative degree of Igh and Igκ locus compaction and de-compaction is depicted. (B) Indicated are the nuclear positions of the TCRβ and TCRα alleles during thymopoieisis. Lymphoid progenitors and the DN cell stage are shown in blue. DP compartment is indicated in green. SP cell stage is shown in red. Blue cluster of loops indicates V regions. Red clusters of loops represent D/J/C coding elements. Dark dots represent the centromeric heterochromatin. Relative degree of TCRα and TCRβ compaction and de-compaction are depicted.

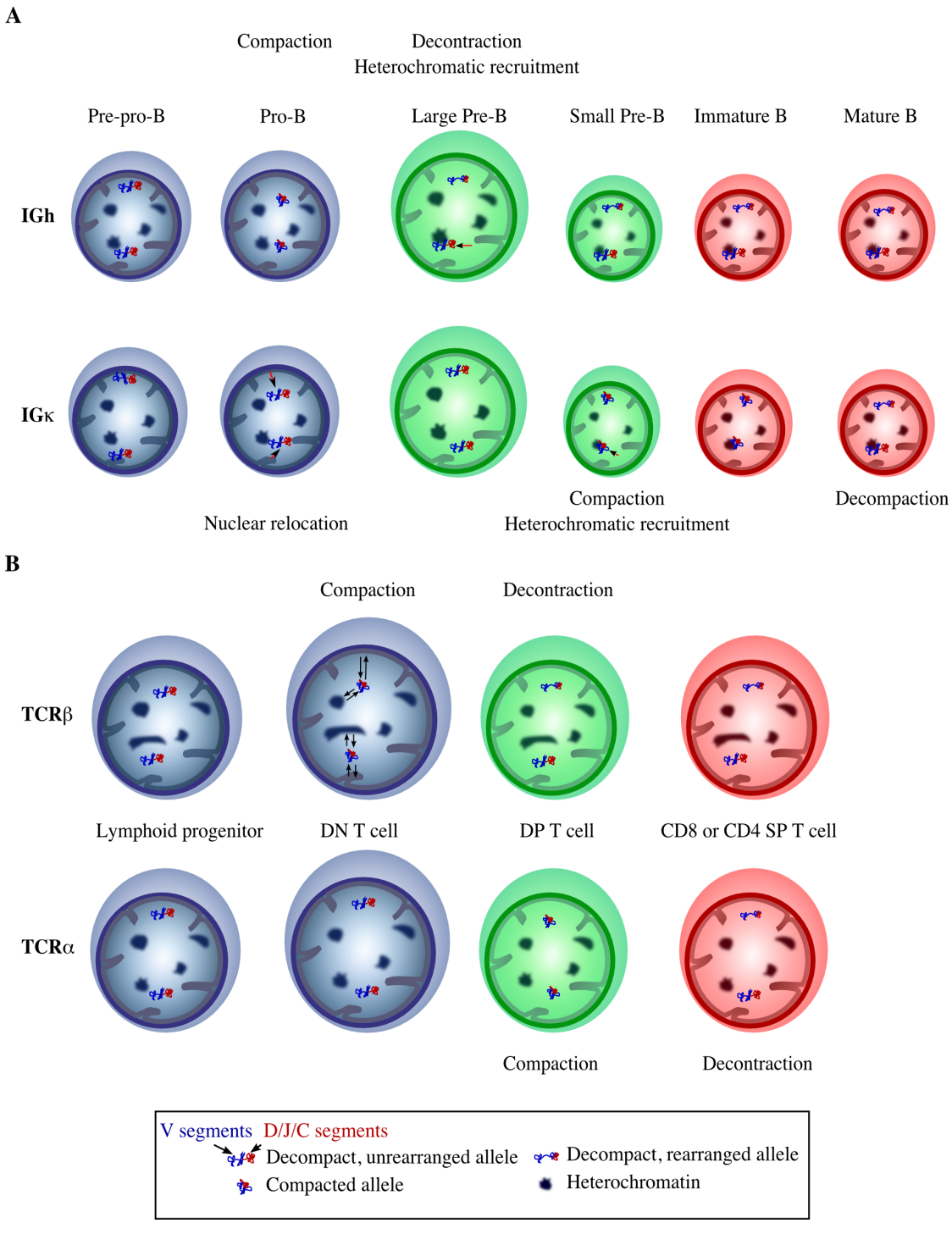
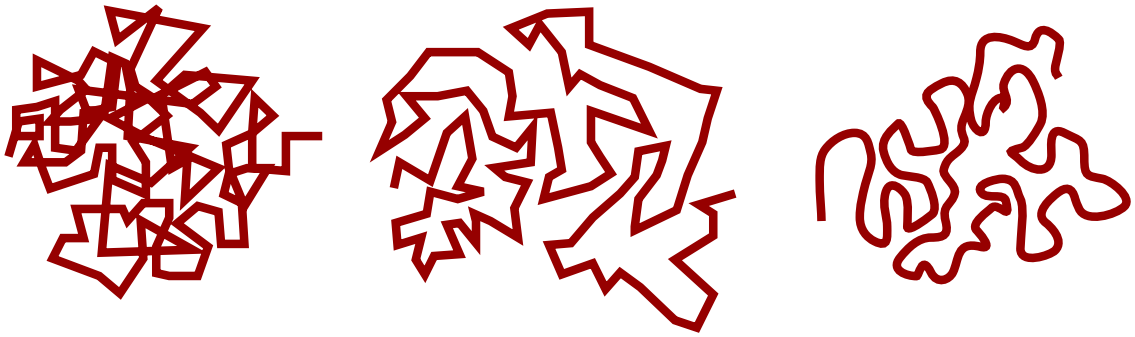

A
Compaction
Decontraction
Heterochromatic recruitment
Pre-pro-B
Pro-B
Large Pre-B
Small Pre-B
Immature B
Mature B
IGh
IGκ
Nuclear relocation
Compaction
Heterochromatic recruitment
Decompaction
B
Compaction
Decontraction
TCRβ
Lymphoid progenitor
DN T cell
DP T cell
CD8 or CD4 SP T cell
TCRα
Compaction
Decontraction
V segments
D/J/C segments
Decompact, unrearranged allele
Decompact, rearranged allele
Compacted allele
Heterochromatin

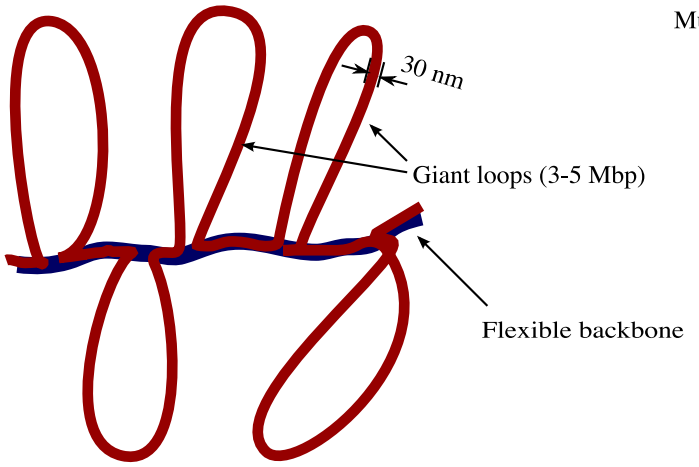
Figure 1.4: Flexible Polymer Chain Models in Free and Confined Environments

(A) Freely jointed chain. A freely jointed chain consists of a series of rigid segments connected by flexible hinges. (B) Self avoiding random walk. In a self-avoiding chain, a segments cannot intersect any other segment. (C) Worm-like chain. In contrast to the freely-jointed chain, which is flexible within the hinges that separate the segments, the worm-like polymer chain is continuously flexible. (D) Random Walk/Giant Loop model. Giant loops of 3-5 Mbp are tethered to a backbone. The DNA within the loops and the backbone itself follow a random walk. (E) Multi-Loop Subcompartment model. Chromatin is organized into 1-2 Mbp subcompartments. Each subcompartment consists of a bundle of loops that are attached to a common loop base. Linkers connect the chromatin subcompartments. Both loops and linkers undergo random walk behavior. (F) Random Loop model. Dynamic loops of all sizes are formed at random intervals on the chromosome. Both individual loops and bundles of loops are shown.

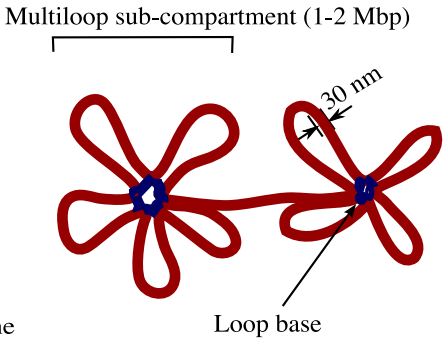
A. Freely jointed chain **B. Self-avoiding chain** **C. Worm-like chain**



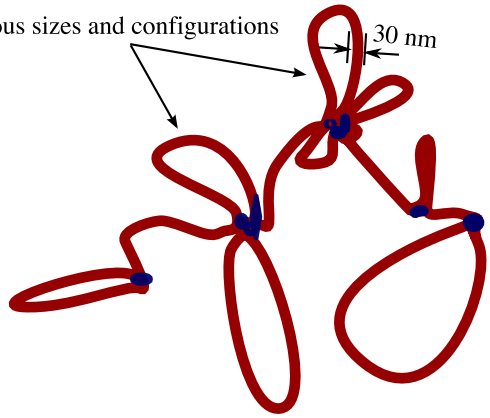
D. RW/GL model



E. MLS model



F. Random Loop model



Chapter 2

Visualization of Looping Involving the Immunoglobulin Heavy Chain Locus in Developing B Cells

The murine Igh locus undergoes nuclear re-localization and structural changes during B cell development in the bone marrow. In non-B lineage cells the locus is in an extended configuration, which is possibly the normal state of chromatin at the locus. At the pro-B cell stage, the locus is in a contracted state, conceivably promoting interaction between V_H segments and the $D_H J_H$ region[11].

The V_H segments at the Igh locus are sub-divided into at least 15 different families based on sequence similarity[36]. Furthermore, they can also be sub-divided into two groups, distal and proximal V_H segments, based on proximity to the $D_H J_H$ segments. The distal V_H family mainly consists of the $V_H 1$ family, also known as the

V_HJ558 family. A functional division between the distal and proximal V_H families is apparent as their regulation is uncoupled. Pax5[12] or Ezh2[37] deficient pro-B cells have significantly reduced rearrangement involving distal V_H segments, whereas rearrangement of proximal V_H segments is only modestly affected. Using ChIP and ChIP-on-chip, A.J. Feeney and colleagues have shown that the active histone modification H3K36me2 is most highly associated with distal V_H segments and the repressive histone modification H3K27me3 is exclusively present on proximal V_H segments[38]. IL-7 receptor signaling, Pax5 and Ezh2 are involved in maintaing these histone marks at the appropriate locations. In short, proximal and distal V_H families are differently regulated for VDJ rearrangement in pro-B cells.

Here, we visualize three domains present in the Igh locus, the constant region, the proximal V_H region and the distal V_H segments, using 3D-fluorescence in situ hybridization. We detect looping of the Igh locus in developing B cells involving the juxtaposition of Igh V regions into proximity of the Igh D/J regions. Interestingly, looping of V-regions was also observed in RAG-deficient pro-B cells. These data indicate that in developing B cells, V_H regions loop in proximity of the $D_H J_H$ cluster prior to the onset of Ig gene recombination. We propose that prior to Ig gene rearrangement, Ig loci are repositioned by a looping mechanism to facilitate joining of Ig gene segments.

2.1 Long-Range 3D-Complexity in Lymphocytes

To explore the mechanisms contributing to the large-scale contraction of the Igh locus observed in developing B cells[11, 39], 4-color 3D-immunofluorescence in situ hybridization (3D-FISH) was used to directly visualize three regions of the Igh locus. Three specific BAC probes were simultaneously used to label the 3' region of V_H1 (green), the V_H 4 and V_H 9 ($V_H9/4$) region (blue), and the Ig Cα exons (red) (Figure 2.1A). Lamin-specific antibodies were used to label the nuclear membrane. Four cell types were used in these experiments: In vitro expanded pro-B cells derived from either wild-type mice or RAG2-deficient mice, in vitro activated $CD8^+$ splenic T cells, and E2A-deficient hematopoietic progenitor cells. Nuclei were hybridized with the Ig specific probes and analyzed by deconvolution microscopy. A fraction of nuclei derived from pro-B cells did not show hybridization to either the V_H 9/4 or to both V_H1 and $V_H9/4$ probes. These were most likely derived from pro-B cells that had completed Igh VDJ rearrangement, since such nuclei were not observed in $RAG2^{-/-}$ pro-B cells and T cells (data not shown). Only cells in which all three signals could be detected from both alleles were included in the analysis. To precisely calculate the distances separating each probe, the 3D-coordinates of the center of mass of each probe were obtained by fitting each signal in each z-section into a polygon (Figure 2.6D, G, J and M). Note that often only one out of two Igh allele are shown because of

the optical sectioning. Polygons obtained for each probe were then interpolated into a 3D-object for which the coordinates of the center of mass were calculated. This analysis revealed striking features of the interphase organization of the Igh locus in the various cell types. The distances separating the V_H1 and $V_H9/4$ regions were significantly reduced in wild-type and RAG2-deficient pro-B cells ($p<0.001$), representing a 41% and 31% contraction respectively, when compared to the distance separating these probes in T cells (Figure 2.1B, 2.6B-D, 2.7A-C, and Table 2.1). Additionally, a significant long-range contraction ($p<0.001$) of the average distance separating the $V_H9/4$ and Cα regions was observed in wild-type (27%) and RAG2-deficient pro-B cells (46%) when compared to that of T lineage cells (Figure 2.1C, Table 2.1). Consistently, we observed a smaller proportion of Igh alleles in pro-B cells showing distances >0.4 μm and an increase in the frequency of alleles showing a distance between 0-0.2 μm when compared to $CD8^+$ T cells (Figure 2.2). Additionally, the relative distribution of distances separating each of the three regions appeared equally scattered in pro-B cells and T cells, suggesting that the Igh locus undergoes contraction in most pro-B cells (Figure 2.3).

It could be argued that Igh locus contraction in pro-B cells was caused by nuclear size differences. To address this question, the nuclear diameter of the cells was examined. The diameter of $CD8^+$ T cells and pro-B cells derived from RAG-deficient mice is comparable, 7.47 $\pm$ 1.00 and 7.74 $\pm$ 0.83 μm respectively. In contrast, the diameter of wild-type pro-B cells was significantly reduced to 5.72

$\pm$ 0.62 when compared to pro-B cells derived from RAG-deficient mice or $CD8^+$ T cells (Table 2.2). To determine whether contraction of the Igh locus in pro-B cells derived from wild-type mice is caused by a reduction in nuclear size, we measured the distances separating two control probes, RP23-132L4 and RP23-478G17, located at 12A1.3 on the arm of chromosome 12. We note that the region delineated by the two control probes comprises at least one house-keeping gene (Figure 2.4). Interestingly, the distances separating these loci, on average 1.76 Mb per μm, were equivalent in all three cell types (Table 2.2). Taken together, these data indicate that locus contraction in developing B cells is not a general property of pro-B cell nuclei.

If the spatial organization of the Igh locus were to be linear, then it would be expected that the sum of the distances separating the V_H1 and $V_H9/4$ probes and $V_H9/4$ and Cα probes to be equal to the distances separating the V_H1 and Cα regions. However, in all cell types analyzed, the distance separating the V_H1-Cα probes was substantially less than predicted from a linear configuration (Figure 2.1D, Table 2.1). Taken together, these observations indicate that the Igh chain gene locus displays long-range 3D-complexity in both T and B lymphocytes.

2.2 Condensation of the Igh Locus in Lymphocytes

Two possible configurations that can account for long-range 3D complexity, locus compaction and looping, were distinguished from the low-complexity configuration and examined in detail (Figure 2.6A). Both the observations presented here and in previous studies showed large-scale contraction of the Igh locus in developing B cells[11]. To examine whether the reduction in distances reflects chromatin condensation, we used the following numerical criteria. Locus compaction was defined when the distance separating the V_H1 from the $V_H9/4$ regions was less than 0.2 μm and the distance separating V_H1 and Cα was less than 0.25 μm, representing at least a 2.5-fold contraction. Using these criteria, our analysis revealed that 5.7% of the $CD8^+$ T cells contained alleles in a compacted conformation. In contrast, pro-B cells derived from RAG-deficient or wild-type mice contained significantly more alleles in a compacted state, the frequency being 18.5% and 14.9% respectively (Figure 2.7D-F). We note that we cannot rule out the possibility that wild-type pro-B cells contained alleles in which recombination occurred downstream of the $V_H9/4$ region.

2.3 Looping at the Igh Locus in Developing B Cells

Upon visualizing the Igh locus in nuclei derived from pro-B cells and $CD8^+$ T, we observed a substantial fraction of alleles, 24.6% and 19.6% respectively, showing the V_H1 (labeled in green) region in closer proximity to the Cα (labeled in red) region relative to the $V_H9/4$ (labeled in blue) region. This observation suggested the presence of looped Igh domains in a large fraction of both B and T-lineage cells (Figure 2.5). Interestingly, closer examination revealed that, in contrast to $CD8^+$ T cells, a substantial proportion of looped Igh alleles derived from wild-type pro-B cells showed the V_H1 and Cα regions (labeled with green and red respectively) localized in close proximity, with the two signals often overlapping, whereas the $V_H9/4$ (labeled in blue) region was distinctly separated (Figure 2.6A,E-J). In order to quantitatively determine the degree of looped structures in which the V_H1 regions was localized in close proximity to the Cα region and the Igh DJ cluster, we used the following numerical criteria: (1) The distance separating V_H1 and Cα being less than 0.2 μm, corresponding to approximately 250 kb. The latter is determined from the distance separating the $V_H9/4$ from the Cα region, estimated at 1.28 Mb per μm, measured in $CD8^+$ T cells which represent a cell type in which the Igh locus is present in relatively low-complexity de-compacted 3D-configuration. We note that the distance separating the Igh D cluster from the Igh

Cα exons is approximately 200 kb. Thus, 0.2μm is consistent with localization of the V_H1 region in proximity to the Igh D cluster. (2) The distance separating V_H9/4 and Cα being greater than the distance separating V_H1 and Cα by at least 0.15μm. Using these criteria, looping was observed in approximately 14% of interphase nuclei of pro-B cells (Figure 2.1E, 2.6E-J). We note that such loops occurred mono-allelically. Using a BAC probe located at the telomeric border of the V_H1 region, we observed that this region also underwent looping in developing B cells indicating that different V_H1 regions have the ability to form loops with DNA segments located in the proximity of the Igh Cα domain (data not shown). Employing the numerical criteria described above, such loops were also detected in interphase nuclei of CD8 T-lineage cells (Figure 2.1E). However, the frequency of loops in T cells was significantly lower, 2.4% versus 14.0% respectively, when compared to the one observed in pro-B cells ($P<0.0001$) (Figure 2.1E). Similarly, loops in E2A deficient hematopoietic progenitor cells were detected at low frequency (1.5%, C.S. unpublished observations). Taken together, these observations demonstrate that loops involving Igh V regions and DNA segments localized in close proximity to the Igh DJ cluster can be detected with relatively high frequency in developing B cells.

2.4 Igh Locus Undergoes Looping in RAG Deficient B-Lineage Cells

In vitro studies have demonstrated that the RAG proteins have the ability to recruit recombination substrates into a synaptic complex. In this complex, RAG-mediated DNA double-stranded breaks are introduced and the broken-ends subsequently fused by general DNA repair mechanisms [40, 41]. Synapse formation has not been directly observed in B-lineage cells. Nevertheless, if such complexes occurred, then recruitment of V segments belonging to the V_H1 family into synaptic complexes could give rise to loop structures as detected by our assay. To determine whether looping of the V_H1 region occurs as a consequence of RAG-mediated synapse formation during Igh V-DJ rearrangement, we analyzed the 3D-organization of the Igh locus in interphase nuclei derived from RAG2-deficient pro-B cells. Interestingly, the fraction of pro-B nuclei derived from RAG2-deficient mice exhibiting looped alleles (6.7%), while significantly lower than that found in wild-type pro-B cells (p=0.0197) (14.0%), was significantly greater than the fraction observed in $CD8^+$ T cells (p=0.0224) (2.4%) (Figure 2.1E, Figure 2.7G-I). It is conceivable that the difference in the proportion of looped alleles observed in wild-type pro-B cells and $RAG2^{-/-}$ pro-B cells is caused by the absence of D-J joints which normally precede Igh rearrangement and/or by the deficiency in RAG activity. Nevertheless, these observations indicate that looping in the Igh locus

occurs in B cells, poised to undergo rearrangement, even in the absence of RAG activity.

2.5 Looping of Ig Variable Regions in Developing B Cells

Recent studies have begun to elucidate the mechanisms by which long-range interactions of cis-regulatory elements promote the regulation of gene expression. A model has been proposed for the developmental regulation of expression of β-globin genes where the repressive chromatin state is maintained by attaching regulatory regions to the nuclear matrix, thus preventing long-range promoter-enhancer interactions[42]. Activation of gene expression would require the expression of specific transcription factors, the acetylation of histones at hypersensitive sites, and finally the clustering of distally located regulatory regions through a looping mechanism. The extended structure of the Igh V-region, encompassing $\sim$2 Mb of DNA, is a potential barrier to the efficient synapsis and rearrangement of distally located elements. In support of this, Alt and colleagues have proposed a chromosomal proximity hypothesis, stating that the rearrangement mechanism preferentially favors genes that are in close proximity[43, 44]. This is best demonstrated in pro-B cells derived from IL7R-deficient mice where a gradient of usage of V elements is observed. D-proximal V elements rearranging normally whereas usage of D-distal

V-segments is impaired[45]. Such a mechanism has also been proposed to regulate rearrangement of the T-cell receptor beta gene[46]. We propose that, similarly to the regulation of β-globin gene expression, looping presents a mechanism by which elements that are physically separated by large distances are brought into proximity during Ig gene rearrangement. Our observations indicate that in both B- and T-lineage cells, loops within the Igh locus can be detected with relative high frequency. However, the frequency of loops involving V_H regions that are located in close proximity to the $D_H J_H$ cluster is significantly higher in B cells as compared to T-lineage cells. Based on these observations we propose that the Igh locus is organized in both B and T cells in looped-like structures. Prior to the onset of Igh gene rearrangements, V_H-region segments are moved into close proximity of the $D_H J_H$ cluster to promote Igh V(D)J gene rearrangement. Our observations also demonstrate that looping into close proximity of the Igh DJ cluster is mono-allelic. It is conceivable that mono-allelic looping contributes to allelic exclusion[47]. However, we note that our observations do not exclude the possibility that V_H regions, distinct from the $V_H 1$, undergo looping on the second allele. How are such loops organized and formed? It is plausible that looping of distinct V_H regions may require the interactions of specific matrix attachment regions (MARs) binding proteins with MARs that are interspersed in the Igh locus[48]. Alternatively, reorganization of a matrix to which the Igh locus is tethered may regulate looping. It should now be possible using the strategy described here to examine how Ig

loops are formed and how they are regulated.

Table 2.1: Distances Separating the Igh Subregions

Distances separating the Igh subregions, V_H1, $V_H9/4$, and Cα , in pro-B cells, RAG-deficient pro-B cells, CD8$^+$ T cells and in E2A-deficient hematopoietic progenitor cells. Shown are averages ± S.D expressed in μ m. n refers to the number of cells analyzed for each cell type. + indicates that some cells may have undergone Ig VDJ recombination. *** p<0.001 when compared to CD8$^+$ T cells.

	Pro-B	**RAG2$^{-/-}$**	**CD8$^+$T**	**E2A$^{-/-}$**
n	114+	372	245	200
$V_H1 - V_H9/4$	0.29 ± 0.14***	0.27 ± 0.11***	0.38 ± 0.20	0.33 ± 0.15
$V_H9/4 - C\alpha$	0.33 ± 0.19***	0.29 ± 0.12***	0.42 ± 0.23	0.42 ± 0.23
$V_H1 - C\alpha$	0.38 ± 0.21***	0.37 ± 0.14***	0.56 ± 0.28	0.51 ± 0.24
$V_H1 - V_H9/4$+ $V_H9/4 - C\alpha$	0.62 ± 0.25	0.56 ± 0.17	0.81 ± 0.31	0.74 ± 0.28

Table 2.2: Distances Separating the Control Loci

Distances separating the control loci, indicated as RP23-478G17 and RP23-132L4, in pro-B cells, RAG-deficient pro-B cells, CD8$^+$ T cells. Shown are averages ± S.D expressed in μm. n refers to the number of cells analyzed for each cell type. No statistical significance (ns) was found in the distances separating the probes in the three cell types examined. *** p<0.001 when compared to CD8$^+$ T cells or RAG-deficient pro-B cells. Differences in the diameter of CD8$^+$ T cells when compared to RAG-deficient pro-B cells were not statistically significant.

	Pro-B	**RAG2$^{-/-}$**	**CD8$^+$T**
n	32	28	33
132L4-2A3	0.42 ± 0.21^{ns}	0.39 ± 0.20^{ns}	0.42 ± 0.23^{ns}
diameter	5.72 ± 0.62***	7.74 ± 0.83^{ns}	7.47 ± 1.00^{ns}

Acknowledgements

The text of chapter 2 is partly reprinted from the paper *"Visualization of Looping Involving the Immunoglobulin Heavy-chain Locus in Developing B Cells"*,

Genes and Development, 2005, 19(3). I would like to thank Camil Sayegh to make me an integral part of the project when I joined the lab. Roy Riblet helped us with his expertise in the genetic organization of the genetic locus and provided us with some hybridization probes.

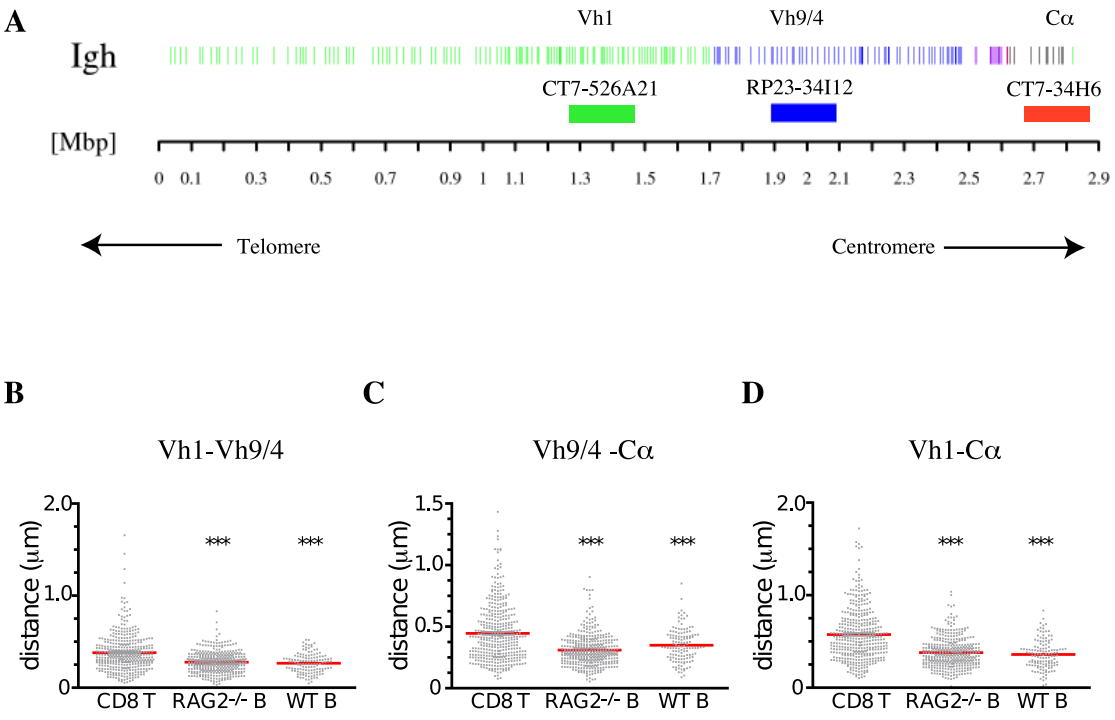


Figure 2.1: Compaction of distinct subregions of the Igh locus

(A) The murine Igh locus and positions of the three BAC probes are indicated (not drawn to scale). The distances separating each of the three BAC probes, CT7-526A21, CT7-34H6, and RP23-24I12 and their positions within the Igh locus were determined using the Ensembl mouse genome database. Note that the exact location of the CT7-526A21 BAC remains to be determined. The colors of the three probes are shown: CT7-526A21 (green), RP23-24I12 (blue) and CT7-34H6 (red). (B-D) Scatter-plots of the distances separating the V and Cα regions in T cells, in vitro cultured wild-type and RAG2-deficient pro-B cells. Y-axis indicates distances separating the loci in μm. Red line indicates the average distance in each group. (***) indicates significant difference ($p<.0001$) of the averages compared to that obtained from T-lineage cells.

Figure 2.2: Relative distributions of the distances separating distinct subregions of the Igh locus
(A) V_H1-$V_H9/4$, (B) $V_H9/4$-Cα , and (C) V_H1-Cα . CD8$^+$ T are indicated by white bars, pro-B cells derived from wild-type mice by black bars or RAG2$^{-/-}$ mice indicated by grey bars.

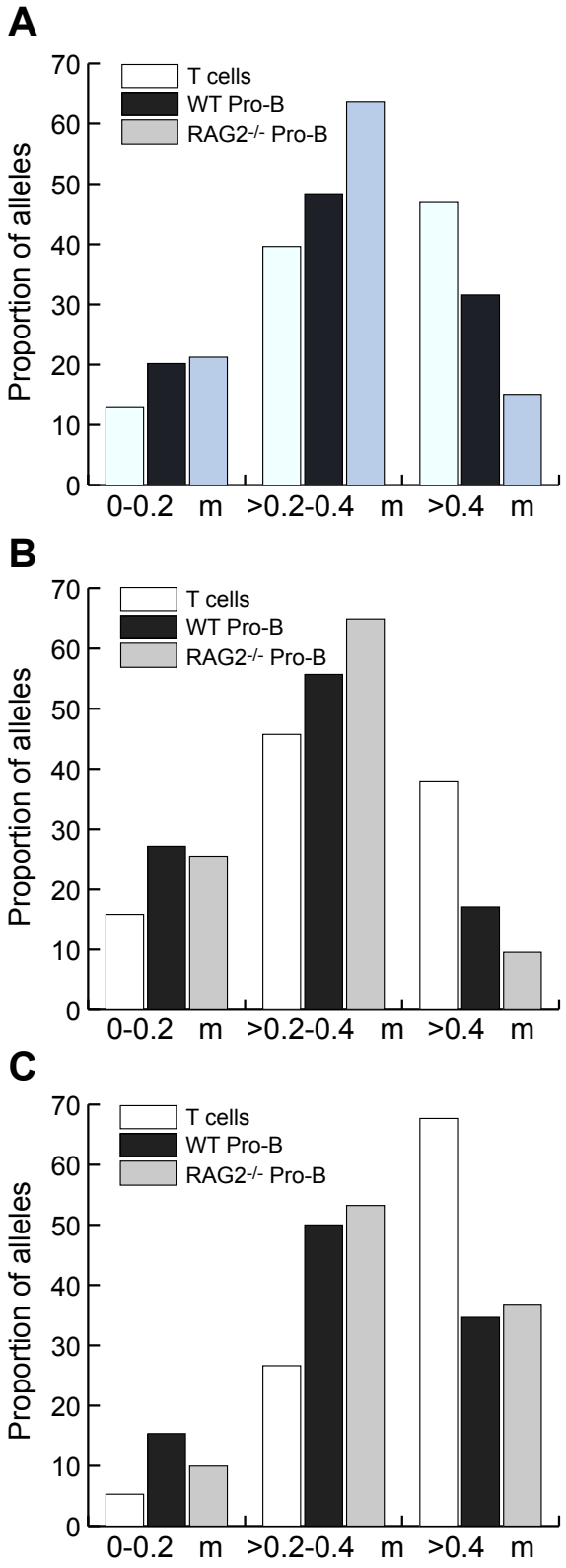
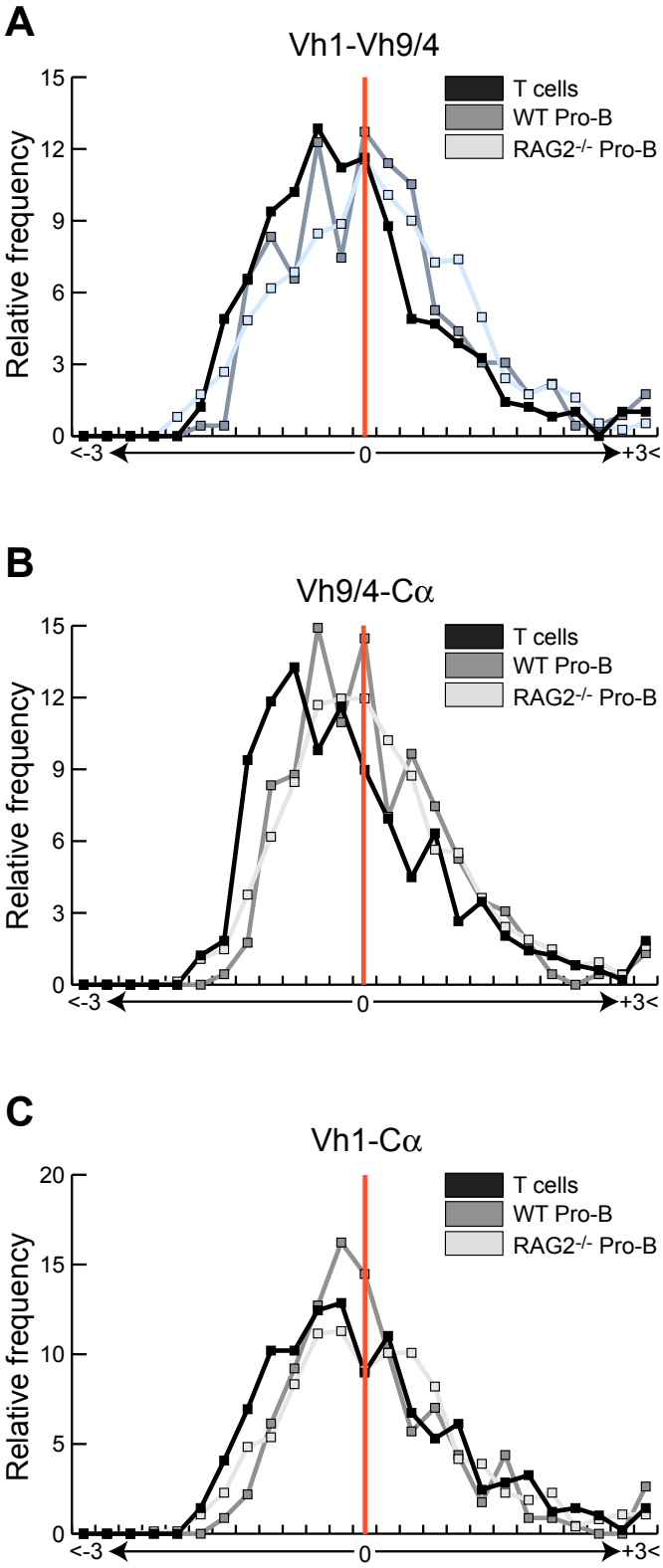
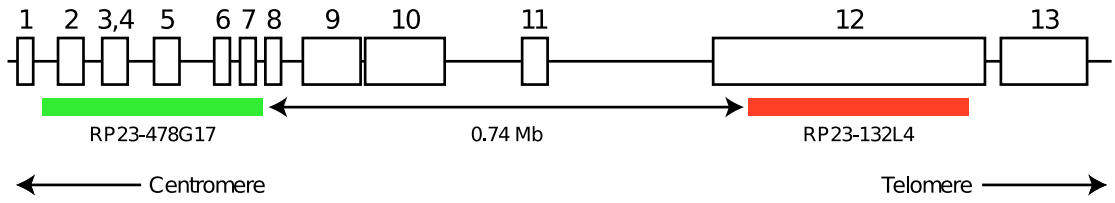

A
B
C
T cells
WT Pro-B
RAG2-/- Pro-B
Proportion of alleles
0
10
20
30
40
50
60
70
0-0.2
m
>0.2-0.4
m
>0.4
m

Figure 2.3: Statistical variation of the distances separating distinct subregions of the Igh locus
(A) V_H1-V_H9/4, (B) V_H9/4-Cα , and (C) V_H1-Cα in CD8$^+$ T (black). Pro-B cells derived from wild-type mice are indicated by dark grey lines. B cells derived from RAG2$^{-/-}$ mice are shown by light grey lines. T cells are indicated by black lines. Curves show the z-score distribution for each cell type. The red line indicates the position at which the z-score is equal to zero. At the x-axis, each interval represents a 0.25 change in the z-core. Last interval to the left shows z-score $<$ -3.0 and to the right z-score $>$ +3.0.


A
Vh1-Vh9/4
Relative frequency
15
12
9
6
3
0
T cells
WT Pro-B
RAG2-/- Pro-B
<-3
0
+3<
B
Vh9/4-Cα
Relative frequency
15
12
9
6
3
0
T cells
WT Pro-B
RAG2-/- Pro-B
<-3
0
+3<
C
Vh1-Cα
Relative frequency
20
15
10
5
0
T cells
WT Pro-B
RAG2-/- Pro-B
<-3
0
+3<

A



B

1 RIKEN cDNA 1700022F17 gene
2 RIKEN cDNA 1110015M06 gene
3 similar to Doublecortin-containing protein 2(RU2S protein)
4 Allantoicase
5 collectin subfamily member 11
6 ribosomal protein s7
7 ribonuclease H1
8 expressed sequence AL024210
9 tetratricopeptide repeat domain 15
10 DNA segmwnr, Chr 12, ERATO Doi 604, expressed
11 similar to SPBPJ 4664.02
12 myelin transcription factor 1-like
13 RIKEN cDNA 2310075M15 gene

Figure 2.4: Map of the control locus

(A) Graphical representation of the two control BAC probes, RP23-132L4 (red) and RP23-478G17 (green). The positions of the two BAC probes and predicted open-reading frames or genes that are located in the region were determined using the Genes_Sequence Database (NCBI). (B) List of genes in the region delineated by the two control BAC probes based on the mouse genome sequence map (build 33.1) at NCBI yields the indicated known and putative genes in the region. Of these, at least Ribonuclease H1 (Rnase H1) has been suggested to be a house-keeping protein.

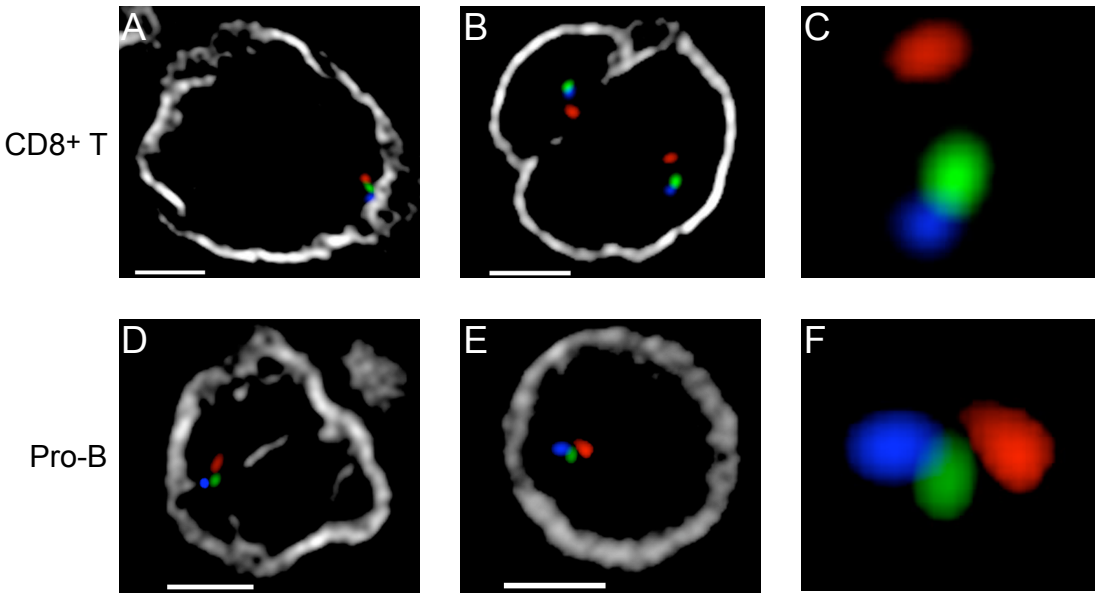


Figure 2.5: Looped chromatin domains involving the distal Igh V_H1 and the Ig Cα regions

The three BAC probes used in this study are as described in Figure 1A. (A-B, D-E) 3D-configuration of the Igh locus resolved using 4-color FISH. (C and F) Digitally magnified pictures of Igh alleles are shown. (A-C) Igh locus in CD8$^+$ T cells and (D-F) wild-type pro-B cells. The nuclear membrane stained using lamin antibodies is shown in grey. Bar=2μm.

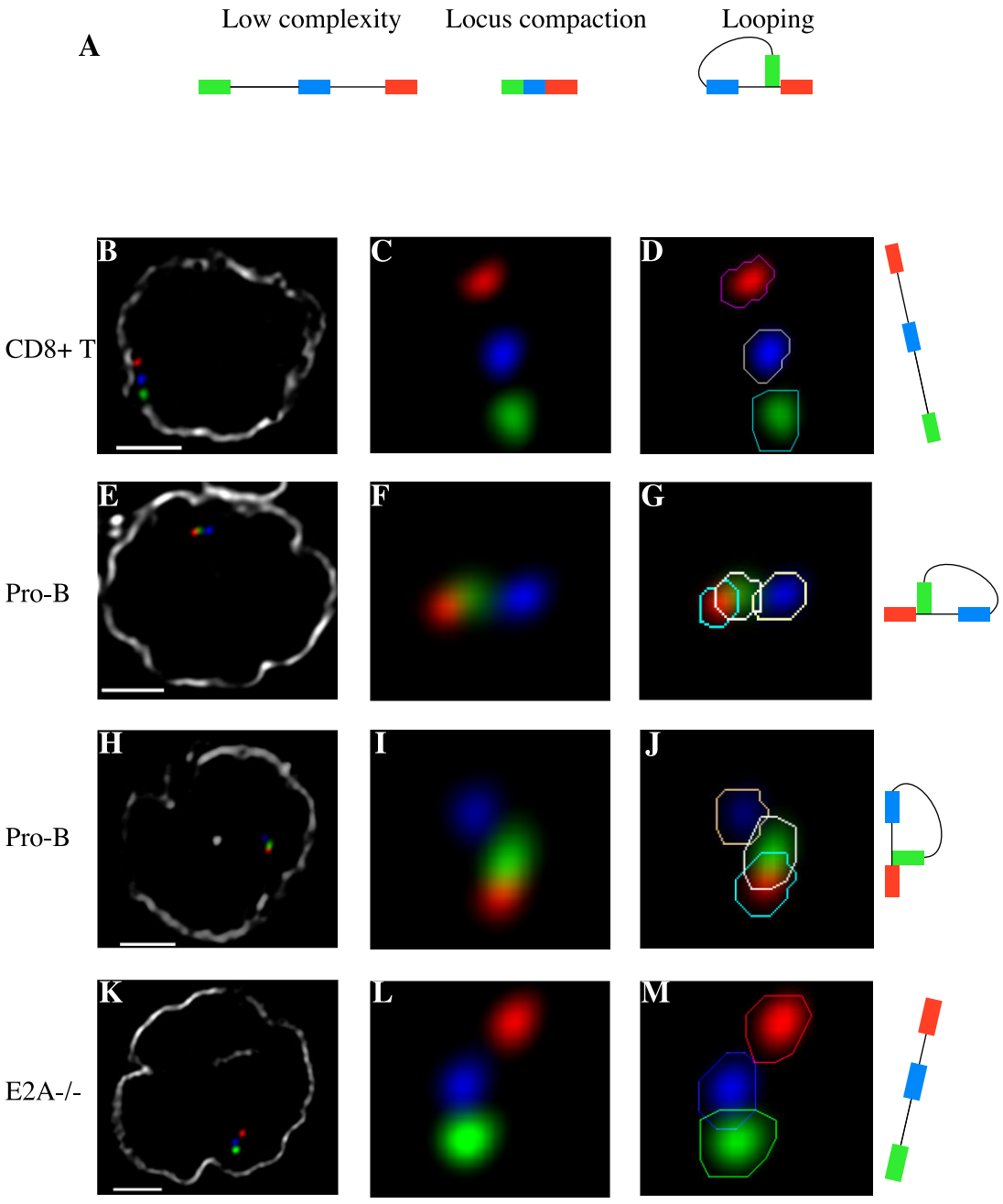


Figure 2.6: Looping involving the distal V_H1 and the C_α regions.

(A) Schematic diagram of three possible distinct Igh configurations. (B-M) 3D-configuration of the Igh locus resolved using 4-color FISH. Digitally magnified pictures of Igh alleles are shown. Polygons (in one selected z-section) that were used to identify the coordinates of the center of mass for each signal are indicated (D, G, J and M). The nuclear membrane stained using lamin antibodies is shown in grey (B, E, H and K). Bar=2μm. (B-D) Igh locus in CD8$^+$ T cells. (E-J) Looping involving the V_H1 and the C_α region in wild-type pro-B cells. (K-M) Configuration of the Igh locus in E2A-deficient hematopoietic progenitor cells.

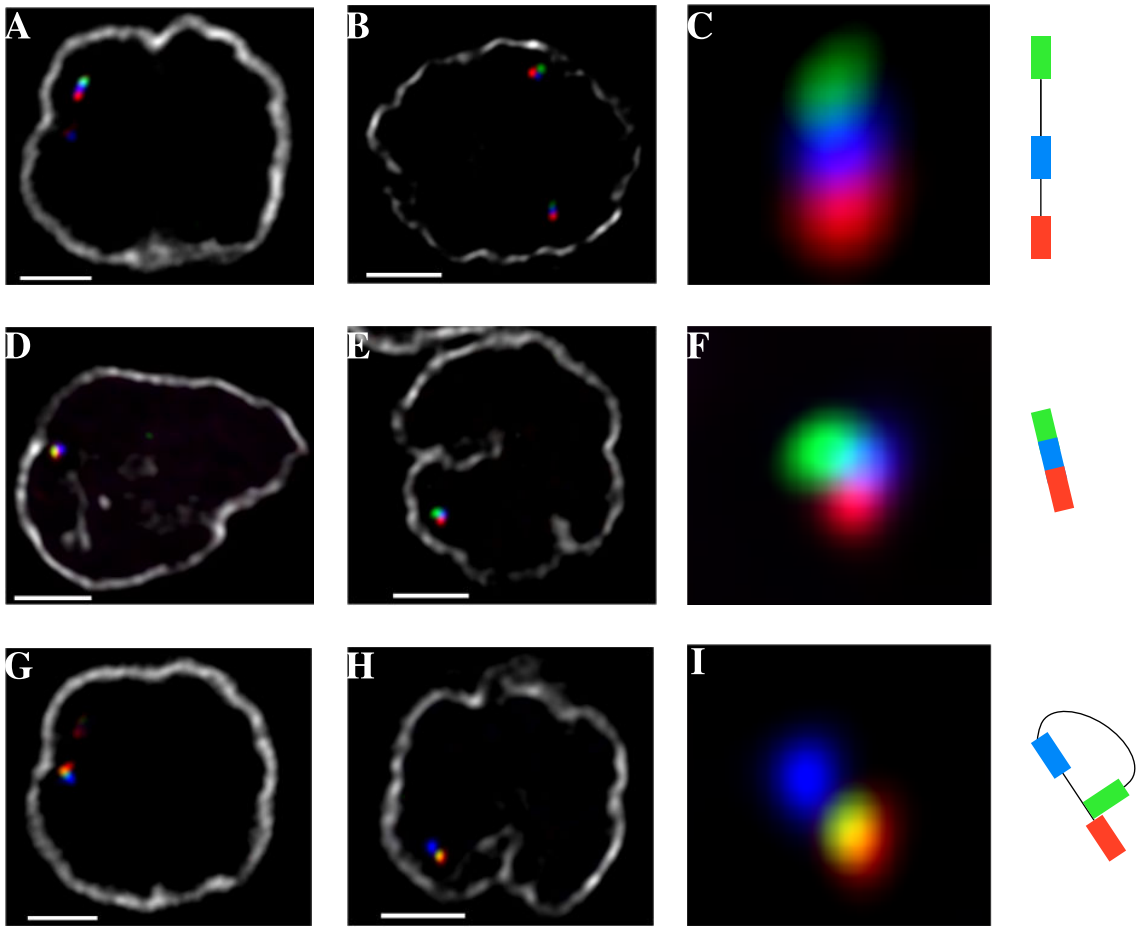


Figure 2.7: Looping involving the Igh locus in RAG deficient pro-B cells

Three distinct configurations of the Igh locus were detected 3D-FISH in nuclei derived RAG deficient pro-B cells. (A-C) low-complexity, (D-F) locus compaction, (G-I) looping. The nuclear membrane is shown in grey. Bar=2μm.

Chapter 3

The 3D-Structure of the Immunoglobulin Heavy Chain Locus

How genes are regulated by spatial rearrangement has been a topic of intensive study. In prokaryotes, transcriptional enhancers act through looping or tracking along the intervening DNA[49, 50]. In eukaryotic cells, chromatin compaction and looping, the presence of DNase I hypersensitive sites at regulatory elements, including transcriptional enhancers, insulators and locus control regions, influence gene expression over large genomic distances[51]. The globin locus control region was shown to act over a large distance by looping the intervening region and physically associating with actively transcribing β-globin genes[52, 53]. Other loci also have been shown to bring distant enhancer elements into proximity of promoter regions by looping, including the Th2 cytokine locus and the interferon gamma gene[54, 55]. As shown above, looping is also observed at the Igh locus.

It is well established that higher order chromatin organization plays a pivotal role in genome function[56]. As a first approach to resolving chromosome conformation, fluorescence in situ hybridization studies, measuring spatial distances in interphase nuclei between genomic markers as a function of genomic separation, suggested a random walk behavior[57]. However, confinement of chromosome arms and bands to territories indicated the presence of spatial constraints. More recent observations showed that the spatial distance depends on the genomic distance according to a power law with exponents of 0.5 below and 0.32 above a genomic separation of 4 Mbp[30, 31, 32, 33]. The constraints and the scaling behavior suggested a Random-Walk/Giant-Loop (RW/GL) configuration[30]. In the RW/GL model, the 30 nm fiber forms 2 to 5 Mbp loops that are attached to a polymer backbone. The backbone and the chromatin fiber within the loops follow random walk dynamics. However, distance measurements between genetic markers with genomic separations of less then 4 Mbp were incompatible with the RW/GL model, but were consistent with another topology, named the Multi-Loop-Subcompartment (MLS) model[34, 33]. The MLS model proposes that the 30 nm fiber is folded into rosettes of small loops, connected by linkers of variable sizes.

Recently computer models have been developed to evaluate and test experimental results, designs and hypotheses about the three-dimensional genome organization[58, 34]. Beyond supporting the chromatin organization into chromosome territory, arm and band domains, these simulations may reveal how the local,

global and dynamic characteristics of cell nuclei are inter-connected[58, 34]. Here, we have used spectral high precision epifluorescence microscopy to determine spatial distance distributions between 12 genomic markers that span the entire Igh locus. These data were analyzed to provide a statistical description of the Igh locus architecture and compared with various computer models to determine what most appropriately describes the locus.

3.1 Methodology for High-Resolution Spatial Distance Measurements between Genomic Markers in B-lineage Cells

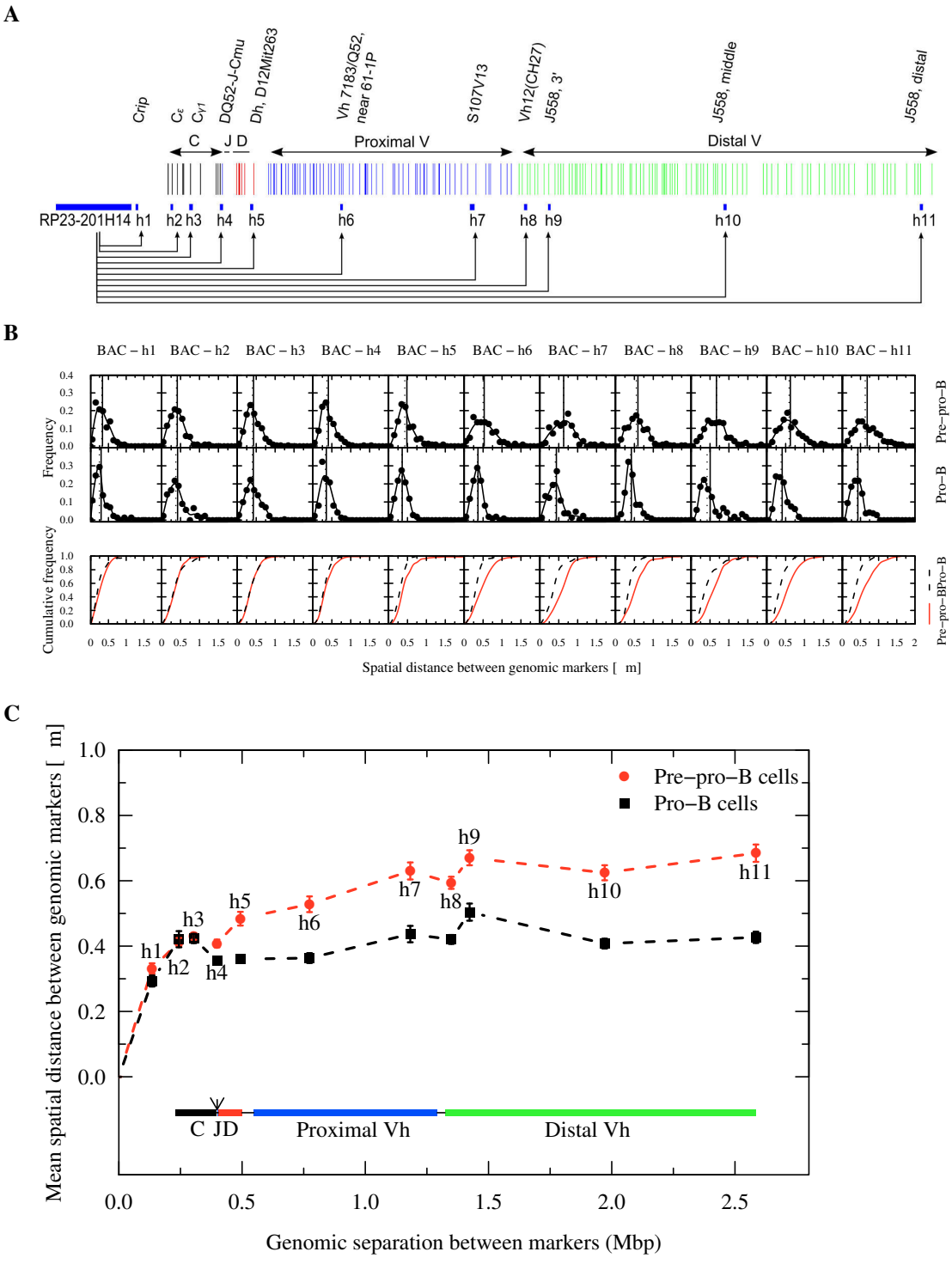
To determine the 3D-architecture of the Igh locus the spatial distances separating multiple markers located throughout the locus were measured. Two cell types were used in these studies: (1) E2A-deficient pre-pro-B cells were examined, since these cells are arrested at the pre-pro-B cell stage and have not yet committed to the B-cell lineage[59]. (2) RAG2-deficient pro-B cells were chosen since they are committed to the B cell lineage but, unlike wild-type pro-B cells, carry the Igh locus in germ-line configuration[60]. Pre-pro-B cells and pro-B cells were fixed, permeabilized and hybridized with two 10 Kbp probes and a BAC probe, located 3' of the Igh locus (Figure 3.1A and 3.5). Although the intensity varied, the

fluorescent signals emitted by the 10 Kbp probes were clearly detectable (Figure 3.1A and 3.5). The effective resolution ('Resolution equivalent', see methods) we obtained using two different colors ranged from 35 to 47 nm for the different combinations of the fluorochromes (Table 3.1 and 3.2). Thus, this approach allowed us to measure spatial distances with a 3D-resolution better than 50 nm.

3.2 3D-Architecture of the Immunoglobulin Heavy Chain Locus

To dissect the topology of the Igh locus, spatial distances were measured between an anchor, BAC probe RP23-201H14, which served as a marker located down-stream of the Igh locus and eleven 10 Kbp probes that span the entire locus (Figure 3.1A). The average spatial distances separating the anchor RP23-201H14 and the markers located within the C_H region cluster (h2 and h3) increased as a function of genomic separation (Figure 3.1B and C, Table 3.1). In pre-pro-B cells, the DNA spanning the Igh locus was 100-1000 fold more compacted as compared to linear DNA, which spans about 3.4 nm per ten bp (Table 3.1). Pro-B cells showed higher compaction values, which were particularly striking in the distal V_H cluster (1284 fold versus 2059 fold between RP23-201H14 and h11) (Table 3.1). In both pre-pro-B and pro-B cells the spatial distances flattened upon increasing the genomic distance (Figure 3.1C). To examine the Igh topology

Figure 3.1: Immunoglobulin Heavy Chain Locus Spatial Distances and Spatial Distributions as a Function of Genomic Separation in Pre-Pro-B and Pro-B cells. (A) Genomic organization of the Igh locus. The anchor and the genomic markers used are indicated. (B) Frequency plots showing the distribution of spatial distances between the probes and the anchor (RP23-201H14). Cumulative frequency distributions are indicated for both pre-pro-B and pro-B cells. (C) Average spatial distances were plotted as a function of genomic separation. Distal and proximal variable regions as well as diversity, joining and constant region segments are shown. Bars indicate standard error. The dotted lines only indicate connectivity. The arrow indicates the position of the intronic enhancer.


A
Crip
Cε
Cγ1
DQ52-J-Cmu
Dh, D12Mit263
Vh 7183/Q52, near 61-1P
S107V13
Vh12(CH27)
J558, 3'
J558, middle
J558, distal
C
J D
Proximal V
Distal V
RP23-201H14 h1
h2 h3
h4
h5
h6
h7
h8 h9
h10
h11
B
BAC – h1
BAC – h2
BAC – h3
BAC – h4
BAC – h5
BAC – h6
BAC – h7
BAC – h8
BAC – h9
BAC – h10
BAC – h11
Frequency
Cumulative frequency
Pre–pro–B
Pro–B
Pre–pro–BPro–B
Spatial distance between genomic markers [m]
C
Mean spatial distance between genomic markers [m]
Pre–pro–B cells
Pro–B cells
h1
h2
h3
h4
h5
h6
h7
h8
h9
h10
h11
C JD
Proximal Vh
Distal Vh
Genomic separation between markers (Mbp)

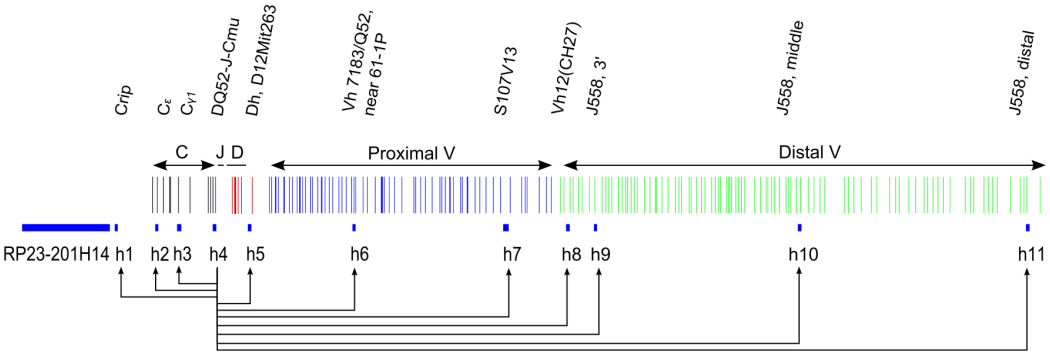
in non-lymphoid cells, the spatial distances as a function of genomic separation in C57Bl/6 embryonic fibroblasts were determined. In embryonic fibroblasts the average spatial distances separating the anchor and the markers located within the C_H region increased substantially as a function of genomic separation when compared to those observed in pre-pro-B and pro-B cells (Figure 3.6A and B). Thus within this region the chromatin fiber appears to be substantially less compacted as compared to lymphoid cells. These data indicate that the Igh locus shows substantial compaction in pre-pro-B cells as compared to non-lymphoid cells, and becomes even more condensed in pro-B cells. Strikingly, however, the spatial distances plateau beyond the D_H elements as a function of increased genomic separation in all three cell-types. To examine the architecture of a genomic region distinct from that of the Igh locus, spatial distances were measured as a function of genomic separation using the RP23-201H14 BAC probe as an anchor and using genomic markers located towards the centromere. The spatial distances again flattened with the topology in pro-B cells being more condensed compared to pre-pro-B cells (Figure 3.6C). However, for genomic distances larger than 4 Mbp the spatial distances in pre-pro-B and pro-B cells merged (Figure 3.6C). In summary, these data indicate: (1) The Igh locus assumes different 3D-architectures in fibroblasts, pre-pro-B and pro-B cells. (2) The spatial distances plateau as the genomic separation increases.

3.3 Long-Range Genomic Separation but Similar Spatial Distances

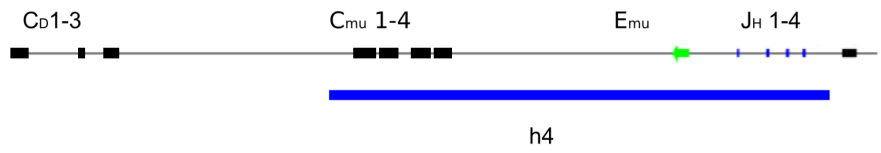
To determine the spatial distances separating the $D_H J_H$ elements from the proximal and distal V_H regions, probe h4, containing the $D_H J_H$ elements, was used as a second anchor (Figure 3.2A-D). The average spatial distances between the $D_H J_H$ elements and the proximal V_H regions (h5 to h8) were similar for pre-pro-B and pro-B cells (Figure 3.2C and 2D, Table 3.1). In contrast, the spatial distances separating the distal V_H (h9 to h11) and $D_H J_H$ regions were substantially reduced in pro-B cells as compared to pre-pro-B cells (Figure 3.2D). Remarkably, although separated by large genomic distances, the spatial distances separating the proximal and distal V_H regions from the $D_H J_H$ cluster were similar in pro-B cells (Figure 3.2D). In fact, the most distal V_H regions were positioned at slightly smaller spatial distances from the $D_H J_H$ elements than the majority of the proximal V_H regions (Figure 3.2D). There are also differences in the spatial distance distributions observed for pre-pro-B and pro-B cells (Table 3.1). The standard deviations were substantially higher in RAG-deficient pro-B cells for the spatial distances separating the majority of the V_H elements from the $D_H J_H$ elements (h4-h6, h4-h7, h4-h8, h4-h9) even though the spatial distances were smaller (Table 3.1). Collectively, these data show: (1) During the transition from the pre-pro-B to the pro-B cell stage, the Igh locus is remodeled to position the entire set of

Figure 3.2: Immunoglobulin Heavy Chain Locus Topology in Pro-B Cells Brings Distal V_H Regions and $D_H J_H$ and Enhancer Elements in Close Spatial Proximity. (A) Genomic organization of the Igh locus. The anchor and the genomic markers used are indicated. (B) Probe h4 contains the J_H segments, the intronic enhancer and Cμ elements. (C) Frequency plots showing the distribution of spatial distances for each genomic marker from the Igh $D_H J_H$ cluster. Cumulative frequency distributions are indicated for both pre-pro-B and pro-B cells. (D) Average spatial distances were plotted as a function of genomic separation for each of the probes. Distal and proximal variable regions as well as diversity, joining and constant region segments are shown. Bars indicate standard error. The dotted lines only indicate connectivity. The arrow indicates the position of the intronic enhancer.

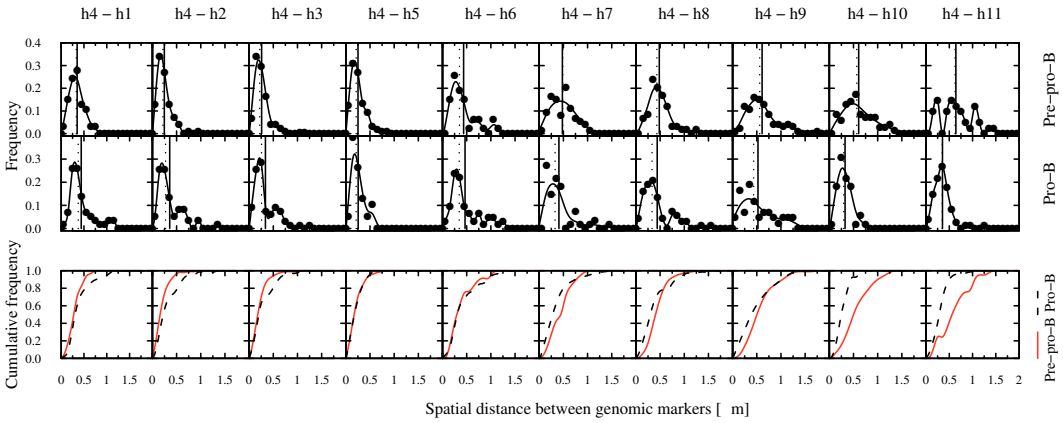
A


Crip
Cε
Cγ1
DQ52-J-Cmu
Dh, D12Mit263
Vh 7183/Q52, near 61-1P
S107V13
Vh12(CH27)
J558, 3'
J558, middle
J558, distal
C
J D
Proximal V
Distal V
RP23-201H14 h1
h2 h3
h4
h5
h6
h7
h8 h9
h10
h11

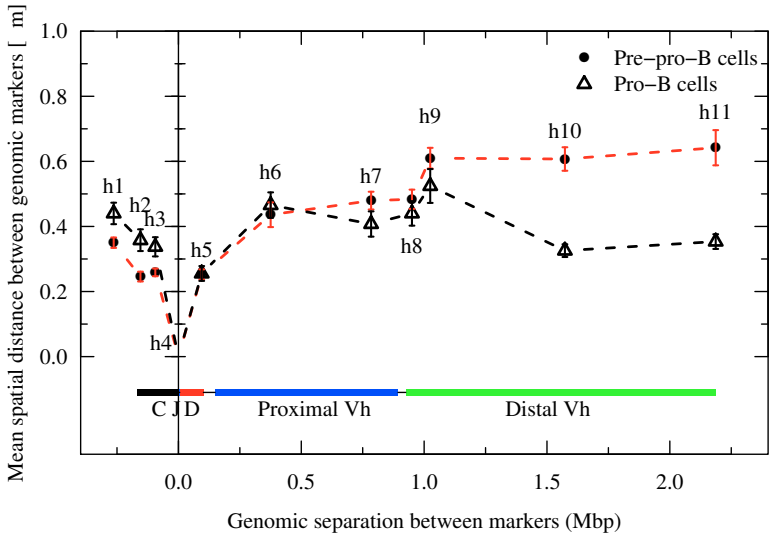
B


CD1-3
Cmu 1-4
Emu
JH 1-4
h4

C


h4 – h1
h4 – h2
h4 – h3
h4 – h5
h4 – h6
h4 – h7
h4 – h8
h4 – h9
h4 – h10
h4 – h11
Frequency
Cumulative frequency
Pre–pro–B
Pro–B
Spatial distance between genomic markers [m]

D


Mean spatial distance between genomic markers [m]
Pre–pro–B cells
Pro–B cells
h1
h2
h3
h4
h5
h6
h7
h8
h9
h10
h11
C J D
Proximal Vh
Distal Vh
Genomic separation between markers (Mbp)

V_H regions at similar distances to the $D_H J_H$ elements. (2) The chromatin fiber that contains the proximal V_H, D_H and J_H elements assumes a wider spectrum of configurations in pro-B versus pre-pro-B cells.

3.4 Long-Range Genomic Interactions and the Probability of V_H to $D_H J_H$ Joining

The physical process of DNA recombination requires that the V_H, D_H and J_H segments be mobile, allowing them to interact with each other with detectable frequencies. To assess the probabilities of such encounters, the cumulative frequencies of the spatial distances that separate the V_H regions from the $D_H J_H$ elements were determined (Figure 3.2C). The cumulative frequency of a marker indicates the fraction of the alleles in which the marker is within a certain distance from the anchor. To compare the probabilities of different V_H regions to associate with $D_H J_H$ elements within the same cell type, the cumulative frequencies were plotted for pre-pro-B and pro-B cells (Figure 3.2C and 3.7). As expected, in pre-pro-B cells the probabilities of V_H regions to encounter $D_H J_H$ elements correlated well with increasing genomic separation (Figure 3.2C and 3.7). However, in pro-B cells the cumulative frequencies were clustered at short spatial distances for the majority of the V_H regions (Figure 3.7). Thus in pro-B cells the probabilities for V_H regions to be localized within close proximity of the $D_H J_H$ elements are similar regardless

of large differences in genomic separation.

3.5 The Immunoglobulin Heavy Chain Topology Cannot be Described as a Self Avoiding Random Walk or Worm-Like Chain

As a first approach towards elucidating the spectrum of Igh topologies, we compared the experimental data with various models of chromatin structure. We found using linear regression that the spatial distance (R) scaled with genomic distance (N) as a power law with exponent of 0.25 for pre-pro-B cells and 0.1 for pro-B cells (data not shown). These values differ significantly from that expected of a free random walk (0.5) and a self-avoiding chain ($\sim$0.6)[28]. We further examined the scaling property of the probability distribution functions (P(R,N)), by plotting $P.N^{\nu.d}$ as a function of R/N^{v} on log-log plots (Figure 3.3A-D). If the Igh locus configurations can be described as a self-avoiding random walk, then all the data points (distance from either anchor (BAC and h4) to any of the genomic markers) are expected to collapse to a single curve when the exponent ν is chosen to be $\sim$0.6[28]. This is clearly not the case for either anchor (BAC and h4) (Figure 3.3A-D). Interestingly for values of ν below 0.3 (0.1$\sim$0.2 for pre-pro-B cells and $\sim$0.1 for pro-B cells), the data points do collapse (Figure 3.8). However, the interpretation

for $\nu < 1/3$ is unclear. To determine whether the Igh topology fits a worm-like chain behavior, the experimental spatial distances as a function of genomic separation were directly compared to the Porod-Kratky chain[61] (see Methods for details). Different values for the persistence length and chromatin density were chosen that previously were shown to describe the physical properties of the yeast chromatin fiber[29]. However, the spectrum of Igh conformations in pre-pro-B cells did not compare well with the worm-like chain, for the different persistence length and chromatin compaction values that were examined (Figure 3.3E and F). The best-fit curve required the parameters to surpass physiological values, and was still not a good fit to the data(Figure 3.3E and F, dotted line). Taken together, these data demonstrate that the Igh locus topology cannot be described in terms of a self-avoiding random walk or worm-like chain.

3.6 Analysis of the Igh 3D-Architecture by Comparison to Structural Computer Models

Simple polymer models do not explain the distribution of spatial distances observed at the Igh locus, as shown above. To explore the Igh 3D-architecture in more detail, computer simulations of potential configurations of chromatin structure were performed using Monte-Carlo and Brownian Dynamics methods (Figure 3.4A)[34]. Two chromatin topologies, the RW/GL and MLS models, were

used. In the RW/GL model, the chromatin fiber is organized into large loops (0.5-5.0 Mbp) that are attached to a fixed backbone (Figure 3.4D). The MLS model implies that the chromatin fiber is organized into rosette-like subcompartments (1-2 Mbp) with smaller loops (60-250 Kbp) connected by linkers of variable sizes (60-250 Kbp) (Figure 3.4D). It is well established that nucleotide content substantially alters the physical properties of the chromatin fiber. However, as a first approach the chromatin fiber was modeled as an elastic homogeneous polymer since it is not known how nucleotide content affects the persistence length of eukaryotic chromatin. To consider both fixed and flexible chromatin architectures, virtual markers were placed in positions that were either dependent or independent of the simulated chromatin architectures[34] (Figure 3.4B). Position-dependent spatial distance measurements assume that loops are fixed structures. Hence, spatial distances may differ substantially due to the relative positions of genomic markers with respect to a loop base. The anchor was placed on the base of the loop and virtual distances were measured (Figure 3.4B, marker 1). The virtual spatial distances were measured from the anchor to other markers in the modeled rosette (red), in the linker (blue), and in the adjacent rosette (green), and then plotted as a function of genomic separation (Figure 3.4C). This analysis showed the characteristic oscillations of spatial distances as a function of genomic separation consistent with a fixed loop structure localized within a rosette (Figure 3.4C, dashed lines and 3.4E). Using a position-independent approach, the virtual anchors

were placed randomly at different topological positions in the simulated chromatin fiber (Figure 3.4B, anchors A-F). The average virtual spatial distances, using the RW/GL and MLS models as starting configurations and randomly placed probes, were determined and plotted as a function of genomic separation (see Methods for details). As expected, the spatial distances for the RW/GL and MLS models, flattened as the linker sizes were decreased (Figure 3.4C, solid lines). The simulated data, using the RW/GL model as a starting configuration, did not compare well with the experimental data obtained from the measurements for both pre-pro-B and pro-B cells (Figure 3.4F and 3.4G, lines b-h). However, the trend of the spatial distances as a function of genomic distances in pre-pro-B cells agreed well with that predicted by the MLS model (Figure 3.4F and 3.4G, compare blue dots and green circles to line A). Thus, in pre-pro-B cells the spatial organization of the Igh locus compares well with a topology in which the chromatin fiber is organized into multi-loop containing subcompartments (1 Mbp), with each loop of ∼126 Kbp, connected by ∼63 Kbp linkers. In contrast, in pro-B cells the Igh fiber showed a substantially more condensed topology than predicted by the MLS model for 126 Kbp loops and a linker size of 63 Kbp (Figure 3.4F and 3.4G, compare pink diamonds and red squares to line A). Collectively, the comparison of the experimental data and the simulations shows that: (1) In pre-pro-B cells, the Igh topology agrees well with the MLS model. (2) In pro-B cells the Igh locus topology is not consistent with the MLS model in which 1 Mbp compartments are separated by linkers that are

similar or larger than 63 Kbp.

3.7 Igh Locus Topology and the Multiple-Loop Subcompartment Model

Our distance distributions observed within the Igh locus in developing B cells are not well fitted by neither the self-avoiding nor the worm-like chain. A better fit was obtained by comparing our experimental values for those obtained applying Monte-Carlo and Brownian Dynamics methods using the MLS model as a starting configuration. The MLS model suggests the presence of rosettes, spanning about 1-2 Mbp of genomic sequence, connected by linkers of variable sizes[33]. In particular, the spatial distributions of the pre-pro-B Igh locus topology compared well with the MLS model (126 Kbp loops and 63 Kbp linkers). Does this comparison imply that the Igh pre-pro-B topology is structured as 126 Kbp loops? We consider this unlikely but rather suggest that loop size within the compartments varies and determined by proteins associating with the bases of loops, including Pax5, YY1, CTCF and Satb1.

The Igh topology in pro-B cells showed a substantially more condensed topology predicted by the modeling of the MLS configuration for the shortest linker (63 Kbp) that was simulated. Since shortening the linker further would essentially merge putative compartments, this latter observation raised the possibility that

Figure 3.3: Immunoglobulin Heavy Chain Topology and Comparison to the Self-Avoiding Random Walk and Worm-like Chain
(A-D) Scaled distributions of spatial distances were plotted for comparison with the self-avoiding walk. Different values for the exponent ν were tested. Graphs are shown for both cell types, and with either BAC RP23-201H14 or h4 probes as anchors. (E) Comparing Igh topology to a worm-like chain with a range of chromatin densities (100-10,000 bp/nm) but with a constant persistence length (200 nm). A 'least square optimization' analysis is also shown (dotted line). (F) Spatial distances were plotted as a function of genomic separation and compared with the Porod-Kratky chain with a range in persistence length of 50-200 nm and a constant chromatin density of 130 bp/nm.

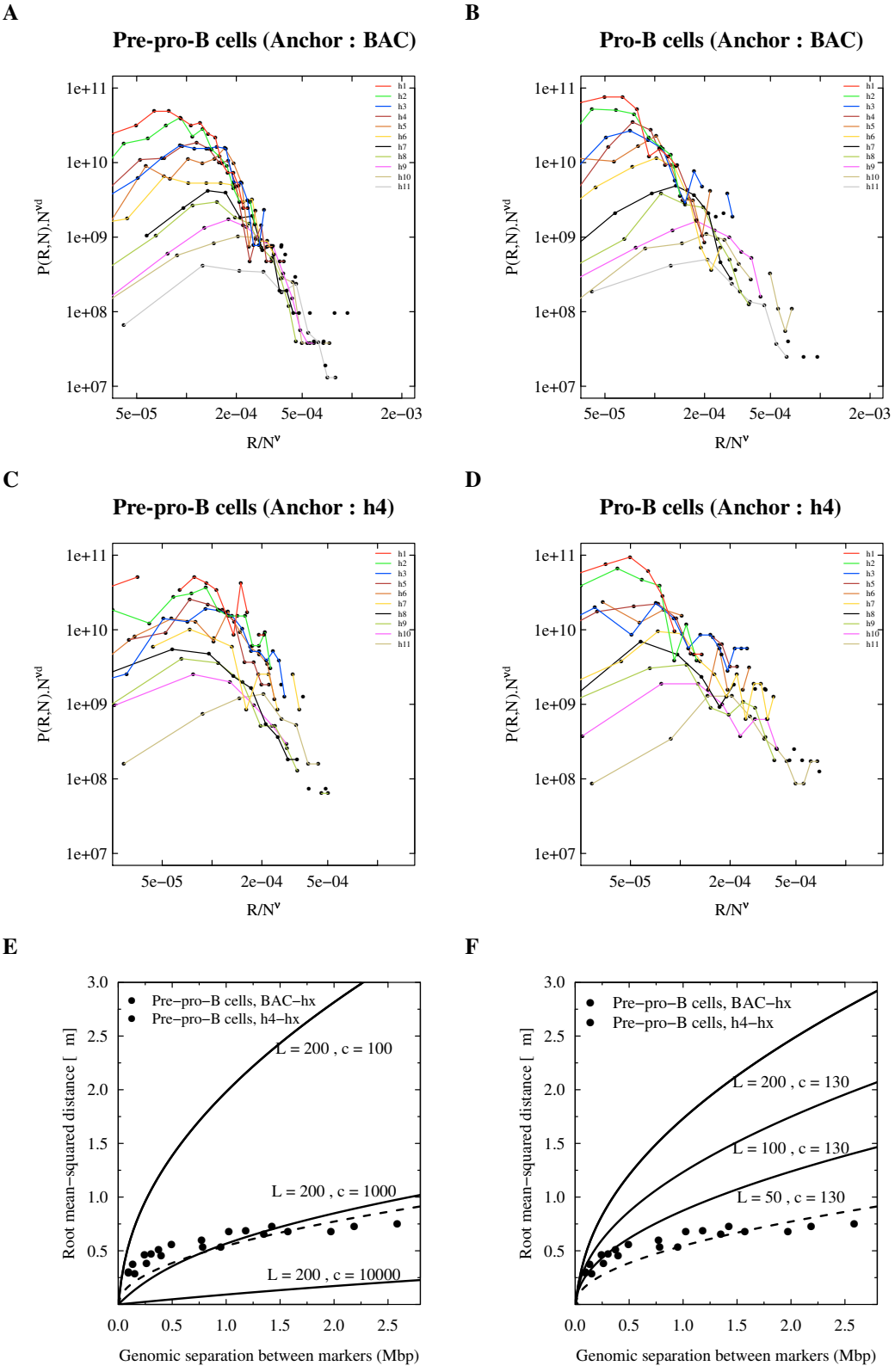
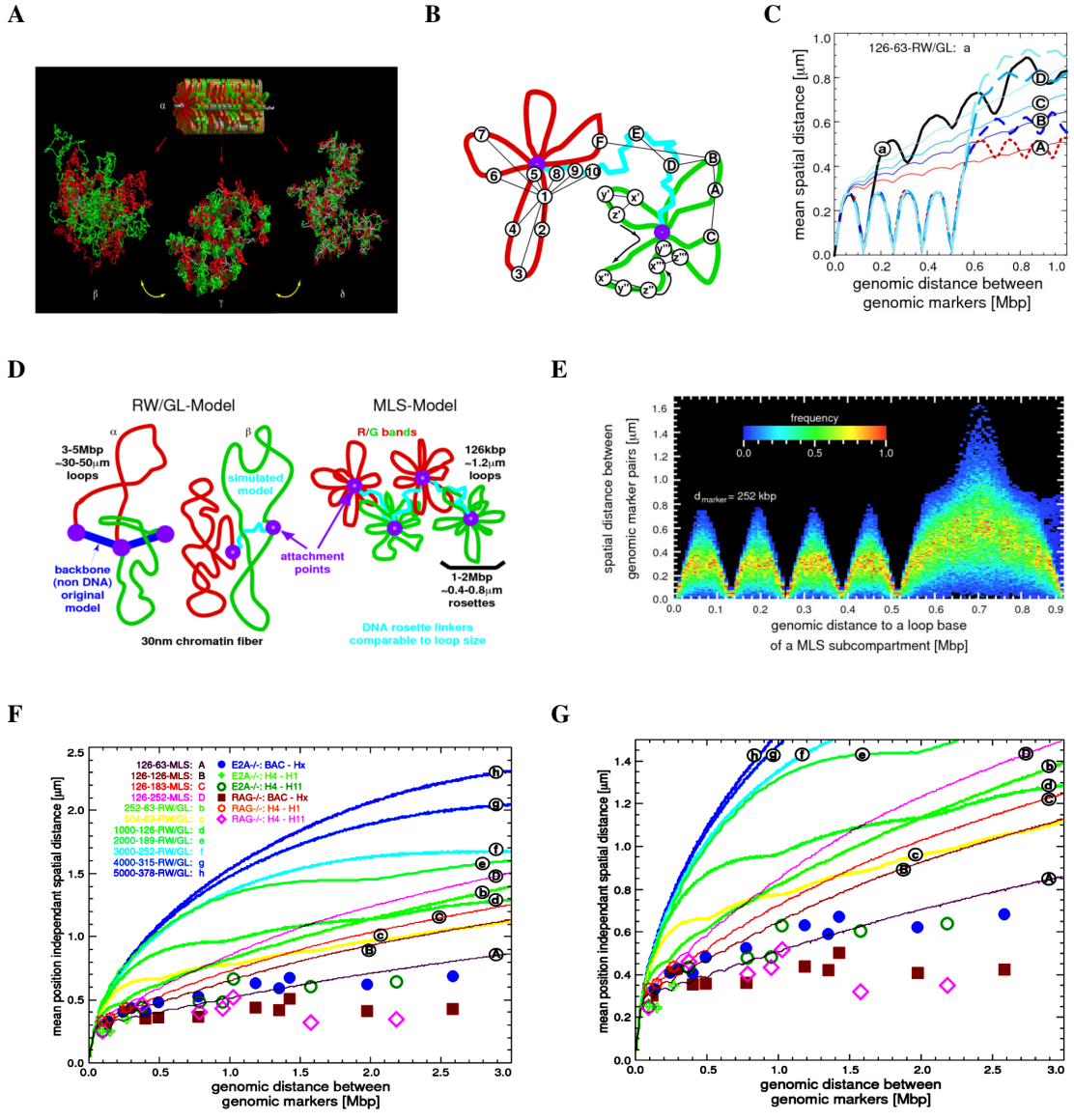

A
Pre-pro-B cells (Anchor : BAC)
B
Pro-B cells (Anchor : BAC)
C
Pre-pro-B cells (Anchor : h4)
D
Pro-B cells (Anchor : h4)
P(R,N).N^vd
R/N^v
h1
h2
h3
h4
h5
h6
h7
h8
h9
h10
h11
E
F
Pre−pro−B cells, BAC−hx
Pre−pro−B cells, h4−hx
L = 200 , c = 100
L = 200 , c = 1000
L = 200 , c = 10000
L = 200 , c = 130
L = 100 , c = 130
L = 50 , c = 130
Root mean−squared distance [m]
Genomic separation between markers (Mbp)

Figure 3.4: Comparison and Evaluation of Spatial Distances between Genomic Markers in the Immunoglobulin Heavy Chain Locus by Computer Simulations
(A) Volume rendered images of simulated Random-Walk/Giant-Loop and Multi-Loop-Subcompartment Models. As a starting conformation with the form and size of a metaphase chromosome (top), rosettes were stacked (α). From such a starting configuration, interphase chromosomes in thermodynamic equilibrium, were decondensed by Monte-Carlo and relaxing Brownian Dynamics steps. A volume rendered image of the simulated Random-Walk/Giant-Loop model containing large loops (5 Mbp) is shown (left, β). Note that the large loops do not form distinct structures but intermingle freely (left, β). In contrast, in a volume rendered image of the simulated Multi-Loop-Subcompartment Model, containing 126 Kbp sized loops and linkers, the rosettes form distinct chromatin territories in which the loops do not intermingle freely (middle, γ). Also is indicated an image of the simulated RW/GL model containing 126 Kbp loops and 63 Kbp linkers (right, δ). Note that the small loops do not intermingle freely. Distinct chromatin territories cannot be detected. (B) Strategy for position-dependent and position-independent virtual spatial distance measurements. For position-dependent virtual distance measurements, the anchor was placed close to the base of the loop (marker 1). The virtual spatial distances were measured from the anchor to other makers in the rosette (1-7) and to a linker (8-10). For position-independent measurements a set of markers separated by the same genomic distance were randomly positioned (x,y,z). (C) Comparison between simulated position-dependent (dotted lines) and position-independent (solid lines) spatial distances. The curves (A-D) indicate simulated MLS models with 126 Kbp loops and different linker sizes. RW/GL is shown for comparison (a). Position-dependent distances (dotted lines) show a stepwise increase in the region where a linker is connecting two chromatin subcompartments, while position-independent distances (solid lines) do not show the stepwise increase in spatial distances as a function of genomic separation. (D) Random-Walk Giant Loop and Multi-Loop-Subcompartment Models. α indicates the RW/GL model in which large loops are attached to a non-DNA backbone. β shows the simulated model containing a chromatin linker between loops. MLS model is shown containing 126 Kbp loops and linkers with individual rosettes spanning 1-2 Mbp. (E) Behavior of a set of markers separated by a fixed genomic separation (256 Kbp) based on their position with respect to a loop base. The simulated loop size was 126 Kbp. Distance distribution between the markers is shown as a function of their shift in 5.2 Kbp steps from a loop base point. (F and G) Comparison between experimental data and computer simulated data obtained from spatial distance measurements in the Igh locus as a function of genomic separation. Nomenclature is loop size [Kbp]-linker size [Kbp]-topology. Experimental spatial distance measurements [μm] were plotted as a function of genomic separation [Mbp] for pre-pro-B cells (blue dots and green circles) and pro-B cells (red squares and pink triangles).


A
B
C
126-63-RW/GL: a
mean spatial distance [µm]
genomic distance between genomic markers [Mbp]
D
RW/GL-Model
MLS-Model
3-5Mbp ~30-50µm loops
backbone (non DNA) original model
simulated model
attachment points
R/G bands
126kbp ~1.2µm loops
1-2Mbp ~0.4-0.8µm rosettes
30nm chromatin fiber
DNA rosette linkers comparable to loop size
E
frequency
d marker = 252 kbp
spatial distance between genomic marker pairs [µm]
genomic distance to a loop base of a MLS subcompartment [Mbp]
F
126-63-MLS: A
126-126-MLS: B
126-183-MLS: C
126-252-MLS: D
252-63-RW/GL: b
504-63-RW/GL: c
1000-126-RW/GL: d
2000-189-RW/GL: e
3000-252-RW/GL: f
4000-315-RW/GL: g
5000-378-RW/GL: h
E2A-/-: BAC - Hx
E2A-/-: H4 - H1
E2A-/-: H4 - H11
RAG-/-: BAC - Hx
RAG-/-: H4 - H1
RAG-/-: H4 - H11
mean position independant spatial distance [µm]
genomic distance between genomic markers [Mbp]
G
mean position independant spatial distance [µm]
genomic distance between genomic markers [Mbp]

the Igh locus in pro-B cells is effectively organized into one compartment.

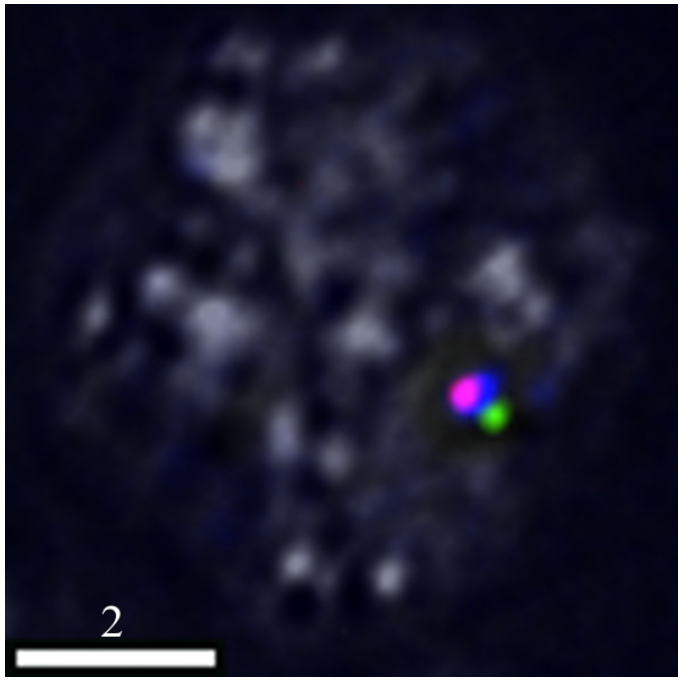
The Igh topology in non-lymphoid cells, for example, mouse embryonic fibroblasts showed a configuration that is even more de-condensed as compared to pre-pro-B cells (Figure 3.6). However, flattening of the spatial distances as a function of genomic separation was also observed in the Igh locus in fibroblasts, indicative of topological confinement as described above for the pre-pro-B and pro-B cell configuration. We note from our observations that the Igh locus in fibroblasts is located in close proximity of the nuclear membrane(data not shown), whereas in B-lineage cells it is positioned in more centrally located domains[11]. It is conceivable that the nuclear location of the Igh locus affects Igh topology. The differences in Igh topology in the different cell types are quite remarkable and it will be of interest to determine how the Igh locus is structured in other cell types as well, including fibroblasts, stem and germ cells.

Acknowledgements

This chapter is partly reprinted from the paper *"The 3D-Structure of the Immunoglobulin Heavy Chain Locus: Implications for Long-Range Genomic Interactions"*, *Cell*, 133(2), 2008. I have spent years working in the dark along with Menno van Zelm and Mandy Peak to generate the data for this project. It would not have been possible to survive, as we call it, the 'dungeon' or the 'beach'

Figure 3.5: 3D-Structure Preserving-Fluorescence in situ Hybridization and Spectral Precision Distance Epifluorescence Microscopy
Three-dimensional fluorescence in situ hybridization in nuclei derived from RAG2-deficient pro-B cells using 10 Kbp probes. Digitally magnified pictures of the Igh locus are shown. Two 10 Kbp-probes (shown in red and green) were labelled with aminoallyl-dUTP using nick-translation followed by incubation with succinimidyl-ester conjugates of Alexa fluorochromes. Nuclei are visualized by DAPI staining.

A



B

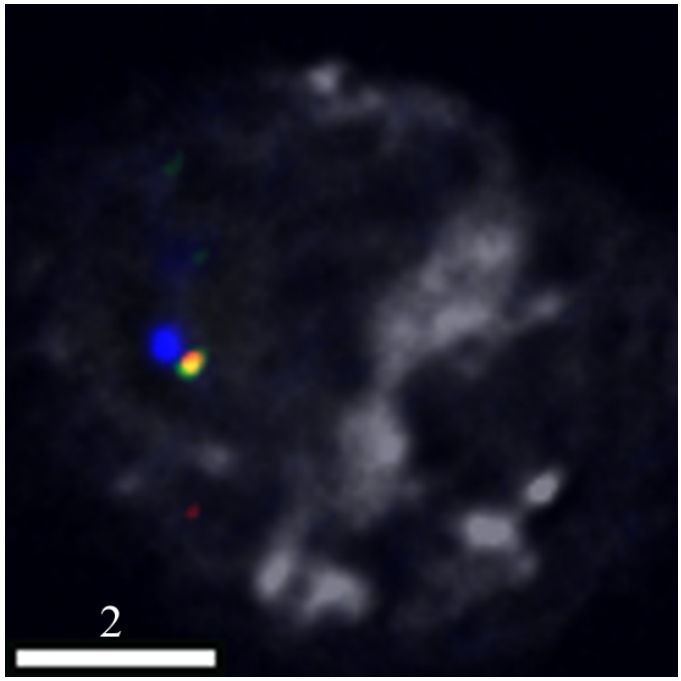
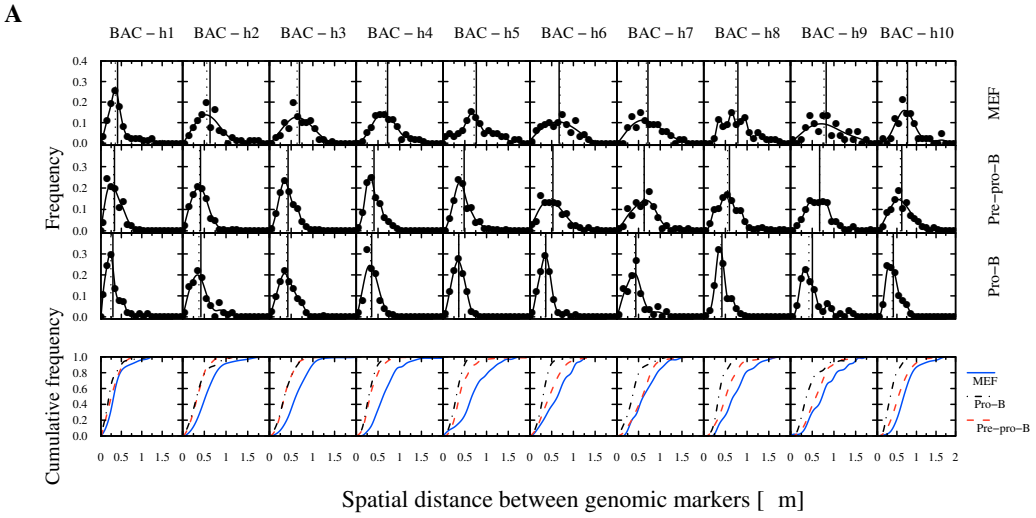
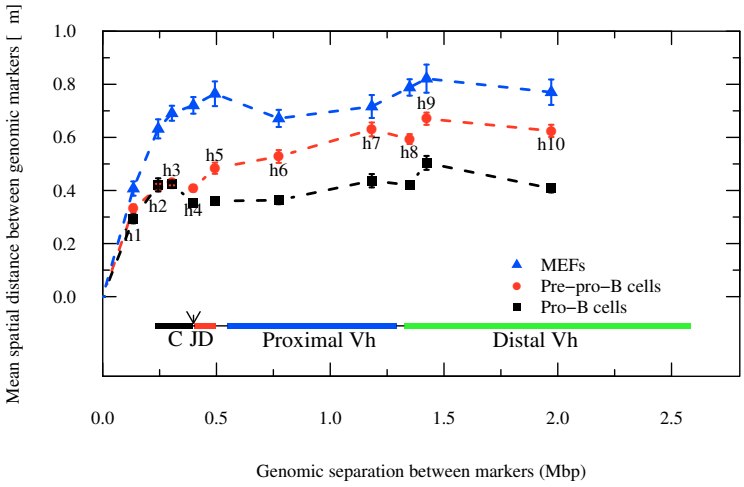


Figure 3.6: Immunoglobulin Heavy Chain Locus Spatial Distances and Spatial Distributions as a Function of Genomic Separation in Murine Embryonic Fibroblasts cells, Pre-Pro-B and Pro-B Cells
(A) Frequency plots showing the distribution of spatial distances between probes and the anchor (RP23-201H14). Cumulative frequency distributions are indicated for murine embryonic fibroblasts, pre-pro-B and pro-B cells. (B) Average spatial distances were plotted as a function of genomic separation for murine embryonic fibroblasts (MEFs), pre-pro-B and pro-B cells. Spatial distances were measured using RP23-201H14 as an anchor as a function of genomic separation towards the telomere. Distal and proximal variable regions as well as diversity, joining and constant region segments are shown. Bars indicate standard error of the mean. The arrow indicates the position of the intronic enhancer. (C) Average spatial distances were plotted as a function of genomic separation for pre-pro-B and pro-B cells. Spatial distances were measured using RP23-201H14 as an anchor as a function of genomic separation towards both the telomere and centromere. Distal and proximal variable regions as well as diversity, joining and constant region segments are shown. Bars indicate standard error of the mean. The arrow indicates the position of the intronic enhancer.


A
BAC – h1
BAC – h2
BAC – h3
BAC – h4
BAC – h5
BAC – h6
BAC – h7
BAC – h8
BAC – h9
BAC – h10
Frequency
Cumulative frequency
MEF
Pre–pro–B
Pro–B
Spatial distance between genomic markers [m]


B
Mean spatial distance between genomic markers [m]
MEFs
Pre–pro–B cells
Pro–B cells
C JD
Proximal Vh
Distal Vh
Genomic separation between markers (Mbp)

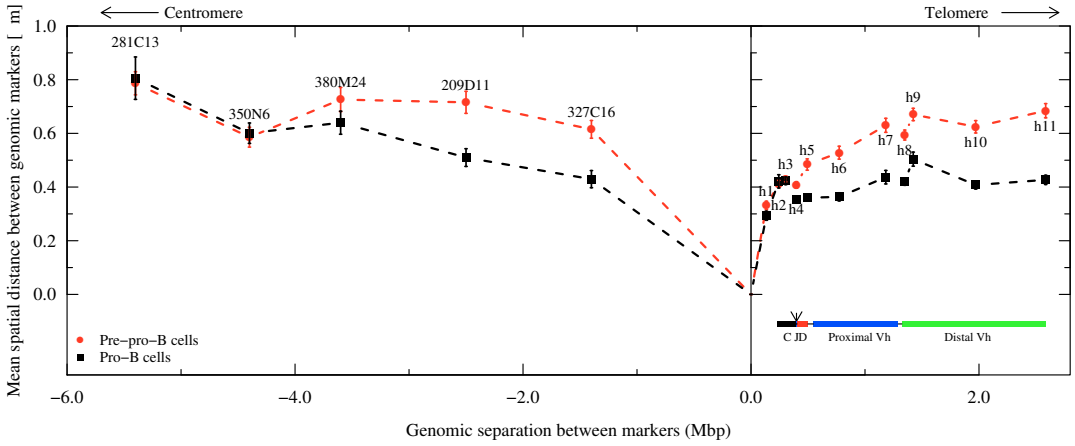
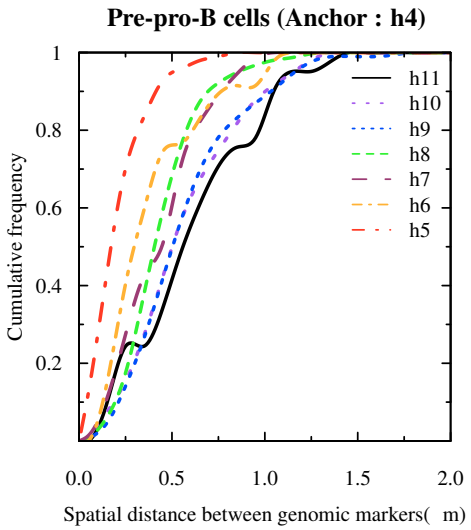
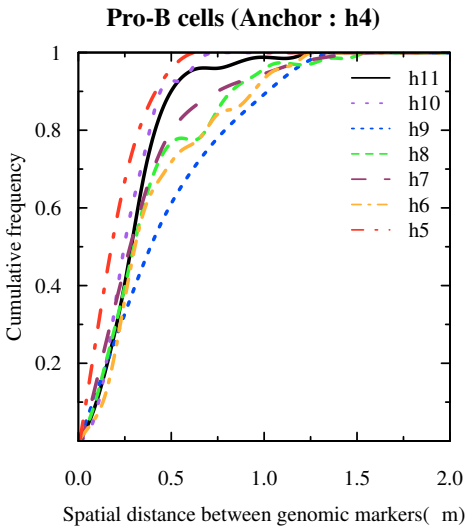

C
Centromere
Telomere
281C13
350N6
380M24
209D11
327C16
Mean spatial distance between genomic markers [m]
Pre–pro–B cells
Pro–B cells
C JD
Proximal Vh
Distal Vh
Genomic separation between markers (Mbp)

Figure 3.7: Cumulative Frequency Distributions as a Function of Genomic Separation
(A) Probabilities for proximal and distal V_H regions to be in spatial proximity to the $D_H J_H$ elements in pre-pro-B cells. Cumulative frequencies were obtained by accumulating the frequency values corresponding to the spatial distances in intervals of 100 nm using the $D_H J_H$ elements (probe h4) as an anchor. (B) Probabilities for V_H regions to be in spatial proximity to the $D_H J_H$ elements in pro-B cells. Cumulative frequencies were obtained by accumulating the frequency values corresponding to the spatial distances in intervals of 100 nm using the $D_H J_H$ elements (probe h4) as an anchor. (C and D) Cumulative frequency distributions separating V_H from $D_H J_H$ elements in pre-pro-B (C) and pro-B cells (D) were plotted for spatial distances separated by less than 250 nm.

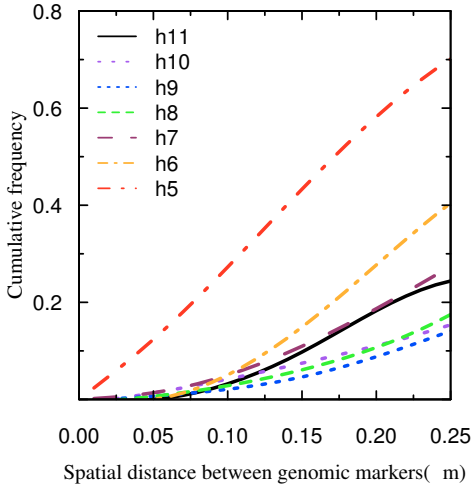
A


Pre-pro-B cells (Anchor : h4)
Cumulative frequency
Spatial distance between genomic markers(m)
h11
h10
h9
h8
h7
h6
h5

B


Pro-B cells (Anchor : h4)
Cumulative frequency
Spatial distance between genomic markers(m)
h11
h10
h9
h8
h7
h6
h5

C


Pre-pro-B cells (Anchor : h4)
Cumulative frequency
Spatial distance between genomic markers(m)
h11
h10
h9
h8
h7
h6
h5

D

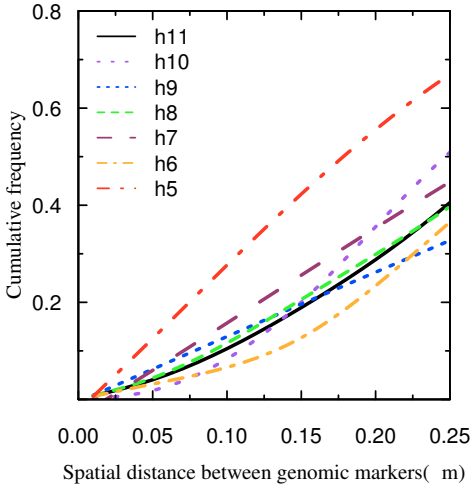
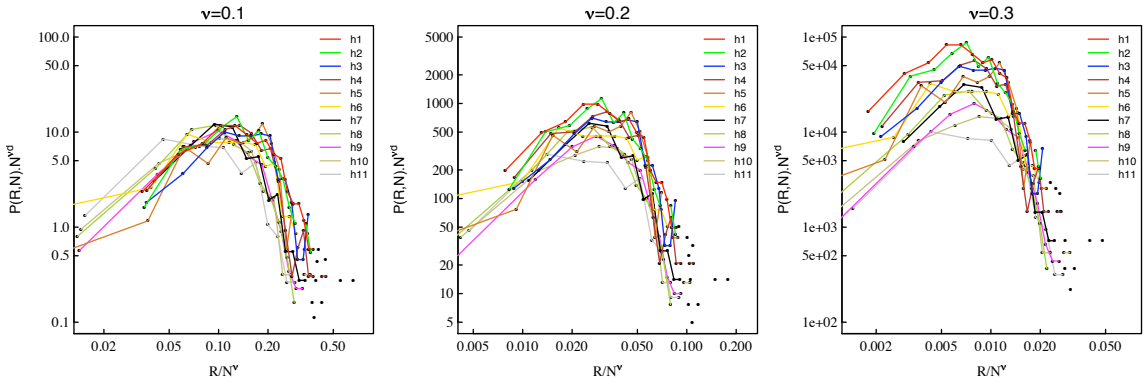

Pro-B cells (Anchor : h4)
Cumulative frequency
Spatial distance between genomic markers(m)
h11
h10
h9
h8
h7
h6
h5

Figure 3.8: Scaled Distribution of Spatial Distances Plotted for Comparison with a Self-Avoiding Random Walk
Probability (P) that two genomic markers separated by 'N' basepairs are at a spatial distance 'R' is indicated. 'd' refers to the number of dimensions (d=3). Different values for the exponent ν (which has the value $\sim$0.6 for a self avoiding random walk) were tested. At lower values of ν (below 0.3), the curves are closer to each other than they are at $\nu = 0.6$. Graphs are shown for both cell types, and with the BAC probe as the anchor.

Pre-pro-B cells, distances from BAC



Pro-B cells, distances from BAC

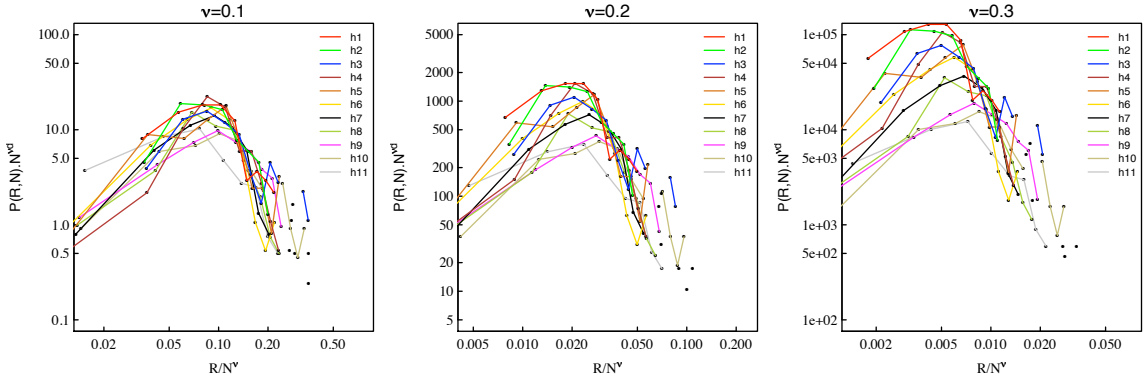


Table 3.1: Average Spatial Distances and Chromatin Compaction as a Function of Genomic Separation in the Immunoglobulin Heavy Chain Locus

n refers to number of alleles analyzed. SD indicates standard deviation. SE refers to standard error. $< x >$ indicates average spatial distance. The average DNA compaction based on the linear contour length of 340 nm for 1.0 Kbp was calculated as follows: $Compaction\ Value = \frac{340[nm].(genomic\ length[kb])}{(1[kb]).(spatial\ distance[nm])}$

marker pair	**[Mbp]**	**E2A$^{-/-}$**				**RAG$^{-/-}$**			
		detection	**nuclei [N]**	spatial distance $< x > \pm\Delta \pm \sigma$ [nm]	**compaction**	**detection**	**nuclei [N]**	spatial distance $< x > \pm\Delta \pm \sigma$ [nm]	**compaction**
BAC - h1	0.14	A647 / A488	131	331.6 ± 181.4 ± 15.8	138	A647/ A488	139	293.2 ± 199.2 ± 16.9	156
BAC - h2	0.24	A647 / A594	262	412.4 ± 203.6 ± 12.6	201	A647 / A594	91	421.1 ± 236.5 ± 24.8	197
BAC - h3	0.30	A647 / A488	391	429.4 ± 196.9 ± 10.0	241	A647 / A488	185	423.4 ± 194.0 ± 14.3	244
BAC - h4	0.40	A647 / A594	300	408.1 ± 206.5 ± 11.9	332	A647 / A594	193	354.1 ± 147.1 ± 10.6	382
BAC - h5	0.49	A647 / A488	185	483.9 ± 286.1 ± 21.0	346	A647 / A488	192	360.2 ± 139.3 ± 10.1	465
BAC - h6	0.77	A647 / A594	135	528.1 ± 278.8 ± 24.0	498	A647 / A594	110	363.4 ± 158.9 ± 15.1	724
BAC - h7	1.18	A647 / A594	114	630.0 ± 278.2 ± 26.1	638	A647 / A594	82	436.8 ± 227.1 ± 25.1	920
BAC - h8	1.35	A647 / A488	227	593.8 ± 278.6 ± 18.5	772	A647 / A488	127	420.3 ± 154.8 ± 13.7	1091
BAC - h9	1.42	A647 / A594	154	670.4 ± 284.9 ± 23.0	722	A647 / A594	125	504.0 ± 293.5 ± 26.2	960
BAC - h10	1.97	A647 / A488	143	624.5 ± 275.0 ± 23.0	1073	A647 / A488	119	407.7 ± 164.9 ± 15.1	1644
BAC - h11	2.59	A647 / A594	143	684.4 ± 317.4 ± 26.5	1284	A647 / A594	115	426.8 ± 185.5 ± 17.3	2059
h4 - h1	0.26	A594 / A488	94	350.1 ± 155.4 ± 16.0	256	A488 / A594	58	440.0 ± 251.4 ± 33.0	204
h4 - h2	0.16	A488 / A594	100	246.0 ± 150.8 ± 15.1	214	A488 / A594	59	357.8 ± 256.6 ± 33.4	147
h4 - h3	0.09	A594 / A488	188	258.8 ± 161.9 ± 11.8	124	A488 / A594	67	337.4 ± 241.5 ± 29.5	95
h4 - h5	0.10	A594 / A488	97	252.3 ± 155.9 ± 15.8	128	A594 / A488	38	255.5 ± 140.7 ± 22.8	126
h4 - h6	0.38	A488 / A594	47	436.0 ± 260.3 ± 38.0	293	A488 / A594	63	465.6 ± 310.7 ± 39.1	274
h4 - h7	0.78	A488 / A594	73	479.4 ± 233.0 ± 27.3	556	A488 / A594	55	407.5 ± 287.5 ± 38.8	654
h4 - h8	0.95	A594 / A488	59	483.5 ± 227.5 ± 29.6	668	A594 / A488	68	440.0 ± 309.6 ± 37.5	734
h4 - h9	1.03	A488 / A594	93	609.8 ± 307.0 ± 31.8	450	A488 / A594	42	524.3 ± 336.8 ± 52.0	665
h4 - h10	1.57	A488 / A594	70	607.2 ± 301.6 ± 36.1	881	A488 / A594	55	325.7 ± 145.7 ± 19.6	1642
h4 - h11	2.19	A488 / A594	41	641.9 ± 346.7 ± 54.1	1158	A488 / A594	74	353.5 ± 194.1 ± 22.6	2103

Table 3.2: Full Width at Half the Maximum (FWHM) Values for the Point Spread Function (PSF) for Each Color

Color	**FWHM**$_X$ [nm]	**FWHM**$_Y$ [nm]	**FWHM**$_Z$ [nm]
A 488	543	489	1800
A 594	380	380	1400
A 647	434	434	1700

Table 3.3: Chromatic Shift 'c' Between Different Fluorophores and Resolution Equivalents RE Before (RE$_b$) and After (RE$_a$) Applying Chromatic Corrections "A" indicates Alexa fluorochromes

Colors	**A 488 / A 594**	**A 488 / A 647**	**A 594 / A 647**
Number of Distances [N]	226	226	226
$< c_x > \pm\sigma \pm \Delta$ [nm]	+31.7± 16.5± 1.1	+53.5± 30.2± 2.0	-21.8± 22.9± 1.5
$< c_y > \pm\sigma \pm \Delta$ [nm]	-14.4± 17.2± 1.1	-2.8± 31.5± 2.1	-11.6± 25.0± 1.7
$< c_z > \pm\sigma \pm \Delta$ [nm]	-133.0± 35.3± 2.3	-159.1± 24.9± 1.7	+26.1± 36.6± 2.4
$< RE_b > \pm\sigma \pm \Delta$ [nm]	139.7± 34.6± 2.3	173.2± 26.2± 1.7	56.3± 24.6± 1.6
$< RE_a > \pm\sigma \pm \Delta$ [nm]	34.9± 24.3± 1.6	46.8± 17.8± 1.2	44.6± 22.1± 1.5

without them. Roy Riblet once again helped us with the map and small probes for the locus. Tobias Knoch was the backbone of the computer simulations for RW/GL and MLS models in this project, and it was his work that inspired us to compare the data to structural models. Steve Cutchin and Amit Chourasia, at the San Diego Supercomputer Center helped us in the trilateration study and made some wonderful 3D movies and figures. I thank Jacques van Dongen, Frank Grosveld and Douglas Forbes for their help with the manuscript. I thank Pete Carlton and John Sedat (UCSF) for their generous help with using their SIM system. Terry Hwa provided us his expert advice in comparing our data to random walk models and inferring compartmentalization.

Chapter 4

Spatial Visualization of the Igh Locus and Compartmentalization

Comparison of the 3D-architecture of the Igh locus, as deciphered using 3D-FISH, with structural models suggests compartmentalization of the Igh locus. It essentially implies that the 30 nm chromatin fiber has a higher order structure, and compartmentalization is an integral component of the organization. Here, we further visualize the Igh locus using different strategies. We use geometry to place the hybridization probes in 3D-space, which reveals the relative positioning of the different subregions. We also visualize compartments using deconvolution microscopy and structural illumination microscopy.

4.1 Geometric Derivation of the Average Structure using Trilateration

By assuming geometric integrity of the average 3D-structure, the relative 3D-coordinates of the V_H, D_H, J_H and C_H elements within the Igh locus can be determined by the technique trilateration. To obtain the 3D-coordinates of genomic markers by trilateration, spatial distances of all the markers from a minimum of four reference probes are needed. Consequently, the spatial distances separating all the probes from each of the three reference probes, BAC RP23-201H14, h3 and h4, were determined (Table 4.1 , See Methods for details). Instead of using a fourth reference point, additional measurements were made for consecutive probe-pairs. Applying this methodology, the average positions for the genomic markers (BAC, h1-h11) that span the entire Igh locus were determined (Table 4.1). Any disagreements between the distances predicted by the obtained 3D structure and observed measurements were treated as errors and the disagreements were minimized by a steepest descent algorithm. The final errors generated by this analysis ranged between 5-100 nm for pro-B cells and 40-150 nm for pre-pro-B cells (Table 4.2). Although it remains to be proven, the error minimization assuming geometric integrity of the average structure may result in a more accurate representation of an average, statistical 3D-structure.

In pre-pro-B cells, relatively large spatial distances separated the distal V_H,

proximal V_H and C_H regions (Figure 4.1A). The C_H elements were clustered (h2-h4) (Figure 4.1A, grey objects). Interestingly, the D_H segments (h4 and h5) were positioned away from the majority of the V_H regions (h6-h11) (Figure 4.1A, red connector). The proximal V_H regions (h5-h8) were located adjacent to the distal portion of the markers that contain D_H elements and were spatially separated from the distal V_H regions (h9-h11) (Figure 4.1A, green objects). The Igh topology in pro-B cells was strikingly different. The pro-B Igh locus appeared to have collapsed for the V_H regions that were visualized (Figure 4.1B). Most remarkable was the merging of the proximal and distal V_H regions (h6-h11) (Figure 4.1B). Whereas in pre-pro-B cells the $D_H J_H$ elements (h4-h5) were positioned away from the majority of V_H regions, in pro-B cells they were juxtaposed and within relative close proximity to the entire V_H region repertoire (Figure 4.1B).

The data described above bring into question whether the relative average 3D-coordinates of the V_H, D_H, J_H and C_H elements determined by trilateration reflect the average trajectory taken by the Igh fiber in a single B cell. The trilateration analysis considers only two-point measurements obtained from average distances that were accumulated from large numbers of cells. Two point measurements, however, do not describe the complete trajectory of the chromatin fiber in a single cell. Three-point measurements reflect a triangular sub-trajectory of the complete trajectory. Thus, as a first approach to determine to what degree the trajectory revealed by trilateration reflects the route taken by the Igh fiber

in single cells, we determined the angular distributions between different combinations of the triple point measurements that were performed. The medians of the experimentally derived angles were then compared to the angles obtained by trilateration. The majority of the median angles observed in pre-pro-B and pro-B cells compared well with those determined by trilateration (Figure 4.2). The agreement is stronger in pre-pro-B cells than in pro-B cells. This indicates that a stable average structure may exist in pre-pro-B cells, whereas the structure may be more unstable from the observed average in pro-B cells.

The trilateration analysis also indicated oscillation of spatial distances within the pro-B compartment containing the V_H cluster. For example, h5 is separated by 1.5 Mbp of DNA from h10 but positioned in relatively close spatial proximity (Figure 4.1B). On the other hand, h5 and h6 are in relatively close genomic proximity but separated by a relatively large spatial distance (Figure 4.1B). These observations reveal how the Igh topology permits DNA elements that are separated by large genomic distances to be localized in relatively close spatial proximity.

We note that although the data points that were analyzed span the entire Igh locus, the number of genomic markers that could be used in this study was restricted. It will be important to provide a higher resolution average trajectory of the Igh locus, but this will require a large set of relatively small probes using a strategy recently described for protein folding[62]. Collectively, these studies show

the relative average spatial positions of the V_H, D_H, J_H and C_H elements in both pre-pro-B and pro-B cells. Furthermore, upon commitment to the B cell fate, striking conformational changes in Igh topology occur, to allow the entire V_H repertoire encounter D_H elements with relatively high and similar frequencies.

The repositioning of the D_H cluster and the merging of the proximal and distal V_H regions during early B cell development are intriguing. We propose that these conformational changes underpin the mechanism by which a diverse antibody repertoire is established. Although separated by large genomic distances, apart from frequent rearrangements involving the two most proximal V_H regions, V_H81X and V_HQ52, little correlation has been observed between V_H region usage and V_H genomic location[63, 16, 43, 64]. Thus, V_H regions despite separated by large genomic distances rearrange with similar frequencies.

How do V_H elements separated by large genomic distances find $D_H J_H$ elements with probabilities that are similar to V_H elements that are localized within close genomic proximity? The trilateration analysis and comparison of experimental and simulated data indicates that it is the folding of the Igh fiber that allows a 2 Mbp genomic region to be packed as bundles of loops, allowing the entire V_H region repertoire, regardless of their genomic location, similar access to the $D_H J_H$ elements.

4.2 Compartmentalization of the Immunoglobulin Heavy Chain Locus

To examine whether probes that are nearby in terms of genomic separation are localized within the same subcompartment or distinct subcompartments, we analyzed whether groups of genomic markers located in spatial proximity move coordinately towards or away from the anchor in a single cell (Figure 4.3A). As an anchor we used BAC RP23-201H14. Triple point spatial distances were derived from single cells separating the anchor and two consecutive probes (h1-h2, h2-h3, h3-h4, h4-h5, h5-h6, h6-h7, h7-h8, h8-h9, h9-h10, h10-h11). We then computed the group correlation coefficients to determine the degree of coordinated movement. A correlation coefficient of 1 indicates perfectly coordinated positioning of the two markers relative to the anchor. On the other hand, a correlation coefficient of 0 demonstrates a lack of linear correlation in the distances that separate the markers from the anchor in single cells.

There was significant correlation between the spatial distances for the majority of consecutive probes separated from the anchor in pre-pro-B cells (Figure 4.3A). Two exceptions were notable, markers h1 and h2 as well as h6 and h7 seem to move independently with respect to the BAC RP23-201H14 anchor, suggesting that these markers are located in separate compartments (Figure 4.3A). Thus, these data suggest that in pre-pro-B cells the Igh locus is organized into at least

three distinct compartments (BAC-h1, h2-h6, h7-h11). On the other hand, the correlation coefficients for pro-B cells were substantially lower when compared to those of pre-pro-B cells (Figure 4.3B). Consequently, these data support a pre-pro-B Igh configuration in which the positions of markers within a chromatin compartment relative to a marker in another compartment are relatively fixed. Hence the macro-trajectory of the chromatin fiber as described by trilateration is more well-defined for pre-pro-B cells whereas the trajectories of the Igh fiber in pro-B cells show more flexibility.

If the Igh topology were to be organized as compartments in pre-pro-B cells, predicted by the comparison of the experimental and simulated data and the grouping analysis, we might be able to visualize chromatin territories if the entire Igh locus were to be fluorescently labeled. Hence, we labeled a set of overlapping BACs that comprise the entire Igh locus (Figure 4.4A). Both pre-pro-B and pro-B cells were hybridized with probes encoding the entire locus and analyzed by epifluorescence microscopy as described above. In pre-pro-B cells 1-3 clusters were visualized whereas in pro-B cells only one cluster was distinguishable (Figure 4.4B and C). Additionally, the linkers connecting the compartments could be visualized.

4.3 Visualization of Compartments by Structural Illumination Microscopy

Visualization of structural details of each compartment using deconvolution or confocal microscopy has inherent limitations due to diffraction of light. This imposes a limit on the resolution of the acquired image. Overcoming this fundamental limit on the resolution imposed by diffraction is a significant focus of current research in light microscopy. John Sedat's group has developed a technique called structural illumination microscopy (SIM), which enhances the resolution by a factor of two[65]. We used this technique to investigate the structural details of compartments of the Igh locus.

We labeled the stretch of the Igh locus spanning the proximal V_H segments and the constant region with overlapping BAC probes(Figure 4.5). In pre-pro-B cells we expected to observe two compartments, the proximal V_H compartment, and the region containing the D_H/J_H segments and the constant segments. Indeed, we see two compartments in most of the cells examined. Focusing on individual compartments reveals that each compartment is made up of at least three lobe-like structures (Figure 4.6). These lobe-like structures might represent loops of the chromatin fiber within a rosette. However, even with this technique, we are at the edge of the resolution required to conclusively determine the fine structural details of a single compartment. The current observations strongly suggest

that a compartment is composed of bundles of loops of DNA. With the advent of higher resolution microscopy techniques, like stimulated emission depletion (STED)[66], photo-activated localization microscopy (PALM)[67], stochastic optical reconstruction microscopy (STORM)[68] it might be possible to understand the detailed structure of each compartment. These techniques, when fully developed and readily available should be of great utility for such studies.

Acknowledgements

This chapter is partly reprinted from the paper *"The 3D-Structure of the Immunoglobulin Heavy Chain Locus: Implications for Long-Range Genomic Interactions"*, *Cell*, 133(2), 2008. I have spent years working in the dark along with Menno van Zelm and Mandy Peak to generate the data for this project. It would not have been possible to survive, as we call it, the 'dungeon' or the 'beach' without them. Roy Riblet once again helped us with the map and small probes for the locus. Tobias Knoch was the backbone of the computer simulations for RW/GL and MLS models in this project, and it was his work that inspired us to compare the data to structural models. Steve Cutchin and Amit Chourasia, at the San Diego Supercomputer Center helped us in the trilateration study and made some wonderful 3D movies and figures. I thank Jacques van Dongen, Frank Grosveld and Douglas Forbes for their help with the manuscript. I thank Pete Carlton and

Table 4.1: Average Spatial Distances [μm] Separating Genomic Markers Spanning the Entire Immunoglobulin Heavy Chain Locus in Pre-Pro-B and Pro-B Cells

Spatial distances were determined between genomic markers using 3D-FISH and epifluorescence microscopy. Both pre-pro-B and pro-B cells were analyzed. Distances separating the various probes are indicated in the matrix and analyzed using trilateration.

		Pre-Pro-B cells											
Pro-B cells		**BAC**	**h1**	**h2**	**h3**	**h4**	**h5**	**h6**	**h7**	**h8**	**h9**	**h10**	**h11**
	BAC		0.3117	0.4128	0.4401	0.4027	0.4350	0.5040	0.5958	0.5351	0.6568	0.5900	0.6059
	h1	0.2788		0.3793	0.3903	0.3348							
	h2	0.3936	0.4182		0.2401	0.2471	0.2799						
	h3	0.4578	0.4757	0.3229		0.2471	0.2687	0.3643	0.5847	0.4141	0.5287	0.5769	0.6433
	h4	0.3378	0.4013	0.3453	0.3353		0.2638	0.4082	0.4631	0.4504	0.7461	0.5726	0.6364
	h5	0.3882			0.3580	0.2589		0.3407					
	h6	0.3512			0.3698	0.4189	0.2876		0.4697				
	h7	0.4009			0.4192	0.3479		0.3722		0.3243			
	h8	0.4208			0.4407	0.4165			0.2944		0.2535	0.3582	0.4093
	h9	0.4519			0.3884	0.5249				0.3239			
	h10	0.4167			0.4107	0.3226					0.4223		0.3435
	h11	0.4086			0.4184	0.3275						0.2540	

Table 4.2: 3D-Cordinates of the Hybridization Probes Obtained by Trilateration Before and After Error Reduction

Pre-pro-B cells									
Before Error reduction					**After error reduction**				
Probe	**x**	**y**	**z**	**Error**	**Probe**	**x**	**y**	**z**	**Error**
BAC	0.0000	0.0000	0.0000	0.0000	**BAC**	0.0103	-0.0084	-0.0005	0.0679
H01	0.1573	0.1148	-0.2434	0.0539	**H01**	0.1374	0.1242	-0.2349	0.0524
H02	0.3481	0.1729	0.1389	0.1164	**H02**	0.3553	0.1155	0.1226	0.0699
H03	0.4401	0.0000	0.0000	0.0267	**H03**	0.4176	-0.0027	-0.0338	0.1521
H04	0.3349	0.2236	0.0000	0.1910	**H04**	0.2834	0.2694	0.0008	0.1357
H05	0.3530	0.1851	-0.1743	0.1883	**H05**	0.4170	0.1599	-0.1436	0.1061
H06	0.3579	0.1337	0.3287	0.2888	**H06**	0.4883	0.0314	0.2303	0.1169
H07	0.2350	0.2951	0.4612	0.2407	**H07**	0.2243	0.3031	0.4538	0.0398
H08	0.3749	0.1164	0.3636	0.2260	**H08**	0.4394	0.0643	0.3564	0.0848
H09	0.3926	-0.2070	0.4842	0.3829	**H09**	0.3926	-0.2070	0.4842	0.1109
H10	0.2374	0.1139	0.5280	0.4605	**H10**	0.3226	-0.1990	0.4624	0.1294
H11	0.1670	0.0740	0.5777	0.2691	**H11**	0.1341	0.0599	0.5767	0.0444

Pro-B cells									
Before Error reduction					**After error reduction**				
Probe	**x**	**y**	**z**	**Error**	**Probe**	**x**	**y**	**z**	**Error**
BAC	0.0000	0.0000	0.0000	0.0000	**BAC**	-0.0193	0.0788	-0.0168	0.0777
H01	0.0666	-0.0045	-0.2707	0.1619	**H01**	0.0687	-0.0422	-0.2492	0.0048
H02	0.2793	0.0161	0.2686	0.2110	**H02**	0.2427	-0.1460	0.1181	0.0381
H03	0.4578	0.0000	0.0000	0.0205	**H03**	0.4659	0.0231	-0.0010	0.0887
H04	0.2307	0.2467	0.0000	0.0252	**H04**	0.2440	0.2106	0.0125	0.0937
H05	0.2535	0.1574	0.2482	0.1378	**H05**	0.2391	0.0853	0.2569	0.0309
H06	0.2143	-0.0949	0.2616	0.1705	**H06**	0.1811	-0.1808	0.1232	0.0269
H07	0.2068	0.1008	0.3203	0.2777	**H07**	0.1980	0.1121	0.3254	0.0293
H08	0.1900	0.0517	0.3718	0.4393	**H08**	0.1170	-0.1642	0.2426	0.1002
H09	0.2872	-0.2143	0.2754	0.0620	**H09**	0.2608	-0.2941	0.0341	0.0589
H10	0.2344	0.1474	0.3114	0.3375	**H10**	0.2808	0.0686	0.3107	0.1072
H11	0.2169	0.1385	0.3128	0.3767	**H11**	0.1936	0.2458	0.2936	0.0816

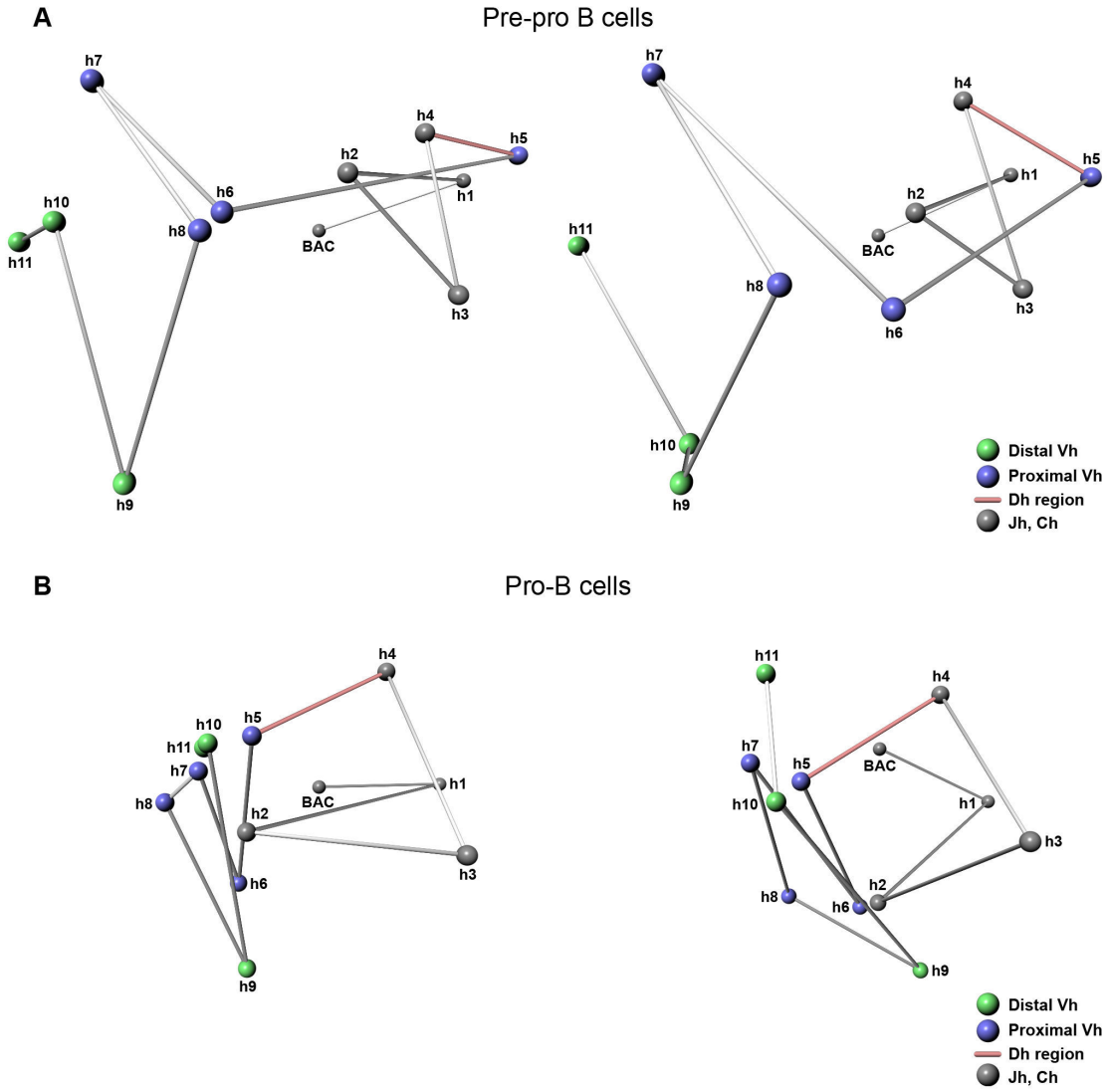
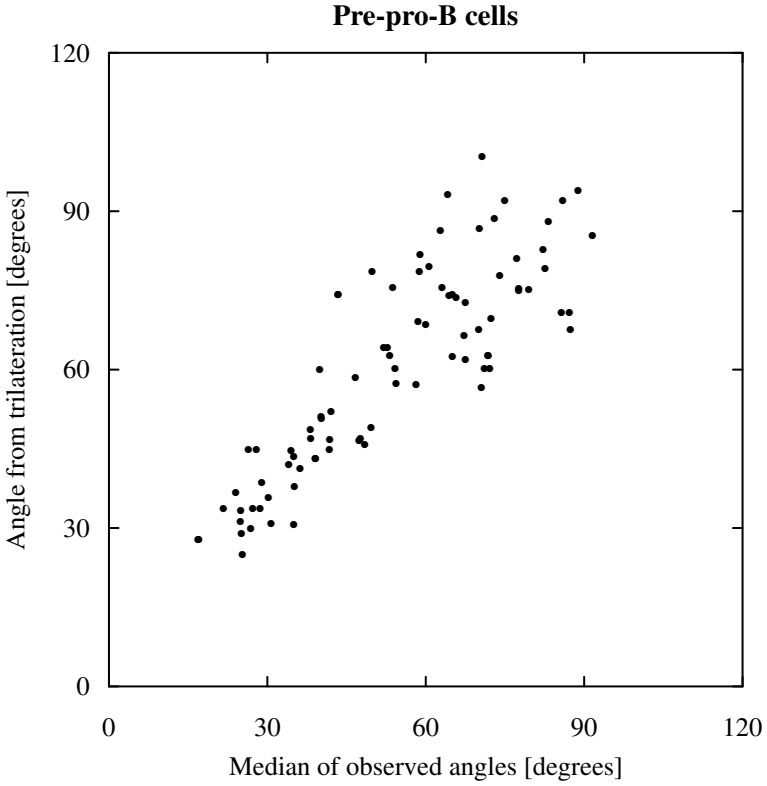


Figure 4.1: 3D-Topology of the Immunoglobulin Heavy Chain Locus by Trilateration

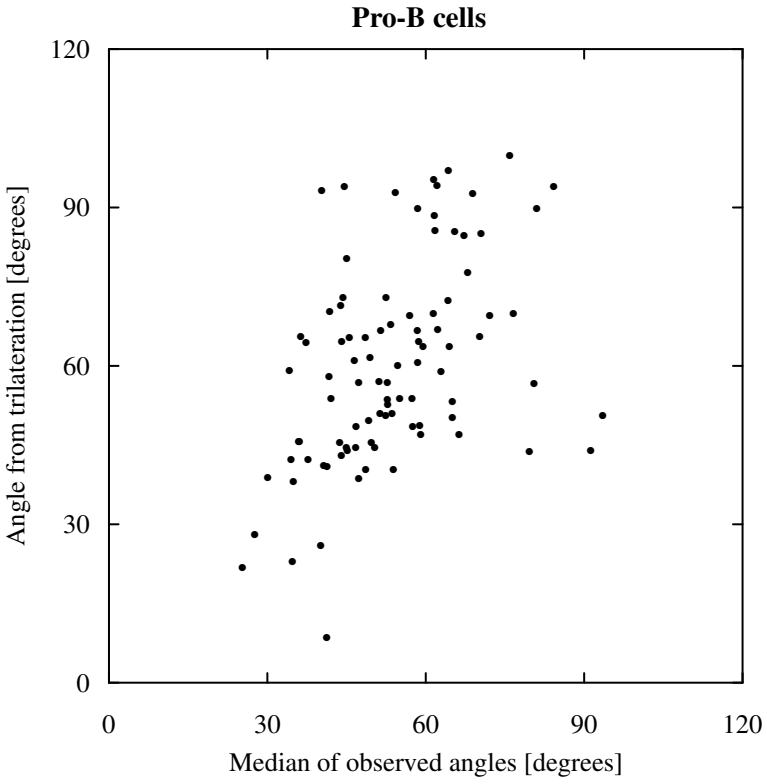
The 3D-topology of the Igh locus in pre-pro-B and pro-B cells was resolved using trilateration. The relative positions of 12 genomic markers spanning the entire immunoglobulin heavy chain locus were computed. Two different views are shown for both cell types. (A) 3D-Topology of the Igh locus in pre-pro-B cells. (B) 3D-Topology of the Igh locus in pro-B cells. Grey objects indicate C_H regions and the 3' flanking region of the Igh locus. Blue objects indicate proximal V_H regions. Green objects indicate distal V_H regions. Red line indicates the linker connecting the proximal V_H and J_H regions. Linkers are indicated only to show connectivity.

Figure 4.2: Comparison of Angles Derived From Three Point Measurements and Trilateration
(A-B) The distribution of angles for different combinations for three probes was determined and the median angles that were observed were compared to those obtained using trilateration.

A


Pre-pro-B cells
Angle from trilateration [degrees]
Median of observed angles [degrees]
0
30
60
90
120

B


Pro-B cells
Angle from trilateration [degrees]
Median of observed angles [degrees]
0
30
60
90
120

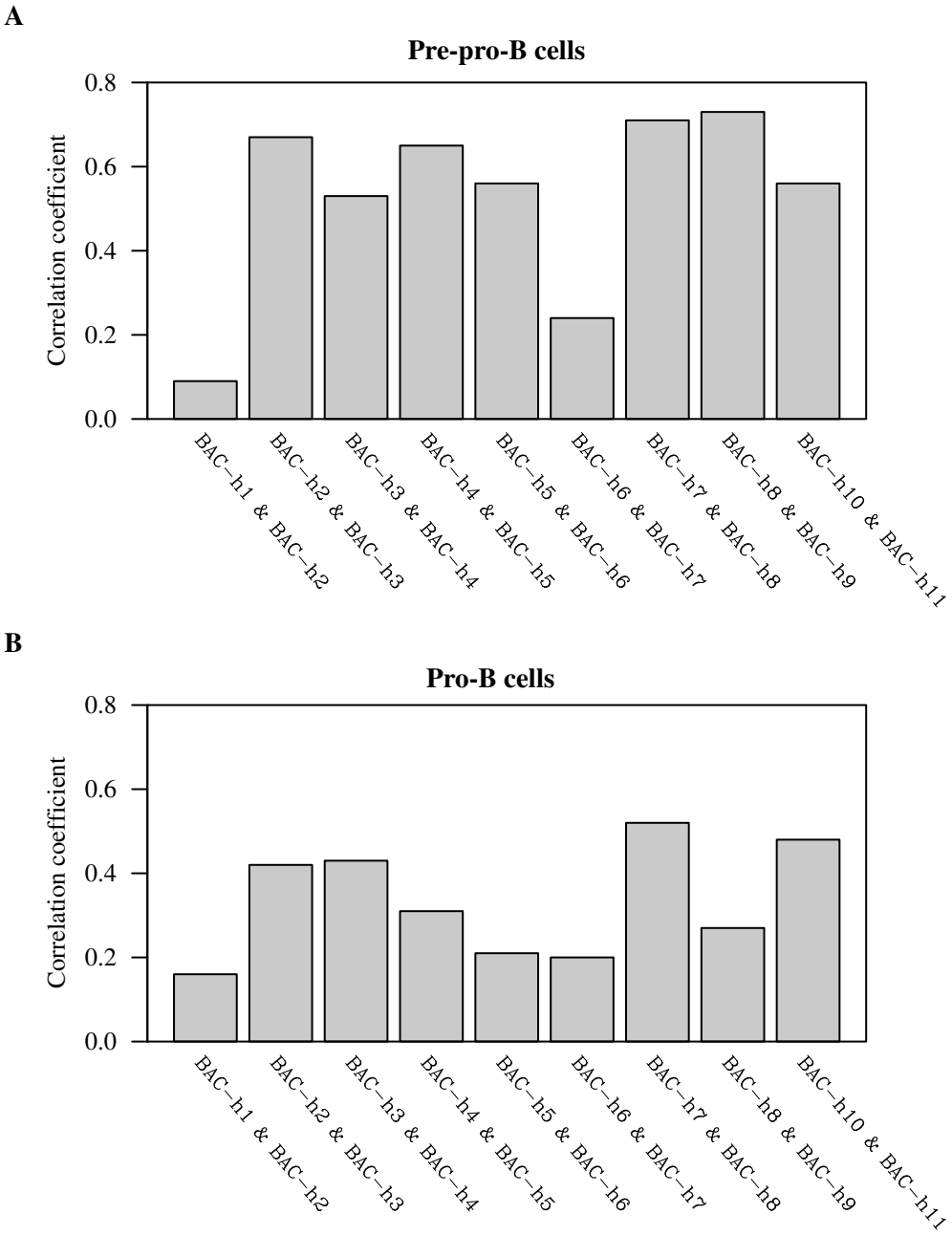
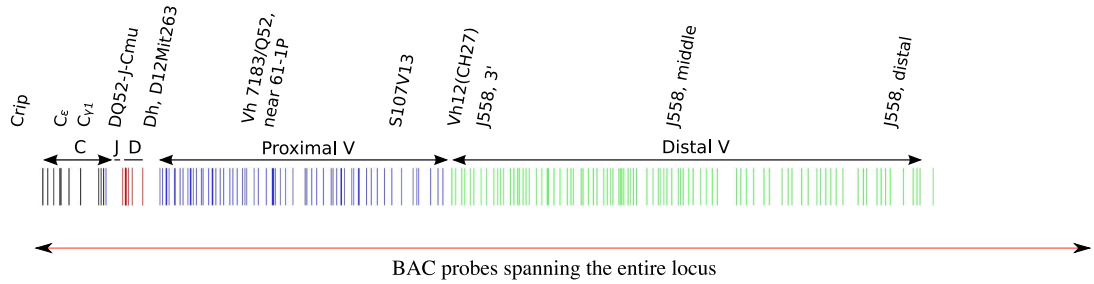


Figure 4.3: DNA Elements within a Chromatin Compartment are Relatively Fixed Correlation Coefficients were computed for spatial distances separating two consecutive genomic markers from an anchor located in a distinct chromatin territory. Correlation coefficients for both pre-pro-B (A) and pro-B (B) cells are shown.

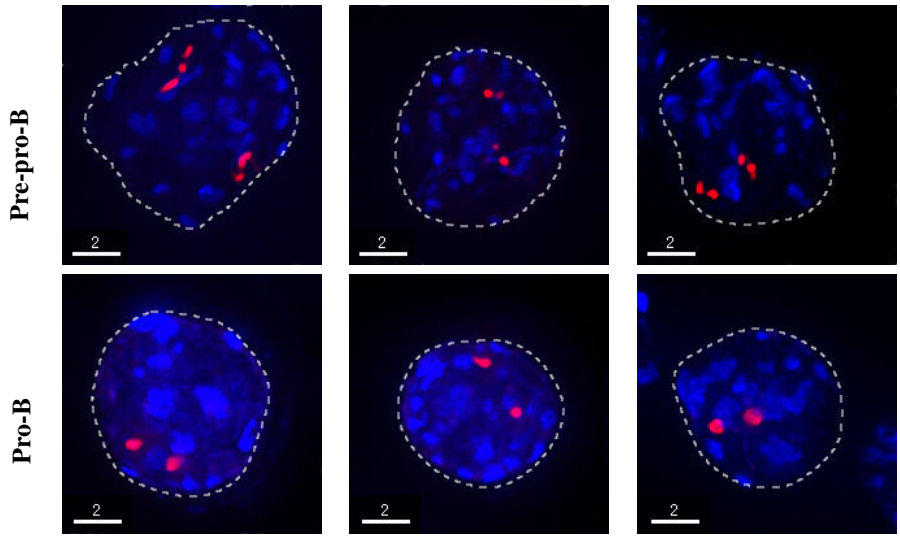
Figure 4.4: Visualization of the Entire Immunoglobulin Heavy Chain Locus

(A) Igh locus was fluorescently labeled using BACs that span the entire locus. (B) 3D- FISH in nuclei derived from pre-pro-B and pro-B cells using fluorescently labeled BACs that span the entire Igh locus. Digitally magnified pictures of the Igh locus are shown. BACs (shown in red) were directly labeled with dUTP conjugated to Alexa 568. Nuclei were visualized by DAPI staining. (C) Fractions of the number of compartments visualized using overlapping set of BAC probes in pre-pro-B and pro-B cells are indicated.

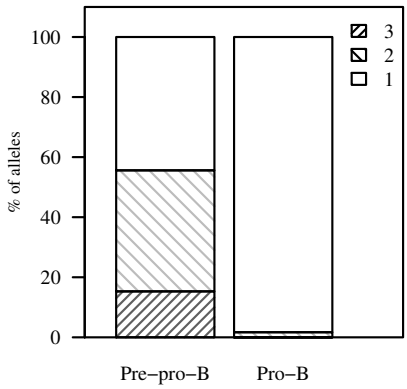
A


Crip
Cε
Cγ1
DQ52-J-Cmu
Dh, D12Mit263
Vh 7183/Q52, near 61-1P
S107V13
Vh12(CH27)
J558, 3'
J558, middle
J558, distal
C
J
D
Proximal V
Distal V
BAC probes spanning the entire locus

B


Pre-pro-B
Pro-B
2

C


100
80
60
40
20
0
% of alleles
3
2
1
Pre−pro−B
Pro−B

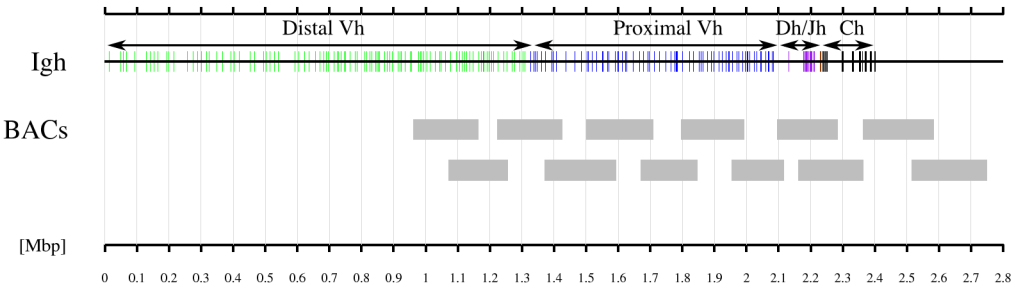


Figure 4.5: Hybridization probes for SIM

Overlapping BAC probes were chosen for the region of the Igh locus spanning from the proximal V_H segments to the 3' end of the locus. The grey boxes represent the BAC probes.

John Sedat (UCSF) for their generous help with using their SIM system. Terry Hwa provided us his expert advice in comparing our data to random walk models and inferring compartmentalization.

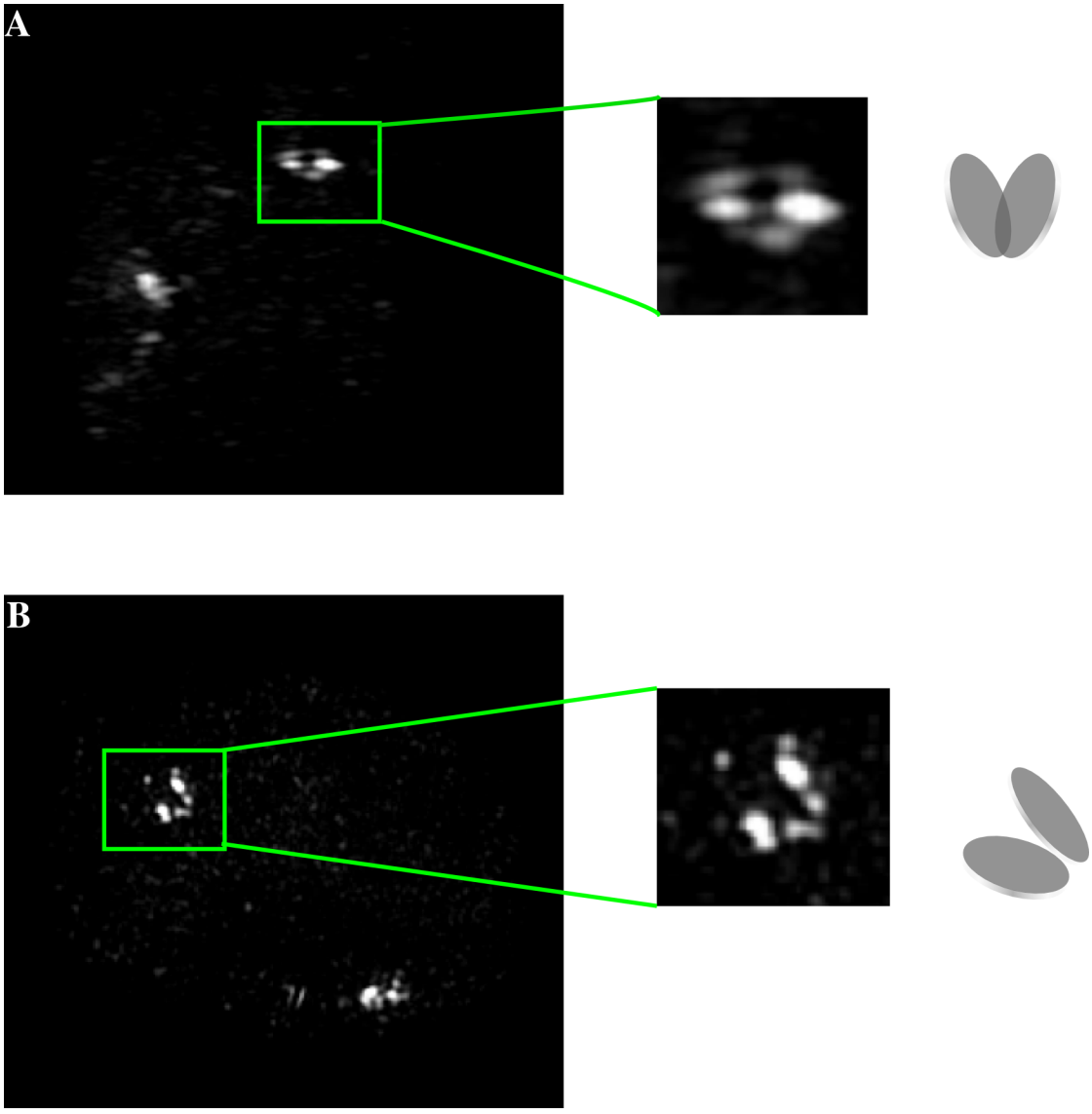


Figure 4.6: Visualization of Proximal V_H and 3' Igh compartments by SIM

Two dimensional projections of two cells hybridized with probes spanning from the proximal V_H segments to the 3' end of the Igh locus are shown (A & B). Both alleles in each cell are readily observable. One allele from each cell, shown by the green box is shown in a zoomed view. The discs at the right depict planar sub-compartments.

Chapter 5

Implications on Long-Range Genomic Interactions

5.1 Spatial Distributions, Besides Average Spatial Proximity, Contribute to Genomic Interactions

In addition to the large conformational changes that seem to accompany early B cell development, we noted changes in the spatial distributions in pre-pro-B versus pro-B cells. Specifically, the standard deviations in spatial distances using the D_H elements (h4) as an anchor were substantially larger in pro-B cells even though the mean spatial distances were smaller. Why is there a wider distribution of configurations in pro-B versus pre-pro-B cells? We suggest that the larger

variation reflects an Igh topology in pro-B cells that is less confined or fixed as compared to pre-pro-B cells and that these differences in the range of conformations reflect the requirement for the Igh fiber in pro-B cells to be flexible in order for the large number of V_H regions that span over 2 Mbp of genomic distance to encounter D_H elements with similar probabilities. Thus we propose that both the average spatial positions as well as the spectrum of conformations change during early B cell development to permit V_H regions scattered over a large genomic region to rearrange with frequencies that seem independent of genomic location.

5.2 An Even Playing Field for All V Segments

Does the spatial proximity of V_H and $D_H J_H$ gene segments directly relate to the efficiency of DNA recombination? In Pax5, YY1 and Ikaros null mutant pro-B cells, deficiencies in V(D)J gene rearrangement correlate well with a decrease in Igh locus contraction[12, 69, 70]. However, indirect effects, perhaps modulation of the persistence length, may decrease the efficiency of V(D)J gene rearrangement and may promote locus de-contraction. Thus, other strategies will be required to directly determine whether locus contraction plays an important role in establishing a diverse antibody repertoire. Altering the size of the contour length, separating the $D_H J_H$ elements from the V_H cluster, would permit assessment of whether the efficiency of DNA recombination directly correlates with spatial separation. Despite

potential pitfalls in pursuing such an approach, it seems essential to demonstrate that physical proximity of V_H and $D_H J_H$ gene segments directly relates to the efficiency of antigen receptor locus assembly. The trilateration analysis shows the average positions of the V_H, D_H, J_H and C_H elements (Figure 4.1). Thus, the structure represents a statistical view of Igh locus topology. However, it is not merely the average positions that impact interaction frequencies since the shape of the probability distributions, peaks and tails, may substantially affect the frequency of direct encounters. In pro-B cells, the variation in spatial distances was substantially larger as compared to distributions observed in pre-pro-B cells, suggesting a change in the spectrum of conformations adopted by the Igh locus fiber. Perhaps the higher degree of variation reflects an ensemble of configurations that are less confined in pro-B cells, permitting the Igh locus to be particularly flexible at a stage in which the V_H and $D_H J_H$ elements are trying to make a connection. Are other antigen receptor loci structured in a similar fashion? Probably. The V gene segments in the Igκ, TCRβ and TCRα loci are also scattered over a vast genomic region. Apart from a V_β element located 3' of the TCRβ $D_\beta - J_\beta$ region, V_β region usage is also randomized. The TCRβ locus has been shown to undergo contraction in the DN thymocyte compartment, but upon developmental progression into DP cells, the locus becomes de-contracted[71]. Where the contraction and de-contraction occurs with the TCRβ locus remains to be resolved but it seems conceivable that the V_β regions have merged into one compartment in cells poised

to undergo TCRβ gene rearrangement, perhaps in a manner similar as described for the Igh locus. In contrast to the Igh and TCRβ loci, V_α and V_κ usage is not random[72, 7]. As aforementioned, $V_\alpha - J_\alpha$ rearrangements are ordered. This, however, does not necessarily imply that the V_α regions do not have equal probabilities of finding a J_α element. Rather, the temporal regulation of $V_\alpha J_\alpha$ rearrangement appears to be controlled, at least in part, by the elongation of non-coding RNAs as well as epigenetic marking[7]. Related epigenetic marking may underpin the non-random usage of V_κ elements. Thus the cardinal point of these findings is that antigen receptor topology permits equal opportunities for all V regions to find a DJ or J element, but epigenetic marking may promote the non-random usage of V regions, such as that described for TCRα loci.

5.3 Compartmentalization Favors Long-Range Genomic Interactions over Simple Polymers

The data described above show that the Igh locus is organized into territories, in which the V_H and $J_H - C_H$ elements are packed at high genomic density in separate compartments. These findings raise the question whether this structure mechanistically permits long-range V_H-$D_H J_H$ encounters to occur with higher frequencies than would be predicted if the chromatin fiber were not compartmentalized. We assume that close spatial proximity directly relates to the probability

of genomic encounters. Thus, we compared the experimental cumulative frequencies obtained from the spatial distance measurements directly to the cumulative frequency distributions as predicted by a 3D-random walk (see Methods for details). Interestingly, the theoretical distance distribution for a 3D-random walk approached the distance distribution observed for the D_H cluster (Figure 5.1, h4-h5). These data indicate that the probabilities for D_H elements to be in close proximity to the J_H elements approach those observed for a random walk. In contrast, for larger genomic separations, the theoretical distance distributions did not compare well with the observed spatial distance distribution, consistent with the presence of chromatin territories and spatial confinement (Figure 5.1, h4-h7, h4-h10 and h4-h11). Consequently, we conclude that it is the Igh topology that mechanistically permits long-range genomic interactions to occur in pro-B cells with relatively high frequency.

5.4 The Special Case of the $D_H J_H$ Region

As shown earlier, the $D_H J_H$ segments show identical spatial distributions in both pre-pro-B cells and in pro-B cells (Figure 5.1, h4-h5). Moreover, these distributions can also be modeled by a simple random walk of the chromatin fiber of mass density of 130 bp/nm, albeit at a much lower persistence length (50 nm) than the normal chromatin fiber (200-300 nm). However, it is possible that the

chromatin fiber might have a lower value of persistence length within a localized region. For example, naked DNA has a persistence length of 50 nm, or 150 bp, and at this value it would not be expected to fold around a histone octamer. But 147 bp of DNA coil around a histone octamer 1.7 times. This is possible due to localized reduction in the persistence length by partial neutralization of the phosphate ions by the positively charged lysine and arginine residues in the histone side-chains. Similarly, localized reduction in the persistence length might be induced at the $D_H J_H$ region to promote genomic interactions.

The observed frequency of interaction between the D_H and J_H segments (Figure 5.1, h4-h5) certainly seems to be high, and it is not a characteristic of the genomic separation between the h4 and h5 probes. This is evident from the fact that the probe-pairs h2-h1 and h2-h3, that are separated by similar genomic separations behave quite differently from the h4-h5 pair (Figure 5.2). In these two cases, the spatial distributions are not identical between pre-pro-B and pro-B cells, and the interactions are less frequent than would be predicted by a simple random walk. It is also interesting to note that in spite of a more contracted overall configuration in pro-B cells, the interaction between the constant segments (h2-h3) and between the constant segments and the 3' flank of the Igh locus (h2-h1) is rarer in pro-B cells than in pre-pro-B cells (Figure 5.2). This shows that identical spatial distributions of distances between the D_H and J_H segments in pre-pro-B pro-B cells are a function of local chromatin topology rather than their genomic

separation, and the local topology at this region might be identical in the two cell types.

5.5 Topological Constraints Manifest as Contraction or De-contraction

The Igh locus in pro-B cells shows a more contracted conformation than in pre-pro-B cells. The contraction brings the $D_H J_H$ segments in close vicinity of the V_H segments. However, the $D_H J_H$ region is de-contracted with respect to the constant segments in pro-B cells. Similarly, de-contraction is observed between constant segments and between constant segments and the 3' flank of the locus (represented by the distance distributions between probe-pairs h2-h1 and h2-h3 respectively, Figure 5.2). What is the basis of this *inversion* of behavior? We propose that the observed contraction and de-contraction is simply a manifestation of introduction of new local topological constraints in the general chromatin topology at the locus. Two topological constraints in the locus can be readily seen in the trilateration results in pro-B cells. The probes h5, h7 and h10 are in close proximity of each other in pro-B cells (Figure 4.1B). Interestingly, these probes lie in three different clusters in pre-pro-B cells (Figure 4.1A). Another constraint evident is that the probe h2 is immersed in the V_H segments in pro-B cells. This constraint on h2 can explain the apparent de-contraction towards the 3' end of the

locus in pro-B cells, as the 3' end is forced to immerse into the V_H region while the remaining constant segments tend to stay away from the V_H segments. It is intriguing that the probe h2 is located very close to the 3' regulatory region of the Igh locus. It is likely that the 3' regulator region is involved in imposition of this constraint and contributes to locus contraction. This model also suggests that the intronic enhancer, located within the probe h4 might not be involved in locus contraction since it does not appear to be constrained. Hence contraction or de-contraction can be explained by imposition of topological constraints.

5.6 Putative DNA-Binding Factors Regulating Chromatin Topology

Establishment and maintenance of chromatin topology would require proteins that can act as tethers for the bases of chromatin loops. Other factors may be required for imposing specific topological constraints. Both types of factors should play an important role in allowing V(D)J recombination in pro-B cells, as loss of function of any of these factors would perturb long-range genomic interactions at the locus. Where are those tethers located in the antigen receptor loci? Putative tethers that span the entire Igh locus have been identified. Prominent among these are YY1 and CTCF.

YY1 is a zinc-finger protein that is evolutionary conserved from Drosophila

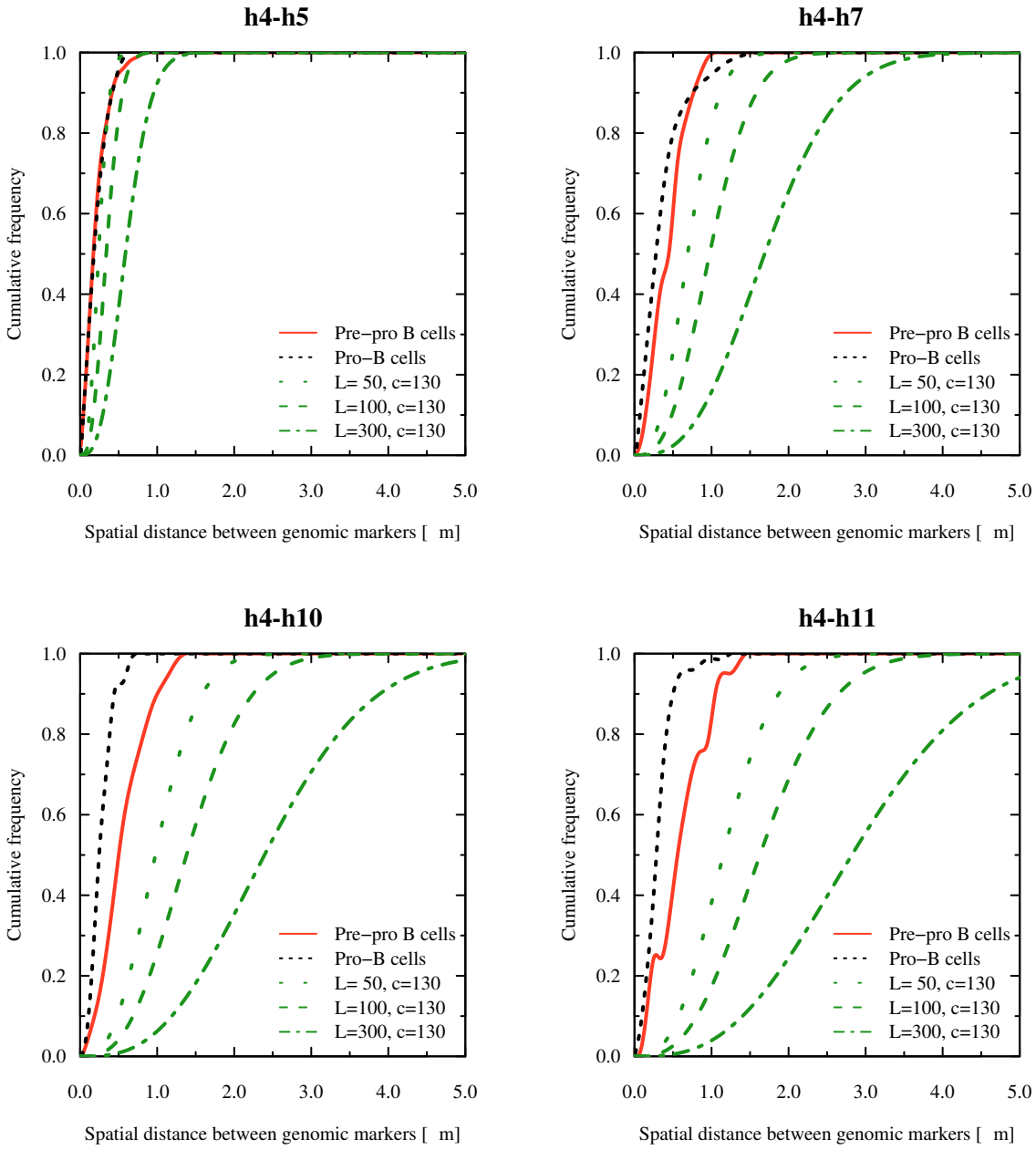


Figure 5.1: Probabilities of Long-Range Interactions in Pre-pro-B and Pro-B Cells Compared to Those Predicted by a 3D-Random Walk
Cumulative frequencies were obtained by accumulating the frequency values corresponding to the spatial distances in intervals of 100 nm using the $D_H J_H$ elements (probe h4) as an anchor. Cumulative frequency distributions for the random walk were determined for different persistence lengths (see Methods). The cumulative frequency distribution for the spatial distance between h4 and h5 is similar for pre-pro-B cells, pro-B cells and for a random walk with a persistence length of 50 nm and chromatin density of 130 bp/nm.

melanogaster to humans. $V_H - D_H J_H$ rearrangement is severely perturbed in YY1-deficient pro-B cells affecting mainly the distal V_H gene segments[69]. Interestingly, the Igh locus is de-contracted in YY1-deficient pro-B cells, raising the possibility that YY1 acts to modulate Igh locus topology to promote distal $V_H - D_H J_H$ rearrangement[69].

How does YY1 promote Igh locus long-range chromatin contraction? The precise mechanism is unknown but YY1 is of interest since it has been demonstrated to interact with CTCF. CTCF is a factor previously shown to promote looping within the β-globin locus[73](Donohoe et al., 2007). In addition to interacting with YY1, recent genome-wide binding studies have revealed that CTCF also interacts with cohesins[74]. These data are intriguing since during DNA replication, the cohesins form a ring-like structure, to surround and stabilize the sister chromatid strands. Cohesins and CTCF proteins may similarly act to stabilize loop formation. Is CTCF a reasonable candidate to act as a bridging factor in antigen receptor loci? The binding pattern of CTCF in the Igh locus in pro-B cells is striking[75]. A total of fifty-three CTCF binding sites span the entire V_H region cluster but are absent within the D_H-J_H and C_H gene segments[75]. A few CTCF sites are located immediately upstream of the $D_H J_H$ cluster and may form a boundary between the V_H cluster and the $D_H J_H$ gene segments. The binding of CTCF to sites present in the Igh locus is not pro-B cell specific and it is unlikely by itself to account for the differences observed in pre-pro-B and pro-B Igh

locus topology. However, Rad21, a component of the cohesin complex was found to preferentially bind with CTCF in pro-B cells as compared to thymocytes and pre-B cells[75, 74]. These data raise the possibility that YY1, CTCF and cohesins act as an anchoring complex to promote the formation of loops. Clearly, it will be critical to deplete CTCF activity in developing pro-B cells and assess whether and how CTCF plays a role in Igh locus topology.

Other candidates for establishing Igh locus topology have also emerged. Pax5-deficient pro-B cells showed decreased Igh locus contraction, again involving the distal V_H regions[12, 13]. How Pax5 promotes contraction remains to be established but it must act with other factors in pro-B cells to modulate Igh locus topology, since in pre-B cells Pax5 abundance is high whereas Igh loci become de-contracted. Ikaros is yet another player involved in modulating Igh locus topology. Ikaros not only activates Rag expression and Igh locus accessibility, but also appears to promote Igh locus contraction[70].

A protein involved in DNA repair, 53BP1 has recently been implicated in modulating antigen receptor topology. 53BP1 initially targets double stranded (ds) DNA breaks upon interacting with H2AX and to methylated histone H4K20 or possibly H3K79, which are constitutive histone marks exposed upon ds DNA break formation. 53BP1 was shown to promote TCRα VJ gene rearrangement[76]. Interestingly, the TCRα locus was found to be in an extended state in 53BP1-ablated as compared to wild-type thymocytes. These data inspire the question as

to how 53BP1 permits DNA elements, separated by large genomic distances, to interact with relatively high probability? Two fundamentally different models have been suggested to underpin 53BP1 activity. 53BP1 may accumulate at ds DNA breaks, and upon homo-oligomerization act as a bridging factor[76]. Alternatively, 53BP1 may directly modulate chromatin dynamics possibly by affecting the persistence length of the chromatin fiber[77].

Thus, a few molecular components have been identified that modulate antigen receptor locus topology. The candidates identified likely represent only the tip of the iceberg. The challenge will be now to identify others and to find out how they act together to promote looping and anchoring, how they modulate the dynamics of loop formation, chromatin assembly and how they promote encounters between V and DJ or J gene segments with the appropriate frequencies.

5.7 Free Polymer Chain Dynamics at $D_H J_H$ segments

Interestingly, CTCF and YY1 binding sites are scattered throughout the V_H region repertoire but the $D_H J_H$ region lacks such elements. Why is the D_H-J_H region devoid of putative loop-attachment sites? We suggest that the D_H-J_H region undergoes free polymer chain dynamics, whereas the V_H repertoire is spatially confined into clusters of loops. This notion is also supported by the comparison of

experimentally derived cumulative frequency spatial distributions and theoretical distributions predicted by random walk behavior, for D_H-J_H and V_H-D_HJ_H gene segments (Figure 5.1, h4-h5). Furthermore, when the locus is scanned for matrix attachment regions and scaffold attachment regions (MAR/SAR), we find that these sites are also scattered throughout the V_H region but absent in the D_HJ_H region (data not shown). Thus, we suggest that whereas the V_H gene segments undergo random walk dynamics, they are doing so within their spatial confinement, whereas D_HJ_H elements are wandering freely, trying to make a connection. This can also explain why D_HJ_H rearrangements are readily detectable in pro-B cells deficient for YY1[69], Pax5[12] or Ikaros[70], or even modestly detectable in T cells.

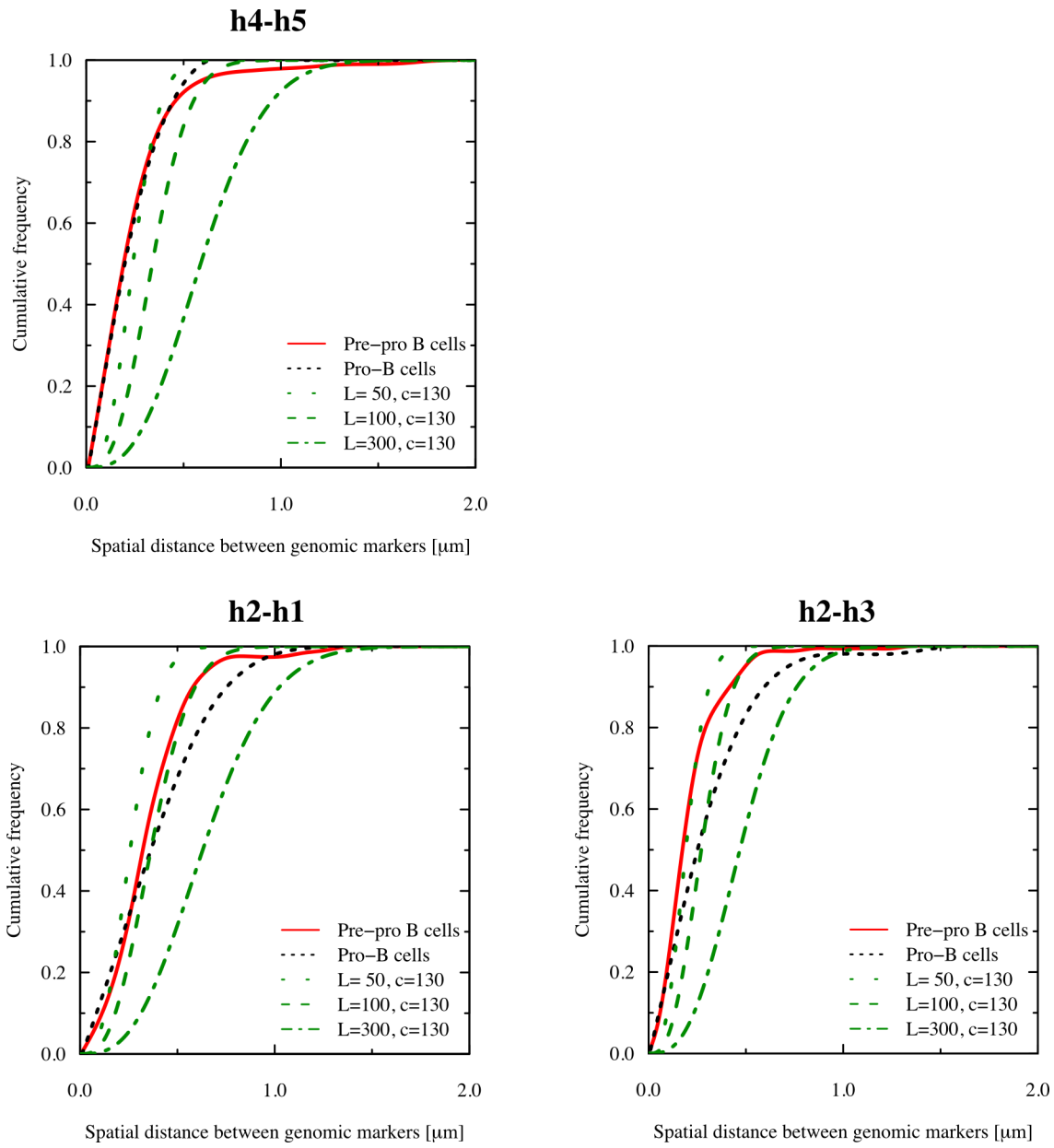


Figure 5.2: Cumulative Frequency Distributions for the Spatial Distances of the Constant Region

Cumulative frequencies were obtained by accumulating the frequency values corresponding to the spatial distances between the probe-pairs h2-h1 and h2-h3, in intervals of 100 nm. Experimental observations are shown for the pre-pro-B cell and pro-B cell data. Cumulative frequency distributions predicted by the random walk were computationally determined for different persistence lengths (see Methods).

The data also bring into question whether the topology described here is restricted to that of the Igh locus, since the Igh locus undergoes genomic rearrangement and has special requirements for proximity. We consider this unlikely since the spatial distances also plateau as a function of genomic separation in the centromeric direction away from the Igh locus. This region is not undergoing long-range DNA recombination. Why do the spatial distances as a function of genomic separation plateau in a region that is not undergoing antigen receptor assembly? We propose that the chromatin fiber in non-rearranging genomic regions is folded into multi-loop-containing compartments to promote high-density packing and to permit long-range genomic interactions[56]. For example, enhancer elements act over large genomic distances[51]. Folding of the chromatin fiber into bundles of loops allows such elements to be in close spatial proximity similar as described here for the Igh locus. Thus, antigen receptor genes may simply have utilized this topology to allow a large number of V_H elements to encounter $D_H J_H$ or J_H elements with relatively high and similar frequencies. Similarly, we propose that enhancer and promoter elements interact over large genomic distances since the organization of the genome into clusters of loops allows genomic elements that are separated by large genomic distances to be located in close spatial proximity.

Acknowledgements

This chapter is partly reprinted from the papers *"The 3D-Structure of the Immunoglobulin Heavy Chain Locus: Implications for Long-Range Genomic Interactions"*, *Cell*, 133(2), 2008 and *"Chromatin architecture and the generation of antigen receptor diversity"*, Cell, 138(3), 2009. I have spent years working in the dark along with Menno van Zelm and Mandy Peak to generate the data for this project. It would not have been possible to survive, as we call it, the 'dungeon' or the 'beach' without them. Roy Riblet once again helped us with the map and small probes for the locus. Tobias Knoch was the backbone of the computer simulations for RW/GL and MLS models in this project, and it was his work that inspired us to compare the data to structural models. Steve Cutchin and Amit Chourasia, at the San Diego Supercomputer Center helped us in the trilateration study and made some wonderful 3D movies and figures. I thank Jacques van Dongen, Frank Grosveld and Douglas Forbes for their help with the manuscript. I thank Pete Carlton and John Sedat (UCSF) for their generous help with using their SIM system. Terry Hwa provided us his expert advice in comparing our data to random walk models and inferring compartmentalization.

Bibliography

1. Dreyer, W. J. and Bennett, J. C. The molecular basis of antibody formation: a paradox. *Proc Natl Acad Sci U S A* **54**(3), 864–869, Sep (1965).

2. Brack, C., Hirama, M., Lenhard-Schuller, R., and Tonegawa, S. A complete immunoglobulin gene is created by somatic recombination. *Cell* **15**(1), 1–14, Sep (1978).

3. Seidman, J. G., Leder, A., Edgell, M. H., Polsky, F., Tilghman, S. M., Tiemeier, D. C., and Leder, P. Multiple related immunoglobulin variable-region genes identified by cloning and sequence analysis. *Proc Natl Acad Sci U S A* **75**(8), 3881–3885, Aug (1978).

4. Weigert, M., Gatmaitan, L., Loh, E., Schilling, J., and Hood, L. Rearrangement of genetic information may produce immunoglobulin diversity. *Nature* **276**(5690), 785–790, Dec (1978).

5. Cedar, H. and Bergman, Y. Choreography of ig allelic exclusion. *Curr Opin Immunol* **20**(3), 308–317, Jun (2008).

6. Jung, D., Giallourakis, C., Mostoslavsky, R., and Alt, F. W. Mechanism and control of v(d)j recombination at the immunoglobulin heavy chain locus. *Annu Rev Immunol* **24**, 541–570 (2006).

7. Krangel, M. S. T cell development: better living through chromatin. *Nat Immunol* **8**(7), 687–694, Jul (2007).

8. Schneider, R. and Grosschedl, R. Dynamics and interplay of nuclear architecture, genome organization, and gene expression. *Genes Dev* **21**(23), 3027–3043, Dec (2007).

9. Akhtar, A. and Gasser, S. M. The nuclear envelope and transcriptional control. *Nat Rev Genet* **8**(7), 507–517, Jul (2007).

10. Ragoczy, T., Bender, M. A., Telling, A., Byron, R., and Groudine, M. The

locus control region is required for association of the murine beta-globin locus with engaged transcription factories during erythroid maturation. *Genes Dev* **20**(11), 1447–1457, Jun (2006).

11. Kosak, S. T., Skok, J. A., Medina, K. L., Riblet, R., Le Beau, M. M., Fisher, A. G., and Singh, H. Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science* **296**(5565), 158–162, Apr (2002).

12. Fuxa, M., Skok, J., Souabni, A., Salvagiotto, G., Roldan, E., and Busslinger, M. Pax5 induces v-to-dj rearrangements and locus contraction of the immunoglobulin heavy-chain gene. *Genes Dev* **18**(4), 411–422, Feb (2004).

13. Roldan, E., Fuxa, M., Chong, W., Martinez, D., Novatchkova, M., Busslinger, M., and Skok, J. A. Locus 'decontraction' and centromeric recruitment contribute to allelic exclusion of the immunoglobulin heavy-chain gene. *Nat Immunol* **6**(1), 31–41, Jan (2005).

14. Sayegh, C. E., Jhunjhunwala, S., Riblet, R., and Murre, C. Visualization of looping involving the immunoglobulin heavy-chain locus in developing b cells. *Genes Dev* **19**(3), 322–327, Feb (2005).

15. Alt, F. W., Yancopoulos, G. D., Blackwell, T. K., Wood, C., Thomas, E., Boss, M., Coffman, R., Rosenberg, N., Tonegawa, S., and Baltimore, D. Ordered rearrangement of immunoglobulin heavy chain variable region segments. *EMBO J* **3**(6), 1209–1219, Jun (1984).

16. Malynn, B. A., Yancopoulos, G. D., Barth, J. E., Bona, C. A., and Alt, F. W. Biased expression of jh-proximal vh genes occurs in the newly generated repertoire of neonatal and adult mice. *J Exp Med* **171**(3), 843–859, Mar (1990).

17. Yancopoulos, G. D. and Alt, F. W. Developmentally controlled and tissue-specific expression of unrearranged vh gene segments. *Cell* **40**(2), 271–281, Feb (1985).

18. Felsenfeld, G., Burgess-Beusse, B., Farrell, C., Gaszner, M., Ghirlando, R., Huang, S., Jin, C., Litt, M., Magdinier, F., Mutskov, V., Nakatani, Y., Tagami, H., West, A., and Yusufzai, T. Chromatin boundaries and chromatin domains. *Cold Spring Harb Symp Quant Biol* **69**, 245–250 (2004).

19. Bates, J. G., Cado, D., Nolla, H., and Schlissel, M. S. Chromosomal position of a vh gene segment determines its activation and inactivation as a substrate for v(d)j recombination. *J Exp Med* **204**(13), 3247–3256, Dec (2007).

20. Sieh, P. and Chen, J. Distinct control of the frequency and allelic exclusion

of the v beta gene rearrangement at the tcr beta locus. *J Immunol* **167**(4), 2121–2129, Aug (2001).

21. Schalch, T., Duda, S., Sargent, D. F., and Richmond, T. J. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* **436**(7047), 138–141, Jul (2005).

22. Sedat, J. and Manuelidis, L. A direct approach to the structure of eukaryotic chromosomes. *Cold Spring Harb Symp Quant Biol* **42 Pt 1**, 331–350 (1978).

23. Paulson, J. R. and Laemmli, U. K. The structure of histone-depleted metaphase chromosomes. *Cell* **12**(3), 817–828, Nov (1977).

24. Rattner, J. B. and Lin, C. C. Radial loops and helical coils coexist in metaphase chromosomes. *Cell* **42**(1), 291–296, Aug (1985).

25. Pienta, K. J. and 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).

26. Belmont, A. S. and Bruce, K. Visualization of g1 chromosomes: a folded, twisted, supercoiled chromonema model of interphase chromatid structure. *J Cell Biol* **127**(2), 287–302, Oct (1994).

27. Langowski, J. and Heermann, D. W. Computational modeling of the chromatin fiber. *Semin Cell Dev Biol* **18**(5), 659–667, Oct (2007).

28. de Gennes, P. *Scaling concepts in polymer physics*. Cornel University Press, (1979).

29. Bystricky, K., Heun, P., Gehlen, L., Langowski, J., and Gasser, S. M. Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. *Proc Natl Acad Sci U S A* **101**(47), 16495–16500, Nov (2004).

30. Sachs, R. K., van den Engh, G., Trask, B., Yokota, H., and Hearst, J. E. A random-walk/giant-loop model for interphase chromosomes. *Proc Natl Acad Sci U S A* **92**(7), 2710–2714, Mar (1995).

31. Warrington, J. A. and 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, Nov (1994).

32. Trask, B. J., Allen, S., Massa, H., Fertitta, A., Sachs, R., van den Engh, G., and

Wu, M. Studies of metaphase and interphase chromosomes using fluorescence in situ hybridization. *Cold Spring Harb Symp Quant Biol* **58**, 767–775 (1993).

33. Munkel, C. and Langowski, J. Chromosome structure described by a polymer model. *Phys. Rev.* **E 57**(5B), 5888–5896 (1998).

34. Knoch, T. A. *Approaching the three-dimensional organization of the human genome: structural-, scaling- and dynamic-properties in the simulation of interphase chromosomes and cell nuclei, long-range correlations in complete genomes, in vivo quantification of the chromatin distribution, construct conversions in simultaneous co-transfections.* PhD thesis, Ruperto-Carola University, Heidelberg, Germany, TAK Press, Tobias A. Knoch, Mannheim, Germany, (2002). ISBN 3-00-009959-X, ISBN 3-00-009960-3.

35. Bohn, M., Heermann, D. W., and van Driel, R. Random loop model for long polymers. *Phys Rev E Stat Nonlin Soft Matter Phys* **76**(5 Pt 1), 051805, Nov (2007).

36. Mainville, C. A., Sheehan, K. M., Klaman, L. D., Giorgetti, C. A., Press, J. L., and Brodeur, P. H. Deletional mapping of fifteen mouse vh gene families reveals a common organization for three igh haplotypes. *J Immunol* **156**(3), 1038–1046, Feb (1996).

37. Su, I.-H., Basavaraj, A., Krutchinsky, A. N., Hobert, O., Ullrich, A., Chait, B. T., and Tarakhovsky, A. Ezh2 controls b cell development through histone h3 methylation and igh rearrangement. *Nat Immunol* **4**(2), 124–131, Feb (2003).

38. Xu, C.-R., Schaffer, L., Head, S. R., and Feeney, A. J. Reciprocal patterns of methylation of h3k36 and h3k27 on proximal vs. distal igvh genes are modulated by il-7 and pax5. *Proc Natl Acad Sci U S A* **105**(25), 8685–8690, Jun (2008).

39. Baxter, J., Merkenschlager, M., and Fisher, A. G. Nuclear organisation and gene expression. *Curr Opin Cell Biol* **14**(3), 372–376, Jun (2002).

40. Eastman, Q. M., Leu, T. M., and Schatz, D. G. Initiation of v(d)j recombination in vitro obeying the 12/23 rule. *Nature* **380**(6569), 85–88, Mar (1996).

41. Steen, S. B., Gomelsky, L., and Roth, D. B. The 12/23 rule is enforced at the cleavage step of v(d)j recombination in vivo. *Genes Cells* **1**(6), 543–553, Jun (1996).

42. Ostermeier, G. C., Liu, Z., Martins, R. P., Bharadwaj, R. R., Ellis, J., Draghici, S., and Krawetz, S. A. Nuclear matrix association of the human beta-globin

locus utilizing a novel approach to quantitative real-time pcr. *Nucleic Acids Res* **31**(12), 3257–3266, Jun (2003).

43. Yancopoulos, G. D., Desiderio, S. V., Paskind, M., Kearney, J. F., Baltimore, D., and Alt, F. W. Preferential utilization of the most jh-proximal vh gene segments in pre-b-cell lines. *Nature* **311**(5988), 727–733, Oct (1984).

44. Yancopoulos, G. D. and Alt, F. W. Regulation of the assembly and expression of variable-region genes. *Annu Rev Immunol* **4**, 339–368 (1986).

45. Chowdhury, D. and Sen, R. Stepwise activation of the immunoglobulin mu heavy chain gene locus. *EMBO J* **20**(22), 6394–6403, Nov (2001).

46. Spicuglia, S., Kumar, S., Yeh, J.-H., Vachez, E., Chasson, L., Gorbatch, S., Cautres, J., and Ferrier, P. Promoter activation by enhancer-dependent and -independent loading of activator and coactivator complexes. *Mol Cell* **10**(6), 1479–1487, Dec (2002).

47. Bergman, Y. and Cedar, H. A stepwise epigenetic process controls immunoglobulin allelic exclusion. *Nat Rev Immunol* **4**(10), 753–761, Oct (2004).

48. Goebel, P., Montalbano, A., Ayers, N., Kompfner, E., Dickinson, L., Webb, C. F., and Feeney, A. J. High frequency of matrix attachment regions and cut-like protein x/ccaat-displacement protein and b cell regulator of igh transcription binding sites flanking ig v region genes. *J Immunol* **169**(5), 2477–2487, Sep (2002).

49. Herendeen, D. R., Kassavetis, G. A., and Geiduschek, E. P. A transcriptional enhancer whose function imposes a requirement that proteins track along dna. *Science* **256**(5061), 1298–1303, May (1992).

50. Rombel, I., North, A., Hwang, I., Wyman, C., and Kustu, S. The bacterial enhancer-binding protein ntrc as a molecular machine. *Cold Spring Harb Symp Quant Biol* **63**, 157–166 (1998).

51. Banerji, J., Rusconi, S., and Schaffner, W. Expression of a beta-globin gene is enhanced by remote sv40 dna sequences. *Cell* **27**(2 Pt 1), 299–308, Dec (1981).

52. Carter, D., Chakalova, L., Osborne, C. S., Dai, Y.-f., and Fraser, P. Long-range chromatin regulatory interactions in vivo. *Nat Genet* **32**(4), 623–626, Dec (2002).

53. Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F., and de Laat, W. Looping

and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell* **10**(6), 1453–1465, Dec (2002).

54. Eivazova, E. R. and Aune, T. M. Dynamic alterations in the conformation of the ifng gene region during t helper cell differentiation. *Proc Natl Acad Sci U S A* **101**(1), 251–256, Jan (2004).

55. Spilianakis, C. G. and Flavell, R. A. Long-range intrachromosomal interactions in the t helper type 2 cytokine locus. *Nat Immunol* **5**(10), 1017–1027, Oct (2004).

56. Cremer, T. and Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* **2**(4), 292–301, Apr (2001).

57. Trask, B. J., Massa, H., Kenwrick, S., and 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, Jan (1991).

58. Knoch, T. A., Münkel, C., and Langowski, J. Three-dimensional organization of chromosome territories in the human interphase nucleus. In High Performance Computing in Science and Engineering, Krause, E. and Jäger, W., editors, 229–238 (Springer, Berlin-Heidelberg-New York, 1999). ISBN 3-540-66504-8.

59. Ikawa, T., Kawamoto, H., Wright, L. Y. T., and Murre, C. Long-term cultured e2a-deficient hematopoietic progenitor cells are pluripotent. *Immunity* **20**(3), 349–360, Mar (2004).

60. Shinkai, Y., Rathbun, G., Lam, K. P., Oltz, E. M., Stewart, V., Mendelsohn, M., Charron, J., Datta, M., Young, F., and Stall, A. M. Rag-2-deficient mice lack mature lymphocytes owing to inability to initiate v(d)j rearrangement. *Cell* **68**(5), 855–867, Mar (1992).

61. Kratky, O. and Porod, G. Röntgenuntersuchung gelöster fadenmoleküle. *Rec. Trav. Chim Pays-Bas* **68**, 1106–1133 (1949).

62. Chen, Y., Ding, F., and Dokholyan, N. V. Fidelity of the protein structure reconstruction from inter-residue proximity constraints. *J Phys Chem B* **111**(25), 7432–7438, Jun (2007).

63. Love, V. A., Lugo, G., Merz, D., and Feeney, A. J. Individual v(h) promoters vary in strength, but the frequency of rearrangement of those v(h) genes

does not correlate with promoter strength nor enhancer-independence. *Mol Immunol* **37**(1-2), 29–39, Jan-Feb (2000).

64. Gu, H., Tarlinton, D., Muller, W., Rajewsky, K., and Forster, I. Most peripheral b cells in mice are ligand selected. *J Exp Med* **173**(6), 1357–1371, Jun (1991).

65. Schermelleh, L., Carlton, P. M., Haase, S., Shao, L., Winoto, L., Kner, P., Burke, B., Cardoso, M. C., Agard, D. A., Gustafsson, M. G. L., Leonhardt, H., and Sedat, J. W. Subdiffraction multicolor imaging of the nuclear periphery with 3d structured illumination microscopy. *Science* **320**(5881), 1332–1336, Jun (2008).

66. Klar, T. A., Engel, E., and Hell, S. W. Breaking abbe's diffraction resolution limit in fluorescence microscopy with stimulated emission depletion beams of various shapes. *Phys Rev E Stat Nonlin Soft Matter Phys* **64**(6 Pt 2), 066613, Dec (2001).

67. Betzig, E., Patterson, G. H., Sougrat, R., Lindwasser, O. W., Olenych, S., Bonifacino, J. S., Davidson, M. W., Lippincott-Schwartz, J., and Hess, H. F. Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**(5793), 1642–1645, Sep (2006).

68. Rust, M. J., Bates, M., and Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (storm). *Nat Methods* **3**(10), 793–795, Oct (2006).

69. Liu, H., Schmidt-Supprian, M., Shi, Y., Hobeika, E., Barteneva, N., Jumaa, H., Pelanda, R., Reth, M., Skok, J., Rajewsky, K., and Shi, Y. Yin yang 1 is a critical regulator of b-cell development. *Genes Dev* **21**(10), 1179–1189, May (2007).

70. Reynaud, D., Demarco, I. A., Reddy, K. L., Schjerven, H., Bertolino, E., Chen, Z., Smale, S. T., Winandy, S., and Singh, H. Regulation of b cell fate commitment and immunoglobulin heavy-chain gene rearrangements by ikaros. *Nat Immunol* **9**(8), 927–936 (2008).

71. Skok, J. A., Gisler, R., Novatchkova, M., Farmer, D., de Laat, W., and Busslinger, M. Reversible contraction by looping of the tcra and tcrb loci in rearranging thymocytes. *Nat Immunol* **8**(4), 378–387, Apr (2007).

72. Feeney, A. J., Lugo, G., and Escuro, G. Human cord blood kappa repertoire. *J Immunol* **158**(8), 3761–3768, Apr (1997).

73. Splinter, E., Heath, H., Kooren, J., Palstra, R.-J., Klous, P., Grosveld, F.,

Galjart, N., and de Laat, W. Ctcf mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev* **20**(17), 2349–2354, Sep (2006).

74. Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H. C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., Cobb, B. S., Yokomori, K., Dillon, N., Aragon, L., Fisher, A. G., and Merkenschlager, M. Cohesins functionally associate with ctcf on mammalian chromosome arms. *Cell* **132**(3), 422–433, Feb (2008).

75. Degner, S. C., Wong, T. P., Jankevicius, G., and Feeney, A. J. Cutting edge: developmental stage-specific recruitment of cohesin to ctcf sites throughout immunoglobulin loci during b lymphocyte development. *J Immunol* **182**(1), 44–48, Jan (2009).

76. Difilippantonio, S., Gapud, E., Wong, N., Huang, C.-Y., Mahowald, G., Chen, H. T., Kruhlak, M. J., Callen, E., Livak, F., Nussenzweig, M. C., Sleckman, B. P., and Nussenzweig, A. 53bp1 facilitates long-range dna end-joining during v(d)j recombination. *Nature* **456**(7221), 529–533, Nov (2008).

77. Dimitrova, N., Chen, Y.-C. M., Spector, D. L., and de Lange, T. 53bp1 promotes non-homologous end joining of telomeres by increasing chromatin mobility. *Nature* **456**(7221), 524–528, Nov (2008).

78. Francke, U. and Oliver, N. Quantitative analysis of high-resolution trypsin-giemsa bands on human prometaphase chromosomes. *Hum Genet* **45**(2), 137–165, Dec (1978).

79. Yunis, E., Leibovici, M., and Quintero, L. Ring (15) chromosome. *Hum Genet* **57**(2), 207–209 (1981).

Appendix A

Methods

A.1 Mice and Cell Culture

All mice used were maintained in a C57BL/6 background. $Rag2^{-/-}$ pro-B cells and $E2A^{-/-}$ hematopoietic progenitors were isolated and grown as described previously (Ikawa et al., 2004; Sayegh et al., 2005). The E2A-deficient and Rag2-deficient targeted mutations were originally engineered in R9 ES cells and subsequently backcrossed on a C57BL/6 background. Since the genomic organization of the Igh locus differs substantially between C57BL/6 and 129 mice strains, genomic DNA was isolated from E2A deficient pre-pro-B and RAG2-deficient pro-B cells to determine their origins. PCR analysis at 4 polymorphic sites, localized throughout the Igh locus, confirmed that in both cell types the Igh locus is derived from C57BL/6 (R.R., data not shown).

A.2 Cloning and Labeling of 10-kb Probes

To identify unique DNA sequences within the Igh locus, a percentage identity plot of the Igh locus was generated. Eleven unique 10 Kbp DNA segments, i.e. regions that were not duplicated within the Igh locus and showed minimal genomic repeats, were identified using a percentage identity plot. PCR primers were designed for these regions and long-range PCR (Eppendorf) was used to amplify the 10-kb stretches. PCR products were cloned into the pGEMTEasy vector system (Promega). The sequences of the clones were confirmed by restriction mapping and by DNA sequencing. DNA probes were labeled with aminoallyl-dUTP (ARES labeling kits, Invitrogen) by nick-translation (Roche). One μg of DNA was labeled in a 20 μl reaction for 3 hours and 45 minutes. Nick-translated 10 Kbp products were size selected on a 0.6% agarose gel and 10-500bp fragments were purified using a Qiagen gel extraction kits. Aminoallyl-modified DNA was fluorescently labeled with succinimidyl ester derivatives of Alexa fluorochromes (Alexa 488, 594 or 647).

A.3 The Self Avoiding Chain and Kratky-Porod Analytical Description of a Polymer

The probability density for the spatial distance **R** between between genomic

markers can be determined as:

$$P(\mathbf{R}, a) = \left(\frac{3c}{2\pi da}\right)^{3/2} e^{-\frac{3c\mathbf{R}^2}{2da}}$$

'a' is the Kuhn length of DNA [nm], 'c' is the DNA mass density [bp/nm] (Grosberg and Khokhlov, 1997) and 'd' is the number of dimensions in which the distances have been measured (here, d=3). The cumulative frequency distribution F(R) is given by:

$$\int_0^R \left(\frac{3c}{2\pi da}\right)^{3/2} e^{-\frac{3c\mathbf{R}^2}{2da}} 4\pi R^2 dR$$

The distribution of spatial distances as a function of genomic separation for the self avoiding chain are described as follows:

$$P(R, N) = N^{-\nu d} f(R/N^{\nu})$$

$$f(x) = x^{g} \ for \ x << 1$$

$$f(x) = e^{-x^{\delta}} \ for \ x >> 1$$

R : Spatial distances separating genomic markers. N : Genomic separation between the markers. P : Frequency value of the spatial distance R for markers separated by N base-pairs. d : Number of dimension in which the distances have been measured. Here, d=3. $\nu = 0.6$ for self-avoiding random walks. For a self avoiding random walk, the constants g and δ are as follows: $g = 0.67 \pm 0.3 \ and \ \delta = 2.5$

The Kratky-Porod description of a semi-flexible polymer was used to

compare the experimental data to that of the worm-like chain[61]:

$$< R^2 >= 2 \times L^2 \times \left(\frac{d}{cL} - 1 + e^{-\frac{d}{cL}} \right)$$

where R is the physical distance separating two genetic markers [nm], d is the genomic separation between genetic markers [bp], c is the DNA mass density [bp/nm], and L is the persistence length [nm]. The cumulative frequencies were determined for different persistence lengths and compared with the experimental data. For theoretic spatial distances and frequency distributions a mass density of 130 bp/nm was used[29]. The experimental frequency distributions were generated after binning 100 nm spatial distances.

A.4 High-Resolution 3D-Structure Preserving Fluorescence in situ Hybridization

40 μl of a 1 x 106 cells/ml suspension of cells was directly attached to coverslips. Cells were fixed in 4% paraformaldehyde for 10 min. at room temperature and carefully placed on a coverslip. Next, cells were incubated with 0.1M Tris-Cl pH7.2 for 10 min. at room temperature and the coverslips were washed with 1X Phosphate buffer saine (PBS). Cells were permeabilized for 10 min. at room temperature with PBS plus 0.1% triton x-100 and 0.1% saponin and then incubated for 20 minutes with 20% glycerol/1X PBS. Subsequently, coverslips

were immersed in liquid nitrogen and thawed for three consecutive times. Cells were washed once in PBS, then treated for 30 minutes with a 0.1 N HCl solution at room temperature, washed once in PBS, and incubated with 100 μg/ml DNase-free RNase in 3% bovine serum albumin (BSA) / 0.1% triton X-100 for 1 hour at 37 oC. Cells were again permeabilized in PBS, 0.5% triton X-100, 0.5% saponin solution for 30 minutes at room temperature. Cells were then washed once in PBS. Nuclear DNA was denatured by incubating coverslips for 2 minutes and 30 seconds at 73 oC in 2X SSC, 70% formamide solution followed by an incubation of 1 minute in 2X SSC plus 50% formamide. Excess liquid was removed and 10 μl of hybridization cocktail was added to each of the coverslips. Coverslips were mounted, sealed with rubber-cement, and incubated overnight at 37 oC. The hybridization solution contained 400 ng of labeled BAC probe, 40 ng each of labeled 10 Kbp probes, 4 μg of mouse Cot-1 DNA, 1 μg of sheared salmon-sperm DNA dissolved in 50% formamide, 4XSSC and 20% dextran sulfate. The probes were denatured at 75^{o}C for 5 minutes and chilled on ice prior to incubation with coverslips. On the following day, coverslips were removed and washed once in 2X SSC and 50% formamide for 15 minutes and 3 times in 2X SSC for a 5 minute period at 37^{o}C with low agitation (100 rpm). Cells were washed once with PBS, excess PBS was removed and coverslips were mounted on slides with 12 μl of Prolong gold anti-fade reagent (Invitrogen) and 400 ng/ml DAPI.

A.5 Image Acquisitions, Distance Calculations, Calibration and Statistics

Images were captured with a DeltaVision epifluorescent deconvolution microscope system (Applied Precision, Inc.) located at the UCSD cancer center microscope facility. Using a 100x (NA 1.4) lens, images of approximately 40 serial optical sections spaced by 0.2 μm were acquired from experiments involving BAC DNA probes only, and 80 serial optical sections spaced by 0.1 μm were acquired from experiments involving 10 Kbp DNA probes. To achieve high precision and analytical correction of the optics the point spread function (PSF) with its full width at half the maximum (FWHM) and the chromatic shift between the fluorophores was measured with fluorescently labeled tetraspeck beads (Invitrogen, 0.2 μm) at 0.1 μm z-intervals. Each bead was uniformly stained with a mixture of fluorescent dyes that emit in the blue, green, orange and dark red wavelengths. For chromatic shift correction, determination of the resolution equivalent and the experimental measurements, the image stacks were deconvolved and optical sections were merged to produce 3D-pictures using SoftWorx software (Applied Precision, Inc) on a Silicon Graphics Octane workstation. The 3D-coordinates of the center of mass of each probe were obtained. The chromatic shift was determined by subtraction of the individual center of mass coordinates of the same bead observed under different colors. To establish the smallest measurable distance

in the analysis described here, the "resolution equivalent" RE was determined by correcting for chromatic shifts the positions of the beads used to measure the chromatic shifts. RE (and also the spatial distance d) was then calculated according to the equation:

$$RE\ or\ d = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2 + (z_a - z_b)^2}$$

with the coordinates x,y, and z of objects a and b. Then the average was taken (Table 3.3). To determine the experimental spatial distances, the initial spatial positions of signals were corrected for the chromatic shift. The average DNA compaction based on the linear contour length of 340 nm for 1.0 Kbp was calculated from

$$\frac{340[nm]}{1[Kbp]} \cdot \frac{genomic\ length[Kbp]}{< spatial\ distance[nm] >}$$

To determine the smallest measurable distance that we could observe using spectral precision epifluorescence microscopy, we measured the point spread function (PSF) and the resolution equivalent. The full width at half maximum values was determined for each fluorescent color from the PSF (lateral 380 to 542 nm, axial 1400 to 1700nm; Table S1; see Methods Section). The approach required correction for chromatic shifts for each of the fluorophore combinations (chromatic shifts: lateral -14 to 53nm, axial -159 to 26 nm; Table S2). From this analysis the smallest measurable distances, the RE values, were determined, ranging from 35 to 47 nm, for each of the fluorophore combinations (Table S2), allowing the

measurement of spatial distances with a 3D-resolution greater than 50 nm.

A.6 Computer Simulations

Using a two step combined Monte Carlo and Brownian Dynamics method, the Random-Walk/Giant-Loop (RW/GL) model and the Multi-Loop Subcompartment (MLS) model were simulated for human interphase chromosome 15. Assuming a flexible polymer chain, the chromatin fiber was split into ∼3,300 segments of 300 nm (∼31 Kbp). To each segment a harmonic stretching potential and between two segments a harmonic bending potential were assigned. To avoid self-crossing of the polymer chain, a short ranged excluded volume potential was introduced, whose potential barrier could be changed to facilitate chain dis-entanglement to speed up simulations. In vivo this is mediated by Topoisomerase-II especially during chromosome de- or condensation. The simulation of single chromosomes necessitated placement into an embedding potential simulating the surrounding nuclear chromosomes. In the RW/GL model loops of 5.0, 4.0, 3.0, 2.0, 1.0, 0.504 and 0.252 Mbp were used and connected by a chromatin linker whose length was adjusted such that the global territory behavior yielded comparable results as in the MLS model. In the MLS model the loop size was 126 Kbp with linker sizes of 63, 126, 189 and 252 kb. The number of loops in the rosettes varied, since the DNA content of the rosettes was assumed to be that of the metaphase ideogram banding

pattern[78] divided by three to account for the transition into interphase ideogram bands[79]. The starting configuration of a chromosome had the approximate form and size of a metaphase-chromosome whose de-condensation resembles the natural process. Typically ∼400,000 Monte Carlo steps were needed to generate enough statistically independent configurations at thermodynamic equilibrium. For comparison with experimental spatial distance measurements between genetic markers as function of their genetic separation, 100 to 150 statistically independent Monte Carlo configurations were taken as starting points for relaxation at higher spatial resolution by Brownian Dynamics methods using a decreased segment length of 50 nm (∼5.2 kb), corresponding to 20,000 segments for chromosome 15. 2,000 Brownian Dynamics steps were performed until equilibration was reached again. The simulated spatial distances were determined position dependently, i.e. the marker pairs were placed in respect to the topological folding of the simulated chromosome and position independently, i.e. the pairs of markers were placed randomly and therefore regardless of any folding topology on the chromosome (for the underlying assumption see results). All pairs of positions for one genomic separation were taken, the resulting spatial distance determined and averaged over 100 to 150 statistically independent chromosome configurations.

A.7 Trilateration Procedure

The BAC probe was assigned as the origin (0,0,0). Probe h3 was placed on the X-axis with the coordinates as $(d_{0,3}, 0, 0)$, where $d_{0,3}$ is the spatial distance between the BAC and h3. Probe h4 was placed in the XY plane and had the coordinates as $(x_4, y_4, 0)$. The x and y coordinates of the probe h4 were calculated as follows:

$$x_4 = \frac{d_{0,3}^2 + d_{0,4}^2 - d_{3,4}^2}{2d_{0,3}}$$

$$y_4 = \sqrt{d_{0,4}^2 - x_4^2}$$

where $d_{0,4}$ is the distance of h4 from the BAC, and $d_{3,4}$ is the distance of h4 from probe h3. The initial estimates of the x, y and z coordinates of a (non-reference) probe i were calculated as follows:

$$x_i = \frac{d_{i,0}^2 - d_{i,3}^2 + x_3^2}{2x_3}$$

$$y_i = \frac{d_{i,0}^2 - d_{i,4}^2 - 2x_i x_4 + x_4^2 + y_4^2}{2y_4}$$

where $d_{i,0},\ d_{i,3}\ and\ d_{i,4}$ are the 3D-distances of probe i from the reference probes BAC, h3 and h4 respectively.

The error in the estimated position of a probe i was determined as follows:

$$\sum_{\substack{j=0 \\ j \neq i}}^{11} = (r_{i,j} - d_{i,j})^2$$

where $r_{i,j}$ is the distance between the probes i and j based on the 3D-coordinates of these probes, and $d_{i,j}$ is the experimental spatial distance measured for this probe

pair. Only the probe-pairs for which we had an experimental spatial distance measurement were considered for error calculation. Thereafter, we used a multiple steepest descent algorithm to iteratively change the 3D-coordinates in order to obtain the set of 3D-coordinates that substantially reduce the error associated with each probe.